GENETICS AND GENOMICS TO ENHANCE CROP PRODUCTION, TOWARDS FOOD SECURITY

EDITED BY: Ajay Kumar, Arron Carter, Reyazul Rouf Mir, Deepmala Sehgal, Pinky Agarwal and Andrew H. Paterson

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88974-048-2
DOI 10.3389/978-2-88974-048-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GENETICS AND GENOMICS TO ENHANCE CROP PRODUCTION, TOWARDS FOOD SECURITY

Topic Editors:
Ajay Kumar, North Dakota State University, United States
Arron Carter, Washington State University, United States
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology, India
Deepmala Sehgal, International Maize and Wheat Improvement Center (Mexico), Mexico
Pinky Agarwal, National Institute of Plant Genome Research (NIPGR), India
Andrew H. Paterson, University of Georgia, United States

Citation: Kumar, A., Carter, A., Mir, R. R., Sehgal, D., Agarwal, P., Paterson, A. H., eds. (2022). Genetics and Genomics to Enhance Crop Production, Towards Food Security. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-048-2

Table of Contents

07 ***Editorial: Genetics and Genomics to Enhance Crop Production, Towards Food Security***
Ajay Kumar, Reyazul Rouf Mir, Deepmala Sehgal, Pinky Agarwal and Arron Carter

13 ***Sources of Resistance to* Fusarium solani *and Associated Genomic Regions in Common Bean Diversity Panels***
Kimberly Zitnick-Anderson, Atena Oladzadabbasabadi, Shalu Jain, Chryseis Modderman, Juan M. Osorno, Phillip E. McClean and Julie S. Pasche

27 ***Development and Application of SSR Markers Related to Genes Involved in Leaf Adaxial-Abaxial Polarity Establishment in Chinese Cabbage (*Brassica rapa *L. ssp.* pekinensis*)***
Ying Gao, Yin Lu, Xiaoguang Li, Na Li, Xiaomeng Zhang, Xiangjie Su, Daling Feng, Mengyang Liu, Shuxin Xuan, Aixia Gu, Yanhua Wang, Xueping Chen, Jianjun Zhao and Shuxing Shen

38 ***Identification of* Fusarium solani *f. sp.* pisi *(*Fsp*) Responsive Genes in* Pisum sativum**
Bruce A. Williamson-Benavides, Richard M. Sharpe, Grant Nelson, Eliane T. Bodah, Lyndon D. Porter and Amit Dhingra

58 ***Dissecting the Molecular Function of* Triticum aestivum STI *Family Members Under Heat Stress***
Shaloo Meena, Sohini Deb, Harsha Samtani and Paramjit Khurana

76 ***Structural Aspects of DNA Repair and Recombination in Crop Improvement***
Prabha Verma, Reetika Tandon, Gitanjali Yadav and Vineet Gaur

106 ***Approximate Genome-Based Kernel Models for Large Data Sets Including Main Effects and Interactions***
Jaime Cuevas, Osval A. Montesinos-López, J. W. R. Martini, Paulino Pérez-Rodríguez, Morten Lillemo and Jose Crossa

120 ***Breeding, Genetics, and Genomics Approaches for Improving Fusarium Wilt Resistance in Major Grain Legumes***
Uday Chand Jha, Abhishek Bohra, Shailesh Pandey and Swarup Kumar Parida

145 ***Identification and Validation of New Stable QTLs for Grain Weight and Size by Multiple Mapping Models in Common Wheat***
Jiajia Cao, Yaoyao Shang, Dongmei Xu, Kangle Xu, Xinran Cheng, Xu Pan, Xue Liu, Mingli Liu, Chang Gao, Shengnan Yan, Hui Yao, Wei Gao, Jie Lu, Haiping Zhang, Cheng Chang, Xianchun Xia, Shihe Xiao and Chuanxi Ma

159 ***Phytohormone-Mediated Molecular Mechanisms Involving Multiple Genes and QTL Govern Grain Number in Rice***
Priyanka Deveshwar, Ankita Prusty, Shivam Sharma and Akhilesh K. Tyagi

181 ***Analysis and Identification of QTL for Resistance to* Sclerotinia sclerotiorum *in Pea (*Pisum sativum *L.)***
Rahil Ashtari Mahini, Ajay Kumar, Elias M. Elias, Jason D. Fiedler, Lyndon D. Porter and Kevin E. McPhee

195 ***Identification of QTLs/Defense Genes Effective at Seedling Stage Against Prevailing Races of Wheat Stripe Rust in India***
Anjan Kumar Pradhan, Sundeep Kumar, Amit Kumar Singh, Neeraj Budhlakoti, Dwijesh C. Mishra, Divya Chauhan, Shikha Mittal, Monendra Grover, Suneel Kumar, Om P. Gangwar, Subodh Kumar, Arun Gupta, Subhash C. Bhardwaj, Anil Rai and Kuldeep Singh

213 ***Gene Expression Dynamics in Rice Peduncles at the Heading Stage***
Manu Kandpal, Chandrapal Vishwakarma, Kushagra Krishnan, Viswanathan Chinnusamy, Ashwani Pareek, Manoj K. Sharma and Rita Sharma

230 ***Haplotype-Based, Genome-Wide Association Study Reveals Stable Genomic Regions for Grain Yield in CIMMYT Spring Bread Wheat***
Deepmala Sehgal, Suchismita Mondal, Leonardo Crespo-Herrera, Govindan Velu, Philomin Juliana, Julio Huerta-Espino, Sandesh Shrestha, Jesse Poland, Ravi Singh and Susanne Dreisigacker

243 ***Strategies Using Genomic Selection to Increase Genetic Gain in Breeding Programs for Wheat***
Biructawit Bekele Tessema, Huiming Liu, Anders Christian Sørensen, Jeppe Reitan Andersen and Just Jensen

255 ***Genome-Wide Association Mapping and Genomic Prediction of Anther Extrusion in CIMMYT Hybrid Wheat Breeding Program via Modeling Pedigree, Genomic Relationship, and Interaction With the Environment***
Anil Adhikari, Bhoja Raj Basnet, Jose Crossa, Susanne Dreisigacker, Fatima Camarillo, Pradeep Kumar Bhati, Diego Jarquin, Yann Manes and Amir M. H. Ibrahim

268 ***Identification of Allelic Variation in Drought Responsive Dehydrin Gene Based on Sequence Similarity in Chickpea (*Cicer arietinum *L.)***
Tapan Kumar, Neha Tiwari, Chellapilla Bharadwaj, Ashutosh Sarker, Sneha Priya Reddy Pappula, Sarvjeet Singh and Mohar Singh

278 ***Genome-Wide Identification and Capsaicinoid Biosynthesis-Related Expression Analysis of the* R2R3-MYB *Gene Family in* Capsicum annuum *L.***
Jin Wang, Yi Liu, Bingqian Tang, Xiongze Dai, Lingling Xie, Feng Liu and Xuexiao Zou

290 ***Pyramiding of High Grain Weight With Stripe Rust and Leaf Rust Resistance in Elite Indian Wheat Cultivar Using a Combination of Marker Assisted and Phenotypic Selection***
Satinder Kaur, Jaspreet Kaur, G. S. Mavi, Guriqbal Singh Dhillon, Achla Sharma, Rohtas Singh, Urmila Devi and Parveen Chhuneja

299 ***Genome-Wide Association Study of Tan Spot Resistance in a Hexaploid Wheat Collection From Kazakhstan***
Alma Kokhmetova, Deepmala Sehgal, Shaukat Ali, Makpal Atishova, Madina Kumarbayeva, Irina Leonova and Susanne Dreisigacker

314 ***Silencing of an Ubiquitin Ligase Increases Grain Width and Weight in* indica *Rice***
Ankit Verma, Geeta Prakash, Rajeev Ranjan, Akhilesh K. Tyagi and Pinky Agarwal

329 ***New QTLs for Spot Blotch Disease Resistance in Wheat (*Triticum aestivum *L.) Using Genome-Wide Association Mapping***
Vipin Tomar, Daljit Singh, Guriqbal Singh Dhillon, Ravi Prakash Singh, Jesse Poland, Arun Kumar Joshi, Pawan Kumar Singh, Pradeep Kumar Bhati, Suneel Kumar, Mokhlesur Rahman, Budhi Sagar Tiwari and Uttam Kumar

345 ***The Barley* HvWRKY6 *Transcription Factor Is Required for Resistance Against* Pyrenophora teres *f.* teres**
Prabin Tamang, Jonathan K. Richards, Shyam Solanki, Gazala Ameen, Roshan Sharma Poudel, Priyanka Deka, Karl Effertz, Shaun J. Clare, Justin Hegstad, Achintya Bezbaruah, Xuehui Li, Richard D. Horsley, Timothy L. Friesen and Robert S. Brueggeman

362 ***Technological Innovations for Improving Cassava Production in Sub-Saharan Africa***
Edwige Gaby Nkouaya Mbanjo, Ismail Yusuf Rabbi, Morag Elizabeth Ferguson, Siraj Ismail Kayondo, Ng Hwa Eng, Leena Tripathi, Peter Kulakow and Chiedozie Egesi

383 ***QTL Mapping and Identification of Candidate Genes for Heat Tolerance at the Flowering Stage in Rice***
Lei Chen, Qiang Wang, Maoyan Tang, Xiaoli Zhang, Yinghua Pan, Xinghai Yang, Guoqing Gao, Ronghua Lv, Wei Tao, Ligeng Jiang and Tianfeng Liang

396 ***Allelic Diversity, Structural Analysis, and Genome-Wide Association Study (GWAS) for Yield and Related Traits Using Unexplored Common Bean (*Phaseolus vulgaris *L.) Germplasm From Western Himalayas***
Reyazul Rouf Mir, Neeraj Choudhary, Vanya Bawa, Sofora Jan, Bikram Singh, Mohd Ashraf Bhat, Rajneesh Paliwal, Ajay Kumar, Annapurna Chitikineni, Mahendar Thudi and Rajeev Kumar Varshney

413 ***Mapping and Validation of Major Quantitative Trait Loci for Resistance to Northern Corn Leaf Blight Along With the Determination of the Relationship Between Resistances to Multiple Foliar Pathogens of Maize (*Zea mays *L.)***
Hosahally Muddrangappa Ranganatha, Hirenallur Chandappa Lohithaswa and Anand Pandravada

426 ***Genomics Armed With Diversity Leads the Way in* Brassica *Improvement in a Changing Global Environment***
Nur Shuhadah Mohd Saad, Anita A. Severn-Ellis, Aneeta Pradhan, David Edwards and Jacqueline Batley

454 ***Comparative Analyses of Full-Length Transcriptomes Reveal* Gnetum luofuense *Stem Developmental Dynamics***
Chen Hou, Huiming Lian, Yanling Cai, Yingli Wang, Dongcheng Liang and Boxiang He

469 ***Screening and Mapping for Head Blast Resistance in a Panel of CIMMYT and South Asian Bread Wheat Germplasm***
Xinyao He, Philomin Juliana, Muhammad R. Kabir, Krishna K. Roy, Rabiul Islam, Felix Marza, Gary Peterson, Gyanendra P. Singh, Aakash Chawade, Arun K. Joshi, Ravi P. Singh and Pawan K. Singh

479 ***Genome-Wide Association Study on Root Traits Under Different Growing Environments in Wheat (*Triticum aestivum *L.)***
Fengdan Xu, Shulin Chen, Xiwen Yang, Sumei Zhou, Junsen Wang, Ziliang Zhang, Yuan Huang, Miao Song, Jun Zhang, Kehui Zhan and Dexian He

491 ***Identification of Main-Effect and Environmental Interaction QTL and Their Candidate Genes for Drought Tolerance in a Wheat RIL Population Between Two Elite Spring Cultivars***

S. M. Hisam Al Rabbi, Ajay Kumar, Sepehr Mohajeri Naraghi, Suraj Sapkota, Mohammed S. Alamri, Elias M. Elias, Shahryar Kianian, Raed Seetan, Ali Missaoui, Shyam Solanki and Mohamed Mergoum

507 ***Genome-Wide Association Mapping for Yield and Related Traits Under Drought Stressed and Non-stressed Environments in Wheat***

S. M. Hisam A. Rabbi, Ajay Kumar, Sepehr Mohajeri Naraghi, Senay Simsek, Suraj Sapkota, Shyam Solanki, Mohammed S. Alamri, Elias M. Elias, Shahryar Kianian, Ali Missaoui and Mohamed Mergoum

520 ***Draft Genome Assembly and Transcriptome Dataset for European Turnip (*Brassica rapa *L. ssp.* rapifera*), ECD4 Carrying Clubroot Resistance***

Sin-Gi Park, Eonji Noh, SuRyun Choi, Boram Choi, In-Gang Shin, Seung-il Yoo, Dong Jin Lee, Sumin Ji, Hae-Suk Kim, Yoon-Jung Hwang, Jung Sun Kim, Jacqueline Batley, Yong Pyo Lim, David Edwards and Chang Pyo Hong

528 ***Identification of Root Rot Resistance QTLs in Pea Using* Fusarium solani *f. sp.* pisi*-Responsive Differentially Expressed Genes***

Bruce A. Williamson-Benavides, Richard M. Sharpe, Grant Nelson, Eliane T. Bodah, Lyndon D. Porter and Amit Dhingra

538 ***Molecular Characterization and Expression Analysis of the Na^+/H^+ Exchanger Gene Family in* Capsicum annuum *L.***

Xirong Luo, Shimei Yang, Yong Luo, Huarong Qiu, Tangyan Li, Jing Li, Xiaocui Chen, Xue Zheng, Yongdui Chen, Jie Zhang, Zhongkai Zhang and Cheng Qin

frontiers in Genetics

EDITORIAL
published: 23 November 2021
doi: 10.3389/fgene.2021.798308

Editorial: Genetics and Genomics to Enhance Crop Production, Towards Food Security

Ajay Kumar[1]*, *Reyazul Rouf Mir*[2], *Deepmala Sehgal*[3], *Pinky Agarwal*[4] and *Arron Carter*[5]

[1]Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [2]Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir (SKUAST-K), Sopore, India, [3]International Maize and Wheat Improvement Center (Mexico), Texcoco, Mexico, [4]National Institute of Plant Genome Research (NIPGR), New Delhi, India, [5]Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States

Keywords: abiotic stress, disease resistance, candidate genes, association mapping, QTL mapping, genomic selection, gene editing, crop improvement

Editorial on the Research Topic

Genetics and Genomics to Enhance Crop Production, Towards Food Security

Twenty first century agriculture faces many challenges including new emerging abiotic and biotic stresses and decreasing arable land. These challenges pose serious threats to food security of an ever-increasing world population. One of the solutions to meet the food demands is to develop high-yielding crop varieties with greater genetic potential and resistance/tolerance to both biotic and abiotic stresses. Just like in mid-nineteenth century, when new genes and methods resulted in the first green revolution, there is a need to combine traditional plant breeding tools with new technologies to bring another green revolution for future food security.

In the last 3 decades, many new genomics tools and technologies (molecular markers, high density marker genotyping platforms, sequencing technologies, genome mapping methodologies, gene editing) have been developed, which have increased our knowledge and capabilities to breed improved varieties in a shorter period of time (Mir et al., 2013; Kumar et al., 2021; Sihag et al., 2021; Tyagi et al., 2021). The advances in genomics, statistical tools, bioinformatics, phenomics, data science and many other related disciplines are helping us achieve desirable genetic gains in crops, in a fast but sustainable manner (Mir et al., 2012; Mir et al., 2019; Salsman et al., 2018). To increase agricultural productivity to meet food demands, there is a need to combine traditional plant breeding tools and techniques with modern technologies like marker assisted/genomic selection and gene editing (e.g., CRISPR-Cas). Therefore, the focus of this topic was to bring together the research and knowledge of application of these modern tools for crop improvement.

This topic includes a total of 35 articles including 30 research articles and five review articles. The articles in this topic focus on a variety of new tools and technologies to develop knowledge and resources to breed high yielding, climate resilient crop varieties. This topic includes articles on genetic dissection of important traits related to yield, biotic and abiotic stress resistance, development of genomic tools to improve genomic predictions, cloning and deployment of important genes, and development of resources for enhancing crop production.

One of the major challenges in crop production is abiotic stresses. Among the abiotic-stresses, heat stress is considered one of the most important factors affecting yield in most crops, including cereals and pulses. This special topic includes several articles on focused on developing knowledge and genomic resources for breeding varieties tolerant to abiotic stresses (mainly drought tolerance) in important crops. In two different studies, Rabbi et al. used both genome-wide association mapping and quantitative trait locus (QTL) mapping combined with Infinium 90 K single-nucleotide polymorphism (SNP) assay to dissect the genetics of drought tolerance in hard red spring wheat

OPEN ACCESS

Edited and reviewed by:
Zefeng Yang,
Yangzhou University, China

***Correspondence:**
Ajay Kumar
ajayguptajammu@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *20 October 2021*
Accepted: *04 November 2021*
Published: *23 November 2021*

Citation:
Kumar A, Mir RR, Sehgal D, Agarwal P and Carter A (2021) Editorial: Genetics and Genomics to Enhance Crop Production, Towards Food Security. Front. Genet. 12:798308. doi: 10.3389/fgene.2021.798308

(*Triticum aestivum* L.) in northern United States. In one of the studies, they evaluated 361 wheat genotypes, which were evaluated in nine different locations for yield and related traits under rain-fed conditions. Association mapping using mixed linear model identified a total of 69 consistent QTL with $p \leq 0.001$. Six potential novel QTL were identified on chromosomes 3D, 4A, 5B, 7A, and 7B. The resources developed in this study could be used in marker-assisted selection for drought-tolerance breeding in spring wheat. In another study, Rabbi et al. used a wheat RIL population derived from a cross between a drought-tolerant cultivar and a high-yielding but drought-susceptible cultivar. They identified a total of 11 consistent QTL for drought tolerance-related traits; six QTL being exclusively identified in drought-prone environments, and five constitutive QTL (identified under both drought and optimal conditions). Major QTL expressed exclusively under drought environments were identified on chromosomes 7B and 2B, while a novel QTL for drought tolerance was identified on chromosome 2D. The authors also conducted *in silico* expression analysis of candidate genes located in the QTL regions. They identified the enrichment of ribosomal and chloroplast photosynthesis-associated proteins showing the most expression variability, thus possibly contributing to stress response. The significant QTL identified in this study will play an important role in the development of drought-tolerant wheat cultivars using genomics-assisted breeding approaches.

In rice (*Oryza sativa* L.), Chen et al. identified a new major QTL for heat tolerance during the flowering stage using an $F_{2:3}$ population. The QTL *qHTT8* was localized to chromosome 8 of rice and contained 65 predicted genes, 10 of which were found to be associated with abiotic stress tolerance. Using qRT-PCR, the differential expression of these 10 putative genes under high temperature conditions was conducted, and it was found that *LOC_Os08g07010* and *LOC_Os08g07440* were highly induced in the heat tolerant cultivar. These results will be used to develop heat tolerant rice cultivars, as well as future efforts to clone *qHTT8* to understand its functionality.

Chickpea (*Cicer arietinum* L.) is one of the most important legumes crops in the world and this crop is mainly grown in rainfed conditions under residual moisture and restricted irrigation. The comprehensive study by Kumar et al. used comparative study of a candidate gene, molecular marker, and physiological traits for screening drought tolerance in chickpea. The abiotic stress-responsive gene "Dehydrin (DHN)" was also identified based on a sequence similarity approach. Most promising drought tolerant genotypes (Pusa1103, ICC4958 and Pusa362) were identified in this study and will prove useful in future chickpea breeding programs aimed at enhancing drought tolerance. In another research article by Meena et al., a set of six members of the *TaSTI* family were identified and among them "TaSTI-2" members were found to express higher as compared to "TaSTI-6" members under heat stress conditions, with *TaSTI-2A* being the most heat responsive member. They showed for the first time that TaSTI-2A interacts with TaHSP90 not only in the nucleus but also in the endoplasmic reticulum and Golgi bodies.

Several articles also focused on genetic dissection of disease resistance and development of genomic resources for molecular breeding. One of the articles focused on identification of novel QTL for spot blotch caused by *Bipolaris sorokiniana,* which is one of the major diseases of wheat under tropical and subtropical conditions (Tomar et al.). Using 139 advanced spring wheat lines and 14,063 polymorphic genotyping-by-sequencing (GBS) markers, Tomar et al. identified eight QTLs associated with spot blotch disease resistance in wheat. The authors also conducted functional annotation of the significant markers, which identified NBS-LRR, MADS-box transcription factor, and 34 other plant-related protein families distributed across the genome. This study identified four new QTLs on chromosomes 1A (497.2 Mb), 1D (89.84 Mb), 2B (421.92 Mb), and 6D (6.84 Mb) associated with several disease resistance protein families. The results of this study could be vital for spot blotch resistance breeding in wheat.

With an objective to transfer white mold resistance (caused by *Sclerotinia sclerotiorum*) into an adapted pea (*Pisium sativum* L.) background, Mahini et al. developed and evaluated two RIL populations (Lifter × PI240515 and PI169603 × Medora) for resistance to white mold by measuring lesion expansion inhibition (LEI) and nodal transmission inhibition (NTI). This study identified a total of 22 short genotypes with partial resistance. Using GBS-derived linkage maps and inclusive composite interval mapping, they identified a total of thirteen QTL associated with white mold resistance traits in both populations. Three of them were co-located with height genes and the other ten QTL were associated with LEI (7) and NTI (3). The resistant lines and genetic resources developed in this study will help accelerate the development of white mold resistance cultivars using molecular breeding approaches.

Common bean (*Phaseolus vulgaris* L.) is considered one of the principal grain legume crops of western-Himalayas of Jammu and Kashmir (Choudhary et al., 2018a; Choudhary et al., 2018b). Huge diversity exists in common bean germplasm from this region. A core set of 96 genotypes selected from a set of 428 genotypes were selected for study of allelic diversity, structural analysis, and genome wide association study (GWAS) for yield and related traits using unexplored common bean germplasm from the western Himalayas (Mir et al.). The study of allelic diversity led to the identification of 691 alleles ranging from 2 to 21 with an average of 7.59 alleles/locus. GWAS for pods/plant, seeds/pod, seed weight, and yield/plant led to the identification of 39 marker-trait associations (MTAs) including 15 major, 15 stable, and 13 both major and stable MTAs. These MTAs will prove useful for common bean breeding programs world-wide.

Breeding disease resistant wheat varieties is one of the most important subject areas of wheat research and wheat blast (*Magnaporthe oryzae* pathotype *Triticum*) is considered one of the most destructive diseases in South American and South Asian countries. GWAS was conducted using a diverse panel of 184 wheat genotypes from South Asia and CIMMYT and genotyping data of 11,401 SNP markers of the Illumina Infinium 15 K BeadChip (He et al.). This study led to the identification of MTAs on chromosomes 1BS, 2AS, 6BS, and 7BL. The most promising resistant accessions/genotypes identified in this study will prove useful as a preemptive strategy to prevent wheat blast outbreaks in South American and South Asian countries.

Breeding for wheat stripe (yellow) rust (*Puccinia striiformis* f. sp. *tritici*) is one of the most important subject areas of wheat research. Therefore, efforts have been made to characterize wheat germplasm for novel sources of resistance and their incorporation into elite cultivars. GWAS was conducted using a set of 391 genotypes and 35 K Axiom® array with the aim to identify QTL for stripe rust resistance at the seedling stage (Pradhan et al.). A total of 40 QTLs were identified, of which 20 QTLs were found to be closely linked with previously reported stripe rust resistance genes/QTLs on chromosomes 1B, 2B, 5B and 6B, whereas the 20 novel QTLs were mapped on chromosomes 2D, 3A, 3D, 5A and 7D. These 20 novel QTL identified in the present study might play a key role in marker-assisted breeding for developing stripe rust resistant wheat cultivars (Pradhan et al.).

Zitnick-Anderson et al. conducted a GWAS on the Andean and Middle American diversity panels of common bean to determine the genetic basis of resistance to Fusarium root rot (*Fusarium solani*) using a greenhouse screening assay developed by them. Important associations were identified in both panels harboring important candidate genes with proven roles in plant defense including Glucan synthase-like protein, NAC domain protein, senescence-associated genes, MAC/Perforin domain-containing gene, and ethylene response factor 1. Most importantly, it was discovered that some common genetic factors might be present to Pv02, Pv09 and Pv11 in the Andean gene pool and to Pv01 and Pv08 in the Middle American gene pool that might be involved in the resistance of both Rhizoctonia and Fusarium root rot in the common bean.

Ranganatha et al. conducted QTL mapping to determine genomic regions controlling resistance to northern corn leaf blight (*Exserohilum turcicum*) in maize (*Zea mays* L.) and identified a large effect QTL on chromosome 8 explaining the highest percentage variation of 16.34%. QTL for two more foliar diseases [sorghum downy mildew (*Peronosclerospora sorghi*) and southern corn rust (*Puccinia polysora*)] were also identified on chromosomes 2 and 10 in order to generate pyramided lines expressing QTL for resistance to all three foliar diseases. Markers linked to these major QTL were deployed in marker-assisted selection and 125 pyramided lines were developed, of which 39 lines exhibited an acceptable level of resistance to the three major diseases.

The net blotch of barley (*Hordium vulgare* L.) caused by the necrotrophic fungal pathogen *Pyrenophora teres* f. *teres* (*Ptt*) has become a serious threat to barley production globally. Tamang et al. employed a mutagenesis approach to identify resistance genes present in barley line CI_{5791}. Seeds of CI_{5791} were irradiated with γ-rays and an M_2 population was generated from which two mutants, $CI_{5791\text{-}\gamma3}$ and $CI_{5791\text{-}\gamma8}$, were sequenced by exome capture. Comparison of their sequences with wild-type CI_{5791} identified independent mutations in the *HvWRKY6* transcription factor located on chromosome 3H, which suggested *HvWRKY6* as a strong candidate gene. Gene silencing experiments conducted in barley line CI_{5791} confirmed the hypothesis that *HvWRKY6* is responsible for imparting resistance to net blotch.

Pyrenophora tritici-repentis, causing tan spot disease in wheat, is a serious threat to wheat production in Kazakhstan. Kokhmetova et al. conducted a GWAS in a collection of spring and winter wheat lines originating from Kazakhstan, Russia, and CIMMYT breeding programs to determine the genetic basis of resistance to tan spot. The most significant finding of this study was the identification of a 3 Mb genomic region on chromosome 6A, where SNPs associated with resistance to the two most prevalent races, Race 1 and Race 5, were identified. This study has opened vistas to initiate molecular marker-assisted breeding activities in Kazakhstan and twenty-five lines carrying novel combinations are being used in the country's breeding programs to improve resistance against both races.

Because of importance of plant roots for water and nutrient acquisition, environmental adaptation, and yield formation, Xu et al. used a total of 196 wheat accessions from the Huang-Huai Wheat Region of China to investigate six root traits. This study showed that the root traits varied most under outdoor pot culture, followed by those under both outdoor hydroponic culture and indoor hydroponic culture, and root elongation under hydroponic culture was faster than that under pot culture. A GWAS using the Wheat 660 K SNP Array identified 12 stable chromosomal regions associated with root traits located on chromosomes 1D, 2A, 4A, 4B, 5B, 6D, and unmapped markers. Linkage disequilibrium blocks identified 27 genes related to root development including *TraesCS4A02G484200, TraesCS4A02G484800*, *TraesCS4A02G493800*, and *TraesCS4A02G493900*, which were involved in cell elongation and differentiation.

Three important articles in this topic focused on development of genomic tools for yield and yield components in wheat. Sehgal et al. reported a comprehensive GWAS and epistatic interactions analysis to investigate the genetic architecture of grain yield (GY) using a large set of 6,461 advanced breeding lines from the CIMMYT breeding program. The authors used a linkage disequilibrium approach to generate 519 genome-wide haplotype blocks and used them in combination with GY data generated under irrigated and stress environments to conduct a haplotypes-based GWAS and analysis of epistatic interactions. In addition to seven hotspots regions reported on chromosomes 1A, 1B, 2B, 4A, 5, 6, and 7B controlling GY, the authors unveiled that a large proportion of the variation was explained by epistatic interactions. To utilize these results for breeding application, the authors constructed heat maps and identified sets of lines having multiple QTL to be included in crossing schemes.

Cao et al. conducted QTL mapping for thousand grain weight (TGW), grain width (GW), and grain length (GL) in wheat using a RIL population derived from the cross between Jing 411 and Hongmangchun 21. The authors not only identified five new stable QTLs for TGW and GL on chromosomes 1B, 2A, 4B, and 7A, but also validated their effects in an independent Chinese mini core collection (CMCC). In addition, *in silico* analysis of the stable QTL in the IWGSC annotation database predicted important candidate genes including *TraesCS4B02G376400* encoding a plasma membrane H^+-ATPase. Investigation of allelic variation in this gene in the CMCC showed significant association of a SNP with TGW, GW and GL.

In order to better understand hybrid wheat seed production, Adhikari et al. performed GWAS and genomic prediction of anther extrusion in the CIMMYT hybrid wheat breeding

program. A set of 603 male lines were phenotyped across three field experiments and genotyped with the 20 K Infinitum iselect SNP Array. GWAS produced five MTA, each with only a small effect. Genomic prediction involved the main effects of lines, environments, genomic and pedigree relationship, as well as their interactions in different statistical models, which were evaluated in three cross-validation scenarios. Genomic prediction appeared to be a very reliable tool for predicting anther extrusion in hybrid wheat breeding, and could be used in selection across breeding cycles, aiding in rapid breeding for this trait.

This research topic Research Topic also includes a few articles in vegetable crops, like Chinese cabbage (*Brassica rapa* L. spp. *Pekinensis*), *Capsicum*, and European turnip (*Brassica rapa* L. ssp. *rapifera*). An important trait in Chinese cabbage is leaf adaxial-abaxial polarity as it relates to the formation of leafy heads. Therefore, Gao et al. identified genes responsible for this trait, as well as their genetic variation, in order to further understand the mechanism of leafy head formation. A total of 41 candidate genes were identified through comparisons with *Arabidopsis thaliana*, which lead to further understanding of gene structure. Based on these 41 candidate genes, 341 simple sequence repeats were detected, and 323 loci were used to design 969 specific primers. A subset of these primers were selected randomly and evaluated using 12 Chinese cabbage accessions with different heading types. A total of 23 new primer pairs were combined with 127 other markers to construct a linkage map from an F_2 population of two cabbage parents with different heading types. These new markers will continue to assist in the understanding of traits associated with the formation of leafy heads in Chinese cabbage as well as assist in the breeding and selection of this trait.

The article by Luo et al. reported a systematic identification and expression-characteristic analysis of *CaNHX* genes at the whole genome level in cultivated and wild pepper (*Capsicum annuum* L.). The study identified a total of 42 CaNHX genes, with nine having complete functional domains of the Na^+/H^+ exchanger gene. The authors also conducted transcriptome analysis which showed that *CaNHX* genes were upregulated under various abiotic stresses. The Na^+/H^+ exchangers (NHXs) play an important role in plant growth and abiotic stress tolerance. Therefore, the findings of this study could play a vital role in breeding stress tolerant pepper.

Park et al. used single-molecule long-read sequencing technology to generate a draft genome assembly of a European turnip, ECD4. This line has strong clubroot resistance, a disease caused by *Plasmodiophora brassicae*. The authors believe that the draft assembly and transcriptome data developed in this study will help the development of genetic and genomic resources for *Brassica* improvement, including the identification of clubroot and other disease resistance genes (*R* genes), the study of diversity and evolution, and markers for genomics assisted breeding in *Brassica*.

Few articles in this topic also focused on transcriptome analyses to identify important candidate genes. In *japonica* rice, *GW2* is a previously identified QTL known as being a negative regulator of grain width. The related SNP is absent in *indica* rice. Verma et al. knocked down *GW2* in *indica* rice and reported increases in grain width, length and weight. This was because of a decrease in cell size but an increase in cell number. The phenotype was stable across three generations of seeds. The starch granules in the endosperm were bigger in size, tightly packed, and showed an increase in polyhedral granules in comparison to the wild type. Transcriptome analyses of seeds from a homozygous line and its comparison with wild type seeds showed 1,426 differentially expressed genes (DEGs), of which 55 have been previously characterized to be involved in rice seed development. Comparison of the DEGs from this analysis with DEGs from seed development transcriptome resulted in 23 seed specific DEGs which were commonly upregulated in both tissues. This implies that *GW2* works upstream of these seed-related genes and regulates their expression.

Elongation of rice peduncles prior to heading determines panicle exsertion, which in turn is directly related to panicle fertility, and hence, yield. Kandpal et al. performed transcriptome analyses on two Indian cultivars contrasting for peduncle elongation. Tissue used for the analyses were elongating peduncles (EP) and non-elongating peduncles (NP) from Swarna (with lesser peduncle elongation) and Pokkali (with higher peduncle elongation). EP of both cultivars had 1,500 common up regulated genes while NP had 1723 genes. Genes differentially expressed in EP coded for cellulase synthases, sucrose synthases, invertases, sucrose transporters, auxin and brassinosteroid biosynthesis and signaling, and 836 transcription factors. Also, 50 genes that were up-regulated in EP were also involved in pollen formation. The authors conclude that the data generated in their study can be used further for improving peduncle elongation in rice.

Wang et al. reported genome-wide identification and capsaicinoid biosynthesis-related expression analysis of the R2R3-MYB gene family in *Capsicum annuum*. Capsaicinoids are naturally specialized metabolites and main factor causing *Capsicum* pungency. The MYB transcription factors (R2R3-MYB subfamily) are key regulators and candidate factors in capsaicinoid biosynthesis. The authors identified a set of 108 R2R3-MYB genes and conducted interspecies collinearity analysis which revealed that the R2R3-MYB family contains 16 duplicated gene pairs. The highest gene density was observed on chromosome 00 and 03. The *CaR2R3-MYB* genes were differentially expressed, and capsaicinoid-biosynthetic genes (CBGs) in fruit development stages were obtained via RNA-seq and qRT-PCR. The study further revealed that a set of six candidate *CaR2R3-MYB* genes are involved in regulating the synthesis of capsaicin and dihydrocapsaicin.

Hou et al. also performed transcriptome analyses to understand an industrially important species, *Gnetum luofuense*. The tissues used were four developmental stages (GLN01-04) of the stem apex from female and male plants. Amongst the expressed genes, the terms ribosome, spliceosome and carbon metabolism showed the highest hits in the KEGG database. In conjunction, 24,151 alternative splicing events were detected, in which intron retention was the highest. Also, 38,108 complete open-reading frames and 728 lncRNAs were identified. Transcription factors and their putative pathways were also identified from the transcriptome

data. Using K-means clustering, differentially expressed transcripts amongst the four tissues were delineated. Most differentially expressed transcripts were present between GLN01 and GLN04, which belonged to phenylpropanoid biosynthesis, starch and sucrose metabolism, and cyanoamino acid metabolism. The data generated in this study is a compendium of transcriptome changes in stem apex development in *Gnetum*.

Williamson-Benavides et al. have identified differentially expressed genes in *Pisum sativum* in response to *Fusarium solani* infection, which causes root rot. The tissues used were four tolerant and four susceptible pea genotypes, in a time-course experiment (0, 6 and 12 h). In response to *Fsp* infection, 42,905 differentially expressed contigs (DECs) were identified. The susceptible genotypes had more DECs upon infection. As with other transcriptome analyses, all DECs were subject to functional annotation and GO enrichment. The study concluded that DECs could be categorized into seven broad categories including transcriptional regulation, pathogenesis-related, anthocyanin and lignin biosynthesis, phytohormone, cell wall and membrane metabolism, and toxin metabolism. The analyses highlight genes which can be targeted to tackle pea root rot. Further, in another study, Williamson-Benavides et al. delineated QTLs for pea root rot resistance from the DEGs identified in the above study. Two RIL populations (190 individuals each) were used to identify QTLs using composite interval mapping. This resulted in 47 SNP markers. Out of these, five were associated with root disease severity and height. This data was overlaid with the transcriptome data mentioned above. Eventually, seven, 17, 22, 14, and 11 DEGs were identified in different genotypes to be associated with disease response. The study identified a major and four minor QTLs, using a combination of transcriptome analyses and QTL mapping. The underlying genes can further be used for imparting root rot resistance in pea.

Following discovery of genes and QTLs, the next important step in crop improvement is their validation and transfer into different genetic backgrounds and development of improved varieties. In this direction, the study by Kaur et al. pyramided genes of grain weight, stripe rust, and leaf rust (*Puccinia triticina*) into elite Indian bread wheat cultivars (PBW343 and PBW550) using both marker-assisted selection and phenotypic selection. The results showed an increase in grain weight of 8.82–10.27 g in PBW343 and 5.27–7.6 g in PBW550 through conventional backcross breeding and phenotypic selections. The improved genotypes in the PBW550 background possess increased grain weight ranging from 45.0 to 46.2 g and three stacked genes for stripe and leaf rust resistance, which is a valuable resource for breeding elite cultivars.

Traditional marker assisted selection incorporates few markers in the selection scheme and is mostly limited to selection of few genes/QTL with major effects. However, the majority of important traits like yield and abiotic stresses have complex genetic architecture and are controlled by many genes. To expedite the genetic gain for such traits, the concept of whole genome prediction based selection (or genomic selection) is gaining momentum. The study by Tessema et al. attempted to quantify the increase in genetic gain by implementing genomic selection in traditional wheat-breeding programs. They studied the effect of genetic correlation between different traits on genetic gain. The study reported three times higher genetic gain using genomic estimated breeding values compared to phenotypic selection, suggesting the importance of genomic selection for improving genetic gain in wheat-breeding programs. This study showed increased genetic gain when genomic information is incorporated in a conventional breeding program.

The article by Cuevas et al. focused on improving genomic prediction models. Genomic prediction and selection are becoming important tools in both animal and plant breeding. However, these genomic tools encounter several computational challenges when the number of observations (n) is large. Inclusion of genomic × environment interaction and multi-trait kernels in the prediction model further complicates the computational capabilities. The study by Cuevas et al. proposed selecting a small number of lines m ($m < n$) for constructing an approximate kernel of lower rank than the original. Doing so exponentially decreased the required computing time. When the authors applied the proposed method to two different wheat data sets of different sizes (n) using the standard linear kernel Genomic Best Linear Unbiased Predictor (GBLUP) and also using eigen value decomposition, results showed a competitive prediction performance of the approximated methods with a significant reduction in computing time.

This Research Topic also included few review articles focused on application of modern phenomic and genomic based tools for addressing challenges to crop production. For examples, Jha et al. reviewed the progress made in developing genetic and genomic resources for Fusarium wilt (Fusarium oxysporum), whichis the key constraint to grain legume production worldwide. This review article provides a brief recap of classical genetic efforts, as well as the latest technological breakthroughs that have enhanced our understanding of the genetic basis of both plant resistance and pathogenicity. This article presented several examples where modern functional genomic tools like RNA-seq, proteomics, and metabolomics are playing a greater role in illuminating the various aspects of plant-pathogen interaction. At the end, they discussed future prospect for breeding for *Fusarium* wild resistance in grain legumes including the application of modern high throughput phenomics and genomics-based breeding techniques to develop *Fusarium* wild resistant grain legumes.

Grain number is one of the major yield components in cereals and thus important to enhance grain yield. In rice, grain number has been found associated with shoot branching (tillering), panicle branching, panicle length, and seed set percentage. It has also been shown that all these traits are controlled by phytohormones by regulating molecular factors and their cross-interactions. This review by Deveshwar et al. focuses on the molecular machinery, involving several genes and QTL, operational in the plant that governs hormonal control and, in turn, gets governed by the hormones to regulate grain number and yield in rice.

An important review on the development of genetic and genomic resources for molecular breeding and technological advances, especially in the area of genome editing, in cassava (*Manihot esculenta* Crantz) was compiled by Mbanjo et al. The

authors also reviewed various phenotyping protocols that have been used to phenotype cassava germplasm collections for important agronomic traits including nutritional traits (carotenoids, cyanogenic potential), dry matter content (substitute for yield), quality traits, root architecture, and resistance to multiple diseases (cassava mosaic virus, brown streak disease, bacterial blight, root rot). An update was also provided on how outputs of gene discovery and genomic selection studies have been practically utilized in crop improvement.

Saad et al. reviewed the literature for genomics approaches in *Brassica* improvement in a changing global environment. The article reviewed the literature on genetic diversity, genomic resources in *Brassica*, nuclear genome, organelle genomes, marker discovery in *Brassica*, epi-genomics, reverse genetics, genome manipulation, genome editing, transcriptomics, transformation, and genomic selection in crop improvement.

The article by Verma et al. reviews the importance of studying plant DNA repair and recombination (DRR) for crop improvement. The knowledge about plant DRR could help design new strategies for further crop improvement to meet the demand of billions of humans in a sustainable environment. In plants, there are different DRR mechanisms including direct repair, nucleotide excision repair, base excision repair, mismatch repair, non-homologous end joining, and homologous recombination, which interact to continuously repairs DNA damages. In this review, the authors provide an overview of different DRR pathways, their structural and biochemical aspects, and their potential for crop improvement.

In summary, this topic comprises of a diverse collection of both original research and review articles. These articles generated a large amount of genetic and genomic resources in different crops including major cereals (wheat, rice, maize, barley), legumes (chickpea, pea, common bean), and vegetable crops (Chinese cabbage, *Capsicum*, European turnip). The articles in this topic reported the application of many modern genomic tools for crop improvement, like QTL mapping, GWAS, high density marker platforms in wheat (Infinium 90 k and 35 K Axiom® array), genotype by sequencing, marker assisted and genomic selection, single-molecule long-read sequencing technology, transcriptome analyses, candidate gene analyses, etc. We believe that the resources and knowledge generated through articles published in this topic will help expediting crop improvement, thus playing an important role in increasing crop production for future food security.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Choudhary, N., Bawa, V., Paliwal, R., Singh, B., Bhat, M. A., Mir, J. I., et al. (2018b). Gene/QTL Discovery for Anthracnose in Common Bean (*Phaseolus vulgaris* L.) from North-western Himalayas. *Plos One* 13, e0191700. doi:10.1371/journal.pone.0191700

Choudhary, N., Hamid, A., Singh, B., Khandy, I., Sofi, P. A., Bhat, M. A., et al. (2018a). Insight into the Origin of Common Bean (*Phaseolus vulgaris* L.) Grown in the State of Jammu and Kashmir of north-western Himalayas. *Genet. Resour. Crop Evol.* 65, 963–977. doi:10.1007/s10722-017-0588-z

Kumar, S., Kumar, M., Mir, R. R., Kumar, R., and Kumar, S. (2021). "Advances in Molecular Markers and Their Use in Genetic Improvement of Wheat," in *Physiological, Molecular, and Genetic Perspectives of Wheat Improvement*. Editors S. H. Wani, A. Mohan, and G. P. Singh (Cham, Switzerland: Springer Nature), 139–174. doi:10.1007/978-3-030-59577-7_8

Mir, R. R., Hiremath, P. J., Riera-Lizarazu, O., and Varshney, R. K. (2013). "Evolving Molecular Marker Technologies in Plants: from RFLPs to GBS," in *Diagnostics in Plant Breeding*. Editors T. Lübberstedt and R. K. Varshney (New York: Springer), 229–247. doi:10.1007/978-94-007-5687-8_11

Mir, R. R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R., and Varshney, R. K. (2012). Integrated Genomics, Physiology and Breeding Approaches for Improving Drought Tolerance in Crops. *Theor. Appl. Genet.* 125, 625–645. doi:10.1007/s00122-012-1904-9

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-Throughput Phenotyping for Crop Improvement in the Genomics Era. *Plant Sci.* 282, 60–72.

Salsman, E., Kumar, A., AbuHammad, W., Oladzad-Abbasabadi, A., Dobrydina, M., Chao, S., et al. (2018). Development and Validation of Molecular Markers for Grain Cadmium in Durum Wheat. *Mol. Breed.* 38, 28. doi:10.1007/s11032-018-0788-z

Sihag, P., Sagwal, V., Kumar, A., Balyan, P., Mir, R. R., Dhankher, O. P., et al. (2021). Discovery of miRNAs and Development of Heat-Responsive miRNA-SSR Markers for Characterization of Wheat Germplasm for Terminal Heat Tolerance Breeding. *Front. Genet.* 12, 699420. doi:10.3389/fgene.2021.699420

Tyagi, S., Kumar, A., Gautam, T., Pandey, R., Rustgi, S., and Mir, R. R. (2021). Development and Use of miRNA-Derived SSR Markers for the Study of Genetic Diversity, Population Structure, and Characterization of Genotypes for Breeding Heat Tolerant Wheat Varieties. *PLoS One* 16, e0231063. doi:10.1371/journal.pone.0231063

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kumar, Mir, Sehgal, Agarwal and Carter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 June 2020
doi: 10.3389/fgene.2020.00475

Sources of Resistance to *Fusarium solani* and Associated Genomic Regions in Common Bean Diversity Panels

Kimberly Zitnick-Anderson[1], Atena Oladzadabbasabadi[2], Shalu Jain[3], Chryseis Modderman[4], Juan M. Osorno[2], Phillip E. McClean[2] and Julie S. Pasche[5]**

[1] Department of Microbiology, North Dakota State University, Fargo, ND, United States, [2] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [3] Department of Pathology and Entomology, Syngenta, Stanton, MN, United States, [4] Department of Soil, Water, and Climate, University of Minnesota, Morris, Morris, MN, United States, [5] Department of Plant Pathology, North Dakota State University, Fargo, ND, United States

OPEN ACCESS

Edited by:
Deepmala Sehgal, International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Jitendra Kumar, University of Minnesota Twin Cities, United States
Dev Paudel, University of Florida, United States

***Correspondence:**
Shalu Jain
shalu.jain@syngenta.com
Julie S. Pasche
Julie.Pasche@NDSU.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 07 February 2020
Accepted: 16 April 2020
Published: 16 June 2020

Citation:
Zitnick-Anderson K, Oladzadabbasabadi A, Jain S, Modderman C, Osorno JM, McClean PE and Pasche JS (2020) Sources of Resistance to Fusarium solani and Associated Genomic Regions in Common Bean Diversity Panels. Front. Genet. 11:475. doi: 10.3389/fgene.2020.00475

Common bean (*Phaseolus vulgaris* L.) production worldwide is hampered by Fusarium root rot (FRR), which is caused by *Fusarium solani*. Screening for FRR resistance on a large scale is notoriously difficult and often yields inconsistent results due to variability within the environment and pathogen biology. A greenhouse screening assay was developed incorporating multiple isolates of *F. solani* to improve assay reproducibility. The Andean (ADP; n = 270) and Middle American (MDP; n = 280) Diversity Panels were screened in the greenhouse to identify genetic factors associated with FRR resistance. Forty-seven MDP and 34 ADP lines from multiple market classes were identified as resistant to FRR. Greenhouse phenotyping repeatability was confirmed via five control lines. Genome-wide association mapping using ∼200k SNPs was performed on standard phenotyping score 1–9, as well as binary and polynomial transformation of score data. Sixteen and seven significant genomic regions were identified for ADP and MDP, respectively, using all three classes of phenotypic data. Most candidate genes were associated with plant immune/defense mechanisms. For the ADP population, ortholog of glucan synthase-like enzyme, senescence-associated genes, and NAC domain protein, associated with peak genomic region Pv08:0.04–0.18 Mbp, were the most significant candidate genes. For the MDP population, the peak SNPs Pv07:15.29 Mbp and Pv01:51 Mbp mapped within gene models associated with ethylene response factor 1 and MAC/Perforin domain-containing gene respectively. The research provides a basis for bean improvement through the use of resistant genotypes and genomic regions for more durable root rot resistance.

Keywords: *Phaseolus vulgaris*, GWAS, quantitative resistance, Fusarium, root rot

INTRODUCTION

Fusarium root rot (FRR; caused by *Fusarium solani* [Mart.] Sacc. f. sp. *phaseoli* [Burk.] W.C. Snyder & H.M. Hans) is one of the most prevalent soilborne diseases in bean-growing regions of the United States (Coleman, 2016). Originally described as one species, *F. solani* recently has been recognized as the *Fusarium solani* species complex (FSSC). FSSC is comprised of up to 60

phylogenetically distinct species divided into 10 *formae speciales* (f. sp.) based on host specificity (Matuo and Snyder, 1973). However, previous studies that used the *F. solani* f. sp. designation were pathogenic on other hosts (VanEtten, 1978). Currently, *F. solani* species are characterized by DNA sequences and multilocus haplotypes, rather than the f. sp. classification method, and grouped together into three FSSC clades (O'Donnell, 2000; O'Donnell et al., 2008). Among the multiple species that cause root rot on common bean, *F. solani* has been documented as the most damaging root rot pathogen (Coleman, 2016). Symptoms of FRR on common bean manifest as dark brown to rust colored sunken lesions where lateral roots begin to rot (Abawi, 1989). Lesions on the lower hypocotyl coalesce as the disease progresses and results in complete rot of the root system (Abawi, 1989). When left unmitigated, FRR can cause up to 84% yield loss (Schneider et al., 2001).

Managing FRR can be difficult due to the durability and extended viability of chlamydospores in soil and plant debris (Katan, 2017). Current management strategies include the use of seed treatment chemicals, avoiding infested fields, crop rotation, and planting certified seeds. However, the most sustainable and durable approaches for controlling the disease is genetic resistance (Rubiales et al., 2015). While foliar disease resistance is a target for crop improvement, less emphasis has been given to breeding for root rot resistance in common bean and there are fewer sources of root rot resistance available. Although commercial cultivars are known to have limited FRR resistance, multiple studies have characterized and identified sources of resistance within common bean germplasm collections (Román-Avilés and Kelly, 2005; Bilgi et al., 2008; Nicoli et al., 2012; Hagerty et al., 2015; Nakedde et al., 2016; Vasquez-Guzman, 2016).

Common bean is divided into the Middle American and Andean gene pools (Mamidi et al., 2011; Bitocchi et al., 2013; Schmutz et al., 2014). The Middle American gene pool is further divided into four races which include Durango, Jalisco, Mesoamerican, and Guatemala (Singh et al., 1991; Beebe et al., 2001; Blair et al., 2009), while the Andean genepool is comprised of races Nueva Granada, Peru, and Chile (Singh et al., 1991). In North America and the United States, the most frequently cultivated common beans are members of market classes within races Durango and Mesoamerican of the Middle American genepool and Nueva Granada of the Andean genepool (USDBC 2017). The Andean Diversity Panel (ADP; Cichy et al., 2015) and Middle American Diversity Panel (MDP; Moghaddam et al., 2016) reflect modern genetic diversity in two common bean gene pools and were used extensively to study the genetics of abiotic and biotic stresses in common bean (Zuiderveen et al., 2016; Soltani et al., 2017, 2018; Oladzad et al., 2019b). In 2013, 310 ADP genotypes were evaluated in the field in Minnesota for resistance to root rot (Vasquez-Guzman, 2016). The major contributor to disease was *F. solani* f. sp. *phaseoli* and only five genotypes were considered resistant to FRR.

From a genetic perspective, a QTL mapping study of Middle American recombinant inbred lines developed from landrace Puebla 152 and the commercial black cultivar 'Zorro' detected one QTL on Pv05 associated with a resistance to FRR (Nakedde et al., 2016). Hagerty et al. (2015), placed a QTL associated with FRR resistance on Pv03 using a snap bean RIL population. Nine QTL explaining 5–53% of the phenotypic variation were identified in two inbred back cross line populations (IBL) developed from crosses between a Mesoamerican black bean with an Andean kidney bean and an Andean cranberry bean (Román-Avilés and Kelly, 2005). Finally, one genome-wide association study (GWAS) used field screening data of the ADP and 3,525 single nucleotide polymorphism (SNPs) and detected a genomic region on Pv04 centered at 3.3 Mbp associated with root rot resistance (Vasquez-Guzman, 2016).

A crucial step in developing resistant varieties is a reproducible protocol to screen for pathogen resistance under controlled conditions. Abiotic factors, including soil moisture content and temperature, can dramatically influence pathogen colonization or root development (Peña et al., 2013), resulting in inconsistent phenotypic evaluations. Improving the phenotypic methods of screening for resistance provides more robust and accurate phenotypic data that increases the power of GWAS to identify and map resistance QTL. GWAS results can vary based on methods either quantitative or qualitative, used to classify phenotypic data (Oladzad et al., 2019b). Therefore, another important aspect of this study was to develop GWAS results from quantitative, three-class, and binary scoring systems.

The objectives of this research were to: (1) develop a reproducible greenhouse evaluation system for FRR in common bean, (2) identify highly resistant genotypes to FRR in the ADP and MDP, and (3) discover genomic regions and potential candidate genes associated with resistance using GWAS. This is the first large scale-study to evaluate the ADP and MDP in the greenhouse and identify genomic regions involved in resistance to FRR.

MATERIALS AND METHODS

Greenhouse Assay Development and Phenotyping

Results from preliminary experiments indicated that very little to no disease symptoms were observed when only one *F. solani* isolate was used for inoculations (data not shown). Therefore, nine isolates of *F. solani* obtained from diseased dry beans in the Red River Valley region were chosen for inoculation. The isolates used were 09/RG/BF212, 08/RG/BF128, 09/RG/BF261, 08/RG/BF199, 09/RG/BF307, 09/RG/BF279, Fs101.5ND15, 08/RG/BF133, and 09.113.03. Isolates were confirmed as *F. solani* via amplification of the translation elongation factor alpha 1 (TEF-1α) with primers EF-1 and EF-2 (Knutsen et al., 2004). Isolates were grown for 1 week on 0.5 × potato dextrose agar (Difco^{TM} Potato Dextrose Media, BD) in 60 × 15 mm plates. Macroconidial spore suspensions were prepared for each isolate. Under sterile conditions, agar containing fungal growth from one 100 mm Petri plate from each isolate was cut into approximately 1 cm square pieces and added to a 2 L Erlenmeyer flask containing 1 L of CarboxyMethly-Cellulose broth (CMC: 15.0 g of CarboxyMethyl-Cellulose, 1.0 g NH_4NO_3,

1.0 g KH_2PO_4 monobasic, 0.5 g $MgSO_47H_2O$, 1.0 g yeast extract, 1 L distilled H_2O; Tuite, 1969). Flasks were swirled at 90 rpm for 7–9 days under continuous fluorescent light at room temperature. The macroconidial concentration for each isolate was adjusted to 1×10^6 in distilled H_2O using a hemacytometer. The macroconidial suspension from each isolate was combined in equal proportions.

Common bean genotypes from the MDP (280) and ADP (270) were evaluated for resistance to *F. solani* under greenhouse conditions. For each genotype, three seeds were planted in a single 4-inch plastic pot with drainage holes containing general-purpose PRO-MIX BX General Purpose (Quakertown, PA, United States) potting soil using a randomized complete block design with three replicates (1 replicate = 1 pot). Soil was saturated with water once daily. Inoculations were conducted when hypocotyl arches broke the soil surface by pipetting 5 mL of *F. solani* macroconidial suspension directly to the base of the seedling. Because the genotypes evaluated vary greatly in root architecture and color, a non-inoculated control, one pot containing three seeds, was included as a reference for disease severity ratings for each genotype. Soil was not watered again until plants reached the 80% wilting point by weight. Pots were watered every 2–3 days thereafter to maintain the soil at an 80% wilting point. All plants were maintained in a greenhouse under 16 h of light at 25°C ± 2°C with 90% relative humidity.

To measure assay reproducibility, the ADP was screened in two sub-groups each consisting of 135 lines and the MDP was divided into two sub-groups consisting of 140 lines (**Supplementary Tables S1, S2**). Susceptible (Montcalm and Cabernet), moderate (Dynasty and Talon), and resistant (VAX3) control lines were included when screening genotypes from each sub-group of the ADP and MDP. These lines were selected based on previously published data and preliminary trials (Bilgi et al., 2008; Vasquez-Guzman, 2016). Therefore, each control line was screened four times and each genotype was screened twice.

Fusarium Root Rot Evaluation and Data Analyses

Two weeks after inoculation, plants were harvested and roots were washed and evaluated for disease using a 1–9 disease rating scale; 1 = no visible disease symptoms, 3 = light discoloration without necrotic lesions or 10% of the hypocotyl/root tissues covered in root lesions, 5 = approximately 25% of the hypocotyl/root tissue is covered with lesions but the tissue remains firm, 7 = approximately 50% of the hypocotyl/root tissue is covered with lesions with considerable softening and rotting, 9 = approximately 75% or more of the hypocotyl/root tissue is affected with advanced stages of rotting along with significant reduction in root system (Van Schoonhoven and Pastor-Corrales, 1987). Infection in the control lines in each experiment was confirmed to be *F. solani* by isolating the fungus from roots. Roots were surface sterilized in a 0.8% NaOCl solution for 30 s and placed onto 0.5 × potato dextrose agar amended with streptomycin and neomycin, both at a concentration of 50 mg/L. Cultures were morphologically identified to species 1 week following hyphal tipping onto 0.5 × potato dextrose agar (Leslie and Summerell, 2006). The translation elongation factor alpha 1 (TEF-1α) was sequenced as described above to verify morphological identification (Knutsen et al., 2004).

Fusarium root rot severity from the control lines was utilized to evaluate assay reproducibility (Oladzad et al., 2019b). Mean, standard error (SE) of the mean, and coefficients of variability (SE of the mean/mean) were calculated from root rot scores. A one-way ANOVA ($\alpha = 0.05$) was conducted across the four MDP and ADP sub-group evaluations for each control line (Wong and Wilcox, 2000). Estimated relative treatment effects (ranging from 0 to 1), confidence intervals, and *P*-values were used to determine statistical differences across control lines within each sub-group evaluation (Shah and Madden, 2004; Oladzad et al., 2019b). Significant differences from control lines VAX3, Talon, and Montcalm in MDP and ADP genotypes were based on *P*-values generated from relative effects and associated 95% confidence intervals were calculated using the LD_CI macro in SAS (Domhof and Langer, 2002; Shah and Madden, 2004). Genotypes with relative effects not significantly different from VAX3 were classified as resistant to FRR.

Genome Wide Association Analysis and Candidate Genes Analysis

Two sets of approximately 200 k imputed SNPs for each diversity panel generated from genotype-by-sequencing (GBS) reads of 325 ADP and 469 MDP genotypes were used for association mapping (Oladzad et al., 2019a). The SNPs were filtered for minor allele frequency $\geq$ 5% for GWAS analysis. Initially the original data was used from the 1 to 9 scoring system as described earlier. However, different disease classifications identify different genetic factors associated with the resistance response (Oladzad et al., 2019a). Therefore, the disease score data was also evaluated as a binary distribution (score < 3 as resistant and score > 3 as susceptible), and as a three-class polynomial distribution (score < 2.5 as resistant, score = 2.5–3.5 as moderate, and score > 3.5 as susceptible). GEMMA was used for the GWAS analysis because the algorithms programmed in GEMMA can model different types of data distributions (Zhou and Stephens, 2012). For each run, random and mixed models were tested. A kinship matrix generated from the centered relatedness procedure in GEMMA was used as a random effect variable in the random model. A structure matrix generated from principle component analysis (PCA) using Prcom function in R (Price et al., 2006) was used as a fixed effect and together with kinship matrix were tested for a mixed model. Three and four PCAs were employed in model analyses for the ADP and MDP, respectively, accounting for 25–50% variation in each gene pool. *P-wald* test (the improved calibrated *P*-value in GEMMA) was calculated for the given model. The bootstrap distributions of *P*-values were estimated based on 10,000 resamples to determine the significance cutoff at the 0.01 and 0.1 frequency. The best-fitting model was chosen for each of the three phenotypic distributions based on the mean of the squared differences (MSD; Mamidi et al., 2011). The mhtplot function from R package gap was used to create Manhattan and QQ-plots (Zhao, 2007). To estimate the amount of phenotypic variation explained by

significant SNPs/regions, a likelihood-ratio-based (R2LR) was calculated using GenABEL package in R (Sun et al., 2010). Finally, the candidate genes were identified based on the best hit on *Arabidopsis thaliana* within a ±50 kb window of the significant SNPs or interval.

RESULTS

Fusarium Root Rot Greenhouse Assay Reproducibility

Mean disease severity (MDS) for the control lines across the two sub-group evaluations within each panel were not significantly different (**Table 1**). VAX3 was significantly more resistant to *F. solani* than the other control lines, with the exception of Dynasty in one of four sub-group evaluations (**Figure 1**). Montcalm was significantly more susceptible to FRR than all other control lines, except for Cabernet in one evaluation. Cabernet was significantly more susceptible than Talon and Dynasty for two and three of four sub-group evaluations, respectively. No significant difference in FRR was observed between Talon and Dynasty across all sub-group evaluations.

TABLE 1 | Fusarium root rot (FRR) mean disease severity (MDS) across two sub-group evaluations for the Andean Diversity Panel (ADP) and Middle American Diversity Panel (MDP).

Control line	Reaction to FRR[a]	ADP sub-groups		MDP sub-groups	
		MDS[b]	*p*-value[c]	MDS[b]	*p*-value[c]
VAX3	Resistant	1.5	0.45	1.3	1.00
Talon	Moderate	2.5	0.16	2.3	0.90
Dynasty	Moderate	2.7	0.29	2.3	0.52
Cabernet	Susceptible	3.2	0.63	3.8	0.20
Montcalm	Susceptible	4.5	0.58	4.5	0.45

[a]*Reaction to FRR based on previous literature (Vasquez-Guzman, 2016) and preliminary studies.* [b]*MDS based on a 1 to 9 scale (Van Schoonhoven and Pastor-Corrales, 1987).* [c]*One way ANOVA ($\alpha < 0.05$) comparing MDS for each control line within diversity panel across sub-group evaluations.*

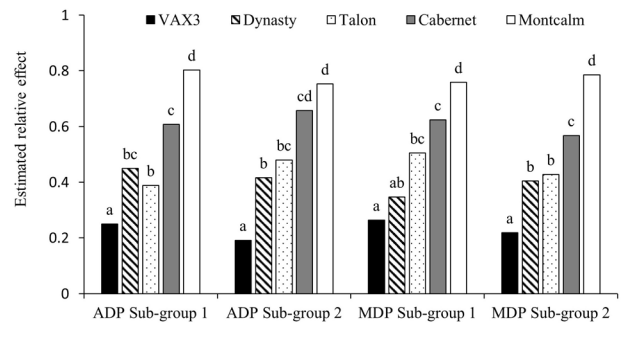


FIGURE 1 | Assay reproducibility of the five control lines when screened for root rot caused by *Fusarium solani* within sub-group evaluations of *Phaseolus vulgaris* Andean Diversity Panel (ADP) and Middle American Diversity Panel (MDP).

Fusarium Root Rot Resistant Genotypes

Relative effects generated from the FRR ratings for the ADP and MDP resulted in normal distributions (**Figure 2**). Root rot severity from single plants of genotypes in both the ADP and MDP ranged from 1 to 9. The MDS for the ADP was 2.7 with a range of 1.1–5.4 (**Figure 2A**). The mean relative effect for the ADP was 0.49 and the range was from 0.18 to 0.86 (**Figure 2B**). The MDS for the MDP was 2.3 with a range of 1.0–4.8 (**Figure 2C**). The relative effects for the MDP ranged from 0.24 to 0.87 (**Figure 2D**) with a mean of 0.50. The estimated relative effects for 34 genotypes from the ADP were not significantly different from the resistant control VAX3 (**Table 2** and **Supplementary Table S1**). Among these 34 were 10 kidney and seven yellow-seeded genotypes. Among the 47 MDP genotypes classified as resistant to FRR were nine from the pinto market class, 11 black, and 10 navy (**Table 3** and **Supplementary Table S2**). All lines from both panels statistically similar to the resistant control VAX3 were significantly different from the susceptible control Montcalm (**Supplementary Tables S1, S2**). Nine ADP lines were statistically similar to both VAX3 and Talon. All MDP lines displaying root rot statistically similar to VAX3 displayed significantly less root rot than Talon.

Genome Wide Association Analysis and Candidate Genes in the ADP

HapMap SNPs totaled 260,670 and 205,293 for the ADP and MDP, respectively[1]. After filtering for MAF, 219,056 SNPs for ADP and 125,745 SNPs for MDP were used for association study. Three GWAS analyses based on different phenotypic distributions detected common and unique SNPs or intervals associated with FRR. Two common genomic regions (Pv08:0.04–0.18 Mbp and Pv07:38.5 Mbp) were associated in the ADP in both the score and three-class phenotyping system at the $P < 0.01$ significance level (**Figure 3**). These two regions cumulatively explained 17 and 19% of the phenotypic variation in score and three-class analyses, respectively. However, a large significant genomic interval on Pv11:9.10–9.47 was detected in three-class (11% phenotypic variation explained) and binary (14%) analyses but not in score analysis. Unique SNPs at this threshold were also identified independently in each analysis (**Table 4**). Overall, binary data explained the most cumulative phenotypic variation (37%) and the significant SNPs detected in three-class data explained the most individual phenotypic variation associated with FRR in the ADP. When looking at the less stringent criteria (0.1% cutoff level), 151 SNPs were discovered to be common at least between two phenotyping system (**Supplementary Table S3**). From this, 17 SNPs were identified in all three systems. Fourteen SNPs were common between the score and binary analyses, 93 SNPs were common between score and three-class analyses, and 27 SNPs were common between binary and three-class analyses. Most of the SNPs on Pv11 were detected in binary and three-class analyses, while most of the SNPs on Pv02 and Pv08 were detected in score and three-class. When comparing the three phenotypic scoring

[1]http://arsftfbean.uprm.edu/beancap/research/

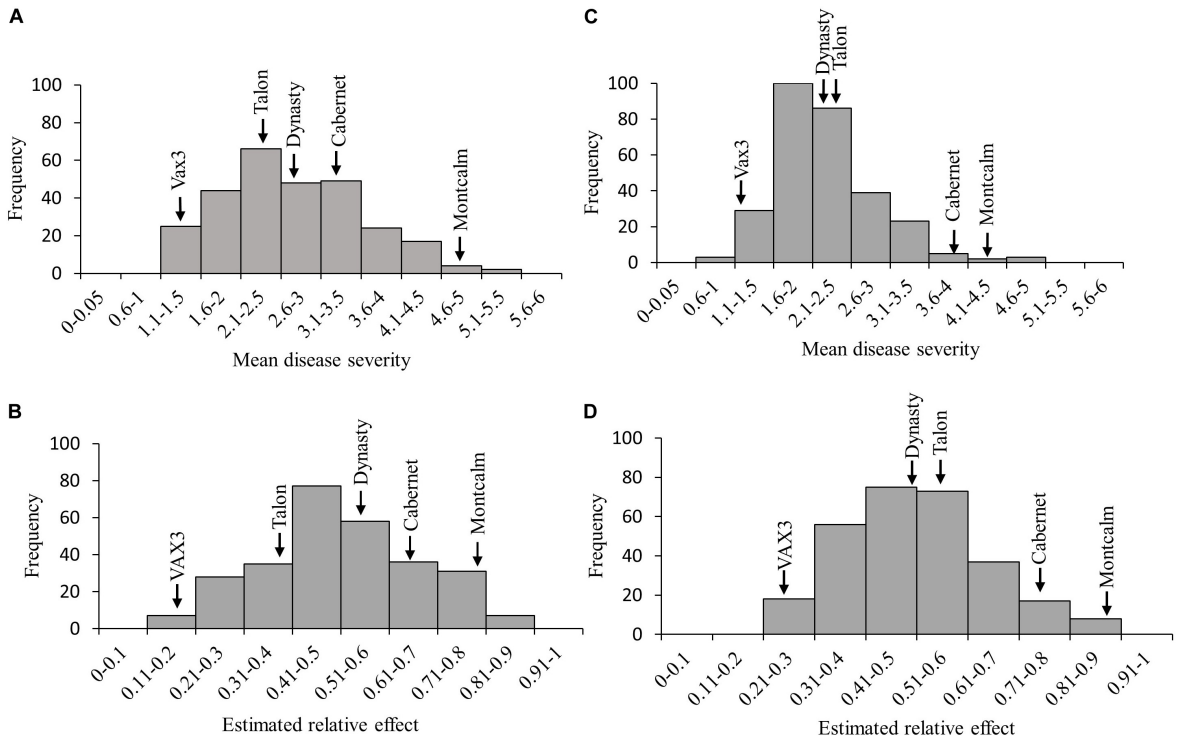


FIGURE 2 | Frequency distribution of **(A)** mean disease severity (MDS) and **(B)** the estimated relative effects of root rot caused by *Fusarium solani* in the Andean Diversity Panel (ADP). Frequency of **(C)** average disease score and **(D)** the estimated relative effects of *Fusarium solani* in the Middle American Diversity Panel (MDP). Arrows indicate the reactions of the five control lines.

analyses in the ADP GWAS analysis, the three-class scoring had the most frequent SNPs in common with another scoring system.

Candidate genes were searched for the peak SNPs at the 0.01 cutoff level and 52 potential candidate genes were identified from all three analyses (**Supplementary Table S4**). Twenty bean gene models were associated with genomic regions Pv08:0.04–0.18 Mbp, and 13 were previously characterized in Arabidopsis. The peak SNP at this interval was located inside the gene model *Phvul.008G001300* which encodes an ortholog of the glucan synthase-like enzyme (*GSL*) associated with the deposition of callose in papillae at pathogen wound sites (Ellinger and Voigt, 2014). A second gene, *Phvul.008G002250*, located 59 kb down stream of this peak SNP, also encodes a glucan synthase-like protein. Two other gene models, *Phvul.008G001100* and *Phvul.008G001200*, both orthologs of senescence-associated genes (*SAG*) involved in disease defense in plants, were also located within this interval. Both *SAG* and *GSL* genes clustered with the NAC domain protein *Phvul.008G001000,* which is also associated with fungal disease (Hickman et al., 2013). The SNP peak at Pv07:38.5 Mbp was located inside Homogentisate prenyltransferase gene encoding for plasoquinon. Eighteen common bean gene models were associated with Pv11:9.10–9.47 genomic interval. *Phvul.011G092600* maps within this interval, a member of the Subtilisin-like serine endopeptidase family protein that is involved in plant-pathogen interactions (Figueiredo et al., 2014). Additionally, eight candidate genes unique to the binary analysis and six candidate genes unique to the three-class analysis were associated with FRR. *Phvul.002G329400*, an ortholog of the pathogenesis-related thaumatin, and *Phvul.003G009700*, a member of the pentatricopeptide repeat (*PPR*) superfamily, were located in significant regions. Presumed functions of these are related to the plant pathogen resistance response (Geddy and Brown, 2007). WRKY DNA-binding protein, P-loop containing nucleoside triphosphate hydrolases, and Leucine-rich receptor-like protein kinase family proteins associated with gene models *Phvul.008G251700*, *Phvul.004G086200*, and *Phvul.009G260500* in binary analysis are also noted for their role in plant disease resistance (Ellis et al., 2000; Yu et al., 2001).

Genome Wide Association Analysis and Candidate Genes in the MDP

The Pv01:51.03–51.07 Mbp and Pv07:15.29 intervals were shared between the score and three-class analyses at the 0.01 cutoff (**Figure 4** and **Table 5**). These intervals explained 8 and 9% of the phenotypic variation in score and three-class analyses, respectively. The binary analysis did not share any significant SNP at this significance level with the other two phenotyping analyses. The largest cumulative phenotypic effect, 27%, was observed for the significant SNPs when using phenotypic score. Significant regions on Pv04 were detected when using the binary analysis. However, more shared SNPs were detected among all three analyses when SNPs significant at the 0.1 cutoff were considered (**Supplementary Table S5**). A total of 51 significant SNPs were shared between at least two scoring systems, of which eight SNPs were shared among all three, six SNPs were

TABLE 2 | Common bean lines from the Andean Diversity Panel (ADP) classified as resistant when compared to the resistant control VAX3 based on overlapping confidence intervals and *P*-values ($\alpha < 0.05$)[a].

Line[b]	Marketclass/seed color	Genotype[c]	Mean rank	Est. relative effect[d]	Confidence interval (95%) for relative effect[e]	
					Lower limit	Upper limit
VAX3	Resistant Control	VAX3	52.8	0.19	0.18	0.20
ADP626	Light Red Kidney	Badillo	73.3	0.26	0.14	0.42
ADP640	White Kidney	Beluga	82.0	0.30	0.17	0.46
ADP43	Dark Red Kidney	BWANA_SHAMBA	83.5	0.30	0.17	0.46
ADP99	Dark Red Kidney	BwanaShamba	83.5	0.30	0.17	0.46
ADP511	Yellow	Canario	51.2	0.19	0.18	0.20
ADP513	Yellow	Canario	72.1	0.26	0.18	0.36
ADP186	Red	G1368	52.8	0.19	0.18	0.20
ADP214	Black	G5087	72.3	0.26	0.14	0.42
ADP444	Red Mottled	HondoValle25	72.3	0.26	0.14	0.42
ADP612	Dark Red Kidney	ICA Quimbaya	62.6	0.22	0.16	0.30
ADP683	Pink Mottled	IJR	62.6	0.22	0.16	0.30
ADP621	Yellow	JaloEEP558	73.3	0.26	0.14	0.42
ADP88	Purple Speckled	KABLANKETI	52.8	0.19	0.18	0.20
ADP81	Purple Speckled	KABLANKETI	83.5	0.30	0.17	0.46
ADP519	Sugar	Katarina Cela	72.7	0.26	0.14	0.42
ADP4	Full	KILOMBERO	62.3	0.22	0.16	0.30
ADP94	Yellow	LUSHALA	72.3	0.26	0.14	0.42
ADP684	Dark Red Kidney	Majesty	82.6	0.30	0.17	0.46
ADP514	Yellow	MantegaAmarela	52.8	0.19	0.18	0.20
ADP21	Yellow	MBULAMTWE	81.4	0.30	0.17	0.46
ADP42	Dark Red Kidney	MKOKOLA	73.3	0.26	0.14	0.42
ADP391	Light Red Kidney	PI308894	82.9	0.30	0.17	0.46
ADP392	Sugar	PI309701	72.3	0.26	0.14	0.42
ADP481	Red Mottled	PI449428	62.3	0.22	0.16	0.30
ADP474	Red Mottled	PI527519	72.3	0.26	0.14	0.42
ADP462	Yellow	PI527540B	83.9	0.30	0.17	0.46
ADP429	Pink Cranberry	PR9920_171	72.7	0.26	0.14	0.42
ADP1	Red Mottled	ROZI_KOKO	52.8	0.19	0.18	0.20
ADP602	Light Red Kidney	Sacramento	78.7	0.27	0.14	0.47
ADP112	Red	Uyole96	71.6	0.26	0.14	0.42
ADP111	Sugar	Uyole98	52.8	0.19	0.18	0.20
ADP2	Purple Speckled	W6_16444	72.9	0.26	0.14	0.42
ADP15	Dark Red Kidney	W6_16495	71.6	0.26	0.14	0.42
ADP91	Manteca	W6_16560	73.3	0.26	0.14	0.42

[a]P-values ($\alpha = 0.05$) were generated as previously described (Altman and Bland, 2011). [b]ADP as previously described (Cichy et al., 2015). [c]VAX3 was utilized as the resistant control. [d]Estimated relative effect was determined using the mean rank for each bean line among all the observations within the experiment. Relative effects range between the values of 0 and 1.0 (Shah and Madden, 2004). [e]95% confidence intervals were generated from the estimated relative effect values.

shared between score and binary systems, 29 SNPs were shared between score and three-class systems, and eight SNPs were shared between binary and three-class. In addition to the shared SNPs, each analysis discovered unique SNPs associated with FRR (**Supplementary Table S5**).

Candidate genes were searched for the peak SNPs at the 0.01 cutoff level and 27 potential candidate genes were identified across all three analyses (**Supplementary Table S6**). Ten gene models were associated with the genomic interval Pv01:51.03–51.07 Mbp. Peak SNP Pv01:51.03 Mbp was located 2 kb upstream of gene model *Phvul.001G263600*, an ortholog of 15-*cis*-zeta-carotene isomerase (*Z-ISO*). Two gene models, *Phvul.001G263700* and *Phvul.001G263800*, orthologs of Immunoglobulin E-set superfamily protein and MAC/Perforin domain-containing protein respectively, were identified upstream of this peak SNP and were detected only in the three-class analysis. Both genes are involved in plant disease defense (Morita-Yamamuro et al., 2005; Zhang et al., 2014). The SNP peak Pv07:15.2 Mbp was located 10kb upstream of gene model *Phvul.007G127800*, an ortholog of ethylene response factor 1 (*ERF*) which regulates plant resistance to some soil-born fungi (Berrocal-Lobo and Molina, 2004). A cluster of cytochrome P450/family 96/subfamily A/polypeptide 10 (*CYP96A10*) was associated with Pv04:37.0–40.34 Mbp in binary analysis.

TABLE 3 | Common bean lines from the Middle American Diversity Panel (MDP) classified as resistant as compared to the resistant control VAX3 based on overlapping confidence intervals and *P*-values ($\alpha < 0.05$)[a].

Line[b]	Marketclass/seed color	Genotype[c]	Mean rank	Est. relative effect[d]	Confidence interval (95%) for relative effect[e]	
					Lower limit	Upper limit
VAX3	Resistant Control	VAX3	65.9	0.24	0.23	0.26
MDP113	Pinto	Fargo	65.9	0.24	0.23	0.26
MDP126	Black	Loreto	71.2	0.24	0.23	0.26
MDP134	Navy	Navigator	65.8	0.24	0.23	0.26
MDP140	Small Red	Ember	65.9	0.24	0.23	0.26
MDP216	Black	I9365_31	65.9	0.24	0.23	0.26
MDP267	Pink	Victor	65.9	0.24	0.23	0.26
MDP332	Black	CDC_Jet	65.9	0.24	0.23	0.26
MDP349	Black	Harrowhawk	65.9	0.24	0.23	0.26
MDP52	Pinto	I06_2575_17	65.9	0.24	0.23	0.26
MDP55	Navy	Sanilac	67.3	0.24	0.23	0.26
MDP9	Small Red	AC_Redbond	65.9	0.24	0.23	0.26
MDP142	Pink	ROG_312	79.5	0.30	0.21	0.41
MDP167	Pinto	UI_126	78.8	0.30	0.21	0.41
MDP302	Pinto	ND_307	75.5	0.30	0.21	0.41
MDP32	Black	DPC_4	78.8	0.30	0.21	0.41
MDP329	Great Northern	CDC_Crocus	78.5	0.30	0.21	0.41
MDP331	Black	CDC_Expresso	80.9	0.30	0.21	0.41
MDP403	Navy	McHale	78.5	0.30	0.21	0.41
MDP129	Navy	Voyager	80.7	0.31	0.20	0.44
MDP131	Pink	Pink_Floyd	79.8	0.31	0.20	0.44
MDP187	Great Northern	GN_Star	82.9	0.32	0.19	0.49
MDP290	Tan	BAT_477	85.3	0.32	0.19	0.49
MDP13	Navy	AC_Polaris	88.8	0.33	0.19	0.51
MDP14	Great Northern	AC_Resolute	88.0	0.33	0.19	0.51
MDP146	Black	Black_Knight	87.2	0.33	0.19	0.51
MDP15	Small Red	AC_Earlired	88.9	0.33	0.19	0.51
MDP159	Small Red	UI_37	88.9	0.33	0.19	0.51
MDP201	Great Northern	NE1_09_20	89.1	0.33	0.19	0.51
MDP203	Pinto	NE2_09_1	88.8	0.33	0.19	0.51
MDP239	Pinto	USPT_CBB_5	89.1	0.33	0.19	0.51
MDP268	Pink	USWA_61	88.8	0.33	0.19	0.51
MDP286	Pink	A285	88.9	0.33	0.19	0.51
MDP3	Pinto	BelDakMi_RR_5	88.8	0.33	0.19	0.51
MDP383	Pinto	Apache	87.2	0.33	0.19	0.51
MDP395	Black	Black_Velvet	88.0	0.33	0.19	0.51
MDP43	Small Red	TARS09_RR007	88.0	0.33	0.19	0.51
MDP61	Navy	Neptune	87.2	0.33	0.19	0.51
MDP66	Black	C_20	87.2	0.33	0.19	0.51
MDP7	Great Northern	BelNeb_RR_1	89.1	0.33	0.19	0.51
MDP78	Navy	Mackinac	89.1	0.33	0.19	0.51
MDP96	Black	Cornell 49-242	88.8	0.33	0.19	0.51
MDP99	Pink	S08418	87.2	0.33	0.19	0.51
MDP133	Navy	Medalist	93.8	0.35	0.18	0.56
MDP339	Navy	Nautica	92.9	0.35	0.18	0.56
MDP382	Pinto	Sequoia	92.9	0.35	0.18	0.56
MDP392	Black	B05055	96.0	0.35	0.18	0.56
MDP90	Navy	Albion	91.3	0.35	0.18	0.56

[a]P-values ($\alpha = 0.05$) were generated as previously described (Altman and Bland, 2011). [b]MDP as previously described (Moghaddam et al., 2016). [c]VAX3 was utilized as the resistant control. [d]Estimated relative effect was determined using the mean ranks for each bean line among all the observations within the experiment. Relative effects range between the values of 0 and 1.0 (Shah and Madden, 2004). [e]95% confidence intervals were generated from the estimated relative effect values.

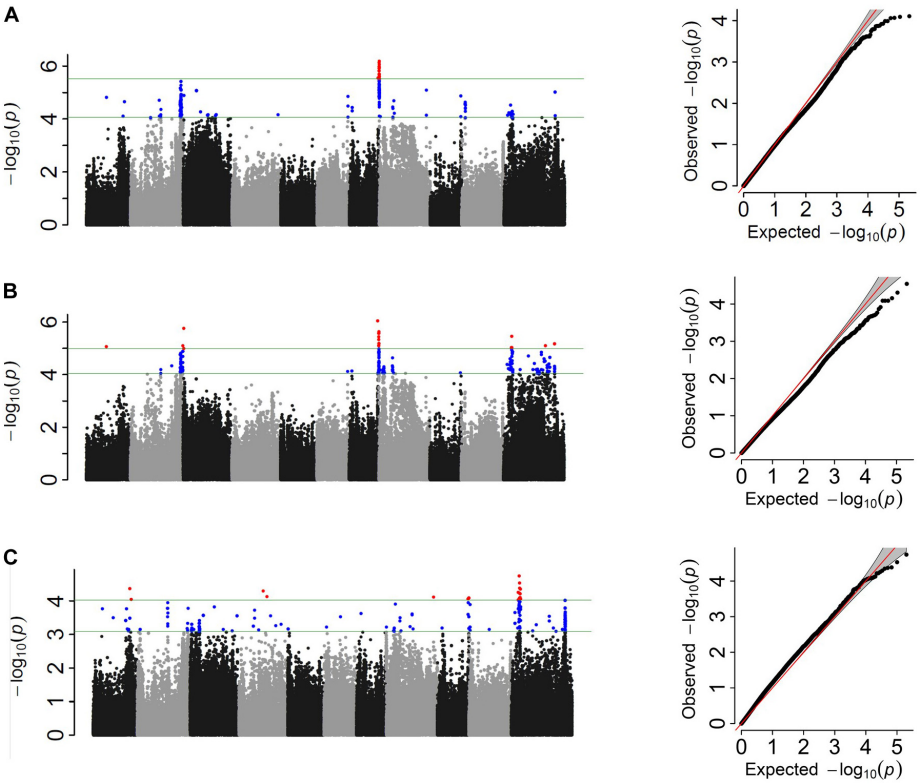


FIGURE 3 | Manhattan and corresponding Q–Q plots representing the genetic architecture of *Fusarium solani* resistance from GWAS analysis of **(A)** quantitative, **(B)** three-class, and **(C)** binary scoring systems in the Andean Diversity Panel (ADP).

TABLE 4 | Significant genomic regions/SNPs from GWAS in Andean Diversity Panel ($P < 0.01$).

Phenotypic data	Interval		Peak SNP		*R* square	Cumulative R square
	Chrom	Genomic interval (Mb)	Position (Mb)	−Log10(P)		
Score	8	0.04-0.18	S08_184683	6.18	0.16	0.17
	7	38.5	S07_38504717	5.53	0.14	
Three class	1	26.86	S01_26860673	5.05	0.12	0.32
	2	49.43	S02_49439131	5.01	0.13	
	3	1.07	S03_1079664	5.75	0.15	
	7	38.5	S07_38504717	6.04	0.15	
	8	0.04–0.18	S08_184683	5.64	0.14	
	11	9.10–9.47	S11_9423668	5.46	0.14	
	11	37.06	S11_9472666	5.03	0.12	
	11	44.79	S11_44799085	5.16	0.13	
Binary	11	8.13	S11_8136772	4.25	0.10	0.37
	11	9.10–9.47	S11_9106599	4.7	0.11	
	11	10.01–10.04	S11_10397908	4.34	0.09	
	1	42.16	S01_42160134	4.37	0.10	
	1	44.34	S01_44346150	4.03	0.08	
	4	24.87	S04_24875044	4.29	0.09	
	4	28.34	S04_28344173	4.12	0.09	
	8	60.07	S08_60070729	4.11	0.09	
	9	38.06	S09_38068347	4.07	0.09	

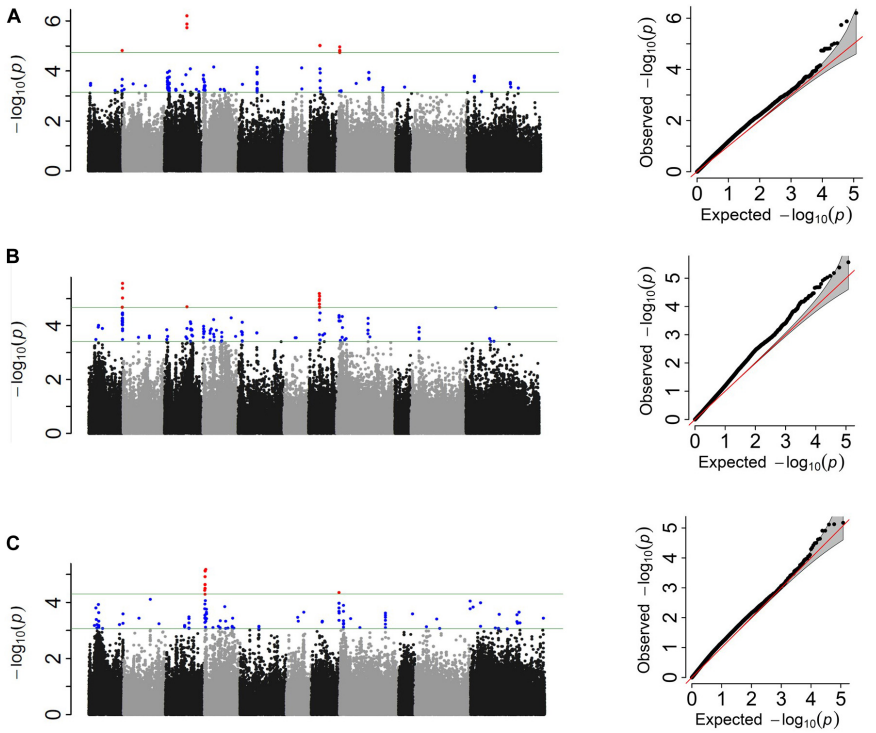


FIGURE 4 | Manhattan and corresponding Q–Q plots representing the genetic architecture of *Fusarium solani* resistance from GWAS analyses using **(A)** quantitative, **(B)** three-class, and **(C)** binary scoring systems in Middle American Diversity Panel (MDP).

TABLE 5 | Significant genomic regions/SNPs from GWAS in Middle American Diversity Panel ($P < 0.01$).

Phenotypic data	Interval		Peak SNP		*R* square	Cumulative R square
	Chrom	Genomic interval (Mb)	Position (Mb)	−Log10(P)		
Score	3	25.26	S03_25262205	6.2	0.10	0.27
	7	15.29	S07_15295231	5.02	0.08	
	8	5.10-5.11	S08_5112094	4.9	0.08	
	1	51.03	S01_51036396	4.82	0.07	
Three class	1	50.79-50.82	S01_50822097	4.68	0.07	0.20
	1	51.03-51.07	S01_51036396	5.56	0.09	
	7	15.29	S07_15295263	5.18	0.08	
Binary	4	0.36-0.43	S04_376163	5.12	0.08	0.09
	4	1.31	S04_1315429	5.17	0.08	

Finally, gene models *Phvul.003G098500* and *Phvul.008G057500,* orthologs of DUF679 domain membrane protein (*DMP*) and GRIP-related ARF-binding domain-containing protein (*GDAP*) respectively were associated with Pv03:25.26 Mbp, Pv08:5.10–5.11 Mbp when using the score system analysis.

DISCUSSION

This research provides a reproducible greenhouse assay for the evaluation of resistance to FRR. Greenhouse assays were demonstrated as reproducible with the consistent results from five control lines. The inclusion of three of these control lines is recommended for continuing FRR evaluations and ongoing comparisons across research studies. VAX3 and Montcalm have been previously documented as appropriate resistant and susceptible controls, respectively (Bilgi et al., 2008; Vasquez-Guzman, 2016). Those observations were consistent with our results; therefore, we recommend the inclusion of these two lines in future FRR evaluations. Additionally, VAX3 and Montcalm were resistant and susceptible to Rhizoctonia root rot, respectively, making them excellent choices for control lines to be included in field studies where both *Fusarium* spp. and *Rhizoctonia solani* Kühn (teleomorph *Thanatephorus cucumeris*) are likely playing a role in the root rot complex (Oladzad et al.,

2019b). In addition to resistant and susceptible controls, the inclusion of a line consistently displaying a moderate reaction to a quantitative trait like root rot resistance is important to fully characterize lines evaluated in all studies. Talon, Dynasty, and Cabernet generally all displayed a moderate reaction to FRR in the current evaluations. FRR severity of these three lines was significantly different from both VAX3 and Montcalm with only one exception; therefore, any of these would be appropriate for inclusion in future studies.

It is particularly difficult to generate reproducible results when evaluating quantitatively inherited traits. Environmental parameters, microbial interactions, and screening methods present numerous opportunities for data inconsistencies. These variables are exacerbated when handling a soil-borne pathogen complex. Soil-borne plant pathogens exist simultaneously with a variety of non-pathogenic soil microorganisms, creating complexes, and therefore, the focus of research concerning soil-borne plant pathogens is being redirected to considering complexes rather than individual strains of individual species (Lamichhane and Venturi, 2015; Abdullah et al., 2017). Similar to other *Fusarium* species complexes, species within the FSSC display varying degrees of aggressiveness (Chitrampalam and Nelson, 2016). Therefore, the approach taken here to incorporate multiple isolates/strains from the FSSC in phenotypic evaluations is more representative of field conditions and contributed to assay reproducibility.

Successful infection by a soil-borne pathogen is heavily dictated by environmental parameters. Controlling or predicting the soil environment is complicated; therefore, determining host resistance against soil-borne pathogens may not be applicable across environments. Of the numerous soil parameters that dictate the infection rate and severity of FRR pathogens, the effects of soil temperature and moisture have been most documented (Porch et al., 2014; Teixeira et al., 2015; Macedo et al., 2017). In North Dakota, variability in growth rates and aggressiveness were observed across a range of temperatures for 96 species in the FSSC (Chitrampalam and Nelson, 2016). Most *Fusarium* spp. require some type of plant stress to incite infection on pulse crops; such stress may include droughts or flooding (Leslie et al., 1990). Soil moisture also has been documented to influence the presence and inoculum density of FSSC (Macedo et al., 2017). Flooding and drought conditions have been reported to substantially weaken the dry bean root system and allow infection caused by FRR (Gossen et al., 2016; Macedo et al., 2017). To date, this is the first FRR screening method to account for soil moisture and the microbial interactions of numerous strains within the FSSC, likely contributing to assay reproducibility.

While field trials are the truest test of pathogen resistance, they are laden with many challenges. Soil-borne pathogens are rarely evenly distributed throughout naturally infested fields, contributing to potentially inconsistent results, particularly when screening a large number of lines. In addition to affecting host susceptibility, the level of disease pressure is greatly affected by environmental conditions. Field trials are constantly threatened by weather events including hail, flooding, or drought. In contrast, greenhouse evaluations somewhat ignore interactions across the soil microbiome and soil environmental factors that affect disease development. However, greenhouse screening can provide the opportunity for consistently identifying genotypes with resistance to a single pathogen over a short time period, as is demonstrated here. The FRR resistance data presented here is more dynamic due to the implementation of a new greenhouse screening method utilizing a mixture of isolates and drought stress to promote disease development on a large set of genotypes (ADP = 270; MDP = 280). Thirty-four ADP and 47 MDP resistant genotypes were identified from the most important market classes in both the Middle American and Andean gene pools including pinto, black, great northern, and dark red kidney. The identification of these genotypes will provide breeding programs with valuable germplasm for incorporation of resistance to this important soil-borne pathogen.

Previous research for screening of FRR resistance either utilized single isolates of *F. solani* or were conducted in naturally pathogen-infested fields (Román-Avilés and Kelly, 2005; Bilgi et al., 2008; Nicoli et al., 2012; Conner et al., 2014; Hagerty et al., 2015; Nakedde et al., 2016; Vasquez-Guzman, 2016). However, the FRR reaction of some genotypes evaluated in this study were supported by results from previous research. The ADP genotype PI5275408B was characterized as resistant, Etna, Fox Fire, Pink Panther, W6_6534, and 46_1 were characterized as moderately resistant, and Montcalm were characterized as susceptible (Bilgi et al., 2008; Conner et al., 2014; Vasquez-Guzman, 2016). Similarly, some previously screened MDP genotypes, AC_Polaris, AC_Resolute, and AC_Earlired (resistant), Navigator, CDC_Jet, AC_Redbond, AC_Island, Black Violetand Zorro (moderately resistant), and Beryl, Envoy, Matterhorn, and Othello (susceptible) were confirmed with same response in our study (Bilgi et al., 2008; Conner et al., 2014; Nakedde et al., 2016).

Some of the FRR resistant genotypes identified in this study also displayed resistance to Rhizoctonia root rot. ADP genotypes ROZI_KOKO, W6_16495, and HondoValle25 and MDP genotypes USWA_61, Nautica, B05055 previously described as resistant to *R. solani* AG2-2 are also resistant to FRR in the present study (Oladzad et al., 2019b). These genotypes provide a unique opportunity for the incorporation of resistance to at least two important soil-borne pathogens of the common bean. Two genomic regions on Pv02 and PV11 identified in this study were also associated *R. solani* resistance in ADP panel (Oladzad et al., 2019b). However, further investigation is needed to determine if the resistance to these two pathogens is due to root architecture or some feature other than host resistance. In a previous study, FRR was determined to be controlled by root genotype and root vigor played an important part in resistance (Cichy et al., 2007). Wang et al. (2018) demonstrated that resistant lines had a slightly higher root biomass and hypothesized that some of the QTL associated with FRR resistance are more likely related to root biomass. High resistance consistently associated with root architecture, which indicates that these may be dependent traits and need to be considered when selecting lines for resistance breeding (Strock et al., 2019).

This study is the first to utilize ∼200 K SNPs to identify SNPs closely associated with FRR resistance in major common bean gene pools. These high-quality SNPs obtained through GWAS in this study will provide the foundation for confident subsequent analyses of candidate genes and can be converted into breeder friendly markers to aid in the incorporation of FRR resistance in high yielding lines through marker assisted selection. In the current study, the genomic regions associated with response to *F. solani* in the common bean were identified independently in two diversity panels representing modern germplasm from the two common bean gene pools. As previous studies suggested, different genetic factors might be involved for the same traits in each common bean gene pool (Schmutz et al., 2014; Soltani et al., 2017, 2018; McClean et al., 2018; Oladzad et al., 2019a,b). Previous work on *R. solani* (Oladzad et al., 2019b) demonstrated that shared and/or unique genomic regions were significantly associated with disease severity when the data was considered on a 1–9 scale, or transformed into binary or multinomial distributions. For this reason, three independent GWAS analyses were performed using the score data and three-class and binary phenotypic data for each diversity panel. As expected, each GWAS analysis detected SNPs specific to each phenotypic data set as well as shared significant regions. Identifying unique, as well as shared, SNPs using three phenotypic approaches is consistent with GWAS results previously observed for *R. solani* resistance in the common bean (Oladzad et al., 2019b). In general, for both the ADP and MDP, genomic intervals discovered with the three-class phenotypic distribution data were shared with results obtained GWAS results for the binary and the 1–9 distribution data. However, in searching for candidate genes, it was observed that each phenotyping data set used for GWAS analysis offered specific important genetic regions associated with FRR. The ADP GWAS identified 52 potential candidate genes from all three analyses, and of these, nine candidates were previously documented for their roles in plant pathogen response. During evolution, plants developed successful immune/defense mechanisms against pathogen infections. One of these mechanisms was discovered from expression studies on Glucan synthase-like protein (*GSL*), which is a potential ADP candidate gene (*Phvul.008G001300*). *GSL* regulates callose synthesis, which are abundant components of the papillae structure at sites of fungal penetration (Voigt, 2014). The formation of this complex structure appears at the earliest phase of the plant defense response to pathogen infection. It has been shown in Arabidopsis that the elevated cell wall callose polymers in papillae provides complete pathogen penetration resistance (Ellinger et al., 2013). In our study, *GSL* were found in the vicinity of NAC domain and senescence-associated genes (*SAG*). NAC domains generally have important roles in the regulation of both biotic and abiotic stresses in plants. However, significant progress in NAC domain function studies revealed the important role of these domains in activating plant's defense responses. It has been shown that both positive and negative regulatory roles of NAC transcription factors (TFs) in the alteration of gene expression are the key mechanisms employed by plants during pathogen attack (Nuruzzaman et al., 2013). Most of these genes regulate signaling of plant hormones during the immune response (Yuan et al., 2019). More importantly, *NAC* TFs and *SAG* genes seem to be closely related to the biotic stress response. Pathogen infection is one of the factors that affects the signaling pathway of senescence in plants. NAC subfamily proteins are involved in altering the regulation of senescence signaling pathway (Guo and Gan, 2006), and these *NAC* TFs regulate salicylic acid (*SA*) and jasmonic acid (*JA*) signaling pathways, both known to accelerate developmental senescence in plants (Hickman et al., 2013). Taken together, detecting a cluster of these three genes in the vicinity of each other on Pv08 supports the idea that these are the most important potential candidate genes associated with *F. solani* response in the ADP.

Subtilisin-like serine and WRKY DNA-binding protein (*WRKY*) genes also mapped to significant genomic regions in the ADP on Pv11 and Pv08, respectively. The first evidence of subtilisin-like serine related to plant pathogenesis was reported in tomatoes (Tornero et al., 1996). Subsequently it was shown that, in soybeans, a plant defense peptide signal (*GmSubPep*) is embedded in subtilisin-like protein (*Glyma18g48580*) and upon pathogen attack, this peptide is accessible to activate defense-related genes (Pearce et al., 2010). In grapes, a subgroup of subtilisin-like proteins exhibit slight structural modifications in varieties resistant to *Plasmopara viticola* to affect plant programmed cell death (PCD) at the site of pathogen attack, such that the pathogen is not able to recognize this protein and prevents the PCD response that is a component of the plant immune resistance mechanism (Gindro et al., 2012; Figueiredo et al., 2014). WRKY DNA binding proteins act upstream of pathogenesis-related genes (*PRs*) and positively regulate their expression upon plant infection by a pathogen (Yu et al., 2001). One PR gene model (*Phvul.002G329400*) was detected in the ADP three-class GWAS analysis, which could be the target of this *WRKY* gene (*Phvul.008G251700*).

In MDP GWAS analysis, 27 potential candidate genes from all three GWAS analysis were associated with resistance to FRR. Many are candidate genes previously documented for their roles in plant pathogen response. On Pv04, a significant peak was located in a cluster of *CYP450* genes. These are members of one of the largest protein families in plants, and they are involved in many diverse biological processes, including metabolism pathways and hormonal responses to biotic and abiotic stresses (Bak et al., 2011). This cluster maps adjacent to a *NB-ARC* disease resistance gene cluster. Both clusters might be potential candidate regions associated with FRR. An ethylene response factor 1 (*ERF1*) was associated with peak SNP on Pv07. *ERF1* is a transcriptional factor (TF) with a role in plant resistance to *Fusarium* spp. and necrotrophic fungi (Lorenzo et al., 2003; Berrocal-Lobo and Molina, 2004; van Loon et al., 2006; Van der Ent and Pieterse, 2018). Upon pathogen attack, *ERF1* TFs trigger the activation of *PR* genes through a positive regulation of *JA* gene expression. A Pv01 candidate, a MAC/Perforin domain-containing gene, is part of the perforine membrane attack complex which plays a key role in both plant and animal innate immunity. In Arabidopsis, it has been shown that *MACPF* is encoded by *constitutively activated cell death 1* gene (*CAD1*) that is negatively regulated by the *SA* signaling pathway. Therefore, in resistant plants, the

mutant form (*cad1*) activates expression of *PR* genes, leading to *SA* enhancement of the programmed cell death and thus restricting pathogen growth (Morita-Yamamuro et al., 2005; Fukunaga et al., 2017).

Some common SNPs were found to be closely linked between Rhizoctonia and Fusarium root rot resistance in GWAS studies across the MDP and ADP (Oladzad et al., 2019b). The significant genomic region Pv02:49.43 Mb detected in this study is in the vicinity of genomic region Pv02:48.38–49.41 Mb associated with Rhizoctonia resistance in the ADP. Moreover, Pv09:38.06 and Pv11:8.13 Mb associated with Fusarium resistance in this study were near Rhizoctonia resistance genomic region Pv09:31.61 and Pv11:7.84 Mb in the same gene pool. Genomic region detected at Pv01:50.82 Mb was 10.62 Mb upstream and genomic region Pv08:5.11 Mb was 12.74 Mb downstream of the significant genomic regions were reported in Rhizoctonia resistance in the MDP. Therefore, based on our evaluations, some common genetic factors may be involved in the resistance of both Rhizoctonia and Fusarium root rot in the common bean.

To our knowledge, this is the first GWAS study on over 500 genotypes across the main gene pools in the common bean using a large number of SNP markers for studying genetic basis of resistance to FRR. Overall, most of the candidate genes detected in both gene pools seems to be involved in signaling pathways such as *SA* and *JA* through activating the expression of *PR*. However, the significant SNPs detected in each gene pool can be used in common bean breeding to speed up and lower the cost of selecting for resistance to this pathogen.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the http://arsftfbean.uprm.edu/beancap/research/.

AUTHOR CONTRIBUTIONS

JP, JO, and KZ-A designed and conceived the experiments. KZ-A and CM performed the experiments in the greenhouse. AO, KZ-A, and SJ performed the data analysis and drafted the manuscript. PM and JP supervised the data analysis. JP, KZ-A, SJ, and AO discussed the results and interpretation of the final data. PM, JP, and JO provided suggestions to improve it. All authors read and approved the final manuscript.

FUNDING

This publication was supported by the U.S. Department of Agriculture's (USDA) Agricultural Marketing Service through grant 15-SCBGPND-0026. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USDA. Northarvest Bean Growers Association also generously provided funding for this project.

ACKNOWLEDGMENTS

The authors thank Dr. Linda Hanson, USDA-ARS Research Plant Pathologist in the Sugarbeet and Bean Research Unit, East Lansing, MI, United States for her expertise in the development of the greenhouse screening assay.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00475/full#supplementary-material

REFERENCES

Abawi, G. S. (1989). "Root rots," in *Bean Production Problems in the Tropics*, 2nd Edn, eds H. F. Schwartz and M. A. Pastor-Corrales (Cali: International Agricultural Center for Tropical Agriculture (CIAT)), 119–120.

Abdullah, A. S., Moffat, C. S., Lopez-Ruiz, F. J., Gibberd, M. R., Hamblin, J., and Zerihun, A. (2017). Host–multi-pathogen warfare: pathogen interactions in co-infected plants. *Front. Plant Sci.* 8:1806. doi: 10.3389/fpls.2017.01806

Altman, D. G., and Bland, J. M. (2011). How to obtain the P value from a confidence interval. *BMJ* 343:d2304. doi: 10.1136/bmj.d2304

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., et al. (2011). Cytochromes p450. *Arabidopsis Book* 9:e0144. doi: 10.1199/tab.0144

Beebe, S., Rengifo, J., Gaitan, E., Duque, M. C., and Tohme, J. (2001). Plant genetic resources. *Crop Sci.* 41, 854–862.

Berrocal-Lobo, M., and Molina, A. (2004). Ethylene response factor 1 mediates *Arabidopsis* resistance to the soilborne fungus *Fusarium oxysporum*. *Mol. Plant Microbe Int.* 17, 763–770. doi: 10.1094/mpmi.2004.17.7.763

Bilgi, V. N., Bradley, C. A., Khot, S. D., Grafton, K. F., and Rasmussen, J. B. (2008). Response of dry bean genotypes to *Fusarium* root rot, caused by *Fusarium solani* f. sp. phaseoli, under field and controlled conditions. *Plant Dis.* 92, 1197–1200. doi: 10.1094/PDIS-92-8-1197

Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., et al. (2013). Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in mesoamerica and the andes. *New Phytol.* 197, 300–313. doi: 10.1111/j.1469-8137.2012.04377.x

Blair, M. W., Diaz, L. M., Buendia, H. F., and Duque, M. C. (2009). Genetic diversity, seed size associations and populations structure of a core collection of common beans (*Phaseolus vulgaris* L). *Theor. Appl. Genet.* 119, 955–972. doi: 10.1007/s00122-009-1064-8

Chitrampalam, P., and Nelson, B. (2016). Multilocus phylogeny reveals an association of agriculturally important *Fusarium solani* species complex (FSSC) 11, and clinically important FSSC 5 and FSSC 3+ 4 with soybean roots in the north central United States. *Antonie Van Leeuwenhoek* 109, 335–347. doi: 10.1007/s10482-015-0636-7

Cichy, K. A., Porche, T., Beaver, J. S., Cregan, P. B., Fourie, D., Glahn, R., et al. (2015). A *Phaseolus vulgaris* diversity panel for andean bean improvement. *Crop Sci.* 55, 2149–2160. doi: 10.2135/cropsci2014.09.0653

Cichy, K. A., Snapp, S. S., and Kirk, W. W. (2007). *Fusarium* root rot incidence and root system architecture in grafted common bean lines. *Plant Soil* 300, 233–244. doi: 10.1007/s11104-007-9408-0

Coleman, J. J. (2016). The *Fusarium solani* species complex: ubiquitous pathogens of agricultural importance. *Mol. Plant Pathol.* 17, 146–158. doi: 10.1111/mpp.12289

Conner, R. L., Hou, A., Balasubramanian, P., McLaren, D. L., Henriquez, M. A., Chang, K. F., et al. (2014). Reaction of dry bean cultivars grown in western Canada to root rot inoculation. *Canadian J. Plant Sci.* 94, 1219–1230. doi: 10.4141/cjps2013-416

Domhof, S., and Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*, Vol. 406. Hoboken, NJ: Wiley-Interscience.

Ellinger, D., Naumann, M., Falter, C., Zwikowics, C., Jamrow, T., Manisseri, C., et al. (2013). Elevated early callose deposition results in complete penetration resistance to powdery mildew in *Arabidopsis*. *Plant Physiol.* 161, 1433–1444. doi: 10.1104/pp.112.211011

Ellinger, D., and Voigt, C. A. (2014). Callose biosynthesis in *Arabidopsis* with a focus on pathogen response: what we have learned within the last decade. *Ann. Bot.* 114, 1349–1358. doi: 10.1093/aob/mcu120

Ellis, J., Dodds, P., and Pryor, T. (2000). The generation of plant disease resistance gene specificities. *Trends Plant Sci.* 5, 373–379. doi: 10.1016/s1360-1385(00)01694-0

Figueiredo, A., Monteiro, F., and Sebastiana, M. (2014). Subtilisin-like proteases in plant–pathogen recognition and immune priming: a perspective. *Front. Plant Sci.* 5:739. doi: 10.3389/fpls.2014.00739

Fukunaga, S., Sogame, M., Hata, M., Singkaravanit-Ogawa, S., Piślewska-Bednarek, M., Onozawa-Komori, M., et al. (2017). Dysfunction of *Arabidopsis* MACPF domain protein activates programmed cell death via tryptophan metabolism in MAMP-triggered immunity. *Plant J.* 89, 381–393. doi: 10.1111/tpj.13391

Geddy, R., and Brown, G. G. (2007). Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* 8:130. doi: 10.1186/1471-2164-8-130

Gindro, K., Berger, V., Godard, S., Voinesco, F., Schnee, S., Viret, O., et al. (2012). Protease inhibitors decrease the resistance of Vitaceae to *Plasmopara viticola*. *Plant Physiol. Biochem.* 60, 74–80. doi: 10.1016/j.plaphy.2012.07.028

Gossen, B. D., Conner, R. L., Chang, K. F., Pasche, J. S., McLaren, D. L., Henriquez, M., et al. (2016). Identifying and managing root rot of pulses on the northern great plains. *Plant Dis.* 100, 1965–1978. doi: 10.1094/PDIS-02-16-0184-FE

Guo, Y., and Gan, S. (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *Plant J.* 46, 601–612. doi: 10.1111/j.1365-313x.2006.02723.x

Hagerty, C. H., Cuesta-Marcos, A., Cregan, P. B., Song, Q., McClean, P., Noffsinger, S., et al. (2015). Mapping *Fusarium solani* and *Aphanomyces euteiches* root rot resistance and root architecture quantitative trait loci in common bean. *Crop Sci.* 55, 1969–1977. doi: 10.2135/cropsci2014.11.0805

Hickman, R., Hill, C., Penfold, C. A., Breeze, E., Bowden, L., Moore, J. D., et al. (2013). A local regulatory network around three NAC transcription factors in stress responses and senescence in *Arabidopsis* leaves. *Plant J.* 75, 26–39. doi: 10.1111/tpj.12194

Katan, J. (2017). Diseases caused by soilborne pathogens: biology, management and challenges. *J. Plant Pathol.* 99, 305–315.

Knutsen, A. K., Torp, M., and Holst-Jensen, A. (2004). Phylogenetic analyses of the *Fusarium poae*, *Fusarium sporotrichioides* and *Fusarium langsethiae* species complex based on partial sequences of the translation elongation factor-1 alpha gene. *Int. J. Food Microbiol.* 95, 287–295. doi: 10.1016/j.ijfoodmicro.2003.12.007

Lamichhane, J. R., and Venturi, V. (2015). Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Front. Plant Sci.* 6:385. doi: 10.3389/fpls.2015.00385

Leslie, J. F., Pearson, C. A., Nelson, P. E., and Toussoun, T. A. (1990). *Fusarium* spp. from corn, sorghum, and soybean fields in the central and eastern United States. *Ecol. Stud.* 44:66.

Leslie, J. F., and Summerell, B. A. (2006). *The Fusarium Laboratory Manual*. Hoboken, NJ: John Wiley & Sons.

Lorenzo, O., Piqueras, R., Sánchez-Serrano, J. J., and Solano, R. (2003). ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *Plant Cell* 15, 165–178. doi: 10.1105/tpc.007468

Macedo, R., Sales, L. P., Yoshida, F., Silva-Abud, L. L., and Junior, M. L. (2017). Potential worldwide distribution of *Fusarium* dry root rot in common beans based on the optimal environment for disease occurrence. *PloS One* 12:e0187770. doi: 10.1371/journal.pone.0187770

Mamidi, S., Rossi, M., Annam, D., Moghaddam, S., Lee, R., Papa, R., et al. (2011). Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38, 953–967.

Matuo, T., and Snyder, W. C. (1973). Use of morphology and mating populations in the identification of Formae speciales in *Fusarium solani*. *Phytopathol* 63, 562–565.

McClean, P. E., Bett, K., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., et al. (2018). White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytol.* 219, 1112–1123. doi: 10.1111/nph.15259

Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., et al. (2016). Genome-wide association study identifies candidate loci underlying agronomic traits in a middle American diversity panel of common bean. *Plant Genome* 9, 1–21. doi: 10.3835/plantgenome2016.02.0012

Morita-Yamamuro, C., Tsutsui, T., Sato, M., Yoshioka, H., Tamaoki, M., Ogawa, D., et al. (2005). The *Arabidopsis* gene CAD1 controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. *Plant Cell Physiol.* 46, 902–912. doi: 10.1093/pcp/pci095

Nakedde, T., Ibarra-Perez, F. J., Mukankusi, C., Waines, J. G., and Kelly, J. D. (2016). Mapping of QTL associated with *Fusarium* root rot resistance and root architecture traits in black beans. *Euphytica* 212, 51–63. doi: 10.1007/s10681-016-1755-6

Nicoli, A., Zambolim, L., Paula Júnior, T. J., Vieira, R. F., Teixeira, H., and Carneiro, J. E. S. (2012). Resistance of advanced common bean lines to *Fusarium* root rot. *Tropical Plant Pathol.* 37, 393–398. doi: 10.1590/s1982-56762012000600003

Nuruzzaman, M., Sharoni, A. M., and Kikuchi, S. (2013). Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Front. Microbiol.* 4:248. doi: 10.3389/fmicb.2013.00248

O'Donnell, K. (2000). Molecular phylogeny of the *Nectria haematococca-Fusarium solani* species complex. *Mycologia* 92, 919–938. doi: 10.1080/00275514.2000.12061237

O'Donnell, K., Sutton, D. A., Fothergill, A., McCarthy, D., Rinaldi, M. G., and Brandt, M. E. (2008). Molecular phylogenetic diversity, multilocus haplotype nomenclature, and in vitro antifungal resistance within the *Fusarium solani* species complex. *J. Clin. Microbiol.* 46, 2477–2490. doi: 10.1128/JCM.02371-07

Oladzad, A., Porch, T., Rosas, J. C., Moghaddam, S. M., Beaver, J., Beebe, S. E., et al. (2019a). Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. *G3 Genes Genomes Genet.* 9, 1881–1892. doi: 10.1534/g3.119.400072

Oladzad, A., Zitnick-Anderson, K., Jain, S., Simons, K., Osorno, J. M., McClean, P. E., et al. (2019b). Genotypes and genomic regions associated with *Rhizoctonia solani* resistance in common bean. *Front. Plant Sci.* 10:956. doi: 10.3389/fpls.2019.00956

Pearce, G., Yamaguchi, Y., Barona, G., and Ryan, C. A. (2010). A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14921–14925. doi: 10.1073/pnas.1007568107

Peña, P. A., Steadman, J. R., Eskridge, K. M., and Urrea, C. A. (2013). Identification of sources of resistance to damping-off and early root/hypocotyl damage from *Rhizoctonia solani* in common bean (*Phaseolus vulgaris* L.). *Crop Prot.* 54, 92–99. doi: 10.1016/j.cropro.2013.04.014

Porch, T. G., Valentin, S., Estevez de Jensen, C., and Beaver, J. S. (2014). Identification of soil-borne pathogens in a common bean root rot nursery in Isabela, Puerto Rico. *J. Agric. Univ. Puerto Rico* 98. doi: 10.46429/jaupr.v98i1.216

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847

Román-Avilés, B., and Kelly, J. D. (2005). Identification of quantitative trait loci conditioning resistance to *Fusarium* root rot in common bean. *Crop Sci.* 45, 1881–1890. doi: 10.2135/cropsci2005.0028

Rubiales, D., Fondevilla, S., Chen, W., Gentzbittel, L., Higgins, T. J. V., Castillejo, M. A., et al. (2015). Achievements and challenges in legume breeding for pest and disease resistance. *Crit. Rev. Plant Sci.* 34, 195–236. doi: 10.1080/07352689.2014.898445

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Gent.* 46, 707–713. doi: 10.1038/ng.3008

Schneider, K. A., Grafton, K. F., and Kelly, J. D. (2001). QTL analysis of resistance to *Fusarium* root rot in bean. *Crop Sci.* 41, 535–542. doi: 10.2135/cropsci2001.412535x

Shah, D. A., and Madden, L. V. (2004). Nonparametric analysis of ordinal data in designed factorial experiments. *Phytopathol* 94, 33–43. doi: 10.1094/PHYTO.2004.94.1.33

Singh, S. P., Gutrierrez, J. A., Molina, A., Urrea, C., and Gepts, P. (1991). Genetic diversity in cultivated common bean: II marker-based analysis of morphological and agronomic traits. *Crop Sci.* 31, 23–29. doi: 10.2135/cropsci1991.0011183x003100010005x

Soltani, A., MafiMoghaddam, S., Oladzad-Abbasabadi, A., Walter, K., Kearns, P. J., Vasquez-Guzman, J., et al. (2018). Genetic analysis of flooding tolerance in an Andean diversity panel of dry bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 9:767. doi: 10.3389/fpls.2018.00767

Soltani, A., MafiMoghaddam, S., Walter, K., Restrepo-Montoya, D., Mamidi, S., Schroder, S., et al. (2017). Genetic architecture of flooding tolerance in the dry bean Middle-American diversity panel. *Front. Plant Sci.* 8:1183. doi: 10.3389/fpls.2017.01183

Strock, C. F., Burridge, J., Massas, A. S. F., Beaver, J., Beebe, S., and Camilo, S. A. (2019). Seedling root architecture and its relationship with seed yield across diverse environments in *Phaseolus vulgaris*. *Field Crops Res.* 237, 53–64. doi: 10.1016/j.fcr.2019.04.012

Sun, G., Zhu, C., Kramer, M. H., Yang, S. S., Song, W., Piepho, H. P., et al. (2010). Variation explained in mixed-model association mapping. *Heredity* 105, 333–340. doi: 10.1038/hdy.2010.11

Teixeira, H., Júnior, T. J. P., Vieira, R. F., Lehner, M. S., Lima, R. C., and Del Ponte, E. M. (2015). Seasonal dynamics of soil-borne inoculum and severity of *Fusarium* root rot of common beans affected by sequential planting of legume or cereal crops. *Tropical Plant Pathol.* 40, 335–338. doi: 10.1007/s40858-015-0047-3

Tornero, P., Conejero, V., and Vera, P. (1996). Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: similarity of functional domains to subtilisin-like endoproteases. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6332–6337. doi: 10.1073/pnas.93.13.6332

Tuite, J. (1969). *Plant Pathological Methods: Fungi and Bacteria*. Minneapolis MN: Burgess Publishing Company, 25.

Van der Ent, S., and Pieterse, C. M. (2018). Ethylene: multi-tasker in plant–attacker interactions. *Annu. Plant Rev. Online* 44, 343–377. doi: 10.1002/9781119312994.apr0485

van Loon, L. C., Geraats, B. P., and Linthorst, H. J. (2006). Ethylene as a modulator of disease resistance in plants. *Trends Plant Sci.* 11, 184–191. doi: 10.1016/j.tplants.2006.02.005

Van Schoonhoven, A., and Pastor-Corrales, M. (1987). *Standard System for the Evaluation of Bean Germplasm*. Cali: Centro Internacional de Agricultura Tropical, 36–38.

VanEtten, H. D. (1978). Identification of additional habitats of *Nectria haematococca* mating population VI. *Phytopathol.* 68, 1552–1556.

Vasquez-Guzman, J. (2016). *Identifying Dry Bean Genotypes and Genomic Regions Associated With Root rot Resistance With Emphasis on Fusarium solani f. sp. Phaseoli*. PhD thesis, North Dakota State University, Fargo ND.

Voigt, C. A. (2014). Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. *Front. Plant Sci.* 5:168. doi: 10.3389/fpls.2014.00168

Wang, W., Jacobs, J. L., Chilvers, M. I., Mukankusi, C. M., Kelly, J. M., and Cichy, K. A. (2018). QTL analysis of *Fusarium* root rot resistance in an andean X Middle American common bean RIL population. *Crop Sci.* 58, 1166–1180. doi: 10.2135/cropsci2017.10.0608

Wong, F. P., and Wilcox, W. F. (2000). Distribution of baseline sensitivities to azoxystrobin among isolates of *Plasmopara viticola*. *Plant Dis.* 84, 275–281. doi: 10.1094/PDIS.2000.84.3.275

Yu, D., Chen, C., and Chen, Z. (2001). Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *Plant Cell* 13, 1527–1540. doi: 10.1105/tpc.010115

Yuan, X., Wang, H., Cai, J., Li, D., and Song, F. (2019). NAC transcription factors in plant immunity. *Phytopathol. Res.* 1:3.

Zhang, L., Li, L., Zhu, Y., Zhang, G., and Guo, X. (2014). Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). *Marine Biotechnol.* 16, 17–33. doi: 10.1007/s10126-013-9526-z

Zhao, J. H. (2007). Gap: genetic analysis package. *J. Stat. Softw.* 8, 1–18.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genet.* 44, 821–824. doi: 10.1038/ng.2310

Zuiderveen, G. H., Padder, B. A., Kamfwa, K., Song, Q., and Kelly, J. D. (2016). Genome-wide association study of anthracnose resistance in andean beans (*Phaseolus vulgaris*). *PLoS One* 11:e0156391. doi: 10.1371/journal.pone.0156391

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zitnick-Anderson, Oladzadabbasabadi, Jain, Modderman, Osorno, McClean and Pasche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 July 2020
doi: 10.3389/fgene.2020.00773

Development and Application of SSR Markers Related to Genes Involved in Leaf Adaxial-Abaxial Polarity Establishment in Chinese Cabbage (*Brassica rapa* L. ssp. *pekinensis*)

Ying Gao[1,2†], Yin Lu[1†], Xiaoguang Li[3], Na Li[1], Xiaomeng Zhang[1], Xiangjie Su[1], Daling Feng[1], Mengyang Liu[1], Shuxin Xuan[1], Aixia Gu[1], Yanhua Wang[1], Xueping Chen[1], Jianjun Zhao[1] and Shuxing Shen[1]**

[1] State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory of Vegetable Germplasm Innovation and Utilization of Hebei, Collaborative Innovation Center of Vegetable Industry in Hebei, College of Horticulture, Hebei Agricultural University, Baoding, China, [2] Agriculture and Rural Affairs Bureau of Xindu District, Xingtai, China, [3] Institute of Forestry and Fruits, Xingtai Academy of Agricultural Sciences, Xingtai, China

In Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*), leaf adaxial-abaxial (ad-ab) polarity is tightly related to leaf incurvature, an essential factor for the formation of leafy heads. Therefore, identification of the genes responsible for leaf ad-ab polarity and studying their genetic variation may clarify the mechanism of leafy head formation. By comparing the sequences of the genes regulating leaf ad-ab polarity development in *Arabidopsis thaliana* (*A. thaliana*), 41 candidate genes distributed on 10 chromosomes were found to be responsible for the establishment of ad-ab polarity in Chinese cabbage. Orthologous genes, including 10 single copies, 14 double copies, and one triple copies, were detected in the Chinese cabbage. The gene structure and conserved domain analyses showed that the number of exons of the 41 candidate genes range from one to 25, and that most genes share the conserved motifs 1, 6, and 10. Based on the 41 candidate genes, 341 simple sequence repeats (SSRs) were detected, including five replicated types: single, double, triple, quintuple, and sextuple nucleotide replications. Among these sequence repeat (SSR) loci, 323 loci were used to design 969 specific primers, and 362 primer pairs were selected randomly and evaluated using 12 Chinese cabbage accessions with different heading types. 23 primer pairs resulting with clear, polymorphic bands, combined with other 127 markers, was used to construct a linkage map by using an F_2 population containing 214 lines derived from the hybrid of the overlapping heading Chinese cabbage "14Q-141" and the outward curling heading Chinese cabbage "14Q-279." The result showed that the sequences of markers in the genetic linkage map and the physical map was consistent in general. Our study could help to accelerate the breeding process of leafy head quality in Chinese cabbage.

Keywords: Chinese cabbage, leaf ad-ab polarity, SSR, genetic linkage map, marker-assisted selection

OPEN ACCESS

Edited by:
Arron Carter, Washington State University, United States

Reviewed by:
Xiaoming Song, North China University of Science and Technology, China
Xiaowu Wang, Biotechnology Research Institute (CAAS), China

***Correspondence:**
Jianjun Zhao
jjz1971@aliyun.com
Shuxing Shen
shensx@hebau.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *10 April 2020*
Accepted: *30 June 2020*
Published: *23 July 2020*

Citation:
Gao Y, Lu Y, Li X, Li N, Zhang X, Su X, Feng D, Liu M, Xuan S, Gu A, Wang Y, Chen X, Zhao J and Shen S (2020) Development and Application of SSR Markers Related to Genes Involved in Leaf Adaxial-Abaxial Polarity Establishment in Chinese Cabbage (Brassica rapa L. ssp. pekinensis). Front. Genet. 11:773. doi: 10.3389/fgene.2020.00773

INTRODUCTION

Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is an important leafy vegetable grown worldwide and one of the most consumed vegetables in Asia. With the improvement of life quality, the leafy head appearance of Chinese cabbage is of increasing concern to both the consumers and breeders (Mao et al., 2014; Liang et al., 2016). Chinese cabbage goes through three developmental stages to produce a leafy head, namely, seedling, rosette, and heading (Yu et al., 2013). The leaves grow flat in the seedling and rosette stages, whereas in the heading stage they curve inwardly and show a large abaxial surface. Leaf incurvature is influenced by leaf ad-ab polarity, and is a precondition for the formation of leafy heads (Mao et al., 2014; Liang et al., 2016). Therefore, studies on leaf ad-ab polarity is helpful for improving the commercial traits of Chinese cabbage (Yu et al., 2013; Mao et al., 2014).

Leaf heading is a complicated quantitative trait controlled by various of genes (Yu et al., 2013; Wang et al., 2014). With the completion of the whole genome sequencing of the model plant *A. thaliana* and a variety of other plants (Initiative, 2000; Wang et al., 2011), as well as the deep research in molecular biology and genetics, it has been found that the regulatory network controlling ad-ab polarity mainly involves abaxial patterning genes, adaxial patterning genes, *WUSCHEL-RELATED HOMEOBOX* (*WOX*) genes, *YABBY* genes, and small RNAs genes (Yamaguchi et al., 2012). In angiosperm models, the adaxial polarity is regulated by genes of the class III homeodomain-leucine zipper (HD-ZIPIII) family [*REVOLUTA* (*REV*), *ATHB8*, *PHAVOLUTA* (*PHV*), and *PHB*] (Mcconnell et al., 2001; Emery et al., 2003), Myb, and LOB domain transcription factors *ASYMMETRIC LEAVES1* (*AS1*) and *AS2* (Lin et al., 2003; Iwakawa et al., 2007). In contrast, the abaxial polarity is specified by the *AUXIN RESPONSE FACTORS* (*ARF3* and *ARF4*) (Pekker et al., 2005), *KANADI* family genes (*KAN1*, *KAN2*, and *KAN3*) (Eshed et al., 2001, 2004; Kerstetter et al., 2001), and small RNAs miR165/166 (Kidner and Martienssen, 2004). *WUS-related homeobox* (*WOX*) genes (*WOX1*, *WOX3*), and *YABBY* genes are involved in adaxial/abaxial patterning and subsequent flat leaf growth (Kidner and Timmermans, 2007; Vandenbussche et al., 2009; Nakata et al., 2012). In addition, leaf ad-ab polarity is also regulated by microRNAs and tasiRNAs, which include *ARGONAUTE1* (*AGO1*), *ARGONAUTE7* (*AGO7*), RNA-*DEPENDENT RNA POLYMERASE 6* (*RDR6*), *SUPPRESSOR OF GENE SILENCING 3* (*SGS3*), *SERRATE* (*SE*), *DICER-LIKE1* (*DCL1*) and *DICER-LIKE4* (*DCL4*) (Moon and Hake, 2011). Furthermore, 45 genes (Version 1.5) involved in leaf adaxial-abaxial polarity establishment were detected in Chinese Cabbage (Liang et al., 2016). Chinese cabbage and *Arabidopsis thaliana* both belong to Cruciferae, giving the collinearity of the two genomes, most genes involved in leaf ad-ab polarity establishment are concluded to be conserved between these two genome.

Developing and utilizing molecular markers related to leaf ad-ab polarity establishment may lead to a better marker-assisted selection in breeding. SSR or microsatellites, are iterations of 1–6 bp nucleotide motifs. SSR molecular markers are widely distributed in genomes, have a high level of polymorphisms, are inherited co-dominantly and can be easily analyzed by PCR. Therefore, it becomes one of the most popular molecular markers and have been used widely to construct genetic linkage map, indentify varieties, and carry out diverse genetic analyses (Zhao et al., 2008; Gao et al., 2012). With the accomplishment of genome sequencing in Chinese cabbage (Wang et al., 2011), it is possible to develop the molecular markers related to leaf ad-ab polarity based on the information of regulators from *Arabidopsis*.

In this study, we developed specific SSR markers that correlate with the establishment of leaf ad-ab polarity. The repeat units and distribution characteristics of these SSR loci as well as the amplification efficiency and polymorphism level were analyzed. Our study will benefit to clarify the molecular mechanism of leafy head formation, and help to accelerate breeding process in leafy head quality in Chinese cabbage.

MATERIALS AND METHODS

Plant Materials

An F_2 population was used as the mapping population in this study. The female parent of the population was the overlapping heading Chinese cabbage "14Q-141," and the male parent was the outward curling heading Chinese cabbage "14Q-279" (**Figure 1**). F_1 was obtained from a cross between "14Q-141" and "14Q-279." 214 F_2 plants were gained after F_1 selfing. The experiment was carried out at Hebei Agricultural University in Hebei, China. The seeds of the 214 F_2 and their parental lines were sown in greenhouse and the seedlings were transplanted to an open field in September 2016, and kept growing until November 2016 in natural conditions.

The Sequences Updates of Genes Responsible for Leaf ad-ab Polarity in Chinese Cabbage

The sequences of genes responsible for leaf ad-ab polarity in Chinese cabbage was published in version 1.5 (Liang et al., 2016). The BRAD database[1] (Cheng et al., 2011) was used to update the sequences information of these genes from version 1.5 to version 3.0. **Table 1** showed the genes responsible for the establishment of leaf ad-ab polarity in Chinese cabbage.

Motif Display and Phylogenetic Analyses of Genes Responsible for Leaf ad-ab Polarity in Chinese Cabbage

To search the conserved motifs of the proteins, the amino acid sequences of the 41 genes responsible for leaf ad-ab polarity in Chinese cabbage were uploaded and analyzed using the online tool MEME Suite 5.1.1[2] (Bailey et al., 2009). The amino acid sequences of these genes in *A. thaliana* were downloaded from the TAIR database[3] (Rhee et al., 2003). The ClustalW 2.0

[1]http://brassicadb.org/brad/

[2]http://meme-suite.org/

[3]http://www.arabidopsis.org/

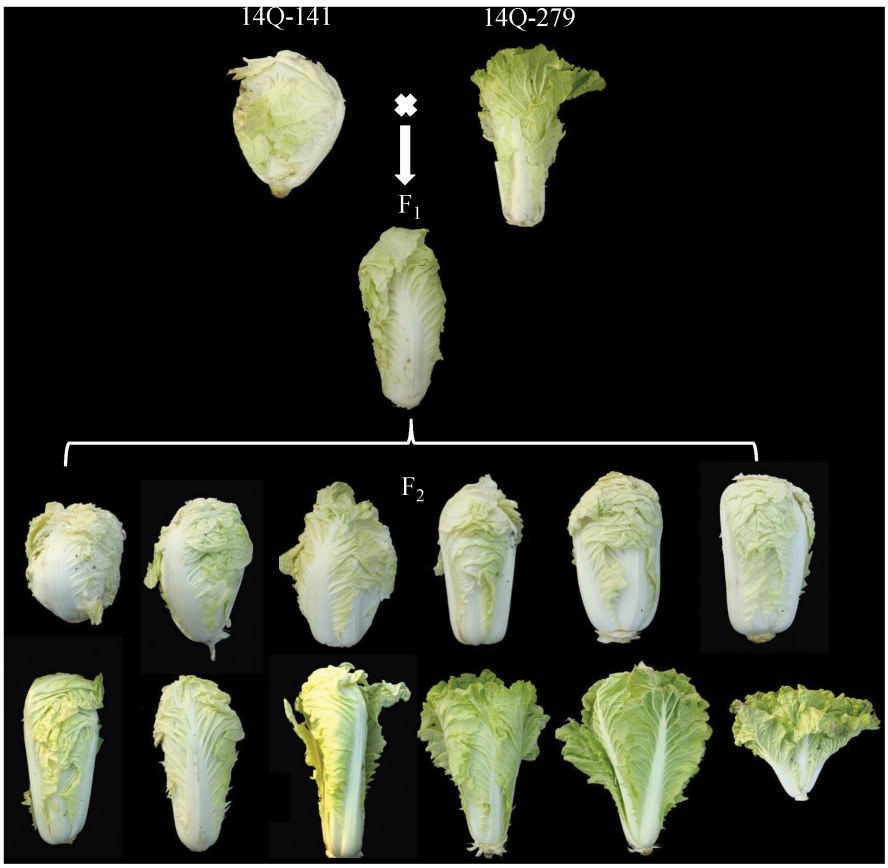


FIGURE 1 | Parents, F_1 and representatives in F_2 population.

(Larkin et al., 2007) and MEGA6.0 (Tamura et al., 2013) software were used to construct the phylogenetic tree. The analysis of the conserved motifs was conducted using Pfam[4] (Sara et al., 2019) and SMART (Letunic et al., 2009). The analyses of collinearity and visualization for the genes responsible for leaf ad-ab polarity were done using MCScan X (Wang et al., 2012) and TBtoolse[5] (Chen et al., 2020), respectively.

SSR Identification and Primer Design

Based on the target gene and its upstream (5 kb) and downstream (5 kb) sequences (Gong et al., 2014). SSR primers were identified and located by MIcroSAtellite (MISA) (Thiel et al., 2003) software with Perl. The search criteria was set to be as follow: ≥ten repeat units for mononucleotides, ≥seven repeat units for dinucleotides, and ≥five repeat units for tri-, tetra-, penta-, and hexanucleotides. The interrupted compound SSRs were also listed as search targets when the interval was less than or equal to 10 bp. Primer pairs were designed based on the sequences of the identified SSR using Primer 3.0 with Perl. The parameters of Primer 3.0 were set as follows: (a) melting temperature between 55°C and 65°C; (b) GC content between 40% and 60%; (c) primer length between 18 and 27 bases; and (d) PCR products length between 100 and 300 bp. The other parameters were set with default values. 969 SSR primer pairs were designed in all.

DNA Extraction

DNA was extracted from young leaves of the F_2 population by the CTAB method (Murray and Thompson, 1980; Rogers and Bendich, 1985). 5–10 μL genomic DNA was used to assess the sample quality using 1.0% agarose gel. A NanoDrop2000 spectrometer was used to assess the DNA quality and concentration.

Assessment of SSR Polymorphisms

The validation of polymorphic primers was done using 12 Chinese cabbage accessions with different heading traits selected from the F_2 population. 362 pairs of SSR primers were selected from the newly designed primers and were used to detect the SSR polymorphisms among the 12 accessions. The primers were synthesized by Sangon Biotech, Shanghai, China. A total volume of 10.0 μL was used to perform PCR, including 1.0 μL of genomic DNA (50 ng/μL), 0.8 μl of dNTPs (2.5 mmol/L), 0.5 μL of each of forward and reverse primers (50 ng/μL), 1 μL of 10 × PCR buffer (Mg_2^+), 0.1 μL of Taq DNA polymerase (2.5 U/μL, TaKaRa, Dalian, China), and 6.1 μL

[4]http://pfam.xfam.org/

[5]https://github.com/CJ-Chen/TBtools

TABLE 1 | Leaf ad-ab polarity–controlling genes in Chinese cabbage.

Regulation pathway	Gene	*Arabidopsis thaliana* genes	Chinese cabbage genes	Location in Linkage group
Adaxial determinants	*REV*	AT5G60690	*BrREV1* (BraA10g018460.3C)	chr10: 13671299..13675382
			BrREV2 (BraA02g010200.3C)	chr02: 4908251..4904176
	PHB	AT2G34710	*BrPHB1* (BraA05g010360.3C)	chr05: 5534983..5539344
			BrPHB2 (BraA04g024760.3C)	chr04: 17922205..17918117
	PHV	AT1G30490	*BrPHV* (BraA09g034560.3C)	chr09: 27178071..27182489
	ATHB8	AT4G32880	*BrHB8.1* (BraA01g005160.3C)	chr01: 2462816..2467248
			BrHB8.2 (BraA08g016820.3C)	chr08: 13619843..13624628
	AS1	AT2G37630	*BrAS1.1* (BraA05g007920.3C)	chr05: 4047371..4048456
			BrAS1.2 (BraA03g019630.3C)	chr03: 9309263..9308220
	AS2	AT1G65620	*BrAS2* (BraA02g016770.3C)	chr02: 8953550..8954158
Middle domain determinants	*YAB1*	AT2G45190	*BrYAB1.1* (BraA04g030760.3C)	chr04: 21056868..21059180
			BrYAB1.2 (BraA03g023630.3C)	chr03: 11342365..11340257
	YAB2	AT1G08465	*BrYAB2.1* (BraA06g005620.3C)	chr06: 3309705..3311007
			BrYAB2.2 (BraA08g033150.3C)	chr08: 21874379..21873237
			BrYAB2.3 (BraA09g062760.3C)	chr09: 43480055..43477632
	YAB3	AT4G00180	–	–
	YAB5	AT2G26580	*BrYAB5* (BraA03g025390.3C)	chr03: 12518810..12521194
	WOX1	AT3G18010	*BrWOX1.1* (BraA05g029540.3C)	chr05: 21857860..21859470
			BrWOX1.2 (BraA03g038230.3C)	chr03: 18917799..18915771
	WOX3	AT2G28610	*BrWOX3.1* (BraA04g020450.3C)	chr04: 15389303..15390390
			BrWOX3.2 (BraA03g024810.3C)	chr03: 12196864..12195825
Abaxial determinants	*KAN1*	AT2G28610	*BrKAN1* (BraA10g023200.3C)	chr10: 16161361..16168134
	KAN2	AT1G32240	*BrKAN2.1* (BraA09g032840.3C)	chr09: 25471282..25477168
			BrKAN2.2 (BraA05g023490.3C)	chr05: 17380156..17384804
	KAN3	AT4G17695	*BrKAN3.1* (BraA01g009320.3C)	chr01: 4661501..4658601
			BrKAN3.2 (BraA08g012540.3C)	chr08: 10833475..10830673
	ARF3	AT2G33860	*BrARF3.1* (BraA05g011080.3C)	chr05: 5950051..5952569
			BrARF3.2 (BraA04g024390.3C)	chr04: 17725613..17723081
	ARF4	AT5G60450	*BrARF4.1* (BraA10g018230.3C)	chr10: 13539330..13535905
			BrARF4.2 (BraA02g010280.3C)	chr02: 5008076..5008416
Small RNA pathway	*AGO1*	AT1G48410	*BrAGO1.1* (BraA05g020200.3C)	chr05: 13144642..13150296
	AGO7	AT1G69440	*BrAGO7* (BraA07g030370.3C)	chr07: 22224668..22227518
	AGO10	AT5G43810	*BrAGO10.1* (BraA06g043560.3C)	chr06: 28421795..28417452
			BrAGO10.2 (BraA09g019960.3C)	chr09: 12912026..12912769
	SGS3	AT5G23570	*BrSGS* (BraA06g031000.3C)	chr06: 21377862..21380043
	RDR6	AT3G49500	*BrRDR6* (BraA01g023640.3C)	chr07: 13537572..13541477
	HYL1	AT1G09700	*BrHYL1.1* (BraA06g006590.3C)	chr06: 3729485..3727977
			BrHYL1.2 (BraA08g032470.3C)	chr08: 21597984..21599798
	DCL1	AT1G01040	*BrDCL1* (BraA10g000840.3C)	chr10: 428538.423842
	DCL4	AT5G20320	*BrDCL4* (BraA10g020250.3C)	chr10: 14712105.14721149
	SE	AT2G27100	*BrSE1* (BraA07g018000.3C)	chr07: 15424372.15427821
			BrSE2 (BraA04g019460.3C)	chr04: 14704649.14701441

H_2O. The reactions conditions were as follows: 3 min at 94°C; then 35 cycles of 45 s at 94°C, 30 s at the proper annealing temperature, 45 s at 72°C, and a final elongation 5 min at 72°C. After the PCR, the products were separated using 7% denaturing polyacrylamide gels and visualized by silver nitrate staining.

Genetic Map Construction

The markers to construct the F_2 genetic map were selected from 581 insertion-deletion (InDel) markers, 60 single nucleotide polymorphism (SNP) markers, 123 random SSR markers, and 362 ad-ab polarity related SSR markers. The genotypes of parents "CC-48" and "PC-101" were screened using the 60K *B. napus* array developed by array TraitGenetics (Germany), resulting in 5,795 polymorphic SNPs. The SNP markers were selected from the TraitGenetics dataset. The BRAD database were used to design the InDel markers and Primer Premier 5.0 (Lalitha, 2004) were used to design the random SSR markers. Native-PAGE method were used to analyze the genotypes of F_2 population with InDel and SSR markers. A 96-well LightScanner instrument was

used for SNP genotyping by high-resolution melting analysis of small amplicons.

The genotypes data were classified as type "a" or "b" based on whether they were the parents of "14Q-141" or "14Q-279," the undefined and missing data were showed by "−." JoinMap version 4.0 software was used to construct genetic maps for F_2 populations (Van Ooijen, 2006). After creating the population nodes, the markers were assigned into the linkage groups (LGs) based on the LOD value of 8.0–10.0. With the method of Kosambi, frequencies of recombination were transformed into centiMorgans (cM) to calculate genetic distance. Mapchart 2.32 was used to draw the map (Voorrips, 2002). A comparison map of genetic linkage distance and physical position was constructed using the ALLMAPS software[6] (Tang et al., 2015).

RESULTS

Classification and Collinearity Analysis of Genes

In *A. thaliana*, a total of 26 genes responsible for leaf ad-ab polarity were identified, including genes of transcription factors and small RNA pathways (Liang et al., 2016). By comparing with these *Arabidopsis* genes, 41 orthologs genes were detected in Chinese cabbage, including 10 transcription factors for adaxial determination, 10 transcription factors for middle domain determination, 9 transcription factors for abaxial determination, and 12 small RNAs for ad-ab polarity. According to the genomic sequence information of *A. thaliana* and Chinese cabbage, orthologous genes related to leaf ad-ab polarity establishment, except for AT4G00180, were all obtained in Chinese cabbage. These genes includes 10 single copies, 14 double copies, and one triple copies (**Figure 2**). A total of four single copies, 11 double copies, and one triple copies were detected in the transcription factor pathway, with the proportion of single copy to multiple copies being 1:3. In the small RNA pathway, a total of six single copies, three double copies, and no triple copies were detected, with the proportion of single copy to multiple copies being 2:1 (**Figure 3**). 60% of the genes exist as double or triple copies in Chinese cabbage.

[6]https://github.com/tanghaibao/jcvi/wiki/ALLMAPS

Distribution of the Genes Responsible for the Leaf ad-ab Polarity and the Characterization of SSR Loci in Chinese Cabbage

Genes responsible for the leaf ad-ab polarity are distributed on 10 chromosomes in Chinese cabbage, with most of them (6, 14.63%) occurring on chromosome A05 and only a very few (2, 4.88%) on chromosome A01. Genes on other chromosomes are distributed evenly. A total of 341 SSR loci were developed for the genes responsible for leaf ad-ab polarity in Chinese cabbage. Among the 41 genes, four contain 1–3 SSR loci, nine contain 4–6 SSR loci, 12 contain 7–9 SSR loci, 13 contain 10–12 SSR loci, and only three contain more than 12 SSR loci (**Figure 4**). 55 SSR loci are located within the genes. According to the principle of SSR primer design, 323 SSR loci were used to design the corresponding SSR primers. A total of 362 SSR primer pairs were selected for this study.

Structure and Motif Composition of Genes Responsible for Leaf ad-ab Polarity in Chinese Cabbage

The exon-intron organization of all the identified genes responsible for leaf ad-ab polarity was analyzed to get more insight into their evolution in Chinese cabbage. As shown in **Figure 5**, the length of the genes responsible for leaf ad-ab polarity is mostly within 6 kb, with only *BrDCL1* and *BrDCL4* reaching 7 and 9 kb, respectively. All the genes possess 1–25 exons. Genes *BrAS2*, *BrAS1.1*, and *BrAS1.2* have only one exon and genes *BrDCL4* has 25 exons. The distributions of exons and sequence length among these genes present obvious distinctions. A schematic representing the structure of four different regulatory pathway proteins is shown in **Figure 5**; the 30 motifs are named Motifs 1–30. Motifs 1, 6, and 10 are widely distributed domains. The members in the same groups are usually composed by a similar motif composition. For example, *BrREV*, *BrPHB*, *BrPHV*, and *BrATHB8* in the adaxial determinant pathway contain relatively more motifs and show a highly similar motif distribution. The number of motifs varies greatly. In adaxial determinant pathway, genes *BrPHV*, *BrPHB.1*, and *BrPHB.2* have the largest number of motifs (24), whereas gene *BrAS2* contains the fewest motifs, only Motif 4. Among the homologous

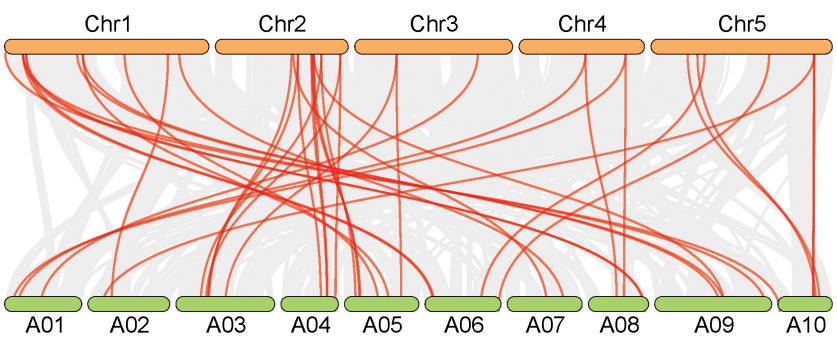


FIGURE 2 | Synteny analysis of leaf ad-ab polarity–controlling genes between Chinese cabbage and *Arabidopsis*.

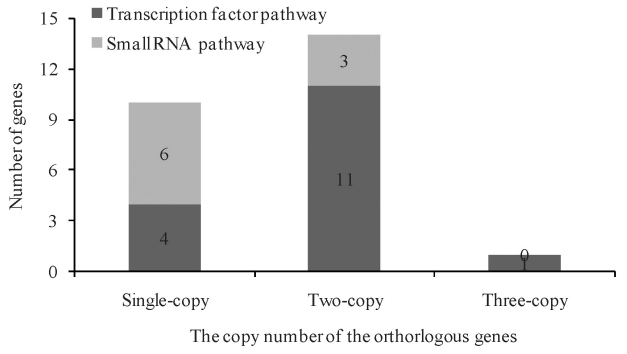


FIGURE 3 | The copy number of orthorlogous of leaf ad-ab polarity–controlling genes between *Arabidopsis* and Chinese cabbage.

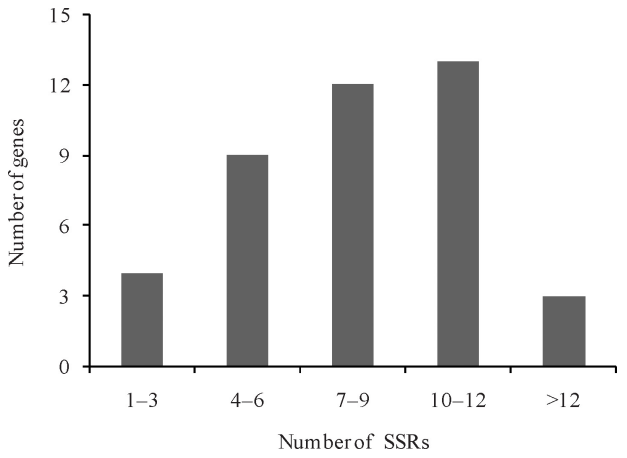


FIGURE 4 | Number of SSRs in leaf ad-ab polarity–controlling genes in Chinese cabbage.

genes *BrARF4.1*, *BrARF4.2*, *BrAGO10.1,* and *BrAGO10.2*, the conserved motifs of one gene are partially lost in another. A total of two loci are located in the START conserved domain, one in gene *BrPHV* and the other in *BrREV1*.

SSR Loci Analysis of the Genes Responsible for Leaf ad-ab Polarity in Chinese Cabbage

The repeat-type distributions of 341 SSR loci are shown in **Table 2**. Repeat types of SSR loci of the genes responsible for leaf ad-ab polarity are abundant in Chinese cabbage. Five repeat types are identified: mononucleotide (231) are the most abundant repeats (67.74%), and the following are dinucleotides (19.65%) and trinucleotides (9.38%); pentanucleotides and hexanucleotides are the fewest (0.30%, respectively). No tetranucleotides are found in the repeats. A/T, AT/TA, AAG, CAATG, and TAGATA are the predominant motifs. Mononucleotide repeats contain seven double-mononucleotide motifs, $(A)_{11}g(A)_{10}$, $(G)_{10}tgtagc(T)_{16}$, $(A)_{10}tttatc(T)_{11}$, $(T)_{15}(G)_{12}$, $(T)_{10}attttgg(A)_{10}$, $(A)_{10}gttcac(T)_{11}$, and $(A)_{10}tt(A)_{13}$. Dinucleotide repeats contain two double-dinucleotide motifs, $(TG)_{10}(TA)_8$ and $(TC)_9(TA)_8$.

Screening and Polymorphic Analysis of SSR Primers

362 pairs of primers were selected at random from the 969 pairs for evaluation using 12 Chinese cabbage accessions with different heading types. The results show that clear amplification products can be obtained by 213 (58.84%) primer pairs, of which 89 (24.59%) primer pairs resulted in polymorphic bands (**Supplementary Figure S1A**), and 68 (18.78%) primer pairs gave no polymorphism (**Supplementary Figure S1B**). In addition, 149 (41.16%) primer pairs failed to give any amplification products (**Supplementary Figure S1C**), and 56 (15.47%) primer pairs resulted in non-specific bands (**Supplementary Figure S1D**).

Among the 89 polymorphic SSR primers, 23 pairs with high polymorphism and clear bands were selected to analyze the 214 Chinese cabbage accessions. These SSR markers distribute on chromosomes A01, A02, A03, A04, A05, A06, A08, A09, and A10. These primers amplified 53 polymorphic bands, and the numbers of amplified fragments varied from two to four. On average, each primer amplifies 2.3 bands. Two alleles were detected by 17 SSR primers, three alleles by five SSR primers, and four alleles by one SSR primer. The primer S119 located on chromosome A05 detect the highest number of alleles (four). 78.6% of the accessions were amplified with one allele (**Supplementary Table S1**).

Genetic Map Construction

34 SSRs (23 specific SSR primers and 11 random SSR primers), 13 SNPs, and 103 InDel markers distributed over 10 linkage groups (LGs) were used to construct the linkage map. The total map length is 1747.57 cM with an average distance of 11.65 cM between adjacent markers (**Figure 6**). The most markers (23) were in linkage group (LG) A02 and the fewest (7) in LG A04. The length of each LG ranged from 90.32 to 235.61 cM. The biggest gap in the genetic map was 24.63 cM in LG A04, and the smallest gap was 5.65 cM in LG A05. The result showed that the sequences of markers in genetic linkage map and physical map are generally consistent, with only three groups of markers distribute in the LG A01, A05, and A08 showing different sequences (**Figure 7**).

DISCUSSION

Analysis of the Evolution of the Relationship Between SSR Loci, Gene Structure, and Conserved Motifs

During plant growth and development, the leaf shapes of different species and the same plant at different growth stages may show diverse morphology. It was found that the leaf ad-ab patterning can affect leaf curvature (Kidner and Timmermans, 2010; Yamaguchi et al., 2012). Based on 26 genes responsible for leaf ad-ab polarity in *A. thaliana*, we identified 41 orthologous genes in Chinese cabbage by a comparative genomic analysis. 75.6% of these genes contains double or triple copies in Chinese cabbage, indicating that they went through duplication and were

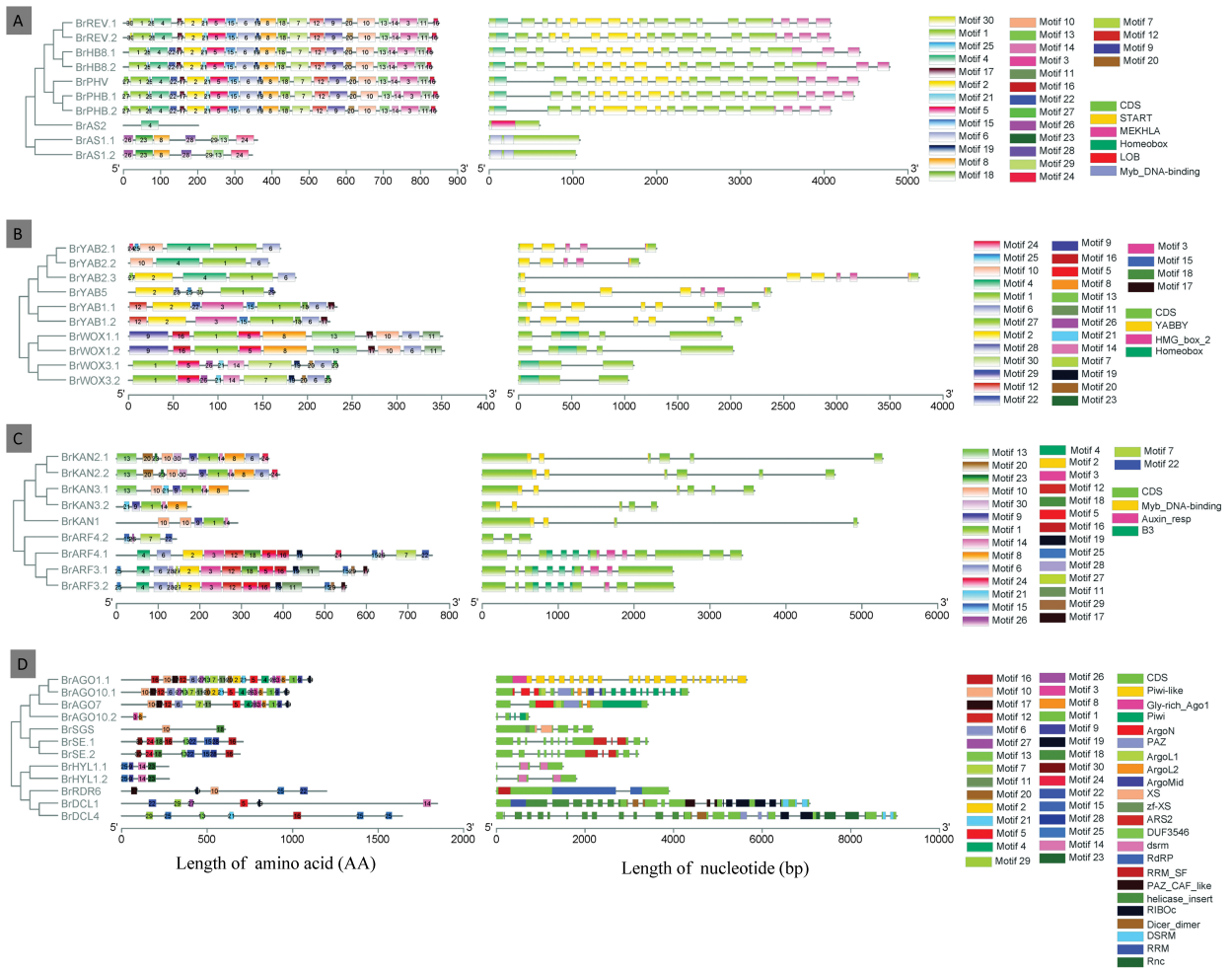


FIGURE 5 | The gene structure and motif analyses of leaf ad-ab polarity–controlling genes in Chinese cabbage. The exon-intron structure of leaf ad-ab polarity–controlling genes were showed on the left, and schematic diagram of amino acid motifs of leaf ad-ab polarity–controlling proteins were showed on the right. **(A)** Adaxial-controlling genes in Chinese cabbage; **(B)** Middle domain-controlling genes in Chinese cabbage; **(C)** Abaxial-determinant genes in Chinese cabbage; **(D)** Small RNA-controlling genes in Chinese cabbage.

reserved after whole genome triploidization. Multiple copy genes were also detected in the research of genes for glucosinolate (GS) biosynthesis in Chinese cabbage (Gong et al., 2014). However, among 12 orthologs genes detected in the small RNA pathway responsible for leaf ad-ab polarity, 42% genes have one or no copies in Chinese cabbage, and none has more than two copies. It was speculated that Chinese cabbage may have experienced a triplication of the whole genome and then undergone diploidy to delete one or two gene copies (Cheng et al., 2012; Tang et al., 2012). We inferred that there was a large amount of gene lost or rearranged during this process (Wang et al., 2011; Cheng et al., 2013). It remains to be further verified whether the leaf morphology establishment is affected by the different copies of leaf ad-ab polarity genes in various breeding materials.

Gene structure analysis is useful in mining the relationship between gene family evolution and gene duplication. The number and distribution of introns and exons may be related to gene evolution. We found that obvious differences exist in the distribution of exons and sequence length among four group of genes. We speculate that these genes may play distinct roles in leaf development. Through motif analysis, we found that Motifs

TABLE 2 | Type, number and frequency of SSRs from leaf ad-ab polarity–controlling genes in Chinese cabbage.

Repeat types	Types of repeat motif	Total number of repeat motif	Percentage(%)	Type of predominant motif
Mononucleotide	4	238	69.79	T/A
Dinucleotide	8	69	20.23	TA/AT
Trinucleotide	26	32	9.38	AAG
Tetranucleotide	0	0	0	–
Pentanucleotide	1	1	0.30	GAATG
Hexanucleotide	1	1	0.30	TAGATA
Total	40	341	100	

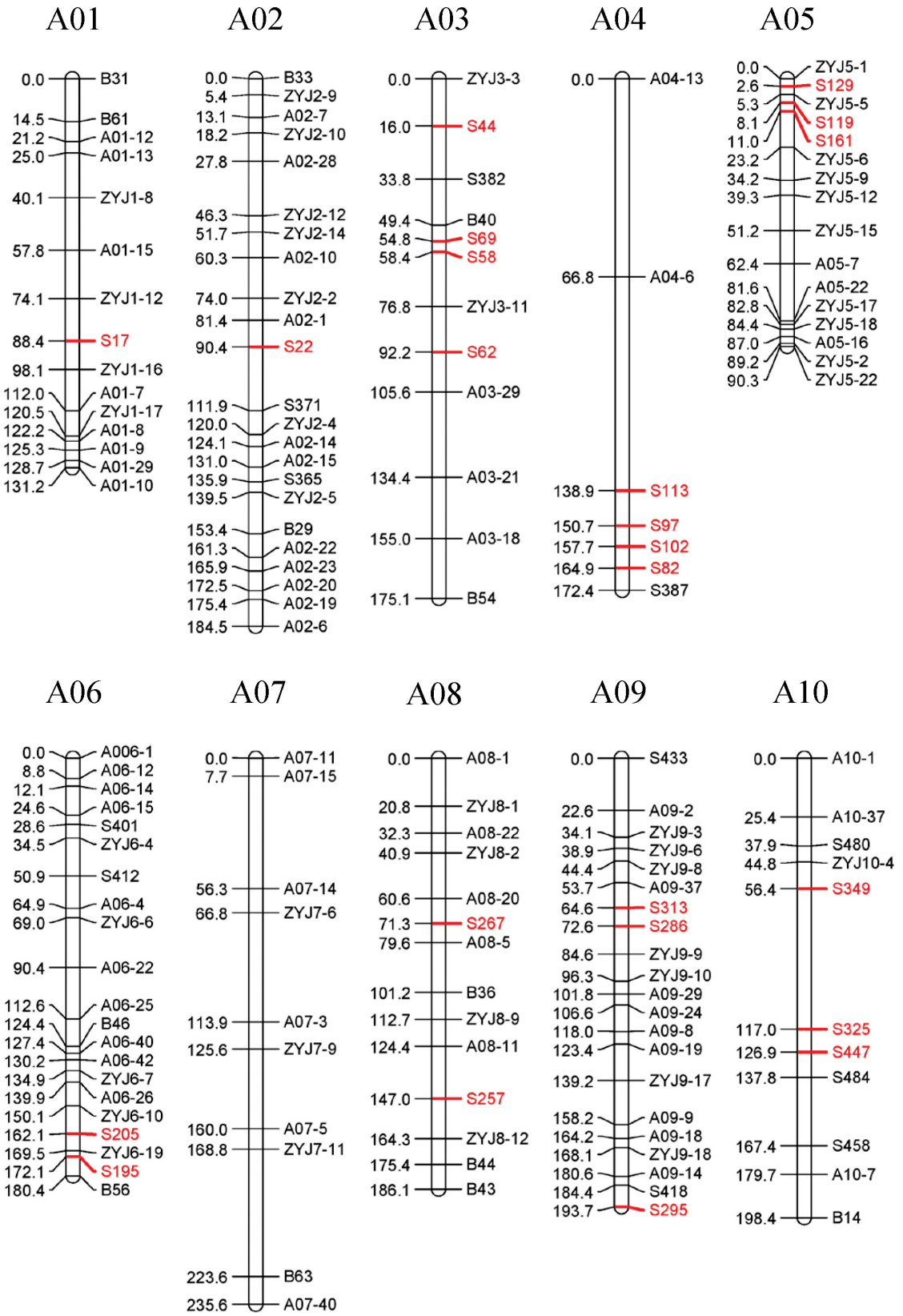


FIGURE 6 | A genetic linkage map of the F_2 population. Recombination distances (cM) are showed on the left, marker names on the right of each linkage group.

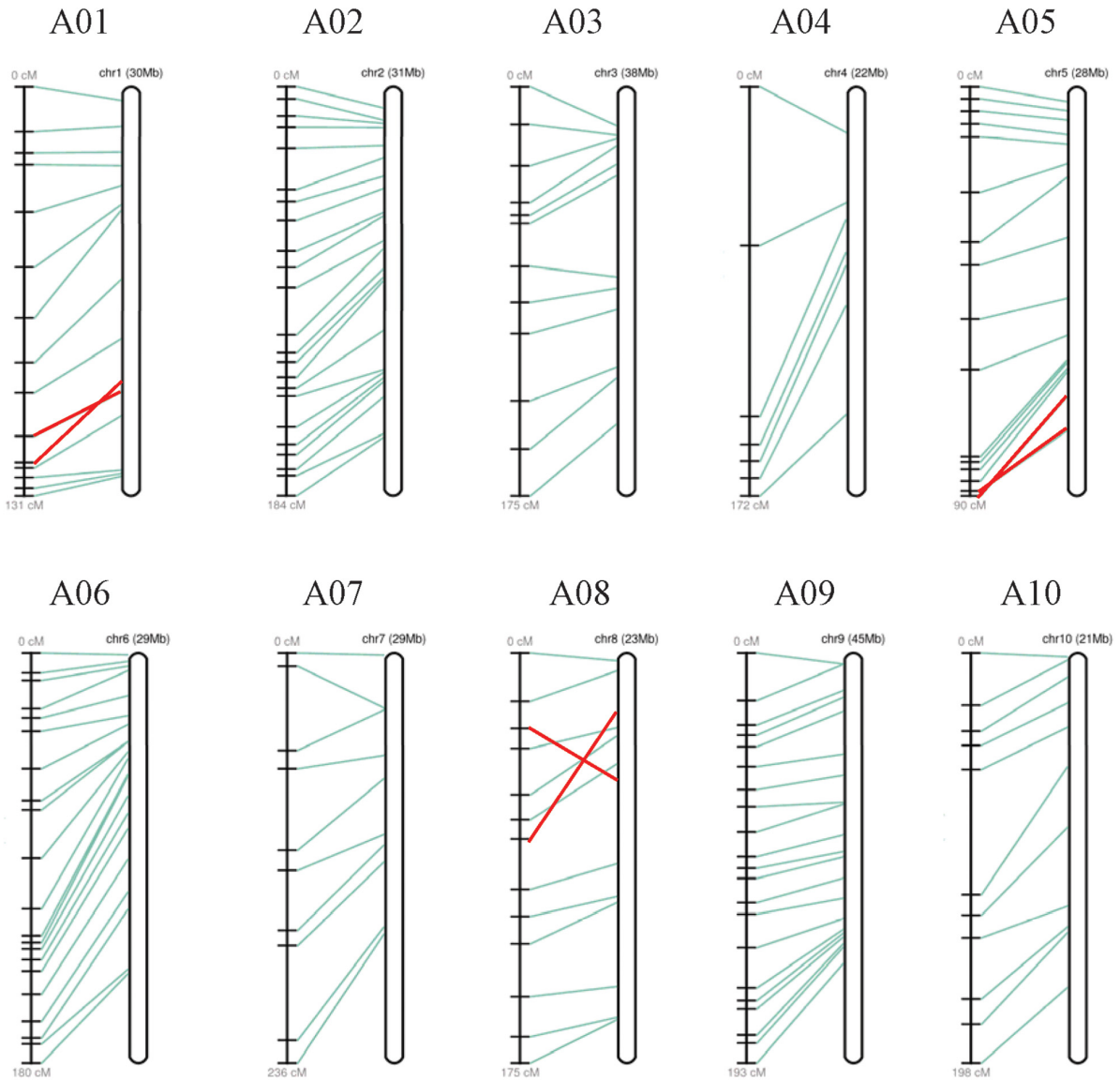


FIGURE 7 | A genetic linkage map and a physical map constructed based on F_2 population. In each group, recombination distances (cM) are showed on the left, physical distances (Mb) are showed on the right, and their correspondences are showed with lines in the middle.

1, 6, and 10 are widely distributed domains, suggesting that these domains may be very important for the gene function. Between the homologous genes *BrARF4.1*, *BrARF4.2*, *BrAGO10.1*, and *BrAGO10.2*, the conserved motifs of one gene was partially lost in another, suggesting a possibility of fragments lost during evolution. The striking dissimilarity between four groups of genes provides useful evidence for the study of genome duplication and phylogenetic evolution.

Development of Specific SSR Markers for Genes Controlling Leaf ad-ab Polarity

Simple sequence repeats markers are widely used because they are simple, rapid, low-cost, and have good repeatability. In molecular-assisted selection breeding, analyzing the diversity by using the SSR markers based on the target functional gene sequences or their upstream and downstream regions can overcome the problem of identification bias for non-gene region. In this study, 341 SSR loci were developed based on the 41 genes responsible for leaf ad-ab polarity, with the mononucleotide repeats (242, 70.96%) the most common and showing a strong bias toward A/T. 969 specific primers were designed from these SSR loci. These studies laid a foundation for further research on head-related traits in Chinese cabbage.

Application of Specific SSR Markers

Simple sequence repeats markers with wide genomic distribution and a high degree of polymorphism are widely used in genetic analyses. A total of 6 jujube cultivars and wild types were used

to verify the polymorphism of 1000 SSR primers; it was observed that, among the germplasms, 725 pairs (72.5%) were clear and effective and 511 pairs (51.1%) were polymorphic (Xiao et al., 2015). Eight Chinese cabbage inbred lines were used to screen the polymorphism of 77 SSR primers; it was shown that 49 SSR primers (63.64%) could amplify clear bands and 16 of them (20.8%) were polymorphic (Gong et al., 2014). In this study, 12 Chinese cabbage accessions with different heading types were used to evaluate the polymorphism of 362 SSR primer pairs, and 23 pairs (6.35%) resulting in high polymorphism and clear bands were selected to construct the linkage map. QTL analysis for 4 leafy head traits of Chinese cabbage was performed, which were head top shape (HTS), Head height (HH), Head Weight (HWe), Plant weight (PWe). The result showed that the sequences of markers in the linkage map were generally consistent with that of the physical map, 12 QTLs were mapped including 6 QTLs with 7 specific SSR coseparation, indicating that these specific SSR markers can be used in the further study and the candidate genes could be quickly indentified with them. However, the polymorphism of these SSRs was low, possibly because the genes responsible for leaf ad-ab polarity were highly conserved during evolution. Therefore, development of primers from other SSR loci of functional genes and other molecular markers such as InDel and SNP are still needed.

As the gene sequences are conserved between Chinese cabbage and other *Brassica* plants, SSR primers from Chinese cabbage are transferable to relative species. In our study, 55 SSR loci were located with the genes and two were in START conserved domain. The molecular genetic map constructed by these genes specific SSR markers, especially markers within the genes, can provide reference for linkage and QTL mapping of target genes among related species. Furthermore, as the reference markers, these SSR markers can be used to integrate the molecular genetic maps among relative species, thus providing a new method for comparative genomics research.

CONCLUSION

Establishment of leaf ad-ab polarity is tightly related to leaf incurvature in Chinese cabbage, which is a main factor result in the formation of a leafy head. In this study, by comparing the sequences of genes controlling leaf ad-ab polarity in *A. thaliana,* 41 candidate orthologous genes were found in Chinese cabbage. Based on these gene, 341 SSRs were detected and characterized. A genetic linkage map were constructed by SSR and other molecular markers screened. The result showed that the sequences of markers in genetic linkage and physical map was generally consistent, suggesting that these SSR markers were valuable for assisted-selective breeding in Chinese cabbage.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

YG and YL performed the research and wrote the manuscript. XL, NL, and XS surveyed the morphological characteristics. XZ and DF performed the assessment of SSR polymorphisms. ML, SX, and AG performed the motif display and phylogenetic analyses of genes. YW and XC performed the genetic map construction analysis. NL and JZ reviewed the manuscript. JZ and SS designed the research and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key R&D of China (Grant Nos. 2017YFD0101802 and 2016YFD0100204-17), the National Natural Science Foundation of China (Grant Nos. 31930098, 31902005, 31872100, and 31801857), the Natural Science Foundation of Hebei (C2018204135), the Science and Technology Support Program of Hebei (16226304D-2), and the International Cooperation Project in the Science and Technology Support Program of Hebei (Grant Nos. 17396315D and 2019YX023A).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00773/full#supplementary-material

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools - an integrative toolkit developed for interactive analyses of big biological data. *bioRxiv* [Preprint]. doi: 10.1101/289660

Cheng, F., Liu, S., Wu, J., Fang, L., Sun, S., Liu, B., et al. (2011). BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biol.* 11:136. doi: 10.1186/1471-2229-11-136

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M. A., and Wang, X. (2013). Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell* 25, 1541–1554. doi: 10.1105/tpc.113.110486

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., et al. (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7:e36442. doi: 10.1371/journal.pone.0036442

Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., et al. (2003). Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Curr. Biol.* 13, 1768–1774. doi: 10.1016/j.cub.2003.09.035

Eshed, Y., Baum, S. F., Perea, J. V., and Bowman, J. L. (2001). Establishment of polarity in lateral organs of plants. *Curr. Biol.* 11, 1251–1260. doi: 10.1016/s0960-9822(01)00392-x

Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., and Bowman, J. L. (2004). Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development* 131, 2997–3006. doi: 10.1242/dev.01186

Gao, Y., Luo, S. X., Wang, Y. H., Gu, A. X., Zhao, J. J., Chen, X. P., et al. (2012). Association analysis of bolting and flowering time with SSR and InDel markers in Chinese cabbage. *Acta Hortic. Sin.* 39, 1081–1089.

Gong, H. Y., Luo, S. X., Gu, A. X., Wang, Y. H., Zhao, J. J., and Shen, S. X. (2014). Development of SSR markers linked to glucosinolate synthesis genes in chinese cabbage (*Brassica rapa* ssp. pekinensis). *J. Agric. Biotechnol.* 22, 1232–1241.

Initiative, T. A. G. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692

Iwakawa, H., Iwasaki, M., Kojima, S., Ueno, Y., Soma, T., Tanaka, H., et al. (2007). Expression of the ASYMMETRIC LEAVES2 gene in the adaxial domain of *Arabidopsis* leaves represses cell proliferation in this domain and is critical for the development of properly expanded leaves. *Plant J.* 51, 173–184. doi: 10.1111/j.1365-313X.2007.03132.x

Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K., and Poethig, R. S. (2001). KANADI regulates organ polarity in *Arabidopsis*. *Nature* 411, 706–709. doi: 10.1038/35079629

Kidner, C. A., and Martienssen, R. A. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature* 428, 81–84. doi: 10.1038/nature02366

Kidner, C. A., and Timmermans, M. C. (2007). Mixing and matching pathways in leaf polarity. *Curr. Opin. Plant Biol.* 10, 13–20. doi: 10.1016/j.pbi.2006.11.013

Kidner, C. A., and Timmermans, M. C. (2010). Signaling sides adaxial-abaxial patterning in leaves. *Curr. Top. Dev. Biol.* 91, 141–168. doi: 10.1016/S0070-2153(10)91005-3

Lalitha, S. (2004). Primer Premier 5. *Biotech Softw. Int. Rep.* 1, 270–272. doi: 10.1089/152791600459894

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, D229–D232. doi: 10.1093/nar/gkn808

Liang, J. L., Liu, B., Wu, J., Cheng, F., and Wang, X. W. (2016). Genetic variation and divergence of genes involved in leaf adaxial-Abaxial polarity establishment in *Brassica rapa*. *Front. Plant Sci.* 7:94. doi: 10.3389/fpls.2016.00094

Lin, W. C., Shuai, B., and Springern, P. S. (2003). The *Arabidopsis LATERAL ORGAN BOUNDARIES*-domain gene *ASYMMETRIC LEAVES2* functions in the repression of *KNOX* gene expression and in adaxial-abaxial patterning. *Plant Cell* 15, 2241–2252. doi: 10.1105/tpc.014969

Mao, Y., Wu, F., Yu, X., Bai, J., Zhong, W., and He, Y. (2014). MicroRNA319a-targeted *Brassica rapa* ssp. pekinensis TCP genes modulate head shape in chinese cabbage by differential cell division arrest in leaf regions. *Plant Physiol.* 164, 710–720. doi: 10.1104/pp.113.228007

Mcconnell, J. R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M. K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* 411, 709–713. doi: 10.1038/35079635

Moon, J., and Hake, S. (2011). How a leaf gets its shape. *Curr. Opin. Plant Biol.* 14, 24–30. doi: 10.1016/j.pbi.2010.08.012

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326. doi: 10.1093/nar/8.19.4321

Nakata, M., Matsumoto, N., Tsugeki, R., Rikirsch, E., Laux, T., and Okada, K. (2012). Roles of the middle domain-specific WUSCHEL-RELATED HOMEOBOX genes in early development of leaves in *Arabidopsis*. *Plant Cell* 24, 519–535. doi: 10.1105/tpc.111.092858

Pekker, I., Alvarez, J. P., and Eshed, Y. (2005). Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *Plant Cell* 17, 2899–2910. doi: 10.1105/tpc.105.034876

Rhee, S. Y., Beavis, W. D., Berardini, T. Z., Chen, G., Dixon, D. A., Doyle, A., et al. (2003). The *Arabidopsis* information resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* 31, 224–228. doi: 10.1093/nar/gkg076

Rogers, S. O., and Bendich, A. J. (1985). Extraction of DNA from milligram amounts of fresh, herbarium and mummifified plant tissues. *Plant Mol. Biol.* 5, 69–76. doi: 10.1007/BF00020088

Sara, E. G., Jaina, M., Alex, B., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Tamura, K., Stecher, G., Peterson, D. S., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis Version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tang, H., Woodhouse, M. R., Cheng, F., Schnable, J. C., Pedersen, B. S., Conant, G., et al. (2012). Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190, 1563–1574. doi: 10.1534/genetics.111.137349

Tang, H. B., Zhang, X. T., Miao, C. Y., Zhang, J. S., Ming, R., Schnable, J. C., et al. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol.* 16:3. doi: 10.1186/s13059-014-0573-1

Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum Vulgare* L.). *Theor Appl Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Van Ooijen, J. W. (2006). *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma BV.

Vandenbussche, M., Horstman, A., Zethof, J., Koes, R., Rijpkema, A. S., and Gerats, T. (2009). Differential recruitment of WOX transcription factors for lateral development and organ fusion in Petunia and *Arabidopsis*. *Plant Cell* 21, 2269–2283. doi: 10.1105/tpc.109.065862

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). *Brassica rapa* genome sequencing project consortium. the genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wang, Y., Wu, F., Bai, J., and He, Y. (2014). BrpSPL9 (*Brassica rapa* ssp. pekinensis SPL9) controls the earliness of heading time in Chinese cabbage. *Plant Biotechnol J.* 12, 312–321. doi: 10.1111/pbi.12138

Xiao, J., Zhao, J., Liu, M. J., Liu, P., Dai, L., and Zhao, Z. H. (2015). Genome-wide characterization of simple sequence repeat (SSR) Loci in Chinese Jujube and Jujube SSR primer transferability. *PLoS One* 10:e0127812. doi: 10.1371/journal.pone.0127812

Yamaguchi, T., Nukazuka, A., and Tsukaya, H. (2012). Leaf adaxial-abaxial polarity specification and lamina outgrowth: evolution and development. *Plant Cell Physiol.* 53, 1180–1194. doi: 10.1093/pcp/pcs074

Yu, X., Wang, H., Zhong, W., Bai, J., Liu, P., and He, Y. (2013). QTL mapping of leafy heads by genome resequencing in the RIL population of *Brassica rapa*. *PLoS One* 8:e76059. doi: 10.1371/journal.pone.0076059

Zhao, J., Diaan, C. L., Lou, P., Wang, Y., Wu, J., Wang, X., et al. (2008). Quantitative trait loci analysis of phytate and phosphate concentrations in seeds and leaves of *Brassica rapa*. *Plant Cell Environ.* 31, 887–900. doi: 10.1111/j.1365-3040.2008.01791.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gao, Lu, Li, Li, Zhang, Su, Feng, Liu, Xuan, Gu, Wang, Chen, Zhao and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 August 2020
doi: 10.3389/fgene.2020.00950

Identification of *Fusarium solani* f. sp. *pisi* (*Fsp*) Responsive Genes in *Pisum sativum*

*Bruce A. Williamson-Benavides[1], Richard M. Sharpe[2], Grant Nelson[1], Eliane T. Bodah[2†], Lyndon D. Porter[3] and Amit Dhingra[1,2]**

[1] Molecular Plant Sciences, Washington State University, Pullman, WA, United States, [2] Department of Horticulture, Washington State University, Pullman, WA, United States, [3] USDA-ARS, Grain Legume Genetics and Physiology Research Unit, Prosser, WA, United States

OPEN ACCESS

Edited by:
Pinky Agarwal, National Institute of Plant Genome Research (NIPGR), India

Reviewed by:
Tom Warkentin, University of Saskatchewan, Canada
Adnan Šišić, University of Kassel, Germany

***Correspondence:**
Amit Dhingra
adhingra@wsu.edu

†Present address:
Eliane T. Bodah, Pure Line Seeds Inc., Warden, WA, United States

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *13 May 2020*
Accepted: *29 July 2020*
Published: *18 August 2020*

Citation:
Williamson-Benavides BA, Sharpe RM, Nelson G, Bodah ET, Porter LD and Dhingra A (2020) Identification of Fusarium solani f. sp. pisi (Fsp) Responsive Genes in Pisum sativum. Front. Genet. 11:950. doi: 10.3389/fgene.2020.00950

Pisum sativum (pea) is rapidly emerging as an inexpensive and significant contributor to the plant-derived protein market. Due to its nitrogen-fixation capability, short life cycle, and low water usage, pea is a useful cover-and-break crop that requires minimal external inputs. It is critical for sustainable agriculture and indispensable for future food security. Root rot in pea, caused by the fungal pathogen *Fusarium solani* f. sp. *pisi* (*Fsp*), can result in a 15–60% reduction in yield. It is urgent to understand the molecular basis of *Fsp* interaction in pea to develop root rot tolerant cultivars. A complementary genetics and gene expression approach was undertaken in this study to identify *Fsp*-responsive genes in four tolerant and four susceptible pea genotypes. Time course RNAseq was performed on both sets of genotypes after the *Fsp* challenge. Analysis of the transcriptome data resulted in the identification of 42,905 differentially expressed contigs (DECs). Interestingly, the vast majority of DECs were overexpressed in the susceptible genotypes at all sampling time points, rather than in the tolerant genotypes. Gene expression and GO enrichment analyses revealed genes coding for receptor-mediated endocytosis, sugar transporters, salicylic acid synthesis, and signaling, and cell death were overexpressed in the susceptible genotypes. In the tolerant genotypes, genes involved in exocytosis, and secretion by cell, the anthocyanin synthesis pathway, as well as the DRR230 gene, a pathogenesis-related (PR) gene, were overexpressed. The complementary genetic and RNAseq approach has yielded a set of potential genes that could be targeted for improved tolerance against root rot in *P. sativum*. *Fsp* challenge produced a futile transcriptomic response in the susceptible genotypes. This type of response is hypothesized to be related to the speed at which the pathogen infestation advances in the susceptible genotypes and the preexisting level of disease-preparedness in the tolerant genotypes.

Keywords: pea, root rot, tolerance, susceptible, RNAseq, transcriptomics, gene expression, *Pisum sativum* L.

INTRODUCTION

The food industry contributes significantly to the world's total greenhouse gas emissions (Poore and Nemecek, 2018). About 18% of the global greenhouse gas emissions are caused by livestock production, which supplies the majority of the dietary protein (Stehfest et al., 2009). Proposed mitigation efforts include a shift to plant-based protein as it is an environmentally sustainable option. The demand for plant-based protein is on the rise due to its health benefits (World Health Organization, 2015), as well as due to the ethical concerns related to exploiting animals as a source of protein (Johansson, 2019). The global plant-based protein market is expected to keep growing at a compound annual growth rate of 8.1% from 2019 to 2025 (Research and Markets, 2019). Popular plant-based meats from Impossible Foods and Beyond Meat have already reached some of the biggest food and retail brands in the United States.

Pea (*Pisum sativum* L.; Family Fabaceae) is a major contributor to this plant-derived protein market and has gained particular interest lately due to its high protein content (20–30%), especially lysine; overall high nutritional value; and relatively low cost (do Carmo et al., 2016; Peng et al., 2016; Xiong et al., 2018). The market for pea protein is expected to be $34.8 million in 2020 due to growing consumer interest in plant-based proteins as an alternative to animal-derived protein (Grand View Research, 2015; Pietrysiak et al., 2018). Pea also plays a critical role in sustainable agriculture due to its nitrogen-fixing capability, short life cycle, and low water usage; all of which make it a useful cover-and-break crop requiring minimal external inputs.

The United States is one of the world's major pea producers. In the United States, harvested area of peas has increased by over 300% during the last 25 years; however, the yields have decreased an average of 7.5% throughout this timespan. This is likely due to the fact that the cultivars that were originally developed in the Pacific Northwest are being increasingly grown in less productive areas such as Montana and the Dakotas (Vandemark et al., 2014). Sustainable production of pea has been negatively affected by several diseases, predominantly root rots (Akhtar and Azam, 2014; Bodah et al., 2016). Root rots are the diseases of greatest impact to crop production worldwide (Kumari and Katoch, 2020). Frequently, root rot diseases involve more than one pathogen; therefore, the disease is known as root rot complex. Pathogens such as oomycetes and fungi are commonly involved in this root rot complex (Xu et al., 2012; Chittem et al., 2015; Gossen et al., 2016; Bodah, 2017).

One of the predominant causal agent of root rots in *P. sativum* is the soil fungus, *Fusarium solani* f. sp. *pisi* (*Fsp*). *F. solani* is a fungal soil-borne facultative parasite that is present worldwide (Zhang et al., 2006). The yields of *P. sativum* cultivars can be reduced by 15–60% by *Fsp* (Seaman, 1976; Grünwald et al., 2003; Porter et al., 2014). Over the years, hundreds of pea cultivars and germplasm core collections have been screened for *Fsp* resistance, and lines have been developed that demonstrate partial resistance to selected *Fsp* races (Coyne et al., 2008). An effort to identify tolerance to root rot in wild pea germplasm resulted in the identification of eight accessions with high levels of partial resistance (Porter, 2010). These accessions have been utilized for developing new cultivars. However, in tests replicated in the greenhouse and/or the field with derived selections, complete tolerance to *Fsp* has not been obtained (Grünwald et al., 2003; Porter et al., 2014; Bodah et al., 2016).

Understanding the genetic basis of tolerance to *Fsp* in a wide array of different pea breeding lines and cultivars has been pursued in several studies. The first QTL for *Fsp* tolerance was reported from a field study utilizing various parental lines that showed resistance to multiple root rots (Kraft, 1992; Feng et al., 2011). Recent studies conducted under controlled conditions have reported three QTLs; QTL *Fsp-Ps 2.1* explains 44.4–53.4% of the phenotypic variance within a 1.2 cM confidence interval. The other two QTLs, *Fsp-Ps 3.2,* and *Fsp-Ps 3.3* explain 3.6–4.6% of the phenotypic variance related to *Fsp* root rot tolerance (Coyne et al., 2015, 2019). While the genes underlying these QTLs have not yet been identified, there is a reason for optimism given the recent release of the pea reference genome (Kreplak et al., 2019). It is expected to facilitate the characterization of potential transcription factors, stress-associated phytohormone genes, Pathogenesis-related (PR) proteins, or pea phytoalexin Pisatin (Kendra and Hadwiger, 1984) in the interaction between pea and *Fsp*.

While genetic approaches for identifying disease tolerance or resistance genes are common, gene expression approaches to identify key genes in response to pathogen challenge remain scarce. A report of *Aphanomyces euteiches*-mediated root rot of pea was investigated using a gene expression approach, and novel genes responsive during the pathogenic interaction with *Medicago truncatula* were reported (Nyamsuren et al., 2003). Besides the expected induction of PR and defense genes, several novel genes were also reported to be overexpressed during the plant-pathogen interaction.

To the best of our knowledge, a gene expression approach to identify genes involved in *Fsp* tolerance in pea is yet to be reported. For gaining a comprehensive insight into the transcriptomic responses during the *Fsp* challenge, a comparative time-course RNAseq expression analysis was performed on four tolerant and four susceptible *P. sativum* genotypes that were selected from a preceding study (Bodah et al., 2016). Data analysis reaffirmed the role of Disease-Resistance Response 230 (DRR230) and sugar transporters, as well as expression patterns of genes associated with receptor-mediated endocytosis and exocytosis, cell death, and anthocyanin synthesis. Interestingly, several previously uncharacterized genes were also identified to be differentially expressed in both tolerant and susceptible genotypes, which may help illuminate the novel mechanism of pea-*Fsp* interaction.

MATERIALS AND METHODS

Plant Material and *Fsp* Isolates

A total of eight, white-flowered pea genotypes were selected for pathogen challenge (**Table 1**). Four tolerant genotypes – 00-5001, 00-5003, 00-5004, and 00-5007 – were selected from the *Fsp* tolerant 5000 series (Porter et al., 2014). The 5000 series pea breeding lines were previously developed

TABLE 1 | Selected white-flowered pea genotypes for time-course transcriptome analysis in response to *Fusarium solani* f. sp. *pisi* (*Fsp*) challenge.

Genotype	Source[a]	*Fsp* tolerance level[b]	Other disease resistance[c]	100 seed weight	Leaf type[d]	Market Class
00-5001	USDA-ARS VFCRU	*	Fop races 1, 2 and 5	22.7	af	Green fresh
00-5003	USDA-ARS VFCRU	*	Fop races 1, 2 and 5	15.9	af	Green fresh
00-5004	USDA-ARS VFCRU	*	Fop races 1, 2 and 5	20.8	af	Green fresh
00-5007	USDA-ARS VFCRU	*	Fop races 1, 2 and 5	22.2	P	Green fresh
Aragorn'	ProGene	***	Fop races 1, 2; PSBMV	19.5	af	Green dry
Banner'	ProGene	***	Fop race 2, PM	18.7	af	Green dry
Bolero'	AsGrow	****	Fop race 1, PM, Pythium, EMV	20.12	P	Green fresh
DSP'	Canner Seed	***	–	20.9	P	Green fresh

*The table summarizes pea genotypes, source, Fsp tolerance level, other disease resistance, 100 seed weight, leaf type, and market class. [a]AsGrow = AsGrow Seed Co., San Juan Bautista, CA, United States; Canner Seed = Canner Seed Co., Idaho Falls, ID, United States; ProGene = ProGene LLC Plant Research, Othello, WA, United States; USDA-ARS VFCRU = USDA-ARS, Vegetable and Forage Crops Research Unit, Prosser, WA, United States. [b]Fsp tolerance = Fsp tolerance resulted from phenotyping, ranging from most tolerant (*) to most susceptible (****) (Bodah et al., 2016). [c]Fop = Fusarium oxysporum; PSBMV = Pea Seed-Borne Mosaic Virus, PM = Powdery Mildew; EMV = Enation Mosaic Virus. [d]af (Afila) = semi-leafless; P (Perfection) = normal leaf type. It is a gradient scale where * is tolerant, ** is less tolerant, *** is susceptible, and **** represents completely susceptible.*

via single-seed descent from crosses initiated in a USDA–ARS greenhouse in Prosser, WA, in 1998. The parentage of 00-5001 is PH14-119/M7477// Coquette/3/86-2197/74-410-2 (Kraft, 1989; USDA–ARS NGRP, 2020). The parentage of 00-5003 is 69PH42-691004/Recette//Popet/3/PH14-119/ DL-1/3/B563-429-2/PI257593//DSP TAC (USDA–ARS NGRP, 2020). The parentage of 00-5004 is 79-2022/ICI 1203-1//Menlo/3/PI189171/ DL-2//75-786 (Kraft and Tuck, 1986; USDA–ARS NGRP, 2020). The parentage of 00-5007 is 00-5005/00-5006. 00-5005 parentage is B669-87-0/M7477//Blixt B5119/3/00-5001/74SN5/3/PH14-119/DL-1//74SN3/Recette/5/ FR-725 (Kraft and Giles, 1976; USDA–ARS NGRP, 2020). 00-5006 parentage is 00-5003/00-5004.

Four susceptible genotypes – "Aragorn," "Banner," "Bolero," and "DSP" – were identified among frequently used commercial pea varieties. These eight genotypes were selected based on their contrasting root disease severity index reported in a preceding study (Bodah et al., 2016). The eight genotypes were classified as either tolerant or susceptible post-*Fsp* challenge by phenotyping of the root disease severity index, plant height, shoot dry weight, and root dry weight (Bodah et al., 2016). The 5000 series pea breeding lines were found to be the most tolerant lines among the white-flowered pea lines.

The *Fsp* isolates Fs 02, Fs 07, and Fs 09, were obtained from infected pea roots collected in the Palouse Region of Washington and Idaho, United States soils by Dr. Lyndon Porter, USDA-ARS Vegetable and Forage Crops Research Unit, Prosser, WA. The three isolates were single-spored, identified as *Fsp*, based on partial translation elongation factor 1-a sequences using Fusarium-ID (Geiser et al., 2004), and their pathogenicity to pea was confirmed (Bodah et al., 2016). The three isolates were grown on pentachloronitrobenzene (PCNB) selective media for 6 days (Nash and Snyder, 1962). Cultures were transferred to KERR's media (Kerr, 1963), and incubated on a shaker at 120 rpm under continuous light for 6 days at 23–25°C. The spore concentration of each isolate was determined using a hemocytometer and diluted to 1×10^6 spores/ml of water. A spore suspension inoculum containing equal parts by volume of each of the three isolates was created.

Fsp Disease Challenge

Fusarium solani f. sp. *pisi* disease challenge was performed as reported previously (Bodah et al., 2016). Briefly, seeds of each pea genotype were sterilized in a 0.6% sodium hypochlorite solution and rinsed in sterile distilled H_2O. Seeds were then soaked for 16 h in either the *Fsp* spore suspension (inoculated set) or in sterile H_2O (control set). At the end of 16 h, seeds from the control and inoculated set were separated from the respective solutions and maintained on perlite under ambient light and controlled greenhouse conditions with temperature ranging between 21 and 25°C for either 6- or 12-h duration. The 0-h time point was represented by the endpoint of the 16-h soaking period. The embryo containing region of the seed was excised and harvested from 600 seeds per genotype per time point (0, 6, and 12 h), immediately frozen under liquid nitrogen and transferred to storage at −80°C for subsequent RNA extraction. The experiment was repeated three times in tandem under identical conditions.

RNA Isolation, cDNA Library Construction and Sequencing

The frozen seed material from two out of the three pathogen challenge experiments was randomly selected and pulverized in a SPEX SamplePrep 6870 FreezerMill (SPEX SamplePrep, Metuchen, NJ, United States) for five cycles. Each cycle consisted of cooling for 2 min and grinding at 15 counts per second for 4 min. Total RNA was isolated from the pulverized tissue using the RNeasy Plant RNA Extraction Kit (Qiagen, Hilden, Germany). A Nanodrop ND-8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States) and a Qubit Fluorometer (Life Technologies, Camarillo, CA, United States) were used to quantify the extracted RNA. Contaminating DNA was removed using the TURBO DNA-free Kit (Life Technologies, Camarillo, CA, United States) using the manufacturer's instructions. RNA quality was verified via electrophoresis on a 1% agarose gel.

Equimolar amounts of RNA samples from tolerant and susceptible genotypes were bulked for each time point before the construction of RNAseq libraries. RNAseq libraries were

constructed using 1 μg of RNA, and the Illumina TruSeq kits (Illumina Inc., San Diego, CA, United States). RNA was purified with an Oligo(dT) cellulose affinity matrix and subsequently fragmented into short pieces of an average size of 450 base pairs with Ampure XP beads (Beckman Coulter, Brea, CA, United States). All libraries were quantified on a Qubit Fluorometer (Life Technologies, Camarillo, CA, United States) and analyzed on an Agilent BioAnalyzer (Agilent Technologies, Santa Clara, CA, United States) to determine the concentration, final size, and purity of the library. A total of 24 libraries were sequenced using the HiSeq2000 configuration 100 PE (Illumina Inc., San Diego, CA, United States) at the Michigan State University Genomics core laboratory.

RNAseq Data Processing and Statistical Analysis

The generated fastq files were analyzed for quality with CLC Bio Genomics Workbench 6.0.1 (CLC Bio, Aarhus, Denmark) and trimmed with trimmomatic (Bolger et al., 2014). *De novo* RNAseq assembly was performed using data from all 24 samples to obtain a master assembly with the software Trinity v2.8.4 (Grabherr et al., 2011). Dependencies for Trinity, Bowtie2 v1.2.3 (Langmead and Salzberg, 2012), Salmon v0.12.0 (Patro et al., 2017), and JELLYFISH v2.2.3 (Marçais and Kingsford, 2011), were used during assembly. Bowtie2 and Salmon were used for abundance estimation, and JELLYFISH was used as a k-mer counting software.

The software Kalisto was used for transcript quantification (Bray et al., 2016). The reads were quantified for each of the two biological replicates of the tolerant or susceptible genotypes at three time points, 0, 6, 12 h after inoculation, and for each control or treatment. This analysis resulted in 24 separate quantification groups that were used for comparison. Differentially expressed contigs (DECs) with p-value of <0.001 and a greater than two-fold change in expression were identified using the Baggerley's test for data from each of the two biological replicates generated for each genotype group, time point, and inoculation treatment. The RPKM (Reads Per Kilobase of transcript per Million mapped reads) expression values were also ascertained for each contig. Heat maps showing fold-change of RPKM values between control (C) and inoculated (I) sets and among genotypes were created in Microsoft Excel 365 ProPlus (Microsoft Corporation, Redmond, WA, United States).

Functional Annotation, Statistical Gene Ontology (GO) Enrichment, and Pathway Analysis

Functional annotation of the master assembly and DECs was conducted via BLAST in BLAST2GO v. 3.3. (Conesa and Götz, 2008). Default parameters were used for the functional annotation, as well as for Gene ontology (GO) mapping, and InterProScan. The two-tailed Fisher's exact test (FDR < 0.05) was used to ascertain over- and under-represented functions during the *Fsp* challenge. A heat map representing "biological process" GO terms over-represented in the *Fsp* inoculated treatment was created in Microsoft Excel 365 ProPlus (Microsoft Corporation, Redmond, WA, United States). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed to identify pathways represented by the set of DECs for each time point and genotype.

Real-Time Quantitative PCR

RNA was extracted utilizing the RNeasy Plant DNA Extraction Kit (Qiagen, Mainz, Germany) from the same sampled tissues utilized for the RNAseq analysis. After DNase treatment, equimolar amounts of RNA from the tolerant and susceptible genotypes were bulked for each time point. First-strand cDNA synthesis was performed using 1,500 ng of each bulked RNA sample with the SuperScript Vilo kit (Thermo Fisher Scientific, Waltham, MA, United States). Nine genes were randomly selected from the list of DECs for RT-qPCR analysis (**Supplementary Table S1**). Primers for RT-qPCR were designed with the Primer3 software (Rozen and Skaletsky, 2000) with the corresponding transcriptome contig as the query sequence for each primer set. The *P. sativum* root border cell-specific protein (GenBank accession AF1139187.1) was used as an internal reference control as it showed invariant expression across genotypes and treatments in the RNAseq data.

The QUBIT 3.0 fluorometer (Invitrogen, Carlsbad, CA, United States) was used to quantify cDNA library concentration. For each reaction, 16 ng of cDNA was used with the iTaq Universal SYBR Green Supermix (Bio-Rad, Berkeley, CA, United States). Each RT-qPCR reaction was performed in triplicate for each of the three biological replicates using the Stratagene M×3005P (Thermo Fisher Scientific, Waltham, MA, United States). The amplification profile consisted of an initial denaturation at 95°C for 150 s, 40 cycles of 20 s at 95°C for denaturation, 20 s at 60°C for annealing, and 20 s at 72°C for extension. A melting curve analysis was performed post amplification to ensure the presence of a unique amplicon and performed with an initial denaturation at 95°C for 1 min and a decrease of temperature to 50°C for annealing. The temperature was then increased in 0.5°C increments at 5 s/step from 50 to 95°C for fluorescence readings. Raw fluorescence data was used as input for crossover threshold (Ct) calculations and reaction efficiencies adjusted with LinRegPCR 2012.0 software (Ruijter et al., 2009). The $\Delta\Delta$Ct method offered by PE Applied Biosystems (Perkin Elmer, Forster City, CA, United States) was used to obtain relative differential expression values after reaction efficiencies were adjusted with the LinRegPCR 2012.0 software (Pfaffl, 2001).

Functional Annotation of QTL Associated With *Fsp* Tolerance in Pea

Fsp-Ps 2.1, the major QTL found to be associated with *Fsp* tolerance in pea (Coyne et al., 2019), was annotated using the transcriptome data generated in this study to determine if there are any differentially expressed genes located in the selected genomic region. *Fsp-Ps 2.1* explains 44.4–53.4% of the phenotypic variance, and it is located on chromosome II within a 1.2 cM confidence interval of marker Ps900203 (Duarte et al., 2014; Coyne et al., 2019). The genomic sequence

of the *Fsp-Ps2.1* ± 1.2 cM (±201,800 nt) was obtained from the pea genome (Kreplak et al., 2019). The length of the QTL sequence was calculated based on the distance between the Ps900203 and Ps000075 markers. Marker Ps000075 is 0.6 cM = 168,166 nt away from Ps900203. The transcriptome data from this study was aligned via BLAST against the *Fsp-Ps2.1* ± 1.2 cM sequence in CLC Bio Genomics Workbench 6.0.1 (CLC Bio, Aarhus, Denmark).

RESULTS

Assembly of Transcriptome Data and Identification of Differentially Expressed Contigs (DECs)

A total of 850 million reads were generated after the sequencing of the 24 libraries (**Supplementary Table S2**). For these 24 libraries, the mean Q score ranged from 34.00 to 34.88; these Q scores validate the quality of the assay (**Supplementary Table S2**). After QC and trimming of low-quality reads, 69.49% of reads were used for assembly of the master transcriptome. The master transcriptome, generated in this study, was composed of 185,721 contigs (**Supplementary Table S3**), which had a mean contig length of 1503.15 nucleotides (nt) with a length range of 184–18,990 nt.

Mapping of reads to the master assembly showed a different number of contigs with zero mapped reads for each time point (0, 6, 12 h) and genotype (tolerant and susceptible). After expunging contigs with zero mapped reads, the total number of contigs ranged from 102,382 to 141,530, or 55.13 to 76.21% of the total 185,721 contigs, respectively (**Supplementary Table S2**). The total number of reads mapped to each contig for each time point and genotype and RPKM values are summarized in **Supplementary Table S3**.

For each contig, twelve pairwise comparisons were performed (Tables S3, S4). In order to identify which genes were differentially expressed in response to the *Fsp* challenge, six comparisons were performed that evaluated the expression values of genes in non-inoculated versus *Fsp*-inoculated treatments, as listed in **Supplementary Table S4**-comparisons 1 to 6 (**Figure 1A**). In order to identify which genes responded differentially to *Fsp* between the tolerant (bulked-tolerant) and susceptible (bulked-susceptible) genotypes, six additional pairwise comparisons were performed and are summarized in **Supplementary Table S4**-comparisons 7 to 12 (**Figure 1B**). The twelve pairwise comparisons resulted in the identification of 42,905 DECs out of the 185,721 contigs.

Pairwise comparisons 1 to 6 that compare non-inoculated versus *Fsp*-inoculated treatments, yielded the number of upregulated DECs obtained for each time point (**Figure 2**). For the *Fsp* inoculated tolerant genotypes, the number of upregulated DECs varied between 1,200 and 1,460 DECs across 0, 6-, and 12-h time points. The number of suppressed (under-expressed) DECs in the inoculated tolerant genotypes was larger (2795-4453 DECs). For the *Fsp* inoculated susceptible genotypes, the total number of upregulated DECs was 5–7 times larger than the total number for the *Fsp* inoculated tolerant genotypes (**Figure 2**). The number of suppressed genes in both sets of genotypes was similar at 0 h. However, for the susceptible genotypes, these numbers were 1.34 and 2.21 times larger than the tolerant genotypes at 6 and 12 h, respectively. A total of 5.7 and 5.4% of the DECs were shared across the three time points for the tolerant and susceptible genotypes, respectively (**Supplementary Table S3**). From the total upregulated DECs in the *Fsp* inoculated sets, 10 (0.10%), 48 (0.52%), and 12 (0.12%) DECs are shared among the tolerant and susceptible genotypes at the 0, 6, and 12-h time points, respectively (**Supplementary Table S3**).

Functional Annotation of DECs and GO-Term Enrichment

From the 185,721 contigs of the master assembly, 120,132 returned positive BLAST hits when queried to the NCBI database (National Center for Biotechnology Information, 2020). A total of 4,734 contigs or 2.55% of the total contigs were annotated as proteins of unknown function or hypothetical proteins. The top BLAST hits showed similarity to *M. truncatula*, *Trifolium pratense*, *Cicer arietinum*, and *Trifolium subterraneum* with a distribution of 34.0, 16.1, 16.0, and 14.6% respectively. *P. sativum*, with a 3.6% match, was fifth in the rankings. The low percentage of hits was most likely due to the relatively small number of *P. sativum* RNAseq data represented in the NCBI database. Of the 42,905 DECs, 36,923 (86.1%) returned positive BLAST hits when aligned to the NCBI database. Interestingly, 33 contigs (0.09%) of the 36,923 hits were classified as proteins of unknown function or hypothetical proteins, which could be useful candidates for understanding the pea-*Fsp* interaction.

The GO enrichment analysis identified significant over- and under-represented GO terms for each of the three *Fsp* inoculated time points (**Supplementary Table S5**). GO terms related to routine DNA processes, such as DNA metabolic process, DNA biosynthetic process, and DNA integration were significantly underrepresented at different time points in the tolerant and susceptible genotypes. DNA integration and DNA metabolic process terms were underrepresented across the three times in the susceptible genotypes, but only at 0 and 6 h for the tolerant genotypes. Nucleic acid phosphodiester bond hydrolysis was also underrepresented only during certain times points in the susceptible genotypes but not in the tolerant genotypes.

Figure 3 represents a heat map with the overrepresented biological process GO terms for each time point in the inoculated treatment for the tolerant and susceptible genotypes. These terms provide a comparative perspective of biological processes that responded to *Fsp* in the two subsets of genotypes.

Quantitative RT-PCR Verification

To verify the expression results obtained from RNAseq data, RT-qPCR analysis was performed on nine randomly selected genes at all time points for the tolerant and susceptible genotypes. The expression trends of eight out of nine genes (89%) correlated with the RPKM values indicating the robustness of RNAseq results (**Figure 4**). The internal reference control (GenBank accession

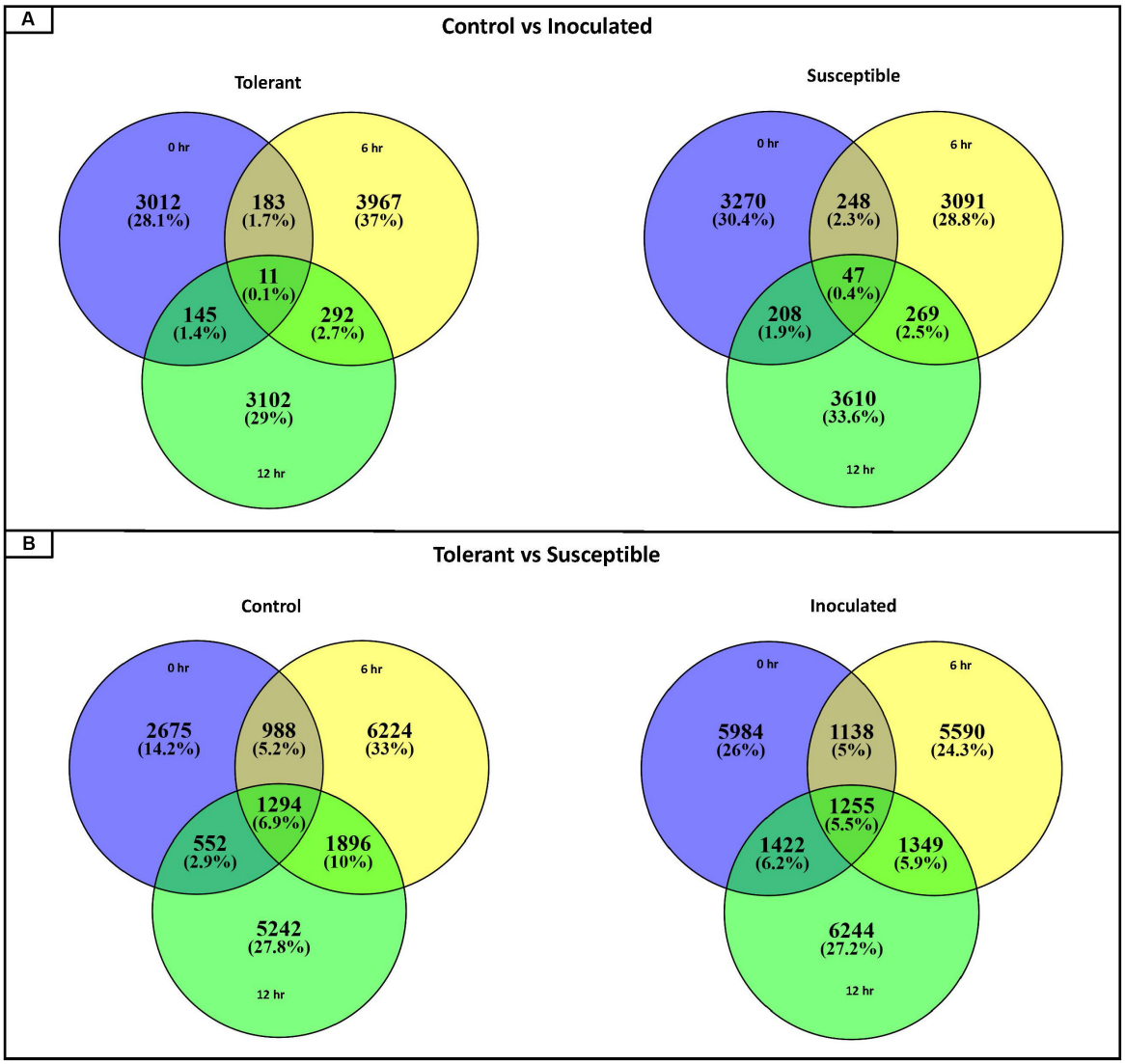


FIGURE 1 | Venn diagrams representing the number of DECs (Differentially Expressed Contigs) for the 12 pairwise comparisons. **(A)** Number of DECs for pairwise comparisons between control and inoculated samples collected at 0, 6, and 12-h time points for the tolerant and the susceptible genotypes. **(B)** Number of DECs for pairwise comparisons between the tolerant and susceptible genotypes for each time point (0, 6, and 12 h) for control and inoculated conditions.

AF1139187.1) showed invariant expression across genotypes and treatments in this RT-qPCR analysis.

Fsp-Induced Differential Gene Expression

A closer look at the 42,905 DECs showed that most of these genes (86.1%) returned positive BLAST hits. The genes were placed into seven broad categories for understanding their potential role in responding to *Fsp* challenge: (1) Expression of signaling-related genes (**Supplementary Table S3**), (2) Genes involved in transcriptional regulation (**Figure 5** and **Supplementary Table S3**), (3) PR genes (**Figure 6** and **Supplementary Table S3**), (4) Anthocyanin and lignin biosynthetic pathway genes (**Figure 7** and Tables S3, S6), (5) Sugar metabolism (**Supplementary Table S3**), (6) Phytohormones (**Figure 8** and **Supplementary Table S3**), (7) Cell wall and membrane metabolism and toxin metabolism (**Figure 9** and **Supplementary Table S3**).

DISCUSSION

Assembly of Transcriptome Data and Identification of Differentially Expressed Contigs (DECs)

The susceptible genotypes showed a larger number of upregulated DECs that represented a more diverse and higher number of transcriptionally regulated genes, in contrast to the

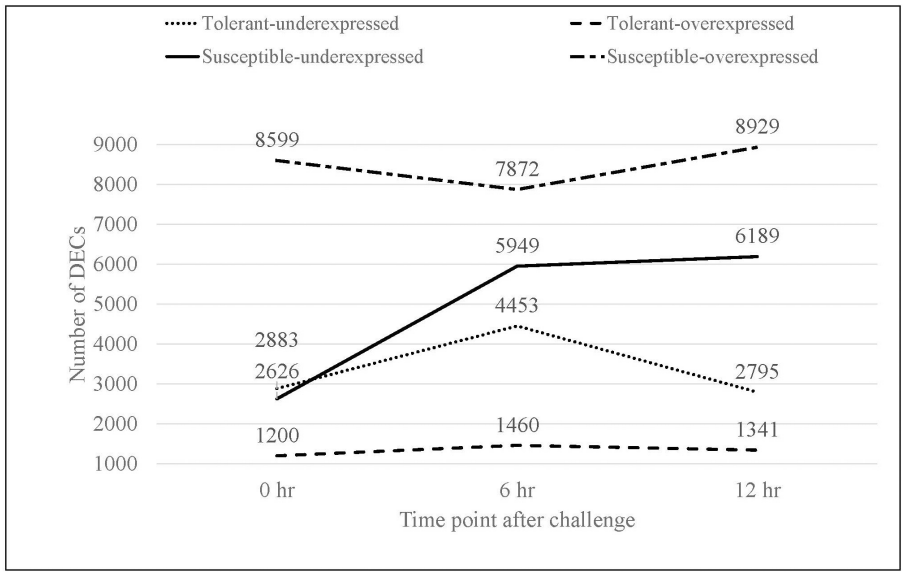


FIGURE 2 | The total number of overexpressed and underexpressed DECs in the inoculated treatments for tolerant and susceptible genotypes at each time point in response to the *Fsp* challenge.

tolerant genotypes. This higher number of upregulated DECs in the susceptible genotypes may be due to involvement of various biological processes associated with successful *Fsp* infection in pea embryonic tissue. This observation is consistent with previous studies, which showed that pathogen attack engages a broader range of pathways and a larger proportion of genes in the susceptible genotypes compared to the resistant ones (Bagnaresi et al., 2012; Zheng et al., 2013; Matič et al., 2016).

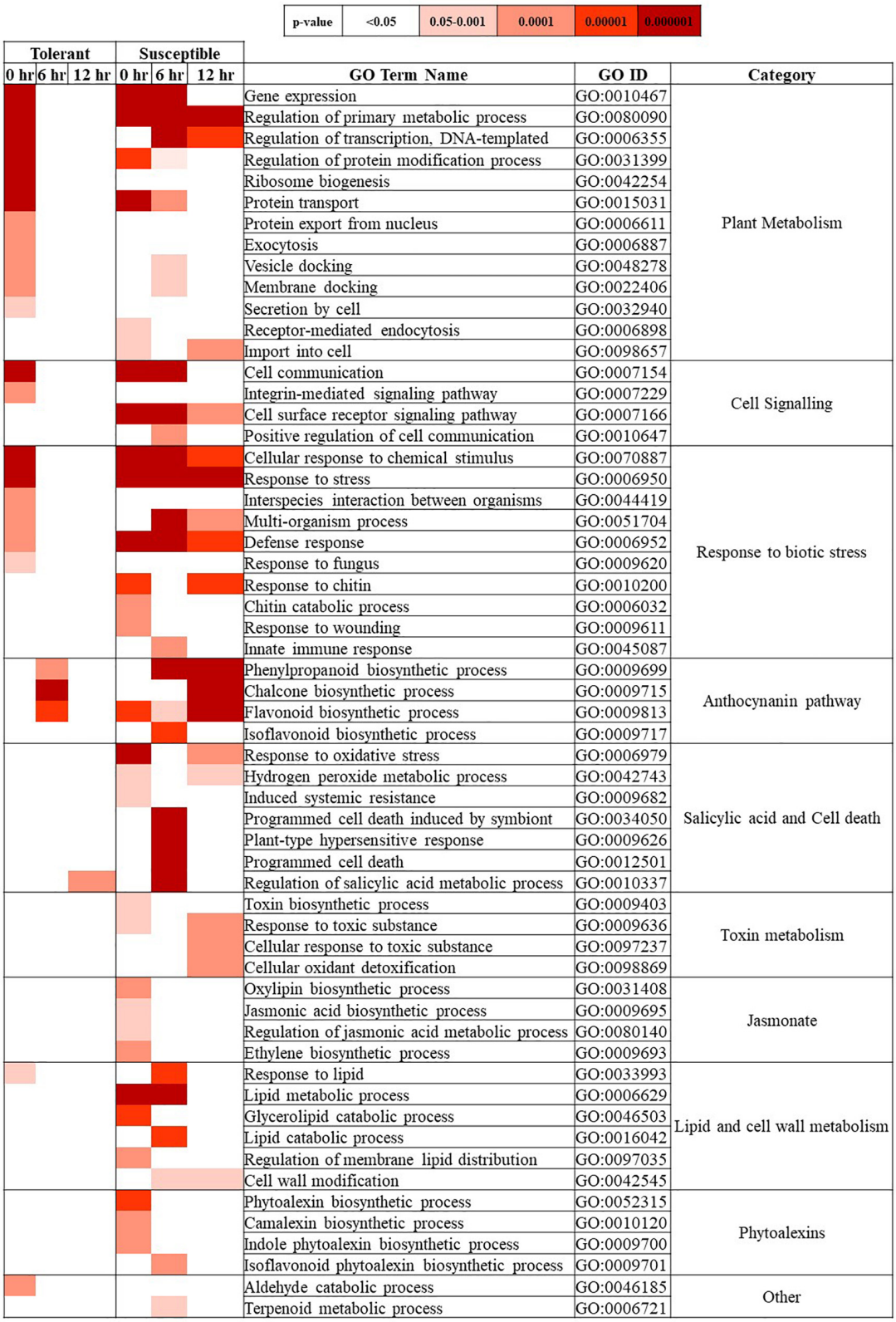
Only 10 (0.10%), 48 (0.52%), and 12 (0.12%) DECs are shared among the tolerant and susceptible genotypes at the 0, 6, and 12-h time points, respectively. Therefore, the response mechanisms involved in the *Fsp* challenge were divergent between the tolerant and susceptible genotypes. The large difference in overexpressed DECs could explain the difference in tolerance between these sets of genotypes. The genes that were overexpressed in susceptible genotypes were numerically different from the ones in the tolerant genotypes.

Functional Annotation of DECs and GO-Term Enrichment

The GO-term enrichment analysis showed that common terms related to basic plant metabolism, such as gene expression, regulation of primary metabolism, transcription, and protein synthesis were over-represented at the 0-h time point in the tolerant genotypes. Conversely, these terms were upregulated in at least two time points in the susceptible genotypes. Therefore, these results showed the tolerant genotypes may have moved toward a basal metabolic state after the 0-h time point, while the susceptible genotypes were responding to *Fsp* throughout the entire time course presented in this study.

Gene ontology terms such as ribosome biogenesis, protein export from nucleus, exocytosis, and secretion by cell were over-represented at 0 h in the tolerant genotypes. Plants are known to transport antimicrobial molecules, such as peptides and/or secondary metabolites, outside the cell to function in plant immunity (Kwon and Yun, 2014). In contrast, the susceptible genotypes seemed to be importing substances inside the cell since terms such as receptor-mediated endocytosis and import to the cell were over-represented. Endocytosis seems not only to play a role in pathogen-associated molecular patterns (PAMP)-triggered immunity and effector-triggered immunity but also in susceptibility. Vesicle endocytosis can be manipulated by pathogens and can import pathogen-derived effectors into the plant cell (Driouich et al., 1997; Kwon and Yun, 2014). In this study, while the tolerant genotypes seemed to export antimicrobial molecules potentially to counter *Fsp*, the susceptible genotypes seemed to import substances. Thus, it is hypothesized that the suppression of exocytosis mechanisms in the susceptible genotypes might block or delay the transport and release of antimicrobial substances against pathogens.

Gene ontology terms in the cell signaling, and response to biotic stress categories also showed an early and unique response at 0 h in the tolerant genotypes, while this response was present throughout the entire experiment (0, 6, and 12 h) or at later stages (6 or 12 h) in susceptible genotypes. Several studies have analyzed how pea responds to *Fsp* and the non-host pathogen *F. solani* f. sp. *phaseoli* (*Fsph*). These studies concluded that the major difference is the speed at which the pea plants react. The type of response exhibited by pea varies with the rate of induction of PR genes and other associated biochemical pathways. In the case of either *Fsph* or *Fsp* infection, the fungus releases DNAses, which localize to the host nuclei and digest the nuclear DNA



GO Term Name	GO ID	Category
Gene expression	GO:0010467	Plant Metabolism
Regulation of primary metabolic process	GO:0080090	
Regulation of transcription, DNA-templated	GO:0006355	
Regulation of protein modification process	GO:0031399	
Ribosome biogenesis	GO:0042254	
Protein transport	GO:0015031	
Protein export from nucleus	GO:0006611	
Exocytosis	GO:0006887	
Vesicle docking	GO:0048278	
Membrane docking	GO:0022406	
Secretion by cell	GO:0032940	
Receptor-mediated endocytosis	GO:0006898	
Import into cell	GO:0098657	
Cell communication	GO:0007154	Cell Signalling
Integrin-mediated signaling pathway	GO:0007229	
Cell surface receptor signaling pathway	GO:0007166	
Positive regulation of cell communication	GO:0010647	
Cellular response to chemical stimulus	GO:0070887	Response to biotic stress
Response to stress	GO:0006950	
Interspecies interaction between organisms	GO:0044419	
Multi-organism process	GO:0051704	
Defense response	GO:0006952	
Response to fungus	GO:0009620	
Response to chitin	GO:0010200	
Chitin catabolic process	GO:0006032	
Response to wounding	GO:0009611	
Innate immune response	GO:0045087	
Phenylpropanoid biosynthetic process	GO:0009699	Anthocynanin pathway
Chalcone biosynthetic process	GO:0009715	
Flavonoid biosynthetic process	GO:0009813	
Isoflavonoid biosynthetic process	GO:0009717	
Response to oxidative stress	GO:0006979	Salicylic acid and Cell death
Hydrogen peroxide metabolic process	GO:0042743	
Induced systemic resistance	GO:0009682	
Programmed cell death induced by symbiont	GO:0034050	
Plant-type hypersensitive response	GO:0009626	
Programmed cell death	GO:0012501	
Regulation of salicylic acid metabolic process	GO:0010337	
Toxin biosynthetic process	GO:0009403	Toxin metabolism
Response to toxic substance	GO:0009636	
Cellular response to toxic substance	GO:0097237	
Cellular oxidant detoxification	GO:0098869	
Oxylipin biosynthetic process	GO:0031408	Jasmonate
Jasmonic acid biosynthetic process	GO:0009695	
Regulation of jasmonic acid metabolic process	GO:0080140	
Ethylene biosynthetic process	GO:0009693	
Response to lipid	GO:0033993	Lipid and cell wall metabolism
Lipid metabolic process	GO:0006629	
Glycerolipid catabolic process	GO:0046503	
Lipid catabolic process	GO:0016042	
Regulation of membrane lipid distribution	GO:0097035	
Cell wall modification	GO:0042545	
Phytoalexin biosynthetic process	GO:0052315	Phytoalexins
Camalexin biosynthetic process	GO:0010120	
Indole phytoalexin biosynthetic process	GO:0009700	
Isoflavonoid phytoalexin biosynthetic process	GO:0009701	
Aldehyde catabolic process	GO:0046185	Other
Terpenoid metabolic process	GO:0006721	

FIGURE 3 | Biological process-GO terms over-represented in the *Fsp* inoculated treatment for tolerant and susceptible genotypes at 0, 6, and 12-h time points. The tone of colors in the heatmap denotes the *p*-value calculated using two-tailed Fisher's exact test. Significantly over-represented GO terms showed a $p > 0.05$. *p*-values.

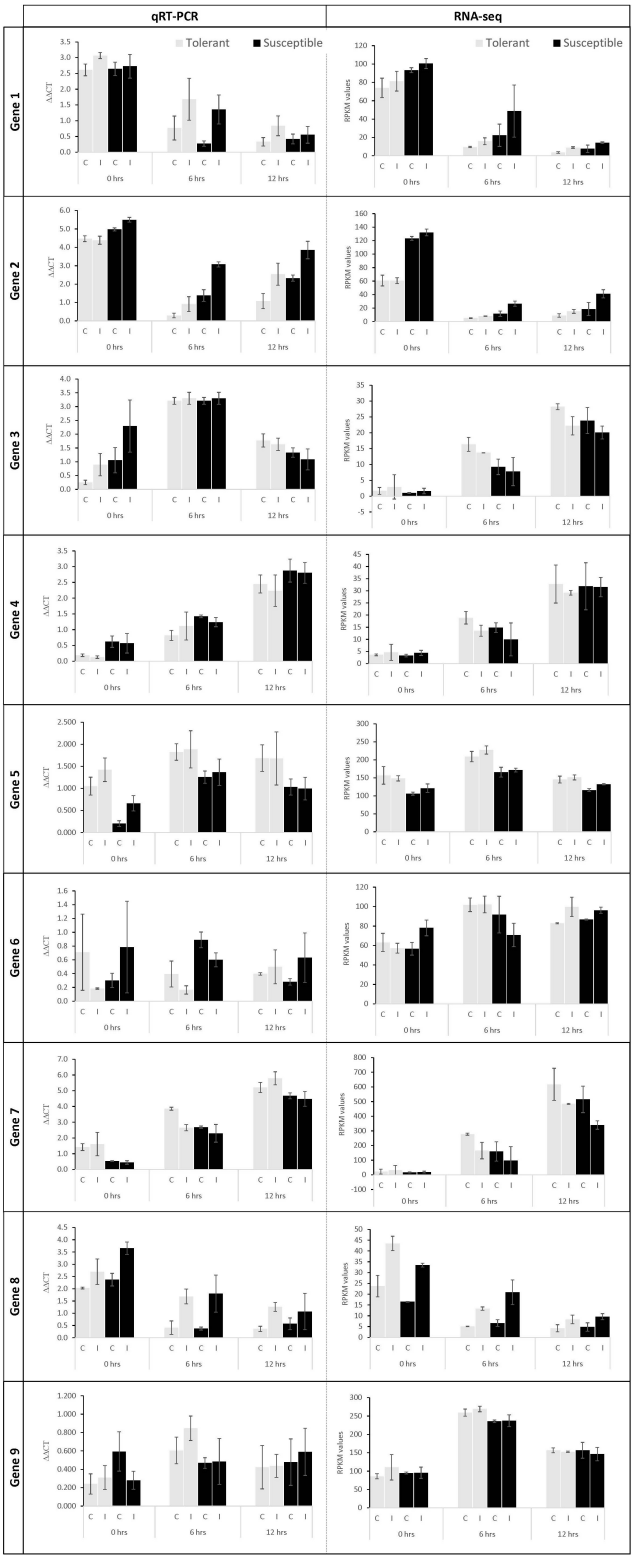


FIGURE 4 | Continued

FIGURE 4 | RT-qPCR validation of select genes in control (C) and Fsp inoculated (I) plants. The gray and black bars represent relative gene expression for the tolerant and susceptible genotypes, respectively. First column: RT-qPCR data show the average relative expression of three biological samples with three technical replicates each. Second column: RPKM values calculated for each gene. The error bars represent the standard error between replicates in RT-qPCR analysis. Gene 1: TRINITY_DN2419_c0_g1_i5 (BLAST accession: XM_003592027.3), Gene 2: TRINITY_DN2754_c0_g1_i11 (BLAST accession: XM_004502933.3), Gene 3: TRINITY_DN5727_c0_g1_i1 (BLAST accession: MK618561.1), Gene 4: TRINITY_DN6240_c0_g1_i9 (BLAST accession: XM_003592048.3), Gene 5: TRINITY_DN2169_c1_g1_i2 (BLAST accession: XM_004506541.3), Gene 6: TRINITY_DN1232_c0_g1_i11 (BLAST accession: XM_004504351.3), Gene 7: TRINITY_DN5529_c0_g1_i9 (BLAST accession: XM_024782286.1), Gene 8: TRINITY_DN8631_c0_g1_i1 (BLAST accession: XM_013611166.2), Gene 9: TRINITY_DN1795_c0_g1_i2 (BLAST accession: XM_004514502.3).

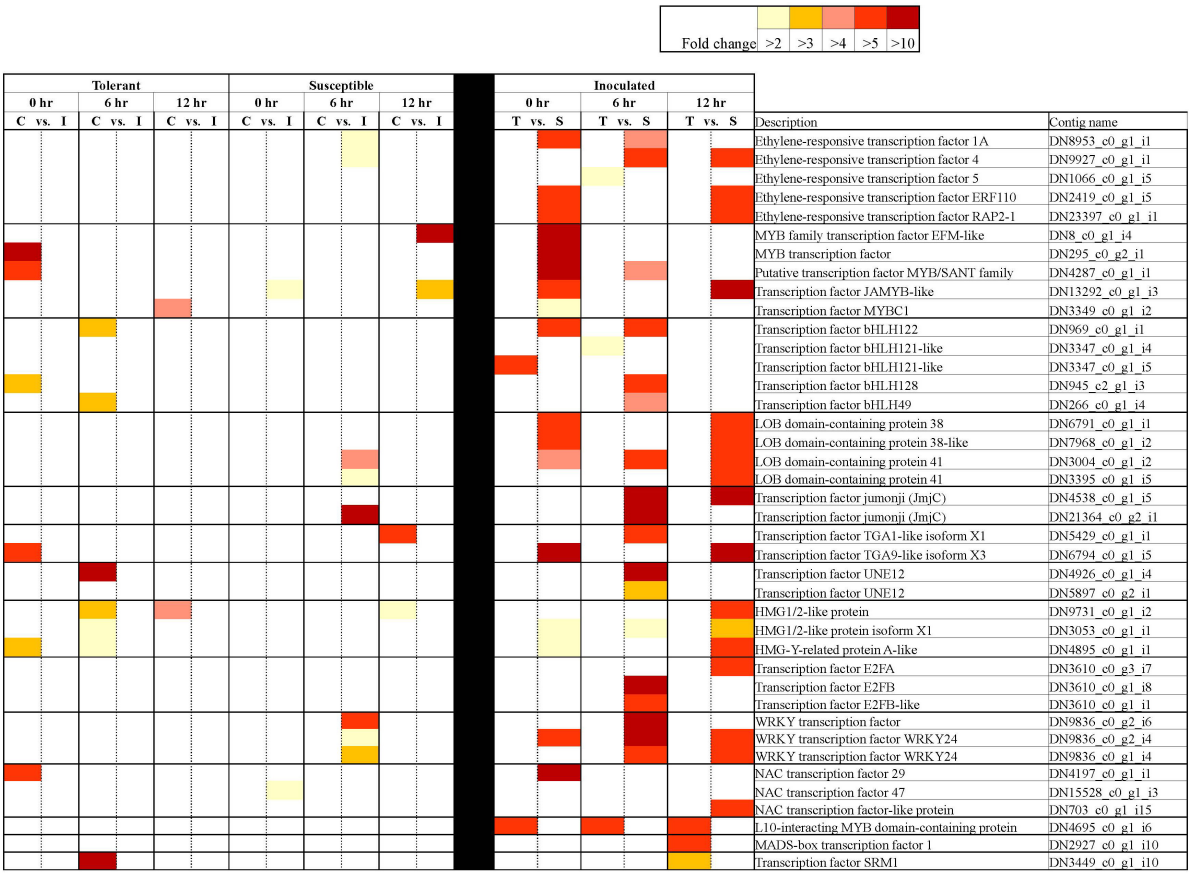
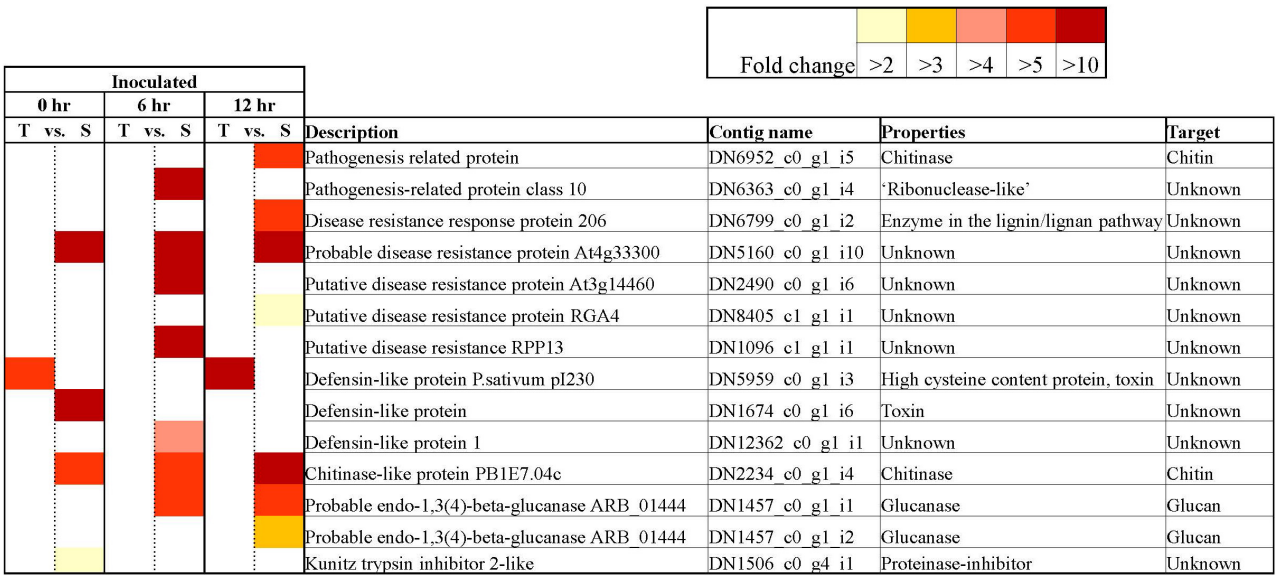


FIGURE 5 | Differentially expressed transcription factors in tolerant and susceptible pea genotypes in response to *Fsp* challenge. The color key denotes fold-change. Pairwise comparisons that displayed greater than two-fold difference ($p < 0.001$) in expression were identified with a color that ranges from light yellow (fold change >2) to dark red (fold change >10).

(Hadwiger and Adams, 1978; Hadwiger, 2008, 2015). Fungal DNAses can also impact the nuclei in the fungal mycelia and trigger their deterioration (Hadwiger, 2008). In the case of a compatible interaction (successful infection leading to disease) between *Fsp* and pea, the slower reaction rate of the pea host allows *Fsp* to protect a small number of its nuclei from fungal DNAses. The slower reaction allows the growth of *Fsp* to resume after 12 h post-inoculation (Klosterman et al., 2001; Hadwiger, 2015). In contrast, the relatively rapid response generated in the host against *Fsph* terminates the growth of the fungi at 6 h. post-inoculation (Hadwiger, 2008, 2015). Given this information, it is hypothesized the speed of reaction to the pathogen may be one of the mechanisms of tolerance in the tolerant genotypes.

The GO terms associated with salicylic acid and cell death category were, in most cases, overrepresented only in the susceptible genotypes. Therefore, the susceptible genotypes were expected to have a more intense response to *Fsp* through induced systemic resistance, host programmed cell death, and plant-type hypersensitive response.

The GO terms associated with the production of flavanones, flavones, flavonols, proanthocyanidins, and anthocyanins showed over-representation in both genotypes. In the tolerant genotypes, these terms were only over-represented at 6 h. In the susceptible genotypes, GO terms from this category were overrepresented at the 0, 6, and 12-h time points, indicating a more intense response. GO terms associated with the



Description	Contig name	Properties	Target
Pathogenesis related protein	DN6952_c0_g1_i5	Chitinase	Chitin
Pathogenesis-related protein class 10	DN6363_c0_g1_i4	'Ribonuclease-like'	Unknown
Disease resistance response protein 206	DN6799_c0_g1_i2	Enzyme in the lignin/lignan pathway	Unknown
Probable disease resistance protein At4g33300	DN5160_c0_g1_i10	Unknown	Unknown
Putative disease resistance protein At3g14460	DN2490_c0_g1_i6	Unknown	Unknown
Putative disease resistance protein RGA4	DN8405_c1_g1_i1	Unknown	Unknown
Putative disease resistance RPP13	DN1096_c1_g1_i1	Unknown	Unknown
Defensin-like protein P.sativum pI230	DN5959_c0_g1_i3	High cysteine content protein, toxin	Unknown
Defensin-like protein	DN1674_c0_g1_i6	Toxin	Unknown
Defensin-like protein 1	DN12362_c0_g1_i1	Unknown	Unknown
Chitinase-like protein PB1E7.04c	DN2234_c0_g1_i4	Chitinase	Chitin
Probable endo-1,3(4)-beta-glucanase ARB_01444	DN1457_c0_g1_i1	Glucanase	Glucan
Probable endo-1,3(4)-beta-glucanase ARB_01444	DN1457_c0_g1_i2	Glucanase	Glucan
Kunitz trypsin inhibitor 2-like	DN1506_c0_g4_i1	Proteinase-inhibitor	Unknown

FIGURE 6 | Key differentially expressed Pathogenesis-related contigs in tolerant and susceptible pea genotypes in response to *Fsp* challenge. The color key denotes fold-change. Pairwise comparisons that displayed greater than two-fold difference ($p < 0.001$) in expression were identified with a color that ranges from light yellow (fold change >2) to dark red (fold change >10).

jasmonate pathway, phytoalexin synthesis, toxin metabolism, and lipid metabolism were overrepresented only in the susceptible genotypes at all time points (**Figure 3** and **Supplementary Table S5**).

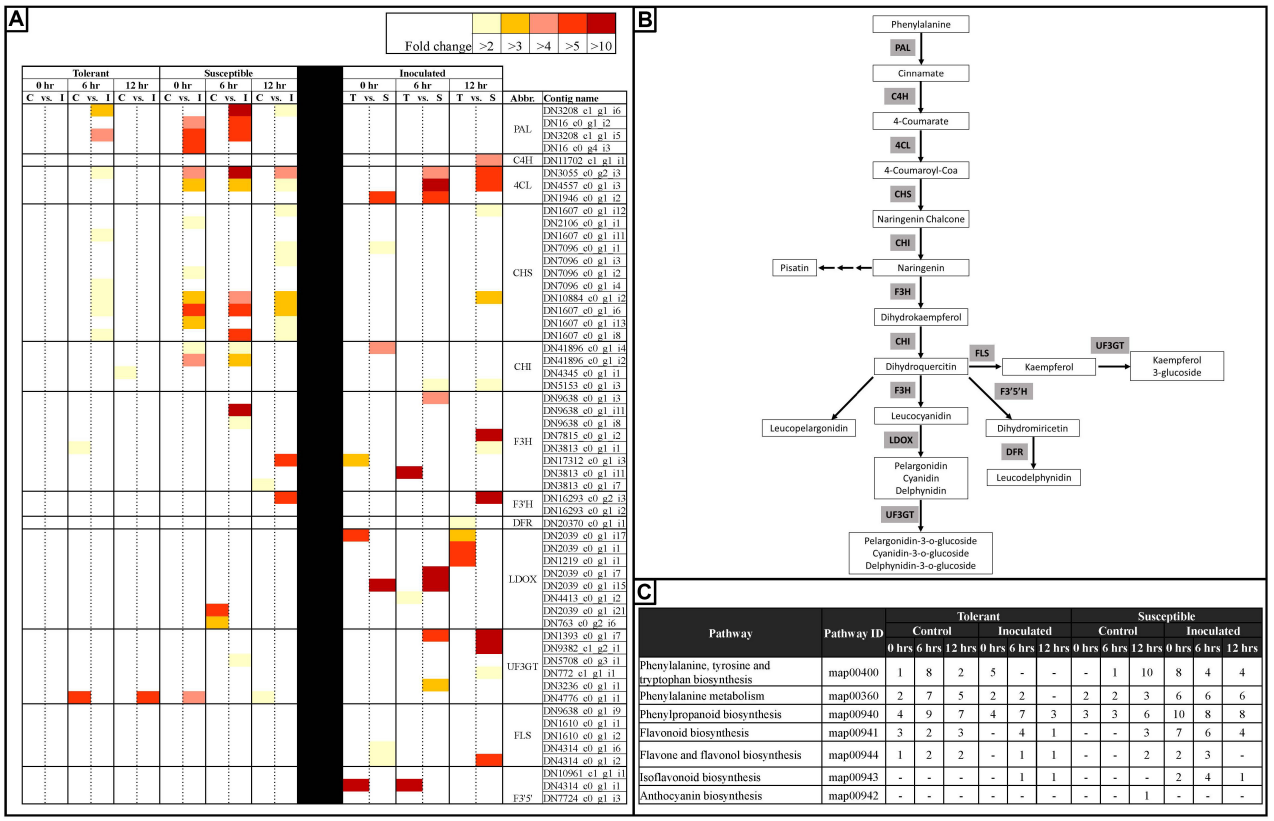
Fsp-Induced Differential Gene Expression

Expression of Signaling-Related Genes

Plants detect pathogens via host sensors known as pattern-recognition receptors (PRR), which act by detecting PAMPs and triggering PAMP-triggered immunity (PTI). Beck et al. (2012) identified three PRRs that have been proven to be specific to fungi: the Chitin Elicitor Binding Protein (CEBiP), the chitin elicitor receptor kinase I (CERK1), and the ethylene-inducing xylanase (Eix2). CEBiP and CERK1 cooperatively regulate chitin elicitor signaling to activate plant defense system (Hayafune et al., 2014). Interestingly, CEBiP was not identified in this study. Contigs corresponding to CERK1 and Eix2 genes were identified, however they were not differentially expressed in any of the genotypes.

In this study, several LRR-RLK receptors were found to be differentially expressed in response to *Fsp* challenge (**Supplementary Table S3**). Most of these genes were upregulated in the susceptible rather than in the tolerant genotypes. The L-type lectin-domain containing receptor kinase, proline-rich receptor-like protein kinase, cysteine-rich receptor-like protein kinase and the wall-associated receptor kinases are also upregulated in the susceptible genotypes, following the same trend as above. The activity of Mitogen-activated protein kinases (MAPK), MAPK kinases (MAPKK), and MAPKK kinases (MAPKKK) were, in most cases, upregulated in the tolerant and susceptible genotypes after the challenge with *Fsp* when compared to the control. However, the expression was, in most cases, significantly higher in the susceptible genotypes throughout the time-course when compared to the tolerant genotypes.

Two contigs identified as receptors were found to change significantly in expression after the *Fsp* challenge in the susceptible genotypes only. Contig DN1290_c0_g1_i9 and Contig DN7023_c0_g2_i5 were identified as a receptor-like cytoplasmic kinase 176 and CC-NBS-LRR resistance protein (**Supplementary Table S3**), respectively. The receptor-like cytoplasmic kinase 176 acts downstream of the CERK1 gene in the fungal chitin signaling pathways that mediates innate immunity responses such as reactive oxygen species generation, defense gene expression, and callose deposition (Ao et al., 2014). The CC–NBS–LRR proteins initiate a resistance response that often includes a type of cell death known as the hypersensitive response (HR) (Moffett et al., 2002). In the susceptible genotypes, contig DN1290_c0_g1_i9 was significantly overexpressed at 6 h (FC = 3.19) after the *Fsp* challenge. Expression of Contig DN7023_c0_g2_i5 was found to be lowered at 12 h (FC = −3.59) after the *Fsp* challenge. Interestingly, no change was observed in the expression of the two contigs in the tolerant genotypes. These two contigs also showed higher expression in the susceptible genotypes when compared to their expression in the tolerant genotypes. Contigs DN1290_c0_g1_i9 and DN7023_c0_g2_i5 were significantly upregulated at 0 (FC = 3.49), 6 (FC = 5.08), and 12 h (FC = 6.77), and 6 h (FC = 71.69), respectively, in the inoculated treatments of the susceptible genotypes when their expression values were compared to the tolerant genotypes. The data on the observed induction of genes coding for receptor-like cytoplasmic kinase 176 and CC-NBS-LRR resistance protein in



Pathway	Pathway ID	Tolerant						Susceptible					
		Control			Inoculated			Control			Inoculated		
		0 hrs	6 hrs	12 hrs	0 hrs	6 hrs	12 hrs	0 hrs	6 hrs	12 hrs	0 hrs	6 hrs	12 hrs
Phenylalanine, tyrosine and tryptophan biosynthesis	map00400	1	8	2	5	-	-	-	1	10	8	4	4
Phenylalanine metabolism	map00360	2	7	5	2	2	-	2	2	3	6	6	6
Phenylpropanoid biosynthesis	map00940	4	9	7	4	7	3	3	3	6	10	8	8
Flavonoid biosynthesis	map00941	3	2	3	-	4	1	-	-	3	7	6	4
Flavone and flavonol biosynthesis	map00944	1	2	2	-	1	1	-	-	2	2	3	-
Isoflavonoid biosynthesis	map00943	-	-	-	-	1	1	-	-	-	2	4	1
Anthocyanin biosynthesis	map00942	-	-	-	-	-	-	-	-	1	-	-	-

FIGURE 7 | Heatmap representation of changes in the expression of genes associated with the anthocyanin biosynthetic pathway in tolerant and susceptible pea genotypes after *Fsp* challenge. **(A)** A subset of differentially expressed genes involved in the anthocyanin biosynthetic pathway in tolerant and susceptible pea genotypes after the *Fsp* challenge. The color key denotes fold-change. Pairwise comparisons that displayed greater than two-fold difference ($p < 0.001$) in expression were identified with a color that ranges from light yellow (fold change >2) to dark red (fold change >10). **(B)** Anthocyanin biosynthesis pathway (Adapted from Solfanelli et al., 2006). **(C)** KEGG pathway analysis of metabolic processes related to the anthocyanin biosynthetic pathway. Abbreviations (Abbr.) PAL, phenylalanine ammonia lyase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone-3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; DFR, dihydroflavonol 4-reductase; LDOX, leucoanthocyanidin dioxygenase; UF3GT, UDP glucose-flavonoid 3-o-glucosyltransferase; FLS, Flavonol synthase; F3′5′H, Flavonoid 3′,5′-hydroxylase.

the susceptible genotypes suggest that the pathogen likely recruits oxygen species generation, hypersensitive response, defense gene expression, and callose deposition to establish infection. It raises a question if loss of function mutation in these genes in the susceptible genotypes could confer tolerance to *Fsp*. A loss-of-function mutation in receptor-like cytoplasmic kinases and CC-NBS-LRRs genes has proven to confer tolerance to different pathogens (Lorang et al., 2007; Sweat and Wolpert, 2007; Zhang et al., 2019).

Genes Involved in Transcriptional Regulation

The following TFs were found to be differentially expressed between the tolerant and susceptible genotypes and/or were influenced by *Fsp* challenge: bZIP, ERF, MYB, GATA, MADS-box, NAC, PLATZ, KAN2, PosF21, WRKY, C2H2, bHLH, DIVARICATA, E2F, GLABRA, ICE, IIIB, Jumonji, PIF, RF2a, SRM1, TCP19, TGA, UNE12, and HMG. From this list of TFs, bZIP, ERF, MYB, MADS-box, NAC, WRKY, C2H2, bHLH, E2F, Jumonji, PIF, RF2a, TCP19, TGA, and HMG have been reported to be master regulators of defense responses against pathogens (Pontier et al., 2001; Vailleau et al., 2002; Dong et al., 2003; Pré et al., 2008; Isaac et al., 2009; Wang et al., 2009; Kiełbowicz-Matuk, 2012; Alves et al., 2013; Li et al., 2013; Song et al., 2013; Chandran et al., 2014; Khong et al., 2015; Li, 2015; Paik et al., 2017; Im et al., 2019). However, the involvement of following TFs in response to pathogen challenge has not been reported previously – GATA, PLATZ, KAN2, PosF21, DIVARICATA, GLABRA, ICE, IIIB, SRM1, UNE1.

In the tolerant genotypes, TFs were either not differentially expressed when the control and inoculated samples were compared, or their expression was significantly suppressed after *Fsp* challenge. In the susceptible genotypes, however, the expression of TFs remained the same or increased after *Fsp* challenge. When the inoculated treatments in the tolerant and the susceptible genotypes were compared at either time point, TFs were overexpressed in the susceptible genotypes (**Figure 5**). It is well documented that overexpression of certain TFs causes susceptibility to certain pathogens (Kim et al., 2006; Lai et al., 2008; Thatcher et al., 2012). Some examples include the enhanced susceptibility of the

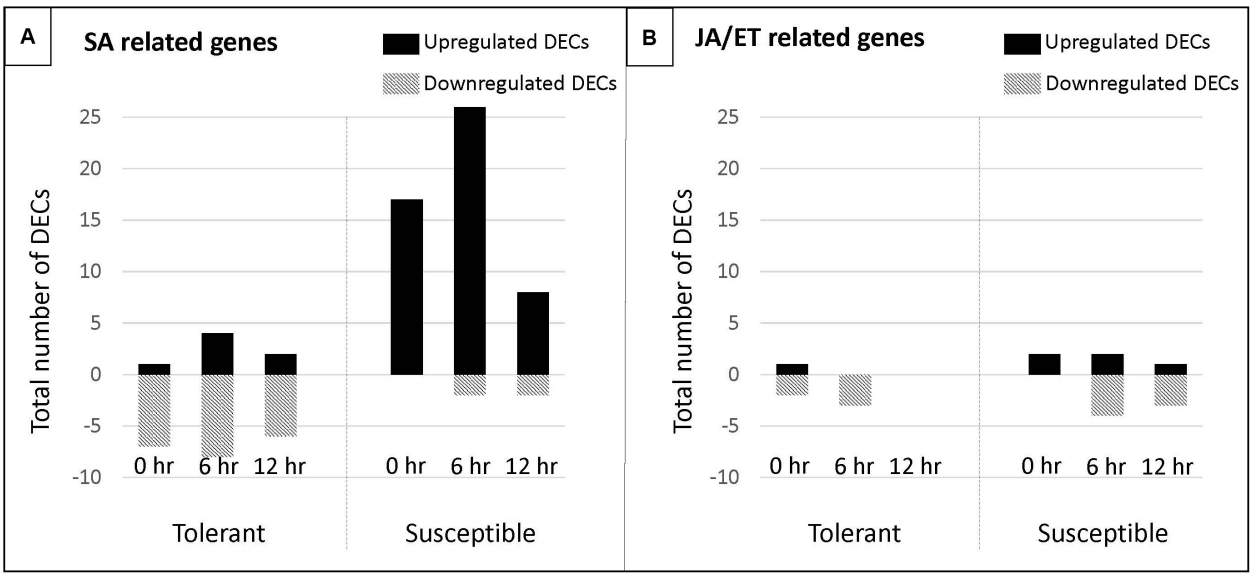
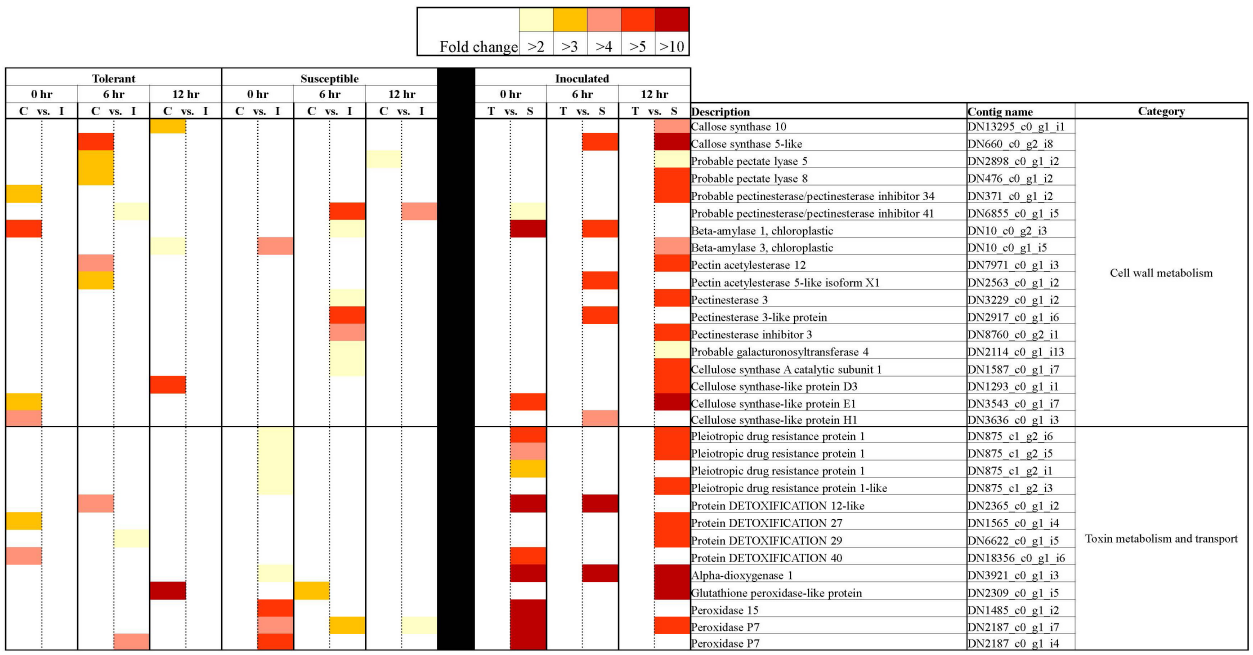


FIGURE 8 | Number of DECs associated with the salicylic (SA) and jasmonate/ethylene (JA/ET) biosynthetic and signaling pathway in a pea tolerant and susceptible genotype after *Fsp* challenge. **(A)** DECs associated with the salicylic (SA) biosynthetic and signaling pathway. **(B)** DECs associated with the jasmonate/ethylene (JA/ET) biosynthetic and signaling pathway.



Description	Contig name	Category
Callose synthase 10	DN13295_c0_g1_i1	Cell wall metabolism
Callose synthase 5-like	DN660_c0_g2_i8	
Probable pectate lyase 5	DN2898_c0_g1_i2	
Probable pectate lyase 8	DN476_c0_g1_i2	
Probable pectinesterase/pectinesterase inhibitor 34	DN371_c0_g1_i2	
Probable pectinesterase/pectinesterase inhibitor 41	DN6855_c0_g1_i5	
Beta-amylase 1, chloroplastic	DN10_c0_g2_i3	
Beta-amylase 3, chloroplastic	DN10_c0_g1_i5	
Pectin acetylesterase 12	DN7971_c0_g1_i3	
Pectin acetylesterase 5-like isoform X1	DN2563_c0_g1_i2	
Pectinesterase 3	DN3229_c0_g1_i2	
Pectinesterase 3-like protein	DN2917_c0_g1_i6	
Pectinesterase inhibitor 3	DN8760_c0_g2_i1	
Probable galacturonosyltransferase 4	DN2114_c0_g1_i13	
Cellulose synthase A catalytic subunit 1	DN1587_c0_g1_i7	
Cellulose synthase-like protein D3	DN1293_c0_g1_i1	
Cellulose synthase-like protein E1	DN3543_c0_g1_i7	
Cellulose synthase-like protein H1	DN3636_c0_g1_i3	
Pleiotropic drug resistance protein 1	DN875_c1_g2_i6	Toxin metabolism and transport
Pleiotropic drug resistance protein 1	DN875_c1_g2_i5	
Pleiotropic drug resistance protein 1	DN875_c1_g2_i1	
Pleiotropic drug resistance protein 1-like	DN875_c1_g2_i3	
Protein DETOXIFICATION 12-like	DN2365_c0_g1_i2	
Protein DETOXIFICATION 27	DN1565_c0_g1_i4	
Protein DETOXIFICATION 29	DN6622_c0_g1_i5	
Protein DETOXIFICATION 40	DN18356_c0_g1_i6	
Alpha-dioxygenase 1	DN3921_c0_g1_i3	
Glutathione peroxidase-like protein	DN2309_c0_g1_i5	
Peroxidase 15	DN1485_c0_g1_i2	
Peroxidase P7	DN2187_c0_g1_i7	
Peroxidase P7	DN2187_c0_g1_i4	

FIGURE 9 | Heatmap representation of differentially expressed genes associated with cell wall metabolism, toxin metabolism and transport in tolerant and susceptible pea genotypes in response to *Fsp* challenge. The color key denotes fold-change. Pairwise comparisons that displayed greater than two-fold difference ($p < 0.001$) in expression were identified with a color that ranges from light yellow (fold change >2) to dark red (fold change >10).

WRKY7-overexpressing Arabidopsis plants to *Pseudomonas syringae* infection. Overexpression of the AtWRKY4 gene in Arabidopsis enhances susceptibility toward the biotrophic bacterium *P. syringae* (Lai et al., 2008). In a disease screen with *Fusarium oxysporum*, it was found that disruption of the *LATERAL ORGAN BOUNDARIES* (*LOB*) *DOMAIN* (*LBD*) TF led to increased resistance to *F. oxysporum* root-rot disease in *Arabidopsis thaliana*.

From the extensive list of differentially expressed TFs identified in this study, only the HMG A has been previously

shown to participate during pea-*Fsp* interaction (Hadwiger, 2008). HMG A has been shown to complex with promoter areas of Disease-resistance response (DRR) genes in pea after *Fsp* infection (Klosterman et al., 2003). Transient expression experiments implicate the HMG-I/Y abundance in the down-regulation of DRR206 gene expression in pea (Klosterman et al., 2003). In this study, contigs DN4895 and DN12030 were identified as HMG-I/Y genes. Interestingly, the expression of the DN4895 contig decreases significantly at 0 h (FC = −3.59) and 6 h (FC = −2.65) after *Fsp* challenge in the tolerant genotypes. On the other hand, the expression of this contig increased significantly at 6 h (FC = 3.51) after *Fsp* challenge in the susceptible genotypes. Furthermore, the expression of contig DN4895 was significantly higher under inoculated conditions at 0 h (FC = 2.56) and 12 h (FC = 7.91) for the susceptible genotypes compared to these expression values in the tolerant genotypes. Contig DN12030 is also significantly overexpressed under inoculated conditions in the susceptible genotypes at 0 h (FC = 3.23) and 12 h (FC = 5.70). It is plausible that HMG-I/Y mediates the reduction of expression of defense-related genes in the susceptible genotypes; however, this assumption will need to be further evaluated.

Pathogenesis-Related (PR) Genes

The inoculation of the tolerant and susceptible genotypes with *Fsp* generated changes in the expression of PR genes. All the PR protein encoding genes that were identified in this experiment were overexpressed in the susceptible genotypes over the tolerant genotypes except one (**Figure 6**). The contig DN5959_c0_g1_i3 was identified as a defensin named *P. sativum* pI230 mRNA (*e*-value: 0.0, percentage identity: 97.23%). The expression of the DN5959_c0_g1_i3 contig was significantly higher in the tolerant genotypes over the susceptible genotypes under control conditions at 0 (FC = −69.21) and 6 h (FC = −151.78), and under inoculated conditions at 0 (FC = −7.93) and 12 h (FC = −43.5). The pI230 mRNA is the precursor for the DRR230 protein, which is a disease resistance response protein identified previously in *P. sativum*. DRR230 defensin was first identified in pea pods in response to infection by *Fsp* (Chiang and Hadwiger, 1991). This defensin was also found to co-localize with a major QTL (*mpIII-4*) involved in resistance to *Mycosphaerella pinodes* in pea (Prioul-Gervais et al., 2007). DRR230 gene does not co-localize with any of the major QTLs identified by Coyne et al. (2019).

Disease-resistance response 230 was isolated by Almeida et al. (2000) and characterized as a small cysteine-rich polypeptide. Almeida et al. (2000) also determined that DRR230 is very effective as a fungal growth inhibitor against *Aspergillus niger*, *Aspergillus versicolor*, *Fsph*, and *Neurospora crassa*. The specific function of DRR230 is not yet known, however plant defensins form a characteristic structure known as the cysteine-stabilized α/β motif, a feature that is also shared by several toxins from insects, scorpions, honeybees, and spider venoms (Hadwiger, 2008). The pea DRR230 was overexpressed in canola, and these plants inhibited were significantly more resistant to *Leptosphaeria maculans*, a hemibiotrophic fungus (Wang et al., 1999). The transcriptome analysis presented in this study reinforces recent and preceding studies that suggest that DRR230 may play a key role in resistance or tolerance to *Fsp* induced root rot.

Anthocyanin and Lignin Biosynthetic Pathway Genes

The germplasm utilized in this study consisted of white-flowered lines that present partial tolerance to *Fsp*. RNAseq analysis showed that the white-flowered lines contain a large set of DECs that participate in the anthocyanin biosynthetic pathway (**Figure 7**). Genes coding for enzymes in the phenylalanine ammonia lyase (PAL) to chalcone isomerase (CHI) biochemical pathway were upregulated in the inoculated treatments in both the tolerant and susceptible genotypes. Furthermore, some of these enzymes were either overexpressed in the susceptible genotypes or were expressed at a similar level between the tolerant and susceptible genotypes (**Figure 7**).

The expression patterns were more variable for genes coding for enzymes from the flavanone-3-hydroxylase (F3H) to leucoanthocyanidin dioxygenase (LDOX), as well as for the flavonoid 3′,5′-hydroxylase (F3′5′H), UDP glucose-flavonoid 3-o-glucosyltransferase (UF3GT), and flavonol synthase (FLS) enzymes (**Figures 7A,B**). Some isoforms of F3H were overexpressed in the susceptible genotypes, but also some other isoforms were overexpressed in the tolerant genotypes. F3H, UF3GT, and FLS were upregulated in the susceptible genotypes, but most or all isoforms of LDOX, and F3′5′ were upregulated in the tolerant genotypes. In the susceptible genotypes, expression of some isoforms of LDOX was suppressed after challenge with *Fsp* (**Figures 7A,B**).

Figure 7C also shows the differential expression of genes at each time point, treatment, and genotype in this study. Early steps in the phenylpropanoid pathway such as phenylalanine biosynthesis, phenylalanine metabolism, phenylpropanoid biosynthesis, flavonoid biosynthesis, flavone and flavonol biosynthesis are active in the control replicates of the tolerant genotypes at the three times points. In the susceptible controls, fewer, or none of the enzymes were overexpressed. This observation suggests that the tolerant genotypes had a higher level of expression of genes in the phenylpropanoid pathway under the basal conditions and, therefore, it was potentially better prepared to defend against *Fsp*.

Coyne et al. (2019) reported a significant QTL (*Fsp-Ps2.1*) that accounts for 44.4 to 53.4% of the phenotypic variance for resistance to *Fsp* and this QTL shows a confidence interval of 1.2 cM. *Fsp-Ps2.1* was mapped within the interval of the pigmented flower/anthocyanin pigmentation gene called as gene A in that study. However, the gene was mapped in a white flower cross. One hypothesis is that the resistance gene(s) may have been linked in the genome with gene A in the pigmented lines. The white-flowered, resistant parent may have been obtained through a linkage break between *Fsp-Ps2.1* and *A*. Alternatively, a metabolite, possibly a colorless one, in the anthocyanin pathway might be the one that provides this resistance. Fine mapping or gene knockouts are necessary to test this hypothesis.

The transcriptome data, generated in this study, was aligned via BLAST against the *Fsp-Ps2.1* ± 1.2 cM sequence (**Supplementary Table S6**). BLAST analysis returned 500

positive blast hits when queried to the transcriptome data. A total of 156 contigs showed differential expression after *Fsp* challenge or when the tolerant and susceptible genotypes expression values are compared (**Supplementary Table S6**). A total of 22 of the 156 contigs were annotated as proteins of unknown function or hypothetical proteins. Only the contig TRINITY_DN4823 was identified as a disease-related gene, soyasaponin III rhamnosyltransferase. However, this contig was overexpressed in the susceptible genotypes when compared to the tolerant genotypes. No genes associated with pigmentation were identified in the *Fsp-Ps2.1* region during this analysis.

Legumes contain the isoflavone synthase enzyme, which redirects phenylpropanoid pathway intermediates, such as naringenin, to the synthesis of isoflavonoid phytoalexins (Sreevidya et al., 2006). The isoflavonoid phytoalexins are low molecular weight antimicrobial compounds (Smith and Banks, 1986; Jeandet et al., 2014). Pisatin is an extensively studied phytoalexin from pea. In pea, the presence of *Fsp* increases the production of pisatin (Hadwiger and Beckman, 1980). The 6a-hydroxymaackiain-3-O-methyltransferase, enzyme directly upstream from the synthesis of pisatin, was expressed in both the tolerant and susceptible genotypes but overexpressed only in the susceptible genotypes at 0 h. Degradation of pisatin is an important mechanism by which *Fsp* resists pisatin and a crucial factor in the pathogenicity of *Fsp* in pea (Mackintosh et al., 1989). However, based on the results in this and previous studies, pisatin does not seem to play a role in the tolerance to *Fsp* (Mackintosh et al., 1989; Hadwiger, 2008).

The lignin biosynthetic pathway involves the central phenylpropanoid biosynthetic pathway. Genes involved in the lignin biosynthetic pathway, such as PAL, 4CL, trans-cinnamate 4-monooxygenase (C4M), and caffeoyl-o-methyltransferase (COMT), were overexpressed upon *Fsp* inoculation in both genotypes but at a significantly higher level and more consistently in the susceptible genotypes. It is well documented that the lignin biosynthetic pathway produces lignin rapidly in response to cell wall structure perturbations (Caño-Delgado et al., 2003; Tronchet et al., 2010; Sattler and Funnell-Harris, 2013; Miedes et al., 2014). Therefore, it seems the susceptible genotypes are responding to the aggressive *Fsp* invasion with a late and futile effort that involves a higher level of lignin synthesis and deposition.

Sugar Metabolism

This study identified DECs involved in sugar transport such as sugar transporter ERD6-like 6, sugar carrier protein C-like, sucrose transport protein SUC3, sugar transport protein 13, probable alkaline/neutral invertase D, bidirectional sugar transporter SWEET2-like, and invertase inhibitor-like protein. These sugars transporters are upregulated in the susceptible genotypes (**Supplementary Table S3**).

Certain pathogens are known to modulate the expression and activity of sugar transporters during their interaction with the plant host. Bacterial and fungal pathogens induce the overexpression of different sugar efflux transporters, such as the SWEET genes; this overexpression results in sucrose accumulating in the apoplast for use in pathogen nutritional gain and growth (Chen et al., 2010; Lanubile et al., 2015). The results of this study are in concordance with the literature, suggesting that the active mobilization of sucrose in the *Fsp*-inoculated susceptible genotypes supported successful infection by *Fsp*. Of all the DECs identified in the susceptible genotypes, 78% (25 genes) were overexpressed, and only 22% (7 genes) were suppressed after the *Fsp* challenge. In the tolerant genotypes, 15% (3 genes) were overexpressed, and 85% (17 genes) were suppressed after the *Fsp* challenge. These data would support the scenario explained previously; *Fsp* is either manipulating sugar metabolism or taking advantage of the active mobilization of sucrose in the susceptible genotypes.

Phytohormones

A large group of DECs were identified that were involved in the synthesis and signaling of salicylic acid (SA), jasmonic acid (JA), and ethylene (ET). The GO enrichment analysis also showed overrepresented GO terms related to the synthesis and signaling of these three hormones. From this set of DECs, a few were overexpressed in the tolerant genotypes, but the vast majority were overexpressed in the susceptible genotypes after the *Fsp* challenge.

It is generally assumed that the trophic nature of the pathogen determines which signal transduction pathway (SA or JA/Et) becomes activated in the plant host (Halim et al., 2006). Biotrophic pathogens generally elicit the defense responses via the salicylic acid signaling pathway, while necrotrophs activate a JA-dependent defense response (McDowell and Dangl, 2000; Dangl and Jones, 2001; Thomma et al., 2001; Van Wees et al., 2003; Grant and Lamb, 2006; Halim et al., 2006; Trusov et al., 2009; El Rahman et al., 2012). SA production promotes cell death, and that in turn promotes additional SA production. SA-signaling travels through the plant activating systemic acquired resistance (SAR) (Glazebrook, 2005). Necrotrophic pathogens have been shown to hijack plant defense responses to produce SA to further promote disease development. The necrotrophic fungi, *Botrytis cinerea*, and *Alternaria solani* use the SA-signaling pathway to exacerbate the disease in tomato (El Rahman et al., 2012). Both pathogens use the SA-signaling pathway through NPR1, a master regulator of SA signaling, and TGA1a TF to promote disease development in tomato. NPR1 and TGA1a suppress the expression of proteinase inhibitors, which in turn suppress the expression of two JA-dependent defense genes (El Rahman et al., 2012).

In this study, differential expression of genes associated with SA synthesis and signaling, cell death and HR in both the tolerant and susceptible genotypes was observed. However, when comparisons were made between the tolerant and susceptible genotypes, these DECs were observed to be overexpressed in the susceptible genotypes. The contig DN5429 in this study was identified as TGA1.a transcription factor (involved in SA signaling) and its isoform DN5429_c0_g1_i1 was overexpressed 6.82-fold in the susceptible genotypes over the tolerant genotypes. These are important observations, at least at the gene expression level, since the current understanding portends that overexpression of SA-related genes should not be observed in interactions between *Fsp* and pea, as cell death in plant hosts does not limit pathogen growth (Glazebrook, 2005).

SA-dependent responses and SAR are not predicted to play a role in tolerance against *Fsp*, whereas responses mediated by JA and ET are expected to do so.

The overexpression of genes associated with JA/ET synthesis and signaling also happens in both the tolerant and susceptible genotypes. When comparisons were made between the tolerant and susceptible genotypes, 22 DECs (80% of JA/ET-associated genes) were observed to be overexpressed in the susceptible genotypes. Based on these data, it is difficult to draw conclusions on the effect of SA and/or JA/ET in the response of the tolerant and susceptible genotypes to *Fsp*. However, a comparison was made using the number of DECs associated with the SA and JA/ET biosynthetic and signaling pathway in the tolerant and susceptible genotypes after the *Fsp* challenge (**Figure 8**). The susceptible genotypes showed an upsurge in the overexpression of genes related to SA biosynthesis and signaling (**Figure 8**). Alternatively, in the tolerant genotypes, the majority of genes related to SA biosynthesis and signaling were suppressed after the *Fsp* challenge. These changes related to the SA-pathway genes in the susceptible genotypes might have (1) deteriorated the action of the JA-signaling pathway, (2) increased the cell death, and therefore, (3) facilitated successful infection by *Fsp*. Genetic mapping or gene knockouts are needed to evaluate the involvement of the SA-signaling pathway in the development of root rot disease caused by *Fsp* in the susceptible genotypes. Some potential targets for gene knockout experiments would be NPR1 and TGA1.a.

Others: Cell Wall and Membrane Metabolism and Toxin Metabolism

Cell wall and membranes play important roles in plant defense as they act as a barrier that prevents pathogen invasion (Vorwerk et al., 2004). In this study, genes related to the cell wall and membrane modification, and callose deposition, were mostly upregulated in the susceptible genotypes (**Figure 9**). The genes for cell membrane transporters, proteins that work on the detoxification (antiporter activity) of substances, and proteins that break down toxins accumulated in the plant host were also overexpressed in the susceptible genotypes (**Figure 9**). Interestingly, after the *Fsp* challenge, the majority of genes associated with cell wall metabolism, toxin metabolism and transport were suppressed in the tolerant genotypes, while they were overexpressed in the susceptible genotypes (**Figure 9**). These responses, at least at the gene expression level, indicate that the response of the susceptible genotypes was delayed as the pathogen had already infested the tissues, and therefore, the host made a futile effort in response to the pathogen attack. Most likely, the tolerant genotypes already possessed physical and biochemical barriers, and the expression of the genes related to these pathways were actually being suppressed.

CONCLUSION

The time course RNAseq results presented in this study provided a comprehensive insight into the transcriptomic changes that accompany *Fsp* infection in tolerant and susceptible *P. sativum* genotypes. Eight different pea genotypes used in this study represent four most tolerant and four most susceptible genotypes identified in a preceding study based on root disease severity (RDS) index (Bodah et al., 2016). While phenotypically they were either tolerant or susceptible, four genotypes in each category represent different parentage. It is hypothesized that each of the genotypes possesses different tolerance or susceptible alleles. The identification of several key disease-related genes including, DRR230 (Chiang and Hadwiger, 1991), a known *Fsp*-tolerance gene, is indicative that a complementary approach used in this study could be utilized for other self-pollinating species for the identification of stimulus-responsive genes.

The observed changes in the expression of genes are associated with various physiological and biochemical processes that are known to be involved in plant disease response against pathogens. *Fsp* challenge produced a more intense and diverse overexpression of genes, across the entire time-course, in the susceptible genotypes compared to the tolerant genotypes. This type of response is hypothesized to be related to the speed at which the pathogen infestations advances in the susceptible genotypes and the preexisting level of disease-preparedness in the tolerant genotypes. The transcriptomic effort demonstrated by the susceptible genotypes seems futile and lacked key specific responses that were present in the tolerant genotypes. In contrast, the tolerant genotypes showed a fine-tuned response: fewer changes in the expression of defense-related genes that helps preserve energy and a faster reset to a basal metabolic state.

This RNAseq analysis helped identify alternate strategies and potential genes that could be evaluated to confer improved tolerance against root rot in *P. sativum*. Specific genes or pathways that might have a key role in tolerance or susceptibility to *Fsp* are as follows: receptor-like cytoplasmic kinase 176, CC-NBS-LRR resistance protein, *WRKY7* TF, WRKY4 TF, *LBD* TF, HMG A TF, anthocyanin biosynthetic pathway, SWEET genes, JA/ET-signaling pathway, cell death, NPR1, and TGA1.a. SA-signaling genes, and most importantly, the DRR230 protein. Functional characterization of these genes is expected to provide mechanistic information regarding pea-*Fsp* interaction, as well as gene targets for breeding. Since *Fsp* tolerance is a polygenic trait, pyramiding *Fsp* tolerance into pea cultivars is laborious but remains the most sustainable option, and it can be pursued rapidly via molecular breeding (Bodah et al., 2016) or gene editing (Ghogare et al., 2019). The rapid development of *Fsp* tolerant pea cultivars is critical for reducing yield losses and address the increasing demand for pea-derived protein for human nutrition.

DATA AVAILABILITY STATEMENT

The raw sequencing data used for RNAseq analysis is publicly available from the NCBI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under the accession number PRJNA630497.

AUTHOR CONTRIBUTIONS

AD, RS, and BW-B designed the study. AD supervised the study. BW-B and EB performed the experiments and generated the data. BW-B and GN analyzed the data. LP provided the tolerant pea genotypes and *Fsp* isolates. All authors read and approved the final manuscript.

FUNDING

BW-B acknowledges graduate research assistantship support from Washington State University Graduate School. Work in the Dhingra lab was supported in part by Washington State University Agriculture Research Center Hatch Grant WNP00011 and USA Dry Pea & Lentil Council. The authors acknowledge that this study was in part supported by funding from ProGene Plant Research. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

The authors are grateful to Kurt Braunwart, CEO, ProGene Plant Research for critical discussions and his support for the project. This manuscript has been released as a pre-print at bioRxiv (Willamson-Benavides et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00950/full#supplementary-material

REFERENCES

Akhtar, M. S., and Azam, T. (2014). Effects of PGPR and antagonistic fungi on the growth, enzyme activity and fusarium root-rot of pea. *Arch. Phytopathol. Plant Prot.* 47, 138–148. doi: 10.1080/03235408.2013.805487

Almeida, M. S., Cabral, K. M. S., Zingali, R. B., and Kurtenbach, E. (2000). Characterization of two novel defense peptides from pea (Pisum sativum) seeds. *Arch. Biochem. Biophys.* 378, 278–286. doi: 10.1006/abbi.2000.1824

Alves, M. S., Dadalto, S. P., Gonçalves, A. B., De Souza, G. B., Barros, V. A., and Fietto, L. G. (2013). Plant bZIP transcription factors responsive to pathogens: a review. *Int. J. Mol. Sci.* 14, 7815–7828. doi: 10.3390/ijms14047815

Ao, Y., Li, Z., Feng, D., Xiong, F., Liu, J., Li, J., et al. (2014). Os CERK 1 and Os RLCK 176 play important roles in peptidoglycan and chitin signaling in rice innate immunity. *Plant J.* 80, 1072–1084. doi: 10.1111/tpj.12710

Bagnaresi, P., Biselli, C., Orrù, L., Urso, S., Crispino, L., Abbruscato, P., et al. (2012). Comparative transcriptome profiling of the early response to Magnaporthe oryzae in durable resistant vs susceptible rice (Oryza sativa L.) genotypes. *PLoS One* 7:e51609. doi: 10.1371/journal.pone.0051609

Beck, M., Heard, W., Mbengue, M., and Robatzek, S. (2012). The INs and OUTs of pattern recognition receptors at the cell surface. *Curr. Opin. Plant Biol.* 15, 367–374. doi: 10.1016/j.pbi.2012.05.004

Bodah, E. T. (2017). Root rot diseases in plants: a review of common causal agents and management strategies. *Agric. Res. Tech. Open Access J.* 5:555661.

Bodah, E. T., Porter, L. D., Chaves, B., and Dhingra, A. (2016). Evaluation of pea accessions and commercial cultivars for fusarium root rot resistance. *Euphytica* 208, 63–72. doi: 10.1007/s10681-015-1545-6

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Caño-Delgado, A., Penfield, S., Smith, C., Catley, M., and Bevan, M. (2003). Reduced cellulose synthesis invokes lignification and defense responses in Arabidopsis thaliana. *Plant J.* 34, 351–362. doi: 10.1046/j.1365-313x.2003.01729.x

Chandran, D., Rickert, J., Huang, Y., Steinwand, M. A., Marr, S. K., and Wildermuth, M. C. (2014). Atypical E2F transcriptional repressor DEL1 acts at the intersection of plant growth and immunity by controlling the hormone salicylic acid. *Cell Host Microbe* 15, 506–513. doi: 10.1016/j.chom.2014.03.007

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X.-Q., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* 468, 527–532.

Chiang, C. C., and Hadwiger, L. A. (1991). The Fusarium solani-induced expression of a pea gene family encoding high cysteine content proteins. *Mol. Plant Microbe Interact.* 4, 324–331.

Chittem, K., Mathew, F. M., Gregoire, M., Lamppa, R. S., Chang, Y. W., Markell, S. G., et al. (2015). Identification and characterization of Fusarium spp. associated with root rots of field pea in North Dakota. *Eur. J. Plant Pathol.* 143, 641–649. doi: 10.1007/s10658-015-0714-8

Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832

Coyne, C. J., Pilet−Nayel, M., McGee, R. J., Porter, L. D., Smıkal, P., and Grünwald, N. J. (2015). Identification of QTL controlling high levels of partial resistance to Fusarium solani f. sp. pisi in pea. *Plant Breed.* 134, 446–453.

Coyne, C. J., Porter, L. D., Boutet, G., Ma, Y., McGee, R. J., Lesné, A., et al. (2019). Confirmation of Fusarium root rot resistance QTL Fsp-Ps 2.1 of pea under controlled conditions. *BMC Plant Biol.* 19:98. doi: 10.1186/s12870-019-1699-9

Coyne, C. J., Porter, L. D., Inglis, D. A., Grünwald, N. J., McPhee, K. E., and Muehlbauer, F. J. (2008). Registration of W6 26740, W6 26743, and W6 26745 green pea germplasm resistant to fusarium root rot. *J. Plant Regist.* 2, 137–139. doi: 10.3198/jpr2007.12.0674crg

Dangl, J. L., and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161

do Carmo, C. S., Nunes, A. N., Silva, I., Maia, C., Poejo, J., Ferreira-Dias, S., et al. (2016). Formulation of pea protein for increased satiety and improved foaming properties. *RSC Adv.* 6, 6048–6057. doi: 10.1039/c5ra22452g

Dong, J., Chen, C., and Chen, Z. (2003). Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* 51, 21–37.

Driouich, A., Jauneau, A., and Staehelin, L. A. (1997). 7-Dehydrobrefeldin A, a naturally occurring brefeldin A derivative, inhibits secretion and causes a cis-to-trans breakdown of Golgi stacks in plant cells. *Plant Physiol.* 113, 487–492. doi: 10.1104/pp.113.2.487

Duarte, J., Rivière, N., Baranger, A., Aubert, G., Burstin, J., Cornet, L., et al. (2014). Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics* 15:126. doi: 10.1186/1471-2164-15-126

El Rahman, T. A., El Oirdi, M., Gonzalez-Lamothe, R., and Bouarab, K. (2012). Necrotrophic pathogens use the salicylic acid signaling pathway to promote disease development in tomato. *Mol. Plant Microbe Interact.* 25, 1584–1593. doi: 10.1094/mpmi-07-12-0187-r

Feng, J., Hwang, R., Chang, K. F., Conner, R. L., Hwang, S. F., Strelkov, S. E., et al. (2011). Identification of microsatellite markers linked to quantitative trait loci controlling resistance to Fusarium root rot in field pea. *Can. J. Plant Sci.* 91, 199–204. doi: 10.4141/CJPS09176

Geiser, D. M., del Mar Jiménez-Gasco, M., Kang, S., Makalowska, I., Veeraraghavan, N., Ward, T. J., et al. (2004). FUSARIUM-ID v. 1.0: a DNA sequence database for identifying Fusarium. *Eur. J. plant Pathol.* 110, 473–479. doi: 10.1007/978-1-4020-2285-2_2

Ghogare, R., Williamson-Benavides, B., Ramírez-Torres, F., and Dhingra, A. (2019). CRISPR-associated nucleases: the Dawn of a new age of efficient crop improvement. *Transgenic Res.* 29, 1–35. doi: 10.1007/s11248-019-00181-y

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 43, 205–227. doi: 10.1146/annurev.phyto.43.040204.135923

Gossen, B. D., Conner, R. L., Chang, K. F., Pasche, J. S., McLaren, D. L., Henriquez, M. A., et al. (2016). Identifying and managing root rot of pulses on the northern great plains. *Plant Dis.* 100, 1965–1978. doi: 10.1094/PDIS-02-16-0184-FE

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644. doi: 10.1038/nbt.1883

Grand View Research (2015). *Pea Protein Market by Type (Isolates, Concentrates, Textured), Application, Textured Pea Protein by Form (Dry, Wet), & by Region – Global Trends & Forecast to 2020*. Pune: Grand View Research.

Grant, M., and Lamb, C. (2006). Systemic immunity. *Curr. Opin. Plant Biol.* 9, 414–420.

Grünwald, N. J., Coffman, V. A., and Kraft, J. M. (2003). Sources of partial resistance to Fusarium root rot in the Pisum core collection. *Plant Dis.* 87, 1197–1200. doi: 10.1094/pdis.2003.87.10.1197

Hadwiger, L. A. (2008). Pea–Fusarium solani interactions contributions of a system toward understanding disease resistance. *Phytopathology* 98, 372–379. doi: 10.1094/phyto-98-4-0372

Hadwiger, L. A. (2015). Anatomy of a nonhost disease resistance response of pea to Fusarium solani: PR gene elicitation via DNase, chitosan and chromatin alterations. *Front. Plant Sci.* 6:373. doi: 10.3389/fpls.2015.00373

Hadwiger, L. A., and Adams, M. J. (1978). Nuclear changes associated with the host-parasite interaction between Fusarium solani and peas. *Physiol. Plant Pathol.* 12, 63–72. doi: 10.1016/0048-4059(78)90019-x

Hadwiger, L. A., and Beckman, J. M. (1980). Chitosan as a component of pea-Fusarium solani interactions. *Plant Physiol.* 66, 205–211. doi: 10.1104/pp.66.2.205

Halim, V. A., Vess, A., Scheel, D., and Rosahl, S. (2006). The role of salicylic acid and jasmonic acid in pathogen defence. *Plant Biol.* 8, 307–313. doi: 10.1055/s-2006-924025

Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., et al. (2014). Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. *Proc. Natl. Acad. Sci. U.S.A.* 111, E404–E413.

Im, J. H., Lee, S. G., Lee, E., Park, S. R., Ahn, I., and Hwang, D.-J. (2019). OsbZIP75 positively regulates plant defense against the bacterial leaf blight pathogen Xanthomonas oryzae pv. oryzae. *Plant Biotechnol. Rep.* 13, 645–651. doi: 10.1007/s11816-019-00561-0

Isaac, J., Hartney, S. L., Druffel, K., and Hadwiger, L. A. (2009). The non-host disease resistance response in peas; alterations in phosphorylation and ubiquitination of HMG A and histones H2A/H2B. *Plant Sci.* 177, 439–449. doi: 10.1016/j.plantsci.2009.07.007

Jeandet, P., Hébrard, C., Deville, M.-A., Cordelier, S., Dorey, S., Aziz, A., et al. (2014). Deciphering the role of phytoalexins in plant-microorganism interactions and human health. *Molecules* 19, 18033–18056. doi: 10.3390/molecules191118033

Johansson, M. (2019). *Pea protein-Systems for Plant Based Protein Products, The Effect of Insoluble and Soluble Lentil Fractions on Rheology, Microstructure and Gelation Properties of Heat Induced Pea Protein Gels*. Göteborg: Chalmers tekniska högskola.

Kendra, D. F., and Hadwiger, L. A. (1984). Characterization of the smallest chitosan oligomer that is maximally antifungal to Fusarium solani and elicits pisatin formation in Pisum sativum. *Exp. Mycol.* 8, 276–281. doi: 10.1016/0147-5975(84)90013-6

Kerr, A. (1963). The root rot-Fusarium wilt complex of peas. *Aust. J. Biol. Sci.* 16, 55–69.

Khong, G. N., Pati, P. K., Richaud, F., Parizot, B., Bidzinski, P., Mai, C. D., et al. (2015). OsMADS26 negatively regulates resistance to pathogens and drought tolerance in rice. *Plant Physiol.* 169, 2935–2949.

Kiełbowicz-Matuk, A. (2012). Involvement of plant C2H2-type zinc finger transcription factors in stress responses. *Plant Sci.* 185, 78–85. doi: 10.1016/j.plantsci.2011.11.015

Kim, K.-C., Fan, B., and Chen, Z. (2006). Pathogen-induced Arabidopsis WRKY7 is a transcriptional repressor and enhances plant susceptibility to *Pseudomonas* syringae. *Plant Physiol.* 142, 1180–1192. doi: 10.1104/pp.106.082487

Klosterman, S. J., Chen, J., Choi, J. J., Chinn, E. E., and Hadwiger, L. A. (2001). Characterization of a 20 kDa DNase elicitor from Fusarium solani f. sp. phaseoli and its expression at the onset of induced resistance in Pisum sativum. *Mol. Plant Pathol.* 2, 147–158. doi: 10.1046/j.1364-3703.2001.00062.x

Klosterman, S. J., Choi, J. J., and Hadwiger, L. A. (2003). Analysis of pea HMG—I/Y expression suggests a role in defence gene regulation. *Mol. Plant Pathol.* 4, 249–258. doi: 10.1046/j.1364-3703.2003.00171.x

Kraft, J. M. (1989). Registration of 86-638, 86-2197, 86-2231, and 86-2236 PEA Germplasms. *Crop Sci.* 29, 494–495.

Kraft, J. M. (1992). Registration of 90—2079—2131, and 90—2322 PEA Germplasms. *Crop Sci.* 32:1076. doi: 10.2135/cropsci1992.0011183x003200040063x

Kraft, J. M., and Giles, R. A. (1976). Registration of 74SN3, 74SN4, and 74SN5 PEA Germplasm1 (Reg. No. GP 15 to GP 17). *Crop Sci.* 16:126. doi: 10.2135/cropsci1976.0011183x001600010042x

Kraft, J. M., and Tuck, J. A. (1986). Registration of 75-786, 84-1638, and 84-1930 Pea Germplasms. *Crop Sci.* 26, 1262–1263. doi: 10.2135/cropsci1986.0011183x002600060053x

Kreplak, J., Madoui, M. A., Cápal, P., Novák, P., Labadie, K., Aubert, G., et al. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* 51, 1411–1422. doi: 10.1038/s41588-019-0480-1

Kumari, N., and Katoch, S. (2020). " "Wilt and root rot complex of important pulse crops: their detection and integrated management," in *Management of Fungal Pathogens in Pulses*, eds B. P. Singh, G. Singh, K. Kumar, S. Nayak, and N. Srinivasa (Berlin: Springer), 93–119. doi: 10.1007/978-3-030-35947-8_6

Kwon, C., and Yun, H. S. (2014). Plant exocytic secretion of toxic compounds for defense. *Toxicol. Res.* 30, 77–81. doi: 10.5487/tr.2014.30.2.077

Lai, Z., Vinod, K. M., Zheng, Z., Fan, B., and Chen, Z. (2008). Roles of ArabidopsisWRKY3 and WRKY4 transcription factors in plant responses to pathogens. *BMC Plant Biol.* 8:68. doi: 10.1186/1471-2229-8-68

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357. doi: 10.1038/nmeth.1923

Lanubile, A., Muppirala, U. K., Severin, A. J., Marocco, A., and Munkvold, G. P. (2015). Transcriptome profiling of soybean (Glycine max) roots challenged with pathogenic and non-pathogenic isolates of Fusarium oxysporum. *BMC Genomics* 16:1089. doi: 10.1186/s12864-015-2318-2

Li, S. (2015). The Arabidopsis thaliana TCP transcription factors: a broadening horizon beyond development. *Plant Signal. Behav.* 10:e1044192. doi: 10.1080/15592324.2015.1044192

Li, T., Chen, X., Zhong, X., Zhao, Y., Liu, X., Zhou, S., et al. (2013). Jumonji C domain protein JMJ705-mediated removal of histone H3 lysine 27 trimethylation is involved in defense-related gene activation in rice. *Plant Cell* 25, 4725–4736. doi: 10.1105/tpc.113.118802

Lorang, J. M., Sweat, T. A., and Wolpert, T. J. (2007). Plant disease susceptibility conferred by a "resistance" gene. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14861–14866. doi: 10.1073/pnas.0702572104

Mackintosh, S. F., Matthews, D. E., and VanEtten, H. D. (1989). Two additional genes for pisatin demethylation and their relationship to the pathogenicity of Nectria haematococca on pea. *Mol. Plant-Microbe Interact* 2, 354–362.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

Matič, S., Bagnaresi, P., Biselli, C., Carneiro, G. A., Siciliano, I., Valé, G., et al. (2016). Comparative transcriptome profiling of resistant and susceptible rice genotypes in response to the seedborne pathogen Fusarium fujikuroi. *BMC Genomics* 17:608. doi: 10.1186/s12864-016-2925-6

McDowell, J. M., and Dangl, J. L. (2000). Signal transduction in the plant immune response. *Trends Biochem. Sci.* 25, 79–82. doi: 10.1016/s0968-0004(99)01532-7

Miedes, E., Vanholme, R., Boerjan, W., and Molina, A. (2014). The role of the secondary cell wall in plant resistance to pathogens. *Front. Plant Sci.* 5:358. doi: 10.3389/fpls.2014.00358

Moffett, P., Farnham, G., Peart, J., and Baulcombe, D. C. (2002). Interaction between domains of a plant NBS–LRR protein in disease resistance−related cell death. *EMBO J.* 21, 4511–4519. doi: 10.1093/emboj/cdf453

Nash, S. M., and Snyder, W. C. (1962). Quantitative estimations by plate counts of propagules of the bean root rot Fusarium in field soils. *Phytopathology* 52, 567–572.

National Center for Biotechnology Information (2020). *National Library of Medicine (US)*. Bethesda, MD: National Center for Biotechnology Information.

Nyamsuren, O., Colditz, F., Rosendahl, S., Tamasloukht, M., Bekel, T., Meyer, F., et al. (2003). Transcriptional profiling of Medicago truncatula roots after infection with Aphanomyces euteiches (oomycota) identifies novel genes upregulated during this pathogenic interaction. *Physiol. Mol. Plant Pathol.* 63, 17–26. doi: 10.1016/j.pmpp.2003.09.001

Paik, I., Kathare, P. K., Kim, J.-I., and Huq, E. (2017). Expanding roles of PIFs in signal integration from multiple processes. *Mol. Plant* 10, 1035–1046. doi: 10.1016/j.molp.2017.07.002

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14:417. doi: 10.1038/nmeth.4197

Peng, W., Kong, X., Chen, Y., Zhang, C., Yang, Y., and Hua, Y. (2016). Effects of heat treatment on the emulsifying properties of pea proteins. *Food Hydrocoll.* 52, 301–310. doi: 10.1016/j.foodhyd.2015.06.025

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.* 29:e45.

Pietrysiak, E., Smith, D. M., Smith, B. M., and Ganjyal, G. M. (2018). Enhanced functionality of pea-rice protein isolate blends through direct steam injection processing. *Food Chem.* 243, 338–344. doi: 10.1016/j.foodchem.2017.09.132

Pontier, D., Miao, Z., and Lam, E. (2001). Trans−dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses. *Plant J.* 27, 529–538. doi: 10.1046/j.1365-313x.2001.01086.x

Poore, J., and Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science* 360, 987–992. doi: 10.1126/science.aaq0216

Porter, L. (2010). Identification of tolerance to Fusarium root rot in wild pea germplasm with high levels of partial resistance. *Pisum Genet.* 42, 1–6.

Porter, L. D., Kraft, J. M., and Grünwald, N. J. (2014). Release of pea germplasm with Fusarium resistance combined with desirable yield and anti-lodging traits. *J. Plant Regist.* 8, 191–194. doi: 10.3198/jpr2013.07.0041crg

Pré, M., Atallah, M., Champion, A., De Vos, M., Pieterse, C. M. J., and Memelink, J. (2008). The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiol.* 147, 1347–1357. doi: 10.1104/pp.108.117523

Prioul-Gervais, S., Deniot, G., Receveur, E. M., Frankewitz, A., Fourmann, M., Rameau, C., et al. (2007). Candidate genes for quantitative resistance to Mycosphaerella pinodes in pea (Pisum sativum L.). *Theor. Appl. Genet.* 114, 971–984. doi: 10.1007/s00122-006-0492-y

Research and Markets (2019). *Plant Based Protein Market by Type (Soy Protein, Wheat Protein, Pea Protein, Potato Protein, Rice Protein, Corn Protein) and Application (Foods and Beverage, Animal Feed, Nutrition and Health Supplements, Pharmaceuticals) - Global Forecast to 2025*. Dublin: Research and Markets.

Rozen, S., and Skaletsky, H. (2000). "Primer3 on the WWW for general users and for biologist programmers," in *Bioinformatics Methods and Protocols*, eds S. Misener and S. A. Krawetz (Berlin: Springer), 365–386. doi: 10.1385/1-59259-192-2:365

Ruijter, J. M., Van der Velden, S., and Ilgun, A. (2009). *LinReg PCR (11.0). Analysis of Quantitative RT-PCR Data*. Amsterdam: Heart Failure Research Center.

Sattler, S., and Funnell-Harris, D. (2013). Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens? *Front. Plant Sci.* 4:70. doi: 10.3389/fpls.2013.00070

Seaman, W. L. (1976). Yield loss conversion factors for fusarium root rot of peal. *Can. Plant Dis. Surv.* 56, 25–32.

Smith, D. A., and Banks, S. W. (1986). Biosynthesis, elicitation and biological activity of isoflavonoid phytoalexins. *Phytochemistry* 25, 979–995. doi: 10.1016/s0031-9422(00)81542-3

Solfanelli, C., Poggi, A., Loreti, E., Alpi, A., and Perata, P. (2006). Sucrose-specific induction of the anthocyanin biosynthetic pathway in Arabidopsis. *Plant Physiol.* 140, 637–646. doi: 10.1104/pp.105.072579

Song, S., Qi, T., Fan, M., Zhang, X., Gao, H., Huang, H., et al. (2013). The bHLH subgroup IIId factors negatively regulate jasmonate-mediated plant defense and development. *PLoS Genet.* 9:e1003653. doi: 10.1371/journal.pgen.1003653

Sreevidya, V. S., Srinivasa Rao, C., Sullia, S. B., Ladha, J. K., and Reddy, P. M. (2006). Metabolic engineering of rice with soybean isoflavone synthase for promoting nodulation gene expression in rhizobia. *J. Exp. Bot.* 57, 1957–1969. doi: 10.1093/jxb/erj143

Stehfest, E., Bouwman, L., Van Vuuren, D. P., Den Elzen, M. G. J., Eickhout, B., and Kabat, P. (2009). Climate benefits of changing diet. *Clim. Change* 95, 83–102. doi: 10.1007/s10584-008-9534-6

Sweat, T. A., and Wolpert, T. J. (2007). Thioredoxin h5 is required for victorin sensitivity mediated by a CC-NBS-LRR gene in Arabidopsis. *Plant Cell* 19, 673–687. doi: 10.1105/tpc.106.047563

Thatcher, L. F., Powell, J. J., Aitken, E. A. B., Kazan, K., and Manners, J. M. (2012). The lateral organ boundaries domain transcription factor LBD20 functions in Fusarium wilt susceptibility and jasmonate signaling in Arabidopsis. *Plant Physiol.* 160, 407–418. doi: 10.1104/pp.112.199067

Thomma, B. P. H. J., Penninckx, I. A. M. A., Cammue, B. P. A., and Broekaert, W. F. (2001). The complexity of disease signaling in *Arabidopsis*. *Curr. Opin. Immunol.* 13, 63–68. doi: 10.1016/s0952-7915(00)00183-7

Tronchet, M., Balague, C., Kroj, T., Jouanin, L., and Roby, D. (2010). Cinnamyl alcohol dehydrogenases−C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in *Arabidopsis*. *Mol. Plant Pathol.* 11, 83–92. doi: 10.1111/j.1364-3703.2009.00578.x

Trusov, Y., Sewelam, N., Rookes, J. E., Kunkel, M., Nowak, E., Schenk, P. M., et al. (2009). Heterotrimeric G proteins−mediated resistance to necrotrophic pathogens includes mechanisms independent of salicylic acid−, jasmonic acid/ethylene−and abscisic acid−mediated defense signaling. *Plant J.* 58, 69–81. doi: 10.1111/j.1365-313x.2008.03755.x

USDA–ARS NGRP (2020). *Germplasm Resources Information Network- (GRIN)*. Washington, DC: USDA–ARS NGRP.

Vailleau, F., Daniel, X., Tronchet, M., Montillet, J.-L., Triantaphylides, C., and Roby, D. (2002). A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10179–10184. doi: 10.1073/pnas.152047199

Van Wees, S. C. M., Chang, H.-S., Zhu, T., and Glazebrook, J. (2003). Characterization of the early response of *Arabidopsis* to *Alternaria brassicicola* infection using expression profiling. *Plant Physiol.* 132, 606–617. doi: 10.1104/pp.103.022186

Vandemark, G. J., Brick, M. A., Osorno, J. M., Kelly, J. D., and Urrea, C. A. (2014). *Edible Grain Legumes. Yield Gains Major US F. Crop*. Scottsbluff, NE: University of Nebraska Panhandle Research & Extension Center, 87–124.

Vorwerk, S., Somerville, S., and Somerville, C. (2004). The role of plant cell wall polysaccharide composition in disease resistance. *Trends Plant Sci.* 9, 203–209. doi: 10.1016/j.tplants.2004.02.005

Wang, X., Basnayake, B. M. V. S., Zhang, H., Li, G., Li, W., Virk, N., et al. (2009). The Arabidopsis ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens. *Mol. Plant-Microbe Interact.* 22, 1227–1238. doi: 10.1094/mpmi-22-10-1227

Wang, Y., Nowak, G., Culley, D., Hadwiger, L. A., and Fristensky, B. (1999). Constitutive expression of pea defense gene DRR206 confers resistance to blackleg (Leptosphaeria maculans) disease in transgenic canola (Brassica napus). *Mol. Plant Microbe Interact.* 12, 410–418. doi: 10.1094/mpmi.1999.12.5.410

Willamson-Benavides, B. A., Sharpe, R. M., Nelson, G. M., Bodah, E. T., Porter, L. D., and Dhingra, A. (2020). Identification of Fusarium solani f. sp. pisi (Fsp) responsive genes in *Pisum sativum*. *bioRxiv*[Preprint] doi: 10.1101/2020.05.12.091892

World Health Organization (2015). *Obesity and overweight*. Geneva: WHO.

Xiong, T., Xiong, W., Ge, M., Xia, J., Li, B., and Chen, Y. (2018). Effect of high intensity ultrasound on structure and foaming properties of pea protein isolate. *Food Res. Int.* 109, 260–267. doi: 10.1016/j.foodres.2018.04.044

Xu, L., Ravnskov, S., Larsen, J., and Nicolaisen, M. (2012). Linking fungal communities in roots, rhizosphere, and soil to the health status of *Pisum sativum*. *FEMS Microbiol. Ecol.* 82, 736–745. doi: 10.1111/j.1574-6941.2012.01445.x

Zhang, N., O'Donnell, K., Sutton, D. A., Nalim, F. A., Summerbell, R. C., Padhye, A. A., et al. (2006). Members of the Fusarium solani species complex that cause

infections in both humans and plants are common in the environment. *J. Clin. Microbiol.* 44, 2186–2190. doi: 10.1128/jcm.00120-06

Zhang, Y., Liu, Q., Zhang, Y., Chen, Y., Yu, N., Cao, Y., et al. (2019). LMM24 Encodes Receptor-Like Cytoplasmic Kinase 109, Which Regulates Cell Death and Defense Responses in Rice. *Int. J. Mol. Sci.* 20, 3243. doi: 10.3390/ijms20133243

Zheng, W., Ma, L., Zhao, J., Li, Z., Sun, F., and Lu, X. (2013). Comparative transcriptome analysis of two rice varieties in response to rice stripe virus and small brown planthoppers during early interaction. *PLoS One* 8:e82126. doi: 10.1371/journal.pone.0082126

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Williamson-Benavides, Sharpe, Nelson, Bodah, Porter and Dhingra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 August 2020
doi: 10.3389/fgene.2020.00873

Dissecting the Molecular Function of *Triticum aestivum STI* Family Members Under Heat Stress

*Shaloo Meena, Sohini Deb†, Harsha Samtani and Paramjit Khurana**

Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi, India

OPEN ACCESS

Edited by:
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Manoj Prasad, National Institute of Plant Genome Research (NIPGR), India
Cheng Qin, Zunyi Vocational and Technical College, China
Sachin Rustgi, Clemson University, United States

***Correspondence:**
Paramjit Khurana
param@genomeindia.org

†Present address:
Sohini Deb, Centre for Cellular and Molecular Biology, Hyderabad, India

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 07 March 2020
Accepted: 16 July 2020
Published: 19 August 2020

Citation:
Meena S, Deb S, Samtani H and Khurana P (2020) Dissecting the Molecular Function of Triticum aestivum STI Family Members Under Heat Stress. Front. Genet. 11:873. doi: 10.3389/fgene.2020.00873

STI/HOP functions as a co-chaperone of HSP90 and HSP70 whose molecular function has largely been being restricted as an adaptor protein. However, its role in thermotolerance is not well explored. In this article, we have identified six members of the *TaSTI* family, which were named according to their distribution on group 2 and group 6 chromosomes. Interestingly, *TaSTI-2* members were found to express higher as compared to *TaSTI-6* members under heat stress conditions, with *TaSTI-2A* being one of the most heat-responsive member. Consistent with this, the heterologous expression of *TaSTI-2A* in *Arabidopsis* resulted in enhanced basal as well as acquired thermotolerance as revealed by the higher yield of the plants under stress conditions. Similarly in the case of rice, *TaSTI-2A* transgenics exhibited enhanced thermal tolerance. Moreover, we demonstrate that TaSTI-2A interacts with TaHSP90 not only in the nucleus but also in the ER and Golgi bodies, which has not been shown till now. Additionally, TaHSP70 was also found to interact with TaSTI-6D specifically in the cytosol. Thus, these data together suggested that the *TaSTI* family members might play different roles under heat stress conditions in order to fine-tune the heat stress response in plants.

Keywords: co-chaperone, STI, heat stress, thermotolerance, endoplasmic reticulum (ER)

INTRODUCTION

Plants unlike animals cannot escape from unfavorable environmental conditions. Changes in ambient temperature become the most common form of stress encountered by plants. When the temperature reaches higher than that required for optimum growth, it leads to irreversible damage and is generally known as heat stress (HS) (Wahid et al., 2007). At the cellular level, HS causes the accumulation of misfolded proteins, which not only leads to loss in activity but also causes protein aggregation, thereby disturbing cell homeostasis (Nakajima and Suzuki, 2013). Thus, cells are equipped with a Quality Control (QC) system consisting of chaperones and degradative complexes. The QC system not only helps in the folding of the newly synthesized proteins but also recognizes the misfolded proteins and facilitates their refolding. However, if any protein is unable to fold correctly, then it is targeted for degradation (Buchberger et al., 2010; Houck et al., 2012). In plants, the expression of these chaperones is enhanced by a conserved mechanism known

as heat stress response (HSR) (Hahn et al., 2011). Among the different chaperones present in the cell, HSP70 and HSP90 play a major role in QC in the cytoplasm under HS conditions. HSP70 accumulates abundantly in the cells and binds to the hydrophobic patches of the misfolded proteins in an ATP-dependent manner, thereby preventing their aggregation (Mayer and Bukau, 2005; Hahn et al., 2011), whereas HSP90 has been found to regulate the activity of different transcription factors, kinases and steroid hormone receptors (Hahn et al., 2011). The activity of these chaperones is regulated by the binding of different co-chaperones that modulate their structural conformation and ATPase activity and, hence, the substrate selection (Li et al., 2012; Prodromou, 2012).

The HSP90/70 organizing protein (HOP) is the co-chaperone that functions as an adaptor in HSP90 and HSP70 machinery (Chen and Smith, 1998; Scheufler et al., 2000; Chen et al., 2010). HOP, also known as STI (stress-induced protein), was first identified in yeast during a genetic screen for proteins involved in HSR (Nicolet and Craig, 1989). Unlike HSP90 and HSP70, STI does not possess any chaperone activity (Bose et al., 1996; Freeman et al., 1996), but it consists of several tetratricopeptide repeat (TPR) domains which are involved in protein–protein interactions. The N-terminus of the TPR1 domain binds to the C-terminus of HSP70, and the central TPR2A domain binds to the C-terminus of HSP90 (Scheufler et al., 2000; Hernandez et al., 2002).

Various functional roles of STI has been reported in animals (Baindur-Hudson et al., 2015). STI has been isolated and characterized in humans as transformation-sensitive human protein (IEF-SSP3521), and its potential role has been suggested in cell proliferation or gene regulation (Honore et al., 1992). STI also interacts with cellular prions and is involved in neuritogenesis and leads to neuroprotection (Zanata et al., 2002; Lopes et al., 2005). In the case of *C. elegans*, the expression of *STI1* was found to be induced by HS. Also, the mutants of *STI1* in *C. elegans* showed a shorter life span and decreased fertility rate, thereby indicating its role in stress responses and in aging processes (Song et al., 2009). A recent report by Karam et al. has highlighted the unique role of STI in transposon silencing and Piwi-interacting RNA biogenesis (Karam et al., 2017).

In the case of plants, *STI* homologs have been described in *Arabidopsis thaliana, Glycine max*, and *Oryza sativa* (Zhang et al., 2003; Fellerer et al., 2011; Fernandez-Bautista et al., 2018). In *Arabidopsis*, the role of HOP has been described in alleviating ER stress response (Fernandez-Bautista et al., 2017). Also, STI was found as one of the partner proteins in HSP90-chloroplast preprotein complex, indicating its role in chaperoning preproteins in plant cytosol (Fellerer et al., 2011). Interestingly, many studies have uncovered its role in plant-pathogen defense responses. STI, along with HSP90, has been shown to interact with rice chitin receptor (OsCERK1) and is involved in innate immunity response against rice blast fungus (Chen et al., 2010). Similarly, STI has been found to act as a major cellular determinant for the mitochondrial *Carnation Italian Ringspot Tombusvirus* (CIRV) and the *Potato Virus Y* proliferation in *Nicotiana benthamiana* and tobacco, respectively (Xu et al., 2014; Lamm et al., 2017). Apart from these, the function of *STI* in various abiotic stresses still remains unclear. A recent study has shown the role of *Arabidopsis STI* family in long term-acquired thermotolerance (Fernandez-Bautista et al., 2018). In the case of *Triticum aestivum, STI* was found to be induced by HS (Chauhan et al., 2011). Thus, these evidences suggest that STI might have broader functions rather than being merely an adaptor protein.

In this study, we have identified six members of *STI* from wheat and analyzed their possible role under HS. Expression profiling of these members revealed their differential regulation, and their localization studies showed their presence in the nucleus, cytoplasm, and ER-Golgi complex. Interestingly, one of the most heat-responsive members, TaSTI-2A, was found to interact with TaHSP90 in the nucleus as well as in the ER and Golgi bodies. TaHSP70 was also found to interact with a different member of the STI gene family, i.e., TaSTI-6D in the cytosol, which has not yet been reported in the case of plants. The overexpression of *TaSTI-2A* promoted thermotolerance in both *A. thaliana* and *O. sativa*. Thus, our results suggest that *TaSTI* could serve as a potential gene for heat tolerance enhancement in crops.

MATERIALS AND METHODS

Identification and Structural Analysis of *STI* Gene Family Members From Wheat

STI protein sequences of *A. thaliana* were taken from the TAIR database (AT1G12270, AT1G62740, AT4G12400). These protein sequences were further used for Blast search against available *T. aestivum* genome (*T. aestivum*-Ensembl Genomes 41), and a total of six members were identified. One of the TaSTI members was previously identified in the lab from the HS cDNA library (GD189073), which was submitted to the NCBI database (Chauhan et al., 2011). This STI member was also found in the Ensembl database, and we have named it as TaSTI-2A. These protein sequences were checked for conserved STI and TPR domains using the NCBI CDD search and SMART domain tool analysis. Multiple sequence alignment was carried out by the CLUSTALW program, and the phylogenetic tree was constructed using the MEGA7 program by Neighbor-Joining method.

Moreover, the Phyre2 web portal was used for the prediction of the three-dimensional structure of *Ta*STI and *At*HOP proteins.

Expression Analysis of *TaSTI* Members in Different Genotypes of Wheat and in *A. thaliana*

Bread wheat (*T. aestivum*) cultivar PBW343 and C306 and *Arabidopsis* ecotype Col-0 were used in this study. Wheat seeds were surface-sterilized with 4% sodium hypochlorite for 20 min followed by five to six washes with autoclaved sterile water. After sterilization, seeds were grown on a cotton tray in a growth chamber (Conviron, Canada) maintained at 22 ± 1°C with a 16-h photoperiod. *Arabidopsis* plants were grown on half strength Murashige and Skoog (MS) media. For the stress treatment of wheat, 10-day-old seedlings were subjected to different abiotic

TABLE 1 | *STI* members identified in wheat and their basic information.

Gene name	Gene ID	Nucleotide (bp)	Exon	Intron	Amino acids	Mw (KDa)	Theoretical PI	Chromosomal location
TaSTI-2A	TraesCS2A02G386800	2275	6	5	577	64.61	6.77	2A
TaSTI-2B	TraesCS2B02G404400	2422	6	5	576	64.38	7.10	2B
TaSTI-2D	TraesCS2D02G383600	2270	6	5	572	64.14	6.77	2D
TaSTI-6A	TraesCS6A02G238600	2412	7	6	581	65.04	5.78	6A
TaSTI-6B	TraesCS6B02G285800	2415	7	6	580	65	5.83	6B
TaSTI-6D	TraesCS6D02G221000	2498	7	6	577	64.73	5.78	6D

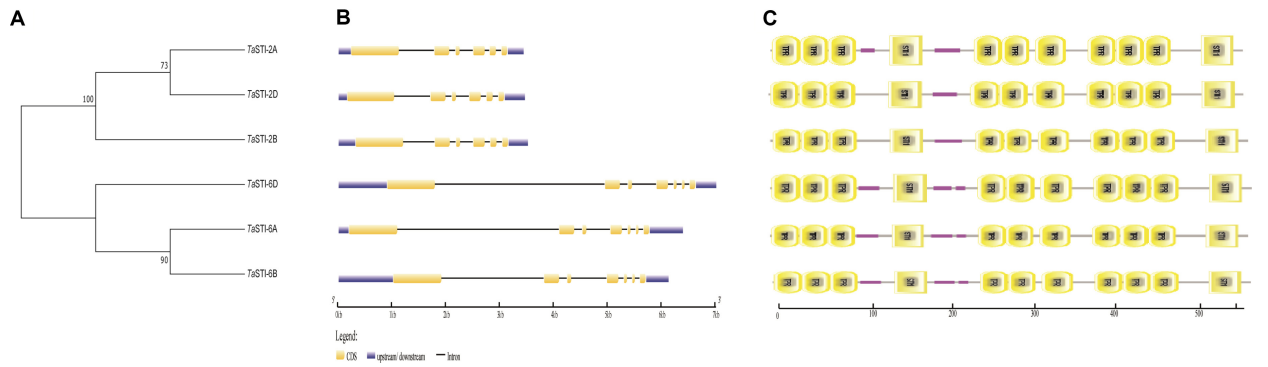


FIGURE 1 | Structural and domain analysis of TaSTI members. **(A)** Phylogenetic relationship between the *TaSTI* gene family members. **(B)** Gene structure analysis of *TaSTI* family members. Exons are depicted using solid boxes and introns are shown using lines. **(C)** TPR and STI domains of all the TaSTI members. Low complexity regions are depicted in solid pink line.

stresses such as heat (42°C for 2 h), cold (4°C for 24 h), salt (200 mM for 24 h), and drought (200 mM mannitol for 24 h) (Peng et al., 2009; Chauhan et al., 2011; Amoah et al., 2019; Hamdi et al., 2020). Ten-day-old seedlings of genotype PBW343 and C306 were subjected to different temperatures ranging from 25°C to 45°C. After the stress treatment, seedlings of control and treated plants were frozen in liquid nitrogen and stored at −80°C until RNA isolation. *Arabidopsis* plants were germinated on half MS media and then transferred to soilrite for further analysis.

Total RNA was isolated using the RNeasy plant mini kit (Qiagen, Germany) according to the manufacturer's instructions, including on-column DNaseI treatment to remove genomic DNA contamination. Two micrograms of the total RNA was used as a template to synthesize cDNA employing the High Capacity cDNA Archive kit (Applied Biosystems, United States) and mixed with 200 nM of each primer and SYBR Green PCR Master Mix (Applied Biosystems) for real-time PCR analysis, using the ABI Prism 7000 Sequence Detection System and Software (PE Applied Biosystems) according to the manufacturer's protocol. Relative fold change was calculated, and *Actin* was used as a housekeeping gene. Graphs were plotted using three biological and three technical replicates. The wheat expression database (Borrill et al., 2016) hosted at http://wheatexpression.com was used to analyze the expression profile of the *STI* members. This database was developed as an expression visualization and integration platform. Further, this database also hosts the normalized data for many development and stress treated wheat samples. Therefore, transcripts per kilobase million (TPM) values for all the *STI* members were downloaded from this database. TPM values were log-transformed (Log2X) in order to generate a heatmap using the gplots package and Rcolorbrewer package.

Cis Element Analysis of TaSTI Promoter Sequences

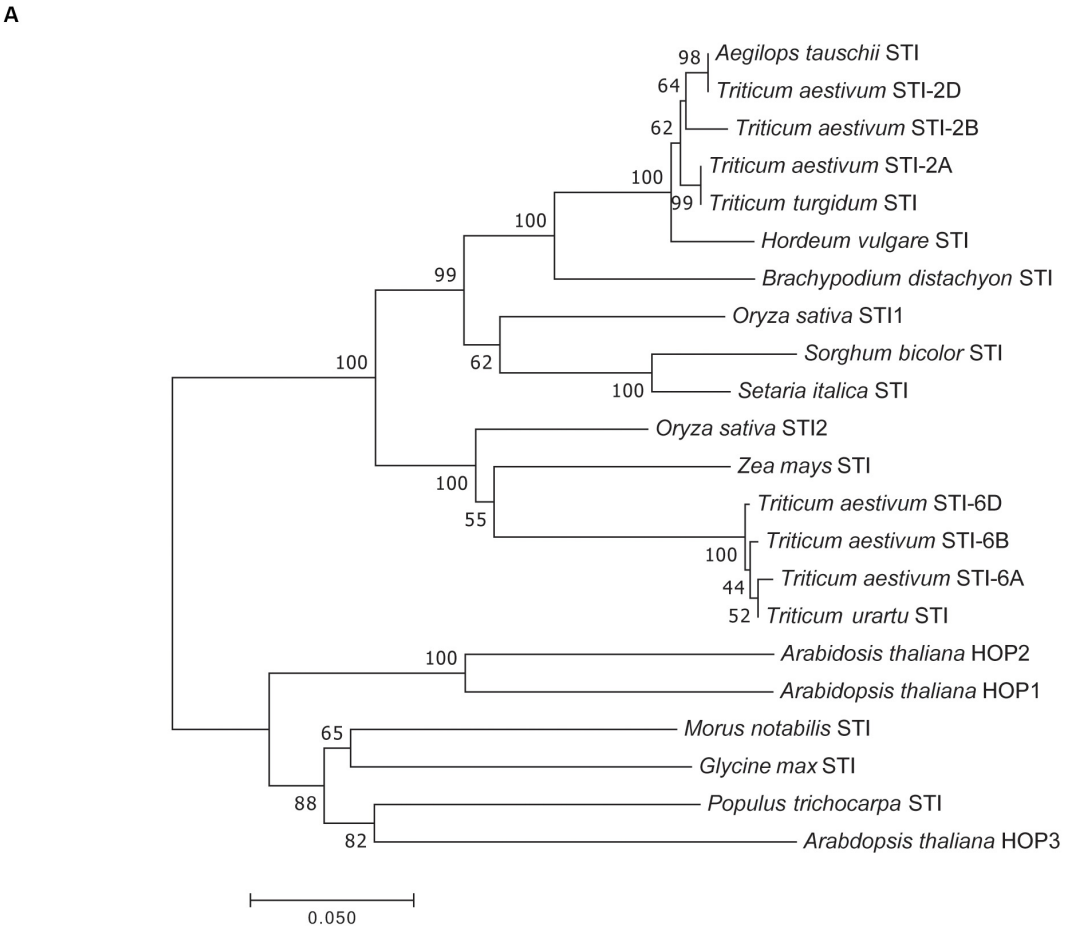
For the analysis of URRs (upstream regulatory regions), the 2-kb upstream region of all the *TaSTI* members was extracted from the Ensembl database (*T. aestivum*-Ensembl Genomes 41). The *cis*-elements in promoters were subsequently searched using the PLACE and PlantCare database (Higo et al., 1999; Lescot et al., 2002).

Yeast-2-Hybrid Assay

For the measurement of the interaction, the putative interactants, *TaSTI1-2A* and *TaHSP90* were cloned into pENTR/D-TOPO vectors and further into pDEST-GADT7 and pDEST-GBKT7 vectors (Clontech, United States). The recombinant plasmids were transformed into the yeast strain AH109 harboring the ADE3 and HIS3 reporter genes. The reporter gene activity was confirmed by a viability test on a medium lacking histidine, leucine, and tryptophan (-HLW) along with 0.5 mM 3-AT (3-Amino-1,2,4-triazole).

Complementation Assay

To check for complementation, the full-length *TaSTI*-2A gene was cloned into a pYES2 vector between the *Eco*RI and *Hin*dIII sites. For the confirmation of complementation, the colonies obtained after transformation were dotted with increasing



B

1. Aegilops tauschii STI	L K E R Q G K A M Q D P E I Q N I L T D P V M R Q V L I D F Q E N P R A A Q D H L K N P G V K Q K I Q K L V S A
2. Brachypodium distachyon STI	L K E R Q D K A M H D P E I Q N I L T D P V M R Q V L I D F Q E N P R S A Q D H L K N P G V K Q K I M K L V S A
3. Sorghum bicolor STI	L K E R Q N K A M Q D P E I Q N I L T D P I M R Q V L T D L Q E N P R A A Q A H L K N P G V M Q K I Q K L V S A
4. Setaria italica STI	L K E R Q N K A M Q D P E I Q N I L T D P I M R Q V L T D L Q E N P R A A Q A H L K N P G V M Q K I Q K L V S A
5. Zea mays STI	L Q E R Q N K A M Q D P E I Q N I L T D P I M R Q V L I D F Q E N P S A A Q E H L K N P G V M Q K I Q K L V S A
6. Morus notabilis STI	L K E R Q A K G M Q D P E I Q N I L S D P V M R Q V L I D F Q E N P K A A Q E H T K N P M V M T K I Q K L V S A
7. Glycine max STI	L K E R Q A K A M Q D P E I Q N I L Q D P V M R Q V L I D F Q E N P K A A Q E H T K N P M V M N K I Q K L V S A
8. Populus trichocarpa STI	L K E R Q A K A M Q D P E I Q N I L S D P V M R Q V L V D F Q E N P K A A Q E H T K N P M V M S K I Q K L V H A
9. Arabidosis HOP2	L K E R Q A K G M Q D P E I Q N I L T D P V M R Q V L S D L Q E N P A A A Q K H M Q N P M I M N K I Q K L I S S
10. Arabidopsis HOP1	L K E R Q A K G M Q D P E I Q N I L T D P V M R Q V L S D L Q E N P S A A Q K H M Q N P M V M N K I Q K L I S A
11. Arabdopsis HOP3	L K E R Q A K A M Q D P E V Q N I L S D P V M R Q V L V D F Q E N P K A A Q E H M K N P M V M N K I Q K L V S A
12. Oryza sativa STI1	L Q E R Q N K A M Q D P E I Q N I L K D P I M Q Q V L T D F Q E N P K A A Q A H L K N P G V M Q K I Q K L V S A
13. Oryza sativa STI2	I Q E R Q N K A M Q D P E I Q N I L T D P I M R Q V L V D L Q E N P R A S Q E H L K N P G V M Q K I Q K L V S A
14. Triticum aestivum STI-2A	L K E R Q G K A M Q D P E I Q N I L T D P V M R Q V L I D F Q E N P R A A Q E H L K N P G V K Q K I Q K L V S A
15. Triticum aestivum STI-2B	L K E R Q G K A M Q D P E I Q S I L T D P V M R Q V L I D F Q E N P R A A Q D H L K N P G V K Q K I Q K L V S A
16. Triticum aestivum STI-2D	L K E R Q G K A M Q D P E I Q N I L T D P V M R Q V L I D F Q E N P R A A Q D H L K N P G V K Q K I Q K L V S A
17. Triticum aestivum STI-6A	L Q E K Q S K A M Q D P E I Q N I L T D P I M R Q V L M D F Q E N P R A A Q D H L K D P G V A Q K I Q K L I N A
18. Triticum aestivum STI-6B	L Q E K Q S K A M Q D P E I Q N I L T D P I M R Q V L M D F Q E N P R A A Q D H L K D P G V A Q K I Q K L I N A
19. Triticum aestivum STI-6D	L Q E K Q S K A M Q D P E I Q N I L T D P I M R Q V L M D F Q E N P R A A Q D H L K D P G V A Q K I Q K L I N A
20. Triticum urartu STI	L Q E K Q S K A M Q D P E I Q N I L T D P I M R Q V L M D F Q E N P R A A Q D H L K D P G V A Q K I Q K L I N A
21. Triticum turgidum STI	L K E R Q G K A M Q D P E I Q N I L T D P V M R Q V L I D F Q E N P R A A Q E H L K N P G V K Q K I Q K L V S A
22. Hordeum vulgare STI	L K E R Q G K A M Q D P E I Q N I L T D P V M R Q V L I D F Q E N P K A A Q E H L K N P G V K Q K I Q K L V S A

FIGURE 2 | Phylogenetic analysis and sequence alignment of TaSTI members and STI proteins from other plant species. **(A)** Phylogenetic tree based on amino acid sequences depicting the relationship of TaSTIs with other plant STI proteins. **(B)** Amino acid sequence alignment of STI domain of TaSTI members and other STI proteins.

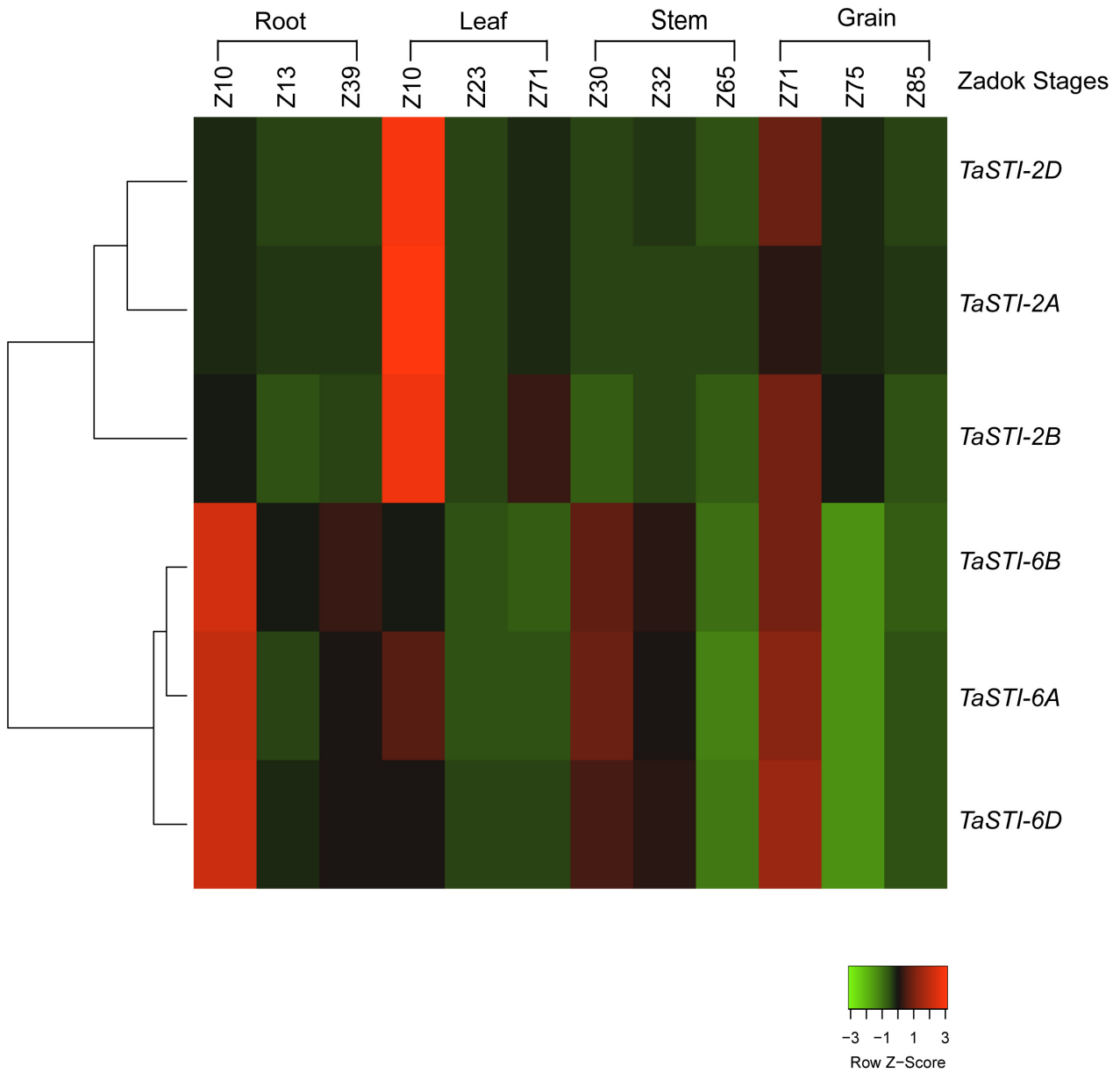


FIGURE 3 | Expression analysis of *TaSTI* family members in different developmental stages of wheat. Heat map showing the relative expression profile of the *TaSTI* family members across the various zadok stages of wheat.

dilutions on –U plates and kept inverted at 30°C (permissive temperature for growth of the yeast *sti* mutant) and 37°C (restrictive temperature for growth of yeast *sti* mutant) for 3 days for the growth of the yeast colonies.

Subcellular Localization and BiFC Assay

For localization studies, the presence of NLS was predicted using the cNLS mapper online tool (Kosugi et al., 2009). Since it has also been reported that STI localizes to ER, the presence of an even ER signal peptide was therefore also checked in the protein sequences (Chen et al., 2010). Both NLS and ER signal peptides were found in all the STI members (**Supplementary Figure S3**). To further confirm the localization pattern of these members, the complete ORF sequence along with the signal peptides were cloned for subcellular localization in onion peel. The genes were first cloned in pENTR-topo vector and then mobilized into the destination vector, i.e., pSITE-3CA, pSITE-nEYFP, and pSITE-cEYFP, under a CaMV35S promoter. The ORF of all the *TaSTI* members were fused in frame with the C terminal of YFP. Onion epidermal cells were used for bombardment by using the PDS-1000 bombardment system (Bio-Rad, Canada) at a pressure of 1100 psi with gold particles coated with plasmid constructs (Lee et al., 2008). Transformed onion peels were kept for incubation at 27°C for 16 h in

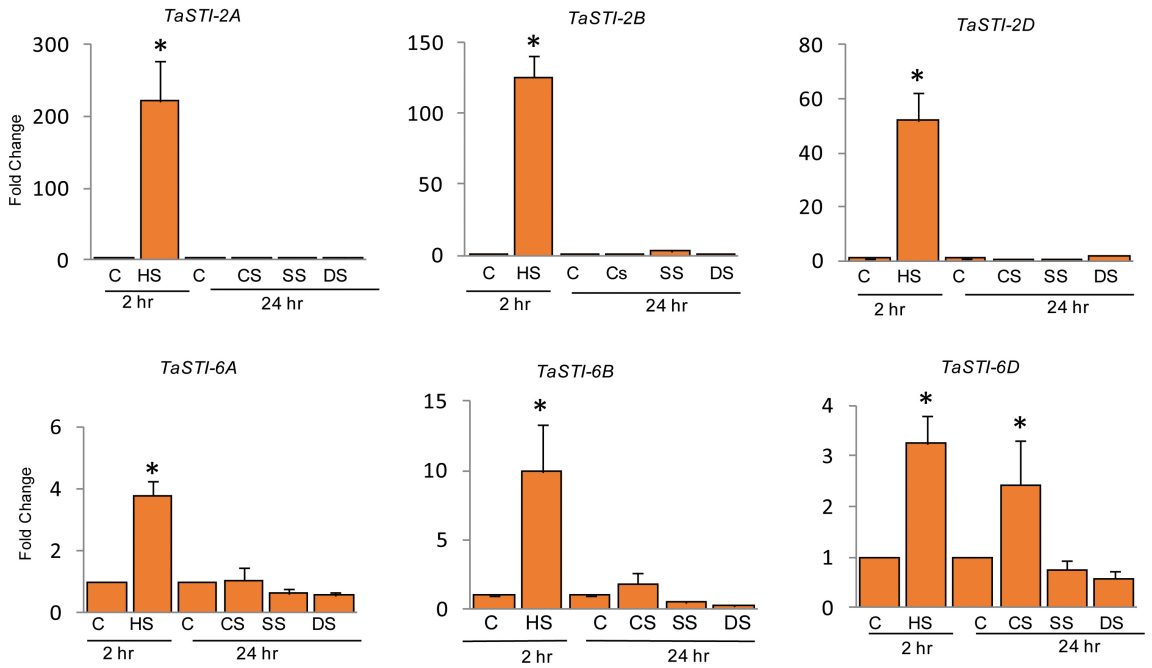


FIGURE 4 | Expression analysis of *TaSTI* family members by real time qPCR. Ten-day-old seedlings of PBW343 were subjected to different abiotic stresses [C-control, HS-Heat stress (42°C for 2 h), CS-cold stress (4°C for 24 h), SS-Salt stress (200 mM NaCl for 24 h), DS-Drought stress (200 mM Mannitol for 24 h)]. Relative fold change was calculated and *TaActin* was used as a housekeeping gene. Graphs were plotted using three biological and three technical replicates. Error bars indicate values ± SD. Asterisks on top of the error bars represent the significance levels (Students *t*-test; $p \leq 0.05$).

dark condition, and fluorescence was observed in a confocal microscope (Leica, Germany).

TaSTI-2A Cloning and Overexpression in *A. thaliana*

For the generation of *Arabidopsis* overexpression plants, the full CDS of 1.7 kb was amplified with gene specific primers using cDNA as a template isolated from a 10-day-old seedling of *T. aestivum* (PBW343 wheat cultivar). The amplified product was then cloned in an entry vector (pENTR^{TM}/D-TOPO) and then in a destination vector pMDC32 under a CaMV35S promoter following the Gateway^{TM} cloning strategy (Directional TOPO cloning kit and LR clonase Enzyme mix II kit, Invitrogen Inc., United States). The GV3101 strain of *Agrobacterium tumefaciens* harboring pMDC32-*TaSTI-2A* was used for transformation in *A. thaliana* through the floral dip method (Clough and Bent, 1998). The T_1 seeds were selected on MS-agar plates supplemented with 50 μg/μl of hygromycin, and the resistant plants were transferred to pots. Further, these overexpressing transgenics were confirmed by PCR using hygromycin and gene-specific primers. Selected plants were further grown up to the T_3 homozygous stage. The plants were confirmed by PCR (**Supplementary Figure S4**).

TaSTI-2A Overexpression in *O. sativa*

Rice (*O. sativa* indica) seeds (variety PB1) were obtained from IARI. Seeds were surface-sterilized by using 0.1% $HgCl_2$ (v/v) for 15 min and washed repeatedly with autoclaved sterile water, and then imbibed in water at 28°C for 16 h. Rice transgenics were generated using the protocol described by Toki et al. (2006). Seeds of the Indica rice variety PB1 were grown under light on an NB medium (Himedia labs, cat no. PT107) at 32°C. Co-cultivation was performed with 7-day-old calli with the EHA105 strain of *A. tumefaciens*. These calli were washed after 3 days of co-cultivation and kept on a selection medium containing hygromycin. The positive calli were then transferred to a regeneration medium until plantlets were formed. These plantlets were transferred to rooting medium for 10 days and then on the rice growth medium.

Histochemical ROS Detection

In order to check the amount of reactive oxygen species (ROS) produced in response to HS in transgenic *Arabidopsis* and WT plants, staining with nitro blue tetrazolium (NBT) was done (Agarwal and Khurana, 2018). For this, 2-week-old seedlings of *Arabidopsis* WT and overexpression transgenics were subjected to HS (42°C for 2 h) after which overnight staining of the plants was done by incubating them in NBT (2 mM NBT powder, 20 mM phosphate buffer). The seedlings were washed with water on the next day and subjected to removal of chlorophyll by dipping them in bleaching solution (ethanol, acetic acid, and glycerol in a ratio of 3:1:1). The plants were then visualized under a bright field light microscope (Leica), and pictures were taken for the comparison of ROS in transgenics and the WT *Arabidopsis* plants after heat stress treatment.

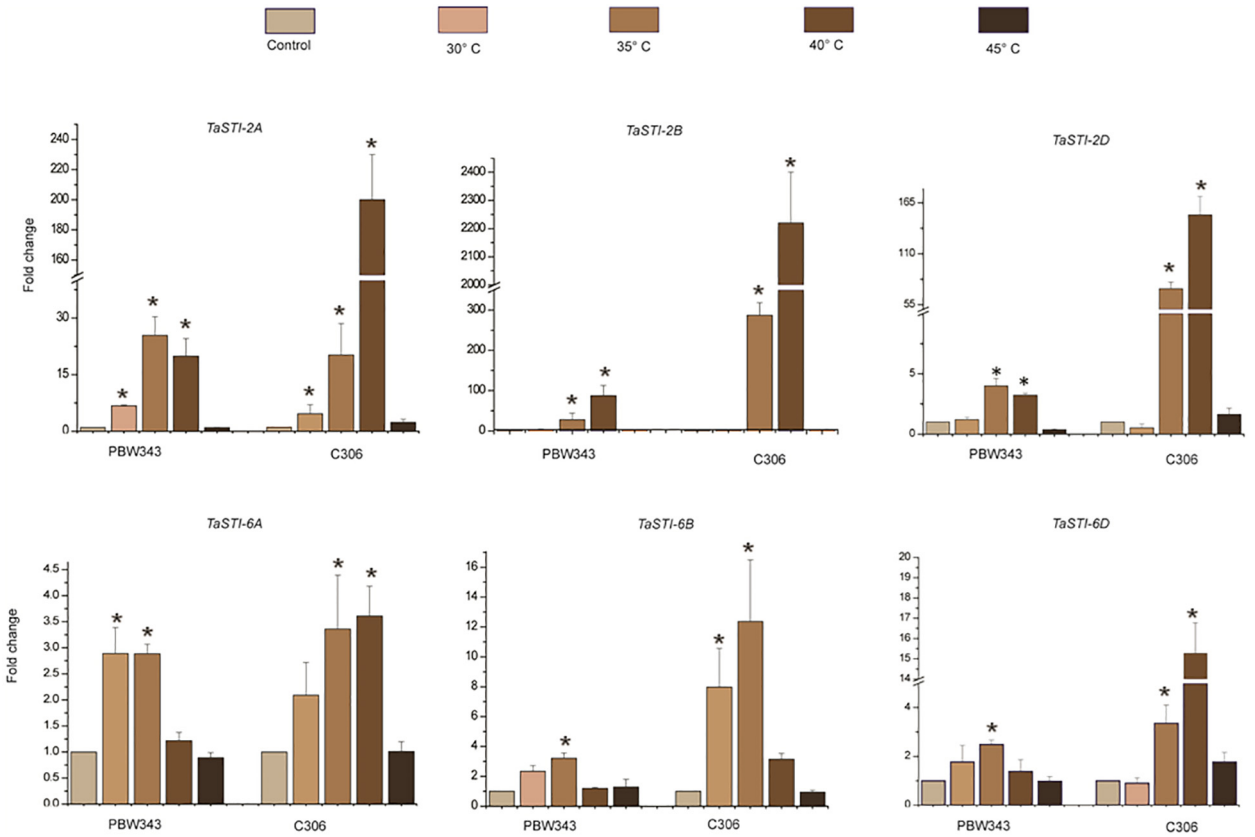


FIGURE 5 | Expression analysis of *TaSTI* family members in PBW343 and C306. Expression was checked in shoot tissue using real-time qPCR. Ten-day-old seedlings were subjected to HS at (30°C, 35°C, 40°C, 45°C) for 2 h. Relative fold change was calculated, and *TaActin* was used as a housekeeping gene. Graphs were plotted using three biological and three technical replicates. Error bars indicate values ± SD. Asterisks on top of the error bars represent the significance levels (Students *t*-test; $p \leq 0.05$).

For rice, 1-month-old plants were taken, and a similar protocol was followed for the comparison of ROS in rice WT and overexpression transgenic lines. NBT staining was done after giving them heat stress.

Physiological Analysis of Heat Stress in *A. thaliana* and *O. sativa*

Photosynthetic Efficiency (Fv/Fm)

Measurements of modulated chlorophyll fluorescence emission from the upper surface of the leaf were made using a pulse amplitude modulation fluorometer (Junior-PAM chlorophyll fluorometer, H. Walz, Germany). Leaves of plants were dark-adapted for 20 min before measuring the induction of fluorescence. Measurements of the PSII function of maximum photosynthetic efficiency (Fv/Fm) was recorded in rosette leaves after stress treatments in at least 10 plants per line viz. WT and transgenics. The same protocol was followed with the rice transgenics.

Estimation of Chlorophyll Content

Two-week-old wild-type and transgenic plants were subjected to HS. For chlorophyll estimation, 100 mg of leaf tissue was taken in a tube containing 2.5 ml of DMSO. Tubes were incubated overnight for chlorophyll bleaching. Absorbance was taken at 645 nm and 663 nm in a UV–Vis spectrophotometer (Hitachi U-2810, Tokyo, Japan), and chlorophyll content was estimated accordingly (Arnon, 1949).

Membrane Stability Index

For Membrane Stability Index (MSI) analysis, 2-week-old stressed and non-stressed seedlings were used. MSI was determined by measuring electrical conductivity with an EC-meter (Eutech, Singapore); 100 mg of leaf tissue was dipped in 10 ml of double distilled water. The tubes were kept at 30°C for 30 min, and conductivity was measured (C1). The seedlings were then autoclaved for 15 min, and electrical conductivity was measured again (C2) in the supernatant. Cellular injury was determined accordingly (Agarwal and Khurana, 2018).

Yield Analysis

Arabidopsis plants were grown on half-strength MS medium in petri plates at 20°C with 16-h photoperiod in daily cycle. Two-week-old plants were subjected to HS at 42°C for 4 h and kept back to recovery. To check for acquired thermotolerance, 2-week-old plants were subjected to continuous heat stress at 30°C for

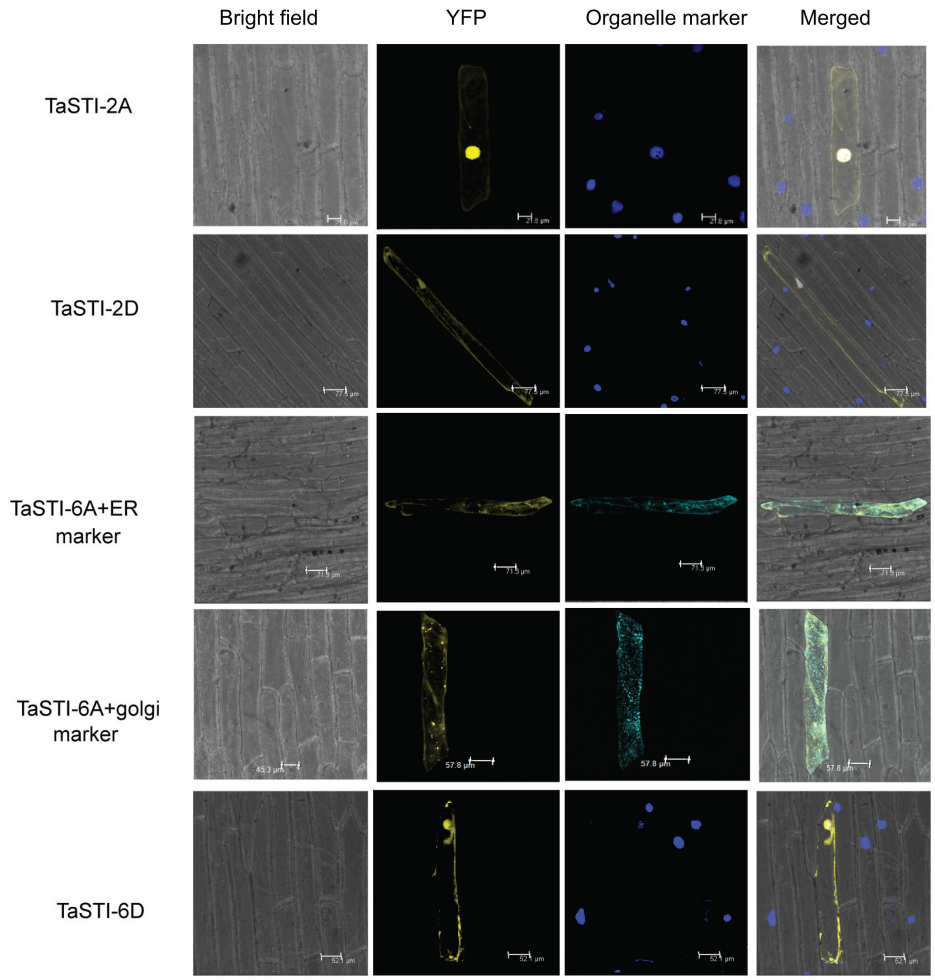


FIGURE 6 | Subcellular localization of selected *TaSTI* genes. CDS of different *TaSTI* genes was cloned in frame with YFP protein and observed under confocal microscope. Different organelle markers were used to locate the gene members in onion epidermis cells. DAPI stain was used to highlight the nucleus along with YFP. ER (Endoplasmic reticulum marker); YFP (yellow fluorescent protein).

its growth and development. Different yield parameters were checked by observing the seed weight, silique number, and silique length of three lines, respectively.

RESULTS

Identification of *TaSTI* Gene Members From *T. aestivum*

To identify the putative STI members in wheat genome, we searched the database with the known STI proteins as query. In total, we obtained six STI genes in wheat, which were mapped to group 2 and group 6 chromosomes. Therefore, the genes were named according to their chromosomal locations (**Table 1**). These members were searched for the presence of conserved domains i.e., STI domain and the TPR domain using the CDD search and SMART tools. Domain analysis revealed that all the members had the same number of STI and TPR domains (**Figure 1C**). The gene structure analysis showed that group 2 *TaSTI* members had a similar organization of introns and exons. On the other hand, all the group 6 *TaSTI* members were found to have longer introns as compared to *TaSTI-2* family members (**Figure 1B**). *TaSTI-6D* was found to be the longest coding member (2498 bp), whereas *TaSTI-2D* was found to have the shortest coding sequence (2270 bp). These six *STI* members with complete coding sequences were used for further analysis.

Phylogenetic Analysis and Multiple Sequence Alignment of TaSTI Members

To analyze the evolution of the TaSTIs, we constructed a Neighbor-Joining (NJ) tree based on a total of 22 STI members from different plants (**Figure 2A**). All the TaSTI members that were present on group 2 chromosomes, were found to belong to a single clade wherein TaSTI-2D was found to be closely related to STI of *Aegilops tauschii*. TaSTI-2A was observed to be close to *Triticum turgidum* STI. However, the TaSTI members that were present on group 6 chromosomes represented a

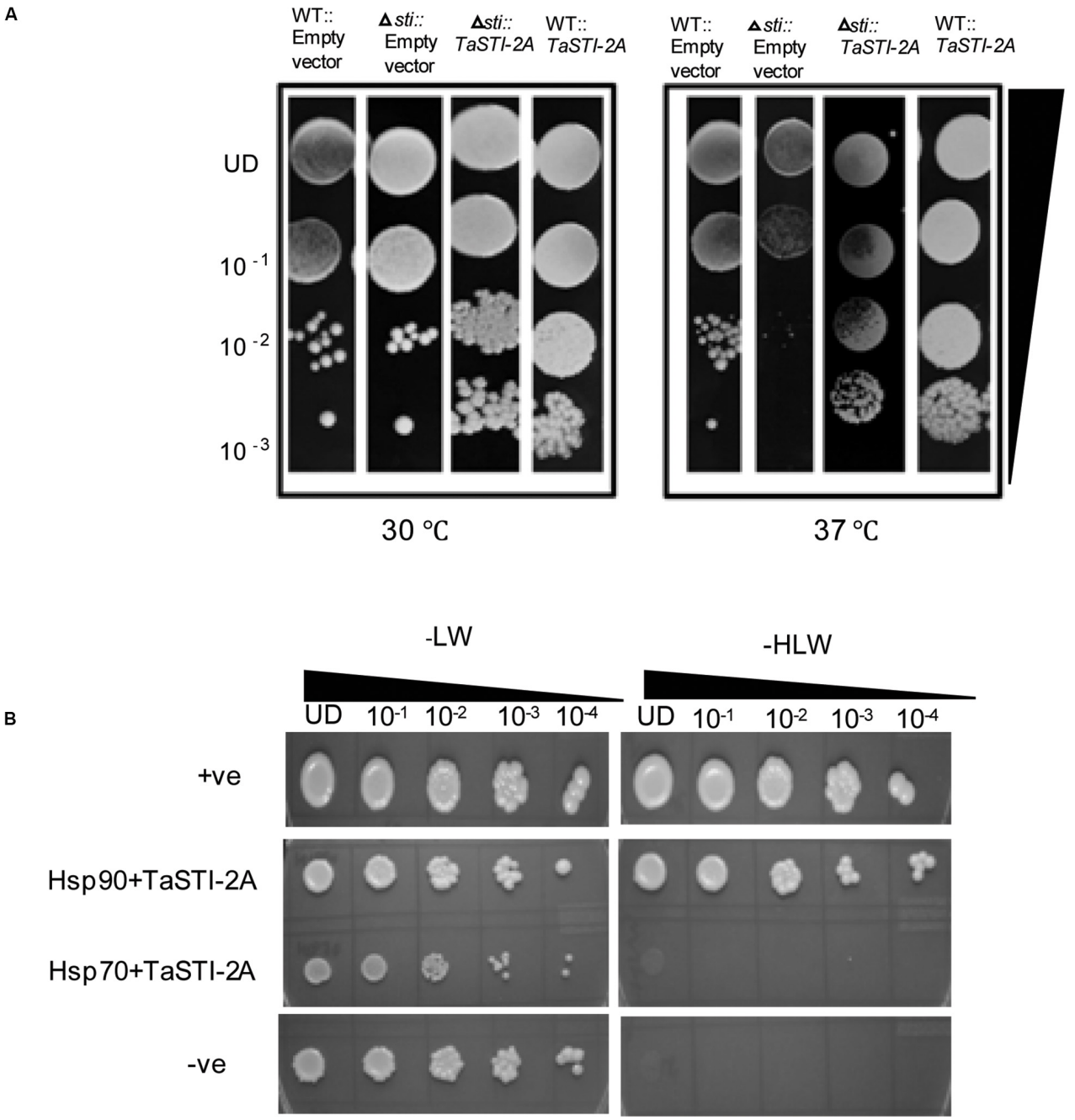


FIGURE 7 | Yeast complementation and yeast two hybrid. **(A)** For the yeast complementation assay, WT and the yeast mutant of the *STI* gene were transformed with empty vector (pYES2) or the fusion construct (pYES2:*TaSTI-2A*). Their growth phenotype was checked at permissive temperature (30°C) or conditional temperature (37°C). **(B)** Yeast-2-hybrid assay was performed to check the interaction of TaSTI-2A with TaHSP90 and TaHSP70 proteins. The growth of yeast cells co-transformed with fusion constructs pDEST-GBKT7:*TaSTI-2A*, pDEST-GADT7:*TaHSP90*, and pDEST-GBKT7:*TaSTI-2A;* pDEST-GADT7:*TaHSP70* was analyzed on SD/-Leu/-Trp (-LW) medium and on SD/-Leu/-Trp/-His (-HLW) medium. The growth of co-transformed cells was monitored by a drop assay on media lacking histidine, leucine, and tryptophan (-HLW) along with 0.5 mM 3-aminotriazole (3AT).

different clade along with *Triticum urartu*. Thus, this suggested the divergent evolution of group 2 and group 6 members in wheat, and probably, these two loci of *STI* genes have existed from the beginning in the *Triticeae* lineage. Further, the multiple sequence alignment of the core STI domain revealed its conserved nature in most of the plants (**Figure 2B**). Moreover, the three-dimensional structure of TaSTI-2A was predicted and compared with the STI/HOP members of the *Arabidopsis* (**Supplementary Figure S2**). Interestingly, it exhibited similarity with AtHOP3, which is the one of the HOP members known to

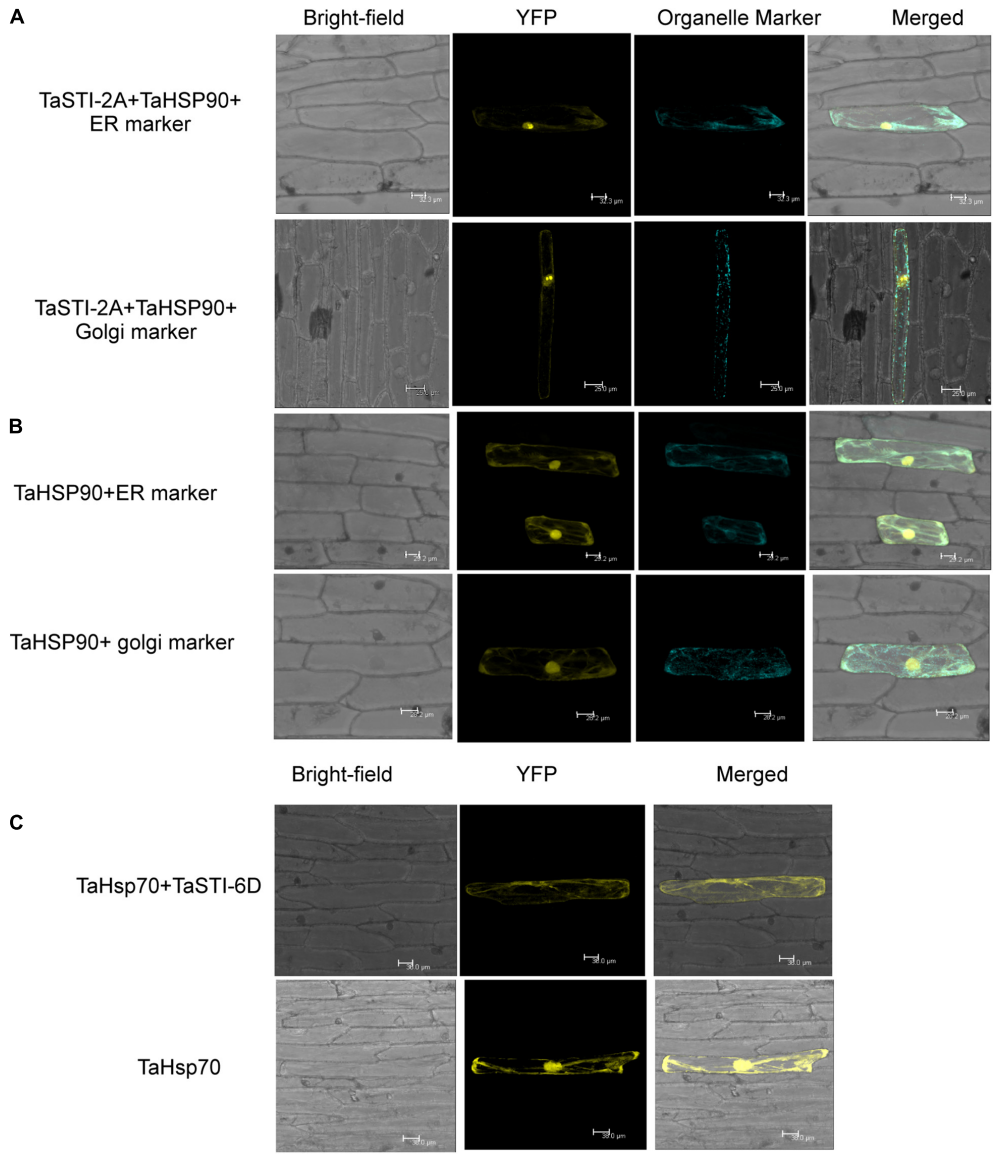


FIGURE 8 | Interaction of TaSTI-2A and TaSTI-6D with TaHSP90 and TaHSP70, respectively. **(A)** BiFC assay was performed to study the interaction between TaSTI-2A and TaHSP90. Transient expression was checked in onion epidermal cells. The localization of the BiFC signal was compared with the ER and Golgi marker localization. Positive interaction could be detected in ER and Golgi bodies along with the nucleus. **(B)** Subcellular localization of the selected TaHSP90 protein. CDS of *TaHSP90* gene was cloned in frame with YFP and observed under confocal microscope. Organelle markers were used to locate the protein in onion epidermis cells. ER (endoplasmic reticulum marker); YFP (yellow fluorescent protein). **(C)** TaSTI-6D and TaHSP70 interaction was observed in cytoplasm using the BiFC assay. For subcellular localization of TaHSP70, CDS of *TaHSP70* gene was cloned in frame with YFP, and transient expression was checked in onion epidermal cells under a confocal microscope.

function in HS in *Arabidopsis* (Fernandez-Bautista et al., 2018). This in turn suggested that TaSTI-2A might have a role in high temperature stress response and they may even bind to the same client proteins, which are known to bind AtHOP3.

Promoter Analysis and Expression Profile of *TaSTI* Family Members

The presence of specific *cis*-elements in the promoter region plays an important role in the regulation of gene expression and thus helps in responding to environmental conditions (Walther et al., 2007). Various *cis*-elements were identified in the 2-kb upstream region of all the *TaSTI* gene members by using the PlantCare and PLACE databases. However, in order to assess their role specifically in thermotolerance, the presence of HSE, STRE (stress responsible elements), and CCAATBOX1 elements was analyzed in the promoter regions. HSE forms an essential component of HSR as HSFs bind to these elements to enhance the expression of HSPs and other heat-responsive genes. Similarly, CCAATBOX1 has been known to contribute

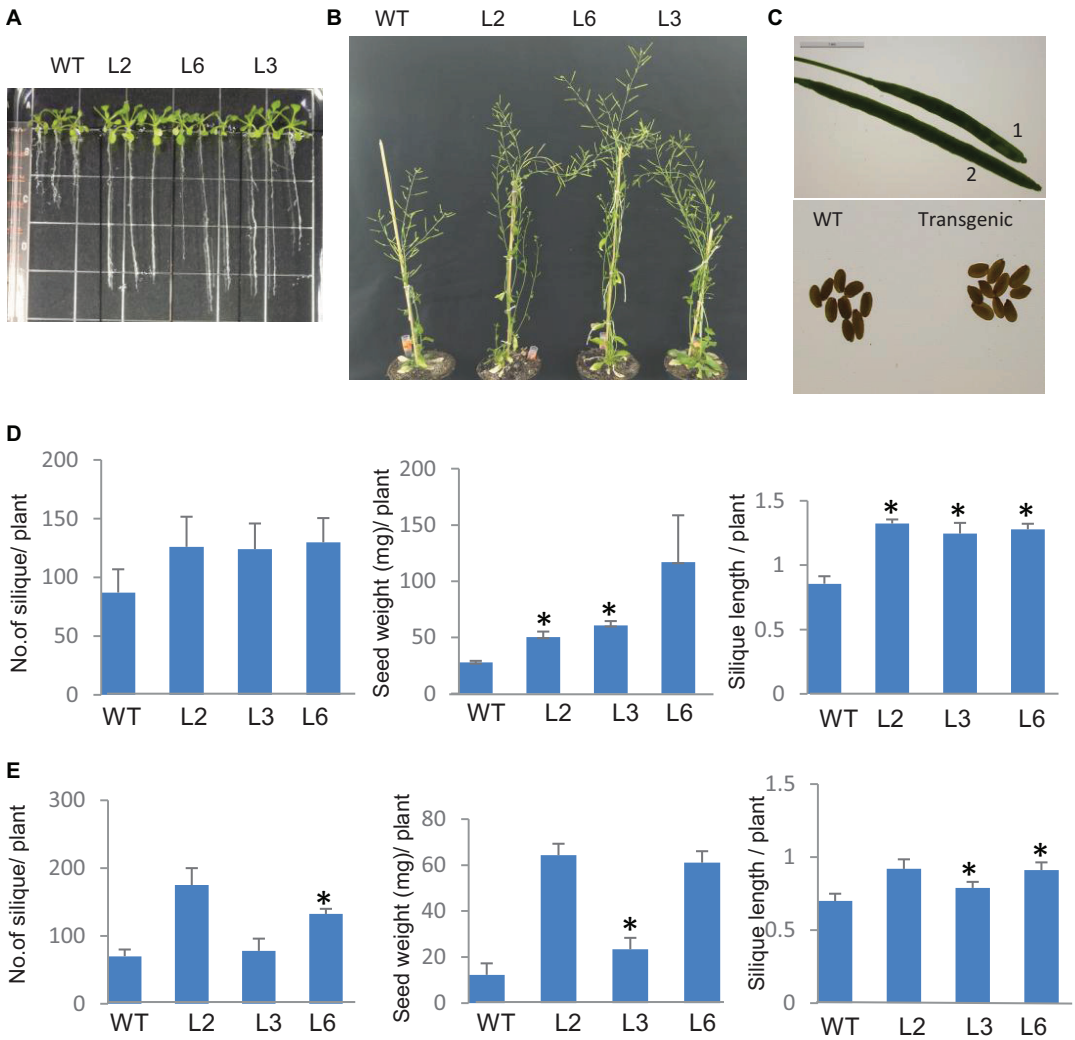


FIGURE 9 | Phenotype of *TaSTI* overexpression transgenic plants of *Arabidopsis thaliana*. Two-week-old plants were analyzed under HS. **(A)** Transgenics showed increased root length in comparison to WT. **(B)** Overexpression line showed better plant growth, **(C)** seed, and silique size (1:WT, 2:OE). **(D)** Graphical representation of silique number, silique length, and seed weight in transgenics and WT at 42°C for 4 h and **(E)** mild heat stress i.e., plants were grown at continuous 30°C. Asterisks on top of the error bars represent the significance levels (Students *t*-test; $p \leq 0.05$).

with HSE elements, and the STRE element has been reported to be involved in various stresses (Khurana et al., 2013). Moreover, it has been documented that the deletion of these motifs in the promoter of *TaHSP26* gene results in the decrease of promoter activity (Khurana et al., 2013). In the case of *TaSTI* gene family members, it was observed that group 2 *TaSTI* members found to possess more HSE elements in their URR as compared to group 6 *TaSTI* members (**Supplementary Figure S1**). Among the group 2 members, *TaSTI-2B* had two HSEs, whereas *TaSTI-2A* and *TaSTI-2D* were found to have only one HSE in their promoter region. Also, the CCAATBOX1 elements were higher in group 2 members (**Supplementary Figure S1**). Thus, more heat responsiveness of *TaSTI* group 2 members can be speculated from these data.

To investigate the functional role of *TaSTI* members, it was important to analyze their expression profile in wheat. For this purpose, the wheat expression database, which comprises RNA-seq data, was used (Borrill et al., 2016). Also, this database offers a method to employ all the published resources in a more meaningful and customizable manner. Thus, we used this database to study the expression pattern and clustering of the identified *STI* members. TPM values of different developmental stages, namely, root, leaf, stem, and grain were analyzed, and it is shown in the form of clustered heat maps (**Figure 3**). A tissue preferential expression pattern was observed, wherein group 2 *TaSTI* members had higher expression in leaves whereas group 6 *TaSTI* members had higher expression in roots. This indicated that these *STI* members apart from providing protection against HS might have a developmental role in plants.

To elucidate the role of these family members during different abiotic stress responses, expression analysis was done by quantitative real-time PCR (qPCR) under four different abiotic

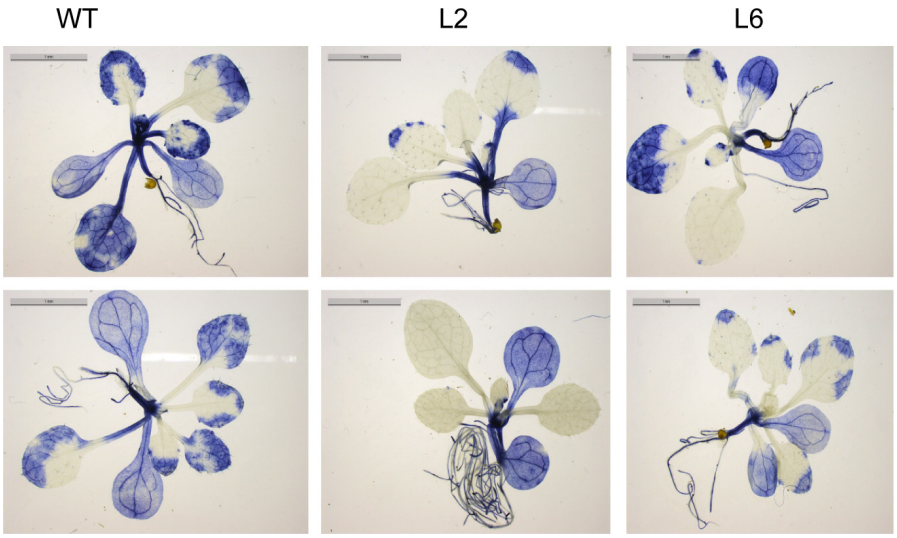


FIGURE 10 | NBT staining. Two-week-old seedling of WT and transgenics were given heat stress at 42°C for 2 h and were then stained with NBT to analyze the superoxide anion accumulation after the stress treatment.

stresses such as heat, cold, salt, and drought conditions. Ten-day-old PBW343 seedlings were subjected to these abiotic stresses. As shown in **Figure 4**, these members were highly upregulated by HS and did not show any significant change in the expression under other stresses. Moreover, *STI* members of group 2 chromosome showed higher expression as compared to the members of the group 6 chromosome, which was in accordance to their promoter analysis. An increase of up to 200 folds was observed in the expression of *TaSTI*-2A under HS conditions.

A comparative profiling of these members was carried out in thermosensitive and thermotolerant wheat cultivars such as PBW343 and C306, respectively (Hairat and Khurana, 2015), post exposure to increasing temperatures (**Figure 5**). Higher expression of these members was observed in thermotolerant wheat cultivar C306 as compared to the thermosensitive cultivar PBW343. This indicated a differential regulation of these members in a varietal specific manner and their differential sensitivity to HS. Moreover, in case of group 2 *TaSTI* members, a concomitant increase in expression was observed with increasing temperatures, but their level decreased at 45°C. However, no specific pattern of expression could be observed in the case of group 6 *TaSTI* members. Since *TaSTI*-2A was one of the members, which showed highest induction under HS (**Figure 4**) and has been earlier identified in our HS library (Chauhan et al., 2011), we selected this gene for further in-depth validation and molecular characterization.

Subcellular Localization of *TaSTI* Gene Members

Subcellular localization of proteins is an important factor in providing the physiological context of their function. Therefore, in order to determine the subcellular location of *TaSTI* gene members, the pSITE-3CA:*TaSTI* constructs were used to bombard epidermal peels of onion (**Figure 6**). Two members each from chromosome 2 (TaSTI-2A, TaSTI-2D) and chromosome 6 (TaSTI-6A, TaSTI-6D) were taken for the localization study. YFP signals for TaSTI-2A, TaSTI-2D, and TaSTI-6D were observed in the nucleus and in the cytoplasm. Interestingly, TaSTI-6A was found to accumulate in the cytoplasmic structures resembling ER. To corroborate this result, co-localization was performed with the ER organelle-specific marker. As observed in **Figure 6**, YFP signals were detected in ER structures that co-localized with CFP-ER. This was in accordance with the result reported by Fernandez-Bautista et al. (2017), wherein AtHOP3 was also found to partially localize to ER.

Ta*STI-2A* Functionally Complements the Yeast *sti* Mutant

To test the functionality of *TaSTI*-2A, a full-length gene was subcloned into the pYES2 vector, and the construct was transformed into the yeast *sti* mutant (Thermo Scientific). The *STI* mutation in yeast causes mild growth defect at 37°C, which is the restrictive temperature for the growth of yeast (Fernandez-Bautista et al., 2018). The colonies obtained after transformation were thus dotted on SD/-Ura plates and incubated at 30 and 37°C. Vector-transformed WT yeast (AH109, Clontech) was found to grow at both 30 and 37°C, although the colony diameter at 37°C was less than at 30°C (**Figure 7A**). The vector-transformed mutant, however, showed less growth at 37°C as compared to that at 30°C (**Figure 7A**). The complementation of the *TaSTI-2A* gene in the mutant improved its growth at 37°C (**Figure 7A**). Moreover, the overexpression of the gene in the WT resulted in more vigorous growth of even the WT yeast in the restrictive growth condition (**Figure 7A**). The *TaSTI*-2A gene could thus complement the absence of the yeast *STI* gene in the heterologous system and confer HS tolerance to the mutant. It even provided a growth advantage to WT yeast under the HS condition.

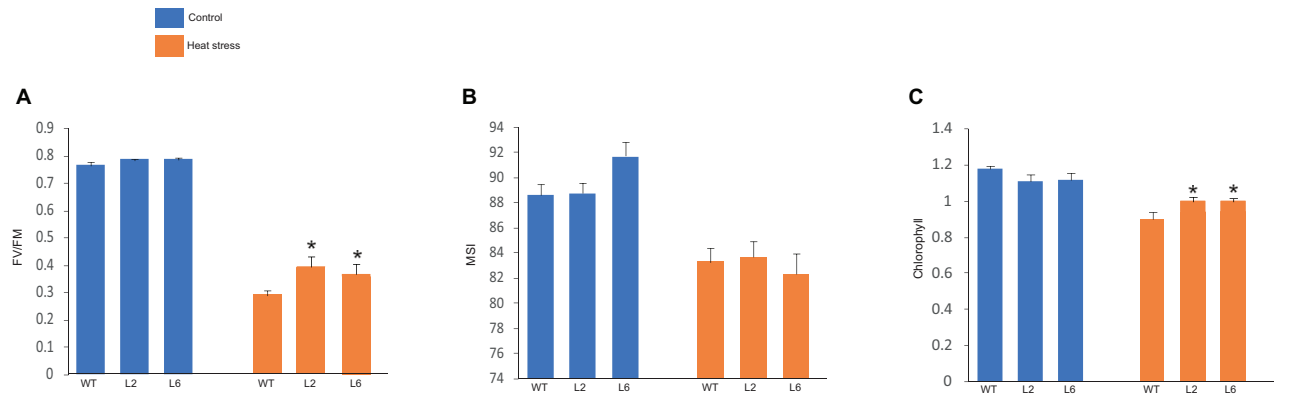


FIGURE 11 | Graphical representation of physiological assays. **(A)** Photosynthetic efficiency, **(B)** cellular membrane stability, **(C)** chlorophyll levels of wild-type (WT) and transgenic lines under control and high temperature stress conditions. Error bars indicate values ± SD. Asterisks on top of the error bars represent the significance levels (Students *t*-test; $p \leq 0.05$).

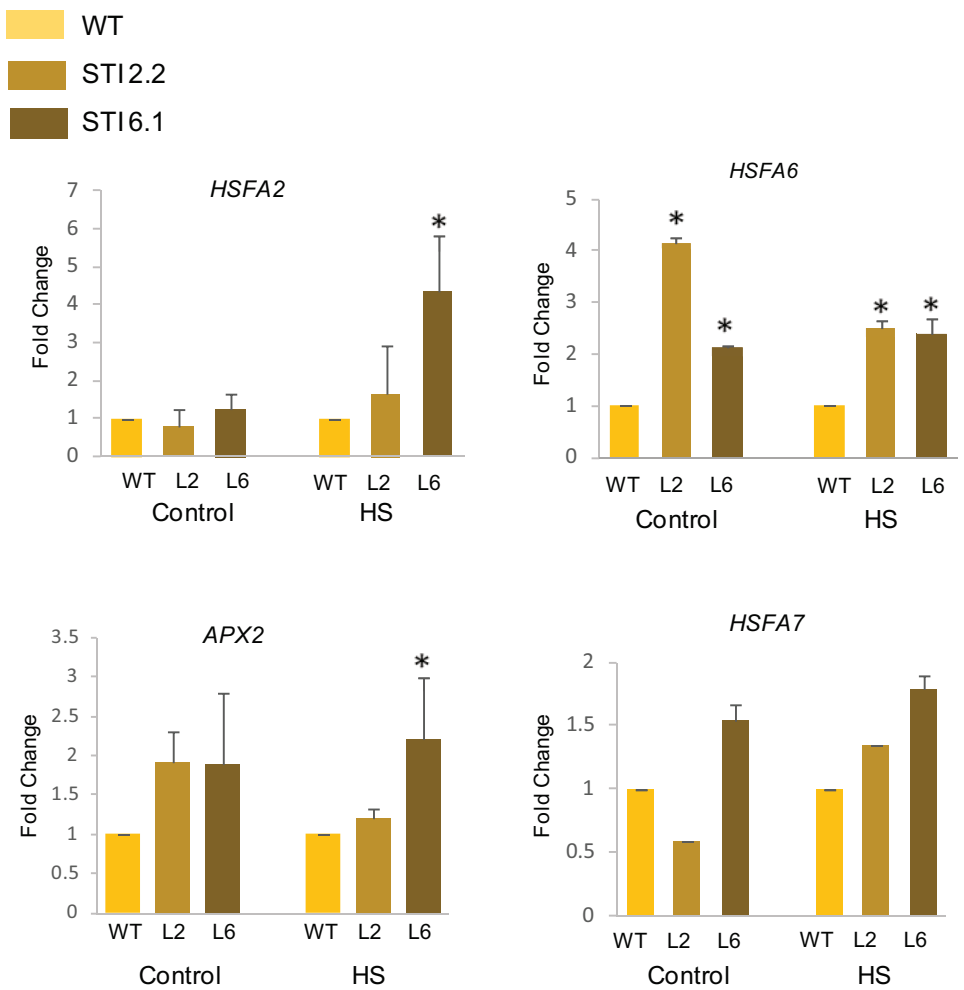


FIGURE 12 | Expression analysis of heat stress marker genes. Transcript analysis of heat stress marker genes in WT and transgenics under control and heat stress. Transcript levels were normalized to WT, and *AtActin* was used as a housekeeping gene. Values represent data from three biological replicates and three technical replicates. Error bars indicate values ± SD. Asterisks on top of the error bars represent the significance levels (Student's *t*-test; *p*-value ≤ 0.05).

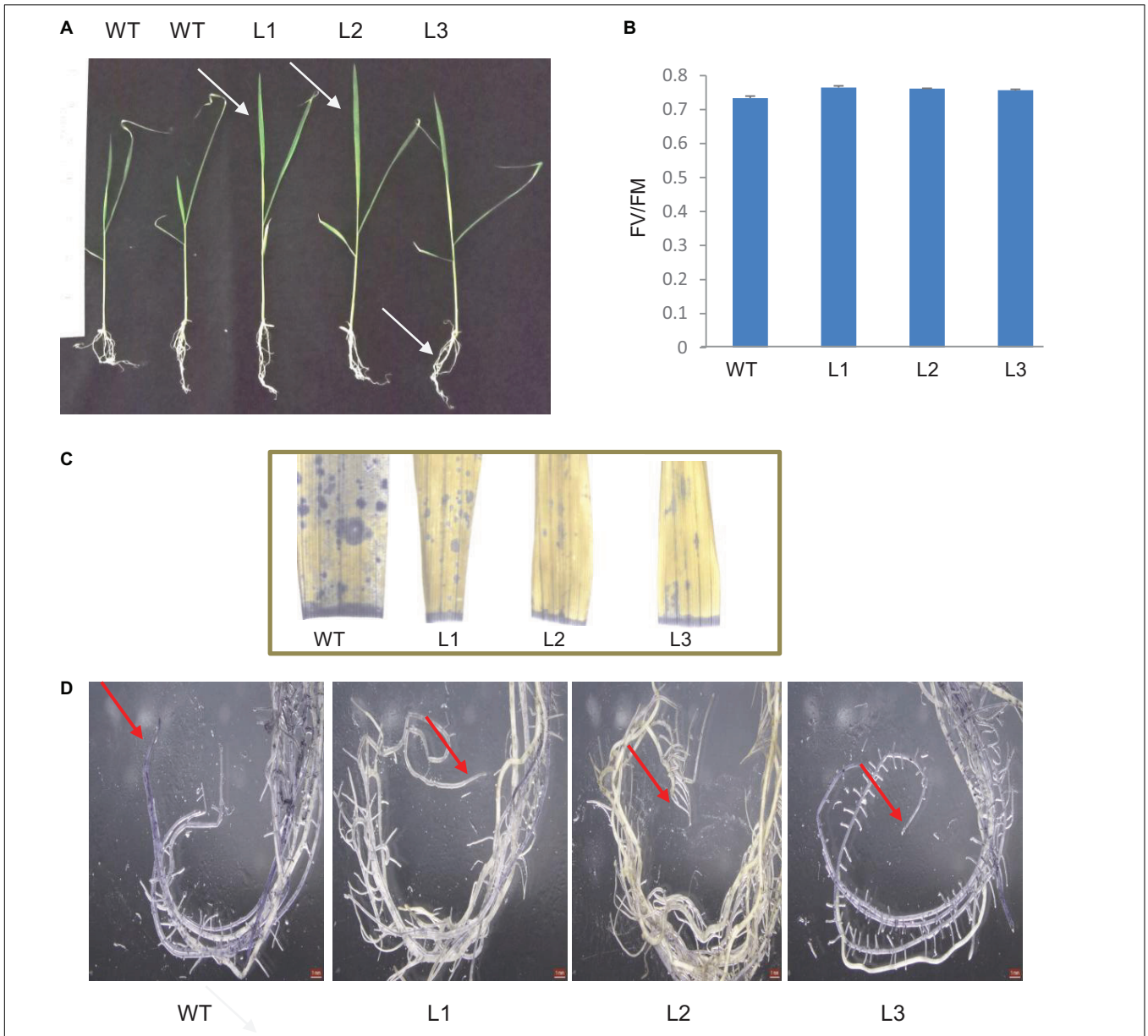


FIGURE 13 | Analysis of rice transgenics under heat stress. **(A)** Five-day-old seedlings were subjected to 42°C for 5 h, and the phenotype was observed after 7 days. **(B)** Photosynthetic efficiency of WT transgenics under HS. **(C)** Localization of superoxide anion in rice leaves and roots by using NBT stain. One-month-old rice plants were subjected to 42°C for 5 h and were then stained with NBT. Blue patches represent the accumulation of ROS in leaves and **(D)** root tissue.

In vivo Interaction of *TaSTI* Members With *TaHSP90* and *TaHSP70*

Earlier studies have reported that STI is a co-chaperone of HSP90 and HSP70; therefore, it forms a protein complex in order to assist these chaperones to carry out the folding of client proteins (Chen and Smith, 1998; Scheufler et al., 2000). Therefore, we examined the interaction of TaSTI-2A with TaHSP90 and TaHSP70 in yeast by using the yeast two-hybrid system. The interaction was found between TaSTI-2A and TaHSP90. All colonies, which had appeared on the −LW media, showed vigorous growth even at 10^{-3} dilution on −HLW media at 0.5 mM 3AT, indicating the interaction to be specific and strong. Interestingly, as shown in **Figure 7B**, TaSTI-2A did not interact with TaHSP70 in the yeast system. This interaction was further validated by BiFC analysis (**Figure 8A**). Positive BiFC signals were observed in the nuclei as well as in the cytoplasmic structures of the onion epidermal peels. Further, co-localization of the BiFC signals with the ER and Golgi marker also confirmed that the interaction was occurring in the ER and Golgi complex, apart from the nucleus. This could be further explained by the fact that TaHSP90 localized ubiquitously in the cell (**Figure 8B**). Since the TaHSP90 protein was found in the ER, Golgi, and nucleus,

it is therefore probable to find the interaction between the two in these organelles.

Interaction of TaSTI-2D, TaSTI-6A, and TaSTI-6D was also checked with TaHSP90 and TaHSP70 by using BiFC analysis. Surprisingly, interaction between TaSTI-6D and TaHSP70, occurring specifically in the cytosol, was observed. This unique interaction could be again explained by the ubiquitous distribution of TaHSP70 throughout the cell (**Figure 8C**).

Overexpression of *TaSTI*-2A Enhanced Heat Tolerance in *A. thaliana*

To elucidate the functional role of *TaSTI-2A* in plants, overexpression transgenic lines of *Arabidopsis* were generated, and the overexpression in these lines was confirmed using hygromycin-specific PCR and real-time PCR (**Supplementary Figure S4**). Since this gene was highly upregulated by heat, we therefore observed the growth of the *Arabidopsis* overexpression lines under both basal HS and acquired HS conditions. For basal HS, 2-week-old transgenic plants were subjected to 42°C for 4 h and then returned to their original growth conditions for recovery. After a 15-day recovery period, transgenic plants revived earlier and showed faster and more robust growth as compared to WT (**Figure 9B**). Also, the transgenic lines showed increased plant height, silique size, and silique number per plant, with respect to WT (**Figures 9B,C**). Yield parameters like silique number, seed weight, and silique length were found to be better in transgenics as compared to WT (**Figure 9D**). For acquired thermotolerance, 2-week-old overexpression lines were grown continuously at 30°C. In this case also, transgenic plants performed better than the WT as displayed by their higher number of siliques per plant, silique length, and silique weight (**Figure 9E**).

As the roots of the plants are equally sensitive to HS as shoots (Giri et al., 2017), root elongation assays were therefore also done in order to check the measure of thermotolerance in overexpression plants. Seven-day-old seedlings were subjected to HS conditions, and it was observed that, in comparison to WT plants, the transgenic plants had longer roots (**Figure 9A**).

It is well documented that HS accelerates the ROS accumulation in plants and thus leads to oxidative damage (Suzuki and Mittler, 2005). Therefore, the levels of ROS were checked in both WT and the *TaSTI-2A* overexpression lines after HS conditions. The level of super oxide ions as measured by the NBT staining was found to be lower in the transgenic lines in comparison to the WT (**Figure 10**). Moreover, in all the aspects among the three transgenic lines L2, L6 performed better than L3 which could be justified by the better ectopic expression of *TaSTI-2A* in the L2 and L6 lines (**Supplementary Figure S4**). Therefore, these two lines were taken for the further physiological assays.

The leaf photosynthesis is highly sensitive to high temperature stress (Law and Crafts-Brandner, 1999). Among the whole photosynthetic machinery, photosystem II (PSII) is the most heat susceptible (Havaux, 1992). Therefore, the effect of HS was investigated by measuring the maximum photosynthetic efficiency (Fv/Fm) of PSII. Fv/Fm of transgenics were found to be higher under HS as compared to WT (**Figure 11A**). The chlorophyll content was also analyzed as chlorophyll synthesis is known to be receptive toward stress and serves as a good indicator (Tewari and Tripathy, 1998). Chlorophyll content (i.e., total chlorophyll content) was found to be higher in transgenics than in WT after stress (**Figure 11C**). Similarly, membrane stability, which is a measure of ion leakage from the tissue, was used to analyze the damage caused to the members due to HS (Niu and Xiang, 2018). However, no significant difference was found in membrane stability between the WT and transgenic plants (**Figure 11B**).

Transcription Profiling of Stress Marker Genes

Heat stress transcription factors (HSFs) functions as the major regulators of the HSR as they regulate the expression of various small HSPs and other HSFs (von Koskull-Doring et al., 2007). Interestingly, in case of tomato, it has been shown that HSP70 and HSP90 could regulate the HSR via the interactions with HSFA1, HSFA2, and HSFB1 (Hahn et al., 2011). Thus, in order to assess thermotolerance in *Arabidopsis TaSTI-2A* overexpression lines, we checked the expression of various HSFs in control and HS conditions. It was observed that the relative expression of *AtHSFA2*, *AtHSFA6*, and *AtHSFA7* was found to be higher in overexpression lines than the WT under both control as well as after HS conditions (**Figure 12**). Apart from these, the expression of antioxidant enzymes were checked as ROS detoxification aids in HS adaptation and in possession of thermotolerance (Yu et al., 2019). The expression of *AtAPX2* was found to be upregulated in overexpression lines under HS conditions in comparison to the WT (**Figure 12**). Interestingly, even under the control condition, the levels of *AtAPX2* were more in the overexpression lines as compared to WT.

Phenotypic Analysis of *TaSTI*-2A Transgenic Rice Lines Showed Improved Heat Tolerance

TaSTI-2A transgenic rice lines were generated using *Agrobacterium*-mediated transformation, and they were confirmed by PCR (**Supplementary Figure S5**). Five-day-old transgenic plants that were subjected to 42°C for 5 h followed by 7 days of recovery performed better than the WT. The transgenic plants had better leaf and root growth as compared to the WT (**Figure 13A**). Also, the photosynthetic efficiency was found to be enhanced in the transgenics (**Figure 13B**). Furthermore, transgenics in comparison to WT were found to have lesser oxidative load after HS as evident by the NBT staining (**Figures 13C,D**).

DISCUSSION

To maintain cellular homeostasis, all eukaryotic cells are equipped with mechanisms to prevent the aggregation of misfolded proteins. HSP90 and HSP70 represent two main chaperone proteins, which play an important role in protein folding, especially under HS conditions (Hahn et al., 2011).

However, the binding of different co-chaperone proteins regulates their substrate selection. Although the main chaperones have been deeply characterized in plants (Xu et al., 2012; Usman et al., 2017), our knowledge about the co-chaperone proteins still remains limited. In plants, STI (also known as HOP) serves as one of the co-chaperones for the HSP proteins (Zhang et al., 2003). In the present work, we have identified six members of the *STI* gene family in *T. aestivum*. This number is comparatively higher than the number of *STI* gene family members identified in other plants (Chen et al., 2010; Fernandez-Bautista et al., 2018). The analysis of the STI and TPR domains showed that they share a high similarity with other plant STIs indicating its conserved nature. The members were found to be located to the long arm of group 2 and group 6 chromosomes, and thus, they were named according to their chromosomal locations. Since the redundancy of genes in wheat could be attributed to its hexaploid nature, high similarity was therefore found between these members.

It is noteworthy that the *TaSTI* family members displayed tissue specific transcription pattern, particularly specific to leaves and roots. It is well known that HS not only primarily affects leaf photosynthesis but can also cause damage to root and decrease nutrient uptake (Giri et al., 2017; Wang et al., 2018). Therefore, it is likely that either these members might be involved in protection against HS in these particular tissues or they may be required in developmental pathways in these tissues. In the case of animals, it has already been reported that HOP plays an important role during embryonic development (Baindur-Hudson et al., 2015). Moreover, all the *TaSTI-2* members possessed more heat-responsive elements in their promoter regions and, thus, displayed very high expression under HS. This suggested their indispensable role in QC, which is especially high in HS inside the cells. Also, a differential expression pattern of these *TaSTI-2* members was observed in thermotolerant varieties in comparison to thermosusceptible varieties. These results suggest that the higher expression of these co-chaperone family members might represent one molecular aspect of imparting the tolerant phenotype to the C306 variety.

STI members have been previously studied in *Arabidopsis* and in rice, and these members have been found to localize to ER as well as in the nucleus (Chen et al., 2010; Fernandez-Bautista et al., 2018). In accordance to this, our study also displayed that most of the members are localized to the nucleus and the cytoplasm, but only TaSTI-6A was found to reside in the ER and the Golgi complex. Therefore, we postulate that TaSTI-6A might be one of the members, which might help in maintaining QC specifically in ER.

It is well known that STI interacts with HSP70 and HSP90, but interestingly, our work highlights that TaSTI-2A interacts with TaHSP90 not only in the nucleus but also in the ER and Golgi bodies. Thus, we speculate that TaSTI-2A interacts with TaHSP90 in the nucleus and then translocates to the ER. Fernandez-Bautista et al. (2018) also showed similar results in *Arabidopsis* wherein HOP family members localize to the cytoplasm and translocate to the nucleus after the heat treatment (Fernandez-Bautista et al., 2018). Interestingly, TaHSP70 was found to interact with TaSTI-6D (and not with TaSTI-2A) in the cytosol, which in turn highlights the differential role of *TaSTI* family members in HS in wheat. Previous studies in plants have only reported the interaction between HOP3 and BIP proteins (which are ER resident HSP70 proteins) in the ER (Fernandez-Bautista et al., 2017). Taking all these evidences into consideration, we hypothesize that under HS conditions, TaSTI-2A moves form the nucleus to the ER, and TaSTI-6D moves from the nucleus to the cytosol to help in the protein folding response in their respective places. However, further experiments need to be conducted to elucidate the role of these two *TaSTI* family members in HS. Moreover, in this study, only one member of the TaHSP90 and TaHSP70 family was taken for checking the interactions with the TaSTI family members. Therefore, there exists a possibility that the rest of the members of TaSTI may interact with other members of the TaHSP90 and TaHSP70 family, thereby providing each TaSTI protein a specific role inside the cell.

Besides the well-characterized role of *STI* in biotic stresses, their role in response to abiotic stress tolerance remains to be explored. Only a solitary recent report in *Arabidopsis* highlights the role of the HOP family in long-term acquired thermotolerance rather than serving as an adaptor protein (Fernandez-Bautista et al., 2018). Similarly, in the present study overexpression of *TaSTI*-2A in *Arabidopsis* enhances the basal and acquired thermotolerance of the transgenic plants. The overexpression lines showed decreased root growth inhibition, enhanced plant height, high chlorophyll content, and better photosynthetic activity under HS conditions in comparison to WT. The level of oxidative damage to plants after HS was found to be lower in the transgenics. This in turn corroborated with the higher expression of *APX2* in the transgenics, which is a well-known ROS scavenging enzyme and an oxidative stress marker. The higher expression of *HSFA2*, *HSFA6*, and *HSFA7* in the transgenic lines could be one of the reasons for their better performance.

Similarly, in the case of rice, the overexpression of *TaSTI-2A* promoted better growth of the plants after the exposure to high temperatures as seen in their recovery. The transgenics were observed to have lower ROS levels and better photosynthetic efficiency. Interestingly, both in the case of *Arabidopsis* and rice, particularly better root growth of the transgenics was observed after HS. Therefore, it might be speculated that since STI is a co-chaperone, it may bind to different master regulators and thus protect them under HS condition. Overall, it suggests that *TaSTI-2A* helps in imparting thermal stress tolerance and can be considered as a suitable gene to improve crop plants under extreme environmental stress conditions.

In conclusion, among the *TaSTI* gene family members, *TaSTI-2A, TaSTI-2B*, and *TaSTI-2D* were found to be the heat responsive *STI* members under HS conditions in wheat. Overexpression of *TaSTI*-2A in *Arabidopsis* and rice conferred thermal stress tolerance to the transgenic plants. TaSTI-2A showed *in vivo* interaction with TaHSP90 in the nucleus as well as in the ER and the Golgi complex. In the future, it will be of interest to explore how these two proteins interact in the ER-Golgi complex

and the functional implications of this interaction. Also, the role of *TaSTI*-6D in HS and the molecular mechanism behind its interaction with TaHSP70 represents an area of future research.

DATA AVAILABILITY STATEMENT

STI protein sequences of *Arabidopsis thaliana* were taken from Tair database (AT1G12270, AT1G62740, and AT4G12400). One of the TaSTI members was previously identified in the lab from the HS cDNA library (GD189073) which was submitted to the NCBI database (Chauhan et al., 2011).

AUTHOR CONTRIBUTIONS

SM, SD, and HS contributed to experimental validation and PK gave the idea, concept, and facilities for the same. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been supported by grants from Department of Biotechnology (DBT) and University Grant Commission (UGC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00873/full#supplementary-material

FIGURE S1 | Distribution of different heat stress responsive *cis*-acting elements in the 2-kb URR of *TaSTI* gene family members. *Cis*-acting elements were identified using PLACE and PlantCare databases. Different elements are depicted in different colors.

FIGURE S2 | Prediction of a three-dimensional structure of STI protein from *T. aestivum* and *Arabidopsis thaliana*. The structures were predicted using the Pyre2 web portal. TaSTI-2A protein was found to be similar to AtHOP3 protein.

FIGURE S3 | Depiction of NLS and ER signal motifs. The presence of ER retention signals, dilysine motifs (KKXX or KXKXX), and RXR motif in the TaSTI protein sequences are depicted by the highlighted residues in red color. The presence of bipartite NLS sequences are depicted by the underlined residues.

FIGURE S4 | Confirmation of *Arabidopsis* overexpression lines. *Arabidopsis* transgenics for the *TaSTI-2A* gene were confirmed by using **(A)** gene-specific and **(B)** hygromycin-specific PCR. **(C)** Expression profile of *TaSTI-2A* in WT and overexpression transgenic lines of *Arabidopsis*. The transcription level in WT was normalized as 1.0 and the results shown are the means ± SD of at least three independent experiments.

FIGURE S5 | Confirmation of *rice* overexpression lines. Rice transgenics for the *TaSTI-2A* gene were confirmed by using a gene-specific forward primer and vector-specific reverse primer.

REFERENCES

Agarwal, P., and Khurana, P. (2018). Characterization of a novel zinc finger transcription factor (TaZnF) from wheat conferring heat stress tolerance in *Arabidopsis*. *Cell Stress Chaperones* 23, 253–267. doi: 10.1007/s12192-017-0838-1

Amoah, J. N., Ko, C. S., Yoon, J. S., and Weon, S. Y. (2019). Effect of drought acclimation on oxidative stress and transcript expression in wheat (*Triticum aestivum* L.). *J. Plant Interact.* 14, 492–505. doi: 10.1080/17429145.2019.1662098

Arnon, D. I. (1949). Copper enzymes in isolated chloroplast. Polyphenoloxidase in *Beta vulgaris*. *Plant Physiol.* 24, 1–15. doi: 10.1104/pp.24.1.1

Baindur-Hudson, S., Edkins, A. L., and Blatch, G. L. (2015). Hsp70/Hsp90 organising protein (hop): beyond interactions with chaperones and prion proteins. *Subcell Biochem.* 78, 69–90. doi: 10.1007/978-3-319-11731-7_3

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170, 2172–2186. doi: 10.1104/pp.15.01667

Bose, S., Weikl, T., Bugl, H., and Buchner, J. (1996). Chaperone function of Hsp90-associated proteins. *Science* 274, 1715–1717. doi: 10.1126/science.274.5293.1715

Buchberger, A., Bukau, B., and Sommer, T. (2010). Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell.* 40, 238–252. doi: 10.1016/j.molcel.2010.10.001

Chauhan, H., Khurana, N., Tyagi, A. K., Khurana, J. P., and Khurana, P. (2011). Identification and characterization of high temperature stress responsive genes in bread wheat (*Triticum aestivum* L.) and their regulation at various stages of development. *Plant Mol. Biol.* 75, 35–51. doi: 10.1007/s11103-010-9702-8

Chen, L., Hamada, S., Fujiwara, M., Zhu, T., Thao, N. P., Wong, H. L., et al. (2010). The Hop/Sti1-Hsp90 chaperone complex facilitates the maturation and transport of a PAMP receptor in rice innate immunity. *Cell Host Microb.* 7, 185–196. doi: 10.1016/j.chom.2010.02.008

Chen, S., and Smith, D. F. (1998). Hop as an adaptor in the heat shock protein 70 (Hsp70) and hsp90 chaperone machinery. *J. Biol. Chem.* 273, 35194–35200. doi: 10.1074/jbc.273.52.35194

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Fellerer, C., Schweiger, R., Schongruber, K., Soll, J., and Schwenkert, S. (2011). Cytosolic HSP90 cochaperones HOP and FKBP interact with freshly synthesized chloroplast preproteins of *Arabidopsis*. *Mol. Plant* 4, 1133–1145. doi: 10.1093/mp/ssr037

Fernandez-Bautista, N., Fernandez-Calvino, L., Munoz, A., and Castellano, M. M. (2017). HOP3, a member of the HOP family in *Arabidopsis*, interacts with BiP and plays a major role in the ER stress response. *Plant Cell Environ.* 40, 1341–1355. doi: 10.1111/pce.12927

Fernandez-Bautista, N., Fernandez-Calvino, L., Munoz, A., Toribio, R., Mock, H. P., and Castellano, M. M. (2018). HOP family plays a major role in long-term acquired thermotolerance in *Arabidopsis*. *Plant Cell Environ.* 41, 1852–1869. doi: 10.1111/pce.13326

Freeman, B. C., Toft, D. O., and Morimoto, R. I. (1996). Molecular chaperone machines: chaperone activities of the cyclophilin Cyp-40 and the steroid aporeceptor-associated protein p23. *Science* 274, 1718–1720. doi: 10.1126/science.274.5293.1718

Giri, A., Heckathorn, S., Mishra, S., and Krause, C. (2017). Heat stress decreases levels of nutrient-uptake and -assimilation proteins in tomato roots. *Plants* 6:6. doi: 10.3390/plants6010006

Hahn, A., Bublak, D., Schleiff, E., and Scharf, K. D. (2011). Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. *Plant Cell* 23, 741–755. doi: 10.1105/tpc.110.076018

Hairat, S., and Khurana, P. (2015). Improving photosynthetic responses during recovery from heat treatments with brassinosteroid and calcium chloride in Indian bread wheat cultivars. *Am. J. Plant Sci.* 6, 1827–1849. doi: 10.4236/ajps.2015.611184

Hamdi, K., Brini, F., Kharrat, N., Masmoudi, K., and Yakoubi, I. (2020). Abscisic Acid, stress, and ripening (TtASR1) gene as a functional marker for salt tolerance in durum wheat. *Biomed. Res. Intern.* 2019. doi: 10.1155/2020/7876357

Havaux, M. (1992). Stress tolerance of photosystem II in vivo: antagonistic effects of water, heat, and photoinhibition stresses. *Plant Physiol.* 100, 424–432. doi: 10.1104/pp.100.1.424

Hernandez, M. P., Sullivan, W. P., and Toft, D. O. (2002). The assembly and intermolecular properties of the hsp70-Hop-hsp90 molecular chaperone complex. *J. Biol. Chem.* 277, 38294–38304. doi: 10.1074/jbc.m206566200

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant Cis-Acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297

Honore, B., Leffers, H., Madsen, P., Rasmussen, H. H., Vandekerckhove, J., and Celis, J. E. (1992). Molecular cloning and expression of a transformation-sensitive human protein containing the TPR motif and sharing identity to the stress-inducible yeast protein STI1. *J. Biol. Chem.* 267, 8485–8491.

Houck, S. A., Singh, S., and Cyr, D. M. (2012). Cellular responses to misfolded proteins and protein aggregates. *Methods Mol. Biol.* 832, 455–461. doi: 10.1007/978-1-61779-474-2_32

Karam, J. A., Parikh, R. Y., Nayak, D., Rosenkranz, D., and Gangaraju, V. K. (2017). Co-chaperone Hsp70/Hsp90-organizing protein (Hop) is required for transposon silencing and Piwi-interacting RNA (piRNA) biogenesis. *J. Biol. Chem.* 292, 6039–6046. doi: 10.1074/jbc.c117.777730

Khurana, N., Chauhan, H., and Khurana, P. (2013). Wheat chloroplast targeted sHSP26 promoter confers heat and abiotic stress inducible expression in transgenic *Arabidopsis* plants. *PLoS One* 8:e54418. doi: 10.1371/journal.pone.0054418

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10171–10176. doi: 10.1073/pnas.0900604106

Lamm, C. E., Kraner, M. E., Hofmann, J., Bornke, F., Mock, H. P., and Sonnewald, U. (2017). Hop/Sti1 - a two-faced cochaperone involved in pattern recognition receptor maturation and viral infection. *Front. Plant Sci.* 8:1754. doi: 10.3389/fpls.2017.01754

Law, R. D., and Crafts-Brandner, S. J. (1999). Inhibition and acclimation of photosynthesis to heat stress is closely correlated with activation of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Plant Physiol.* 120, 173–182. doi: 10.1104/pp.120.1.173

Lee, L. Y., Fang, M. J., Kuang, L. Y., and Gelvin, S. B. (2008). Vectors for multi-color bimolecular fluorescence complementation to investigate protein-protein interactions in living plant cells. *Plant Methods* 4:24. doi: 10.1186/1746-4811-4-24

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant Cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Li, J., Soroka, J., and Buchner, J. (2012). The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim. Biophys. Acta* 1823, 624–635. doi: 10.1016/j.bbamcr.2011.09.003

Lopes, M. H., Hajj, G. N., Muras, A. G., Mancini, G. L., Castro, R. M., Ribeiro, K. C., et al. (2005). Interaction of cellular prion and stress-inducible protein 1 promotes neuritogenesis and neuroprotection by distinct signaling pathways. *J. Neurosci.* 25, 11330–11339. doi: 10.1523/jneurosci.2313-05.2005

Mayer, M. P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol. Life Sci.* 62, 670–684. doi: 10.1007/s00018-004-4464-6

Nakajima, Y., and Suzuki, S. (2013). Environmental stresses induce misfolded protein aggregation in plant cells in a microtubule-dependent manner. *Int. J. Mol. Sci.* 14, 7771–7783. doi: 10.3390/ijms14047771

Nicolet, C. M., and Craig, E. A. (1989). Isolation and characterization of STI1, a stress-inducible gene from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9, 3638–3646. doi: 10.1128/mcb.9.9.3638

Niu, Y., and Xiang, Y. (2018). An overview of biomembrane functions in plant responses to high-temperature stress. *Front. Plant Sci.* 9:915. doi: 10.3389/fpls.2017.00915

Peng, Z., Wang, M., Li, F., Lv, H., Li, C., and Xia, G. (2009). A Proteomic study of the response to salinity and drought stress in an introgression strain of bread wheat. *Mol. Cell. Proteom.* 8, 2676–2686. doi: 10.1074/mcp.m900052-mcp200

Prodromou, C. (2012). The 'active life' of Hsp90 complexes. *Biochim. Biophys. Acta* 1823, 614–623. doi: 10.1016/j.bbamcr.2011.07.020

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., et al. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* 101, 199–210.

Song, H. O., Lee, W., An, K., Lee, H. S., Cho, J. H., Park, Z. Y., et al. (2009). *C. elegans* STI-1, the homolog of Sti1/Hop, is involved in aging and stress response. *J. Mol. Biol.* 390, 604–617. doi: 10.1016/j.jmb.2009.05.035

Suzuki, N., and Mittler, R. (2005). Reactive oxygen species and temperature stresses: a delicate balance between signaling and destruction. *Physiol. Plant.* 126, 45–51. doi: 10.1111/j.0031-9317.2005.00582.x

Tewari, A. K., and Tripathy, B. C. (1998). Temperature-stress-induced impairment of chlorophyll biosynthetic reactions in cucumber and wheat. *Plant Physiol.* 117, 851–858. doi: 10.1104/pp.117.3.851

Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S., et al. (2006). Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. *Plant J.* 47, 969–976. doi: 10.1111/j.1365-313x.2006.02836.x

Usman, M. G., Rafii, M. Y., Martini, M. Y., Yusuff, O. A., Ismail, M. R., and Miah, G. (2017). Molecular analysis of Hsp70 mechanisms in plants and their function in response to stress. *Biotechnol. Genet. Eng. Rev.* 33, 26–39. doi: 10.1080/02648725.2017.1340546

von Koskull-Doring, P., Scharf, K. D., and Nover, L. (2007). The diversity of plant heat stress transcription factors. *Trends Plant Sci.* 12, 452–457. doi: 10.1016/j.tplants.2007.08.014

Wahid, A., Gelani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: an overview. *Environ. Exp. Bot.* 61, 199–223. doi: 10.1016/j.envexpbot.2007.05.011

Walther, D., Brunnemann, R., and Selbig, J. (2007). The regulatory code for transcriptional response diversity and its relation to genome structural properties in *A. thaliana*. *PLoS Genet.* 3:e11. doi: 10.1371/journal.pone.000011

Wang, Q. L., Chen, J. H., He, N. Y., and Guo, F. Q. (2018). Metabolic reprogramming in chloroplasts under heat stress in plants. *Int. J. Mol. Sci.* 19:849. doi: 10.3390/ijms19030849

Xu, K., Lin, J. Y., and Nagy, P. D. (2014). The hop-like stress-induced protein 1 cochaperone is a novel cell-intrinsic restriction factor for mitochondrial tombusvirus replication. *J. Virol.* 88, 9361–9378. doi: 10.1128/jvi.00561-14

Xu, Z. S., Li, Z. Y., Chen, Y., Chen, M., Li, L. C., and Ma, Y. Z. (2012). Heat shock protein 90 in plants: molecular mechanisms and roles in stress responses. *Int. J. Mol. Sci.* 13, 15706–15723. doi: 10.3390/ijms131215706

Yu, W., Wang, L., Zhao, R., Sheng, J., Zhang, S., Li, R., et al. (2019). Knockout of SlMAPK3 enhances tolerance to heat stress involving ROS homeostasis in tomato plants. *BMC Plant Biol.* 19:354. doi: 10.1186/s12870-019-1939-z

Zanata, S. M., Lopes, M. H., Mercadante, A. F., Hajj, G. N., Chiarini, L. B., Nomizo, R., et al. (2002). Stress-inducible protein 1 is a cell surface ligand for cellular prion that triggers neuroprotection. *EMBO J.* 21, 3307–3316. doi: 10.1093/emboj/cdf325

Zhang, Z., Quick, M. K., Kanelakis, K. C., Gijzen, M., and Krishna, P. (2003). Characterization of a plant homolog of hop, a cochaperone of Hsp90. *Plant Physiol.* 131, 525–535. doi: 10.1104/pp.011940

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Meena, Deb, Samtani and Khurana. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 11 September 2020
doi: 10.3389/fgene.2020.574549

Structural Aspects of DNA Repair and Recombination in Crop Improvement

*Prabha Verma, Reetika Tandon, Gitanjali Yadav and Vineet Gaur**

National Institute of Plant Genome Research, New Delhi, India

The adverse effects of global climate change combined with an exponentially increasing human population have put substantial constraints on agriculture, accelerating efforts towards ensuring food security for a sustainable future. Conventional plant breeding and modern technologies have led to the creation of plants with better traits and higher productivity. Most crop improvement approaches (conventional breeding, genome modification, and gene editing) primarily rely on DNA repair and recombination (DRR). Studying plant DRR can provide insights into designing new strategies or improvising the present techniques for crop improvement. Even though plants have evolved specialized DRR mechanisms compared to other eukaryotes, most of our insights about plant-DRRs remain rooted in studies conducted in animals. DRR mechanisms in plants include direct repair, nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), non-homologous end joining (NHEJ) and homologous recombination (HR). Although each DRR pathway acts on specific DNA damage, there is crosstalk between these. Considering the importance of DRR pathways as a tool in crop improvement, this review focuses on a general description of each DRR pathway, emphasizing on the structural aspects of key DRR proteins. The review highlights the gaps in our understanding and the importance of studying plant DRR in the context of crop improvement.

Keywords: DNA repair and recombination, photolyases, glycosylases, structure-specific endonucleases, transcriptomic interventions, crop improvement

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University,
United States

Reviewed by:
Jose Sebastian,
Indian Institute of Science Education and Research Berhampur (IISER), India
Subodh Kumar Sinha,
Indian Council of Agricultural Research (ICAR), India

***Correspondence:**
Vineet Gaur
vgaur@nipgr.ac.in;
vineetgaur1982a@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *20 June 2020*
Accepted: *25 August 2020*
Published: *11 September 2020*

Citation:
Verma P, Tandon R, Yadav G and Gaur V (2020) Structural Aspects of DNA Repair and Recombination in Crop Improvement. Front. Genet. 11:574549. doi: 10.3389/fgene.2020.574549

INTRODUCTION

Agriculture played an essential role in directing human evolution from hunter-gatherer to agro-pastoralist lifestyle (Hervella et al., 2012), which in turn resulted in changed feeding habits (Luca et al., 2010) and steep increase in population growth rates (Zahid et al., 2016). However, agriculture is now threatening various ecosystems (Foley et al., 2011; Tilman et al., 2011). The combined effect of exponentially increasing global population and concomitant increase in malnutrition has put considerable strains on agriculture. The green revolution in the 1960s considerably enhanced crop production (Conway and Toenniessen, 1999) but was limited to a few species and geographical regions. However, present-day crops are more vulnerable to stress, with greater dependence on chemical pesticides, and productivity is still unable to meet the demand. Therefore, agriculture needs a second revolution (Wollenweber et al., 2005) to increase productivity without increasing the cultivable land (Foley et al., 2011). Conventional

breeding methods (Breseghello and Coelho, 2013) or an understanding of genetic engineering (Genetic modification and Genome editing) (Zhang et al., 2018) can assist in realizing these goals. Even though conventional plant breeding (hybridization and selection to achieve rearranging of the genome) is the preferred approach, but it is time-consuming and labor-intensive (Blary and Jenczewski, 2019; Zhang et al., 2018). On the other hand, technological advancement, and availability of gene sequences have enabled researchers to either insert a DNA sequence (genetic modification, GM) or precisely edit any gene sequence of the plant (genome editing). Coupling genetic modification and genome editing with conventional plant breeding can expedite the research for crop improvement.

Genetic modification involves the transfer of a foreign nucleic acid (transgenic, cisgenic, or intragenic) into a plant of economic importance resulting in generating an entirely new trait (*e.g.,* tolerance against various biotic and abiotic stresses) (Bawa and Anilakumar, 2013; Kamle et al., 2017). However, GM crops are associated with controversies of social, environmental, and human health-related aspects (Bawa and Anilakumar, 2013; Jones, 2015; Ali et al., 2019). Genome editing, in contrast, has emerged in recent years as a more acceptable alternative to transgenic modification since the introduced changes mimic natural changes to a large extent. Genome editing employs site-directed nucleases (Zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system) to precisely make alterations in a pre-determined site in a sequence-specific manner to alter the function of the target gene (Voytas, 2013; Chen and Gao, 2014; Zhang et al., 2018). The basic principle of crop improvement using site-specific endonucleases relies on the generation of double-stranded breaks (DSBs), which are then repaired by the internal DNA repair and recombination (DRR) machinery of the plant itself (Symington and Gautier, 2011).

DNA can undergo damage due to various exogenous (Ionizing radiations, UV-radiations, alkylating agents) or endogenous (Intracellular metabolism and DNA metabolism) factors resulting in a variety of different DNA lesions as listed in **Table 1**. These DNA lesions, if not repaired, can result in impaired cellular processes, and lead to genome toxicity. Therefore, all organisms have evolved a range of DRR mechanisms. Plants being sedentary have further evolved DRR mechanisms (Singh et al., 2010) and, thus, an in-depth and objective study on DRR in plants is crucial. Among plants, DRR mechanisms operate in tissue-specific, developmentally regulated, and cell-cycle dependent manner. Some DRR mechanisms are antagonistic, while others are redundant with entirely different outcomes. Some DRR pathways are efficient, while others are inherently more error-prone. Plant DRR mechanisms constitute a delicately regulated process; they can slow down with the aging of plants (Britt, 1999; Hefner et al., 2006; Uchiyama et al., 2004; Manova and Gruszka, 2015). Many of these pathways play an essential role in DNA repair in somatic cells, whereas the same pathways play an important role in genetic recombination (Schuermann et al., 2005). Most of our understanding regarding DRR mechanisms in plants comes from structural and biochemical studies in prokaryotes, yeast, and animals. Plant DRR related genes and proteins have been identified through homology-based searches, and there is still a wide gap in their structural and biochemical studies. Therefore, the information about plant DRR is available only in bits and pieces.

DRR is an important life process involved in the maintenance of genome stability and is equally vital for application-based work such as crop improvement. Regardless of the approach: traditional plant breeding or targeted, the success rate is mostly dependent upon the complex interplay between various DRR pathways (Manova and Gruszka, 2015). This review aims to provide an overview of different DRR pathways, with emphasis on scope and extent of available knowledge in the plant kingdom, as well as structural and biochemical aspects of various DRR mechanisms and their potential for crop improvement.

DNA REPAIR AND RECOMBINATION IN PLANTS

Plants have achieved substantial specializations in their DNA repair and recombination methods compared to other living organisms due to their sedentary lifestyle and inability to avoid environmental factors that could ultimately result in DNA damage (**Table 1**). The main mechanisms of DRR in plants are direct repair (Photoreactivation), base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), non-homologous end joining (NHEJ), homologous recombination (HR), and translesion DNA synthesis (TLS) as shown in **Figure 1** (Britt, 1999; Singh et al., 2010; Manova and Gruszka, 2015). Understanding of DRR mechanisms in plants comes from studies in other organisms (bacteria, yeast, and animals). The following section provides a brief account of the various mechanisms of DRR in plants and the differences exhibited by plant DRR compared to other living organisms.

Direct Repair

Direct repair involves a direct reversal of DNA lesions by enzymatic reactions; therefore, it is an error-free pathway. Ultraviolet radiations are the most common DNA damaging agent. UV radiation mostly generates cyclobutane pyrimidine dimers (CPDs), and [6-4]pyrimidinone dimers (also called (6-4) photoproducts) (Britt, 1999). These lesions can be repaired either by light-dependent direct repair (photoreactivation) or by light-independent excision of the lesion (dark repair). Photoreactivation is carried out by a class of enzymes called photolyases, which shows activity in the presence of light (360–420 nm) (Brettel and Byrdin, 2010). Plants have two types of photolyases: Class II CPD photolyase and (6-4) photolyase. Photoreactivation is one of the well-studied mechanisms of DNA repair in plants (Maul et al., 2008; Hitomi et al., 2009, 2012). Spontaneously occurring photolyase variants are

TABLE 1 | Types of DNA lesions and their repair mechanisms.

DNA lesion	Source	Repair mechanisms[#]
Mismatch lesion 	*Endogenous sources:* Nucleotide misincorporation during DNA Replication, Homologous recombination, Spontaneous hydrolytic deamination of DNA bases *Exogenous sources:* Oxidizing agents, Alkylating agents	MMR BER
Single strand breaks 	*Endogenous sources:* Endogenous reactive oxygen species, Enzymatic cleavage of phosphodiester bond during BER, Abortive DNA Topoisomerase I activity *Exogenous Sources:* Ionizing radiations, Anti-cancer drugs	BER
Double strand breaks 	*Endogenous sources:* Endogenous reactive oxygen species, DNA replication and repair, Meiosis I, T-cell receptor formation Immunoglobulin class switching Excision of transposable element *Exogenous Sources:* Ionizing radiations, Radiomimetic compounds Engineered nucleases	HR NHEJ
DNA intrastrand crosslinks 	*Endogenous sources:* Reactive aldehydes, Endogenous reactive oxygen species *Exogenous Sources:* Ultraviolet radiation, Ionizing radiations, Cisplatin, Mitomycin C	Photoreactivation NER TLS
DNA interstrand crosslinks 	*Endogenous sources:* Nitrous acid, Reactive aldehydes, *Exogenous Sources:* Alkylating agents, Cisplatin, Mitomycin C, Psoralens	NER HR NHEJ TLS
DNA-Protein crosslinks 	*Endogenous sources:* Reactive aldehydes, metals, Enzymatic DNA-Protein crosslinks *Exogenous sources:* Ultraviolet radiation, Ionizing radiations, Chemotherapeutic agents	NER HR Crosslink hydrolysis Proteolysis
Base modification 	*Endogenous sources:* Nitric oxide, Superoxide, Spontaneous hydrolysis of N-glycosidic bond Spontaneous hydrolytic deamination of DNA bases DNA methylation *Exogenous sources:* Ultraviolet radiation, Ionizing radiations, Alkylating agents	BER NER TLS

[#]*MMR, Mismatch Repair; BER, Base Excision Repair; HR, Homologous Recombination; NHEJ, Non-Homologous End Joining; NER, Nucleotide Excision Repair; TLS, Translesion DNA Synthesis. *Site of DNA damage.*

associated with differing plant growth and productivity (Hidema et al., 1997; Hitomi et al., 2012). AGT (O^6-Alkylguanine-DNA-alkyltransferase) is another enzyme involved in the direct repair. AGT transfers the alkyl group of alkylated bases to a Cys residue of the enzyme in an irreversible reaction. Interestingly, till now, AGT homologs have not been found in plants (Pegg, 2011).

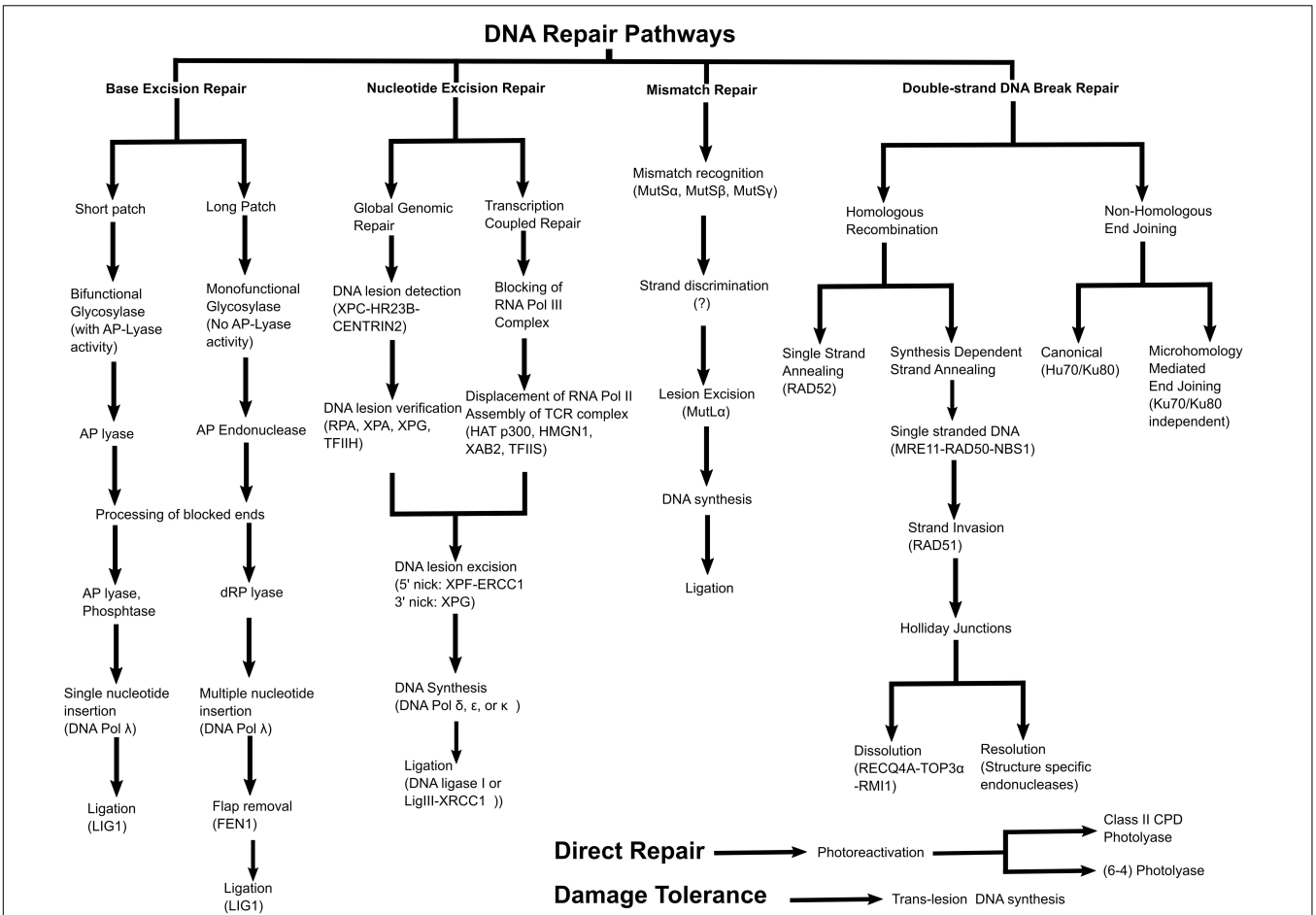


FIGURE 1 | Summary of various DNA repair and recombination pathways.

Base Excision Repair (BER)

Base Excision Repair repairs DNA lesions resulting from oxidation, alkylation, or deamination of the nitrogenous base of DNA (Roldan-Arjona et al., 2019). The first step involves damage recognition and cleavage of N-glycosidic bond resulting in the generation of an abasic site (AP site) by DNA glycosylases. Spontaneous hydrolysis of the N-glycosidic bond can also generate an AP site (Sheppard et al., 2000). The AP site thus generated is further processed either by AP lyase (short patch BER) or AP endonuclease (long patch BER). Both AP lyase and AP endonuclease generate blocked ends that need to be processed further to facilitate the DNA polymerase and DNA ligase activities. Short patch BER inserts only one nucleotide, whereas long patch BER inserts multiple nucleotides resulting in a 5′ flap structure, requiring further processing by flap endonuclease (FEN1) before ligation. The mechanism of BER in plants is almost identical to other organisms with specific differences (Manova and Gruszka, 2015; Roldan-Arjona et al., 2019). Among mammals, DNA polymerase β and DNA ligase III play a significant role in short patch BER (Kubota et al., 1996). However, both these proteins are absent in plants. DNA polymerase λ may be participating in short patch BER in plants. The role of replicative polymerases in long patch BER among plants needs further scrutiny (Roldan-Arjona et al., 2019). In the absence of DNA ligase III (LIG 3) in plants, an ortholog of mammalian ligase I (*At*LIG1) play a role in both short patch as well as long patch BER (Cordoba-Canero et al., 2011).

Nucleotide Excision Repair (NER)

Nucleotide Excision Repair is the most conserved DNA repair mechanism among all the eukaryotes. It primarily repairs UV induced DNA lesions and bulky DNA adducts. Based on the mode of DNA lesion identification, NER can operate as (1) Global genomic repair (repair influenced by a lesion-induced change in DNA structure) or (2) Transcription coupled repair (repair initiated by lesion-induced transcription inhibition). The NER in plants is like other eukaryotes (Xu et al., 1998; Gallego et al., 2000; Liu et al., 2000; Molinier et al., 2004; Canturk et al., 2016). The NER involves lesion recognition, verification, excision, DNA synthesis, and finally, ligation of the repaired stretch. In global genome repair, a heterotrimeric complex (XPC/HR23B/Centrin2, **Table 2**) detects DNA lesion as a local distortion in the DNA structure. Verification of the lesion and identification of the DNA strand harboring the lesion is carried

TABLE 2 | Frequently used abbreviations in the article.

BER	Base Excision Repair
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDM1	Decrease in DNA methylase 1
DRR	DNA Repair and Recombination
DSBs	Double Stranded DNA Breaks
EME1	Essential Meiotic Structure-Specific Endonuclease 1
ERCC1	Excision Repair Cross-Complementation Group 1
FEN1	Flap Endonuclease 1
GEN1	Flap Endonuclease GEN Homolog 1
HAT P300	Histone Acetyltransferase P300
HJs	Holliday Junctions
HMGN1	High Mobility Group Nucleosome Binding Domain 1
HR	Homologous Recombination
HR23B	Human homologue of Rad23B
iHJ	Intact Holliday Junction
Ku70	Lupus Ku autoantigen protein p70
Ku80	Lupus Ku autoantigen protein p80
LIG3	DNA Ligase 3
MET1	Methyltransferase 1
MMEJ	Microhomology Mediated End Joining
MMR	Mismatch Repair
MRE11	Meiotic Recombination 11
MUS81	MMS and UV Sensitive
nHJ	Nicked Holliday Junction
NBS1	Nibrin
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
PARP	Poly(ADP-Ribose) Polymerase
RECQ4A	RecQ helicase
RMI1	RecQ Mediated Genome Instability 1
RTR	RECQ4A-TOP3α-RMI1
SDNs	Site directed nucleases
SDSA	Synthesis-Dependent Strand Annealing
SEND1	Single-strand DNA endonuclease 1
SLX1	Synthetic Lethal of Unknown (X) function
SLX4	Synthetic Lethal of Unknown (X) function
SSA	Single-Strand Annealing
SSEs	Structure-Specific Endonucleases
TALENs	Transcription Activator-Like Effector Nucleases
TFIIH	Transcription factor II Human
TLS	Translesion DNA Synthesis
TOP3α	DNA Topoisomerase III Alpha
XAB2	XPA Binding Protein 2
XPA	Xeroderma pigmentosum, complementation group A
XPC	Xeroderma pigmentosum, complementation group C
XPF	Xeroderma pigmentosum, complementation group F
XPG	Xeroderma pigmentosum, complementation group G
XRCC1	X-Ray Repair Cross Complementing 1
ZFNs	Zinc Finger Nucleases

out by a multi-protein complex (RPA/XPA/XPG/TFIIH, **Table 2**). Verification is followed by nicks on 5′ and 3′ sides of the lesion by XPF (Xeroderma pigmentosum, complementation group F)-ERCC1 (Excision Repair Cross-Complementation Group 1) and XPG (Xeroderma pigmentosum, complementation group G) nucleases, respectively. The downstream events involve DNA synthesis (DNA polymerases δ, ε, or κ) and ligation (DNA ligase I or DNA ligase III/XRCC1). Transcription coupled repair initiates when a DNA lesion in the template strand halts the progression of RNA Pol II complex. Repair machinery (HAT p300/HMGN1/XAB2/TFIIS, **Table 2**) displaces halted RNA Pol II complex. XPF-ERCC1 and XPG incise the lesion, followed by DNA synthesis and ligation. In addition to photoreactivation, players of NER are also important targets for enhancing UV tolerance among plants. There is a crosstalk between NER and HR through Centrin2. A defect in Centrin2, a key player in detecting DNA lesion in global genome repair, results in an enhanced somatic HR (Dubest et al., 2004; Molinier et al., 2004, 2008). Therefore, NER is one of the attractive targets for crop modification since it can be used to fine-tune the recombination frequency.

Mismatch Repair (MMR)

Mismatch Repair corrects mismatches due to misincorporation of nucleotides during DNA replication and recombination. The prokaryotic MMR provides the basic MMR template for deciphering the molecular events involved in eukaryotic MMR (Harfe and Jinks-Robertson, 2000; Li, 2008). The essential players in prokaryotic MMR are homodimeric MutS (with an ATPase activity) (Lamers et al., 2003), homodimeric MutL (with an ATPase activity) (Spampinato and Modrich, 2000), and monomeric MutH (Ban and Yang, 1998). MutS identifies the lesion (mismatches and loops arising from insertion and deletion) and recruits MutL. MutL, in turn, recruits and activate MutH. MutH distinguishes the daughter strand from the parental strand by binding the nearest (either on 3′ or 5′ side) hemimethylated dGATC sequence to the lesion (Lahue et al., 1989). Binding of MutH is followed by an incision in the daughter strand and recruitment of UvrD (helicase II). UvrD unwinds the DNA duplex towards the mismatch lesion (Matson and Robertson, 2006). SSB (single-stranded DNA binding protein) protects the single-stranded DNA. Based on the location of incision relative to the lesion either a 3′-5′ exonuclease (ExoI or ExoX) or a 5′-3′ exonuclease (ExoVII or RecJ) removes the nicked strand. DNA Polymerase III and DNA ligase finally complete the repair process (Harfe and Jinks-Robertson, 2000; Li, 2008). Among eukaryotes, MutS has given rise to 6 genes (7 genes in plants) called MSH (MutS homologs) (Fishel and Wilson, 1997; Pochart et al., 1997; Culligan and Hays, 2000) and MutL gene gave rise to MLH (MutL homologs) and PMS (Post Meiotic Segregation) (Kolodner, 1996; Wang et al., 1999). In eukaryotes, mismatch recognition is carried out by MutSα (MSH2/MSH6 heterodimer), and MutSβ (MSH2/MSH3 heterodimer). Plants also have a third heterodimeric complex MutSγ (MSH2/MSH7). All these complexes have various overlapping substrate specificities (Wu et al., 2003; Emmanuel et al., 2006; Gomez and Spampinato, 2013; Chirinos-Arias and Spampinato, 2020). MutLα (MLH1/PMS2) heterodimer carries out the function of MutL. Unlike prokaryotic MutL, eukaryotic MutLα also has an endonuclease activity (Kadyrova and Kadyrov, 2016). Since eukaryotes lack methylation and MutJ homologs, strand-specific cleavage in eukaryotic MMR is not

clear (Li, 2008). DNA re-synthesis and DNA ligation follow the excision of the lesion. It is noteworthy that, besides repair of mismatches, MMR repairs UV induced lesions and inhibits HR between divergent sequences, therefore, maintain the interspecies barrier (Tham et al., 2016). Correction of UV lesions and inhibition of HR have implications in the efficiency of methods used in crop improvement.

Double-Stranded DNA Break Repair (DSB Repair)

Double-stranded DNA breaks (DSBs) are the most lethal type of DNA damages, and they are also the most important lesions in terms of crop-improvement. Two independent mechanisms repair DSBs: homologous recombination (HR) and non-homologous end joining (NHEJ). Repair of DSBs via HR depends upon the information from homologous sequences, whereas NHEJ does not require any sequence information and therefore is more error-prone compared to HR. In plants, NHEJ is more frequent than HR: a safeguard mechanism to prevent recombination between two non-allelic sequences (Puchta and Fauser, 2014; Manova and Gruszka, 2015).

Homologous recombination plays an essential role in DNA repair in somatic cells and the generation of diversity in meiotic cells (Schuermann et al., 2005). HR is also involved in repairing interstrand crosslinks and helps in the restart of the stalled replication fork (Li and Heyer, 2008; Manova and Gruszka, 2015). HR can proceed through one of the two mechanisms: (1) Single-strand annealing (SSA), and (2) synthesis-dependent strand annealing (SDSA) (Kimura and Sakaguchi, 2006; Li and Heyer, 2008; Manova and Gruszka, 2015; Wright et al., 2018). Repair by SSA takes place when DSB occurs in between homologous sequences with the help of RAD52. SSA involves the removal of non-homologous overhangs, and therefore, this process is error-prone but plays a significant role in molecular evolution (Puchta, 2005). SDSA, on the other hand, relies on homologous sequences present on the sister chromatid or homologous chromosome. SDSA initiates with the formation of a single-stranded DNA (MRN complex: MRE11/RAD50/NBS1, **Table 2**) (Akutsu et al., 2007) followed by strand invasion with the help of RAD51, resulting in the formation of a D-loop (Abe et al., 2005; Osakabe et al., 2005). One of the intermediates in this process is the formation of a four-way DNA junction, known as Holliday Junction (HJ). These HJs need to be processed into two independent DNA duplexes. The processing of HJs is carried out by two independent mechanisms: (1) dissolution: requiring action of helicase (RECQ4A), topoisomerase (TOP3α) and a structural protein RMI1 (RecQ Mediated Genome Instability 1) (Bagherieh-Najjar et al., 2005; Rohrig et al., 2016), and (2) resolution: requiring action of structure-specific endonucleases (SSEs) (Matos and West, 2014). Dissolution gives rise to non-crossovers, but resolution can give rise to both non-crossovers as well as crossovers depending upon how the SSEs make nick at the crossover point of HJs. SSEs have been characterized extensively in bacteria, yeast, and animals. The function of SSEs is not restricted to HJ resolution only; a few also participate in other DRR mechanisms like interstrand crosslink repair and NER. Sequence homology allowed the identification of various SSEs in plants (GEN1, SEND1, MUS81-EME1, SLX1, FEN1, XPF-ERCC1, **Table 2**). Many of the SSEs from plants need to be characterized further for an in-depth understanding of HR in plants (**Table 3**).

Non-homologous end joining repairs DSBs along with HR. Even though NHEJ is sequence-independent and error-prone, it is still the most efficient method of DSB repair among plants (Puchta, 2005). One of the applications of NHEJ is during the integration of T-DNA into the genome during transformation (Park et al., 2015). In the canonical method of NHEJ, Ku70/Ku80 heterodimers bind to DSB ends to prevent further degradation and bring them in proximity. In mammals, Ku70/Ku80 (Lupus Ku autoantigen protein p70/ Lupus Ku autoantigen protein p80) recruits DNA-PKcs (DNA dependent protein kinases), followed by the action of nucleases, DNA polymerases, and ligases. *At*KU70, *At*KU80, LIG4 proteins have been found in plants, while DNA-PKc kinase has not been identified so far (Nishizawa-Yokoi et al., 2012). A competing Ku70/Ku80 independent, alternate NHEJ pathway, MMEJ (Microhomology mediated end joining), is also known to repair DSBs. MMEJ utilizes microhomologous sequences at the DNA ends (Wang and Xu, 2017). MMEJ requires removal of flap strands after microhomology based sequence alignment, followed by ligation. Since this method involves trimming of the ends of DNA, MMEJ is highly mutagenic. The mechanism involves PARPs (Poly(ADP-Ribose) Polymerases), XRCC1, XPF, MRE11, LIG3 (Jia et al., 2013).

Translesion DNA Synthesis (TLS)

Photoreactivation, BER, NER, or MMR take care of lesions arising from UV exposure or oxidative damage. However, all these processes are not capable of completely removing these lesions. Any of these lesions left behind stalls the progression of replicative DNA polymerase. A stalled replication fork can cause genome toxicity. However, plants and other organisms have evolved mechanisms to restart these stalled replication forks by bypassing the DNA lesions. Bypassing of the DNA lesions involves the removal of replicative DNA polymerase and binding of specialized DNA polymerases called TLS polymerases (Lehmann et al., 2007). TLS polymerases are known to have a spacious active site to accommodate bulky DNA lesions (Plosky and Woodgate, 2004). TLS polymerases mostly belong to the Y-family of DNA polymerases. These polymerases have been extensively studied biochemically and structurally in bacteria, yeast, and humans. Most of the TLS polymerases from plants (DNA polymerase ζ, η, κ, and Rev1) need to be characterized structurally and biochemically (Kimura and Sakaguchi, 2006; Roldan-Arjona and Ariza, 2009) (**Table 3**).

CROP IMPROVEMENT: PRESENT TECHNIQUES AND FUTURISTIC APPROACHES

Genetic diversity is necessary for crop improvement to generate novel combinations of genes to achieve desired phenotypes (Glaszmann et al., 2010). A crop genome accumulates

TABLE 3 | A summary of key proteins and their functions that have potential to be used as tool kits for crop improvement.

Name of the protein	Name of the gene[#]	Protein family	Mutation/ overexpression studies	Miscellaneous information	Structural information from plants*	References
Photolyases (Potential use in increasing UV resistance)						
Class II CPD photolyase	*OsPHR* (OSNPB_100167600) *PHR1* (AT1G12370)	PHR2 superfamily	*PHR1* mutants show no differences from wild type in the absence of UV light. In the presence of UV light, PHR*1* mutants show growth inhibition and leaf necrosis. *PHR1* overexpression resulted in enhanced DNA repair in *A. thaliana.*	Repair UV induced cyclobutane pyrimidine dimers. Essential for the survival in the presence of UV-B. Widespread expression in somatic tissues and during later flower development. No expression of *PHR1* in dark grown etiolated seedlings. More UV tolerant rice cultivars have Gln296 in class II CPD photolyase, while less tolerant cultivars have His at this position. Norin 1 is UV sensitive because of Gln to Arg change at position 126.	Class II CPD photolyase from *O. sativa* (PDB: 3UMV)	(Hidema et al., 2000) (Yamamoto et al., 2007) (Hitomi et al., 2012) (Willing et al., 2016) (Ahmad et al., 1997) (Landry et al., 1997) (Kaiser et al., 2009)
(6-4) photolyase	*OsUVR3* (OSNPB_020204400) *AtUVR3* (AT3G15620)	PhrB superfamily	A nonsense mutation results in photoreactivation defect	Repair UV induced (6-4) photoproducts. Expression of *AtUVR3* is downregulated by light in photosynthesis dependent manner. Structural studies show that local differences in the amino acids contribute toward the major functional differences presented by PHR/CRY family instead of large structural changes.	(6-4) photolyase from *A. thaliana* (PDB: 3FY4)	(Katarzyna Banas et al., 2018) (Hitomi et al., 2009) (Nakajima et al., 1998) (Jiang et al., 1997)
DNA glycosylases (Potential use in increasing resistance to abiotic and biotic stresses and altering efficacy of targeted mutagenesis)						
Uracil DNA glycosylase (UDG)	*AtUNG* (AT3G18630)	UDG superfamily	*AtUNG* mutant plants display a normal phenotype and increased resistance against 5-fluorouracil (5-FU) cytotoxicity.	Removes Uracil from DNA. Strict substrate specificity in contrast to bacterial or human enzymes.	No structure known. 54.55% identity with human UDG (PDB: 3TKB)	(Cordoba-Canero et al., 2010)
Alkylpurine DNA glycosylase (AAG)	*AtMAG* (AT3G12040)	AAG Superfamily	-	Removes alkylated purines. A role in DNA replication and cell growth. High expression of *AtMAG* in rapidly dividing tissues and growing leaves.	No structure known.	(Santerre and Britt, 1994) (Shi et al., 1997)

(Continued)

TABLE 3 | Continued

Name of the protein	Name of the gene[#]	Protein family	Mutation/ overexpression studies	Miscellaneous information	Structural information from plants*	References
8-oxoguanine-DNA glycosylase 1 (OGG1)	*AtOGG1* (AT1G21710)	HhH Superfamily	*AtOGG1* overexpression improves seed longevity and enhances abiotic stress tolerance. *AtOGG1* mutants show no phenotypic differences in comparison to the wild type.	Removes 7,8-dihydro-8-oxoguanine (8-oxo-G) from DNA.	No structure known. 38.91% identity with human 8-oxoguanine-DNA glycosylase (PDB: 1KO9)	(Chen et al., 2012) (Morales-Ruiz et al., 2003) (Garcia-Ortiz et al., 2001) (Dany and Tissier, 2001) (Murphy, 2005)
Methyl-CPG-binding domain 4-like (MBD4L)	*AtMBD4* (AT3G63030)	HhH Superfamily	*AtMBD4* overexpression induces the expression of LIG1 (also involved in BER). *AtMBD4* overexpression enhances tolerance to oxidative stress. Mutation at *AtMBD4* locus exhibits altered root architecture.	Removes uracil and thymine mispaired with G. CpG sequence context preferred. AtMBD4L lacks the signature MBD domain but possesses a conserved glycosylase domain important for conferring substrate specificity. *AtMBD4* expresses in perivascular leaf tissues, flowers, and the apex of immature siliques. AtMBD4L negatively regulates a subset of phosphate starvation genes.	No structure known. 34.43% identity with human MBD4 (PDB: 3IHO)	(Ramiro-Merina et al., 2013) (Nota et al., 2015) (Parida et al., 2019)
Endonuclease III (NTH)	*AtNTH1* (AT2G31450) *AtNTH2* (AT1G05900)	HhH Superfamily	-	AtNTH1 and AtNTH2 are structural and functional homologues of *E. coli* endonuclease III and protects plant against oxidative damages. AtNTH1 and AtNTH2 appear to be targeted to the chloroplast.	No structure known. 32.24% (AtNTH1) and 30.85% (AtNTH2) identity with *E. coli* endonuclease III (PDB: 2ABK)	(Gutman and Niyogi, 2009) (Roldan-Arjona et al., 2000)
Demeter (DME)	*AtDME* (AT5G04560)	DML family of HhH Superfamily	Loss-of-function mutations in *AtDME* cause seed abortion. Knocking down *AtDME* expression in a triple mutant background (*ros1 dml2 dml3*) has enhanced disease susceptibility to *Fusarium oxysporum* infection.	Excises 5-methylcytosine (5-meC) from DNA. DME shows a preference for 5-meC over thymine in CpG sequence context. Regulates genomic imprinting through its interaction with histone H1. Active DNA demethylation by DME also requires HMG domain containing SSRP1. DME is required for endosperm gene imprinting and seed viability. DME demethylates similar kind of genes in both vegetative and central cells in male and female gametophytes.	No structure known. 30.86% identity with *Deinococcus radiodurans* endonuclease III (PDB: 1ORN)	(Rea et al., 2012) (Morales-Ruiz et al., 2006) (Schumann et al., 2019) (Ikeda et al., 2011) (Choi et al., 2002) (Gehring et al., 2006) (Ohr et al., 2007) (Schoft et al., 2011)

(Continued)

TABLE 3 | Continued

Name of the protein	Name of the gene[#]	Protein family	Mutation/ overexpression studies	Miscellaneous information	Structural information from plants*	References
Repressor of silencing 1 (ROS1)	*AtROS1* (AT2G36490)	DML family of HhH Superfamily	*AtROS1* overexpression increases demethylation of promoter and coding regions of the genes involved in flavonoid biosynthesis and antioxidant pathways under salt stress. In *AtROS1* mutant plants many of the CpXpG and CpXpX sites become heavily methylated. ROS1 mutation causes transcriptional silencing of many specific genes.	ROS1 shows a preference for 5-meC over thymine in CpG sequence context. ROS1 preferentially targets transposable elements and intergenic regions. ROS1 positively regulates stress responsive genes. ROS1 regulates seed dormancy by negatively regulating DOGL4. ROS1 has been implicated in immune responsiveness of Arabidopsis. ROS1 antagonizes RNA dependent DNA methylation. Expression of ROS1 is promoted by DNA methylation and antagonizes by DNA demethylation. ROS1 is a slow-turnover enzyme, a feature that helps avoiding generation of double stranded breaks. ROS1 interacts with RPA2 during DNA repair. DNA damage binding protein (DDB2) is a critical regulator of ROS1 activity.	No structure known. 25.27% identity with *Deinococcus radiodurans* endonuclease III (PDB: 1ORN)	(Morales-Ruiz et al., 2006) (Tang et al., 2016) (Zhu et al., 2018) (Bharti et al., 2015) (Williams et al., 2015) (Ponferrada-Marin et al., 2009) (Zhu et al., 2007) (Agius et al., 2006) (Kapoor et al., 2005b) (Gong et al., 2002) (Zhu et al., 2018) (Lopez Sanchez et al., 2016) (Yamamuro et al., 2014) (Li et al., 2012) (Xia et al., 2006) (Kapoor et al., 2005a) (Cordoba-Canero et al., 2017) (Le et al., 2014)
DME like (DML)	*DML2* (AT3G10010) *DML3* (AT4G34060)	DML family of HhH Superfamily	Mutations in *DML2* and/or *DML3* lead to hypermethylation of cytosines that are normally unmethylated or weakly methylated, and hypomethylation of cytosines that are generally hypermethylated.	DML2 and DML3 are 5-meC DNA glycosylases. DML2 and DML3 play a role in maintaining methylation marks. DML2 and DML3 along with ROS1 play a role in resistance against fungal diseases.	No structure known.	(Ortega-Galisteo et al., 2008) (Le et al., 2014) (Brooks et al., 2014)
Formamido-pyrimidine DNA glycosylase (FPG)	*AtFPG1* (AT1G52500)	H2TH superfamily	Mutation in *AtFPG1* is compensated by *AtOGG1* and *vice versa*. A double mutant result in an increase in oxidative damage.	AtFPG1 excises oxidatively modified purines. Presence of a protein segment (213-229) enables AtFPG1 to process 8-oxoG in addition to other oxidative modifications. *AtFPG1* produces two transcripts arising from alternate splicing and encoding two proteins: AtFPG-1 and AtFPG-2. The two proteins exhibit differences in cleaving double stranded oligonucleotide containing 8-oxoG.	AtFPG1 (PDB: 3TWL, 3TWK)	(Duclos et al., 2012) (Gao and Murphy, 2001) (Murphy and Gao, 2001) (Cordoba-Canero et al., 2014)

(Continued)

TABLE 3 | Continued

Name of the protein	Name of the gene[#]	Protein family	Mutation/ overexpression studies	Miscellaneous information	Structural information from plants*	References
Structure Specific Endonucleases (Potential use in assisting site-directed nucleases)						
FEN1	*SAV6* (AT5G26680) *OsFEN-1a* (OSNPB_050540100) *OsFEN-1b* (OSNPB_030834000)	Rad2/XPG	FEN1 mutation causes hypersensitivity to methyl methanesulfonate, UV-C and reduced telomere length. No effect of FEN1 mutation on chemicals that block DNA replication.	FEN1 removes 5′ flaps and important for maintaining genome stability. FEN1 is abundant in tissues rich in proliferating cells. FEN1 in plants either lacks or exhibits a weak exonuclease activity. Rice possess two FEN1 homologs: OsFEN1a and OsFEN1b. OsFEN1 physically interacts with OsPCNA.	No structure known. 54.69% identity with human FEN1 (PDB: 1UL1)	(Zhang J. et al., 2016) (Zhang Y. et al., 2016) (Kimura et al., 2003) (Kimura et al., 2001) (Kimura et al., 2000) (Kimura et al., 1997)
GEN1	*AtGEN1* (AT1G01880) *OsGEN-L* (OSNPB_090521900)	Rad2/XPG	Silencing of *OsGEN-L* results in low fertility, male-sterility, absence of mature pollens. Loss of function also results in persistent double strand breaks resulting in programmed cell death of the male gametes.	Like other eukaryotic GEN1, plant GEN1 is also a canonical Holliday junction resolvase. There are indications of sequence specificity exhibited by AtGEN1. *OsGEN-L* plays a role in early microspore development in rice.	No structure known. 31.35% identity with human GEN1 (PDB: 5T9J)	(Moritoh et al., 2005) (Wang et al., 2017) (Bauknecht and Kobbe, 2014)
SEND1	*AtSEND1* (AT3G48900) *OsSEND1* (OSNPB_080101600)	Rad2/XPG	Depletion of *OsSEND1* has no effect on plant development. *OsSEND1* depletion does not enhance the defect caused by *OsGEN1* depletion. Combined absence of *AtMUS81* and *AtSEND1* results in developmental defects, spontaneous cell death, and genome instability.	Like other eukaryotic GEN1, plant SEND1 is also a canonical Holliday junction resolvase. There are indications of sequence specificity exhibited by AtSEND1. AtSEND1 is as an essential backup to MUS81.	No structure known. 28.71% identity with human GEN1 (PDB: 5T9J)	(Wang et al., 2017) (Bauknecht and Kobbe, 2014) (Olivier et al., 2016) (Furukawa et al., 2003)
MUS81	*AtMUS81* (AT4G30870) *OsMUS81* (OSNPB_010948100)	XPF	Mutation in *AtMUS81* shows normal growth and no meiotic impairment because of the presence of alternative parallel pathways. Disruption of *AtMUS81* increases sensitivity to DNA damaging agents and moderate decrease in meiotic recombination. Mutation of *AtMUS81* in combination with mutation in *AtRecQ4A* results in synthetic lethality. This synthetic lethality can be suppressed by disrupting *RAD51C*.	MUS81 is the catalytic partner in MUS81-EME1 complex. MUS81-EME1 plays an important role in meiotic DNA damage and meiotic recombination in addition to somatic DNA repair and recombination. It is also an important player in interstrand cross-link repair. *AtMUS81* transcript is present in all tissues, but almost 9-fold higher concentration in anthers. *OsMUS81* gene produces two alternative transcripts.	No structure known. 32.12% identity with human MUS81 (PDB: 4P0P)	(Kurzbauer et al., 2018) (Higgins et al., 2008) (Berchowitz et al., 2007) (Hartung et al., 2006) (Mannuss et al., 2010) (Mimida et al., 2007)

(Continued)

TABLE 3 | Continued

Name of the protein	Name of the gene[#]	Protein family	Mutation/ overexpression studies	Miscellaneous information	Structural information from plants*	References
EME1	*AtEME1A* (AT2G21800) *AtEME1B* (AT2G22140) *OsEME1* (OSNPB_040648700)	XPF	-	EME1 is the non-catalytic partner in MUS81-EME1 complex. MUS8-EME1 plays an important role in meiotic DNA damage and meiotic recombination in addition to somatic DNA repair and recombination. It is also an important player in interstrand cross-link repair. Arabidopsis has two EME1 homologs: AtEME1A and AtEME1B. Both the homologs form distinct complexes with MUS81: MUS81-EME1A and MUS81-EME1B. Both the complexes have same cleavage patterns on DNA substrates with slight differences in processing intact Holliday junctions.	No structure known.	(Kurzbauer et al., 2018) (Higgins et al., 2008) (Berchowitz et al., 2007) (Hartung et al., 2006) (Mannuss et al., 2010) (Mimida et al., 2007) (Geuting et al., 2009)
SLX1	*T9D9.16* (AT2G30350)	GIY-YIG	-	Also known as HIGLE. Interacts with HYL1 and Serrate proteins.	No structure known. 22.61% identity with *Candida glabrata* SLX1 (PDB: 4XLG)	(Cho et al., 2017)
Holliday Junction dissolution (potential use in manipulating the frequency of HR)						
RecQ helicase	*AtRECQ4A* (AtRECQ4A) *AtRECQ4B* (AT1G60930) *OsRECQL4* (OSNPB_040433800)	Superfamily II DNA helicases	*OsRECQL4* and *AtRECQ4A* mutant plants are hypersensitive to DNA damaging agents and exhibit high HR frequency. *AtRECQ4B* mutant plants are not mutagen sensitive but can impair HR (antagonistic to *AtRECQ4A)*. A double mutant for *AtRecQ4A* and *AtMUS81* is lethal. Mutation of *AtRecQ4A* but not *AtRecQ4B* can suppress the lethal phenotype of *AtTop3*α mutant.	AtRecQ4A plays a role in removing inter-chromosomal telomeric connections during meiotic recombination.	No structure known. 40.83% (AtRECQ4A) and 38.64% (AtRECQ4B) identity with human ATP dependent helicase Q1 (PDB: 4U7D)	(Kwon et al., 2013) (Hartung et al., 2007a) (Hartung et al., 2006) (Higgins et al., 2011)
TOP3α	*AtTOP3alpha* (AT5G63920)	Type IA topoiso-merase	Disruption of *AtTOP3alpha* results in fragmented chromosomes and sensitivity to camptothecin. Mutation of *AtRecQ4A* can suppress the lethal phenotype of *AtTop3*α mutant.	Essential for meiotic recombination. A defect in TOP3α results in sterile flowers.	No structure known. 43.86% identity with human Topoisomerase3α (PDB: 4CGY)	(Hartung et al., 2008) (Seguela-Arnaud et al., 2015)

(Continued)

TABLE 3 | Continued

Name of the protein	Name of the gene[#]	Protein family	Mutation/ overexpression studies	Miscellaneous information	Structural information from plants*	References
RMI1	*AtRMI1* (AT5G63540)	OB-fold proteins	Disruption of *AtRMI1* exhibit phenotype like the disruption of *AtTOP3alpha.*	RMI1 is a structural protein. RMI1 is essential of meiotic recombination in plants. A role in DNA cross-link repair has been implicated.	No structure known. 40.00% identity with human RMI1 (PDB: 4CGY)	(Hartung et al., 2008) (Bonnet et al., 2013) (Chelysheva et al., 2008)
Translesion DNA synthesis (Potential use in increasing UV tolerance)						
Polη	*AtPOLH* (AT5G44740)	Y- family DNA polymerase	*AtPOLH* overexpression increases UV resistance. Disruption of *AtPOLH* increases mutation frequency.	Alternative splicing was detected. Expresses ubiquitously in plants.	No structure known. 41.20% identity with human Polη (PDB: 4O3S).	(Nakagawa et al., 2011) (Jesus Santiago et al., 2008) (Santiago et al., 2009) (Santiago et al., 2008) (Santiago et al., 2006)
Polκ	*AtPOLK* (AT1G49980)	Y- family DNA polymerase	-	Deletion of 193 amino acids from C-terminal markedly increases the activity and processivity of Polκ. Highly expressed in variety of tissues.	No structure known. 52.94% identity with human Polκ (PDB: 2OH2)	(Garcia-Ortiz et al., 2007) (Garcia-Ortiz et al., 2004)
Rev3	*AtREV3* (AT1G67500)	B- family DNA polymerase	Disruption of *AtREV3* decreases mutation frequency Disruption of *AtREV3* causes hypersensitivity to UV-B and gamma rays.	Catalytic subunit of Polζ. Rev3 cooperates with MUS81 in response to interstrand crosslinks and alkylated bases.	No structure known. 28.30% identity with human Polδ (PDB: 6S1M)	(Nakagawa et al., 2011) (Kobbe et al., 2015) (Sakamoto et al., 2003)
Rev7	*AtREV7* (AT1G16590)	HORMA (Hop1, Rev7 and Mad2) family	*AtREV7* disrupted mutant is sensitive to DNA cross-linker but no sensitivity toward UV-B and gamma rays.	Non-catalytic subunit of Polζ	No structure known. 33.33% identity with human MAD2B (PDB: 6BI7)	(Takahashi et al., 2005)
Rev1	*AtREV1* (AT5G44750)	Y- family DNA polymerase	Disruption of *AtREV1* decreases mutation frequency. *AtREV1* disrupted mutant is moderately sensitive to UV-B, gamma rays, and DNA cross-linkers. Plants with moderate over-expression of *AtREV1* could be obtained indicating toxic nature of Rev1 at high levels.	Alternative splicing was detected. Expresses ubiquitously in plants. AtRev1 is a deoxycytidyl transferase.	No structure known. 31.23% identity with human REV1 (PDB: 3GQC)	(Santiago et al., 2009) (Nakagawa et al., 2011) (Santiago et al., 2008) (Takahashi et al., 2005) (Jesus Santiago et al., 2008) (Takahashi et al., 2007)
Polι	Not identified	–	–	–	–	–

[#]*Values in parentheses are locus tags. Only genes from Arabidopsis thaliana (At) and Oryza sativa (Os) are included. *Closest protein for which structure is available. Percentage sequence identity refers to sequence identity between A. thaliana proteins as query sequence and the non-plant proteins for which crystal structures are available.*

spontaneous mutations (reactive oxygen species, replication errors, transposable elements, ionizing radiation, *etc.*), therefore contributing diversity in the already existing genetic pool. However, the process is too slow to keep pace with ever-increasing demand. Meanwhile, many present-day crop plants have lost genetic diversity, compelling the intervention of crop improvement methods (Esquinas-Alcazar, 2005). Different crop improvement approaches are in use. These methods broadly fall into three categories: (1) chemical and physical mutagenesis, (2) transgenics, and (3) genome editing (**Table 4**). In the past few decades, these approaches have been used successfully for improving various traits of economically important plants: resistance to biotic and abiotic stresses, seed quality, crop yield, *etc.* and resulted in many fruitful outcomes. Intriguingly, DRR takes the central stage in all the crop improvement techniques.

Chemical and Physical Mutagenesis

One of the earliest methods for generating genetic diversity is inducing random mutations through ionizing radiations (*e.g.*, X-ray, γ-rays, neutron, and high-energy ion beams) or chemical mutagens (*e.g.*, alkylating agents, dyes, nitrous acid, *etc.*) (Shu et al., 2012; Holme et al., 2019). These methods result in random double-strand breaks, single-strand breaks, or base modifications, ensuing repair through specialized DRR mechanisms (**Tables 1**, **4**). DRR mechanism thus activated dictates the outcome of mutagenesis: substitution or deletion (Oladosu et al., 2016). The advantage of this method is the cost-effectiveness, no need for the prior knowledge of gene function or sequence, and the technique is beyond the purview of GMO regulations. This method generates random mutations. Therefore, obtaining a mutation of interest is governed by chance events, pausing a significant limitation in terms of larger population size for mutagenesis and a robust screening methodology. The generation of chimeras is another limitation of this methodology in the case of vegetatively propagated plants (Geier, 2012). *In vitro* mutagenesis (mutation induced by treating an explant) and high-throughput mutation screening techniques (DNA molecular markers, TILLING: Targeting Induced Local Lesion In Genomes, HRM: High Resolution Melting, EMAIL: Endonucleolytic Mutation Analysis by Internal Labeling, *etc.*) overcame these limitations (Parry et al., 2009; Chen et al., 2014; Oladosu et al., 2016; Simko, 2016). This method has resulted in crops with various improved traits, *e.g.*, enhanced nutritional traits of soybean (Espina et al., 2018), agronomic traits of rice (Wu J.L. et al., 2005), biotic resistance in wheat (Hussain et al., 2018), *Medicago truncatula* seed size improvement (Ge et al., 2016), *etc*.

Transgenics (Plant Transformation)

Transgenics involves introducing a foreign gene of a known function into the genome of a plant cell followed by the selection of transformed cells and, eventually, regenerating an entire transgenic plant. Transgenics, therefore, results in a genetically modified plant (Newell, 2000). Many transgenic plants have been created, since the mid-1980s (Caplan et al., 1983), using two techniques: *Agrobacterium tumefaciens* mediated transformation (Hwang et al., 2017) or gene gun (microprojectile bombardment) mediated transformation (Klein et al., 1992). Unlike chemical or physical mutagenesis, transgenics result in dominant traits. This method's main advantage is the flexibility of introducing a gene of known function into a host plant. The gene often gets integrated randomly with the host genome either through microhomologies-mediated double-strand break repair or single-strand gap repair (Tzfira et al., 2004). Despite unlimited applications, strict GMO regulations due to environmental (invasiveness, intraspecific, and interspecific hybridization) and health (food toxicity and allergenicity) biosafety concerns (Stewart et al., 2000), are the significant limitations. Developing a genetically modified crop is therefore expensive as every GM crop must be assessed for environmental and health biosafety (Giraldo et al., 2019). Transgenics resulted in crops with improved traits, *e.g.*, nutritional value (Broun et al., 1999; Hirschberg, 1999), tolerance to various abiotic stresses (Holmberg and Bulow, 1998; Koh et al., 2007), herbicide resistance (Anthony et al., 1999), insect resistance (Dempsey et al., 1998), modified flower color (Mol et al., 1999), *etc*. Transgenics also has immense pharmaceutical potential in generating human therapeutic proteins in plants (Rodgers et al., 1999; Staub et al., 2000).

Genome Editing (Targeted Mutagenesis)

Genome editing overcomes the disadvantages of random mutagenesis by physical and chemical mutagens. Since the changes introduced by genome editing mimic the natural changes, it is more acceptable than transgenics. Therefore, genome editing is the most promising technology for crop improvement in the current scenario (Abdallah et al., 2015). Genome editing relies upon tailor-made site-directed nucleases (SDN): ZFNs, TALENs, and CRISPR/Cas system. These nucleases make precise nicks in a sequence-specific manner resulting in DSBs. The outcome of genome editing relies upon the mode of DSB repair, which can either proceed through NHEJ or HR. NHEJ can result in the insertion or deletion of nucleotides (SDN1), while HR can facilitate the exchange of DNA sequence with an exogenously provided donor DNA containing the desired sequence/mutation (SDN2 and SDN3). The primary limitations of genome editing are off-target cleavages, a prerequisite knowledge of genomic information, and an efficient delivery method (Podevin et al., 2013). Genome editing successfully generated herbicide tolerance in maize (Shukla et al., 2009) and tobacco (Townsend et al., 2009), improved quality of soybean oil (Haun et al., 2014), stress tolerance in maize (Shi et al., 2017), improved yield traits in tomato (Soyk et al., 2017), *etc*. While most genome editing proceeds through NHEJ resulting in gene knockouts, genome editing's true potential in the future relies on HR to generate traits that are difficult to achieve through conventional breeding (Abdallah et al., 2015).

Even though DRR is central to all crop improvement techniques, more studies on plant DRR mechanisms are important to innovate the crop improvement methodologies. Studying various DRR proteins from at least five different classes can greatly contribute toward developing new approaches for crop improvement: photolyases, DNA glycosylases, RTR complex, structure-specific endonucleases, and TLS DNA

TABLE 4 | Comparison of various crop improvement techniques.

	Chemical/Physical mutagen	Transgenics	Genome editing
Mechanism	Use of chemical mutagens or ionizing radiations to induce random mutations.	*Agrobacterium tumefaciens* mediated transformation or gene gun (microprojectile bombardment) to introduce a foreign gene.	Use of site-directed nucleases (SDNs) to introduce DSBs
Outcome	Random mutations, mostly recessive in nature.	Introduction of a foreign gene through random insertion. Mostly resulting in a dominant trait.	Site-specific mutations through insertions or deletions. The introduced changes could be loss of function or gain of function.
DRR pathways involved	Base excision repair Nucleotide excision repair Translesion DNA synthesis Mismatch repair Non-homologous end joining Homologous recombination	Double-strand break repair Single-strand gap repair	Non-homologous end joining (SDN1) Homologous recombination (SDN2 and SDN3)
Advantages	Easiest method of introducing random mutations. Excluded from GMO regulations.	Introduces novel traits. A tool to study gene functions (loss of function/gain of function)	Site-directed. Partially exempted from strict GMO regulations.
Limitations	Obtaining a mutation of interest is dictated by chance event. Large population size is required for inducing mutation. Robust screening methodologies are required.	Environment and health safety concerns. Strong GMO regulations.	Tailor-made site-specific nuclease required. Generation of off-site cleavages.
Screening Methods	DNA markers, TILLING, HRM	PCR based detection methods, phenotypic characterization, DNA sequencing, ELISA	PCR based methods, Sanger sequencing, ELISA, MALDI-TOF, DNA Microarray, NMR
Scope of improvement	Mutations in mismatch repair pathway or overexpression of certain translesion DNA polymerase can reduce the need of using chemical or physical mutagens to introduce random mutations.	The efficiency of gene targeting can be increased by manipulating the efficacy of Homologous recombination.	Since the outcome of this technique rely on the mechanisms of DSB repair, a better control over NHEJ and HJ could open enormous application possibilities.

polymerases (**Table 3**). The knowledge of DRR can contribute toward crop improvement in four possible ways: (1) targeting DRR genes directly, (2) manipulation of HR frequency, (3) modification of gene-editing techniques, and (4) Computational systems biology and precision agriculture.

Targeting DRR Genes Directly

Many of the DRR genes are known to be associated with biotic and abiotic stresses, and merely targeting these genes by altering their expression or protein structure might help generating improved crop plants (**Table 3**). Photolyases and many TLS DNA polymerases are associated with UV tolerance. Structure-based studies highlighted the importance of single amino acid substitution in stabilizing the overall structure of photolyase and subsequently affecting the UV sensitivity of the plant (Landry et al., 1997; Hidema et al., 2000; Yamamoto et al., 2007; Hitomi et al., 2012). Overexpression of CPD photolyase in *A. thaliana* resulted in increased DNA repair and enhanced UV tolerance (Kaiser et al., 2009). Altering the expression of TLS polymerases in *A. thaliana* showed an influence on UV tolerance. An overexpression of *AtPOLH* (coding Polη) increases the UV resistance (Jesus Santiago et al., 2008; Nakagawa et al., 2011), disruption of *AtREV3* (coding the catalytic subunit of Polζ) causes hypersensitivity to UV-B and γ-rays (Sakamoto et al., 2003), and disruption in *AtREV1* (coding REV1) moderately increases the sensitivity to UV-B and γ-rays (Jesus Santiago et al., 2008; Nakagawa et al., 2011; **Table 3**). Structural and biochemical characterization of TLS polymerases, similar to photolyases, are essential to understand the structural basis of repair of photoproducts and UV tolerance among plants in order to target these proteins for crop improvement. However, overexpressing TLS polymerase has its limitations since these polymerases are highly error-prone, and an overexpression might result in an increase in mutation frequency (Jesus Santiago et al., 2008; Nakagawa et al., 2011). Many DNA glycosylases play an essential role in tolerating oxidative stress. Overexpression of *AtOGG1* (codes for AtOGG1 DNA glycosylase) improves seed longevity (Chen et al., 2012), and overexpression of *AtMBD4* (codes for MBD4L DNA glycosylases) enhances tolerance to oxidative stress (Nota et al., 2015). Overexpression of *AtROS1* (codes for ROS1) activates the expression of genes coding for antioxidant pathways under salt stress (Bharti et al., 2015). DNA glycosylases belonging to the DEMETER family are critical for resistance against fungal diseases (Le et al., 2014; Schumann et al., 2019). The studies carried out in Arabidopsis provide clues about altering the expression of DNA glycosylases to improve tolerance against biotic and abiotic stresses. Similar studies need to be carried out in the crop plants to utilize the full potential of DNA glycosylases in crop improvement against various biotic and abiotic stresses.

Manipulation of Homologous Recombination Frequency

Homologous recombination is the fundamental driving force behind generating diversity and new allelic combinations. An in-depth understanding of HR and its crosstalk with other DRR mechanisms can provide a robust tool for controlling HR events as and when required during a crop improvement program. Besides, it will also provide tools to facilitate homeologous recombination between divergent sequences. Any HR event starts with the generation of DSBs. During meiosis generation of DSBs is programmed and mediated by SPO11 proteins (Grelon et al., 2001; Hartung et al., 2007b). HR and NHEJ are the two competing pathways involved in DSB repairs. In fact, in plants, NHEJ (more error-prone) is the preferred DSB repair mechanism (Puchta, 2005). Two competing processes further reduce the generation of crossover species: HJ dissolution by RTR complex and HJ resolution by structure-specific endonucleases (Knoll et al., 2014). MMR also suppresses the frequency of HR (Tam et al., 2011). Furthermore, HR is developmentally regulated in plants (Boyko et al., 2006). Therefore, tinkering with NHEJ, processing of HJs, and MMR can increase the overall frequency of HR. Suppression of Ku70/Ku80 or LIG4 resulted in enhanced HR (Nishizawa-Yokoi et al., 2012). *AtRECQ4A* and *OsRECQL4* knockouts result in high HR frequency in Arabidopsis and rice, respectively (Hartung et al., 2007a; Kwon et al., 2013). Targeting MMR in plants is another very promising avenue to increase the frequency of HR and to enable homeologous recombination among related species. The loss of key MMR proteins (MSH2, MSH7, and PMS1) in Arabidopsis and tomato, increased homeologous recombination between divergent sequences (Emmanuel et al., 2006; Li et al., 2006, 2009; Tam et al., 2011). Besides being antagonistic to HR, MMR is a crucial player in maintaining genomic integrity. Arabidopsis with defective MMR proteins show a significant increase in the number of single-nucleotide variants in the gene. Therefore, MMR mutant plants could be used to introduce random mutations, thus replacing the need for chemical and physical mutagens (Hoffman et al., 2004; Van Marcke and Angenon, 2013; Belfield et al., 2018). Not all regions of a chromosome undergo HR with equal frequency. HR hotspots are mostly concentrated in euchromatin compared to heterochromatin (also comprises of functional genes). Therefore, DNA demethylation could be an approach to promote HR in heterochromatin regions. MET1 (Methyltransferase 1) and DDM1 (Decrease in DNA methylation 1) are involved in CG methylation maintenance. A loss of MET1 or DDM1 can restructure the distribution of crossing over hot spots along the chromosomes (Melamed-Bessudo and Levy, 2012; Yelina et al., 2012, 2015; Choi, 2017). The main limitation of any of these approaches will be to restore the silenced pathways since they are essential in maintaining genome stability.

Modification of Genome Editing Techniques

Site-directed nucleases (SDN) revolutionized the field of genome editing. There are three different site-directed nuclease techniques: SDN1, SDN2, and SDN3. SDN1 method relies upon error-prone repair of DSBs introduced by SDNs through NHEJ, resulting in the deletion or addition of nucleotides. SDN2 and SDN3, in contrast to SDN1, rely on the repair of DSBs through HR. In the case of SDN2, a donor DNA (externally supplied DNA template) carrying a sequence of interest is used to facilitate HR resulting in gene modification at a predetermined site. In the case of SDN3, the donor DNA often contains an insert as big as a transgene (Podevin et al., 2013). The main limitation with SDN2 and SDN3 is low HR frequencies (Pattanayak et al., 2011; Van Vu et al., 2019). Overexpressing proteins facilitating HR can increase the HR frequency. Expressing prokaryotic proteins: RecA (a homolog of eukaryotic RAD51) (Reiss et al., 1996) and RuvC [structure-specific endonuclease (SSE)] increased the HR frequency in plants (Shalev et al., 1999). Interestingly, plants have many endogenous SSEs (GEN1, SEND1, MUS81-EME1, and SLX1), the over-expression of which could be used along with SDNs to increase the efficacy of exchange of gene segments with the donor sequence. An increased HR activity will also facilitate the repair of off-target cleavages (another limitation of SDNs) through HR instead of NHEJ; therefore, reducing the toxicity associated with SDNs. Even though the use of SSEs with SDNs appears to be an attractive approach to direct gene modification/gene incorporation at a specific site, there are limitations of using SSEs. Studies from prokaryotes, yeast, and humans indicate that SSEs can be harmful if not regulated because of their potential to cleave genomic DNA indiscriminately, resulting in genotoxicity (Minocherhomji and Hickson, 2014; Dehe and Gaillard, 2017). SSEs from plants are poorly characterized in terms of their regulation of the activity and substrate specificities, therefore, pausing a major hurdle in using them at the present moment to assist SDNs. Structural and biochemical characterization of plant-specific SSEs is indispensable in fully understanding the molecular basis of HJ resolution, substrate specificities, mapping of cleavage patterns, and regulation of the catalytic activity before they could be implemented in any crop improvement technique.

Computational Systems Biology and Precision Agriculture

The past decade has witnessed a surge of data science techniques involving a strong component of digital inferences for precision agriculture, that take into account the whole system instead of individual genes or proteins. One of the essential tasks of systems biology is to create explanatory and predictive models of complex systems encompassing important physiological processes. The DRRs offer an untapped opportunity to explore the extent to which known genes and complexes can be used to predict the occurrence, distribution, regulation, and evolution of this machinery across wild and cultivated crop varieties. About 92 fully sequenced and annotated plant genomes are currently available in the Phytozome (Goodstein et al., 2012), in addition to over one thousand vegetative transcriptomes in the public domain (One Thousand Plant Transcriptomes, 2019), as well as an exponentially increasing time-resolved and condition-specific gene expression datasets (both microarrays and RNA-Seq). The next-generation sequencing (NGS) and other high

throughput (HTP) experimental datasets have paved the way for reconstructing direct or indirect regulatory connections between various genes and gene products. Transcriptional regulatory inferences from genomic datasets of the known DRR genes across the plant kingdom will enable identification of novel genes and regulators in DRR pathways and the reconstruction of gene regulatory networks that can provide insights into the biological process of DNA repair and recognition.

DNA REPAIR AND RECOMBINATION PROTEINS AS POTENTIAL TOOLS IN CROP IMPROVEMENT

An understanding of various proteins participating in different pathways of DRR has potential to contribute significantly to crop improvement by targeting endogenous plant proteins. In some cases, this information has already translated to model crop plants. This section provides an update regarding the available information on selected proteins participating in DRR and that have the potential for crop improvement (**Table 3**).

Photolyase

Photolyase catalyzes the light-dependent direct reversal of the UV induced lesions through photoreactivation. Photolyases are flavoproteins (FAD cofactor) belonging to the photolyase (PHR)/Cryptochrome (CRY) family (Ozturk, 2017). Members of this family are present in bacteria to humans and perform diverse functions. PHR proteins participate in DNA repair, whereas CRY proteins regulate plant development (Liu et al., 2011), and associated with biological rhythms in both plants and animals (Thresher et al., 1998; Vitaterna et al., 1999; Cashmore, 2003). Photolyase based DNA repair is, however, absent in placental mammals (Essen and Klar, 2006). Photolyases belong to two classes: CPD photolyases (substrate: CPDs) and (6-4) photolyases (substrate: (6-4) photoproducts). Structurally, CPD photolyases belong to two subclasses: Class I and Class II CPD photolyases. Plants have two types of photolyases: Class II CPD photolyase and (6-4) photolyase. Interestingly, (6-4) photolyase are closer to Class I CPD photolyase irrespective of different substrate specificities. On the other hand, Class I and Class II CPD photolyases provide the case of convergent evolution (Hitomi et al., 2012).

Photolyases are by far the most well-studied and structurally characterized DRR proteins in plants. The structures of proteins from all three classes from diverse organisms are available. In general, the overall structure of photolyases shows conservation across all three classes. All photolyases have an N-terminal α/β domain and a C-terminal FAD-binding helical domain. A long linker connects both domains. FAD binds in an unusual U-shaped conformation bringing isoalloxazine and adenine rings in proximity (Park et al., 1995; Hitomi et al., 2009, 2012; **Figure 2**). A Trp electron transfer pathway restores PHR activity by reduction of FAD in all the three classes. Substrate binding requires flipping of the photoproduct into the photoproduct binding cavity. Amino acid residues from PHR further stabilize the complementary undamaged strand. Although Class I CPD photolyases and (6-4) photolyases are evolutionarily related, there are significant differences in the size of photoproduct binding site (active site of (6-4) photolyase being narrower), mode of catalysis [(6-4) photolyase requires two His for catalysis], and second cofactor binding site (Hitomi et al., 2009). Class II CPD photolyase has low sequence similarity with class I CPD photolyase, but it still adopts the same overall fold. Marked differences are present in the C-terminal substrate-binding region and the mode in which proteins from the two classes interact with the DNA (Hitomi et al., 2012). Class II CPD photolyase structure provides clues regarding how a single amino acid substitution can improve the UV resistance of certain strains of rice varieties over others. The more UV tolerant rice cultivars have Gln296 in class II CPD photolyase, while less tolerant cultivars have His at this position. Gln indeed helps stabilize the overall structure of the protein (Hitomi et al., 2012). Norin 1 is one of the most widely cultivated varieties of rice is UV sensitive because of single Gln to Arg change at position 126, resulting in a less stable photolyase (Hidema et al., 2000). Both these examples provide classic examples of structure-based understanding to engineer the photolyases for crop improvement.

DNA Glycosylases

DNA glycosylases catalyze the first reaction in BER. There are two types of DNA glycosylases: monofunctional and bifunctional DNA glycosylases. Monofunctional glycosylases cleave only N-glycosidic bonds resulting in an AP site. In contrast, bifunctional glycosylases cleave the N-glycosidic bond and phosphodiester bond (Roldan-Arjona et al., 2019). Bifunctional DNA glycosylases possess an AP lyase activity in addition to the DNA glycosylase activity. Monofunctional DNA glycosylases, in contrast, are devoid of AP lyase activity. Therefore, an AP endonuclease activity follows catalysis by monofunctional DNA glycosylases (Fromme et al., 2004). DNA glycosylases exhibit a wide range of substrate specificities. Almost all eukaryotic DNA glycosylases rely on flipping the damaged or modified base into the active site, followed by the cleavage of the N-glycosidic bond (**Figure 3**). There are five structural superfamilies of plant DNA glycosylases: Uracil DNA glycosylases (UDG), Alkyladenine DNA glycosylase (AAG), helix-hairpin-helix (HhH), and helix-two-turn-helix (H2TH) (reviewed in Roldan-Arjona et al., 2019).

Uracil DNA Glycosylases (UDG)

Uracil DNA glycosylases (UDG) remove Uracil from DNA. UDG superfamily encompasses six subfamilies (Brooks et al., 2013; Roldan-Arjona et al., 2019). Only Family-1 of UDG is present in plants (Talpaert-Borle and Liuzzi, 1982; Warner, 1983; Bensen and Warner, 1987; Bones, 1993). *At*UNG is a UDG isolated and characterized from *A. thaliana* (Cordoba-Canero et al., 2010). Family-1 of UDG, as exemplified by human UDG, comprises a central four-stranded parallel β-sheet, flanked by eight α-helices (**Figure 3A**). These are monofunctional enzymes. Breaking of N-glycosidic bond involves flipping out of the base, followed by the distortion and weakening of the N-glycosidic bond before the cleavage (Parikh et al., 2000). At present, no structural information is available on plant UDG. Plant UDG active site is more specific and selective compared

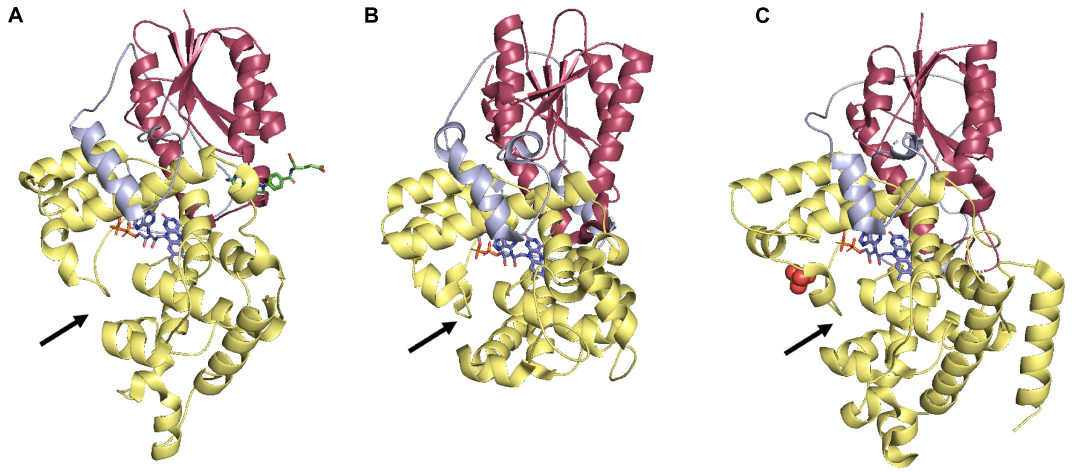


FIGURE 2 | Comparison of photolyases from three different classes. **(A)** Class I CPD photolyase from *E. coli* (PDB: 1DNP), **(B)** Class II CPD photolyase from *O. sativa* (Sasanishiki) (PDB: 3UMV), and **(C)** (6-4) photolyase from *A. thaliana* (PDB: 3FY4). N-terminal α/β domain is shown in Red, C-terminal α-helical domain is shown in Yellow, and interdomain loop is shown in Blue color. The arrows demarcate the substrate binding pocket. FAD molecules are shown in lines in a U-shaped conformation. All the figures were generated using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

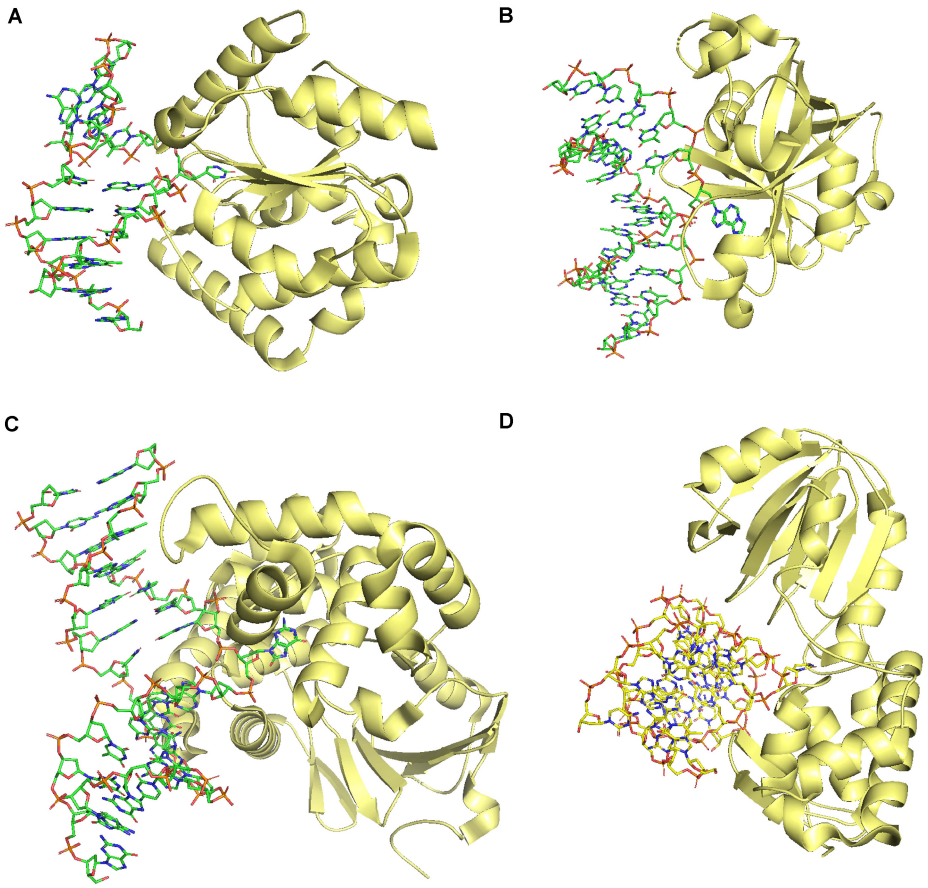


FIGURE 3 | Comparison of DNA glycosylases from four different super families showing a common mechanism of catalysis by base flipping. **(A)** UDG superfamily (PDB: 1EMH), **(B)** AAG superfamily (PDB: 1F4R), **(C)** HhH superfamily (PDB: 2NOI), **(D)** H2TH superfamily (PDB: 1R2Z). All the figures were generated using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

to UDG from other organisms because of the inability of *At*UNG to act upon 5-fluorouracil (5-FU). *Atung* mutant plants do not show any apparent defect; however, inactivation of *At*UNG protects the plant against the cytotoxic effects of 5-FU (Cordoba-Canero et al., 2010).

Alkyladenine DNA Glycosylases

Alkyladenine DNA glycosylases (AAG) or alkylpurine glycosylases are monofunctional enzymes and remove alkylated purines. Alkylated purines may arise because of cellular metabolic processes, mutagenic agents, or environmental mutagens. Alkylated purines may be cytotoxic (3-methyladenine), mutagenic (O^6-methylguanine), or harmless (7-methlylguanine) (Loechler et al., 1984; Larson et al., 1985; Santerre and Britt, 1994). AAG has been extensively characterized biochemically and structurally from humans (hAAG), bacteria, and archaea (Brooks et al., 2013). hAAG cleaves a variety of alkylated bases (adenine and guanine). hAAG is a single domain protein with a core of eight β-strands forming a curved antiparallel β-sheet and a β-hairpin that protrudes into the minor groove of DNA. Flanking loops and α-helices form the remaining DNA binding surface (Lau et al., 1998). The β-hairpin intercalates into the minor groove and flips the modified nucleotide into the active site. A water molecule then makes a nucleophilic attack to cleave the N-glycosidic bond (**Figure 3B**). A combination of the shape, hydrogen-bonding capability, and aromaticity of the modified base dictate the selectivity of AAG towards the damaged nucleotide (Lau et al., 1998, 2000). Incidentally, alkylating agents are the most used mutagens to induce mutations for crop improvement, and still, our understanding regarding the plant AAG is in infancy. 3-methyladenine glycosylase is an AAG from *A. thaliana* (Santerre and Britt, 1994) and requires further structural and biochemical characterization.

HhH Superfamily

Helix-hairpin-helix superfamily comprises of both monofunctional and bifunctional enzymes. They repair a wide range of modified bases resulting from alkylation, oxidation, or spontaneous deamination. The structure consists of N-terminal and C-terminal α-helical domains, connected by a type II β-hairpin (**Figure 3C**). Various homologs of *E. coli* DNA glycosylases acting upon alkylated purines and oxidized bases are present in Arabidopsis (Bjelland et al., 1993; Labahn et al., 1996; Britt, 2002; Roldan-Arjona et al., 2019). Members of this family involved in excising oxidized bases have also been identified and characterized from Arabidopsis. A bifunctional OGG1 protein (8-oxyguanine DNA glycosylase) from Arabidopsis and *Medicago truncatula* acts upon 7-hydro-8-oxoguanine (8-oxoG) (Roldan-Arjona et al., 2000; Dany and Tissier, 2001; Macovei et al., 2011). A monofunctional MBD4-like glycosylase (*At*MBD4L) from Arabidopsis excises uracil or thymine mispaired to guanine (Ramiro-Merina et al., 2013). Further characterization on these glycosylases is required to understand the substrate specificities, and physiological phenotypes arising from the defects in these proteins.

Plants exclusively contain 5-meC DNA glycosylases. They form a separate family of glycosylases (**DEMETER-LIKE (DML) family**) of HhH superfamily (Zhu, 2009; Brooks et al., 2013; Roldan-Arjona et al., 2019). DML family comprises of proteins like DME (DEMETER) (Choi et al., 2002), Ros1 (Repressor of silencing 1) (Gong et al., 2002), DML2 (DME-like 2), and DML3 (DME-like 3) (Ortega-Galisteo et al., 2008) from *Arabidopsis*. All the members of this family are bifunctional enzymes. They take part in crucial processes like regulation of transcription and inhibition of erroneous gene silencing by demethylating DNA through the process of BER (Choi et al., 2002; Gong et al., 2002; Ortega-Galisteo et al., 2008). At present, no structural information regarding the members of the DML family is available except for sequence analysis and modeling studies (Ponferrada-Marin et al., 2011). The salient features distinguishing members of DML family from other members of HhH superfamily are the presence of a [4Fe-S] cluster, a discontinuous catalytic site, and an additional Lys rich N-terminal domain and a C-terminal domain (Ponferrada-Marin et al., 2011). Structure-based studies of these proteins are essential for understanding substrate specificities and catalytic mechanisms of these proteins.

H2TH Superfamily

Helix-two-turn-helix superfamily members are primarily involved in repairing oxidative damages. *E. coli* Formamidopyrimidine DNA glycosylase (Fpg/ MutM) and Endonuclease VIII (Nei) are the typical members of H2TH superfamily (Roldan-Arjona et al., 2019). Members of this superfamily are bifunctional enzymes. They remove a wide range of modified bases, *e.g.,* 5-hydroxyuracil, 5-hydroxycytosine, dihydrouracil, thymine glycol, *etc.* (Sugahara et al., 2000; Kathe et al., 2009; Zhu et al., 2016). Structurally, members of this family comprise an N-terminal domain and a C-terminal domain containing a Zn finger. A flexible hinge connects the two domains (Sugahara et al., 2000; **Figure 3D**). The catalysis involves tautomerization-dependent recognition, flipping of the base, followed by excision (Zhu et al., 2016). Plants have homologs of *E. coli* Fpg with Arabidopsis having seven different isoforms (Ohtsubo et al., 1998). More structural and biochemical characterization of plant FPG proteins are essential to ascertain their role in repairing oxidative damage. More studies are required to determine the relative contributions of plant FPG proteins and OGG1 in oxidative damage repair (Cordoba-Canero et al., 2014).

Since most of the commonly used mutagens, modify the nitrogen bases, which are subsequently repaired by DNA glycosylases, a structure-based study of these DNA glycosylases becomes of utmost importance for increasing the efficacy of mutagenesis during crop improvement. Structural and biochemical studies will provide mechanistic insights into the repair of bases modified by chemical mutagens, amino acid residues contributing to substrate specificities, and overall stability of the proteins. Such studies will enable engineering DNA glycosylases with variable substrate specificities and catalytic mechanisms, providing more control over mutagenesis. Structure-based rational engineering had shown changing substrate specificity and transforming a DNA glycosylases into a site-specific DNA glycosylase (Kwon et al., 2003). DNA

glycosylases also play a role in epigenesis through demethylation with a strict preference for 5-methylcytosine (5-meC) over thymine in the CpG sequence context (Morales-Ruiz et al., 2006; **Table 3**). Structure-based engineered DNA glycosylases can alter demethylation and therefore has potential in epigenetic studies.

Structure-Specific Endonucleases (SSEs)

Endonucleases have played a significant role in crop improvement. Engineered artificial nucleases (ZFN, TALEN, and CRISPR-Cas) are in use for the last two decades (Zhang et al., 2017, 2018; Xu et al., 2019). All these nucleases introduce DSBs in the target gene, which is then repaired by plant internal DRR machinery. Structure-specific endonucleases (SSEs) are among the key players involved in DSB repair. Unlike sequence-specific endonucleases, SSEs recognize the secondary structure of DNA. During DNA metabolism, repair, and recombination, various joint DNA molecules (e.g., 5′ flap, 3′ flap, replication forks, splayed arm DNA, Holliday junctions, *etc.*) appear as intermediates. SSEs process these joint DNA molecules and restore regular DNA duplexes. While engineered nucleases are in use extensively to generate DNA breaks in a sequence-specific manner, knowledge about the structure of SSEs will enhance understanding regarding DRR in plants. Besides, it will provide an additional toolkit for designing innovative methods for crop improvement in the future. Various SSEs participate in different DRR pathways in plants: FEN1, GEN1, SEND1, MUS81-EME1, and SLX1.

Flap Endonuclease 1 (FEN1)

Flap Endonuclease 1 (FEN1) belongs to the Rad2/XPG family of nucleases. Members of the RAD2 family are involved in DNA replication (FEN1), repair (FEN1, XPG, EXO1), and recombination (GEN1, SEND1)(Lieber, 1997; Furukawa and Shimada, 2009; Tsutakawa et al., 2011). FEN1 has endonuclease as well as 5′-3′ exonuclease activities. FEN1 takes part in the removal of 5′ flap intermediates during long patch BER (Huggins et al., 2002; Asagoshi et al., 2010; Sun et al., 2017) and processing of Okazaki fragments during replication by getting associated with PCNA (Gomes and Burgers, 2000; Liu et al., 2004; Dovrat et al., 2014). FEN1 comprises a catalytic domain, and a flexible PCNA interacting C-terminal domain. The crystal structure of human FEN1 in complex with PCNA provides clues regarding the sliding of FEN1 along with PCNA in an inactive conformation and switching to an active conformation on encountering a single-stranded flap DNA (Sakurai et al., 2005). The crystal structure of the catalytic domain of human FEN1 in complex with branched substrates and structural comparison with unliganded FEN1 provides insights into the structural attributes that allow FEN1 to process branched DNA structures (Kim et al., 2001; Sakurai et al., 2005; Tsutakawa et al., 2011). FEN1 has two separate DNA duplex binding sites on its surface (downstream region and upstream region) (characteristic of Rad2/XPG family), which allows DNA bending and therefore facilitate interaction with either branched or nicked DNA. An H2TH motif interacts specifically with the non-nicked strand (or template strand), thus positioning the DNA substrate in a catalytically favorable orientation. Structural features that confer specificity towards flap substrates are the presence of a 3′ flap pocket and a gateway that allows only a single-stranded DNA to pass through it and enters the catalytic site. Catalysis involves double-base flipping and a 2-metal ion active site (Tsutakawa et al., 2011; **Figure 4A**). The characterization of FEN1 from plants is inadequate. FEN1 from *Brassica oleracea*, *Arabidopsis thaliana*, *Oryza sativa*, and *Triticum vulgare* participate in DNA repair, cell growth, organ formation, maintenance of genome stability, and transcriptional gene silencing (Gallego et al., 2000; Kimura et al., 2000; Przykorska et al., 2004; Zhang J. et al., 2016). Similar to other eukaryotes, PCNA appears to coordinate DNA repair and replication by interacting with FEN1 (Kimura et al., 2001). *Arabidopsis* SAV6 is a FEN1 homolog with an endonuclease activity but lacking an exonuclease activity (Zhang J. et al., 2016). In plants, so far, structural, and biochemical studies have not been done on FEN1. However, crystals of AtPCNA in association with PIP (PCNA interacting protein box motif) from FEN1 have recently been obtained (Kowalska et al., 2020). Structural studies on plant FEN1 will further shed light on the role of FEN1 in coupling DNA repair and replication in plants.

GEN1

GEN1 is an HJ resolvase that works in a pathway parallel to SLX1 and MUS81-EME1 mediated HJ resolution (Wyatt et al., 2013). GEN1 belongs to the Rad2/XPG nuclease family and, therefore, structurally related to FEN1 (Ip et al., 2008). GEN1 can process HJs by introducing two symmetrical nicks across the junction and 5′ flaps by cleaving the single-stranded flap. GEN1 can also process replication forks. Unlike FEN1, GEN1 does not have an exonuclease activity (Rass et al., 2010). Like FEN1, GEN1 is monomeric in solution; however, it dimerizes on binding an HJ to facilitate two symmetrical nicks during HJ resolution (Rass et al., 2010). GEN1 structures reveal features common to the members of the Rad2/XPG family: two metal ion catalysis mechanism, two DNA binding surfaces separated by a wedge allowing bending of DNA, and an H2TH motif involved in binding the non-cleaved strand. Nevertheless, GEN1 has specialized structural features that enable GEN1 to dimerize and process HJs. The helical arches forming the gateway in FEN1 are modified to recognize the central portion of an HJ and provide a surface for GEN1 dimerization (**Figure 4B**). Relaxation of HJ and transition of GEN1 active site from disordered to ordered state appear to regulate the activity of GEN1 (Lee et al., 2015; Liu et al., 2015). A chromodomain, a structural feature exclusive to GEN1, further stabilizes the interaction between GEN1 and HJ. Truncation in chromodomain affects nuclease activity (Lee et al., 2015). Plants have two homologs of GEN1: GEN1 and SEND1 (Single-Strand DNA Endonuclease1) (Furukawa et al., 2003; Moritoh et al., 2005; Bauknecht and Kobbe, 2014). AtGEN1 and AtSEND1 from *A. thaliana* can resolve HJs by two symmetrical nicks; however, both the proteins have distinct substrate specificities guided by the structure and sequence of the substrate (Bauknecht and Kobbe, 2014). OsGEN1 can also process an HJ by two symmetrical nicks without any cooperativity in the two nicking events (unlike other known FEN1) (Yang et al., 2012). AtSEND1 (and not AtGEN1), along with MUS81 from *A. thaliana* is

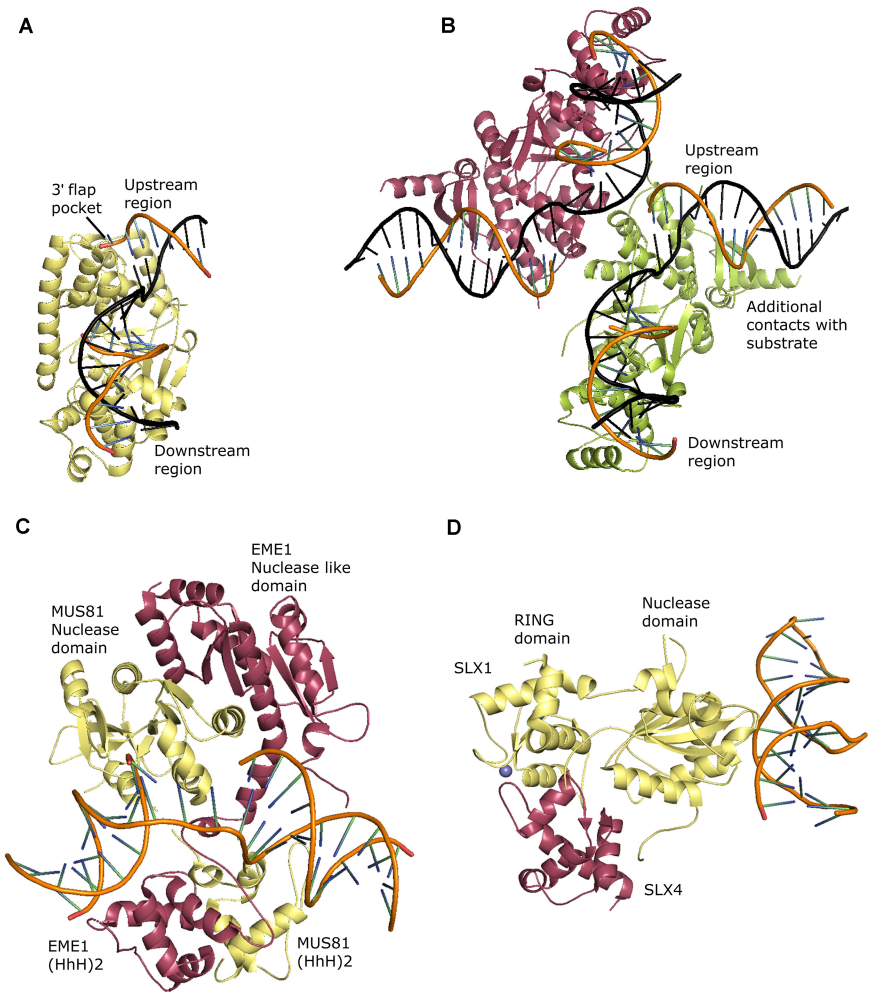


FIGURE 4 | Structure specific endonucleases. **(A)** Crystal structure of human FEN1 in complex with a 5′ flap substrate: non-nicked DNA strand is shown in black color (PDB: 3Q8M), **(B)** Crystal structure of *Chaetomium thermophilum* GEN1 in complex with product DNA: two symmetry related monomers are shown in red and green colors and the non-nicked DNA strand is shown in black color (PDB: 5CO8), **(C)** Crystal structure of human MUS81-EME1 in complex with a 3′ flap substrate: MUS81 and EME1 are shown in yellow and red colors respectively (PDB: 4P0R), **(D)** Crystal structure of *Thermothielavioides terrestris* SLX1-SLX4^{CCD} in complex with a distorted DNA duplex presenting one of the three DNA binding sites of SLX1: SLX1 and SLX4^{CCD} are shown in yellow and red colors respectively, the Zn ions of RING domain are shown in blue color (PDB: 6SEI). All the figures were generated using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

essential for telomere stability (Olivier et al., 2016). Interestingly, in rice, OsGEN1 (and not OsSEND1) plays an essential role in homologous recombination (Wang et al., 2017). More structural insights are imperative to understand HJ resolution by GEN1 and SEND1 from plants and to pinpoint the structural elements responsible for differences in the functions of these two proteins.

MUS81

MUS81 forms a heterodimeric complex with a non-enzymatic partner, EME1 (**Figure 4C**). Both MUS81 and EME1 belong to the XPF family of SSEs (Ciccia et al., 2008). MUS81-EME1 complex plays an important role in resolving HJs, rescuing collapsed replication forks, DSB repair, and interstrand crosslink repair (Boddy et al., 2001; Doe et al., 2002; Hanada et al., 2006). Biochemically, MUS81-EME1 can process nicked HJ (nHJ), 3′ flaps, and replication forks (Fricke et al., 2005). nHJs are the preferred substrates for MUS81-EME1 in comparison to intact HJ (iHJ). MUS81-EME1 coordinates with SLX1-SLX4 complex to resolve HJs by a nick and counter nick mechanism, where SLX1-SLX4 makes the first nick in the iHJ generating an nHJ, which serves as a substrate for MUS81-EME1 (Gaillard et al., 2003; Svendsen et al., 2009; Wyatt et al., 2013). Crystal structure of truncated MUS81-EME1 (N-terminal region removed from both the proteins) complexes from human in unliganded and in complex with DNA substrates provided comprehensive insights into the overall architecture and substrate preferences of MUS81-EME1 (Chang et al., 2008; Gwon et al., 2014). Both MUS81 and EME1 comprises of a nuclease domain and two repeats of the helix-hairpin-helix motif $(HhH)_2$ connected by a linker. The nuclease domain of EME1 (also referred to as nuclease-like domain) is catalytically dead. The domains from both the proteins intertwine with each other. The differences in the

linkers result in a quaternary structure, distinct from the other members of the XPF family (Kim et al., 2008). $(HhH)_2$ domains of both MUS81 and EME1 along with the nuclease domain, play a significant role in binding branched substrates. Substrate binding induces conformational changes in the overall structure of MUS81-EME1, resulting in exposing a hydrophobic wedge and a 5′ end binding pocket. These structural changes are essential for bending the DNA substrate, substrate specificity (nHJ or a 3′ flap), and catalysis (**Figure 4C**; Gwon et al., 2014). In plants, homologs of MUS81, as well as EME1, have been identified from *Arabidopsis* and rice (Berchowitz et al., 2007; Higgins et al., 2008; Mimida et al., 2007; Geuting et al., 2009). Arabidopsis genome has two homologs of EME1: AtEME1a and AtEME1b. Both EME1 homologs are capable of forming a functional enzymatic complex with MUS81 with differences in the processing of iHJ (Geuting et al., 2009). Interestingly, the *OsMUS81* gene of rice produces two alternate transcripts: *OsMUS81*α and *OsMUS81*β differing in the HhH motif at the C-terminal end. Further studies are necessary to characterize MUS81-EME1 complexes from plants.

SLX1

SLX1 is a member of the GIY-YIG family of endonucleases (Dunin-Horkawicz et al., 2006). In fungi and animals, SLX1 forms a heterodimeric complex with a non-enzymatic protein, SLX4. SLX1 interacts with SLX4 through the CCD domain (C-terminal conserved domain) of SLX4 (Fekairi et al., 2009; Gaur et al., 2015). SLX1-SLX4 complex participates in HR and maintenance of telomeres. Biochemically, SLX1-SLX4 complex can cleave HJs, 5′ flap, replication forks, 3′ flaps, and splayed arm DNA substrates (Fricke and Brill, 2003; Gaur et al., 2015, 2019). Crystal structures of SLX1 and SLX1 in complex with CCD domain of SLX4 (unliganded and in complex with DNA substrate) provided insights into the regulation of SLX1 activity and substrate specificity. SLX1 comprises an N-terminal GIY-YIG nuclease domain, and a C-terminal RING domain connected by a long α-helix (Gaur et al., 2015, 2019). In fungi, SLX1 exists in a self-inhibitory homodimeric form. An interaction between SLX1 and SLX4 is critical for the enzymatic activity of SLX1 (Gaur et al., 2015). Unlike other SSEs, SLX1 is devoid of any DNA binding secondary structural features. Instead, SLX1 has DNA binding patches on its surface. SLX1 uses the spatial organization of these DNA binding patches to bend the DNA substrate and identify the branching point as a flexible discontinuity in branched DNA substrates (**Figure 4D**; Gaur et al., 2019). Based on the available sequences, a GIY-YIG containing, SLX1 like protein called HIGLE has been identified in Arabidopsis (Cho et al., 2017). SLX4 has not been identified in plants so far and is an overly exciting avenue for future research. Further studies are required to ascertain the role of plant SLX1 in HJ resolution.

RTR Complex (RECQ4A-TOP3α-RMI1)

Homologous recombination is a central event during a crop breeding event. The key intermediate of HR is HJ (a four-way DNA junction). There are two independent mechanisms of processing an HJ: resolution and dissolution. Dissolution of HJ involves a complex of RecQ helicase (RECQ4A), topoisomerase 3α (TOP3α), and a structural protein RMI1 (Bagherieh-Najjar et al., 2005; Hartung et al., 2007a; Knoll and Puchta, 2011). Similar complexes from humans (BTR complex: BLM helicase, Topoisomerase 3α, RMI1, and RMI2), and yeast (Sgs1, Top3, and Rmi1) provide insights into the molecular architecture and function of RTR complex from plants (Wu L. et al., 2005; Bachrati and Hickson, 2009; Cejka et al., 2010; Bizard and Hickson, 2014). In humans, mutations in BLM helicase result in Bloom syndrome, characterized by increased frequency of sister chromatid exchange (Ellis et al., 1995). A similar phenotype is associated with mutations in yeast Ssg1 (Miyajima et al., 2000). Among plants, mutations associated with various components of the RTR complex result in hypersensitivity to DNA damaging agents and accumulation of unrepaired DSBs (Knoll et al., 2014). RTR complex by participating in dissolution significantly decreases the probability of reshuffling DNA segments by crossing over. RTR complex is one of the critical negative regulators of crossing over along with MMR proteins (Choi, 2017). A disruption of the MSH2 gene has already shown a 40% increase in the crossover rate in *Arabidopsis* (Emmanuel et al., 2006). Similar studies involving structural and biochemical characterizations for plant RTR complex are necessary to open new avenues for crop improvement.

TLS Polymerases

Trans-lesion DNA synthesis polymerases are essential for plant survival as they restart the stalled replication fork. TLS is an error-prone mechanism of tolerating DNA lesions. TLS can have both mutagenic and less mutagenic activities in plants (Sakamoto, 2019). Disruptions of AtPol ζ or AtRev1 decrease homologous recombination frequencies, whereas disruption of AtPol η increases the frequencies of homologous recombination in somatic tissues (Sakamoto, 2019). An overexpression of AtPOLH (coding for AtPolη) increases UV resistance in *Arabidopsis* (Jesus Santiago et al., 2008). REV3 subunit of Pol ζ appears to cooperate with structure-specific endonuclease, MUS81-EME1 (a participant in HJ resolution), however, similar cooperation with RECQ4A (a participant in HJ dissolution) is not known (Kobbe et al., 2015). All these studies have immense potential in designing strategies for crop improvement. The field of plant TLS polymerases is in initial phases. There is much to be done to understand substrate specificities of various plant TLS polymerases, regulation of their expression, coupling with DNA synthesis, and crosstalk with other DRR mechanisms such as photoreactivation, BER, NER, and HR.

Final Considerations

An intricate interplay of various DRR mechanisms continuously repairs DNA damages arising from multiple internal and external DNA damaging agents. Complex crosstalk between different DRR pathways exists in plants. The different DRR pathways are under regulation in the cell-cycle, tissue-specific, and development stage-dependent manner. Incidentally, factors described above likewise impact the success of present-day crop-improvement techniques. Therefore, an understanding of these pathways could further assist in fine-tuning various gene-editing techniques. There are already some attempts to understand the complex crosstalk between various DRR pathways, especially in

the context of HR. *e.g.*, the nature of DSB ends, and the phase of the cell cycle governs the choice between the two pathways involved in DSB repair: HR and NHEJ (Symington and Gautier, 2011). Plant hormones further control the choice between the two pathways. Abscisic acid increases the frequency of HR, while at the same time, suppresses Ku70 (a vital component of NHEJ) (Yin et al., 2009). Interestingly the two pathways are not entirely independent of each other, and various gene rearrangements could be explained based on the cooperative actions of both the pathways (Gorbunova and Levy, 1999). Various environmental factors (*e.g.,* the chemical composition of soil) and different amounts of exposure to mutagen can also influence the frequency of HR (Kovalchuk et al., 2000). Besides crosstalk with NHEJ, HR also crosstalks with the MMR pathway through MSH4 and Rad51 (Higgins et al., 2004). An interconnection between HR and early steps of the NER pathway also exists and mediated by CENTRIN2 (Molinier et al., 2004). A similar level of complexity involves the repair of DNA lesions induced by UV radiation and chemical mutagens. Various pathways ranging from photoreactivation, BER, MMR, NER to the involvement of TLS DNA polymerases cooperate/compete to repair a wide array of DNA lesions. There are indications of crosstalk between BER and other DDR pathways involving post-translational modifications (Limpose et al., 2017). OGG1, a glycosylase that repairs oxidized bases through BER, interacts with HR protein RAD52 with functional implications. This interaction inhibits RAD52 while increasing the turnover rate for OGG1 (de Souza-Pinto et al., 2009). All these studies imply the complex nature of plant DRR, and a necessity to comprehend these complexities for transforming future crop improvement techniques.

Crop improvement methodologies rely on generating DNA lesions through irradiation or chemical mutagenesis, the integration of foreign nucleic acids, and the repair of DSBs generated by site-directed nucleases. The success of any technique involved in crop improvement relies on the intrinsic DRR machinery of the plants and a complex interplay between various DRR pathways. There is an increased emphasis on increasing the crossover frequencies by manipulating the existing crosstalk between different DRR pathways (Blary and Jenczewski, 2019). One of the most successful methods of introducing mutations in a plant is the use of chemical mutagens. Such mutations by-and-large impact single nucleotides that are repaired by various glycosylases. Future crop improvement methodologies could be designed using engineered DNA glycosylases based on in-depth structural, biochemical, and computational studies. A structural-based study of different endonucleases from sources other than plants (*e.g.,* robust tools like CRISPR/Cas) enabled researchers to induce site-specific DSB successfully. Plants have several endogenous structure-specific endonucleases already involved in DRR pathways. The structural mechanism of the catalysis and regulation of endogenous structure-specific endonuclease can unleash their potential for crop improvement. Studying various aspects of DRR, therefore, provides an excellent opportunity to improvise the existing methods or to innovate new ways for efficient and faster crop improvement to meet the demand of billions of humans in an environmentally friendly manner.

AUTHOR CONTRIBUTIONS

PV, RT, GY, and VG contributed significantly to various sections of the manuscript. VG conceived the idea of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors gratefully acknowledge financial support from the NIPGR core grant to GY and VG, Ramalingaswami Fellowship (Department of Biotechnology, Government of India) to VG, DBT EMRF Grant BT/PR22334/BID/7/786/2016 to GY, and CSIR JRF to PV.

ACKNOWLEDGMENTS

The authors thank Dr. Pinky Agarwal for the critical reading of this manuscript. The authors also thank Director NIPGR, for support.

REFERENCES

Abdallah, N. A., Prakash, C. S., and McHughen, A. G. (2015). Genome editing for crop improvement: challenges and opportunities. *GM Crops Food* 6, 183–205. doi: 10.1080/21645698.2015.1129937

Abe, K., Osakabe, K., Nakayama, S., Endo, M., Tagiri, A., Todoriki, S., et al. (2005). Arabidopsis RAD51C gene is important for homologous recombination in meiosis and mitosis. *Plant Physiol.* 139, 896–908. doi: 10.1104/pp.105.065243

Agius, F., Kapoor, A., and Zhu, J. K. (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11796–11801. doi: 10.1073/pnas.0603563103

Ahmad, M., Jarillo, J. A., Klimczak, L. J., Landry, L. G., Peng, T., Last, R. L., et al. (1997). An enzyme similar to animal type II photolyases mediates photoreactivation in Arabidopsis. *Plant Cell* 9, 199–207. doi: 10.1105/tpc.9.2.199

Akutsu, N., Iijima, K., Hinata, T., and Tauchi, H. (2007). Characterization of the plant homolog of Nijmegen breakage syndrome 1: involvement in DNA repair and recombination. *Biochem. Biophys. Res. Commun.* 353, 394–398. doi: 10.1016/j.bbrc.2006.12.030

Ali, S., Ghufran, M., Nawaz, M. A., and Hussain, S. N. (2019). The psychological perspective on the adoption of approved genetically modified crops in the presence of acceptability constraint: the contingent role of passion. *GM Crops Food* 10, 220–237. doi: 10.1080/21645698.2019.1680241

Anthony, R. G., Reichelt, S., and Hussey, P. J. (1999). Dinitroaniline herbicide-resistant transgenic tobacco plants generated by co-overexpression of a mutant alpha-tubulin and a beta-tubulin. *Nat. Biotechnol.* 17, 712–716. doi: 10.1038/10931

Asagoshi, K., Tano, K., Chastain, P. D., Adachi, N., Sonoda, E., Kikuchi, K., et al. (2010). FEN1 functions in long patch base excision repair under conditions of oxidative stress in vertebrate cells. *Mol. Cancer Res.* 8, 204–215. doi: 10.1158/1541-7786.mcr-09-0253

Bachrati, C. Z., and Hickson, I. D. (2009). Dissolution of double Holliday junctions by the concerted action of BLM and topoisomerase IIIalpha. *Methods Mol. Biol.* 582, 91–102. doi: 10.1007/978-1-60761-340-4_8

Bagherieh-Najjar, M. B., de Vries, O. M., Hille, J., and Dijkwel, P. P. (2005). Arabidopsis RecQI4A suppresses homologous recombination and modulates DNA damage responses. *Plant J.* 43, 789–798. doi: 10.1111/j.1365-313x.2005.02501.x

Ban, C., and Yang, W. (1998). Structural basis for MutH activation in *E.coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* 17, 1526–1534. doi: 10.1093/emboj/17.5.1526

Bauknecht, M., and Kobbe, D. (2014). AtGEN1 and AtSEND1, two paralogs in Arabidopsis, possess holliday junction resolvase activity. *Plant Physiol.* 166, 202–216. doi: 10.1104/pp.114.237834

Bawa, A. S., and Anilakumar, K. R. (2013). Genetically modified foods: safety, risks and public concerns-a review. *J. Food Sci. Technol.* 50, 1035–1046. doi: 10.1007/s13197-012-0899-1

Belfield, E. J., Ding, Z. J., Jamieson, F. J. C., Visscher, A. M., Zheng, S. J., Mithani, A., et al. (2018). DNA mismatch repair preferentially protects genes from mutation. *Genome Res.* 28, 66–74. doi: 10.1101/gr.219303.116

Bensen, R. J., and Warner, H. R. (1987). Partial Purification and Characterization of Uracil-DNA Glycosylase Activity from Chloroplasts of Zea mays Seedlings. *Plant Physiol.* 84, 1102–1106. doi: 10.1104/pp.84.4.1102

Berchowitz, L. E., Francis, K. E., Bey, A. L., and Copenhaver, G. P. (2007). The role of AtMUS81 in interference-insensitive crossovers in *A. thaliana*. *PLoS Genet.* 3:e132. doi: 10.1371/journal.pgen.0030132

Bharti, P., Mahajan, M., Vishwakarma, A. K., Bhardwaj, J., and Yadav, S. K. (2015). AtROS1 overexpression provides evidence for epigenetic regulation of genes encoding enzymes of flavonoid biosynthesis and antioxidant pathways during salt stress in transgenic tobacco. *J. Exp. Bot.* 66, 5959–5969. doi: 10.1093/jxb/erv304

Bizard, A. H., and Hickson, I. D. (2014). The dissolution of double Holliday junctions. *Cold Spring Harb. Perspect. Biol.* 6:a016477. doi: 10.1101/cshperspect.a016477

Bjelland, S., Bjoras, M., and Seeberg, E. (1993). Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli*. *Nucleic Acids Res.* 21, 2045–2049. doi: 10.1093/nar/21.9.2045

Blary, A., and Jenczewski, E. (2019). Manipulation of crossover frequency and distribution for plant breeding. *Theor. Appl. Genet.* 132, 575–592. doi: 10.1007/s00122-018-3240-1

Boddy, M. N., Gaillard, P. H. L., McDonald, W. H., Shanahan, P., Yates, J. R., and Russell, P. (2001). Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell* 107, 537–548. doi: 10.1016/s0092-8674(01)00536-0

Bones, A. M. (1993). Expression and occurrence of uracil-DNA glycosylase in higher plants. *Physiol. Plant.* 88, 682–688. doi: 10.1034/j.1399-3054.1993.880422.x

Bonnet, S., Knoll, A., Hartung, F., and Puchta, H. (2013). Different functions for the domains of the *Arabidopsis thaliana* RMI1 protein in DNA cross-link repair, somatic and meiotic recombination. *Nucleic Acids Res.* 41, 9349–9360. doi: 10.1093/nar/gkt730

Boyko, A., Zemp, F., Filkowski, J., and Kovalchuk, I. (2006). Double-strand break repair in plants is developmentally regulated. *Plant Physiol.* 141, 488–497. doi: 10.1104/pp.105.074658

Breseghello, F., and Coelho, A. S. (2013). Traditional and modern plant breeding methods with examples in rice (*Oryza sativa* L.). *J. Agric. Food Chem.* 61, 8277–8286. doi: 10.1021/jf305531j

Brettel, K., and Byrdin, M. (2010). Reaction mechanisms of DNA photolyase. *Curr. Opin. Struct. Biol.* 20, 693–701. doi: 10.1016/j.sbi.2010.07.003

Britt, A. (2002). Repair of damaged bases. *Arabidopsis Book* 1:e0005. doi: 10.1199/tab.0005

Britt, A. B. (1999). Molecular genetics of DNA repair in higher plants. *Trends Plant Sci.* 4, 20–25. doi: 10.1016/s1360-1385(98)01355-7

Brooks, S. C., Adhikary, S., Rubinson, E. H., and Eichman, B. F. (2013). Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* 1834, 247–271. doi: 10.1016/j.bbapap.2012.10.005

Brooks, S. C., Fischer, R. L., Huh, J. H., and Eichman, B. F. (2014). 5-methylcytosine recognition by *Arabidopsis thaliana* DNA glycosylases DEMETER and DML3. *Biochemistry* 53, 2525–2532. doi: 10.1021/bi5002294

Broun, P., Gettner, S., and Somerville, C. (1999). Genetic engineering of plant lipids. *Annu. Rev. Nutr.* 19, 197–216. doi: 10.1146/annurev.nutr.19.1.197

Canturk, F., Karaman, M., Selby, C. P., Kemp, M. G., Kulaksiz-Erkmen, G., Hu, J., et al. (2016). Nucleotide excision repair by dual incisions in plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4706–4710. doi: 10.1073/pnas.1604097113

Caplan, A., Herrera-Estrella, L., Inze, D., Van Haute, E., Van Montagu, M., Schell, J., et al. (1983). Introduction of genetic material into plant cells. *Science* 222, 815–821. doi: 10.1126/science.222.4625.815

Cashmore, A. R. (2003). Cryptochromes: enabling plants and animals to determine circadian time. *Cell* 114, 537–543.

Cejka, P., Plank, J. L., Bachrati, C. Z., Hickson, I. D., and Kowalczykowski, S. C. (2010). Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. *Nat. Struct. Mol. Biol.* 17, 1377–1382. doi: 10.1038/nsmb.1919

Chang, J. H., Kim, J. J., Choi, J. M., Lee, J. H., and Cho, Y. J. (2008). Crystal structure of the Mus81-Eme1 complex. *Gene Dev.* 22, 1093–1106. doi: 10.1101/gad.1618708

Chelysheva, L., Vezon, D., Belcram, K., Gendrot, G., and Grelon, M. (2008). The *Arabidopsis BLAP75/Rmi1* homologue plays crucial roles in meiotic double-strand break repair. *PLoS Genet.* 4:e1000309. doi: 10.1371/journal.pgen.1000309

Chen, H., Chu, P., Zhou, Y., Li, Y., Liu, J., Ding, Y., et al. (2012). Overexpression of AtOGG1, a DNA glycosylase/AP lyase, enhances seed longevity and abiotic stress tolerance in Arabidopsis. *J. Exp. Bot.* 63, 4107–4121. doi: 10.1093/jxb/ers093

Chen, K., and Gao, C. (2014). Targeted genome modification technologies and their applications in crop improvements. *Plant Cell Rep.* 33, 575–583. doi: 10.1007/s00299-013-1539-6

Chen, L., Hao, L. G., Parry, M. A. J., Phillips, A. L., and Hu, Y. G. (2014). Progress in TILLING as a tool for functional genomics and improvement of crops. *J. Integr. Plant Biol.* 56, 425–443. doi: 10.1111/jipb.12192

Chirinos-Arias, M. C., and Spampinato, C. P. (2020). Growth and development of AtMSH7 mutants in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 146, 329–336. doi: 10.1016/j.plaphy.2019.11.035

Cho, S. K., Ryu, M. Y., Poulsen, C., Kim, J. H., Oh, T. R., Choi, S. W., et al. (2017). HIGLE is a bifunctional homing endonuclease that directly interacts with HYL1 and SERRATE in *Arabidopsis thaliana*. *FEBS Lett.* 591, 1383–1393. doi: 10.1002/1873-3468.12628

Choi, K. (2017). Advances towards controlling meiotic recombination for plant breeding. *Mol. Cells* 40, 814–822.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., et al. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* 110, 33–42. doi: 10.1016/s0092-8674(02)00807-3

Ciccia, A., McDonald, N., and West, S. C. (2008). Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu. Rev. Biochem.* 77, 259–287. doi: 10.1146/annurev.biochem.77.070306.102408

Conway, G., and Toenniessen, G. (1999). Feeding the world in the twenty-first century. *Nature* 402, C55–C58.

Cordoba-Canero, D., Cognat, V., Ariza, R. R., Roldan Arjona, T., and Molinier, J. (2017). Dual control of ROS1-mediated active DNA demethylation by DNA damage-binding protein 2 (DDB2). *Plant J.* 92, 1170–1181. doi: 10.1111/tpj.13753

Cordoba-Canero, D., Dubois, E., Ariza, R. R., Doutriaux, M. P., and Roldan-Arjona, T. (2010). Arabidopsis uracil DNA glycosylase (UNG) is required for base excision repair of uracil and increases plant sensitivity to 5-fluorouracil. *J. Biol. Chem.* 285, 7475–7483. doi: 10.1074/jbc.m109.067173

Cordoba-Canero, D., Roldan-Arjona, T., and Ariza, R. R. (2011). Arabidopsis ARP endonuclease functions in a branched base excision DNA repair pathway completed by LIG1. *Plant J.* 68, 693–702. doi: 10.1111/j.1365-313x.2011.04720.x

Cordoba-Canero, D., Roldan-Arjona, T., and Ariza, R. R. (2014). Arabidopsis ZDP DNA 3′-phosphatase and ARP endonuclease function in 8-oxoG repair initiated by FPG and OGG1 DNA glycosylases. *Plant J.* 79, 824–834. doi: 10.1111/tpj.12588

Culligan, K. M., and Hays, J. B. (2000). Arabidopsis MutS homologs-AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7-form three distinct protein heterodimers with different specificities for mismatched DNA. *Plant Cell* 12, 991–1002. doi: 10.2307/3871224

Dany, A. L., and Tissier, A. (2001). A functional OGG1 homologue from *Arabidopsis thaliana*. *Mol. Genet. Genomics* 265, 293–301. doi: 10.1007/s004380000414

de Souza-Pinto, N. C., Maynard, S., Hashiguchi, K., Hu, J., Muftuoglu, M., and Bohr, V. A. (2009). The recombination protein RAD52 cooperates with the excision repair protein OGG1 for the repair of oxidative lesions in mammalian cells. *Mol. Cell Biol.* 29, 4441–4454. doi: 10.1128/mcb.00265-09

Dehe, P. M., and Gaillard, P. H. L. (2017). Control of structure-specific endonucleases to maintain genome stability. *Nat. Rev. Mol. Cell. Biol.* 18, 315–330. doi: 10.1038/nrm.2016.177

Dempsey, D. A., Silva, H., and Klessig, D. F. (1998). Engineering disease and pest resistance in plants. *Trends Microbiol.* 6, 54–61. doi: 10.1016/s0966-842x(97)01186-4

Doe, C. L., Ahn, J. S., Dixon, J., and Whitby, M. C. (2002). Mus81-Eme1 and Rqh1 involvement in processing stalled and collapsed replication forks. *J. Biol. Chem.* 277, 32753–32759. doi: 10.1074/jbc.m202120200

Dovrat, D., Stodola, J. L., Burgers, P. M. J., and Aharoni, A. (2014). Sequential switching of binding partners on PCNA during in vitro Okazaki fragment maturation. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14118–14123. doi: 10.1073/pnas.1321349111

Dubest, S., Gallego, M. E., and White, C. I. (2004). Roles of the AtErcc1 protein in recombination. *Plant J.* 39, 334–342. doi: 10.1111/j.1365-313x.2004.02136.x

Duclos, S., Aller, P., Jaruga, P., Dizdaroglu, M., Wallace, S. S., and Doublie, S. (2012). Structural and biochemical studies of a plant formamidopyrimidine-DNA glycosylase reveal why eukaryotic Fpg glycosylases do not excise 8-oxoguanine. *DNA Repair* 11, 714–725. doi: 10.1016/j.dnarep.2012.06.004

Dunin-Horkawicz, S., Feder, M., and Bujnicki, J. M. (2006). Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7:98. doi: 10.1186/1471-2164-7-98

Ellis, N. A., Lennon, D. J., Proytcheva, M., Alhadeff, B., Henderson, E. E., and German, J. (1995). Somatic intragenic recombination within the mutated locus BLM can correct the high sister-chromatid exchange phenotype of Bloom syndrome cells. *Am. J. Hum. Genet.* 57, 1019–1027.

Emmanuel, E., Yehuda, E., Melamed-Bessudo, C., Avivi-Ragolsky, N., and Levy, A. A. (2006). The role of AtMSH2 in homologous recombination in *Arabidopsis thaliana*. *EMBO Rep.* 7, 100–105. doi: 10.1038/sj.embor.7400577

Espina, M. J., Ahmed, C. M. S., Bernardini, A., Adeleke, E., Yadegari, Z., Arelli, P., et al. (2018). Development and Phenotypic Screening of an Ethyl Methane Sulfonate Mutant Population in Soybean. *Front. Plant Sci.* 9:394. doi: 10.3389/fpls.2018.00394

Esquinas-Alcazar, J. (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat. Rev. Genet.* 6, 946–953. doi: 10.1038/nrg1729

Essen, L. O., and Klar, T. (2006). Light-driven DNA repair by photolyases. *Cell. Mol. Life Sci.* 63, 1266–1277. doi: 10.1007/s00018-005-5447-y

Fekairi, S., Scaglione, S., Chahwan, C., Taylor, E. R., Tissier, A., Coulon, S., et al. (2009). Human SLX4 Is a Holliday Junction Resolvase Subunit that Binds Multiple DNA Repair/Recombination Endonucleases. *Cell* 138, 78–89. doi: 10.1016/j.cell.2009.06.029

Fishel, R., and Wilson, T. (1997). MutS homologs in mammalian cells. *Curr. Opin. Genet. Dev.* 7, 105–113. doi: 10.1016/s0959-437x(97)80117-7

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342.

Fricke, W. M., Bastin-Shanower, S. A., and Brill, S. J. (2005). Substrate specificity of the *Saccharomyces cerevisiae* Mus81-Mms4 endonuclease. *DNA Repair* 4, 243–251. doi: 10.1016/j.dnarep.2004.10.001

Fricke, W. M., and Brill, S. J. (2003). Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. *Gene Dev.* 17, 1768–1778. doi: 10.1101/gad.1105203

Fromme, J. C., Banerjee, A., and Verdine, G. L. (2004). DNA glycosylase recognition and catalysis. *Curr. Opin. Struct. Biol.* 14, 43–49. doi: 10.1016/j.sbi.2004.01.003

Furukawa, T., Kimura, S., Ishibashi, T., Mori, Y., Hashimoto, J., and Sakaguchi, K. (2003). OsSEND-1: a new RAD2 nuclease family member in higher plants. *Plant Mol. Biol.* 51, 59–70.

Furukawa, T., and Shimada, H. (2009). The RAD2 family of nucleases in higher plants. *Jarq Jpn. Agric. Res. Q.* 43, 87–93. doi: 10.6090/jarq.43.87

Gaillard, P. H. L., Noguchi, E., Shanahan, P., and Russell, P. (2003). The endogenous Mus81-Eme1 complex resolves Holliday junctions by a nick and counternick mechanism. *Mol. Cell* 12, 747–759. doi: 10.1016/s1097-2765(03)00342-3

Gallego, F., Fleck, O., Li, A., Wyrzykowska, J., and Tinland, B. (2000). AtRAD1, a plant homologue of human and yeast nucleotide excision repair endonucleases, is involved in dark repair of UV damages and recombination. *Plant J.* 21, 507–518. doi: 10.1046/j.1365-313x.2000.00694.x

Gao, M. J., and Murphy, T. M. (2001). Alternative forms of formamidopyrimidine-DNA glycosylase from *Arabidopsis thaliana*. *Photochem. Photobiol.* 73, 128–134. doi: 10.1562/0031-8655(2001)073<0128:afofdg>2.0.co;2

Garcia-Ortiz, M. V., Ariza, R. R., Hoffman, P. D., Hays, J. B., and Roldan-Arjona, T. (2004). *Arabidopsis thaliana* AtPOLK encodes a DinB-like DNA polymerase that extends mispaired primer termini and is highly expressed in a variety of tissues. *Plant J.* 39, 84–97. doi: 10.1111/j.1365-313x.2004.02112.x

Garcia-Ortiz, M. V., Ariza, R. R., and Roldan-Arjona, T. (2001). An OGG1 orthologue encoding a functional 8-oxoguanine DNA glycosylase/lyase in *Arabidopsis thaliana*. *Plant Mol. Biol.* 47, 795–804.

Garcia-Ortiz, M. V., Roldan-Arjona, T., and Ariza, R. R. (2007). The noncatalytic C-terminus of AtPOLK Y-family DNA polymerase affects synthesis fidelity, mismatch extension and translesion replication. *FEBS J.* 274, 3340–3350. doi: 10.1111/j.1742-4658.2007.05868.x

Gaur, V., Wyatt, H. D. M., Komorowska, W., Szczepanowski, R. H., de Sanctis, D., Gorecka, K. M., et al. (2015). Structural and Mechanistic Analysis of the Slx1-Slx4 Endonuclease. *Cell Rep.* 10, 1467–1476. doi: 10.1016/j.celrep.2015.02.019

Gaur, V., Ziajko, W., Nirwal, S., Szlachcic, A., Gapinska, M., and Nowotny, M. (2019). Recognition and processing of branched DNA substrates by Slx1-Slx4 nuclease. *Nucleic Acids Res.* 47, 11681–11690. doi: 10.1093/nar/gkz842

Ge, L. F., Yu, J. B., Wang, H. L., Luth, D., Bai, G. H., Wang, K., et al. (2016). Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12414–12419. doi: 10.1073/pnas.1611763113

Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., et al. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124, 495–506. doi: 10.1016/j.cell.2005.12.034

Geier, T. (2012). Chimeras: properties and dissociation in vegetatively propagated plants. *Plant Mutat. Breed. Biotechnol.* 191–201. doi: 10.1079/9781780640853.0191

Geuting, V., Kobbe, D., Hartung, F., Durr, J., Focke, M., and Puchta, H. (2009). Two Distinct MUS81-EME1 complexes from Arabidopsis process Holliday junctions. *Plant Physiol.* 150, 1062–1071. doi: 10.1104/pp.109.136846

Giraldo, P. A., Shinozuka, H., Spangenberg, G. C., Cogan, N. O. I., and Smith, K. F. (2019). Safety assessment of genetically modified feed: Is there any difference from food? *Front. Plant Sci.* 10:1592. doi: 10.3389/fpls.2019.01592

Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., and Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Biol.* 13, 167–173. doi: 10.1016/j.pbi.2010.01.004

Gomes, X. V., and Burgers, P. M. J. (2000). Two modes of FEN1 binding to PCNA regulated by DNA. *EMBO J.* 19, 3811–3821. doi: 10.1093/emboj/19.14.3811

Gomez, R., and Spampinato, C. P. (2013). Mismatch recognition function of *Arabidopsis thaliana* MutSgamma. *DNA Repair* 12, 257–264. doi: 10.1016/j.dnarep.2013.01.002

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldan-Arjona, T., David, L., and Zhu, J. K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* 111, 803–814. doi: 10.1016/s0092-8674(02)01133-9

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186.

Gorbunova, V. V., and Levy, A. A. (1999). How plants make ends meet: DNA double-strand break repair. *Trends Plant Sci.* 4, 263–269. doi: 10.1016/s1360-1385(99)01430-2

Grelon, M., Vezon, D., Gendrot, G., and Pelletier, G. (2001). AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J.* 20, 589–600. doi: 10.1093/emboj/20.3.589

Gutman, B. L., and Niyogi, K. K. (2009). Evidence for Base Excision Repair of Oxidative DNA Damage in Chloroplasts of *Arabidopsis thaliana*. *J. Biol. Chem.* 284, 17006–17012. doi: 10.1074/jbc.m109.008342

Gwon, G. H., Jo, A., Baek, K., Jin, K. S., Fu, Y., Lee, J. B., et al. (2014). Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates. *EMBO J.* 33, 1061–1072. doi: 10.1002/embj.201487820

Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J., et al. (2006). The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *EMBO J.* 25, 4921–4932. doi: 10.1038/sj.emboj.7601344

Harfe, B. D., and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. *Annu. Rev. Genet.* 34, 359–399.

Hartung, F., Suer, S., Bergmann, T., and Puchta, H. (2006). The role of AtMUS81 in DNA repair and its genetic interaction with the helicase AtRecQ4A. *Nucleic Acids Res.* 34, 4438–4448. doi: 10.1093/nar/gkl576

Hartung, F., Suer, S., Knoll, A., Wurz-Wildersinn, R., and Puchta, H. (2008). Topoisomerase 3alpha and RMI1 suppress somatic crossovers and are essential for resolution of meiotic recombination intermediates in *Arabidopsis thaliana*. *PLoS Genet.* 4:e1000285. doi: 10.1371/journal.pgen.1000285

Hartung, F., Suer, S., and Puchta, H. (2007a). Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18836–18841.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S., and Puchta, H. (2007b). The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. *Plant Cell* 19, 3090–3099. doi: 10.1105/tpc.107.054817

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* 12, 934–940. doi: 10.1111/pbi.12201

Hefner, E., Huefner, N., and Britt, A. B. (2006). Tissue-specific regulation of cell-cycle responses to DNA damage in Arabidopsis seedlings. *DNA Repair* 5, 102–110. doi: 10.1016/j.dnarep.2005.08.013

Hervella, M., Izagirre, N., Alonso, S., Fregel, R., Alonso, A., Cabrera, V. M., et al. (2012). Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS One* 7:e34417. doi: 10.1371/journal.pone.0034417

Hidema, J., Kumagai, T., and Sutherland, B. M. (2000). UV radiation-sensitive norin 1 rice contains defective cyclobutane pyrimidine dimer photolyase. *Plant Cell* 12, 1569–1578. doi: 10.2307/3871174

Hidema, J., Kumagai, T., Sutherland, J. C., and Sutherland, B. M. (1997). Ultraviolet B-sensitive rice cultivar deficient in cyclobutyl pyrimidine dimer repair. *Plant Physiol.* 113, 39–44. doi: 10.1104/pp.113.1.39

Higgins, J. D., Armstrong, S. J., Franklin, F. C., and Jones, G. H. (2004). The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes Dev.* 18, 2557–2570. doi: 10.1101/gad.317504

Higgins, J. D., Buckling, E. F., Franklin, F. C., and Jones, G. H. (2008). Expression and functional analysis of AtMUS81 in Arabidopsis meiosis reveals a role in the second pathway of crossing-over. *Plant J.* 54, 152–162. doi: 10.1111/j.1365-313x.2008.03403.x

Higgins, J. D., Ferdous, M., Osman, K., and Franklin, F. C. (2011). The RecQ helicase AtRECQ4A is required to remove inter-chromosomal telomeric connections that arise during meiotic recombination in Arabidopsis. *Plant J.* 65, 492–502. doi: 10.1111/j.1365-313x.2010.04438.x

Hirschberg, J. (1999). Production of high-value compounds: carotenoids and vitamin E. *Curr. Opin. Biotechnol.* 10, 186–191. doi: 10.1016/s0958-1669(99)80033-0

Hitomi, K., Arvai, A. S., Yamamoto, J., Hitomi, C., Teranishi, M., Hirouchi, T., et al. (2012). Eukaryotic class II cyclobutane pyrimidine dimer photolyase structure reveals basis for improved ultraviolet tolerance in plants. *J. Biol. Chem.* 287, 12060–12069. doi: 10.1074/jbc.m111.244020

Hitomi, K., DiTacchio, L., Arvai, A. S., Yamamoto, J., Kim, S. T., Todo, T., et al. (2009). Functional motifs in the (6-4) photolyase crystal structure make a comparative framework for DNA repair photolyases and clock cryptochromes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6962–6967. doi: 10.1073/pnas.0809180106

Hoffman, P. D., Leonard, J. M., Lindberg, G. E., Bollmann, S. R., and Hays, J. B. (2004). Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective Arabidopsis. *Gene Dev.* 18, 2676–2685. doi: 10.1101/gad.1217204

Holmberg, N., and Bulow, L. (1998). Improving stress tolerance in plants by gene transfer. *Trends Plant Sci.* 3, 61–66. doi: 10.1016/s1360-1385(97)01163-1

Holme, I. B., Gregersen, P. L., and Brinch-Pedersen, H. (2019). Induced genetic variation in crop plants by random or targeted mutagenesis: convergence and differences. *Front. Plant Sci.* 10:1468. doi: 10.3389/fpls.2019.01468

Huggins, C. F., Chafin, D. R., Aoyagi, S., Henricksen, L. A., Bambara, R. A., and Hayes, J. J. (2002). Flap endonuclease 1 efficiently cleaves base excision repair and DNA replication intermediates assembled into nucleosomes. *Mol. Cell* 10, 1201–1211. doi: 10.1016/s1097-2765(02)00736-0

Hussain, M., Iqbal, M. A., Till, B. J., and ur-Rahman, M. (2018). Identification of induced mutations in hexaploid wheat genome using exome capture assay. *PLoS One* 13:e0201918. doi: 10.1371/journal.pone.0201918

Hwang, H. H., Yu, M., and Lai, E. M. (2017). Agrobacterium-mediated plant transformation: biology and applications. *Arabidopsis Book* 15:e0186. doi: 10.1199/tab.0186

Ikeda, Y., Kinoshita, Y., Susaki, D., Ikeda, Y., Iwano, M., Takayama, S., et al. (2011). HMG Domain Containing SSRP1 Is Required for DNA Demethylation and Genomic Imprinting in Arabidopsis. *Dev. Cell* 21, 589–596. doi: 10.1016/j.devcel.2011.08.013

Ip, S. C. Y., Rass, U., Blanco, M. G., Flynn, H. R., Skehel, J. M., and West, S. C. (2008). Identification of Holliday junction resolvases from humans and yeast. *Nature* 456, 357–U339.

Jesus Santiago, M., Alejandre-Duran, E., Munoz-Serrano, A., and Ruiz-Rubio, M. (2008). Two translesion synthesis DNA polymerase genes, AtPOLH and AtREV1, are involved in development and UV light resistance in Arabidopsis. *J. Plant Physiol.* 165, 1582–1591. doi: 10.1016/j.jplph.2007.11.012

Jia, Q., den Dulk-Ras, A., Shen, H., Hooykaas, P. J., and de Pater, S. (2013). Poly(ADP-ribose)polymerases are involved in microhomology mediated back-up non-homologous end joining in *Arabidopsis thaliana*. *Plant Mol. Biol.* 82, 339–351. doi: 10.1007/s11103-013-0065-9

Jiang, C. Z., Yee, J., Mitchell, D. L., and Britt, A. B. (1997). Photorepair mutants of Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7441–7445. doi: 10.1073/pnas.94.14.7441

Jones, H. D. (2015). Regulatory uncertainty over genome editing. *Nat. Plants* 1:14011.

Kadyrova, L. Y., and Kadyrov, F. A. (2016). Endonuclease activities of MutLalpha and its homologs in DNA mismatch repair. *DNA Repair* 38, 42–49. doi: 10.1016/j.dnarep.2015.11.023

Kaiser, G., Kleiner, O., Beisswenger, C., and Batschauer, A. (2009). Increased DNA repair in Arabidopsis plants overexpressing CPD photolyase. *Planta* 230, 505–515. doi: 10.1007/s00425-009-0962-y

Kamle, M., Kumar, P., Patra, J. K., and Bajpai, V. K. (2017). Current perspectives on genetically modified crops and detection methods. *3 Biotech* 7:219.

Kapoor, A., Agarwal, M., Andreucci, A., Zheng, X., Gong, Z., Hasegawa, P. M., et al. (2005a). Mutations in a conserved replication protein suppress transcriptional gene silencing in a DNA-methylation-independent manner in Arabidopsis. *Curr. Biol.* 15, 1912–1918. doi: 10.1016/j.cub.2005.09.013

Kapoor, A., Agius, F., and Zhu, J. K. (2005b). Preventing transcriptional gene silencing by active DNA demethylation. *FEBS Lett.* 579, 5889–5898. doi: 10.1016/j.febslet.2005.08.039

Katarzyna Banas, A., Hermanowicz, P., Sztatelman, O., Labuz, J., Aggarwal, C., Zglobicki, P., et al. (2018). 6,4-PP Photolyase Encoded by AtUVR3 is Localized in Nuclei. Chloroplasts and Mitochondria and its Expression is Down-Regulated by Light in a Photosynthesis-Dependent Manner. *Plant Cell Physiol.* 59, 44–57. doi: 10.1093/pcp/pcx159

Kathe, S. D., Barrantes-Reynolds, R., Jaruga, P., Newton, M. R., Burrows, C. J., Bandaru, V., et al. (2009). Plant and fungal Fpg homologs are formamidopyrimidine DNA glycosylases but not 8-oxoguanine DNA glycosylases. *DNA Repair* 8, 643–653. doi: 10.1016/j.dnarep.2008.12.013

Kim, C. Y., Park, M. S., and Dyer, R. B. (2001). Human flap endonuclease-1: conformational change upon binding to the flap DNA substrate and location of the Mg2+ binding site. *Biochemistry* 40, 3208–3214. doi: 10.1021/bi002100n

Kim, J. S., Chang, J. H., Kim, J. J., and Cho, Y. (2008). Crystal structure of the Mus81-Eme1 complex. *Acta Crystallogr. A* 64, C301–C301.

Kimura, S., Furukawa, T., Kasai, N., Mori, Y., Kitamoto, H. K., Sugawara, F., et al. (2003). Functional characterization of two flap endonuclease-1 homologues in rice. *Gene* 314, 63–71. doi: 10.1016/s0378-1119(03)00694-2

Kimura, S., Kai, M., Kobayashi, H., Suzuki, A., Morioka, H., Otsuka, E., et al. (1997). A structure-specific endonuclease from cauliflower (Brassica oleracea var. botrytis) inflorescence. *Nucleic Acids Res.* 25, 4970–4976. doi: 10.1093/nar/25.24.4970

Kimura, S., and Sakaguchi, K. (2006). DNA repair in plants. *Chem. Rev.* 106, 753–766.

Kimura, S., Suzuki, T., Yanagawa, Y., Yamamoto, T., Nakagawa, H., Tanaka, I., et al. (2001). Characterization of plant proliferating cell nuclear antigen (PCNA) and flap endonuclease-1 (FEN-1), and their distribution in mitotic and meiotic cell cycles. *Plant J.* 28, 643–653. doi: 10.1046/j.1365-313x.2001.01184.x

Kimura, S., Ueda, T., Hatanaka, M., Takenouchi, M., Hashimoto, J., and Sakaguchi, K. (2000). Plant homologue of flap endonuclease-1: molecular cloning, characterization, and evidence of expression in meristematic tissues. *Plant Mol. Biol.* 42, 415–427.

Klein, R. M., Wolf, E. D., Wu, R., and Sanford, J. C. (1992). High-velocity microprojectiles for delivering nucleic acids into living cells. 1987. *Biotechnology* 24, 384–386.

Knoll, A., and Puchta, H. (2011). The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. *J. Exp. Bot.* 62, 1565–1579. doi: 10.1093/jxb/erq357

Knoll, A., Schropfer, S., and Puchta, H. (2014). The RTR complex as caretaker of genome stability and its unique meiotic function in plants. *Front. Plant Sci.* 5:33. doi: 10.3389/fpls.2014.00033

Kobbe, S., Trapp, O., Knoll, A., Manuss, A., and Puchta, H. (2015). The translesion polymerase zeta has roles dependent on and independent of the nuclease MUS81 and the Helicase RECQ4A in DNA Damage Repair in Arabidopsis. *Plant Physiol.* 169, 2718–2729.

Koh, S., Lee, S. C., Kim, M. K., Koh, J. H., Lee, S., An, G., et al. (2007). T-DNA tagged knockout mutation of rice OsGSK1, an orthologue of Arabidopsis BIN2, with enhanced tolerance to various abiotic stresses. *Plant Mol. Biol.* 65, 453–466. doi: 10.1007/s11103-007-9213-4

Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev.* 10, 1433–1442. doi: 10.1101/gad.10.12.1433

Kovalchuk, O., Arkhipov, A., Barylyak, I., Karachov, I., Titov, V., Hohn, B., et al. (2000). Plants experiencing chronic internal exposure to ionizing radiation exhibit higher frequency of homologous recombination than acutely irradiated plants. *Mutat. Res.* 449, 47–56. doi: 10.1016/s0027-5107(00)00029-4

Kowalska, E., Strzalka, W., and Oyama, T. (2020). A crystallization and preliminary X-ray diffraction study of the *Arabidopsis thaliana* proliferating cell nuclear antigen (PCNA2) alone and in a complex with a PIP-box peptide from Flap endonuclease 1. *Acta Biochim. Pol.* 67, 49–52.

Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E., and Lindahl, T. (1996). Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.* 15, 6662–6670. doi: 10.1002/j.1460-2075.1996.tb01056.x

Kurzbauer, M. T., Pradillo, M., Kerzendorfer, C., Sims, J., Ladurner, R., Oliver, C., et al. (2018). *Arabidopsis thaliana* FANCD2 promotes meiotic crossover formation. *Plant Cell* 30, 415–428. doi: 10.1105/tpc.17.00745

Kwon, K., Jiang, Y. L., and Stivers, J. T. (2003). Rational engineering of a DNA glycosylase specific for an unnatural cytosine : pyrene base pair. *Chem. Biol.* 10, 351–359. doi: 10.1016/s1074-5521(03)00077-2

Kwon, Y. I., Abe, K., Endo, M., Osakabe, K., Ohtsuki, N., Nishizawa-Yokoi, A., et al. (2013). DNA replication arrest leads to enhanced homologous recombination and cell death in meristems of rice OsRecQl4 mutants. *BMC Plant Biol.* 13:62. doi: 10.1186/1471-2229-13-62

Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., and Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell* 86, 321–329. doi: 10.1016/s0092-8674(00)80103-8

Lahue, R. S., Au, K. G., and Modrich, P. (1989). DNA mismatch correction in a defined system. *Science* 245, 160–164. doi: 10.1126/science.2665076

Lamers, M. H., Winterwerp, H. H., and Sixma, T. K. (2003). The alternating ATPase domains of MutS control DNA mismatch repair. *EMBO J.* 22, 746–756. doi: 10.1093/emboj/cdg064

Landry, L. G., Stapleton, A. E., Lim, J., Hoffman, P., Hays, J. B., Walbot, V., et al. (1997). An Arabidopsis photolyase mutant is hypersensitive to ultraviolet-B radiation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 328–332. doi: 10.1073/pnas.94.1.328

Larson, K., Sahm, J., Shenkar, R., and Strauss, B. (1985). Methylation-induced blocks to in vitro DNA replication. *Mutat. Res.* 150, 77–84. doi: 10.1016/0027-5107(85)90103-4

Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, T. (1998). Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* 95, 249–258. doi: 10.1016/s0092-8674(00)81755-9

Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000). Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase. *AAG. Proc. Natl. Acad. Sci. U.S.A.* 97, 13573–13578. doi: 10.1073/pnas.97.25.13573

Le, T. N., Schumann, U., Smith, N. A., Tiwari, S., Au, P. C., Zhu, Q. H., et al. (2014). DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. *Genome Biol.* 15:458.

Lee, S. H., Princz, L. N., Klugel, M. F., Habermann, B., Pfander, B., and Biertumpfel, C. (2015). Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *eLife* 4:e12256.

Lehmann, A. R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J. F., et al. (2007). Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair* 6, 891–899. doi: 10.1016/j.dnarep.2007.02.003

Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85–98. doi: 10.1038/cr.2007.115

Li, L. L., Dion, E., Richard, G., Domingue, O., Jean, M., and Belzile, F. J. (2009). The Arabidopsis DNA mismatch repair gene PMS1 restricts somatic recombination between homeologous sequences. *Plant Mol. Biol.* 69, 675–684. doi: 10.1007/s11103-008-9447-9

Li, L. L., Jean, M., and Belzile, F. (2006). The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *Plant J.* 45, 908–916. doi: 10.1111/j.1365-313x.2006.02657.x

Li, X., and Heyer, W. D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* 18, 99–113. doi: 10.1038/cr.2008.1

Li, X., Qian, W., Zhao, Y., Wang, C., Shen, J., Zhu, J. K., et al. (2012). Antisilencing role of the RNA-directed DNA methylation pathway and a histone acetyltransferase in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11425–11430. doi: 10.1073/pnas.1208557109

Lieber, M. R. (1997). The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays* 19, 233–240. doi: 10.1002/bies.950190309

Limpose, K. L., Corbett, A. H., and Doetsch, P. W. (2017). BERing the burden of damage: pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management. *DNA Repair* 56, 51–64. doi: 10.1016/j.dnarep.2017.06.007

Liu, H., Liu, B., Zhao, C., Pepper, M., and Lin, C. (2011). The action mechanisms of plant cryptochromes. *Trends Plant Sci.* 16, 684–691. doi: 10.1016/j.tplants.2011.09.002

Liu, Y., Kao, H. I., and Bambara, R. A. (2004). Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.* 73, 589–615. doi: 10.1146/annurev.biochem.73.012803.092453

Liu, Y. J., Freeman, A. D. J., Declais, A. C., Wilson, T. J., Gartner, A., and Lilley, D. M. J. (2015). Crystal structure of a eukaryotic GEN1 resolving enzyme bound to DNA. *Cell Rep.* 13, 2565–2575. doi: 10.1016/j.celrep.2015.11.042

Liu, Z., Hossain, G. S., Islas-Osuna, M. A., Mitchell, D. L., and Mount, D. W. (2000). Repair of UV damage in plants by nucleotide excision repair: Arabidopsis UVH1 DNA repair gene is a homolog of *Saccharomyces cerevisiae* Rad1. *Plant J.* 21, 519–528. doi: 10.1046/j.1365-313x.2000.00707.x

Loechler, E. L., Green, C. L., and Essigmann, J. M. (1984). In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proc. Natl. Acad. Sci. U.S.A.* 81, 6271–6275. doi: 10.1073/pnas.81.20.6271

Lopez Sanchez, A., Stassen, J. H., Furci, L., Smith, L. M., and Ton, J. (2016). The role of DNA (de)methylation in immune responsiveness of Arabidopsis. *Plant J.* 88, 361–374. doi: 10.1111/tpj.13252

Luca, F., Perry, G. H., and Di Rienzo, A. (2010). Evolutionary adaptations to dietary changes. *Annu. Rev. Nutr.* 30, 291–314. doi: 10.1146/annurev-nutr-080508-141048

Macovei, A., Balestrazzi, A., Confalonieri, M., Fae, M., and Carbonera, D. (2011). New insights on the barrel medic MtOGG1 and MtFPG functions in relation to oxidative stress response in planta and during seed imbibition. *Plant Physiol. Biochem.* 49, 1040–1050. doi: 10.1016/j.plaphy.2011.05.007

Mannuss, A., Dukowic-Schulze, S., Suer, S., Hartung, F., Pacher, M., and Puchta, H. (2010). RAD5A. RECQ4A, and MUS81 have specific functions in homologous recombination and define different pathways of DNA repair in *Arabidopsis thaliana*. *Plant Cell* 22, 3318–3330. doi: 10.1105/tpc.110.078568

Manova, V., and Gruszka, D. (2015). DNA damage and repair in plants - from models to crops. *Front. Plant Sci.* 6:885. doi: 10.3389/fpls.2015.00885

Matos, J., and West, S. C. (2014). Holliday junction resolution: regulation in space and time. *DNA Repair* 19, 176–181. doi: 10.1016/j.dnarep.2014.03.013

Matson, S. W., and Robertson, A. B. (2006). The UvrD helicase and its modulation by the mismatch repair protein MutL. *Nucleic Acids Res.* 34, 4089–4097. doi: 10.1093/nar/gkl450

Maul, M. J., Barends, T. R., Glas, A. F., Cryle, M. J., Domratcheva, T., Schneider, S., et al. (2008). Crystal structure and mechanism of a DNA (6-4) photolyase. *Angew. Chem. Int. Ed. Engl.* 47, 10076–10080.

Melamed-Bessudo, C., and Levy, A. A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 109, E981–E988.

Mimida, N., Kitamoto, H., Osakabe, K., Nakashima, M., Ito, Y., Heyer, W. D., et al. (2007). Two alternatively spliced transcripts generated from OsMUS81, a rice homolog of yeast MUS81, are up-regulated by DNA-damaging treatments. *Plant Cell Physiol.* 48, 648–654. doi: 10.1093/pcp/pcm029

Minocherhomji, S., and Hickson, I. D. (2014). Structure-specific endonucleases: guardians of fragile site stability. *Trends Cell Biol.* 24, 321–327. doi: 10.1016/j.tcb.2013.11.007

Miyajima, A., Seki, M., Onoda, F., Shiratori, M., Odagiri, N., Ohta, K., et al. (2000). Sgs1 helicase activity is required for mitotic but apparently not for meiotic functions. *Mol. Cell Biol.* 20, 6399–6409. doi: 10.1128/mcb.20.17.6399-6409.2000

Mol, J., Cornish, E., Mason, J., and Koes, R. (1999). Novel coloured flowers. *Curr. Opin. Biotechnol.* 10, 198–201. doi: 10.1016/s0958-1669(99)80035-4

Molinier, J., Lechner, E., Dumbliauskas, E., and Genschik, P. (2008). Regulation and role of Arabidopsis CUL4-DDB1A-DDB2 in maintaining genome integrity upon UV stress. *PLoS Genet.* 4:e1000093. doi: 10.1371/journal.pgen.1000093

Molinier, J., Ramos, C., Fritsch, O., and Hohn, B. (2004). CENTRIN2 modulates homologous recombination and nucleotide excision repair in Arabidopsis. *Plant Cell* 16, 1633–1643. doi: 10.1105/tpc.021378

Morales-Ruiz, T., Birincioglu, M., Jaruga, P., Rodriguez, H., Roldan-Arjona, T., and Dizdaroglu, M. (2003). *Arabidopsis thaliana* Ogg1 protein excises 8-hydroxyguanine and 2,6-diamino-4-hydroxy-5-formamidopyrimidine from oxidatively damaged DNA containing multiple lesions. *Biochemistry* 42, 3089–3095. doi: 10.1021/bi027226u

Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marin, M. I., Martinez-Macias, M. I., Ariza, R. R., and Roldan-Arjona, T. (2006). DEMETER and REPRESSOR OFSILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6853–6858. doi: 10.1073/pnas.0601109103

Moritoh, S., Miki, D., Akiyama, M., Kawahara, M., Izawa, T., Maki, H., et al. (2005). RNAi-mediated silencing of OsGEN-L (OsGEN-like), a new member of the RAD2/XPG nuclease family, causes male sterility by defect of microspore development in rice. *Plant Cell Physiol.* 46, 699–715. doi: 10.1093/pcp/pci090

Murphy, T. M. (2005). What is base excision repair good for? knockout mutants for FPG and OGG glycosylase genes in Arabidopsis. *Physiol. Plant.* 123, 227–232. doi: 10.1111/j.1399-3054.2005.00453.x

Murphy, T. M., and Gao, M. J. (2001). Multiple forms of formamidopyrimidine-DNA glycosylase produced by alternative splicing in *Arabidopsis thaliana*. *J. Photochem. Photobiol. B* 61, 87–93. doi: 10.1016/s1011-1344(01)00172-5

Nakagawa, M., Takahashi, S., Tanaka, A., Narumi, I., and Sakamoto, A. N. (2011). Role of AtPolzeta. AtRev1, and AtPoleta in UV light-induced mutagenesis in Arabidopsis. *Plant Physiol.* 155, 414–420. doi: 10.1104/pp.110.166082

Nakajima, S., Sugiyama, M., Iwai, S., Hitomi, K., Otoshi, E., Kim, S. T., et al. (1998). Cloning and characterization of a gene (UVR3) required for photorepair of 6-4 photoproducts in *Arabidopsis thaliana*. *Nucleic Acids Res.* 26, 638–644. doi: 10.1093/nar/26.2.638

Newell, C. A. (2000). Plant transformation technology - Developments and applications. *Mol. Biotechnol.* 16, 53–65. doi: 10.1385/mb:16:1:53

Nishizawa-Yokoi, A., Nonaka, S., Saika, H., Kwon, Y. I., Osakabe, K., and Toki, S. (2012). Suppression of Ku70/80 or Lig4 leads to decreased stable transformation and enhanced homologous recombination in rice. *New Phytol.* 196, 1048–1059. doi: 10.1111/j.1469-8137.2012.04350.x

Nota, F., Cambiagno, D. A., Ribone, P., and Alvarez, M. E. (2015). Expression and function of AtMBD4L, the single gene encoding the nuclear DNA glycosylase MBD4L in Arabidopsis. *Plant Sci.* 235, 122–129. doi: 10.1016/j.plantsci.2015.03.011

Ohr, H., Bui, A. Q., Le, B. H., Fischer, R. L., and Choi, Y. (2007). Identification of putative Arabidopsis DEMETER target genes by GeneChip analysis. *Biochem. Biophys. Res. Commun.* 364, 856–860. doi: 10.1016/j.bbrc.2007.10.092

Ohtsubo, T., Matsuda, O., Iba, K., Terashima, I., Sekiguchi, M., and Nakabeppu, Y. (1998). Molecular cloning of AtMMH, an *Arabidopsis thaliana* ortholog of the *Escherichia coli* mutM gene, and analysis of functional domains of its product. *Mol. Gen. Genet.* 259, 577–590. doi: 10.1007/s004380050851

Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., et al. (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnol. Biotechnol Equip.* 30, 1–16. doi: 10.1080/13102818.2015.1087333

Olivier, M., Da Ines, O., Amiard, S., Serra, H., Goubely, C., White, C. I., et al. (2016). The Structure-Specific Endonucleases MUS81 and SEND1 Are Essential for Telomere Stability in Arabidopsis. *Plant Cell* 28, 74–86. doi: 10.1105/tpc.15.00898

One Thousand, and Plant Transcriptomes, I. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R., and Roldan-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* 67, 671–681. doi: 10.1007/s11103-008-9346-0

Osakabe, K., Abe, K., Yamanouchi, H., Takyuu, T., Yoshioka, T., Ito, Y., et al. (2005). Arabidopsis Rad51B is important for double-strand DNA breaks repair in somatic cells. *Plant Mol. Biol.* 57, 819–833. doi: 10.1007/s11103-005-2187-1

Ozturk, N. (2017). Phylogenetic and functional classification of the photolyase/cryptochrome family. *Photochem. Photobiol.* 93, 104–111. doi: 10.1111/php.12676

Parida, A. P., Sharma, A., and Sharma, A. K. (2019). AtMBD4: a methylated DNA binding protein negatively regulates a subset of phosphate starvation genes. *J. Biosci.* 44:14.

Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., et al. (2000). Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5083–5088. doi: 10.1073/pnas.97.10.5083

Park, H. W., Kim, S. T., Sancar, A., and Deisenhofer, J. (1995). Crystal structure of DNA photolyase from *Escherichia coli*. *Science* 268, 1866–1872. doi: 10.1126/science.7604260

Park, S. Y., Vaghchhipawala, Z., Vasudevan, B., Lee, L. Y., Shen, Y., Singer, K., et al. (2015). Agrobacterium T-DNA integration into the plant genome can occur without the activity of key non-homologous end-joining proteins. *Plant J.* 81, 934–946. doi: 10.1111/tpj.12779

Parry, M. A. J., Madgwick, P. J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., et al. (2009). Mutation discovery for crop improvement. *J. Exp. Bot.* 60, 2817–2825.

Pattanayak, V., Ramirez, C. L., Joung, J. K., and Liu, D. R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* 8, 765–770. doi: 10.1038/nmeth.1670

Pegg, A. E. (2011). Multifaceted roles of alkyltransferase and related proteins in DNA repair. DNA damage, resistance to chemotherapy, and research tools. *Chem. Res. Toxicol.* 24, 618–639. doi: 10.1021/tx200031q

Plosky, B. S., and Woodgate, R. (2004). Switching from high-fidelity replicases to low-fidelity lesion-bypass polymerases. *Curr. Opin. Genet. Dev.* 14, 113–119. doi: 10.1016/j.gde.2004.02.002

Pochart, P., Woltering, D., and Hollingsworth, N. M. (1997). Conserved properties between functionally distinct MutS homologs in yeast. *J. Biol. Chem.* 272, 30345–30349. doi: 10.1074/jbc.272.48.30345

Podevin, N., Davies, H. V., Hartung, F., Nogue, F., and Casacuberta, J. M. (2013). Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol.* 31, 375–383. doi: 10.1016/j.tibtech.2013.03.004

Ponferrada-Marin, M. I., Parrilla-Doblas, J. T., Roldan-Arjona, T., and Ariza, R. R. (2011). A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. *Nucleic Acids Res.* 39, 1473–1484. doi: 10.1093/nar/gkq982

Ponferrada-Marin, M. I., Roldan-Arjona, T., and Ariza, R. R. (2009). ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. *Nucleic Acids Res.* 37, 4264–4274. doi: 10.1093/nar/gkp390

Przykorska, A., Solecka, K., Olszak, K., Keith, G., Nawrot, B., and Kuligowska, E. (2004). Wheat (Triticum vulgare) chloroplast nuclease ChSI exhibits 5′ flap structure-specific endonuclease activity. *Biochemistry* 43, 11283–11294. doi: 10.1021/bi049947u

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14.

Puchta, H., and Fauser, F. (2014). Synthetic nucleases for genome engineering in plants: prospects for a bright future. *Plant J.* 78, 727–741. doi: 10.1111/tpj.12338

Ramiro-Merina, A., Ariza, R. R., and Roldan-Arjona, T. (2013). Molecular characterization of a putative plant homolog of MBD4 DNA glycosylase. *DNA Repair* 12, 890–898. doi: 10.1016/j.dnarep.2013.08.002

Rass, U., Compton, S. A., Matos, J., Singleton, M. R., Ip, S. C. Y., Blanco, M. G., et al. (2010). Mechanism of Holliday junction resolution by the human GEN1 protein. *Gene Dev.* 24, 1559–1569. doi: 10.1101/gad.585310

Rea, M., Zheng, W. G., Chen, M., Braud, C., Bhangu, D., Rognan, T. N., et al. (2012). Histone H1 affects gene imprinting and DNA methylation in Arabidopsis. *Plant J.* 71, 776–786. doi: 10.1111/j.1365-313x.2012.05028.x

Reiss, B., Klemm, M., Kosak, H., and Schell, J. (1996). RecA protein stimulates homologous recombination in plants. *Proc. Natl. Acad. Sci. U.S.A.* 93, 3094–3098. doi: 10.1073/pnas.93.7.3094

Rodgers, P. B., Hamilton, W. D., and Adair, J. R. (1999). The therapeutic potential of plant-derived vaccines and antibodies. *Expert Opin. Investig. Drugs* 8, 211–227. doi: 10.1517/13543784.8.3.211

Rohrig, S., Schropfer, S., Knoll, A., and Puchta, H. (2016). The RTR Complex Partner RMI2 and the DNA Helicase RTEL1 Are Both Independently Involved in Preserving the Stability of 45S rDNA Repeats in *Arabidopsis thaliana*. *PLoS Genet.* 12:e1006394. doi: 10.1371/journal.pgen.1006394

Roldan-Arjona, T., and Ariza, R. R. (2009). Repair and tolerance of oxidative DNA damage in plants. *Mutat. Res.* 681, 169–179. doi: 10.1016/j.mrrev.2008.07.003

Roldan-Arjona, T., Ariza, R. R., and Cordoba-Canero, D. (2019). DNA base excision repair in plants: an unfolding story with familiar and novel characters. *Front. Plant Sci.* 10:1055. doi: 10.3389/fpls.2019.01055

Roldan-Arjona, T., Garcia-Ortiz, M. V., Ruiz-Rubio, M., and Ariza, R. R. (2000). cDNA cloning, expression and functional characterization of an *Arabidopsis thaliana* homologue of the *Escherichia coli* DNA repair enzyme endonuclease III. *Plant Mol. Biol.* 44, 43–52.

Sakamoto, A., Lan, V. T., Hase, Y., Shikazono, N., Matsunaga, T., and Tanaka, A. (2003). Disruption of the AtREV3 gene causes hypersensitivity to ultraviolet B light and gamma-rays in Arabidopsis: implication of the presence of a translesion synthesis mechanism in plants. *Plant Cell* 15, 2042–2057. doi: 10.1105/tpc.012369

Sakamoto, A. N. (2019). Translesion synthesis in plants: ultraviolet resistance and beyond. *Front. Plant Sci.* 10:1208. doi: 10.3389/fpls.2019.01208

Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., et al. (2005). Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.* 24, 683–693. doi: 10.1038/sj.emboj.7600519

Santerre, A., and Britt, A. B. (1994). Cloning of a 3-methyladenine-DNA glycosylase from *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2240–2244. doi: 10.1073/pnas.91.6.2240

Santiago, M. J., Alejandre-Duran, E., and Ruiz-Rubio, M. (2006). Analysis of UV-induced mutation spectra in *Escherichia coli* by DNA polymerase eta from *Arabidopsis thaliana*. *Mutat. Res.* 601, 51–60. doi: 10.1016/j.mrfmmm.2006.05.036

Santiago, M. J., Alejandre-Duran, E., and Ruiz-Rubio, M. (2009). Alternative splicing of two translesion synthesis DNA polymerases from *Arabidopsis thaliana*. *Plant Sci.* 176, 591–596. doi: 10.1016/j.plantsci.2009.01.018

Santiago, M. J., Ruiz-Rubio, M., Dio, L. D., Gonzalez-Reyes, J. A., and Alejandre-Duran, E. (2008). Ubiquitous expression of two translesion synthesis DNA polymerase genes in Arabidopsis. *Planta* 227, 1269–1277. doi: 10.1007/s00425-008-0698-0

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., et al. (2011). Function of the DEMETER DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8042–8047. doi: 10.1073/pnas.1105117108

Schuermann, D., Molinier, J., Fritsch, O., and Hohn, B. (2005). The dual nature of homologous recombination in plants. *Trends Genet.* 21, 172–181. doi: 10.1016/j.tig.2005.01.002

Schumann, U., Lee, J. M., Smith, N. A., Zhong, C. C., Zhu, J. K., Dennis, E. S., et al. (2019). DEMETER plays a role in DNA demethylation and disease response in somatic tissues of Arabidopsis. *Epigenetics* 14, 1074–1087. doi: 10.1080/15592294.2019.1631113

Seguela-Arnaud, M., Crismani, W., Larcheveque, C., Mazel, J., Froger, N., Choinard, S., et al. (2015). Multiple mechanisms limit meiotic crossovers: TOP3alpha and two BLM homologs antagonize crossovers in parallel to FANCM. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4713–4718. doi: 10.1073/pnas.1423107112

Shalev, G., Sitrit, Y., Avivi-Ragolski, N., Lichtenstein, C., and Levy, A. A. (1999). Stimulation of homologous recombination in plants by expression of the bacterial resolvase ruvC. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7398–7402. doi: 10.1073/pnas.96.13.7398

Sheppard, T. L., Ordoukhanian, P., and Joyce, G. F. (2000). A DNA enzyme with N-glycosylase activity. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7802–7807. doi: 10.1073/pnas.97.14.7802

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* 15, 207–216. doi: 10.1111/pbi.12603

Shi, L., Kent, R., Bence, N., and Britt, A. B. (1997). Developmental expression of a DNA repair gene in Arabidopsis. *Mutat. Res.* 384, 145–156. doi: 10.1016/s0921-8777(97)00023-2

Shu, Q. Y., Forster, B. P., and Nakagawa, H. (eds) (2012). "Principles and applications of plant mutation breeding," in *Plant Mutation Breeding and Biotechnology*, (Wallingford: CABI), 301–325. doi: 10.1079/9781780640853.0301

Shukla, V. K., Doyon, Y., Miller, J. C., DeKelver, R. C., Moehle, E. A., Worden, S. E., et al. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* 459, 437–441. doi: 10.1038/nature07992

Simko, I. (2016). High-Resolution DNA melting analysis in plant research. *Trends Plant Sci.* 21, 528–537. doi: 10.1016/j.tplants.2016.01.004

Singh, S. K., Roy, S., Choudhury, S. R., and Sengupta, D. N. (2010). DNA repair and recombination in higher plants: insights from comparative genomics of Arabidopsis and rice. *BMC Genomics* 11:443. doi: 10.1186/1471-2164-11-443

Soyk, S., Lemmon, Z. H., Oved, M., Fisher, J., Liberatore, K. L., Park, S. J., et al. (2017). Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* 169, 1142–1155.e12.

Spampinato, C., and Modrich, P. (2000). The MutL ATPase is required for mismatch repair. *J. Biol. Chem.* 275, 9863–9869. doi: 10.1074/jbc.275.13.9863

Staub, J. M., Garcia, B., Graves, J., Hajdukiewicz, P. T. J., Hunter, P., Nehra, N., et al. (2000). High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* 18, 333–338. doi: 10.1038/73796

Stewart, C. N. Jr., Richards, H. A. T., and Halfhill, M. D. (2000). Transgenic plants and biosafety: science, misconceptions and public perceptions. *Biotechniques* 29, 832–836,838–843.

Sugahara, M., Mikawa, T., Kumasaka, T., Yamamoto, M., Kato, R., Fukuyama, K., et al. (2000). Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8. *EMBO J.* 19, 3857–3869. doi: 10.1093/emboj/19.15.3857

Sun, H., He, L., Wu, H., Pan, F., Wu, X., Zhao, J., et al. (2017). The FEN1 L209P mutation interferes with long-patch base excision repair and induces cellular transformation. *Oncogene* 36, 194–207. doi: 10.1038/onc.2016.188

Svendsen, J. M., Smogorzewska, A., Sowa, M. E., O'Connell, B. C., Gygi, S. P., Elledge, S. J., et al. (2009). Mammalian BTBD12/SLX4 Assembles A Holliday Junction Resolvase and Is Required for DNA Repair. *Cell* 138, 63–77. doi: 10.1016/j.cell.2009.06.030

Symington, L. S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet.* 45, 247–271. doi: 10.1146/annurev-genet-110410-132435

Takahashi, S., Sakamoto, A., Sato, S., Kato, T., Tabata, S., and Tanaka, A. (2005). Roles of Arabidopsis AtREV1 and AtREV7 in translesion synthesis. *Plant Physiol.* 138, 870–881. doi: 10.1104/pp.105.060236

Takahashi, S., Sakamoto, A. N., Tanaka, A., and Shimizu, K. (2007). AtREV1, a Y-family DNA polymerase in Arabidopsis, has deoxynucleotidyl transferase activity in vitro. *Plant Physiol.* 145, 1052–1060. doi: 10.1104/pp.107.101980

Talpaert-Borle, M., and Liuzzi, M. (1982). Base-excision repair in carrot cells. Partial purification and characterization of uracil-DNA glycosylase and apurinic/apyrimidinic endodeoxyribonuclease. *Eur. J. Biochem.* 124, 435–440. doi: 10.1111/j.1432-1033.1982.tb06611.x

Tam, S. M., Hays, J. B., and Chetelat, R. T. (2011). Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. *Theor. Appl. Genet.* 123, 1445–1458. doi: 10.1007/s00122-011-1679-4

Tang, K., Lang, Z., Zhang, H., and Zhu, J. K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat. Plants* 2:16169.

Tham, K. C., Kanaar, R., and Lebbink, J. H. G. (2016). Mismatch repair and homeologous recombination. *DNA Repair* 38, 75–83. doi: 10.1016/j.dnarep.2015.11.010

Thresher, R. J., Vitaterna, M. H., Miyamoto, Y., Kazantsev, A., Hsu, D. S., Petit, C., et al. (1998). Role of mouse cryptochrome blue-light photoreceptor in circadian photoresponses. *Science* 282, 1490–1494. doi: 10.1126/science.282.5393.1490

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264.

Townsend, J. A., Wright, D. A., Winfrey, R. J., Fu, F., Maeder, M. L., Joung, J. K., et al. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* 459, 442–445. doi: 10.1038/nature07845

Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., et al. (2011). Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. *Cell* 145, 198–211. doi: 10.1016/j.cell.2011.03.004

Tzfira, T., Li, J., Lacroix, B., and Citovsky, V. (2004). Agrobacterium T-DNA integration: molecules and models. *Trends Genet.* 20, 375–383. doi: 10.1016/j.tig.2004.06.004

Uchiyama, Y., Kimura, S., Yamamoto, T., Ishibashi, T., and Sakaguchi, K. (2004). Plant DNA polymerase lambda, a DNA repair enzyme that functions in plant meristematic and meiotic tissues. *Eur. J. Biochem.* 271, 2799–2807. doi: 10.1111/j.1432-1033.2004.04214.x

Van Marcke, I., and Angenon, G. (2013). Genomic stability in Nicotiana plants upon silencing of the mismatch repair gene MSH2. *Plant Biotechnol. Rep.* 7, 467–480. doi: 10.1007/s11816-013-0285-0

Van Vu, T., Sung, Y. W., Kim, J., Doan, D. T. H., Tran, M. T., and Kim, J. Y. (2019). Challenges and perspectives in homology-directed gene targeting in monocot plants. *Rice* 12:95.

Vitaterna, M. H., Selby, C. P., Todo, T., Niwa, H., Thompson, C., Fruechte, E. M., et al. (1999). Differential regulation of mammalian period genes and circadian rhythmicity by cryptochromes 1 and 2. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12114–12119. doi: 10.1073/pnas.96.21.12114

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350. doi: 10.1146/annurev-arplant-042811-105552

Wang, C., Higgins, J. D., He, Y., Lu, P., Zhang, D., and Liang, W. (2017). Resolvase OsGEN1 Mediates DNA Repair by Homologous Recombination. *Plant Physiol.* 173, 1316–1329. doi: 10.1104/pp.16.01726

Wang, H., and Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell Biosci.* 7:6.

Wang, T. F., Kleckner, N., and Hunter, N. (1999). Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13914–13919. doi: 10.1073/pnas.96.24.13914

Warner, H. R. (1983). Base excision repair in the thermophile *Thermus* sp. strain X-1. *J. Bacteriol.* 154, 1451–1454. doi: 10.1128/jb.154.3.1451-1454.1983

Williams, B. P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet.* 11:e1005142. doi: 10.1371/journal.pgen.1005142

Willing, E. M., Piofczyk, T., Albert, A., Winkler, J. B., Schneeberger, K., and Pecinka, A. (2016). UVR2 ensures transgenerational genome stability under simulated natural UV-B in *Arabidopsis thaliana*. *Nat. Commun.* 7:13522.

Wollenweber, B., Porter, J. R., and Lubberstedt, T. (2005). Need for multidisciplinary research towards a second green revolution. *Curr. Opin. Plant Biol.* 8, 337–341. doi: 10.1016/j.pbi.2005.03.001

Wright, W. D., Shah, S. S., and Heyer, W. D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10524–10535.

Wu, J. L., Wu, C. J., Lei, C. L., Baraoidan, M., Bordeos, A., Madamba, M. R. S., et al. (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol. Biol.* 59, 85–97. doi: 10.1007/s11103-004-5112-0

Wu, L., Chan, K. L., Ralf, C., Bernstein, D. A., Garcia, P. L., Bohr, V. A., et al. (2005). The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *EMBO J.* 24, 2679–2687. doi: 10.1038/sj.emboj.7600740

Wu, S. Y., Culligan, K., Lamers, M., and Hays, J. (2003). Dissimilar mispair-recognition spectra of Arabidopsis DNA-mismatch-repair proteins MSH2*MSH6 (MutSalpha) and MSH2*MSH7 (MutSgamma). *Nucleic Acids Res.* 31, 6027–6034. doi: 10.1093/nar/gkg780

Wyatt, H. D. M., Sarbajna, S., Matos, J., and West, S. C. (2013). Coordinated Actions of SLX1-SLX4 and MUS81-EME1 for Holliday Junction Resolution in Human Cells. *Mol. Cell* 52, 234–247. doi: 10.1016/j.molcel.2013.08.035

Xia, R., Wang, J., Liu, C., Wang, Y., Wang, Y., Zhai, J., et al. (2006). ROR1/RPA2A, a putative replication protein A2, functions in epigenetic gene silencing and in regulation of meristem development in Arabidopsis. *Plant Cell* 18, 85–103. doi: 10.1105/tpc.105.037507

Xu, H., Swoboda, I., Bhalla, P. L., Sijbers, A. M., Zhao, C., Ong, E. K., et al. (1998). Plant homologue of human excision repair gene ERCC1 points to conservation of DNA repair mechanisms. *Plant J.* 13, 823–829. doi: 10.1046/j.1365-313x.1998.00081.x

Xu, J., Hua, K., and Lang, Z. (2019). Genome editing for horticultural crop improvement. *Hortic. Res.* 6:113.

Yamamoto, A., Hirouchi, T., Mori, T., Teranishi, M., Hidema, J., Morioka, H., et al. (2007). Biochemical and biological properties of DNA photolyases derived from utraviolet-sensitive rice cultivars. *Genes Genet. Syst.* 82, 311–319. doi: 10.1266/ggs.82.311

Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., et al. (2014). Overproduction of stomatal lineage cells in Arabidopsis mutants defective in active DNA demethylation. *Nat. Commun.* 5:4062.

Yang, Y., Ishino, S., Yamagami, T., Kumamaru, T., Satoh, H., and Ishino, Y. (2012). The OsGEN-L protein from Oryza sativa possesses Holliday junction resolvase activity as well as 5′-flap endonuclease activity. *J. Biochem.* 151, 317–327. doi: 10.1093/jb/mvr145

Yelina, N. E., Choi, K., Chelysheva, L., Macaulay, M., de Snoo, B., Wijnker, E., et al. (2012). Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. *PLoS Genet.* 8:e1002844. doi: 10.1371/journal.pgen.1002844

Yelina, N. E., Lambing, C., Hardcastle, T. J., Zhao, X., Santos, B., and Henderson, I. R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. *Genes Dev.* 29, 2183–2202. doi: 10.1101/gad.270876.115

Yin, H., Zhang, X., Liu, J., Wang, Y., He, J., Yang, T., et al. (2009). Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase epsilon mutation in Arabidopsis. *Plant Cell* 21, 386–402. doi: 10.1105/tpc.108.061549

Zahid, H. J., Robinson, E., and Kelly, R. L. (2016). Agriculture, population growth, and statistical analysis of the radiocarbon record. *Proc. Natl. Acad. Sci. U.S.A.* 113, 931–935. doi: 10.1073/pnas.1517650112

Zhang, H., Zhang, J., Lang, Z., Botella, J. R., and Zhu, J.-K. (2017). Genome Editing—Principles and Applications for Functional Genomics Research and Crop Improvement. *Crit. Rev. Plant Sci.* 36, 291–309. doi: 10.1080/07352689.2017.1402989

Zhang, J., Xie, S., Zhu, J. K., and Gong, Z. (2016). Requirement for flap endonuclease 1 (FEN1) to maintain genomic stability and transcriptional gene silencing in Arabidopsis. *Plant J.* 87, 629–640. doi: 10.1111/tpj.13224

Zhang, Y., Massel, K., Godwin, I. D., and Gao, C. (2018). Applications and potential of genome editing in crop improvement. *Genome Biol.* 19:210.

Zhang, Y., Wen, C., Liu, S., Zheng, L., Shen, B., and Tao, Y. (2016). Shade avoidance 6 encodes an Arabidopsis flap endonuclease required for maintenance of genome integrity and development. *Nucleic Acids Res.* 44, 1271–1284. doi: 10.1093/nar/gkv1474

Zhu, C., Lu, L., Zhang, J., Yue, Z., Song, J., Zong, S., et al. (2016). Tautomerization-dependent recognition and excision of oxidation damage in base-excision DNA repair. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7792–7797. doi: 10.1073/pnas.1604591113

Zhu, H., Xie, W., Xu, D., Miki, D., Tang, K., Huang, C. F., et al. (2018). DNA demethylase ROS1 negatively regulates the imprinting of DOGL4 and seed dormancy in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 115, E9962–E9970.

Zhu, J., Kapoor, A., Sridhar, V. V., Agius, F., and Zhu, J. K. (2007). The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis. *Curr. Biol.* 17, 54–59. doi: 10.1016/j.cub.2006.10.059

Zhu, J. K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* 43, 143–166. doi: 10.1146/annurev-genet-102108-134205

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Verma, Tandon, Yadav and Gaur. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fgene.2020.567757

Approximate Genome-Based Kernel Models for Large Data Sets Including Main Effects and Interactions

*Jaime Cuevas[1], Osval A. Montesinos-López[2], J. W. R. Martini[3], Paulino Pérez-Rodríguez[4], Morten Lillemo[5] and Jose Crossa[3,4]**

[1] Universidad de Quintana Roo, Chetumal, Mexico, [2] Facultad de Telemática, Universidad de Colima, Colima, Mexico, [3] International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [4] Colegio de Postgraduados, Texcoco, Mexico, [5] Department of Plant Sciences (IPV), Norwegian University of Life Sciences (NMBU), Ås, Norway

The rapid development of molecular markers and sequencing technologies has made it possible to use genomic prediction (GP) and selection (GS) in animal and plant breeding. However, when the number of observations (n) is large (thousands or millions), computational difficulties when handling these large genomic kernel relationship matrices (inverting and decomposing) increase exponentially. This problem increases when genomic × environment interaction and multi-trait kernels are included in the model. In this research we propose selecting a small number of lines $m (m < n)$ for constructing an approximate kernel of lower rank than the original and thus exponentially decreasing the required computing time. First, we describe the full genomic method for single environment (FGSE) with a covariance matrix (kernel) including all n lines. Second, we select m lines and approximate the original kernel for the single environment model (APSE). Similarly, but including main effects and G × E, we explain a full genomic method with genotype × environment model (FGGE), and including m lines, we approximated the kernel method with G × E (APGE). We applied the proposed method to two different wheat data sets of different sizes (n) using the standard linear kernel Genomic Best Linear Unbiased Predictor (GBLUP) and also using eigen value decomposition. In both data sets, we compared the prediction performance and computing time for FGSE versus APSE; we also compared FGGE versus APGE. Results showed a competitive prediction performance of the approximated methods with a significant reduction in computing time. Genomic prediction accuracy depends on the decay of the eigenvalues (amount of variance information loss) of the original kernel as well as on the size of the selected lines m.

Keywords: genomic-enabled prediction, approximate kernels, computing time, genotype × environment interaction, large data sets

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University,
United States

Reviewed by:
Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil
Akio Onogi,
National Agriculture and Food Research Organization (NARO), Japan
Marnin D. Wolfe,
Cornell University, United States

****Correspondence:***
Jose Crossa
j.crossa@cgiar.org
orcid.org/0000-0001-9428-5855

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *30 May 2020*
Accepted: *28 August 2020*
Published: *15 October 2020*

Citation:
Cuevas J, Montesinos-López OA, Martini JWR, Pérez-Rodríguez P, Lillemo M and Crossa J (2020) Approximate Genome-Based Kernel Models for Large Data Sets Including Main Effects and Interactions. Front. Genet. 11:567757. doi: 10.3389/fgene.2020.567757

INTRODUCTION

The rapid development of molecular markers and sequencing technologies has made it possible to use genomic prediction (GP) and selection (GS) in animal and plant breeding (Meuwissen et al., 2001), and practical evidence in plant and animal breeding data has shown that GS provides important prediction accuracy for GS-assisted breeding (Meuwissen et al., 2001; Crossa et al., 2010, 2011; de los Campos et al., 2010; Pérez-Rodríguez et al., 2012).

Additive genetic effects can be predicted directly from marker effects by Ridge Regression best linear unbiased prediction (rrBLUP) (Endelman, 2011) and/or by employing Bayesian inference (Meuwissen et al., 2001), and/or developing the genomic relationship linear kernel matrix (***G***) to fit the GBLUP (VanRaden, 2008). The GBLUP has the advantage of mitigating the high dimension problem and is flexible enough to be extended to more complex situations like incorporating genotype × environment interactions (GE) or studying multi-traits and multi-environments with multi-kernel methods (Jarquín et al., 2014; Lopez-Cruz et al., 2015). The ***G*** of the GBLUP method is a linear kernel (***K***), since it models the additive lineal relationship between lines.

Departures from linearity can be assessed by semi-parametric approaches, such as mixed models with non-additive covariance structure defined in the Reproducing Kernel Hilbert Space (RKHS) framework or more complicated prediction methods such as neural networks (Gianola et al., 2006; Gianola and van Kaam, 2008; de los Campos et al., 2010; González-Camacho et al., 2012; Pérez-Rodríguez et al., 2012). Gianola et al. (2006, 2014) suggested using RKHS regression for semi-parametric, genomic-enabled prediction and pointed out that non-parametric methods such as kernel regression are necessary to reduce the dimension of the parametric space, and to be able to capture complex cryptic interaction among markers. The most commonly used nonlinear kernels in the Reproducing Kernel Hilbert Space (RKHS) (Gianola et al., 2006, 2014) is the Gaussian kernel (GK) that can be expressed dually as a marker effect and interaction effect model (epistasis) (Martini et al., 2020). The Gaussian kernel (GK) for estimating genetic values captures more complex relationships between markers using the Euclidean distance as the dissimilarity between lines based on molecular markers and estimating a bandwidth parameter (h) (de los Campos et al., 2010). Thus, a Gaussian kernel function is $K_h(x_i, x_{i'}) = \exp\left(-hd^2_{ii'}/q\right)$, where $x_i, x_{i'}$ are the marker vectors for the ith and i'th individuals, and q is a scale factor that can be fixed by the user with the idea of reducing the value of h; in general it is a percentile of the squared Euclidean distance $d^2_{ii'}$ for example, the fifth percentile of the squared Euclidean distance $d^2_{ii'}$ (Pérez-Rodríguez et al., 2012), or the 50 percentile used by Crossa et al. (2010).

Standard GS models were extended to multi-environments by assessing genomic × environment interaction (GE) (Burgueño et al., 2012). Jarquín et al. (2014) proposed an extension of the GBLUP that is a type of random effects model where the main effects of markers and environmental covariates (ECs), as well as the interactions between markers and ECs, are introduced using covariance structures that are functions of marker genotypes and ECs. The proposed approach can be interpreted as a random effects model on all the markers, all the ECs, and all the interactions between markers and ECs using a multiplicative operator. Lopez-Cruz et al. (2015) proposed a marker × environment interaction model where the marker effects and genomic values are partitioned into components that are stable across environments (main effects) and others that are environment-specific (interactions); this interaction model is useful when selecting for stability and for adaptation to targeted environments. Consistently, genomic prediction accuracy substantially increased when incorporating GE and marker × environment interaction (Crossa et al., 2017). The marker × environment interaction model has some advantages over previous models; it is easy to implement in standard software for GS like the BGLR (de los Campos and Pérez-Rodríguez, 2018) or the BGGE (Granato et al., 2018), and it can also be implemented with any priors commonly used in GS, including not only shrinkage methods (e.g., GBLUP), but also variable selection methods (that could not be directly implemented under the reaction norm model) (Crossa et al., 2016).

Cuevas et al. (2016) applied the marker × environment interaction GS model of Lopez-Cruz et al. (2015) but modeled not only through the standard linear kernel (GBLUP) but also through a nonlinear Gaussian kernel similar to that used in the Reproducing Kernel Hilbert Space with Kernel Averaging (RKHS KA) (de los Campos et al., 2010) and a Gaussian kernel with the bandwidth estimated through an empirical Bayesian method (Pérez-Elizalde et al., 2015). The methods proposed by Cuevas et al. (2016) were used to perform single-environment analyses and extended to account for GE interaction in wheat and maize data sets. Cuevas et al. (2016) concluded that the higher prediction accuracy of the Gaussian kernel models with the GE model is due to more flexible kernels that allow accounting for small, more complex marker main effects and marker-specific interaction effects.

In the Ridge Regression rrBLUP (Kang et al., 2008; Endelman, 2011), the reduced dimensionality advantages of eigen decomposition were used to estimate the variance components by means of maximum likelihood and/or restrictive maximum likelihood (REML) to compute the genomic-enabled predictions. Pérez-Elizalde et al. (2015) also used eigen decomposition with the marginal maximum likelihood to estimate the genetic and the residual variance components. Pérez-Rodríguez and de los Campos (2014) developed a very useful and efficient statistical software for Bayesian Generalized Linear Regression (BGLR) based on Monte Carlo Markov Chain (MCMC). Granato et al. (2018) also used the spectral decomposition with covariance matrices of exact rank when employing a Bayesian approach.

However, in GP, not only is the number of markers large, but also the number of individuals could be high, thus making the complete kernel matrix difficult to manipulate, and computationally very intensive. This significant increase in the number of observations (individuals) is common when the genomic-enabled prediction model includes genotype × environment interaction (GE) with different and large numbers of lines in each environment (or year) (Jarquín et al., 2014). In these models, the covariance matrices of the main effects and interactions usually have ranks smaller than the number of observations (lines). In these cases, exact low rank matrices are commonly employed, as in rrBLUP (Kang et al., 2008; Endelman, 2011), as well as the Bayesian Genomic Genotype × Environment interaction (BGGE) software of Granato et al. (2018); however, eigen decomposition also has a high computational cost when both the number of observations and markers is large.

An alternative way to deal with large data sets is to use methods such as approximate kernels with the objective of reducing the computational processing time without affecting the genomic prediction accuracy very much. This methodology is commonly used in the framework of machine learning and in the Gaussian process (Rasmussen and Williams, 2006), where the main problem is the large number of observations (e.g., several thousands), whereas the number of covariates (markers) is not that large.

Wang et al. (2015) commented that the method of approximate kernels could be useful for GP when the number of observations is large. The application of GBLUP or GK is practically intractable for deriving the eigen decomposition of large n because of the time scale and the storage capacity. In animal genomic selection, Misztal (2016) proposed a method to approximate a linear kernel relationship matrix using a small size of the original large training population with the objective of facilitating the inversion of the genomic matrix and being able to employ a single-step method when predicting the performance of a large number of animals.

Lately, a number of new ideas and algorithms have addressed the problem of determining input that is relevant for predicting the output, that is, it is possible to develop an efficient predictive model that does use all the large n observations, but approximates the kernels with a low rank. The method of approximate kernels seems to achieve this objective by proposing a simple input that originally had a kernel matrix $\boldsymbol{K}_{n,n}$ of order $n \times n$ from where a smaller sub-matrix is selected, $\boldsymbol{K}_{m,m}$ of order $m \times m$ with the restriction that $m < n$, with the objective of finding an approximate matrix $\boldsymbol{Q}$ of rank m, smaller than the rank of the original matrix (Seeger et al., 2003). That is,

$$\boldsymbol{K} \approx \boldsymbol{Q} = \boldsymbol{K}_{n,m} \boldsymbol{K}_{m,m}^{-1} \boldsymbol{K}_{n,m}^{'}$$

where $\boldsymbol{K}_{m,m}$ is a sub-matrix of the initial $\boldsymbol{K} = \boldsymbol{K}_{n,n}$ and can be constructed with m selected lines with p markers where $\boldsymbol{K}_{n,m}$ is a sub-matrix of $\boldsymbol{K}$ with the relation between the total n lines and the m selected ones. Therefore, $\boldsymbol{Q}$ is an approximation of $\boldsymbol{K}$, but of smaller rank (m), so that computational time is significantly saved when performing the required spectral decomposition or/and inversion. Based on this approximation, a large number of methods have been proposed, such as the projected process of Seeger et al. (2003), which assumes *a priori* that the random effects have a covariance matrix of $\sigma_u^2 \boldsymbol{Q}$. Also, Snelson and Ghahramani (2006) proposed correcting the diagonal of $\boldsymbol{Q}$ in order to propose a method of pseudo points. Furthermore, a similar approximate method was proposed and implemented by Misztal et al. (2014) and Misztal (2016), who employed recursive methods from the joint distribution of the random genetic effects when testing a large amount of animal production. Titsias (2009) proposed a variational perspective that maximizes the lower bound of the exact marginal likelihood by incorporating, as a penalized element, the trace of the differences of matrices $\boldsymbol{K}$, $\boldsymbol{Q}$. Hensman et al. (2013) presented a stochastic variational method and found a lower limit than the one reported by Titsias (2009).

In general, approximate kernel methods could be useful when the size of the training set is large and the construction of the matrices and their manipulations in terms of storage, inversion and decomposition are highly computing intensive and practically prohibitive (Rasmussen and Williams, 2006). On the other hand, the main concern is how the quality of the approximations would be in terms of genomic-enabled prediction. According to Wang et al. (2015), the eigenvalue decomposition of these full matrices decays rapidly, thus favoring the use of these approximations (Rasmussen and Williams, 2006). Based on the previous difficulties in assessing efficient computer-scale time of genomic problems when the number of observations is large, we have adopted an approximate kernel method for large data using a Bayesian approach to be used in genomic-enabled prediction R packages like BGLR (Pérez-Rodríguez and de los Campos, 2014). To test our proposed approximate method, we used two wheat data sets, one of which is relatively small and the other very large. We compared the performance of the proposed approximate kernel versus the full kernel based on the genomic-enabled prediction accuracy, which in turn was measured based on the correlations between the observed and predictive values, the mean squared error and the estimation of the magnitude of the residual error. This method is valid for any kind of kernel; however, in this study we used it only with linear kernels.

MATERIALS AND METHODS

Statistical Models and Methods

We named the conventional GBLUP the "full genomic model" (FG) and the approximation model the "genomic sparse kernel approximation model" (AP). Depending on whether the model is for single-environment (SE) analyses or for GE analyses, FG is called FGSE and FGGE, respectively, and the AP method is called APSE and APGE.

The Full Genomic Method Single-Environment Model (FGSE)

To facilitate the description of this model, we first explain the basic parametric genetic model (assuming the fixed effects have been already considered)

$$\boldsymbol{y} = \mu \mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ is the vector of observations of the response variable of size $n \times 1$, μ is the overall mean, $\boldsymbol{X}$ is the matrix of the p markers on the n lines associated with $\boldsymbol{y}$, and $\boldsymbol{\beta}$ is the vector of the p marker effects, which in the Bayesian framework are considered random effects with normal distribution $N(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_n)$. Finally, random vector $\boldsymbol{\varepsilon}$ has normal distribution $N\left(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}_n\right)$, where σ_ε^2 is the variance component of the random errors and $\boldsymbol{I}_n$ is an identity matrix of order $n \times n$.

The previous model can be represented as a GBLUP model

$$\boldsymbol{y} = \mu \mathbf{1}_n + \boldsymbol{u} + \boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{u}$ is the vector of random effects of size $n \times 1$ with $N(\mathbf{0}, \sigma_u^2 \boldsymbol{K})$, σ_u^2 is a scaled parameter to be estimated and $\boldsymbol{K}$ is a known positive semidefinite matrix of order $n \times n$, constructed based on molecular markers $\boldsymbol{X}$ of order $n \times p$, where p denotes the number of markers such that $\boldsymbol{K} = \dfrac{\boldsymbol{X}\boldsymbol{X}'}{p}$ is known as GBLUP

(VanRaden, 2008; Lopez-Cruz et al., 2015). Note that there is no incidence matrix for $\boldsymbol{u}$ because $\boldsymbol{K}$ is constructed directly using the markers of model (1), which are in line with the response vector $\boldsymbol{y}$.

The eigenvalue decomposition of $\boldsymbol{K}$ is $\boldsymbol{U}\boldsymbol{S}^{1/2}\,\boldsymbol{S}^{1/2}\,\boldsymbol{U}'$, substituting $\boldsymbol{u}$ in model (2), is equivalent to

$$\boldsymbol{y} = \mu\mathbf{1}_n + \boldsymbol{P}\boldsymbol{f} + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{f} \sim N\left(\mathbf{0}, \sigma_f^2\,\boldsymbol{I}_r\right)$, (where r is the rank of $\boldsymbol{K}$) and $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{S}^{1/2}$. Note that models (1), (2) and (3) are equivalent. Models (1) and (3) can be fitted by the conventional Ridge regression model. The Ridge regression model can be computationally fitted very quickly, especially in situations where $r < \min\left(n, p\right)$, which is common in multi-environment and/or multi-trait models. It should be noted that only r effects can be summarized and projected for $\boldsymbol{P}$ to explain the n effects without any loss of precision with the available information.

Genomic Approximate Kernel Methods for a Single-Environment Model (APSE)

First, the method considers $\boldsymbol{K}$, based on a smaller sub-matrix $\boldsymbol{K}_{m,m}(m < n)$ constructed with the markers of m lines. When the row vectors are linearly independent, the rank of $K_{m,m}$ is m. Williams and Seeger (2001) showed that the Nyström approximation of the kernel is as follows:

$$\boldsymbol{K} \approx \boldsymbol{Q} = \boldsymbol{K}_{n,m}\,\boldsymbol{K}_{m,m}^{-1}\,\boldsymbol{K}_{n,m}'$$

where $\boldsymbol{Q}$ will have the rank of $\boldsymbol{K}_{m,m}$, that is m. Note, however, that it is not necessary to compute and store the original matrix $\boldsymbol{K}$, only $\boldsymbol{K}_{m,m}$ and $\boldsymbol{K}_{n,m}$.

In this approximation, $\boldsymbol{K}_{m,m}$ is constructed with m lines with all the p markers, that is, $\boldsymbol{X}_{m,p}$. For the case of the GBLUP, $\boldsymbol{K}_{m,m} = \dfrac{\boldsymbol{X}_{m,p}\boldsymbol{X}_{m,p}'}{p}$ and $\boldsymbol{K}_{n,m} = \dfrac{\boldsymbol{X}_{n,p}\boldsymbol{X}_{m,p}'}{p}$ which captures the relationship of all n lines with the m. Note that in the construction of $\boldsymbol{Q}$, all the p markers and all the n lines are considered, but not all their relationships are accounted for; for example, relationships $\boldsymbol{K}_{n-m,n-m} = \dfrac{\boldsymbol{X}_{n-m,p}\boldsymbol{X}_{n-m,p}'}{p}$ are not considered (where $n-m$ represents all the rest of the m lines). To try to explain this, we ordered the elements of matrix $\boldsymbol{K}$ per blocks, such that $\boldsymbol{K}_{n,n} = \begin{bmatrix} \boldsymbol{K}_{m,m} & \boldsymbol{K}_{m,n-m} \\ \boldsymbol{K}_{n-m,m} & \boldsymbol{K}_{n-m,n-m} \end{bmatrix}$.

Rasmussen and Williams (2006) showed that $\boldsymbol{Q}_{m,m} = \boldsymbol{K}_{m,m}$, $\boldsymbol{Q}_{n-m,m} = \boldsymbol{K}_{n-m,m}$, $\boldsymbol{Q}_{m,n-m} = \boldsymbol{K}_{m,n-m}$, and that the difference between $\boldsymbol{K}_{n-m,n-m} - \boldsymbol{Q}_{n-m,n-m}$, that is, $\boldsymbol{K}_{n-m,n-m} - \boldsymbol{K}_{n-m,m}\,\boldsymbol{K}_{m,m}^{-1}\,\boldsymbol{K}_{m,n-m}$ is the Schur complement of $\boldsymbol{K}_{m,m}$ on $\boldsymbol{K}_{n,n}$. Then, because it is assumed that $\boldsymbol{K}_{m,m}$ and $\boldsymbol{K}_{n,n}$ are positive semidefinite, their Schur complement is also positive semidefinite: $\boldsymbol{Q}_{n,n} = \begin{bmatrix} \boldsymbol{K}_{m,m} & \boldsymbol{K}_{m,n-m} \\ \boldsymbol{K}_{n-m,m} & \boldsymbol{Q}_{n-m,n-m} \end{bmatrix}$. Assuming the effects of $\boldsymbol{u}_{n-m}|\,\boldsymbol{u}_m$ are conditional independent, Snelson and Ghahramani (2006) and Misztal et al. (2014) proposed substituting the diagonal of the differences of $\boldsymbol{Q}_{n-m,n-m}$ with the diagonal of $\boldsymbol{K}_{n-m,n-m}$.

In the method called Projected Process, Seeger et al. (2003) theoretically show that using all lines and considering the minimum Kullback-Leibler distance $KL(q(\boldsymbol{u}|y)||p\left(\boldsymbol{u}|\mathrm{y}\right))$ justifies that matrix $\boldsymbol{K}$ in the prior distribution of $\boldsymbol{u}$ (of model 2) can be substituted for the $\boldsymbol{Q}$ approximations from Nyström (Titsias, 2009). That is, the random genetic vectors have a normal distribution $\boldsymbol{u} \sim N(\mathbf{0}, \sigma_u^2\,\boldsymbol{Q})$, where $\boldsymbol{Q} = \boldsymbol{K}_{n,m}\,\boldsymbol{K}_{m,m}^{-1}\,\boldsymbol{K}_{n,m}'$. More details are given in Csató and Opper (2002).

These adjustments in the distribution of the random effects $\boldsymbol{u}$ of model 2 can be done for genome-based prediction. It is common to estimate parameters σ_ε^2 and σ_u^2 of the model with the marginal likelihood by means of numerical methods and then predict them using the inversion lemma, which is fast when the model is for a single environment. However, the purpose of this study is to develop a methodology in order to jointly estimate and predict complex models such as genotype × environment interactions by making the eigen value decomposition transformation so that it allows us to use ridge regression or Bayesian ridge regression, which can be adjusted with diverse software. Furthermore, if matrix $\boldsymbol{Q}$ is directly used with model (2), the advantages (in terms of speed) of the approximate kernel would not apply. Therefore, similar to model (3), what we did is perform an eigen-decomposition of $\boldsymbol{K}_{m,m}^{-1} = \boldsymbol{U}_{m,m}\boldsymbol{S}_{m,m}^{-1/2}\boldsymbol{U}_{m,m}'$, where $\boldsymbol{U}_{m,m}$ are the eigenvectors of order $m \times m$ and $\boldsymbol{S}_{m,m}$ is a diagonal matrix of order $m \times m$ with the eigenvalues ordered from largest to smallest. These values are substituted in $\boldsymbol{Q}$ resulting in $\boldsymbol{u}_n \sim N(\mathbf{0}, \sigma_u^2\,\boldsymbol{K}_{n,m}\boldsymbol{U}_{m,m}\boldsymbol{S}_{m,m}^{-1/2}\,\boldsymbol{S}_{m,m}^{-1/2}\boldsymbol{U}_{m,m}'\,\boldsymbol{K}_{n,m}')$, and thus, due to the properties of the normal distribution, model (1) could be expressed as:

$$\boldsymbol{y} = \mu\mathbf{1}_n + \boldsymbol{P}\boldsymbol{f} + \boldsymbol{\varepsilon} \tag{4}$$

Model (4) is similar to model (3), except that $\boldsymbol{f}$ is a vector of order $m \times 1$ with a normal distribution of the form $\boldsymbol{f} \sim \boldsymbol{N}(\mathbf{0}, \sigma_f^2\,\boldsymbol{I}_m)$, where $\boldsymbol{P} = \boldsymbol{K}_{n,m}\boldsymbol{U}_{m,m}\boldsymbol{S}_{m,m}^{-1/2}$. This implies estimating only the m effects and expanding them in the n dimensional space in order to predict $\boldsymbol{u}_n$ and explain $\boldsymbol{y}_n$. Note that model (4) has a Ridge regression solution, and thus diverse software can be used.

In summary, the approximation described above consists of the following steps:

Step 1. Compute the matrix $\boldsymbol{K}_{m,m}$ from m lines of the training set. The lines are randomly selected.
Step 2. Construct matrix $\boldsymbol{K}_{n,m}$.
Step 3. Compute the eigenvalue decomposition of $\boldsymbol{K}_{m,m}$.
Step 4. Compute matrix $\boldsymbol{P} = \boldsymbol{K}_{n,m}\boldsymbol{U}_{m,m}\boldsymbol{S}_{m,m}^{-1/2}$.
Step 5. Fit the model and make genomic-enabled predictions with Bayesian Ridge Regression or Ridge Regression.

The Full Genomic Method With the Genotype × Environment Model (FGGE)

The model of Jarquín et al. (2014) including GE is described as

$$\boldsymbol{y} = \mu\mathbf{1}_n + \boldsymbol{e} + \boldsymbol{g} + \boldsymbol{g}\boldsymbol{e} + \boldsymbol{\varepsilon} \tag{5}$$

In this case, the response $\boldsymbol{y}$ is a column vector of size $n \times 1$ comprising observations from k environments, that is, $\boldsymbol{y} = (\boldsymbol{y}_{n1}, \ldots, \boldsymbol{y}_{ni}, \ldots, \boldsymbol{y}_{n_K})'$, where $\boldsymbol{y}_{n_i}$ denotes the vector of observations of the ith environment, and n_i is the number of observations in the ith environment, with $n = \sum_{i=1}^{k} n_i$ the total number of observations in k environments. Also, μ is the overall mean, vector $\boldsymbol{e}$ is a random effects of the environments of size $n \times 1$ with a normal distribution $\boldsymbol{e} \sim N\left(\mathbf{0}, \sigma_e^2 \boldsymbol{Z}^e \boldsymbol{E} \boldsymbol{Z}^{e'}\right)$, where $\boldsymbol{E}$ could be an identity matrix of order $k \times k$ (where k represents the number of environments) or a variance-covariance matrix when some lines are repeated in some environments. Matrix $\boldsymbol{Z}^e$ is the incidence matrix of size $n \times k$ that relates the $\boldsymbol{y}$ observations with the environments. Vector $\boldsymbol{g}$ denotes the genetic random main effects of size $n \times 1$ with normal distribution $\boldsymbol{g} \sim N(\mathbf{0}, \sigma_g^2 \boldsymbol{G})$, where $\boldsymbol{G}$ is a matrix of order $n \times n$, which is usually computed as $\boldsymbol{Z}^g \boldsymbol{K} \boldsymbol{Z}^{g'}$, where $\boldsymbol{Z}^g$ is an incidence matrix that relates the genotypes to the observations and $\boldsymbol{K}$ is the genomic similarity kernel matrix of lines. Vector $\boldsymbol{ge}$ represents the random effect of the genotype $\times$ environment interaction of size $n \times 1$ with a normal distribution $\mathbf{ge} \sim N(\mathbf{0}, \sigma_{ge}^2 \boldsymbol{GE})$, where $\boldsymbol{GE}$ is a known matrix of order $n \times n$. Note that matrix $\boldsymbol{GE}$ can be constructed as $\boldsymbol{G}\# \boldsymbol{Z}^e \boldsymbol{E} \boldsymbol{Z}^{e'}$ where # represents the Hadamard product. The vector of random errors with homogeneous variance is normal $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}_n)$.

Genomic Approximate Kernel Methods With a Genotype × Environment Model (APGE)

We will focus on the main effects of the genotypes and the interaction effects to take advantage of the properties of the approximate kernel. Therefore, the approximate method is similar to the case of a single environment, that is, $\boldsymbol{g} \sim N(\mathbf{0}, \sigma_g^2 \boldsymbol{Q}^g)$, where $\boldsymbol{G} \approx \boldsymbol{Q}^g = \boldsymbol{G}_{n,m} \boldsymbol{G}_{m,m}^{-1} \boldsymbol{G}'_{n,m}$, whereas for the random interaction $\mathbf{ge} \sim N(\mathbf{0}, \sigma_{ge}^2 \boldsymbol{Q}^{ge})$, where $\boldsymbol{GE} \approx \boldsymbol{Q}^{ge} = \boldsymbol{GE}_{n,m} \boldsymbol{GE}_{m,m}^{-1} \boldsymbol{GE}'_{n,m}$.

Similarly, for the approximate method for a single environment, we can decompose $\boldsymbol{G}_{m,m}^{-1}$ and $\boldsymbol{GE}_{m,m}^{-1}$ in such a way that model (5) could be approximated as:

$$\boldsymbol{y} = \mu \mathbf{1}_n + \boldsymbol{e} + \boldsymbol{P}^g \boldsymbol{f} + \boldsymbol{P}^{ge} \boldsymbol{l} + \boldsymbol{\varepsilon} \tag{6}$$

where $\boldsymbol{P}^g = \boldsymbol{G}_{n,m} \boldsymbol{U}_{m,m}^g \boldsymbol{S}_{m,m}^{g^{-\frac{1}{2}}}$, $\boldsymbol{P}^{ge} = \boldsymbol{GE}_{n,m} \boldsymbol{U}_{m,m}^{ge} \boldsymbol{S}_{m,m}^{ge^{-\frac{1}{2}}}$, and vectors $\boldsymbol{f}$, $\boldsymbol{l}$ are of order $m \times 1$.

In summary, the suggested approximate method described above can be implemented with the following steps:

Step 1. Randomly select m lines from the training set, extracting the same number of lines for each environment.
Step 2. To construct matrices $\boldsymbol{G}_{m,m}$ and $\boldsymbol{G}_{n,m}$, one could proceed by ordering matrix $\mathbf{X} = (\mathbf{X}_{n_1,p}, .., \mathbf{X}_{n_i,p}, \ldots, \mathbf{X}_{n_k,p})'$, and constructing $\boldsymbol{G}_{m,m} = \frac{\boldsymbol{X}_{m,p} \boldsymbol{X}'_{m,p}}{p}$, $\boldsymbol{G}_{n,m} = \frac{\boldsymbol{X}_{n,p} \boldsymbol{X}'_{m,p}}{p}$. Another way to proceed is to use matrix $\boldsymbol{K}$, if available, and construct matrices $\boldsymbol{G}_{m,m} = \boldsymbol{Z}_{m,c}^g \boldsymbol{K}_{c,c} \boldsymbol{Z}_{m,c}^{g'}$ and $\boldsymbol{G}_{n,m} = \boldsymbol{Z}_{n,c}^g \boldsymbol{K}_{c,c} \boldsymbol{Z}_{m,c}^{g'}$, where c represents the number of lines without replicates.
Step 3. Construct matrices $\boldsymbol{GE}_{m,m} = \boldsymbol{G}_{m,m}{}^{\#} \boldsymbol{Z}_{k,m}^e \boldsymbol{E}_{k,k} \boldsymbol{Z}_{k,m}^e$, $\boldsymbol{GE}_{n,m} = \boldsymbol{GE}_{n,m}{}^{\#} \boldsymbol{Z}_{n,k}^e \boldsymbol{E}_{k,k} \boldsymbol{Z}_{k,m}^e$, $\boldsymbol{P}^g$, $\boldsymbol{P}^{ge}$.
Step 4. With the previous matrices, model (6) can be fitted and the required genomic-enabled predictions can be obtained.

EXPERIMENTAL DATA

To evaluate the performance of the different methods (FS and AP) and models (SE and GE) (FSSE, FSGE, ASE, and APGE), we used two sets of wheat data; the first data set (**data set 1**) is a large data set and the second is a small data set (**data set 2**).

Data Set 1 – Large Data Set

This data set was used by Pérez-Rodríguez et al. (2020) and comprises 45,099 wheat lines and genotypes with 6978 GBS markers. From the total number of 45,099 wheat lines, 7671, 9021, 9501, 9821 and 9015 wheat lines were evaluated in years 2013–2014, 2014–2015, 2015–2016, 2016–2017 and 2017–2018, respectively. Thus, this data set has 5 environments that represent 5 different years, and the lines in different years are different.

Data Set 2 – Small Data Set

This data set includes the wheat data sets used by Crossa et al. (2010), and comprises 599 wheat lines evaluated in four different environments and genotyped with 1279 SNP markers.

Assessing Prediction Accuracy of the Full Genomic and the Genomic Approximate Kernel models for Single-Environment and for GE

To assess the performance of method-model combinations FGSE and APSE, we used models 3 and 4, respectively, and drew 20 random samples, with 80% of the observations used for training and 20% for testing in each sample. We used all the data and made predictions for single environments for both FGSE and APSE methods. However, for the AP method, we used 5 different sample sizes (m); for **data set 1**, $m = 4000$, $m = 2000$, $m = 1000$, $m = 500$, and $m = 100$. The analyses were performed in each case (FGSE and APSE) for five of the cycles included in this study (**Table 1** and **Figure 2**). For **data set 2**, $m = 264$, $m = 132$, $m = 74$, $m = 32$, $m = 15$ (**Table 2** and **Figure 3**). In addition, **Tables 1** and **2** show the % of variation of matrix $\boldsymbol{K}$ that would be explained by taking the first m eigenvalues from the decomposition of $\boldsymbol{K}$, that is, $\varphi = 100 \times \sum_{i=1}^{m} s_i / \sum_{i=1}^{n} s_i$ (as a measure of the decay of the eigenvalues).

Data set 1 is used for fitting the GE models, FGGE and APGE, using training cycles 2013–2014, 2014–2015, 2015–2016, 2016–2017 to predict cycle 2017–2018. For the FGGE model, it was computationally not possible to fit such a model using a standard laptop (computer 1, laptop) since the size of the

TABLE 1 | Data set 1.

Cycle (n = total number of lines)	Sample size of the training m					
	Model FGSE m **= all** φ **= 100**	**Model APSE** m **= 4000** φ **= 99.7**	**Model APSE** m **= 2000** φ **= 98.6**	**Model APSE** m **= 1000** φ **= 97.2**	**Model APSE** m **= 500** φ **= 96.0**	**Model APSE** m **= 100** φ **= 92.5**
CORR						
Cycle 2017_2018 (n = 9015)	0.575 (0.016)	0.575 (0.016)	0.570 (0.015)	0.557 (0.017)	0.534 (0.017)	0.464 (0.02)
Cycle 2016_2017 (n = 9821)	0.483 (0.011)	0.483 (0.011)	0.477 (0.015)	0.465 (0.013)	0.447 (0.011)	0.386 (0.012)
Cycle 2015_2016 (n = 9501)	0.533 (0.013)	0.533 (0.013)	0.522 (0.014)	0.508 (0.013)	0.483 (0.014)	0.402 (0.016)
Cycle 2014_2015 (n = 9021)	0.494 (0.017)	0.493 (0.012)	0.485 (0.020)	0.470 (0.016)	0.441 (0.018)	0.318 (0.021)
Cycle 2013_2014 (n = 7671)	0.572 (0.015)	0.572 (0.015)	0.567 (0.016)	0.549 (0.015)	0.515 (0.016)	0.366 (0.004)
PMSE						
Cycle 2017_2018 (n = 9015)	0.282 (0.009)	0.282 (0.009)	0.284 (0.008)	0.290 (0.009)	0.300 (0.008)	0.336 (0.01)
Cycle 2016_2017 (n = 9821)	0.369 (0.009)	0.369 (0.010)	0.364 (0.010)	0.377 (0.010)	0.385 (0.010)	0.410 (0.011)
Cycle 2015_2016 (n = 9501)	0.304 (0.010)	0.304 (0.010)	0.309 (0.013)	0.315 (0.010)	0.326 (0.010)	0.356 (0.012)
Cycle 2014_2015 (n = 9021)	0.309 (0.012)	0.309 (0.013)	0.311 (0.011)	0.319 (0.013)	0.329 (0.013)	0.368 (0.016)
Cycle 2013_2014 (n = 7671)	0.413 (0.011)	0.413 (0.013)	0.413 (0.014)	0.429 (0.012)	0.451 (0.012)	0.508 (0.011)
$\hat{\sigma}^2_\varepsilon$						
Cycle 2017_2018 (n = 9015)	0.247 (0.003)	0.250 (0.003)	0.262 (0.002)	0.275 (0.002)	0.293 (0.003)	0.330 (0.004)
Cycle 2016_2017 (n = 9821)	0.317 (0.003)	0.323 (0.003)	0.337 (0.003)	0.350 (0.003)	0.365 (0.003)	0.400 (0.003)
Cycle 2015_2016 (n = 9501)	0.255 (0.003)	0.257 (0.003)	0.279 (0.003)	0.297 (0.003)	0.315 (0.004)	0.357 (0.005)
Cycle 2014_2015 (n = 9021)	0.259 (0.003)	0.266 (0.003)	0.280 (0.003)	0.298 (0.003)	0.315 (0.004)	0.366 (0.004)
Cycle 2013_2014 (n = 7671)	0.313 (0.004)	0.324 (0.005)	0.358 (0.005)	0.391 (0.006)	0.424 (0.006)	0.501 (0.006)
TIME (in seconds)						
Cycle 2017_2018 (n = 9015)	3931	1710	707	345	174	47
Cycle 2016_2017 (n = 9821)	4350	1765	768	356	176	48
Cycle 2015_2016 (n = 9501)	4200	1750	759	375	184	49
Cycle 2014_2015 (n = 9021)	3850	1310	695	330	165	51
Cycle 2013_2014 (n = 7671)	2800	1135	533	247	134	44

Prediction of 20 random samples with 80% of observations in the training set and 20% in the testing set for single site models for the 5 cycles. The FGSE model with a training set of size m = all and the % of variation retained of matrix ***K*** *(*φ*) as* φ *= 100, and APSE for m = 4000, m = 2000, m = 1000, m = 500, and m = 100 wheat lines and* φ *= 99.7,* φ *= 98.6,* φ *= 97.2,* φ *= 96.0, and* φ *= 92.5. Average correlation between predictive and observed values (CORR). Predictive Mean Squared Error (PMSE), and residual variance (*$\hat{\sigma}^2_\varepsilon$*). Computing average time required including the time for preparing matrix* ***K*** *plus the time required for 20,000 iterations using the software R package BGLR (standard deviations for CORR, PMSE, and* $\hat{\sigma}^2_\varepsilon$ *in parentheses).*

training set is a $\boldsymbol{G}$ matrix of order 45099 × 45099. Therefore, we used the results from Pérez-Rodríguez et al. (2020; Chapter 13, Table 13.4), who used the same training data to predict cycle 2017–1018. These authors achieved a genomic-enabled prediction accuracy of 0.4263 using only markers. The prediction of the same cycle (2017–2018) used the approximate APGE model with only 25% of the total training lines from each cycle for the m, that is, matrices $\boldsymbol{K}_{m,m}$, $\boldsymbol{K}_{n,m}$ are manageable matrices of order 9021 × 9021 and 45099 × 9021, respectively. **Table 3** shows the genomic prediction of each cycle taking one or more of the previous cycles as training. For model APGE, we used 25% of the total training set for each cycle as the size of m. For fitting model FGGE, we used another computer facility (computer 2) because the laptop (computer 1) could not fit the models.

For the small data set, **data set 2** is predicted with the rest of the environments using the full genomic FGGE (model 5); the variance-covariance matrices are of order 2396 × 2396, a size that does not cause any computational problem. For APGE, m was 25% of the training set of each environment (representing a total of 450 wheat lines, that is, 150 lines in each of the three environments used for training). **Table 4** shows the variance component estimates for model APGE in **data set 2**.

As criteria for all model-method combinations (FGSE, APSE, FGGE, APGE) used to evaluate the prediction accuracy and computing time, we employed: (1) the mean Pearson's correlation between the predictive and the observed values (CORR), where the predictive values are extracted from the mode of the Bayesian predictive distribution; (2) the prediction mean squared error PMSE is the mean of the squared difference between the predictive and the observed value; (3) the fitted models with the residual error variance ($\hat{\sigma}^2_\varepsilon$); and (4) the time (TIME) for constructing the matrices and fitting the model (**Tables 1**–**3** and **Figures 1**, **2**). For model APGE, we estimated the variance components of the main effects σ^2_g, the interaction effects σ^2_{ge} and random error σ^2_ε (**Table 4**).

Software

To fit the models we used Bayesian Ridge Regression from BGLR (de los Campos and Pérez-Rodríguez, 2018), because it is a free software that focuses on genomic predictions, and it is flexible, allowing users to fit complex models including multi-kernels,

TABLE 2 | Data set 2.

Environment	Sample size of m					
	Model FGSE m **= all** φ **= 100**	**Model APSE** m **= 264** φ **= 98.6**	**Model APSE** m **= 132** φ **= 95.8**	**Model APSE** m **= 72** φ **= 95.6**	**Model APSE** m **= 36** φ **= 88.8**	**Model APSE** m **= 15** φ **= 82.2**
CORR						
E1	0.506 (0.046)	0.501 (0.047)	0.468 (0.063)	0.425 (0.073)	0.362 (0.060)	0.266 (0.088)
E2	0.471 (0.068)	0.461 (0.062)	0.439 (0.066)	0.407 (0.071)	0.374 (0.060)	0.283 (0.072)
E3	0.384 (0.046)	0.384 (0.047)	0.381 (0.059)	0.359 (0.053)	0.318 (0.064)	0.267 (0.068)
E4	0.448 (0.051)	0.439 (0.05)	0.420 (0.048)	0.398 (0.053)	0.359 (0.050)	0.302 (0.053)
PMSE						
E1	0.771 (0.074)	0.776 (0.047)	0.806 (0.075)	0.848 (0.085)	0.899 (0.086)	0.957 (0.088)
E2	0.751 (0.08)	0.761 (0.078)	0.782 (0.081)	0.809 (0.092)	0.834 (0.077)	0.891 (0.090)
E3	0.821 (0.085)	0.817 (0.082)	0.822 (0.098)	0.837 (0.087)	0.863 (0.087)	0.892 (0.090)
E4	0.802 (0.098)	0.811 (0.090)	0.827 (0.096)	0.844 (0.097)	0.873 (0.096)	0.912 (0.090)
$\hat{\sigma}^2_\varepsilon$						
E1	0.523 (0.041)	0.572 (0.038)	0.656 (0.035)	0.733 (0.037)	0.819 (0.037)	0.890 (0.040)
E2	0.587 (0.039)	0.635 (0.041)	0.707 (0.036)	0.768 (0.037)	0.840 (0.041)	0.902 (0.046)
E3	0.602 (0.039)	0.691 (0.043)	0.768 (0.048)	0.823 (0.041)	0.877 (0.045)	0.930 (0.048)
E4	0.598 (0.046)	0.652 (0.044)	0.720 (0.040)	0.775 (0.041)	0.833 (0.038)	0.890 (0.044)
TIME (in seconds)						
TE1	17	13.7	11	10.9	9.25	8.6
E2	17	13.7	11	10.9	9.25	8.6
E3	17	13.7	11	10.9	9.25	8.6
E4	17	13.7	11	10.9	9.25	8.6

Prediction of 20 random samples with 80% of observations in the training set and 20% in the testing set for single site models for the 5 cycles. The FGSE model with the training set of size m = all and the % of variation retained of matrix $\boldsymbol{K}$ *(*φ*) as* φ *= 100, and APSE for m = 264, m = 132, m = 72, m = 36, and m = 100 wheat lines and* φ *= 98.6,* φ *= 95.8,* φ *= 95.6,* φ *= 88.8, and* φ *= 82.2. Average correlation between predictive and observed values (CORR). Predictive Mean Squared Error (PMSE), and residual variance (*$\hat{\sigma}^2_\varepsilon$*). Computing average time required including the time for preparing matrix* $\boldsymbol{K}$ *plus the time required for 20,000 iterations using the software R package BGLR (standard deviations for CORR, PMSE, and* $\hat{\sigma}^2_\varepsilon$ *in parentheses).*

main effects and G × E effects. BGLR is very well documented with a large number of clearly explained examples that can be found in https://github.com/gdlc/BGLR-R.

Models were fitted and predictions were made using 20,000 iterations and discarding the first 3000 iterations and using a thinning of 2. Initially the Raftery and Lewis (1992) criteria was employed to determine the minimum of iterations, the "burn in" and the "thin." Also we made visual observations of graphs representing the Monte Carlo Markov Chain to make sure a good mixture was achieved.

Hardware

Computer 1 is a laptop with a processor intel® Core i5™ i5-7300 HQ CPU@ 2.5 GHz 2.5 GHz, RAM 16 GB, Operative System of 64 bit, with processor x64.

Computer 2, vendor_id : AuthenticAMD, cpu family : 16, model: 9, model name;: AMD Opteron(tm) Processor 6140, stepping : 1, microcode : 0x10000c4, cpu MHz : 2600.185, cache size : 512 KB.

Data Repository

The 5 phenotypic and genotypic data sets (cycle 13–14, cycle 14–15, cycle 15–16, cycle 16–17, and cycle 17–18) comprising **data set 1** can be downloaded from the following link: http://hdl.handle.net/11529/10548425. As already mentioned, **data set 2** can be found in Crossa et al. (2010), or as an illustrative example in the BGLR R package (de los Campos and Pérez-Rodríguez, 2018) or in a large number of other genomic-based studies that have used this experimental data set.

RESULTS

Results of FGSE and APSE for Large (Data Set 1) and Small (Data Set 2) Data

For large **data set 1**, **Table 1** and **Figure 2** show the prediction accuracy of 20 random cross-validation partitions, where in each sample, 20% of the wheat lines are predicted from a training set of 80% of the total wheat lines for the 5 cycles. The first column contains the results of the FGSE (model 3) using all wheat lines in each cycle (m = all). It shows the average correlation (CORR) of the 20 random samples of 20% of the wheat lines in the testing set, as well as the mean of the 20 PMSEs and the mean of the 20 estimations of the residuals ($\hat{\sigma}^2_\varepsilon$). Finally, it shows the TIME invested in each sample of the training-testing combination for 20,000 iterations. Columns 2–6 in **Table 1** provide the results of CORR, PMSE, $\hat{\sigma}^2_\varepsilon$, and TIME for m = 4000, m = 2000, m = 1000, m = 500, m = 100 wheat lines of APSE (model 4), randomly selected in order to compute $\boldsymbol{K}_{m,m}$, and $\boldsymbol{K}_{n,m}$.

The behavior of the cycles is similar for FGSE and APSE for 4000 wheat lines for $\boldsymbol{K}_{m,m}$, $\boldsymbol{K}_{n,m}$, but genomic-enabled prediction values are lost as the number of lines included in the

TABLE 3 | The models FGGE and APGE considering the size of m, as 25% of the original training set.

Cycle	Training	CORR	PMSE	$\hat{\sigma}^2_\varepsilon$	TIME (h)
Data set 1 modelFGGE (using *computer 2*)					
Cycle 2014_2015	Cycle 2013_2014	0.222	2.45	0.317	4.96
Cycle 2015_2016	Cycle 2013_2014	0.328	0.525	0.287	11.10
	Cycle 2014_2015				
Cycle 2016_2017	Cycle 2013_2014	0.328	0.480	0.275	23.72
	Cycle 2014_2015				
	Cycle 2015_2016				
Cycle 2017_2018	Cycle 2013_2014	0.426	NA	NA	NA
	Cycle 2014_2015				
	Cycle 2015_2016				
	Cycle 2016_2017				
Data set 1 model APGE (using *computer 1*)					
Cycle 2014_2015	Cycle 2013_2014	0.206	1.08	0.363	0.68
Cycle 2015_2016	Cycle 2013_2014	0.347	0.408	0.309	2.80
	Cycle 2014_2015				
Cycle 2016_2017	Cycle 2013_2014	0.321	0.517	0.29	5.08
	Cycle 2014_2015				
	Cycle 2015_2016				
Cycle 2017_2018	Cycle 2013_2014	0.427	0.618	0.301	8.38
	Cycle 2014_2015				
	Cycle 2015_2016				
	Cycle 2016_2017				
Environment	**Training**	**CORR**	**PMSE**	$\hat{\sigma}^2_\varepsilon$	**TIME (s)**
Data set 2 model FGGE (using *computer 1*)					
E1	E2	-0.166	1.520	0.532	175
	E3				
	E4				
E2	E1	0.511	0.912	0.600	178
	E3				
	E4				
E3	E1	0.469	0.879	0.577	180
	E2				
	E4				
E4	E1	0.311	.940	0.570	187
	E2				
	E3				
Data set 2 model APGE (using *computer 1*)					
E1	E2	-0.188	1.54	0.607	70
	E3				
	E4				
E2	E1	0.491	0.942	0.71	72
	E3				
	E4				
E3	E1	0.445	0.887	0.70	73
	E2				
	E4				
E4	E1	0.281	0.960	0.651	82
	E2				
	E3				

Average correlation between predictive and observed values (CORR), Predictive Mean Squared Error (PMSE), and residual variance ($\hat{\sigma}^2_\varepsilon$). Computing average time required including the time for preparing the matrices G, GE plus the time required for 20,000 iterations using the software R package BGLR.

TABLE 4 | Estimated variance components for model APGE for **data set 1** and **data set 2**.

Testing	Training	$\hat{\sigma}^2_\varepsilon$	$\hat{\sigma}^2_g$	$\hat{\sigma}^2_{ge}$
Data set 1				
Cycle 2014–2015	Cycle 2013–2014	0.3624	0.4680	0.3300
Cycle 2015–2016	Cycle 2013–2014	0.3087	0.2638	0.3337
	Cycle 2014–2015			
Cycle 2016–2017	Cycle 2013–2014	0.2916	0.22705	0.2956
	Cycle 2014–2015			
	Cycle 2015–2016			
Cycle 2017–2018	Cycle 2013–2014	0.3019	0.1886	0.2962
	Cycle 2014–2015			
	Cycle 2015–2016			
	Cycle 2016–2017			
Data set 2				
E1	E1	0.6070	0.3953	0.5576
	E3			
	E4			
E2	E1	0.7102	0.3183	0.1120
	E3			
	E4			
E3	E1	0.7001	0.3053	0.1356
	E2			
	E4			
E4	E1	0.6510	0.2981	0.1985
	E2			
	E3			

training set is reduced; this is reflected in the decrease of the CORR, and the increase in PMSE and ($\hat{\sigma}^2_\varepsilon$). For example, for cycle 2017_2018, FGSE with all observations had a CORR of 0.575, a PMSE of 0.282, and an estimated $\hat{\sigma}^2_\varepsilon$ of 0.247. Interestingly, these results are similar to those found for the APSE when only 4000 wheat lines were used as training (55% of the total original training set), with a CORR of 0.575, a PMSE of 0.282 and an estimated $\hat{\sigma}^2_\varepsilon$ of 0.250. Furthermore, when APSE used only 2000 wheat lines as training (28% of the total original training set), the genomic-enabled prediction accuracy slightly decreased to a CORR of 0.570, and the PMSE had a small increase with PMSE = 0.254 as a result of a less fitted value $\hat{\sigma}^2_\varepsilon = 0.262$ (**Table 1**).

The genomic-enabled prediction decreases for smaller sample sizes (m) of 1000, 500, and 100, where CORR takes values of 0.557, 0.534, and 0.46, respectively, increasing PMSE to 0.290, 0.300, and 0.336, as well as the estimated $\hat{\sigma}^2_\varepsilon$ values to 0.275, 0.293, and 0.330, respectively. The computing TIME decreases almost linearly (3931, 1710, 707, 345, 174, 47 seconds) for the decreasing sample size (m). The results of the different sample sizes of m and the correlations from **Table 1** (**data set 1**) are also displayed in **Figure 1A** where, for example, for cycle 2017–2018 for m/n = 0.22, the average correlation for the genomic-enabled prediction is 0.570, whereas for m/n = 1.0, the average correlation is 0.575. It is interesting to observe that the computational time required decreases linearly as the size of m decreases in relation to the size of n (**Figure 1B**).

The results of the small **data set 2** shown in **Table 2** and **Figure 3** have the same structure as those shown in **Table 1** and

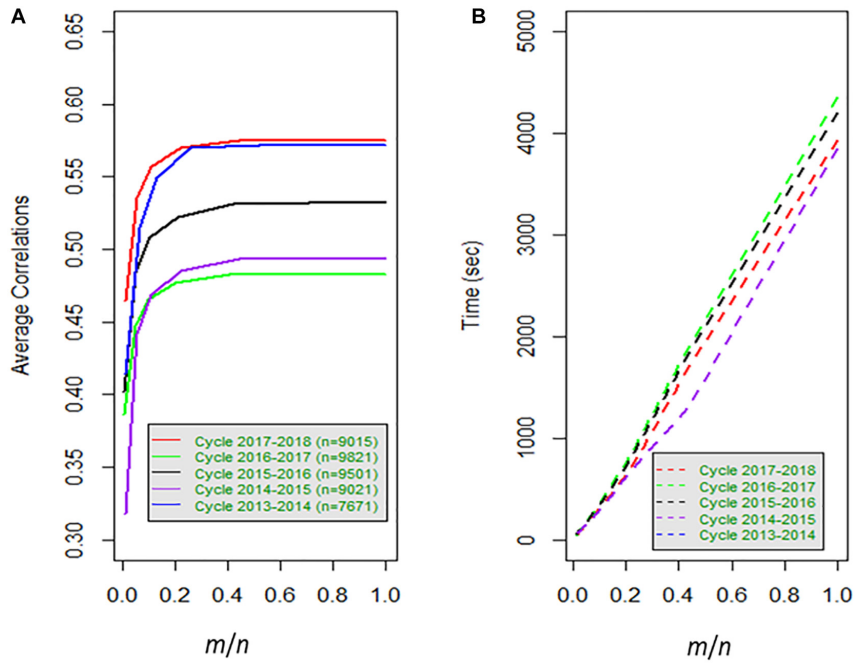


FIGURE 1 | (A) Average correlation for 80% training and 20% testing for 20 random samples for data set1, versus the proportion of size *m* with respect to the total number of observations (lines) *n*; **(B)** time in seconds for each sample versus the proportion of size of *m* over the total number of lines (*n*).

Figure 2 for **data set 1**; however, the results are different. These data have 599 wheat lines all evaluated in 4 environments (E1, E2, E3, and E4) (Crossa et al., 2010). Each of the 20 random samples with 479 wheat lines in the training set and 120 lines in the testing set had varying results; however, compared to the results obtained with the large data set (**data set 1**), these results are quite different. The first column shows the results of the full genomic model (FGSE model 3) using all the data, and when compared with the APSE (model 4) with $m = 264$ (55% of the total training population), the CORR decreased slightly in 3 of the 4 environments; for example, in E1 it decreased from 0.506 to 0.501, whereas in E2 went from 0.47 to 0.461, it stayed the same in E3 and decreased in E4 from 0.448 to 0.439. Similar patterns were found for PMSE and $\hat{\sigma}^2_\varepsilon$.

When $m = 132$ lines (28% of the total original size of the training population), the decrease in CORR was severe in E1 and E2, decreasing to 0.468 and 0.439, respectively, but less so in E3, where it decreased to 0.381 as a consequence of a decrease in the fit with $\hat{\sigma}^2_\varepsilon$ of 0.656, 0.707, 0.766 and 0.720, respectively, in E1, E2, E3 and E4. The decreasing trend in CORR increased as m decreased; for example, in E1, when $m = 74$, or 36 or 15, CORR was 0.425, 0.362 and 0.262, respectively, and $\hat{\sigma}^2_\varepsilon$ increased to 0.730, 0.819 and 0.890. However, in contrast to data set 1, the mean computing time (TIME) for each of the 20 samples of the random cross-validation did not decrease in the same proportion as those due to the size of the sample.

Tables 1 and **2** and **Figure 1** indicate that the differences in genomic prediction with respect to the full models depend more on the size of m, that is, the larger the m, the smaller the differences with the full model (m = all). Another important indicator is φ, because when $\varphi > 98$, the genomic-enabled prediction accuracy of the approximate model is equal to that of the full models; when $\varphi < 98$, the results of the approximate models are less precise than those obtained from the full model.

Results of FGGE and APGE for Large (Data Set 1) and Small (Data Set 2) Data

Table 3 shows the genomic-enabled prediction accuracy for models FGGE y APGE for the two groups of data. To predict cycle 2017–2018 from **data set 1** using the previous 4 cycles with the full genomic GE model (FGGE, model 5), it is necessary to manipulate two large covariance matrices, one for the main effects of the genomic (G) model and another matrix for the interaction (GE) of order 45099 × 45099. It was not possible to manage this matrix size with the current conventional laptop (computer 1) used to analyze these data; therefore, we used the genomic-enabled prediction accuracy recently reported by Pérez-Rodríguez et al. (2020) as a reference. The authors used and reported a genomic prediction accuracy of 0.426 for cycle 2017–2018 using all the other cycles as a training set.

Using the approximate model APGE (model 6) and only 25% of the total training set, that is, $m = 9021$, such that matrices $\boldsymbol{K}_{m,m}$, and $\boldsymbol{K}_{n,m}$, are now of manageable sizes of order 9021 × 9021 and 45099 × 9021, respectively, this gives a genomic prediction accuracy of 0.427, with a residual variance of 0.302, that is, there is no loss of genomic prediction accuracy with respect to the full genomic models with GE (FGGE model 5). The computing time required, including the time

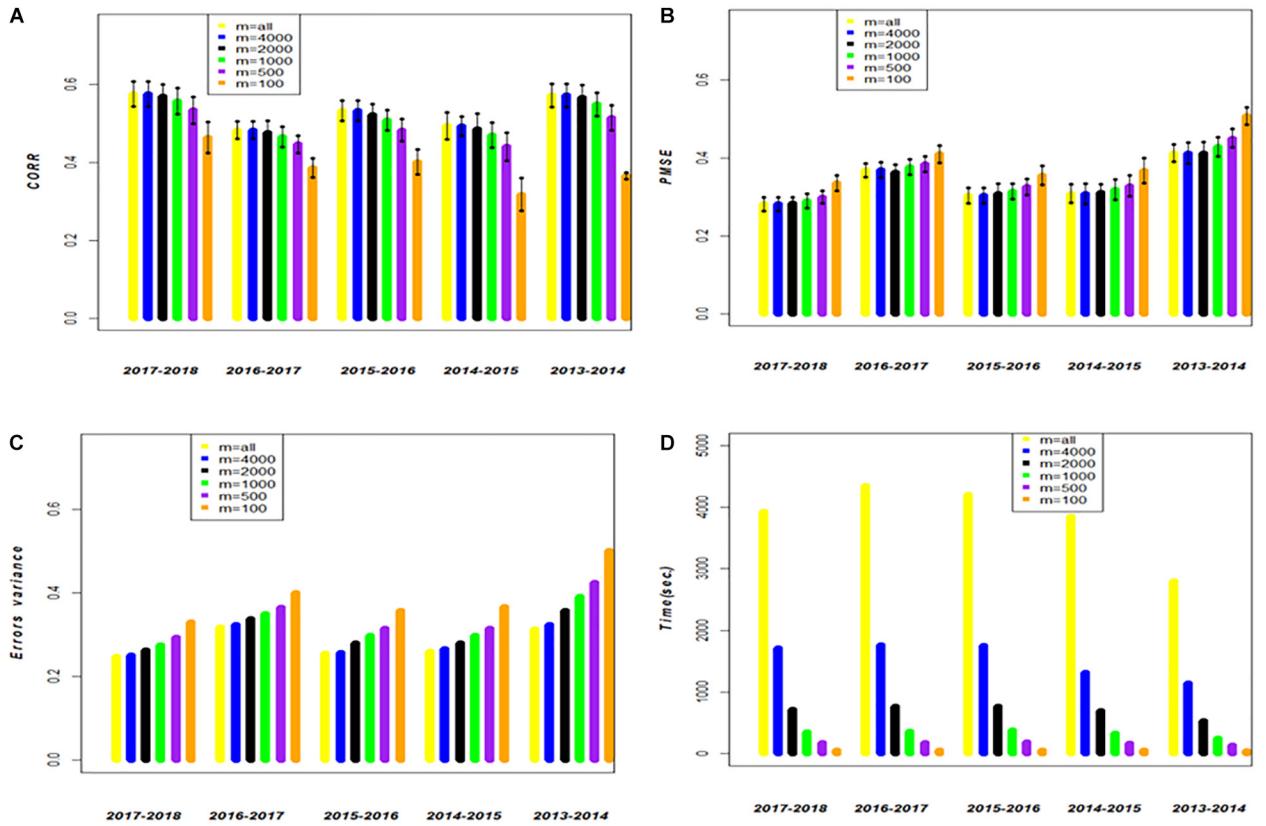


FIGURE 2 | Data set 1. Models FGSE (yellow, m = all lines) and APSE (blue, black, green, purple and orange that correspond to m = 4000, m = 2000, m = 1000, m = 500, m = 100), **(A)** average correlation between observed and predictive values of FGSE and APSE models at different sizes of m; bars indicated 2 standard deviations) **(B)** average prediction mean squared error (PMSE) values of FGSE and APSE models at different sizes of m, **(C)** error variance of FGSE and APSE models ($\hat{\sigma}^2_\varepsilon$) at different sizes of m, and **(D)** time in seconds to fit FGSE and APSE models at different sizes of m.

for preparing the matrices for the approximation method, and the time for the eigenvalue decomposition and the 20,000 iterations, was 30,670 seconds. This was very similar for the prediction of the other cycles, and the only differences were in the computing time consumed between FGGE and APGE; this difference exponentially increased with the total number of training data.

When we used **data set 2** to predict environment E4 using environments E1, E2, and E3 as training and using FGGE, the required covariance matrices were of order 2396 × 2396, which does not pose any problems for their storage and manipulation. The prediction accuracy achieved by the FGGE for the genomic-enabled prediction of E4 was 0.311, with a PMSE of 0.94, a residual variance $\hat{\sigma}^2_\varepsilon$ of 0.57, and a duration time of 187 seconds. When using the approximate model APGE (model 6), we selected 25% of the training set (480 wheat lines) and found a decrease in the genomic prediction accuracy of 0.281 compared to the FGGE, an increase in the PMSE of 0.960, and an increase in the residual variance with respect to model FGGE of $\hat{\sigma}^2_\varepsilon$= 0.651, with a faster computing time (82 seconds) than model FGGE. When predicting the other environments, the results were similar regarding the differences in the correlations between models FGGE and APGE.

Table 4 shows the estimated variance components for model APGE. It can be observed that for **data set 1**, the variance components for the main effects and the interactions were of similar magnitude, indicating the importance of both types of effects. For **data set 2**, the interaction variance component is relatively smaller than the main effects.

DISCUSSION

The main objective of this study was to show that the approximate kernel method offers a good solution for the large data sets usually encountered in genomic-enabled prediction when Bayesian linear mixed models need to be fitted. The usual problem in genomic prediction is that the number of markers (covariates, p) is much larger than the number of observations (n). However, the number of observations is also large, so performing matrix decomposition requires very intense computing in terms of time, storage capacity, etc. Approximate kernels allow matrix manipulation and storage, thus saving storage resources and computing processing time. In some cases, genomic prediction accuracy does not decrease much, but in other cases, the loss of precision is indeed important. This depends mainly on the size of

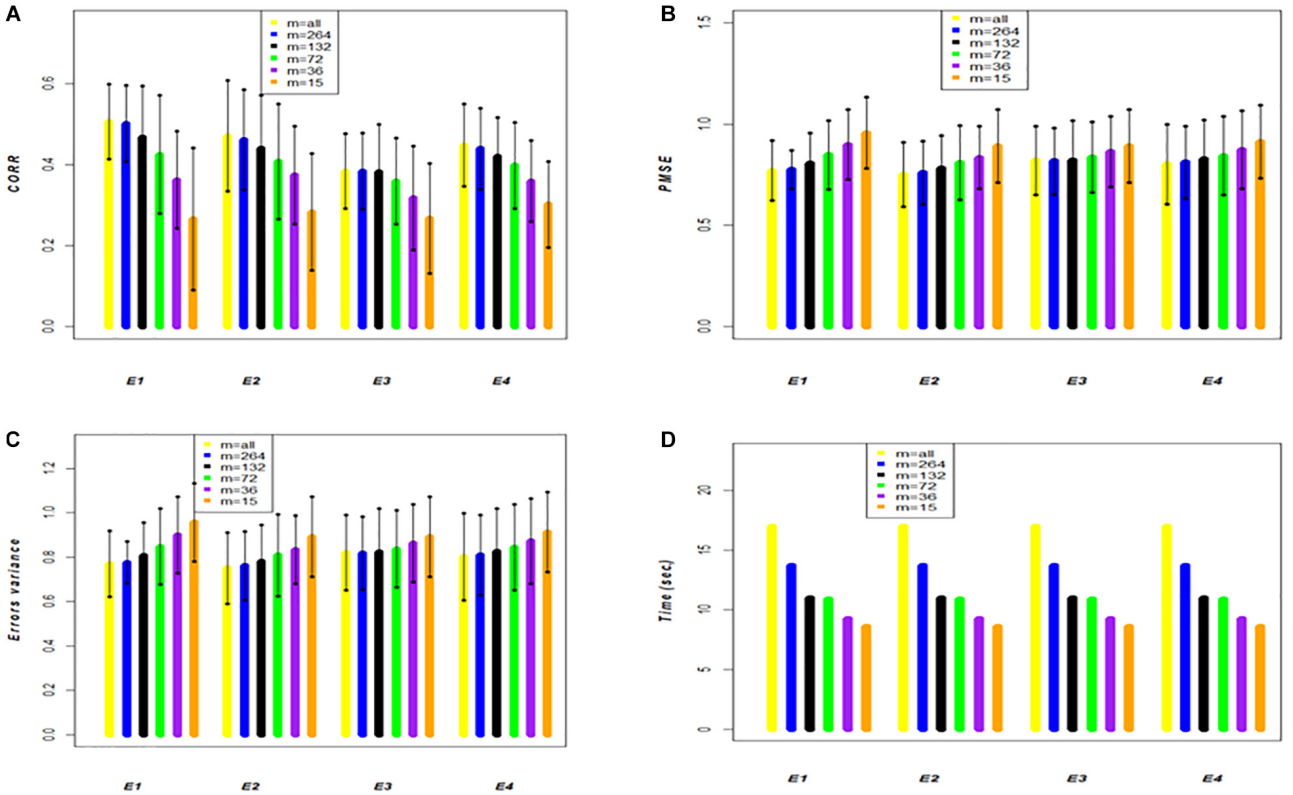


FIGURE 3 | Data set 2. Models FGSE (yellow, m = all lines) and APSE (blue, black, green, purple and orange that correspond to m = 4000, m = 2000, m = 1000, m = 500, m = 100), **(A)** average correlation between observed and predictive values of FGSE and APSE models at different sizes of m; bars indicated 2 standard deviations) **(B)** average prediction mean squared error (PMSE) values of FGSE and APSE models at different sizes of m, **(C)** error variance of FGSE and APSE models ($\hat{\sigma}^2_\varepsilon$) at different sizes of m, and **(D)** time in seconds to fit FGSE and APSE models at different sizes of m.

m and on how fast the decrease in the eigenvalue decomposition of kernel $\boldsymbol{K}$ occurred. A rapid decrease in the eigenvalues indicates that with only a few singular values, a high percentage of the important variation could be retained. The variance retained using φ (Pocrnic et al., 2016) indicated the percentage of variation retained for a certain number of eigenvalues.

Data sets 1 and **2** were fitted using the full genomic (FG) method and the approximate model (AP) for the single-environment model with certain percentages of the points selected from the total training set similar in the two data sets (55, 28, 14, 7, and 2% of the total training set). The size of m influenced the precision of the predictions. In **data set 1**, the genomic prediction accuracy was higher at 55 and 28% and slowly declined as the size of m decreased; this decrease in the prediction accuracy was smoother in **data set 1** than in **data set 2.** One of the reasons for these differences in prediction accuracy between the two data sets could be due to the rank of kernel $\boldsymbol{K}$. For example, in **data set 1** for cycle 2017–2018, kernel $\boldsymbol{K}$ (of order 9015 $\times$ 9015) had a rank of 7017, whereas for **data set 2**, the rank of matrix $\boldsymbol{K}$ was 598; that is, **data set 1** had more degrees of freedom than **data set 2**. A common feature of both data sets is the rapid decline in the singular values of their kernels; this is measured by φ as the percentage of variance retained by $\boldsymbol{K}$ using a certain number of singular values (size of m). The empirical results suggested using $\varphi > 98$ to avoid losing precision. This result is in agreement with that suggested by Misztal (2016). This could be used as a rule of thumb to select the minimum size of m that would return a $\varphi > 98$.

The rapid decline in the singular value of kernel $\boldsymbol{K}$ favors the use of the approximate kernel $\boldsymbol{Q}$, as suggested by Wang et al. (2015). Therefore, the rapid decline in the singular value of kernel $\boldsymbol{K}$ also favors the use of other methods that improve the computer speed, such as principal component regression using the original matrix $\boldsymbol{K}$. However, if the data are large, intense computational efforts are required to construct matrix $\boldsymbol{K}$, with an exponential requirement of computing capacity for eigenvalue decomposition. On the other hand, the approximate method requires a matrix of much lower order. When using an m associated with $\varphi > 98$, we do not expect significant differences in the prediction accuracy of the approximate model and the full model; also, no differences between the approximate model and the principal component regression model are expected using a similar size of m; however, when $\varphi \ll 98$, more differences are expected between the approximate model and the full genomic models but less with the principal component regression model.

In relation with the necessary computing time, the AP method applied to **data set 1** showed that the saving of computing time increases when the size of m decreases, whereas for **data**

set 2, this also occurs but in different proportions because the data are of much lower dimension than those in **data set 1**. In general, the results of this study indicated that the computing time used to fit the full model increases exponentially with the number of observations n; this also applies to the approximate models. These results are in agreement with those of Wang et al. (2015), who commented that "most kernel-based methods have a computational complexity of order $O(n^3)$. This is prohibitive when we have large-scale training samples. The low-rank spectral reconstruction of a kernel can be performed by the Nyström method, which can speed up many regression-oriented algorithms. The approximation quality of these methods is protected by a reasonable and key assumption that the genomic data, like most other large data, live in a lower dimension space and the spectra of the kernel matrices often decay quickly."

Figure 2A (**data set 1**) and **Figure 3A** (**data set 2**) display the predictions of the different years (cycles) for the FGSE (yellow, m = all lines) and APSE (blue m = 4000 lines; black m = 2000 lines; green, m = 1000 lines; purple, m = 500 lines and orange m = 100 lines). The pattern of the predictions are kept similar (2 times the standard deviations) for both models, FGSE and APSE, but changing the average correlations based on the size of m indicating congruence among the predictions of the 20 samples of different sizes of m that were randomly selected to form the training set. Also, it can be observed that for FGSE and APSE models, the sizes of m (yellow, blue and black, m = all lines m = 4000 lines and black m = 2000 lines, respectively) did not change the prediction accuracy of the unobserved wheat lines in the testing set much. In addition, note that in **Figures 2C**, **3C**, the residual variance increased as the size of m decreased, indicating that the AP model does not produce overfitting.

It is indeed in the GE models where approximate kernels could have the greatest utility because the covariance matrix (G) of the main effects of markers and the GE are, in general, large matrices and the fit of the models is very slow computationally. The fit of model APGE for **data set 1**(**Table 3**) did not lose prediction accuracy when fitted with approximate kernels of lower rank as compared with the ones required by the FGGE, with an important reduction in the computing time. The APGE reduced the time required to prepare the matrices and to fit the model with 20,000 iterations to 8.5 h, when it takes days on a big server. For data set 2, the results were not that good; nevertheless, the precision did not decrease much, but the reduction in time was important.

Table 4 shows the variance components of the two data sets for model APGE. The magnitude of the variance components shows that the model captured the main effects as well as the interactions. Although φ is a good indicator for explaining the relationship between the decay of the singular values, unfortunately it is not always possible to estimate the decrease in the prediction accuracy and the adequate size of m.

Using the approximate kernel of this study, authors like Seeger et al. (2003); Snelson and Ghahramani (2006) and Titsias (2009) show examples with large numbers of observations (n), while the covariates (p) are continuous and of low dimensions. The $n \gg p$ implies the existence of redundant information (more degrees of freedom available); this allows using approximate kernels or a sparse Gaussian process (Rasmussen and Williams, 2006) in a very efficient manner. All these propositions emphasize the size of m, but also indicate which observations to choose. To deal with the selection of observations, some authors propose selecting those that minimize the trace of the matrix differences between the original matrix $\boldsymbol{K}$ and the approximate matrix $\boldsymbol{Q}$ (Rasmussen and Williams, 2006). Other authors propose maximizing the marginal likelihood based on the variational inference (Titsias, 2009; Hensman et al., 2013), where m observations are considered hyper-parameters. Nevertheless, for the linear mixed models used in genomic prediction, the high number of covariates (markers) may require investing important additional computing time for selecting the observations comprising m. On the other hand, empirical results show that selecting the observations at random (**Tables 1**–**3**) works all right because the main constraint is the size of m. These results are in line with the approximate kernel developed in animal breeding by means of pedigree and genomic selection for determining the breeding values performance of large numbers of animals (Misztal, 2016). However, in plant breeding, methods for efficiently selecting the observations comprising m need to be studied further, probably by selecting m lines using population substructure and diversity criteria such as the ones proposed by Akdemir (2014); Jeong et al. (2017).

CONCLUSION

The approximate kernel methods used in this study are very promising because they allow a significant reduction in computing time and data manipulation of large data sets, without significant loss of prediction performance.

Results of model APSE for **data set 1** show a good performance on the genomic-enabled prediction accuracy compared with the full models with APSE employing an important decrease in computing time with respect to the full model. This can be explained by the rapid decrease in the singular values and their ability to capture important information, since with only 25% of the singular values, 98% of the total information was retained. For **data set 2**, model APSE does not have the same prediction performance as for **data set 1.** On average, genomic-enabled prediction accuracies decreased rapidly when the size of m decreased; however, the variability of the predictions was maintained with respect to the full model. In **data set 2**, the decay of the singular values was less rapid than that observed for **data set 1**, that is, 25% of the singular values retained 95% of the information.

For the very large **data set 1**, the results of model APGE with the size of m representing only 25% of the total number of lines gave an excellent correlation between predictive and observed values, along with an important saving of computing time. For the small **data set 2**, the APGE model gave better results than model APSE, and the decrease in the correlation was less compared to that of the full model when 25% of the total lines were used in m. In both data sets, the APGE model with fairly large G × E interactions of the variance components indicates that this variability will indeed increase the genomic-enabled

prediction accuracy with respect to models that only include the main effects.

We also observed that the larger the data sets are, the more benefits can be obtained from the approximate kernel methods. However, for their successful implementation, two important factors should be taken into account: (a) the number of lines (m) that need to be used for approximating the kernel, and (b) the amount of information that can be retained in the approximate kernel (φ). According to our empirical study, we observed that for large data sets, a $\varphi > 98\%$ and a size of $m > 50\%$ of the total training observations are required for single-environment analyses, and $m > 25\%$ for GE analyses prevent important decreases in genomic-enabled prediction accuracy while obtaining time savings in computing resources.

Results of this study indicated that the proposed approximation could be an alternative to genomic prediction when the number of observations is large and the construction and storage of the large kernel matrices is difficult and it takes excessive computing time to fit models FGSE and FGGE. Regarding φ, although it is a good indicator of the variance retained by the singular values and thus for determining the adequate size of m, unfortunately, in practice it is not possible to compute it. Therefore, further research on this subject is needed for selecting the size of m. However, the results obtained are promising because they provide a partial solution to an important problem of genome-based prediction models.

DATA AVAILABILITY STATEMENT

The 5 phenotypic and genotypic data sets (cycle 13–14, cycle 14–15, cycle 15–16, cycle 16–17, and cycle 17–18) comprising data set 1 can be downloaded from the following link: http://hdl.handle.net/11529/10548425. As already mentioned, data set 2 can be found in Crossa et al. (2010), or as an illustrative example in the BGLR R package (de los Campos and Pérez-Rodríguez, 2018) or in a large number of other genomic based studies that have used this experimental data set.

AUTHOR CONTRIBUTIONS

JCu developed the idea, ran the analyses, and wrote the article. JCr discussed the original idea, wrote the manuscript, and helped JCu to present the new idea in several data set from CIMMYT. JM discussed the original ideas, and contributed with new insights when writing the article. PP-R contributed with the computing codes for running the software. OM-L contributed to reading and completing the models. ML read the manuscript and edited parts of the several drafts. All authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge the financial support provided by the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) in Norway through NFR grant 267806. We are also thankful for the financial support provided by CIMMYT CRP (maize and wheat), the Bill & Melinda Gates Foundation, as well the USAID projects (Cornell University and Kansas State University) that generated the CIMMYT wheat data analyzed in this study.

ACKNOWLEDGMENTS

We thank all scientists, field workers, and lab assistants from National Programs and CIMMYT who collected the data used in this study.

REFERENCES

Akdemir, D. (2014). Training population selection for (breeding value) prediction.*arXiv*[Preprint]. arXiv:1401.7953

Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.*52, 707–719. doi: 10.2135/cropsci2011.06.0299

Crossa, J., de los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J., and Pérez-Rodríguez, P. (2016). Extending the marker × environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Sci.*56, 2193–2209.

Crossa, J., De Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*186, 713–724. doi: 10.1534/genetics.110.118521

Crossa, J., Pérez, P., de los Campos, G., Mahuku, G., Dreisigacker, S., and Magorokosho, C. (2011). Genomic selection and prediction in plant breeding. *J. Crop Improv.*25, 239–261.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.*22, 961–975. doi: 10.1016/j.tplants.2017.08.011

Csató, L., and Opper, M. (2002). "Sparse online gaussian process," in *Neural Computing Research Group*. Technical Report NCRG/2001/014.

Cuevas, J., Crossa, J., Soberanis, V., Perez-Elizalde, S., PérezRodríguez, P., de los Campos, G., et al. (2016). Genomic prediction of genotype × environment interaction kernel regression models. *Plant Genome*9, 1–20. doi: 10.3835/plantgenome2016.03.0024

de los Campos, G., Gianola, D., Rosa, G. J., Weigel, K. A., and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res.*92, 295–308. doi: 10.1017/S0016672310000285

de los Campos, G., and Pérez-Rodríguez, P. (2018). *BGLR: Bayesian Generalized Linear Regression. R Package Version 1.0.8*. Available online at: https://CRAN.R-project.org/web/packages/BGLR/BGLR.pdf

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*173, 1761–1776. doi: 10.1534/genetics.105.049510

Gianola, D., and van Kaam, J. B. C. H. M. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*178, 2289–2303. doi: 10.1534/genetics.107.084285

Gianola, D., Weigel, K. A., Krämer, N., Stella, A., and Schön, C. C. (2014). Enhancing genome-enabled prediction by bagging genomic BLUP. *PLoS One*9:e91693. doi: 10.1371/journal.pone.0091693

González-Camacho, J. M., de los Campos, G., Pérez-Rodríguez, P., Gianola, D., Cairns, J. E., Mahuku, G., et al. (2012). Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor. Appl. Genet.*125, 759–771. doi: 10.1007/s00122-012-1868-9

Granato, I., Cuevas, J., Luna, F., Crossa, J., Montesinos-López, O., Burgueño, J., et al. (2018). BGGE: a new package for genomic-enabled prediction incorporating genotype × environment interaction models. *G3 Genes Genom. Genet.*8, 3039–3047. doi: 10.1534/g3.118.200435

Hensman, J., Fusi, N., and Lawrence, N. (2013). "Gaussian processes for big data," in *Proceedings of the 29th Conference, UAI 2013, Uncertainty in Artificial Intelligence*, Bellevue, WA.

Jarquín, D., Crossa, J., Lacaze, X., Cheyron, P. D., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.*127, 595–607. doi: 10.1007/s00122-013-2243-1

Jeong, S., Kim, J., Jeong, S., Kang, S., Moon, S., and Kim, N. (2017). GenoCore: a simple and fast algorithm for core subset selection from large genotype datasets. *PLoS One*12:e0181420. doi: 10.1371/journal.pone.0181420

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics*178, 1709–1723. doi: 10.1534/genetics.107.080101

Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., et al. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3 Genes Genom. Genet.*5, 569–582. doi: 10.1534/g3.114.016097

Martini, J. W. R., Toledo, F. H., and Crossa, J. (2020). On the approximation of interaction effect models by Hadamard powers of the additive genomic relationship. *Theor. Population Biol.*132, 16–23. doi: 10.1016/j.tpb.2020.01.004

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*157, 1819–1829.

Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics*202, 401–409. doi: 10.1534/genetics.115.182089

Misztal, I., Legarra, A., and Aguilar, I. (2014). Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.*97, 3943–3952. doi: 10.3168/jds.2013-7752

Pérez-Elizalde, S., Cuevas, J., Pérez-Rodríguez, P., and Crossa, J. (2015). Selection of the bandwidth parameter in a bayesian kernel regression model for genomic-enabled prediction. *J. Agric. Biol. Environ. Stat.*20, 512–532. doi: 10.1007/s13253-015-0229-y

Pérez-Rodríguez, P., Burgueño, J., Montesinos-López, O. A., Singh, S., Juliana, J., Mondal, S., et al. (2020). "Prediction with big data in the genomic and high-throughput phenotyping era: a case study with wheat data," in *Quantitative Genetics, Genomics and Plant Breeding*, 2nd Edn, ed. M. S.Kang (Boston, MA: CABI International).

Pérez-Rodríguez, P., and de los Campos, G. (2014). Genome-wide regression & prediction with the BGLR statistical package. *Genetics*198, 483–495. doi: 10.1534/genetics.114.164442

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J. M., Crossa, J., Manès, Y., and Dreisigacker, S. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes Genom. Genet.*2, 1595–1605. doi: 10.1534/g3.112.003665

Pocrnic, I., Lourenco, D. A., Masuda, Y., Legarra, A., and Misztal, I. (2016). The dimensionality of genomic information and its effect on genomic prediction. *Genetics*203, 573−581. doi: 10.1534/genetics.116.187013

Raftery, A. E., and Lewis, S. M. (1992). "How many iterations in the gibbs sampler?," in *Bayesian Statistics*, Vol. 4, eds J. M.Bernardo, J. O.Berger, A. P.Dawid, and A. F. M.Smith (Oxford: Oxford University Press), 763–773.

Rasmussen, C. E., and Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Seeger, M., Williams, C. K. I., and Lawrence, N. (2003). "Fast forward selection to speed up sparse gaussian process regression," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics*, eds C.Bishop and B. J.Frey (Key West, FL: Congress Ninth).

Snelson, E., and Ghahramani, Z. (2006). "Local and global sparse Gaussian process approximations," in *Proceedings of the Eleven International Workshop on Artificial Intelligence and Statistics, Society for Artificial Intelligence and Statistics*, eds M.Meilia and X.Shen (Madison, WI: Omnipress).

Titsias, M. K. (2009). "Variational learning of inducing variables in sparse Gaussian Processes," in *Proceedings of the Eleven International Workshop on Artificial Intelligence and Statistics, volume 5*, eds D.van Dyk and M.Welling (Clearwater Beach, FL: JMLR W&CP), 567.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.*91, 4414–4423. doi: 10.3168/jds.2007-0980

Wang, X., Xing, E. P., and Schaid, D. J. (2015). Kernel methods for large-scale genomic data analysis. *Brief. Bioinform.*16, 183–192. doi: 10.1093/bib/bbu024

Williams, C. K. I., and Seeger, M. (2001). "Using the nyström method to speed up kernel machines," in *Advances in Neural Information Processing Systems*, Vol. 13, eds T. K.Leen, T. G.Diettrich, and V.Tresp (Cambridge, MA: MIT Press), 682–688.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cuevas, Montesinos-López, Martini, Pérez-Rodríguez, Lillemo and Crossa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 23 October 2020
doi: 10.3389/fgene.2020.01001

Breeding, Genetics, and Genomics Approaches for Improving Fusarium Wilt Resistance in Major Grain Legumes

Uday Chand Jha[1], Abhishek Bohra[1]*, Shailesh Pandey[2] and Swarup Kumar Parida[3]*

[1] ICAR-Indian Institute of Pulses Research, Uttar Pradesh, India, [2] Forest Protection Division, Forest Research Institute, Dehradun, India, [3] National Institute of Plant Genome Research, New Delhi, India

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University, United States

Reviewed by:
Nicolas Rispail,
Institute for Sustainable Agriculture, Spanish National Research Council, Spain
Siddhesh B. Ghag,
UM-DAE Centre for Excellence in Basic Sciences, India
Lekha T. Pazhamala,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

***Correspondence:**
Uday Chand Jha
uday_gene@yahoo.co.in;
u9811981@gmail.com
Abhishek Bohra
abhi.omics@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *11 May 2020*
Accepted: *06 August 2020*
Published: *23 October 2020*

Citation:
Jha UC, Bohra A, Pandey S and Parida SK (2020) Breeding, Genetics, and Genomics Approaches for Improving Fusarium Wilt Resistance in Major Grain Legumes. Front. Genet. 11:1001. doi: 10.3389/fgene.2020.01001

Fusarium wilt (FW) disease is the key constraint to grain legume production worldwide. The projected climate change is likely to exacerbate the current scenario. Of the various plant protection measures, genetic improvement of the disease resistance of crop cultivars remains the most economic, straightforward and environmental-friendly option to mitigate the risk. We begin with a brief recap of the classical genetic efforts that provided first insights into the genetic determinants controlling plant response to different races of FW pathogen in grain legumes. Subsequent technological breakthroughs like sequencing technologies have enhanced our understanding of the genetic basis of both plant resistance and pathogenicity. We present noteworthy examples of targeted improvement of plant resistance using genomics-assisted approaches. In parallel, modern functional genomic tools like RNA-seq are playing a greater role in illuminating the various aspects of plant-pathogen interaction. Further, proteomics and metabolomics have also been leveraged in recent years to reveal molecular players and various signaling pathways and complex networks participating in host-pathogen interaction. Finally, we present a perspective on the challenges and limitations of high-throughput phenotyping and emerging breeding approaches to expeditiously develop FW-resistant cultivars under the changing climate.

Keywords: Fusarium wilt, genomics, molecular markers, genetic variability, grain legume

INTRODUCTION

Grain legumes being a rich source of dietary proteins and essential minerals serve as one of the vital components of human food (Graham and Vance, 2003; Bohra et al., 2015). Besides, grain legumes also supply essential micronutrients to the human population for combating increasing malnutrition related problems worldwide (Mudryj et al., 2014). Global legume production is severely challenged by a variety of fungal diseases (Kaiser et al., 2000), of which wilt caused by *Fusarium oxysporum* is one of the most destructive (Wade, 1929; Haware et al., 1978; Armstrong and Armstrong, 1981; Reddy et al., 1990; Kraft et al., 1998; Smith et al., 1999; Fall et al., 2001). It ranks fifth among top 10 fungal pathogens of research and economic importance (Dean et al., 2012). The fungus is soil-borne, and occurs as pathogenic (plant, animal, and human) as well as non-pathogenic strains (Leslie and Summerell, 2006). The pathogenic strains have been assigned to

forma specialis (plural: *formae speciales*), abbreviated f. sp. (plural ff. spp.) based on host specificity (Baayen et al., 2000; Lievens et al., 2008). For instance, chickpea (*Cicer arietinum* L.) is affected by *F. oxysporum* f. sp. *ciceris* (*Foc*). Interestingly, *Foc* can also invade root tissues of other grain legumes, such as *Vicia faba*, *Lens culinaris*, *Pisum sativum,* and *Cajanus cajan* without causing external symptoms (Jiménez-Díaz et al., 2015). Presently Index Fungorum[1] lists 124 special forms, whereas MycoBank[2] lists 127 special forms of *F. oxysporum*. Following entry of Fusarium wilt (FW) pathogen through plant root, its colonization in the vascular system disrupts plant root-water continuum, leading to wilting symptoms and death of plant (Schäfer, 1994). Upon exposure to the wilt pathogen, plants recognize pathogen-associated molecular patterns (PAMPs) through their receptor protein pattern recognition receptors (PRRs) known as pattern triggered immunity (PTI) and pathogen effector triggered immunity (ETI), two important mechanisms for averting FW attacks (Zipfel and Robatzek, 2010; Langner et al., 2018). Among the various strategies devised to control FW disease, developing host plant resistance through breeding remains the most straightforward, economic, and sustainable approach. Therefore, physiological pathotypes or races, the subspecific ranks applied to *formae speciales* based on cultivar specificity, are extremely important to breeders for resistance breeding. However, race labels have been incoherent (Gerlagh and Blok, 1988; Correll, 1991; Kistler, 1997; Fourie et al., 2011) as numerous different race designation systems have been applied (Gabe, 1975; Risser et al., 1976; Armstrong and Armstrong, 1981) that created confusion (Kistler, 1997). Nevertheless, recent technological advances in molecular biology helped to overcome many of these hurdles. For instance, sequence-characterized amplified region (SCAR) markers facilitated race identification (Lievens et al., 2008; Epstein et al., 2017; Gilardi et al., 2017). However, laborious and prolonged pathogenicity tests are still required for the identification of new emerging races and resistance testing of the newly developed cultivars against the known races (Epstein et al., 2017; Gilardi et al., 2017). Understanding the genetic makeup of host plant resistance is crucial in this regard. In grain legumes, initial studies based on Mendelian genetics have elucidated a number of gene(s)/genetic determinants underlying resistance against FW (Wade, 1929; Hare et al., 1949; Upadhyaya et al., 1983a,b; Pandey et al., 1996; Coyne et al., 2000; Cross et al., 2000; Fall et al., 2001; Kotresh et al., 2006). Subsequent advances in genomics accelerated molecular mapping of FW resistance gene(s)/QTLs. To this end, availability of whole genome sequences of both plant and pathogen has shed deep insights into the host-pathogen relationship through leveraging comparative genomics approach (Varshney et al., 2012, 2013; Schmutz et al., 2014; Williams et al., 2016; Srivastava et al., 2018; Kreplak et al., 2019; Lonardi et al., 2019). In this review, we discuss the targeted improvement of plant resistance against FW using genomics-assisted approaches. The contributions of functional genomics toward delineating genomic regions/candidate gene(s) responsible for FW resistance in major grain legumes are highlighted. The potential of new omics approaches is discussed with respect to molecular players, signaling pathways, and complex networks underlying host-pathogen interactions. Finally, we present a perspective on high-throughput disease screening and emerging novel breeding techniques for developing FW-resistant cultivars under the changing climate.

[1]http://www.indexfungorum.org/

[2]http://www.mycobank.org/

MODE OF FW INFECTION AND POSSIBLE MECHANISM OF HOST PLANT RESISTANCE AGAINST FW

Fusarium oxysporum is considered saprophytic because of its ability to survive on soil organic matter for several years (Alabouvette et al., 1993). The fungus survives in soil through producing chlamydospores that may serve as a reservoir of inoculums (Schippers and Van Eck, 1981). Subsequent to penetrating plant root epidermis, the pathogen invades the xylem vessels to cause wilting symptoms (Srinivas et al., 2019). The complete molecular mechanism of FW pathogenesis remains to be elucidated (Rep and Kistler, 2010). Genome-wide transcriptome profiling of conidial germination of one of the most virulent Indian races (race 1) of *Foc* has revealed germination-related genes and families of genes encoding secreted effectors, cell wall/pectin-degrading enzymes, metabolism related enzymes, transporters, and peptidases. Importantly, qRT-PCR confirmed the up-regulation of metabolism related enzymes during early infection, whereas up-regulation of most transporters and secondary metabolites important for tissue colonization and pathogenicity was confirmed during later stages (Sharma et al., 2016b).

Following establishment of the pathogen on plant roots, root penetration, and hyphal propagation of the FW pathogen causes a compromise in the host defense system (for detail see Rep and Kistler, 2010; Srinivas et al., 2019). At molecular level, the fungal pathogen recognizes a particular host and produces a range of cell wall-degrading enzymes including cellulases, pectinases, polygalacturonases, etc., in response to host plant derived hydrolases (viz., chitinase, glucanases) (Michielse et al., 2009; Swarupa et al., 2014; Husaini et al., 2018). Besides, the FW pathogen is also known to produce various mycotoxins/phytotoxins viz., fusaric acid (FSA), beauvericin, and enniatins in banana (López-Díaz et al., 2017; Liu et al., 2020; Shao et al., 2020). In parallel, the attacking pathogen integrates various signal transduction pathways mediated by mitogen activated protein (MAP) kinase cascades that transduce the signal downstream to the intracellular targets in response to the signal perceived by various receptors at cell surface during host infection (Widmann et al., 1999; Husaini et al., 2018). Thus, MAP plays a key role in regulating FW pathogenicity. Several genes are known to regulate host colonization and pathogenicity, which include *FWO1* (Inoue et al., 2002), *ClcI* (Cañero and Roncero, 2008), *chitin synthase V*, *DCW1*, *mannose- 6- phosphate isomerase* gene, *FOXG_11097* (Michielse et al., 2009), *XlnR* (Calero-Nieto et al., 2007), secreted in xylem (SIX) protein encoding genes

(Thatcher et al., 2016), *SIX1, SIX6,* and *FTF1* genes (Ninþo-Saìnchez et al., 2015) (for details see Husaini et al., 2018). Thus, successful disease occurrence by the FW pathogen demands a compromise in the plant defense system.

The PTI and the ETI activated by the PAMPs and pathogen effectors, respectively constitute the two tiers of the plant defense system. The PTI, presenting the first line of plant defense, ensues by the recognition of PAMPs *via* the host receptor protein PRRs (Zipfel and Robatzek, 2010; Beck et al., 2012; Lanubile et al., 2015). Following this, the plant evokes oxidative burst and ion influx that transduce the signals to different pathways by triggering down-stream signaling networks mediated by the protein kinases viz., mitogen-activated protein kinases (MAPK) and protein phosphorylation (Nakagami et al., 2005; Pedley and Martin, 2005; Boudsocq et al., 2010; Rodriguez et al., 2010; Bigeard et al., 2015; Lanubile et al., 2015). This is accompanied by the activation of multiple TFs that switch on various defense responsive genes including PR genes and the genes involved in hormone biosynthesis and signaling, protein and sugar metabolism (Castillejo et al., 2015; Lanubile et al., 2015; Kumar et al., 2016). As shown in **Figure 1**, once the PTI fails, the ETI forms the second line of immune response in which the plant defends itself against pathogen attacks through immune receptors encoded by the nucleotide binding leucine rich repeat (NB-LRR) class of R genes, thus enabling recognition of the effector molecules (Zipfel and Robatzek, 2010; Bigeard et al., 2015; Ma and Ma, 2016). This in turn triggers the plant innate immunity that inhibits the pathogen attack (Dodds and Rathjen, 2010). These resistance genes are overcome by the more virulent races of pathogen; hence, a broad-spectrum and durable host resistance is highly needed (Yin and Qiu, 2019; Li et al., 2020).

FW RACES AND THEIR EFFECTS ON MAJOR GRAIN LEGUMES

Chickpea

Fusarium oxysporum f. sp. *ciceris* is prevalent in all major chickpea growing countries including India, Iran, Peru, Syria, Ethiopia, Mexico, and the United States (Nene et al., 1989; Halila and Strange, 1996). Up to 40% yield losses have been reported in chickpea due to FW, and the disease may lead to complete crop failure under congenial environment (Haware et al., 1978, 1986; Sharma et al., 2014). In recent times, FW has emerged as a severe threat to chickpea production because of a shift in chickpea area from long prevailing cool season of northern region to warm regions of southern and central India (Sharma and Pande, 2013).

Early wilting is manifested in the form of dull green discoloration within 25 days after sowing accounts for 77–94% yield loss (Jiménez-Díaz et al., 2015). In the case of "late wilt"

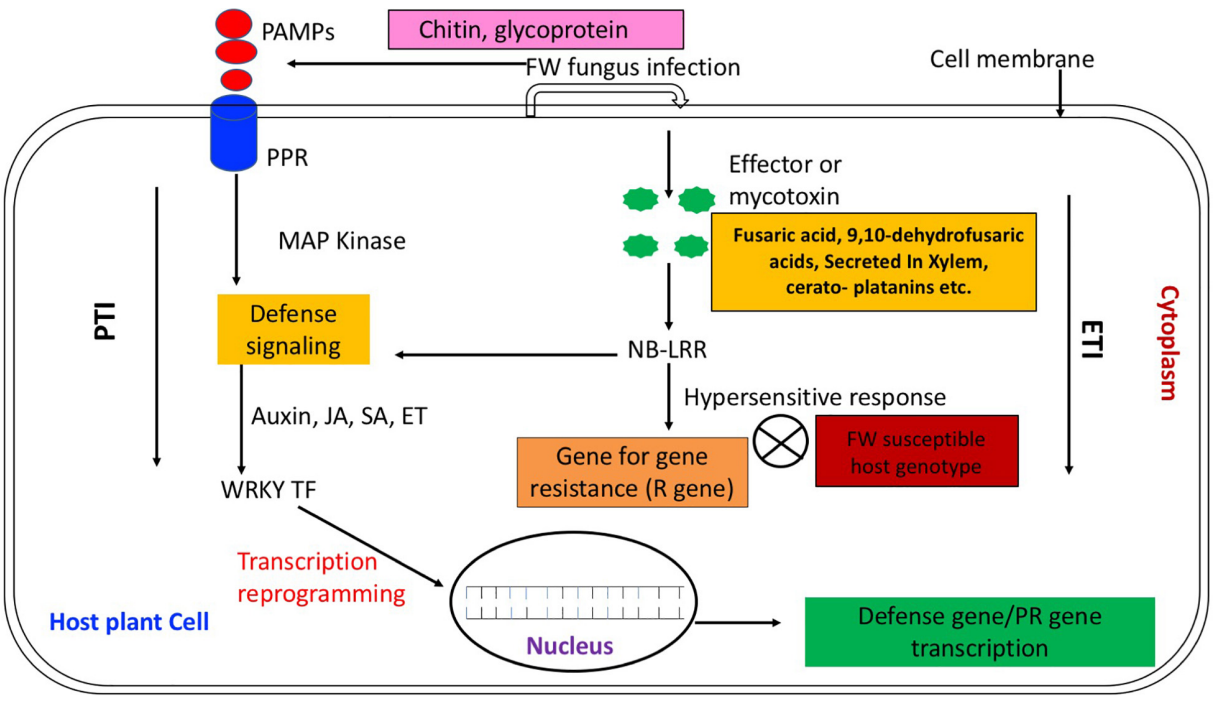


FIGURE 1 | In PTI following the perception of PAMPs by PRRs, plant evokes signaling networks mediated by protein kinases viz., mitogen-activated protein kinases (MAPK) and protein phosphorylation (Nakagami et al., 2005; Boudsocq et al., 2010; Rodriguez et al., 2010; Lanubile et al., 2015). In turn, this is accompanied by activation of multiple TFs genes (via auxin, JA, ET) that ultimately switch on transcription of various defense responsive genes including PR genes (Castillejo et al., 2015). Once the PTI fails, the ETI forms the second line of immune response in which plant defends itself against pathogen attack through encoding immune receptors of the nucleotide binding leucine rich repeat (NB-LRR) class by R genes that enable recognizing the effector molecules (viz., fusaric acid) (Bani et al., 2014) produced by pathogen and initiate hypersensitive response (Zipfel and Robatzek, 2010; Ma and Ma, 2016). In case of susceptible host, the effector molecules remain unrecognized and there is no host pathogen hypersensitive response (Li et al., 2020).

dropping petioles and yellowing of leaf symptoms appear at podding stage, causing 24–65% yield loss (Jiménez-Díaz et al., 2015). Based on the symptoms produced on host plant, *Foc* races are grouped into two categories. Six races (1A, 2, 3, 4, 5, and 6) cause wilting symptoms (see **Table 1**), while two races (0 and 1B/C) cause yellowing symptoms in the plant during infection (Haware and Nene, 1982; Jimenez-Diaz et al., 1993; Kelly et al., 1994; Sharma et al., 2004). Among these *Foc* races, races 2, 3, and 4 are prevalent in India. Races 0, 1B/C, 5, and 6 are reported in Mediterranean and the United States, whereas the race 1A occurs in India, the United States, and the Mediterranean region (Jimenez-Diaz et al., 1993; Jiménez-Gasco et al., 2001; Landa et al., 2006). Gurjar et al. (2009) classified race 3 as *Fusarium proliferatum* based on phylogenetic analysis of translation elongation factor 1 alpha (TEF1-alpha) sequence data. The study demonstrated the reliable identification of Indian *Foc* races using a diverse DNA marker system. Earlier, Jiménez-Gasco and Jiménez-Díaz (2003) developed SCAR markers from race-specific random amplified polymorphic DNA (RAPD) markers for the identification of *Foc* and its pathogenic races. Change in the race scenario of *Foc* has been reported in India. Seventy *Foc* isolates, collected from 13 states and four crop cultivation zones of India, grouped into eight races based on their response to differential cultivars of chickpea. Further, characterization with four different molecular markers (RAPD, universal rice primers, SSR, and ISSR) grouped the isolates into eight clusters, which partially corresponded to the chickpea-growing zones and the racial distribution of the pathogen (Dubey et al., 2012). The presence of 47 isolates of different *Foc* races into one vegetative compatibility group (VCG) suggested its monophyletic origin (Nogales Moncada et al., 2009).

Cowpea

Fusarium oxysporum f. sp. *tracheiphilum* (*Fot*), a soil-borne fungus, severely challenges cowpea production (Armstrong and Armstrong, 1981). The disease is prevalent in all cowpea growing countries across the world including the United States, central valley of California (Hare, 1953; Smith et al., 1999), Australia (Summerell et al., 2011), Brazil and Nigeria (Armstrong and Armstrong, 1980; Assunção et al., 2003). Four *Fot* races (1, 2, 3, and 4) have been identified based on their pathogenic reactions to differential cowpea genotypes. Worldwide occurrence of race

TABLE 1 | List of FW races and their effects on various grain legumes.

Crop	Causal organism	Races	Symptoms	% Yield loss	References
Chickpea	*Fusarium oxysporum* f.sp. *ciceri* (*Foc*)	Eight races (0, 1B/C, 1A, 2, 3, 4, 5, and 6)	Destruction of vascular bundle that leads to inflicting plant water relation, drooping of petioles, rachis and leaflets; Races 0 and 1B/C cause yellowing syndrome	10–40% and even 100% under favorable condition	Haware and Nene, 1982; Jimenez-Diaz et al., 1993; Kelly et al., 1994; Sharma and Muehlbauer, 2007; Castro et al., 2010; Jiménez-Díaz et al., 2015
Pea	*Fusarium oxysporum* Schl. f. sp. *pisi* Snyd. and Hans. (*Fop*)	Four races (1, 2, 5, and 6)	Gray-green discolored foliage, thickening of basal internodes, downward curling of the leaves from the base to the apex, chlorosis with unilateral wilting, yellow to orange discoloration in the vascular tissue wilting and death of the entire aerial part race 1 and race 5 cause complete death of plant	up to 100%	Wade, 1929; Hare et al., 1949; Coyne et al., 2000; Haglund and Kraft, 2001; Jain et al., 2015
Cowpea	*Fusarium oxysporum* f.sp. *tracheiphilum* (*Fot*)	Four races (1, 2, 3, and 4)	Wilting and leaf chlorosis and stunting the entire plant typical symptoms evident during the flowering and early pod development stages causing high mortality in the affected areas with severe overall yield loss	up to 100%	Hare, 1953; Patel, 1985; Smith et al., 1999; Davis and Frate, 2007; Pottorff et al., 2014
Pigeon pea	*Fusarium udum* Butler	–	–	30–100%	Reddy et al., 1990; Okiror, 2002
Common bean	*Fusarium oxysporum* (Schlecht.) f. sp. *phaseoli* Kendrick & Snyder (*FOP*)	Two race (2, 4)	Phloem blockage; chlorosis and wilt in older leaves wilting of younger leaves, and necrosis of the apex and vascular tissue Vascular tissues turn a red-brown color, shortening of plant lifecycle and plant death	up to 100%	Abawi and Pastor-Corrales, 1990; Buruchara and Camacho, 2000; Fall et al., 2001; Xue et al., 2015
Lentil	*Fusarium oxysporum* f. sp. *lentis* (*Fol*)	Eight races/pathotypes	Stunting, wilting, internal vascular discoloration in lower stem, wilting symptoms appear in seedling, flowering and pod stages	up to 100%	Kannaiyan and Nene, 1976; Agrawal et al., 1991; Erskine and Bayaa, 1996; Tosi and Cappelli, 2001; Hiremani and Dubey, 2018
Faba bean	*Fusarium oxysporum* f. *fabae* (*FOF*)	–	–	–	Abdel Rehim, 1962

3 has been reported among various *Fot* races (Hare, 1953; Smith et al., 1999). Broad patches of diseased cowpea plants with visible symptoms of chlorosis, wilting and stunting at the seedling stage or during flowering and early pod development result in high mortality and significant yield loss (Smith et al., 1999; Pottorff et al., 2012; see **Table 1**).

Common Bean

Wilt of common bean caused by *F. oxysporum* f. sp. *phaseoli* (*FOP*) results in substantial yield loss. *FOP* is commonly found in various common bean growing countries across the United States, China, Africa, and Latin America (Harter, 1929; Buruchara and Camacho, 2000). Elena and Papas (2002) reported that most of the isolates among 27 *FOP* isolates from Greece belonged to the same VCG. Earlier, VGC and molecular analysis of 128 *FOP* isolates collected from the field of El Barco de Avila in Spain suggested a difference between *FOP* isolates of Spain and America thus, supporting its pathogenic evolution (Alves-Santos et al., 1999).

The *FOP* invades plants through penetrating roots and colonizes the cortical cell. The hyphae move toward vascular parenchyma cells accompanied by a collapse of the xylem vessel and subsequent disruption of the water uptake from the roots (Ninþo-Saìnchez et al., 2015; Garces-Fiallos et al., 2017). Recent histological evidences support faster colonization of *FOP* in the susceptible genotypes in comparison to the resistant genotypes (Garces-Fiallos et al., 2017). A previous study by Ninþo-Saìnchez et al. (2015) explained the differential pattern of host plant colonization of highly virulent and weakly virulent strains of *FOP* based on expression analysis of different virulence genes viz., *FTF1, SIX1,* and *SIX6*. The infected plant displays wilting of older leaves followed by younger leaves and necrosis of the apex part (Fall et al., 2001; Ninþo-Saìnchez et al., 2015; Batista et al., 2017; Garces-Fiallos et al., 2017), internal stem discoloration (Batista et al., 2016), leading to death of the plant (Fall et al., 2001; Batista et al., 2017).

Lentil

Fusarium oxysporum f. sp. *lentis* (*Fol*) is the causal agent of wilt disease in lentil (Erskine and Bayaa, 1996; Bayaa et al., 1998; Tosi and Cappelli, 2001). In India, 0.7–9.3% mortality has been reported at reproductive stage based on a survey of 116 districts of nine lentil growing states (Chaudhary et al., 2010). Similarly, severe disease incidence was recorded in 21 locations of Pakistan (Rubab et al., 2014). The pathogen survives in plant debris/soil and infects the host through the roots. The ability of the pathogen to survive in soil for long periods through chlamydospores further aggravates the situation. Evidence also suggests transmission of the pathogen through seeds (Erskine et al., 1990). In Algeria, based on the pathogenic variation of 32 isolates of *Fol*, Belabid et al. (2004) reported homogeneous behavior of Algerian *Fol* isolates with no variations in virulence and the existence of one *Fol* race. Although resistant or moderately resistant wilt varieties have been developed, variable responses of these varieties across agro-ecological niches imply the presence of high pathogen variability (Naimuddin and Chaudhary, 2009). No pathotypes within this *formae specialis* were identified until the study of Pouralibaba et al. (2016). Upon inoculation of 28 lentil resistant accessions with six *Fol* isolates with different geographical origins, a highly significant isolate × accession interaction resulted in the identification of four accessions as a putative differential set. The virulence pattern of 52 *Fol* isolates from Iran, Syria, and Algeria allowed the identification of seven pathotypes (1–7). Similarly, in India, Hiremani and Dubey (2018) identified eight races/pathotypes of the pathogen based on resistant and susceptible reactions on a set of differential cultivars. These studies pave the way for developing race specific wilt resistant lentil cultivars and help in the identification of new races from different lentil growing areas worldwide.

Pea

Wilt caused by *F. oxysporum* f. sp. *pisi* (*Fop*), is among the most significant yield reducers in pea growing areas (Kraft et al., 1981; Infantino et al., 2006; Rubiales et al., 2015). Among the four races (1, 2, 5, and 6) of *Fop*, races 1 and 5 are the most destructive, causing death of the plant (Jain et al., 2015). However, race 2 remains less pathogenic and wilting symptom appears in post podding stage. Based on VCG, isolates of *Fop 1* and *Fop 6* remained in a single VCG, whereas *Fop5* belonged to a second VCG and *Fop2* were present in another two VCGs (Kraft, 1994). *Fop* infects the root and interferes with the plant water movement, resulting in wilting symptoms (see **Table 1**). The other notable symptoms of wilt in pea include gray-green discoloration of foliage, chlorosis with unilateral wilting, orange discoloration in the vascular tissue (Jain et al., 2015).

Pigeonpea

Pigeonpea wilt is caused by *Fusarium udum*, a soil and seed borne pathogen, and the disease is reported to cause up to 100% yield loss in the susceptible cultivars (Reddy et al., 1990; Okiror, 2002). Occurrence of FW is reported predominantly in north and central parts of India, Kenya, and Malawi, and also in Ghana, Tanzania and Uganda (Nene, 1980). The annual losses caused by FW in pigeonpea have been estimated to be United States $71 million and 470,000 tons of grain in India and 30,000 tons of grain in Africa (Saxena et al., 2017).

Faba Bean

Faba bean cultivation is severely affected by wilt [*F. oxysporum* f. *fabae* (*FOF*)] worldwide (Abdul Wahid et al., 1998; Dong et al., 2014). This disease was first reported by Abdel Rehim (1962) in Egypt.

PLANT GENETIC RESOURCES FOR FW RESISTANCE

Efforts to control FW disease with chemical agents have met with limited success. In view of the hazardous nature of fungicides to the environment, developing host plant resistance presents the most durable, economic and ecofriendly means to minimize the FW loss in grain legumes (Saxena, 2008; Jain et al., 2015). Therefore, harnessing the variation in plant traits that impart

FW resistance could substantially improve the resistance level of various grain legumes. Considerable genetic variability has been observed in chickpea genotypes for FW response (Haware et al., 1992; Ali et al., 2002). Identification of resistant sources of FW in both kabuli (ICCV 2 and UC 15) and desi types (FLIP 85-20C, FLIP 85-29C, and FLIP 85-30C) by Ali et al. (2002) was consistent with the earlier report of Jimeìnez-Diìaz et al. (1991). More recently, Sharma et al. (2019) identified several chickpea FW resistant genotypes viz., ICCV 98505, ICCV 07105, ICCV 07111, ICCV 07305 based on GGE biplot analysis (see **Table 2**). Previously, Kumar et al. (1985) developed four kabuli resistant genotypes (ICCV 2, ICCV 3, ICCV 4, and ICCV 5) through a pedigree method. Recent advancements in genomic technologies in grain legumes have provided crop breeders with a set of more efficient tools for resistance breeding. Consequently, successful examples of genomics-assisted trait improvement for abiotic and biotic stresses are now available in legume crops (Bohra et al., 2014a,b; Varshney et al., 2015). In chickpea, a marker-assisted back crossing (MABC) scheme has allowed targeted transfer of genomic regions conferring FW resistance (*foc4*) from WR 315 to Annigeri 1 and JG 74, two elite yet FW-sensitive elite chickpea cultivars (Mannur et al., 2019). Thus, MABC derived products in chickpea such as Super Annigeri 1 and JG 74315-14 showed an 8% increase in yield and disease resistance over Annegiri and 53.5% increase in yield over JG74, respectively (Mannur et al., 2019). Likewise, genomic regions underlying *foc1* and *foc*2 resistance were transferred from JG 315 to C 214 (Varshney et al., 2014) and from Vijay to Pusa 256 (Pratap et al., 2017) using MABC. The MABC-bred lines carrying favorable alleles such as ICCX-100175-349-2-2, ICCX-100175-382-4-6, and ICCX-100175-389-3-2 had high to moderate resistance against FW (*foc1*) under field conditions (Varshney et al., 2014). Marker-aided breeding schemes could enable efficient pyramiding of QTL into new cultivars, thus imparting on them durable resistance against multiple FW races. Apart from the cultivated pool, crop wild relatives (CWRs) of chickpea viz., *Cicer reticulatum*, *Cicer echinospermum, Cicer bijugum*, *Cicer judaicum*, *Cicer pinnatifidum*, and *Cicer cuneatum* have also been identified having traits that confer FW resistance (Nene and Haware, 1980; Kaiser et al., 1994; Singh et al., 1998).

Studies have shown significant genetic variability for *Fol* resistance in lentil (Bayaa et al., 1995; Pouralibaba et al., 2015).

Studies have shown significant genetic variability for *Fol* resistance in lentil (Bayaa et al., 1995; Pouralibaba et al., 2015). Screening in both controlled and field conditions led to the identification of three promising lines viz., 81S15, FLIP 2007-42 L, and FLIP 2009-18 L (Mohammadi et al., 2012) and BGE040548, BGE019708, BGE022526, BGE025720 (Pouralibaba et al., 2015) exhibiting *Fol* resistance. Recently, GGE biplot analysis has revealed two lentil genotypes PL101 and L4076 as promising sources for *Fol* resistance (Parihar et al., 2017). Similarly, resistant genotypes ILWL 79 and ILWL 113 (*L. culinaris* ssp. *orientalis*) and ILWL 138 (*Lens nigricans* ssp. *ervoides*) were screened against *Fol* from 219 wild lentils (Bayaa et al., 1995). Interspecific cross between ILL10829 (*L. culinaris* subsp. *culinaris*) and ILWL30 (*Lens ervoides*) has also revealed substantial genetic variability for *Fol* resistance (Singh et al., 2017).

Previously, based on glasshouse screening, Doling (1964) identified Delwiche Commando, New Era and New Season genotypes to be resistant to both *Fop1* and *Fop2* races of FW. Likewise, Kraft (1994) identified 74SN5 pea line to be an important source of resistance against all the four races *Fop1*, *Fop2*, *Fop5,* and *Fop6*. Subsequently, screening of a large set of 452 pea accessions collected from 24 countries resulted in the identification of 62 accessions to be resistant against *Fop2* and 39 accessions out of these 62 also possessed resistant to *Fop1* (McPhee et al., 1999). Additionally, PI 344012 a wild progenitor of pea displayed resistance against both *Fop1* and *Fop2* (McPhee et al., 1999). Given the screening of 117 pea genotypes against *Fop1*, *Fop2*, *Fop5,* and *Fop6* under growth chamber, Neumann and Xue (2003) identified Radley and Princess cultivars exhibiting resistant reaction against *Fop2*, *Fop5,* and *Fop6* races. Likewise, a thorough assessment of eighty accessions of *Pisum* spp. against *Fop2* by recording detailed disease scoring based on the various parameters revealed significant genetic variation for *Fop2* disease reaction (Bani et al., 2012). Further the authors reported eleven accessions namely JI 1412, JI 1559, JI 1760, P 23, P 42, P 614, P 627, P 633, P 639, P 650, and P 656 to be resistant against *Fop2* (Bani et al., 2012). Evaluation of 34 pea genotypes against FW under wilt sick plot and artificially controlled conditions allowed for the identification of GP-6, GP-55, and GP-942 as highly resistant and GP-17, GP-48, GP-473, and GP-941 as resistant donors for *Fop* resistance (Shubha et al., 2016).

Considering the common bean earlier, Ribeiro and Hagedorn (1979) identified the Preto Uberabinha common bean cultivar to be resistant to *FOP* based on disease reaction of plant root inoculated with *FOP* microconidia. Subsequently, rigorous screening of 73 climbing and bush type accessions of common bean against *FOP* led to identification of 19 climbing type and 28 bush type accessions to be resistant against *FOP*. Further, two genotypes RWR 950 and G 685 displayed resistance even at higher inoculum density of *FOP* (Buruchara and Camacho, 2000). Based on the restriction of *FOP* xylem tissue colonization, Pastor-Corrales and Abawi (1987) reported Manteigão Fosco 11 genotype to be resistant against *FOP*. Likewise, UFSC-01 genotype was revealed to be resistant against *FOP* by checking the colonization of *FOP* inside the xylem vessel (Garces-Fiallos et al., 2017). Furthermore, availability of high throughput molecular maker based genome wide association study allowed the identification of 14 highly resistant common bean genotypes against *FOP-SP1* race 6 (Leitão et al., 2020).

In pigeonpea, multi-location and multi-year testing of wider sets of germplasm and advanced breeding lines has revealed several promising lines for regular use in breeding programs. The resistant sources resulting from these evaluations include KPL 43, KPL 44, IPAs 16 F, 8 F, 9 F, and 12 F (Singh et al., 2011a) and ICPLs 20109, 20096, 20115, 20116, 20102, 20106, and 20094 (Sharma et al., 2016a; see **Table 2**). Notably, resistance to wilt is an essential prerequisite for variety identification and release in pigeonpea. Several pigeonpea varieties such as Asha (ICPL

TABLE 2 | List of grain legume genetic resource contributing to FW resistance.

Crop	Resistance source	Name of the institute	References
Chickpea	ICC 11322 (WR 315)	ICRISAT, Patancheru	Singh et al., 1974
Chickpea	*C. bijugum, C. judaicum, C. pinnatifidum, C. reticulatum, C. echinospermum, and C. cuneatum*	ICRISAT, Patancheru	Nene and Haware, 1980; Singh et al., 1998
Chickpea	ICCV2, ICCV3, ICCV4, and ICCV5 (against race I)	ICRISAT, Patancheru	Kumar et al., 1985
Chickpea	ICC 11322, 14424, and 14433 (against to race I)	ICRISAT, Patancheru	Nene et al., 1989
Chickpea	ICC-2862, -9023, -9032, -10803, -11550, and -11551	–	Haware et al., 1990
Chickpea	FLIP 84-43C (against race 0), ILC-5411, FLIP 85-20C (against race 5), FLIP 85-29C, FLIP 85-30C, ILC-127 (against race 0), ILC-219 (against race 0), ILC-237, ILC-267, and ILC-513 (against race 0)	Santaella, Córdoba, Spain	Jimeìnez-Dìaz et al., 1991
Chickpea	*C. cuneatum, C. judaicum* (PI458559 resistant against race 0)*, C. bijugum* (PI458550, PI458552 resistant to race 0, 5)*, Cicer canariense* (PI553457 resistant against race 0)*, Cicer chorassanicum* (PI458553 resistant against race 0)*, C. cuneatum, C. judaicum,* and C. *pinnatifidum* (PI 458555, PI458556 resistant to race 0)		Kaiser et al., 1994
Chickpea	ICCV 2 and UC 15 FLIP 85-20C, FLIP 85- 29C, and FLIP 85-30C	Hudeiba Research Station, Ed-Damer, Sudan	Ali et al., 2002
Chickpea	CA-334.20.4, CA-336.14.3.0, and ICC-14216K (race 5)	–	Navas-Cortes et al., 1998; Castillo et al., 2003
Chickpea	Andoum 1 and Ayala (race 0)	–	Halila and Harrabi, 1990; Landa et al., 2006
Chickpea	Surutato-77, Sonora-80, Tubutama, UC-15 and UC-27, Gavilan	Mexico	Morales, 1986; Buddenhagen et al., 1988; Helms et al., 1992; Sharma et al., 2005
Chickpea	BG-212	India	Sharma et al., 2005
Chickpea	ICC-7520	Iran	Sharma et al., 2005
Chickpea	Annigeri	India	Sharma et al., 2005
Chickpea	ICC 7537 resistant to all races (except race 4)	Ethiopia	Sharma et al., 2005
Chickpea	ICC14194, ICC17109, WR315	ICRISAT, Patancheru	Gaur et al., 2006
Chickpea	CM418-1/01, CM446-1/01, CM499/01, CM499-1/01, CM499-2/01 CM554-1/01, CM554-2/01, CM557-2/01 CM557-5/01, CM557-6/01, CM557-7/01, CM5578/01 and CM499-5/01		Shah et al., 2009
Chickpea	ICCV 09118, ICCV 09113, ICCV 09115, ICCV 09308, ICCV 09314	ICRISAT, Patancheru	Sharma et al., 2010
Chickpea	ICCV 05527, ICCV 05528, ICCV 96818	ICRISAT, Patancheru	Sharma et al., 2012
Chickpea	Three lines derived from MABC based C 214 × WR 315 cross	ICRISAT, Patancheru	Varshney et al., 2014
Chickpea	ICCVs 98505, 07105, 07111, 07305, 08113, and 93706 (highly resistant) ICCVs 08123, 08125, 96858, 07118, 08124, 04514, 08323, and 08117(moderately resistant)	ICRISAT, Patancheru	Sharma et al., 2019
Chickpea	Digvijay	India	Upasani et al., 2016
Chickpea	SCGP-WR 28, H 10-05, GL 10023, IPC 2006-77, and CSJK 72	IARI, New Delhi, India	Dubey et al., 2017
Chickpea	Super annigeri and Improved JG74 (resistant against foc4)	ICRISAT, Patancheru ARS-Kalaburagi JNKVV, Jabalpur	Mannur et al., 2019
Pigeonpea	ICP 9145	ICRISAT, Patancheru	Reddy et al., 1995
Pigeonpea	IC0574574	IIPR, Kanpur	Singh et al., 2011b
Pigeonpea	ICPL 20109, ICPL 20096, ICPL 20115, ICPL 20116, ICPL 20102, ICPL 20106, and ICPL 20094	ICRISAT, Patancheru	Sharma et al., 2016a
Pigeonpea	BDN-2004-1, BDN-2001-9, BWR-133, and IPA-234	–	Singh I. P. et al., 2016
Cowpea	CB46, CB3, 7964, and 8514	University of California, Riverside (UCR)	Roberts et al., 1995
Cowpea	California Blackeye 27, California Blackeye 46, California Blackeye 50 (Fot race 3)	–	Ehlers et al., 2000, 2009
Cowpea	CB27, 524B	University of California, Riverside (UCR)	Muchero et al., 2009
Common bean	HF 465-63-1	–	Pastor-Corrales and Abawi, 1987
Common bean	RWR 950 and G 685	–	Buruchara and Camacho, 2000

(Continued)

TABLE 2 | Continued

Crop	Resistance source	Name of the institute	References
Common bean	CAAS260205	Yunnan, China	Xue et al., 2015
Pea	J1412, JI1760, P633 (*P. sativum* ssp. *arvense*), P42(*P. sativum* ssp. *elatius*) against Fop race 2	–	Bani et al., 2012, 2018a
Lentil	ILL 422, ILL 813, ILL 1220, ILL 1462, ILL 2313, and ILL 2684	ICARDA, Syria	Sarker et al., 2001
Lentil	ILWL 79 and ILWL 113 of *L. culinaris* ssp. *orientalis* ILWL 138 of *L. nigricans* ssp. *ervoides*	Tel Hadya farm, Northern Syria	Bayaa et al., 1995
Lentil	81S15, FLIP2007-42 L and FLIP2009-18 L	–	Mohammadi et al., 2012
Lentil	BGE016363, BGE025720, BGE032290, and BGE040548	–	Pouralibaba et al., 2015
Lentil	PL101 and L4076	AICRP, India	Parihar et al., 2017
Faba bean	Assiut-215, Roomy-3, Marut-2, and Giza-2	Assiut University, Egypt	Mahmoud and Abd El-Fatah, 2020

87119), ICP 8863, BSMR 736, TS3R, IPA 203, BDN 708, BDN 711, etc., show considerable level of resistance to *F. udum* (Singh I. P. et al., 2016; Bohra et al., 2017).

In cowpea, genotypes CB46, CB3, 7964, and 8514 were identified as FW resistant based on three years evaluation at two different locations (Roberts et al., 1995). Similarly, genotypes CB46, CB27, and CB50 could serve as donors for developing FW resistant cowpea (Ehlers et al., 2000, 2009; Muchero et al., 2009). Since screening of large germplasm collections for FW resistance remains time consuming, thus availability of molecular markers linked with FW resistant gene(s) could circumvent the traditional screening methods (Ali et al., 2012; Jain et al., 2015).

In faba bean, variation was reported among 16 lines for FW resistance using inter-simple sequence repeat (ISSR), sequence related amplified polymorphism (SRAP) and SSR markers (Mahmoud and Abd El-Fatah, 2020). Based on the disease severity, Assiut-215, Roomy-3, Marut-2, and Giza-2 were found to be promising for FW resistance in faba bean.

GENETIC BASIS OF HOST PLANT RESISTANCE AGAINST FW IN GRAIN LEGUMES

Inheritance of FW Resistance in Grain Legumes

Initial studies on genetic inheritance of FW resistance in grain legumes relied on Mendelian genetics. Literature in chickpea on FW inheritance suggests its control by major genes (Upadhyaya et al., 1983a,b; Kumar, 1998; Tullu et al., 1999). For example, genetic resistance against *Foc*1 results from the action of three independent loci h_1, h_2, and h_3 (Upadhyaya et al., 1983a,b; Singh et al., 1987a,b). Kumar (1998) also advocated involvement of three separate loci controlling *Foc*2 resistance. Likewise, other researchers have reported digenic inheritance for other races (0 and 2) of *Foc* (Tullu et al., 1999; Rubio et al., 2003). Previously, the digenic (*a, b*) nature of *Foc*2 resistance was also established based on the disease reaction of F_2 and F_3 individuals derived from WR 315 × C 104 (Kumar, 1998). Whereas Tekeoglu et al. (2000) and Sharma et al. (2004) reported monogenic inheritance of *Foc 3* and *Foc* 5 resistance. Concerning resistance against *Foc*4, Tullu et al. (1998, 1999) explained its monogenic and recessive nature in the genotype WR 315, however, its recessive and digenic nature was explained in Surutato 77. Likewise, Sharma et al. (2005) reported that the genetic resistance against each race 1A, 2, 3, 4, and 5 was controlled by a single gene. However, the genetic basis of resistance to races 1B/C and 6 is still to be studied.

Classical genetics studies in pea established that resistance to FW (races 1, 2, 5, and 6) was controlled by different genes of a dominant nature (Wade, 1929; Hare et al., 1949; Coyne et al., 2000; Haglund and Kraft, 2001). Resistance against *Fop* races 1, 5, and 6 was conferred by a single dominant gene, however, resistance to race 2 follows a quantitative pattern (Bani et al., 2012; McPhee et al., 2012). A study by Wade (1929) established that the *Fop*1 resistance was controlled by a single dominant gene (*Fw*), which was subsequently mapped on to LG III (Grajal-Martin and Muehlbauer, 2002).

Research on *FOP* resistance in common bean has shown presence of a single gene (Ribeiro and Hagedorn, 1979; Salgado et al., 1995; Cross et al., 2000; Fall et al., 2001), oligogene (Ribeiro and Hagedorn, 1979; Salgado et al., 1995; Cross et al., 2000; Fall et al., 2001; Batista et al., 2017) and polygenes (Salgado et al., 1995; Cross et al., 2000; Batista et al., 2016, 2017). Considering resistance against *Fop4*, occurrence of a single dominant gene (Salgado et al., 1995; Cross et al., 2000) as well as QTL was proposed (Fall et al., 2001).

Studies on the inheritance of FW resistance in pigeonpea indicate varying patterns such as two complementary genes (Pathak (1970), dominant monogenic (Pawar and Mayee, 1986; Pandey et al., 1996; Kotresh et al., 2006; Karimi et al., 2010), recessive monogenic (Jain and Reddy, 1995; Odeny et al., 2009), one dominant and one recessive gene with dominant suppressive epistatic (Saxena et al., 2012) as well as polygenic inheritance (Pal, 1934). Recent analysis of populations derived from four resistant and four susceptible parents suggested that the FW resistance was governed by one dominant gene each in BDN 2004-1 and BDN 2001-9 in comparison to two duplicate dominant genes in BWR 133 and two dominant complimentary genes in IPA 234 (Singh D. et al., 2016).

In lentil, limited research has been done on understanding the genetic basis of resistance to *Fol*. Kamboj et al. (1990) proposed the presence of five independently segregating genes for *Fol* resistance based on allelic tests of the crosses involving three *Fol* resistant lines (L 234, JL 446, and LP 286) and two susceptible lines (L 9-12 and JL 641). Subsequently, Eujayl et al. (1998)

established monogenic dominant inheritance of *Fol* resistance based on $F_{2:4}$ progenies [ILL 5588 × L 692–16–l(s)].

Identification of Resistance Loci and Molecular Marker for FW Resistance in Grain Legumes

DNA marker technologies have facilitated locating/mapping of gene(s) controlling resistance against various races of *Foc* in chickpea (Mayer et al., 1997; Ratnaparkhe et al., 1998; Tullu et al., 1999; Tekeoglu et al., 2000; Winter et al., 2000; Rubio et al., 2003; Sharma et al., 2004; Cobos et al., 2005). Previously, Cobos et al. (2005) reported the $Foc0_1/foc0_1$ gene on LG 5 flanked by OPJ20600 and TR59 markers. Later, Halila et al. (2009) confirmed a second gene $Foc0_2/foc0_2$ (flanked by TS47 and TA59 markers) on LG2 following analysis of two mapping populations CA 2156 × JG 62 and CA 2139 × JG 62. Earlier, this gene was discovered by Rubio et al. (2003). Likewise, Jendoubi et al. (2016) fine mapped the *Foc01/foc01* gene within an interval of 2 cM on LG5 using nearly isogenic lines (NILs). Of the 27 annotated genes, two candidate genes *LOC101514038* and *LOC101499491* are involved in disease resistance (Jendoubi et al., 2016). An SSR-based QTL analysis of $F_{2:3}$ (C 214 × WR 315) elucidated two QTLs *FW-Q-APR-6-1* and *FW-Q-APR-6-2* on LG6 for *Foc1* resistance (Sabbavarapu et al., 2013; see **Table 3**). The SSR marker TA103 was used for introgression of *Foc1* from WR 315 to C 214 (Varshney et al., 2014). Previously, Gowda et al. (2009) located *Foc1* flanked with SSRs TA110 and H3A12 on LG2. The authors also mapped the *Foc2* (TA96-H3A12) and *Foc3* (TA194- H1B06y) on LG2. However, Jingade and Ravikumar (2015) reported one major QTL *GSSR 18-TC14801* on LG 1 for *Foc*1 resistance, which explained up to 71% phenotypic variation (PV). Subsequently, a major QTL *FW-Q-APR-2-1* on CaLG02 and two other minor QTLs *FW-Q-APR-4-1* and *FW-Q-APR-6-1* on CaLG4 and CaLG6, respectively were identified for resistance against *Foc1* and *Foc3* (Garg et al., 2018). Considering *Foc5*, monogenic/oligogenic nature has been established for resistance loci on LG 2 (Tekeoglu et al., 2000; Winter et al., 2000; Castro et al., 2010). Recently, by using SNP in combination with SSR markers the candidate genomic region on LG2 was narrowed down to 820 kb (Caballo et al., 2019b). The authors also suggested involvement of a putative candidate gene *LOC101511605* encoding CBL-interacting serine/threonine-protein kinase 8 in FW response.

In pea, genetic linkages of AFLP (McClendon et al., 2002), SSR (Loridon et al., 2005), SCAR (Okubara et al., 2005), and TRAP (Kwon et al., 2013) with FW was reported. Jain et al. (2015) identified a CAPS marker at 0.9 cM from the *Fw* locus on LG3, which could be effectively used for screening FW resistance in pea (see **Table 3**). Earlier, McPhee et al. (2012) reported two minor QTLs on LG 3 controlling resistance against *Fop* race 2.

In cowpea, SNP analysis of the population developed from CB27 × 24-125B-1 allowed identification of a 3.56 cM genomic region on LG1 for resistance to *Fot3-1* (Pottorff et al., 2012). These marker-trait associations (MTAs) explained up to 27.8% PV for the resistance. Furthermore, a comparative analysis between cowpea and soybean genomes suggested four candidate genes from the *Fot3-1* genomic region, which were related to leucine-rich repeat serine/threonine protein kinases (Pottorff et al., 2012). Likewise, two QTLs *Fot4-1* and *Fot4-2* imparting *Fot4* resistance were identified on LG 5 and LG 3, respectively (Pottorff et al., 2014). Synteny analysis between soybean and cowpea suggested a role for candidate genes underlying the *Fot4-1* and *Fot4-2* QTLs that code for TIR–NBS–LRR proteins and leucine-rich repeat serine/threonine protein kinases (Pottorff et al., 2014; see **Table 3**).

In lentil, Eujayl et al. (1998) tagged *Fw* locus controlling resistance to *Fol* at 10.8 cM from RAPD marker ($OPK-15_{900}$). Subsequently, Hamwieh et al. (2005) reported *Fw* locus on LG 6 flanked by SSR59-2B and AFLP marker p17m30710.

In pigeonpea, different research groups have found significant MTAs and candidate genes for FW by using SSR (Patil et al., 2017a) and SNP markers (Singh V. K. et al., 2016; Saxena et al., 2017).

Whole genome sequence information in combination with high-throughput DNA marker technologies has divulged massive amounts of genome-wide markers to analyze MTAs in large germplasm sets for traits including FW resistance.

In common bean, GWAS on a diverse collection of 162 Portuguese accessions showed nine significant SNPson chromosomes Pv04, Pv05, Pv07, and Pv08 for *F. oxysporum* f. sp. *phaseoli* strain FOP-SP1 race 6. Authors also reported that the resulting candidate genes are engaged in phytoalexin biosynthesis, hypersensitive response, and plant primary metabolism (Leitão et al., 2020). Similarly, GWAS of 96 genotypes in cowpea revealed 11 significant MTAs (on LG 1, 3, 4, 6, 8, 9, 10, and 11) explaining 4% PV related to leaf damage traits and seven significant MTAs (LG 3, 6, 10, 11) explaining 9.7% PV related to vascular discoloration (Wu et al., 2015). Among the significant MTAs, two SNPs 1_0691 and 1_1369 showed close proximity to the QTL *Fot3-1* and *Fot4-2*, previously identified by Pottorff et al. (2012, 2014). A recent SSR-based association study of 89 pigeonpea lines phenotyped for 3 years in a wilt-sick field provided a set of six SSR markers, which were cross-validated in a biparental population segregating for FW (Patil et al., 2017a,b).

SEQUENCING BASED APPROACHES FOR UNDERSTANDING THE PLANT-WILT INTERACTIONS IN GRAIN LEGUMES

Whole Genome Sequencing of Host and FW Causing Pathogen: New Insights Into the Plant Defense System Against FW

Availability of whole genome sequence information in chickpea (Varshney et al., 2013), common bean (Schmutz et al., 2014), cowpea (Lonardi et al., 2019), pea (Kreplak et al., 2019), and pigeonpea (Varshney et al., 2012) could allow identification of the candidate gene(s)/the genomic regions controlling disease resistance. Concurrently, draft genome assemblies of *Fusarium udum* F02845 (Srivastava et al., 2018) and *Foc* (*Foc-38-1*) and *Fop* (*Fop-37622*) (Williams et al., 2016) have shed new light onto virulence-related genes that enhance our current understanding

TABLE 3 | List of QTLs contributing to FW resistance in various grain legumes.

Crop	Mapping population size and type	QTL/gene	Type of marker and name	LG group	PV%	QTL method analysis	References
Chickpea	–	*H1* locus of *Foc 1*	CS27A(RAPD)	–	–	–	Mayer et al., 1997
Chickpea	C-104 × WR-315 (100 F_5)	Single recessive gene (race 1 and race4)	CS-27_{700}, UBC-170_{550} (RAPD)	–	–	–	Tullu et al., 1998
Chickpea	C. *arietinum* × C. *reticulatum* (130, RIL)	races 4 and 5	STMS and a SCAR	–	–	–	Winter et al., 2000
Chickpea	131 (F_6 RIL)	*foc-0*, *foc-4*, and *foc-5*	CS-27 (STS)	–	–	–	Tekeoglu et al., 2000
Chickpea	CA2156 × JG62 (RIL)	*Foc 01/foc 01*	OPJ20(600) (RAPD)	–	–	Maximum Likelihood method	Rubio et al., 2003
	CA2139 × JG62 (RIL)	*Foc 01/foc 01* and *Foc 02/foc 02*		–	–		
Chickpea	CA2139 × JG62 (RIL)	One gene resistance for Fusarium wilt race 0 (*Foc0*)	OPJ20(600) (RAPD) TR59(STMS)	LG3	–	–	Cobos et al., 2005
Chickpea	WR315 × C104 WR315 × K850 K850 × GW5/7	Three loci (race 2)	–			–	Kumar, 2006
Chickpea	WR-315 × C-104	*foc-3* gene	TA96 and TA27, TA196 (STMS)	–		MAPMAKER program	Sharma et al., 2004;
	100 F_7	*foc-1* [syn. h(1)] and *foc-4*	CS27A (STS)				Winter et al., 2000;
			TA194 (STMS)				Gowda et al., 2009
Chickpea	CA2156 × JG62 CA2139 × JG62	Single gene (race 0)	OPJ20600 (RAPD)			–	Rubio et al., 2003
Chickpea	–	*Foc5*	TA59 and TA96 (SSR)	–	–	–	Cobos et al., 2009
Chickpea	F_9	*Foc1*	H3A12 and TA110 (STMS)			–	Gowda et al., 2009
Chickpea		*Foc2*	TA96 and H3A12 (STMS)				
Chickpea		*Foc3*	H1B06y and TA194				
Chickpea	CA2139 × JG62 (RIL)	$Foc0_2/foc0_2$	TA59 (STMS)	LG2	22–26	Interval mapping, multiple-QTL models (MQM)	Halila et al., 2009
	CA2156 × JG62 (RIL)						
Chickpea	–	*foc-5*	TA59 (STMS)	LG2			Castro et al., 2010
Chickpea	C214 × WR315	*FW-Q-APR-6-1 (Foc1)* and *FW-Q-APR-6-2 (Foc1)*	CaM1402 and CaM1101 (STMS)	LG6	10.4–18.8		Sabbavarapu et al., 2013
						QTLNetwork 2.0 Composite interval mapping	
Chickpea	JG62 × WR315, (94 RIL)	*Wilt 1* (race 1), *Wilt 2* (race 1)	TA27-TA59 (STMS)	LG2	16–36		Patil et al., 2014
			TA27-TA110 (STMS)				
Chickpea	C 214 × WR 315	Genomic region resistance for *foc1* and *foc3*	TR19, TA194, TAA60, GA16, TA110, and TS82	LG2	–	–	Varshney et al., 2014
Chickpea	K850 × WR315 (RIL, 140)	5 QTLs	GSSR 11-EST SSR 3	LG1	56–70		Jingade and Ravikumar, 2015
			TR 24-EST SSR 21				
			EST SSR 21-EST SSR 65				
			GSSR 18-TC14801				
Chickpea	CA2156 × JG62 (RIL, 80)	*Foc01/foc01*	H2I20 and TS43 (STMS)	LG5	–	Interval mapping	Jendoubi et al., 2016

(Continued)

TABLE 3 | Continued

Crop	Mapping population size and type	QTL/gene	Type of marker and name	LG group	PV%	QTL method analysis	References
	ILC3279 × JG62 (RIL, 115)	*LOC101514038* and *LOC101499491*	CaGM20820, CaGM20889				
	JG62 × ILC72 (RIL, 102)	Candidate genes					
Chickpea	Pusa 256 × Vijay	*Foc* 2	TA 37 and TA110	–		–	Pratap et al., 2017
Chickpea	JG62 × ICCV 05530	3 QTL (race 1), *FW-Q-APR-2-1*	TR19 and H2B061	CaLG02	6.6–31.5	QTL-IciMapping	Garg et al., 2018
		FW-Q-APR-4-1, FW-Q-APR-6-1	TA132 and TA46	CaLG04 and CaLG06			
		2 QTLs (race 3)	TA80 and CaM0594	CaLG02 and CaLG04			
		FW-Q-APR-2-1 and *FW-Q-APR-4-1*	CKAM1256 and TS72				
Chickpea	WR315 × ILC3279 (RIL, 103)	*LOC101511605* (*Foc5*)	TA59, CaGM07922, SNPs	LG2	–	–	Caballo et al., 2019b
	ICCL81001 × Cr5-9 (RIL 88)						
Chickpea	Annigeri 1 × WR 315 (BC)	Genomic region conferring resistance against	TA59, TA96, TR19, and TA27	LG2		–	Mannur et al., 2019
	JG 74 × WR315 (BC)	*foc4*	GA16 and TA96				
Common bean	F_2 and F_3	One dominant gene for *Fop* race 4	–	–	–	–	Cross et al., 2000
Common bean	Belneb RR-1 A55	One major QTL	RAPD, U20.750	LG10	63.5	–	Fall et al., 2001
	RILs (76)						
Cowpea	CB5 × CB3, CB5 × 7964, F_1, F_2, and BC	One dominant gene (race 2), one dominant gene (race 3)	–	–	–	–	Rigert and Foster, 1987
Cowpea	California Blackeye 27 × 24-125B-1	*Fot3-1*	SNP 1_0860 and 1_1107 1_1484 and 1_0911	LG6	27.8	Kruskal–Wallis and interval mapping analysis	Pottorff et al., 2012
Cowpea	IT93K-503-1 × CB46	*Fot4-1* and *Fot4-2*. *Fot4-1*	-	LG3	–	Interval mapping	Pottorff et al., 2014
	CB27 × 24-125B-1						
	CB27 × IT82E-18/Big Buff						
Cowpea	A panel of 96 genotypes	17 significant MTAs for Fusarium wilt resistance	SNP 1_0075, 1_1111,1_1147, 1_0251, 1_0895, 1_0691, 1_0897, 1_0298, 1_0410, 1_0857, 1_0981, 1_1369, 1_0691, 1_0330, 1_1062, 1_0629, 1_0318, and 1_1504	LG1, LG3, LG4, LG5, LG6, LG8,LG9, LG10, and LG11	2–4	—	Wu et al., 2015
Cowpea	F_1, F_2, and BC	–	SSR, C13−16			—	Omoigui et al., 2018
Pea	Green arrow × PI179449	*Fw*	ACG:CAT_222(AFLP)		—	—	McClendon et al., 2002
			ACC:CTG(AFLP), Y15_1050(RAPD)	—	—		
Pea	–	*Fw*	–	LG3	–		Grajal-Martin and Muehlbauer, 2002
Pea	Shawnee × Bohatyr	*Fnw*		LG4	–		McPhee et al., 2012
	RILs (187)						

(Continued)

TABLE 3 | Continued

Crop	Mapping population size and type	QTL/gene	Type of marker and name	LG group	PV%	QTL method analysis	References
Pea	Lifter × Radley (393 RIL) and Shawnee × Bohatyr (187 RIL)	*Fw*	THO(CAPS marker)	LG3	70—92	Single–factor ANOVA	Jain et al., 2015
			PRX1TRAP13(TRAP marker)				
			AnMtL6, Mt5_56				
Lentil	ILL 5588 × L 692-16-1(s), (RIL 86)	*Fw gene*	SSR59-2B, p17m30710	LG6	–	–	Hamwieh et al., 2005
Lentil	ILL5588 × L692–16–l(s)	*Fw gene*	OP–BH800 and OP–DI5500	–	–	–	Eujayl et al., 2006
	$F_{2:4}$, RIL						
Pigeonpea	GSlxICPL87119,(254 F_2)	One gene	RAPD (OPM03704 and OPAC11500)	–	–	–	Kotresh et al., 2006
	GS1 × ICP8863						
Pigeonpea	ICPL 20096 × ICPL 332	*C.cajan_03203*	SNP	LG2, LG11	–	Seq–BSA approach	Singh V. K. et al., 2016
	(RIL F_7)	*C.cajan_07078*					
		C.cajan_07124					
		C.cajan_02962					
Pigeonpea	ICPB 2049 × ICPL 99050 (RIL)	*qFW1.1,qFW2.1*	SNP, S1_2827280–S1_4263752	LG1, LG2, LG3, LG4, and LG6	6.5–14	Composite interval mapping	Saxena et al., 2017
	ICPL 20096 × ICPL (332 RIL)	*qFW3.1,qFW4.1*	S2_16115010–S2_15580586				
	ICPL 85063 × ICPL 87119 (F_2)	*qFW6.1,qFW11.1*	S3_18695411–S3_17153283				
		qFW11.2,qFW11.3	S4_597553–S4_1108184				
			S4_597553–S4_1108184				
			S6_22726005–S6_23553522				
			S11_37262913–S11_37133265				
			S11_43777543–S11_37133265				
			S11_20607023–S11_16809228				
			S11_4243778–S11_22408748				

of pathogenicity of FW and evolution of the host-pathogen interaction in the legume species. Further, comparative genomics of *Fusarium* species could also elucidate the host specific gene(s), effector gene(s) and the sequence conservation across legume-infecting isolates and other *Fusarium* spp. (Williams et al., 2016). Additionally, construction of fungal pangenome could offer deeper insights into novel pathogen effectors and how they defeat host plant resistance rapidly (Badet and Croll, 2020). Thus, growing refinements in deep sequencing chemistry have paved the way for whole genome re-sequencing (WGRS) of global legume germplasms for capturing the large structural variations (SVs) including the copy number variations and presence-absence variations controlling various traits of economic importance including disease resistance (Varshney et al., 2013; Thudi et al., 2016). In pigeonpea, a comparison of whole genome sequence information of FW-resistant genotypes (ICPL 87119, ICPL 20097, ICP 8863, and ICPL 99050) and FW-susceptible genotype (ICPB 2049) in combination with Seq-BSA of the resistant and susceptible bulks (ICPL 20096 × ICPL 332) revealed four candidate genes including *C.cajan_03203* (Singh V. K. et al., 2016). These identified markers/candidate genes could be deployed for breeding FW resistance in pigeonpea. Genome-wide analysis using sequencing data of wilt responsive genotypes may help in pinpointing haplotypes responsible for resistance against multiple FW races, thus providing scope for gene pyramiding.

FUNCTIONAL "OMICS" STUDIES TO DELINEATE HOST GENES IMPARTING FW RESISTANCE

Prior to the discovery of digital transcriptome profiling, expressed sequenced tags (ESTs), cDNA-AFLP, and cDNA-RAPD were largely employed to find the gene(s) participating in the plant defense mechanisms and plant-pathogen interactions

(Wise et al., 2007; Ashraf et al., 2009; Gupta et al., 2009; Gurjar et al., 2012; Xue et al., 2015).

In common bean, cDNA-AFLP analysis of FW-resistant and susceptible genotypes revealed differential expression of 423 transcript derived fragments (TDFs), of which 98 TDFs had annotated functions in signal transduction, protein synthesis and processing, RNA and energy metabolism, defense and stress responses (Xue et al., 2015). Furthermore, q-RT-PCR analysis confirmed FW-responsive expression of 19 candidate genes in CAAS 260205 (resistant) and BR 130 (susceptible) genotypes. Some important candidate genes viz., *CBFi28*, *CBFi43* (ubiquitin protein), *CBFi45* (poly-ubiquitin protein), *CBFi76* (peroxidase), *CBFi54*, *CBFi58* (calcium dependent protein kinase), *CBFi83,* and *CBFi171* (NBS-LRR) had abundant expression in the resistant genotype (Xue et al., 2015).

In recent years, RNA-sequencing has enabled genome-wide surveys of transcriptomes to identify FW responsive candidate genes and their biological roles with greater precision and higher resolution (Li C. Y. et al., 2012; Kohli et al., 2014). Transcriptome analysis of four chickpea cultivars JG 62, ICCV 2, K 850, and WR 315 allowed several important "large effect" SNPs and Indels in the genomic regions controlling FW resistance (Jain et al., 2015). The underlying genomic region containing these SNPs and indels was predicted to be associated with defense related activity. Caballo et al. (2019a) functionally validated the genomic region controlling *Foc* (race 5) resistance in chickpea from resistant and sensitive NILs developed from the cross ILC 3279 × WR 315. Differential gene expression analysis at 24 h post inoculation (hpi) suggested two known candidate genes *LOC101499873* (encoding chaperonin) and *LOC101490851* and three novel candidate genes (*LOC101509359*, *LOC101495941,* and *LOC101510206* encoding MADS-box transcription factor, MATE family protein and serine hydroxymethyl-transferase, respectively) to be related to defense activity against FW. Likewise, nine genes viz., *LOC101503802*, *LOC101505941*, *LOC101506693* had significantly higher expression in FW sensitive NIL at 48 hpi than the FW resistant NIL (Caballo et al., 2019a). Transcriptome analysis of JG 62 and WR 315 in response to FW (race 1) infection uncovered abundance of differentially expressed transcripts related to various TFs, cellular transporters, sugar metabolism contributing to activate defense signaling against FW in chickpea (Gupta et al., 2013a; see **Table 4**). Further, network analysis also provided greater insights into the role of genes associated with defense components (MAP kinase, serine threonine kinase, etc.), reactive oxygen (superoxide dismutases, glutathione reductase, thioredoxin reductase, etc.), ATPase (myo-inositol phosphate, carboxylate synthase, etc.), significantly participating in the defense signaling against FW in chickpea (Gupta et al., 2013a). Similarly, differential expression was obtained by transcriptome profiling of chickpea genotypes JG 62 (*Foc* susceptible) and Digvijay (*Foc* resistant) for genes that are involved in lignification, hormonal homeostasis, plant defense signaling, reactive oxygen species (ROS) homeostasis, R-gene mediated defense in response to host-pathogen interaction (Upasani et al., 2017). An integrated analysis of transcriptome and metabolome data from the root samples of control and *FOP* infected seedlings in common bean demonstrated that pathogen establishment occurs after 24 h of infection, which is accompanied by timely induction of the defense mechanism. The study reinforced the proposition that the *FOP* defense system in common bean requires contributions from defense-related proteins such as glycosylphosphatidylinositol-anchored proteins (GPI-APs), signaling pathways mediated by hormones like salicylic acid, jasmonate and ethylene, and flavonoid biosynthesis pathway.

Further, to explore the role of micro RNAs (miRNAs) contributing to FW resistance, RNA-seq analysis of ICC 4958 uncovered known as well as novel miRNAs (car-miRNA008 targeting the chalcone synthase gene) involved in FW resistance in chickpea (Kohli et al., 2014). Among the identified miRNAs, miR530 showed 17-fold high expression, whereas miR156_1 and miR156_10 had slightly higher expression in response to FW infection (see **Table 4**).

Proteomics and Metabolomics to Elucidate Plant Defense Mechanism Against FW in Grain Legumes

A proteomics approach allows for the unraveling of the various proteins engaged in host-pathogen interaction and their role in defending the host plant against pathogen attacks (Castillejo et al., 2015; Kumar et al., 2016). Significant roles for myriads of proteins in host-pathogen interactions have been suggested either during the establishment of the pathogen in the susceptible host plant or protecting the host plant from pathogen invasion (Rep et al., 2002; Berrocal-Lobo and Molina, 2008; Mehta et al., 2008; Castillejo et al., 2010; Palomares-Rius et al., 2011; Gunnaiah et al., 2012). These proteins included chitinases, xylem proteinases, β-1,3-glucanases, proteinase inhibitors, pathogenesis-related (PR) proteins, leucine rich-repeat proteins, proline-rice glycoproteins, cellulose synthases, ankyrin repeat containing protein, thaumatin-like protein PR-5b, syntaxins to subtilin-like proteases in various plant species in response to FW infection (Yang et al., 1997; De Ascensao and Dubery, 2000; Li T. et al., 2012; Castillejo et al., 2015; Kumar et al., 2016; Silvia Sebastiani et al., 2017). In the host plant several enzymes viz., glutathione S-transferases, peroxidases, peroxiredoxin, uinone oxidoreductase, copper amine oxidase, caffeic acid O-methyltransferase, chalcone synthase, chalcone isomerase, isoflavone reductase, phenylalanine ammonia lyase, etc., show change in response to FW invasion (Klessig et al., 1998; Garcia-Limones et al., 2002; Gupta et al., 2013a; Kumar et al., 2016). However, limited information is available in grain legumes on the participation of anti-fungal proteins especially for FW (Palomares-Rius et al., 2011; Scarafoni et al., 2013; Kumar et al., 2016). To establish roles of proteins in disease development or prohibiting pathogen attack for disease progression in pea, Castillejo et al. (2015) performed proteomic analysis using two-dimensional electrophoresis (2-DE) and mass spectrometry (MALDI-TOF/TOF), and the study found 53 proteins engaged in the plant's response to *Fop* race 2 infection. These proteins were found to affect carbohydrate and energy metabolism (viz., fructokinase-like protein, beta-amylase, phosphoglucomutase, cytoplasmic), nucleotide and amino acid

TABLE 4 | Exhaustive list of various DEG and candidate genes contributing to FW resistance in grain legumes.

Crop	Genotype name	Differentially expressed genes/candidate gene	Function	Platform used	References
Chickpea	JG62 and WR315	6272 ESTs related to cell signaling and transcription and RNA processing and modification, cellular transport and homeostasis, hormone responses, cellular redox and energy metabolism, defense, and stress responsive genes	Defense mechanism, cellular metabolism and hormone signaling	RNA blot analysis	Ashraf et al., 2009
Chickpea	JG62 and WR315	Redox signaling genes such as redox regulatory respiratory burst oxidase homolog F (RBOHF), thioredoxin 3 (TRX3), cationic peroxidase 3 (OCP3), flavodoxin-like quinone reductase 1 (FQR1), iron superoxide dismutase 1, NADH cytochrome b5 reductase (CBR), Fe (II) oxidoreductase 7 (FRO7), Genes related to intracellular transportation *ABC transporter* like gene, polyol transporter gene, translocase, heavy metal transporter (detoxifying protein) (FRS6), bZIP, homeodomain leucine zipper, MYB, helix loop helix, zinc finger (CCHC type), and heat shock family protein, sucrose synthase (SUS4), β-amylase (BAM1), serine threonine kinase (CDKB1.1), and vacuolar ATPase (TUF)	Defense activity, ROS activity, and sugar transportation, sugar metabolism	qPCR	Gupta et al., 2013a
Chickpea	ICC4958	car-miRNA008 targets the chalcone synthase miR2118 and car-miR5213 targeting defense gene encoding Toll/Interleukin-1 receptor-nucleotide binding site-leucine-rich repeats miR2111 targets a Kelch repeat-containing F-box protein	Flavonoid biosynthesis involved in defense mechanism	Illumina (NGS) qPCR	Kohli et al., 2014
Chickpea	Digvijay and JG62	Glutamate dehydrogenase-GDH, glutamate synthase Glutamine synthase and asparagine synthase Methionine synthase and AdoMet synthetase CCoAMT, CHS, CHI, iso-flavone 40-O-methyltransferase, IFS, and IFR	Nitrogen mobilization and methionine metabolism, lignin and phytoalexin biosynthetic pathways	qPCR	Kumar et al., 2016
Chickpea	Digvijay and JG62	Chitin synthase VII (*Chs7*) G protein β subunit gene Mitochondrial carrier protein (*Fow1*) Xylanases gene, pectate lyases gene, polygalacturonases gene	Involved in establishment of pathogen in the host plant	qPCR	Upasani et al., 2016
Chickpea	Digvijay and JG62	3816 DEGs	Lignification, hormonal homeostasis, plant defense signaling, ROS homeostasis, R-gene mediated defense	LongSAGE	Upasani et al., 2017
Chickpea	NILs—RIP8-94-5 resistant R/RIP8-94-11 susceptible (S)	Upregulation of LOC101509359, LOC101495941, LOC101510206 genes in resistant NILs and upregulation of LOC101503802, LOC101505941, LOC101506693, LOC101507659, LOC101509037, LOC101510206, LOC101510544, LOC101501552, and LOC101502928 candidate genes in susceptible NILs	Encodes MADS-box transcription factor, encodes protein related to multidrug and toxic compound extrusion (MATE) family, encodes for a serine hydroxymethyl-transferase	qPCR	Caballo et al., 2019a

(Continued)

TABLE 4 | Continued

Crop	Genotype name	Differentially expressed genes/candidate gene	Function	Platform used	References
Chickpea	JG62 and WR315	PR10, pectinesterase, Aquaporin, ATP synthase, mtr3, exosome, immunity associated differentially expressed genes, cystatin and DnaJ, pectinesterase and xyloglucosyl transferase, actin and profilin-like genes, cytochrome P450, cytochrome P450 monooxygenase, and peroxidase	Immunity, ATP, ROS activity, cell wall remodelers, cytoskeleton related function	qPCR, Microarray analysis	Ashraf et al., 2018
Common bean	BRB130 and CAAS260205	122 transcript-derived fragments (redox-related genes, hormone response genes, transport-related genes, defense and stress response-related, signal transduction genes development and cytoskeletal organization). Upregulation of candidate genes viz., CBFi28,CBFi43,CBFi45,CBFi76,CBFi56, CBFi63, and CBFi122 CBFi54, CBFi58, CBFi83 and CBFi171 in resistant genotype	Signal transduction, hormonal response, defense and cellular metabolism, etc.	qPCR, cDNA-AFLP analysis	Xue et al., 2015
Common bean	Liyun No. 2	8269 downregulated genes, 13,771 upregulated genes Upregulated candidate genes *PHAVU_007G070400g*, *PHAVU_004G134300g, PHAVU_011G042100g*, *PHAVU_008G232600g, PHAVU_007G185300g.* Downregulated candidate genes *PHAVU_003G141800g*, *PHAVU_007G0495001g*, and *PHAVU_007G236300g* down regulation of UDP-glucuronic acid decarboxylase and cellulose synthases high accumulation of pectate lyases, pectin methylesterase inhibitors (PMEI), pectin methylesterases (PME), and Polygalacturonases genes, genes responsive to hormone signaling	Structural defense Defense related proteins Hormones signaling pathways Energy metabolism and nitrogen mobilization Flavonoid biosynthesis pathway	Illumina (NGS)	Chen et al., 2009

metabolism (viz., apyrase S-type, adenosine kinase/copper ion binding), signal transduction and cellular process (viz., chalcone O-methyltransferase, 4-3-3-like protein), redox and homeostasis (viz., short-chain alcohol dehydrogenase SAD-C, short-chain alcohol dehydrogenase A), defense (endochitinase A2, beta-1,3-glucanase), and biosynthetic process (viz., NADPH: isoflavone oxidoreductase, glutamate decarboxylase) (Castillejo et al., 2015). In chickpea, the role of various defense-related proteins was observed in restricting FW infection in the genotypes Digvijay (FW resistant) and JG 62 (FW susceptible) (Kumar et al., 2016). Several ROS activating enzymes viz., glutathione peroxidase, glutaredoxin, glutathione S-transferase, ascorbate peroxidase, peroxiredoxin were abundant in Digvijay as compared to JG 62. Likewise, the genotype Digvijay was able to restrict FW pathogen attack than FW susceptible cultivar JG 62 due to the abundance of PR proteins (Kumar et al., 2016). Thus, proteomics could further illuminate our understanding of the unknown proteins involved in various signal transduction pathways for inducing host innate immunity against FW attack in grain legumes. Concurrently, this approach could also enable us to discover the novel pathogen effectors that drive the arms race between host plants and pathogens.

Like proteomics, a metabolomics approach has greatly advanced our understanding about various metabolites, hormonal crosstalks, and signaling molecules that participate in plant defense mechanisms against FW pathogenesis in crop plants including grain legumes (Gupta et al., 2010; Kumar et al., 2016). The various metabolites produced in response to FW include sugars like hexokinase, glucose-6-phosphate, sucrose synthase, trehalose, invertase, β-amylase, etc. (Morkunas and Ratajczak, 2014). These sugars play a key role in plant resistance against pathogen attacks by serving as substrate for supplying energy, causing oxidative burst and ROS generation, enhancing lignification of cell wall, and acting as signaling molecule in concert with various phytohormones to induce plant innate immunity (Morkunas et al., 2005, 2007; Nikraftar et al., 2013; Morkunas and Ratajczak, 2014).

Besides sugars, the other important metabolites that are implicated in FW pathogen attack and entry into host plant include various amino acids, organic acids (pyruvate, lactate, acetate, etc.), nucleotides and their derivatives, antioxidants, phytoalexins (lignans, pisatins), polyphenols, phenolic acids (monomers of lignin), calmodulins, flavonoids, lipids, and phenylpropanoids (Klessig et al., 1998; Rolland et al., 2002; Garcia-Limones et al., 2009; Kumar et al., 2015, 2016; Bani et al., 2018a,b). Several plant hormones that serve as the essential signaling molecules in regulation of host defense response against FW infection are ethylene, salicylic acid, and jasmonic

acid (Kunkel and Brooks, 2002). The plant immunity or plant defense response mediated by these phytohormone is defeated or suppressed by the attacking pathogen through production of toxins or effectors (Ma and Ma, 2016).

Metabolite profiling of common bean in response to *FOP* demonstrated participation of UDP-glucuronic acid decarboxylase, cellulose synthases, and pectate lyases, amino acids, glycosyl phosphatidyl inositol-anchored proteins, and various phytohormones (salicylic acid, jasmonate, and ethylene), polyphenols, anthocyanins, flavanones, and flavones during host plant, and *FOP* interaction. Similarly, an abundance of various proteins contributing in glycolysis and TCA processes, defense related metabolites (endo beta-1,3-glucanase, chitinases, caffeic acid O-methyltransferase, and caffeoyl-CoA O-methyltransferase), phytoalexins (genistein, luteolin, quinone), flavonoids, isoflavonoids, and phenolic compounds was observed in Digvijay than in JG 62 (Kumar et al., 2016). However, significant decrease in various amino acids and sugars viz., sucrose and fructose in susceptible cultivar allows FW pathogens to invade and promote disease development (Kumar et al., 2016). Therefore, further advancements in metabolomics could enable elucidation of intricate network of various signaling molecules and hormonal crosstalk contributing to FW resistance in grain legumes.

A comparison of the different studies that analyzed changes in plant transcriptomes, proteomes and metabolomes in response to FW infection reinforces the role of chitinases, PR proteins, ROS activating enzymes, phenolic compounds, flavonoids, phytoalexins in imparting wilt resistance (Castillejo et al., 2015; Kumar et al., 2016; Upasani et al., 2017; Bani et al., 2018a,b). These studies also highlight the significance of molecules that participate in cellular metabolism including carbohydrate, protein, nucleotides (Castillejo et al., 2015; Xue et al., 2015; Kumar et al., 2016) and signaling pathways involving MAP kinase, serine threonine kinase and various phytohormones (Gupta et al., 2013a,b; Xue et al., 2015).

FUTURE PROSPECT FOR BREEDING FOR FW RESISTANCE IN GRAIN LEGUMES

Focus of Phenomics Capturing Host Pathogen Interaction

The declining cost of genotyping and accumulation of huge genotyping data have allowed pinpointing the targeted genomics regions and the underlying causative gene(s)/genomic regions for a variety of important traits including FW resistance. However, linking of this genomic information with the phenotype still remains a daunting task due to complexity of G × E interactions (Araus and Cairns, 2014). Current state-of-the-art high throughput phenotyping (HTP) approach has the potential to bridge the genotype-phenotype gap for various complex traits (Fiorani and Schurr, 2013; Araus and Cairns, 2014). Recent advances in high-resolution imaging platforms and sensor technologies have revolutionized our capacity to investigate plant disease interaction, screening of disease resistant lines and identifying plant disease at large scale and large field (Lowe et al., 2017; Shakoor et al., 2017; Singh et al., 2018). Several HTP approaches including field-based remote sensing, 3D scanning, unmanned aerial vehicles system in association with multispectral and thermal cameras, RGB based imaging, fluorescence imaging to hyperspectral image sensing are worth mentioning, and are routinely employed for precise understanding of plant-pathogen interaction, detecting early stage disease symptoms and assessing disease severity (Mahlein et al., 2012; Römer et al., 2012; Mahlein, 2016; Lowe et al., 2017; Zhang et al., 2019). Given the cumbersome techniques of identifying disease infected plants and monitoring of plant disease symptoms both manually and visually that may delay preventing disease progression, early detection of disease symptoms through various sophisticated imaging techniques could assist in taking early preventive measures for restricting disease progression and crop yield loss (Mahlein, 2016; Lowe et al., 2017; Ghosal et al., 2018). Among the various fluorescence-based imaging analysis techniques used, the chlorophyll fluorescence imaging technique estimating F_v/F_m remains a reliable phenotyping technique for monitoring plant-pathogen interaction and disease severity with greater precision (Berger et al., 2007; Bauriegel et al., 2011; Rousseau et al., 2013). Monitoring changes in leaf surface temperature of FW responsive genotypes through an infra-red system allowed identification of FW resistant and susceptible genotypes in pea (Rispail and Rubiales, 2015) and in *Medicago truncatula* (Rispail et al., 2015). Likewise, multi and hyperspectral imaging was used for early prediction of disease onset, differentiating healthy and diseased plants, quantifying disease infection and assessing disease severity in various plant species (for details see Lowe et al., 2017; Shakoor et al., 2017). Similarly, high resolution thermal and hyperspectral imaging approaches were used to detect early wilt of olive caused by *Verticillium dahliae* in a large acreage (Calderón et al., 2015). However, disease prediction accuracy completely depends on the curation and interpretation of acquired hyperspectral imaging data (Singh et al., 2018). Therefore, to improve prediction accuracy of various plant diseases, currently machine learning and deep learning (convolutional neural network and artificial neural networks) approaches have been combined with the hyperspectral imaging data (Mohanty et al., 2016; Sladojevic et al., 2016; Ghosal et al., 2018; Singh et al., 2018). Further interpretation of these images could allow us to identify disease with higher precision and accuracy and also assist in proper assessment of disease severity. Thus, in the era of "Crop breeding 4.0 driven by the big data," HTP will be an integral component of genotype, phenotype, and environment-based decision-making models (Wallace et al., 2018; Jiang et al., 2019).

Novel Breeding Techniques for Designing FW Resistant Grain Legumes

Grain legume cultivars with an enhanced level of FW resistance have been bred using conventional breeding approaches for many decades. However, this approach is time consuming, and demands (i) sufficient genetic variation in the breeding material

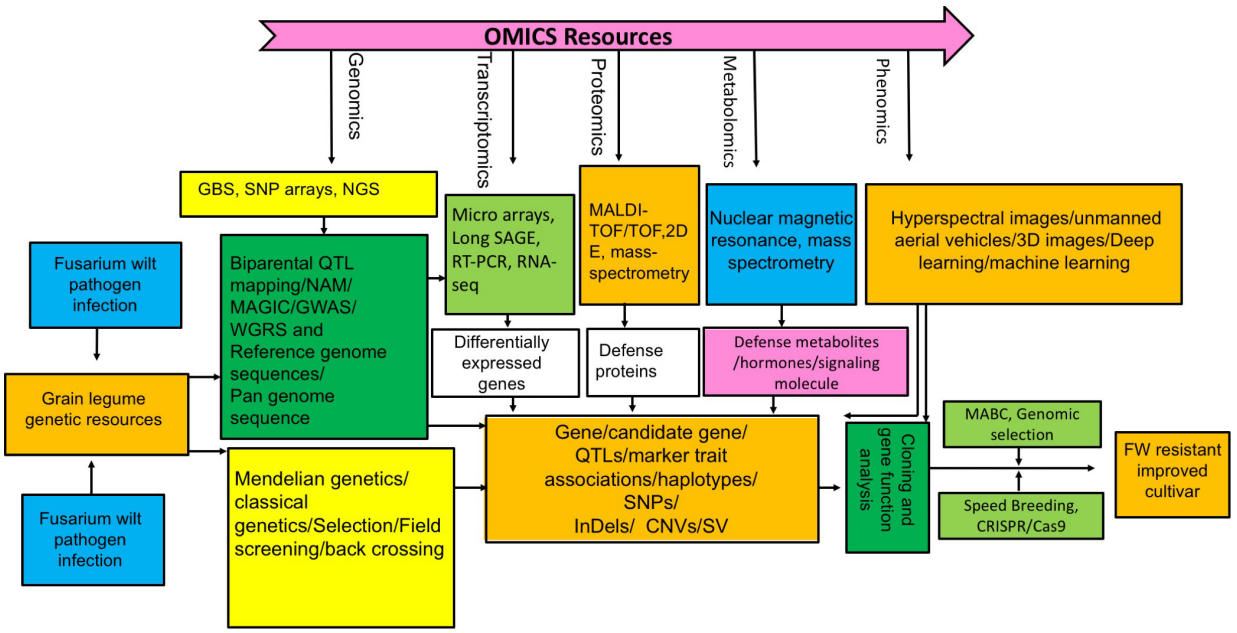


FIGURE 2 | Integrated breeding, genetics, and "omics" scheme illustrating how to combat FW resistance in grain legume.

and (ii) greater manpower for hybridization and handling of segregating population (Yin and Qiu, 2019). A continuous supply of FW-resistant varieties in response to evolving FW pathogens under changing climate demands adoption of efficient breeding technologies (Meuwissen et al., 2001; Gao, 2018; Ghosh et al., 2018).

Explosion of SNP data thanks to the high throughput genotyping platform has offered a great opportunity to adopt genomic selection (GS) in crop plants. GS is used to predict the genomic estimated breeding value of untested individuals using genome-wide marker data. The genotypic and phenotypic information of "training population" is used to train prediction models (Meuwissen et al., 2001; Jannink et al., 2010; Desta and Ortiz, 2014). This approach has been employed for developing disease resistant genotypes in wheat and its scope for breeding disease resistant genotypes has been discussed elsewhere (Poland and Rutkoski, 2016). Genome-wide predictions are yet to be employed for disease resistance in grain legumes.

Speed breeding (SB) or rapid generation advancement presents another promising means to reduce the crop generation time and accelerate the breeding program (Ghosh et al., 2018). This technology has allowed for the recovery of six generations per year in various crops including wheat, barley, chickpea, pea (Ghosh et al., 2018; Hickey et al., 2019). When combined with MAS, SB may dramatically accelerate the screening of the breeding populations against disease or targeted introgression of loci controlling resistance to susceptible genotypes.

To circumvent the cumbersome process of trait manipulation in plants, CRISPR/Cas9 based genome editing technology is revolutionizing plant biology and breeding by precise modification of target gene sequence using customized nucleases (Voytas, 2013; Sander and Joung, 2014; Wang et al., 2014; Langner et al., 2018). This technique has been successfully employed to improve plant resistance against various bacterial and fungal diseases (Li T. et al., 2012; Wang et al., 2014, 2016; Chandrasekaran et al., 2016; Nekrasov et al., 2017; Peng et al., 2017; Zhang et al., 2017; Langner et al., 2018). In future this technique could be harnessed for improving FW resistance in grain legumes. An integrated approach involving various omics technologies and novel breeding schemes for future designing of FW resistant grain legumes has been depicted in **Figure 2**.

CONCLUSION AND PROSPECTS

Severity and frequency of disease occurrence has seen a considerable rise in the wake of changing global climate, thus jeopardizing grain legume production worldwide. Breeding for FW resistance is a key breeding objective of crop improvement programs in grain legumes. Sourcing novel variations of FW resistance from unexploited CWRs and landraces needs greater attention to strengthen the genetic base. In parallel, pyramiding of different resistant gene(s) by adopting both standard backcrossing and DNA marker-aided approaches could expedite breeding of resistant cultivars. Advances in genomic technologies along with increasing genome sequence information could deepen knowledge about the resistant candidate genes/haplotypes to better breed FW-resistant grain legumes. Likewise, functional genomics could allow discovery of candidate loci, their biological functions and the molecular mechanisms underlying host-pathogen interactions. Importantly, emerging HTP phenotyping could illuminate the spatio-temporal aspects of host-pathogen interaction. Targeted and rapid manipulation of genomic loci responsible for FW

resistance in grain legumes could be achieved with adoption of newer techniques like GS, SB, CRISPR/Cas9. An efficient combination of these new approaches paves the way for a steady stream of resistant legume cultivars that yield higher in increasing disease scenarios.

AUTHOR CONTRIBUTIONS

UCJ conceived the idea and wrote the manuscript with AB. UCJ developed the draft of the manuscript. AB and SKP contributed sections on omics technologies. SP contributed section on the FW infection and their effects on major grain legumes. AB edited the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors acknowledge support from Indian Council of Agricultural Research (ICAR), India.

REFERENCES

Abawi, C. S., and Pastor-Corrales, M. A. (1990). *Root Rots of Bean in Latin American and Africa: Diagnosis, Research Methodologies, and Management Strategies*. Cali, CO: Centro Internacional de Agricultura Tropical (CIAT).

Abdel Rehim, M. A. (1962). *Studies on the Organisms Causing Root Rot and Wilt of Horse Beans (Vicia faba var. equina) in U. A. R*. Ph.D. thesis, Alexandria University, Alexandria.

Abdul Wahid, O. A., Ibrahim, M. E., and Omar, M. A. (1998). Oc- currence of soil suppressiveness to *Fusarium* wilt disease of broad bean in Ismailia Governorate. *J. Phytopathol.* 146, 431–435. doi: 10.1111/j.1439-0434.1998.tb04777.x

Agrawal, S. C., Singh, K., and Lal, S. S. (1991). "Plant protection of lentil in India," in *Lentil in South Asia ICAR-ICARDA Seminar*, New Delhi, 147–167.

Alabouvette, C., Lemanceau, P., and Steinberg, C. (1993). Recent advances in the biological control of *Fusarium* wilts. *Pestic. Sci.* 37, 365–373. doi: 10.1002/ps.2780370409

Ali, H., ul Haq, M. A., Shah, T. M., Rahman, M., and Chen, W. (2012). Validation of molecular markers for resistance among Pakistani chickpea germplasm to races of *Fusarium oxysporum* f. sp. Ciceris. *Eur. J. Plant Pathol.* 132, 237–244. doi: 10.1007/s10658-011-9868-1

Ali, M. E. K., Inanga, S., and Sugimoto, Y. (2002). Soursce of resistance to *Fusarium* wilt of chickpea in Sudan. *Phytopathol. Mediterr.* 41, 163–169.

Alves-Santos, F. M., Benito, E. P., Eslava, A. P., and Diaz-Minguez, J. M. (1999). Genetic diversity of *Fusarium oxysporum* strains from common bean fields in Spain. *Appl. Environ. Microbiol.* 65, 3335–3340. doi: 10.1128/aem.65.8.3335-3340.1999

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008

Armstrong, G. M., and Armstrong, J. K. (1980). Cowpea wilt *Fusarium oxysporum* f. sp. tracheiphilum race I from Nigeria. *Plant Dis.* 64, 954–955. doi: 10.1094/PD-64-954

Armstrong, G. M., and Armstrong, J. K. (1981). "Formae speciales and races of *Fusarium oxysporum* causing wilt disease," in *Fusarium: Diseases, Biology and Taxonomy*, eds P. Nelson, T. Toussoun, and R. Cook (University Park, PA: Pennsylvania State University Press), 391–399.

Ashraf, N., Basu, S., Narula, K., Ghosh, S., Tayal, R., and Gangisetty, N. (2018). Integrative network analyses of wilt transcriptome in chickpea reveal genotype dependent regulatory hubs in immunity and susceptibility. *Sci. Rep.* 8:6528. doi: 10.1038/s41598-018-19919-5

Ashraf, N., Ghai, D., Barman, P., Basu, S., Gangisetty, N., Mandal, M. K., et al. (2009). Comparative analyses of genotype dependent expressed sequence tags and stress-responsive transcriptome of chickpea wilt illustrate predicted and unexpected genes and novel regulators of plant immunity. *BMC Genom.* 10:415. doi: 10.1186/1471-2164-10-415

Assunção, I. P., Michereff, S. J., Mizubuti, E. S. G., and Brommonschenkel, S. H. (2003). Influence of *Fusarium* wilt intensity on cowpea yield. *Fitopatol. Bras.* 28, 615–619. doi: 10.1590/S0100-41582003000600006

Baayen, R. P., O'Donnell, K., Bonants, P. J., Cigelnik, E., Kroon, L. P., and Roebroeck, E. J. (2000). Gene genealogies and AFLP Analyses in the *Fusarium oxysporum* complex identify monophyletic and nonmonophyletic formae speciales causing wilt and rot disease. *Phytopathology* 90, 891–900. doi: 10.1094/phyto.2000.90.8.891

Badet, T., and Croll, D. (2020). The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr. Opin. Plant Biol.* 56, 65–73. doi: 10.1016/j.pbi.2020.04.009

Bani, M., Cimmino, A., Evidente, A., Rubiales, D., and Rispai, N. (2018a). Pisatin involvement in the variation of inhibition of *Fusarium oxysporum* f. sp. pisi spore germination by root exudates of Pisum spp. germplasm. *Plant Pathol.* 67, 1046–1054. doi: 10.1111/ppa.12813

Bani, M., Peirez-De-Luque, A., Rubiales, D., and Rispail, N. (2018b). Physical and chemical barriers in root tissues contribute to quantitative resistance to *Fusarium oxysporum* f. sp. pisi in Pea. *Front. Plant Sci.* 9:199. doi: 10.3389/fpls.2018.00199

Bani, M., Rispail, N., Evidente, A., Rubiales, D., and Cimmino, A. (2014). Identification of the main toxins isolated *from Fusarium oxysporum* f. sp. pisi race 2 and their relation with isolates' pathogenicity. *J. Agric. Food Chem.* 62, 2574–2580. doi: 10.1021/jf405530g

Bani, M., Rubiales, D., and Rispail, N. (2012). A detailed evaluation method to identify sources of quantitative resistance to *Fusarium oxysporum* f. sp. pisi race 2 within a Pisum spp. germplasm collection. *Plant Pathol.* 61, 532–542. doi: 10.1111/j.1365-3059.2011.02537.x

Batista, R. O., Oliveira, A. M. C., Silva, J. L. O., Nicoli, A., Carneiro, P. C. S., de Sousa Carneiro, J. E., et al. (2016). Resistance to *Fusarium* wilt in common bean. *Crop Breed. Appl. Biotechnol.* 16, 226–233. doi: 10.1590/1984-70332016v16n3a34

Batista, R. O., Silva, L. C., Moura, L. M., Souza, M. H., Carneiro, P. C. S., and Filho, J. L. S. (2017). Inheritance of resistance to *Fusarium* wilt in common bean. *Euphytica* 213:133. doi: 10.1007/s10681-017-1925-1

Bauriegel, E., Giebel, A., and Herppich, W. B. (2011). Hyperspectral and chlorophyll fluorescence imaging to analyse the impact of *Fusarium culmorum* on the photosynthetic integrity of infected wheat ears. *Sensors* 11, 3765–3779. doi: 10.3390/s110403765

Bayaa, B., Erskine, W., and Hamdi, A. (1995). Evaluation of a wild lentil collection for resistance to vascular wilt. *Genet. Resour. Crop Evol.* 42, 231–235. doi: 10.1007/bf02431257

Bayaa, B., Kumari, S. G., Akkaya, A., and Erskine, W. (1998). Survey of major biotic stresses of lentil in South-East Antalia. *Turkey. Phytopathol. Mediterr.* 37, 88–95.

Beck, M., Heard, W., Mbengue, M., and Robatzek, S. (2012). The INs and OUTs of pattern recognition receptors at the cell surface. *Curr. Opin. Plant Biol.* 15, 1–8. doi: 10.1016/j.pbi.2012.05.004

Belabid, L., Baum, M., Fortas, Z., Bouznad, Z., and Eujayl, I. (2004). Pathogenic and genetic characterisation of Algerian isolates of *Fusarium oxysporum* f. sp. lentis by RAPD and AFLP analysis. *Afr. J. Biotechnol.* 3, 25–31. doi: 10.5897/ajb2004.000-2005

Berger, S., Benediktyová, Z., Matouš, K., Bonfig, K. B., Mueller, M. J., Nedbal, L., et al. (2007). Visualization of dynamics of plant-pathogen interaction by novel combination of chlorophyll fluorescence imaging and statistical analysis: differential effects of virulent and avirulent strains of P. syringae and of oxylipins on A. thaliana. *J. Expt. Bot.* 58, 797–806. doi: 10.1093/jxb/erl208

Berrocal-Lobo, M., and Molina, A. (2008). *Arabidopsis* defense response against *Fusarium oxysporum*. *Trends Plant Sci.* 13, 145–150. doi: 10.1016/j.tplants.2007.12.004

Bigeard, J., Colcombet, J., and Hirt, H. (2015). Signaling mechanisms in pattern-triggered immunity (PTI). *Mol. Plant* 8, 521–539. doi: 10.1016/j.molp.2014.12.022

Bohra, A., Jha, R., Pandey, G., Patil, P. G., Saxena, R. K., Singh, I. P., et al. (2017). New hypervariable SSR markers for diversity analysis, hybrid purity testing and trait mapping in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *Front. Plant Sci.* 8:377. doi: 10.3389/fpls.2017.00377

Bohra, A., Jha, U. C., Kavi Kishor, P. B., Pandey, S., and Singh, N. P. (2014a). Genomics and molecular breeding in lesser explored pulse crops: current trends and future opportunities. *Biotechnol. Adv.* 32, 1410–1428. doi: 10.1016/j.biotechadv.2014.09.001

Bohra, A., Pandey, M. K., Jha, U. C., Singh, B., Singh, I. P., Datta, D., et al. (2014b). Genomics-assisted breeding in the four major pulse crops of developing countries: present status and prospects. *Theor. Appl. Genet.* 127, 1263–1291. doi: 10.1007/s00122-014-2301-3

Bohra, A., Sahrawat, K. L., Kumar, S., Joshi, R., Parihar, A. K., Singh, U., et al. (2015). Genetics- and genomics-based interventions for nutritional enhancement of grain legume crops: status and outlook. *J. Appl. Genet.* 56, 151–161. doi: 10.1007/s13353-014-0268-z

Boudsocq, M., Willmann, M. R., McCormack, M., Lee, H., Shan, L., and He, P. (2010). Differential innate immune signalling via Ca2+ sensor protein kinases. *Nature* 464, 418–423. doi: 10.1038/nature08794

Buddenhagen, I. W., Workneh, F., and Bosque-Pe rez, N. (1988). Chickpea improvement and chickpea diseases in California. *Intern. Chickpea Newslett.* 19, 9–10.

Buruchara, R. A., and Camacho, L. (2000). Common bean reaction to *Fusarium oxysporum* f. sp. phaseoli, the cause of severe vascular wilt in Central Africa. *J. Phytopathol.* 148, 39–45. doi: 10.1111/j.1439-0434.2000.tb04622.x

Caballo, C., Castro, P., Gil, J., Millan, T., Rubio, J., and Die, J. V. (2019a). Candidate genes expression profiling during wilting in chickpea caused by *Fusarium oxysporum* f. sp. ciceris race 5. *PLoS One* 14:e0224212. doi: 10.1371/journal.pone.0224212

Caballo, C., Madrid, E., Gil, J., Chen, W., Rubio, J., and Millan, T. (2019b). Saturation of genomic region implicated in resis- tance to *Fusarium oxysporum* f. sp. ciceris race 5 in chickpea. *Mol. Breed.* 39:16. doi: 10.1007/s11032-019-0932-4

Calderón, R., Navas-Cortés, J. A., and Zarco-Tejada, P. J. (2015). Early detection and quantification of *Verticillium* wilt in olive using hyperspectral and thermal imagery over large areas. *Remote Sens.* 7, 5584–5610. doi: 10.3390/rs70505584

Calero-Nieto, F., Di Pietro, A., Roncero, M. I., and Hera, C. (2007). Role of the transcriptional activator xlnR of *Fusarium oxysporum* in regulation of xylanase genes and virulence. *Mol. Plant Microbe Interact.* 20, 977–985. doi: 10.1094/mpmi-20-8-0977

Cañero, D. C., and Roncero, M. I. (2008). Influence of the chloride channel of *Fusarium oxysporum* on extracellular laccase activity and virulence on tomato plants. *Microbiology.* 154, 1474–1481. doi: 10.1099/mic.0.2007/015388-0

Castillejo, M. A., Curto, M., Fondevilla, S., Rubiales, D., and Jorriìn, J. V. (2010). Two- dimensional electrophoresis based proteomic analysis of the pea (*Pisum sativum*) in response to Mycosphaerella pinodes. *J. Agric. Food Chem* 58, 12822–12832. doi: 10.1021/jf1036917

Castillejo, M. Á, Bani, M., and Rubiales, D. (2015). Understanding pea resistance mechanisms in response to *Fusarium oxysporum* through proteomic analysis. *Phytochemistry* 115, 44–58. doi: 10.1016/j.phytochem.2015.01.009

Castillo, P., Navas-Cortés, J. A., Gomar-Tinoco, D., Di Vito, M., and Jiménez-Díaz, R. M. (2003). Interactions between Meloidogyne artiellia, the cereal and legume root-knot nematode, and *Fusarium oxysporum* f. sp. ciceris race 5 in chickpea. *Phytopathology* 93, 1513–1523. doi: 10.1094/phyto.2003.93.12.1513

Castro, P., Pistón, F., Madrid, E., Milla'n, T., Gil, J., and Rubio, J. (2010). Development of chickpea near-isogenic lines for *Fusarium* wilt. *Theor. Appl. Genet.* 121, 1519–1526. doi: 10.1007/s00122-010-1407-5

Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., and Pearlsman, M. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi: 10.1111/mpp.12375

Chaudhary, R. G., Saxena, D. R., Dhar, V., Singh, R. K., and Namdev, J. K. (2010). Prevalence of wilt-root rot and their associated pathogens at reproductive phase in lentil. *Arch. Phytopathol. Plant Protect.* 43, 996–1000. doi: 10.1080/03235400802214851

Chen, L., Wu, Q., He, W., He, T., Wu, Q., and Miao, Y. (2019). Combined *De Novo* transcriptome and metabolome analysis of common bean response to *fusarium oxysporum* f. sp. *phaseoli* infection. *Int. J. Mol. Sci.* 20:6278. doi: 10.3390/ijms20246278

Cobos, M., Fernandez, M., Rubio, J., Kharrat, M., Moreno, M., Gil, J., et al. (2005). A linkage map of chickpea (*Cicer arietinum* L.) based on populations from Kabuli × Desi crosses: location of genes for resistance to *Fusarium* wilt race 0. *Theor. Appl. Genet.* 110, 1347–1353. doi: 10.1007/s00122-005-1980-1

Cobos, M., Winter, P., Kharrat, M., Cubero, J., Gil, J., Millan, T., et al. (2009). Genetic analysis of agronomic traits in a wide cross of chickpea. *Field Crops Res.* 111, 130–145. doi: 10.1016/j.fcr.2008.11.006

Correll, J. C. (1991). The relationship between formae speciales, races, and vegetative compatibility groups in *Fusarium oxysporum*. *Phytopathology* 81, 1061–1064.

Coyne, C. J., Inglis, D. A., Whitehead, S. J., McClendon, M. T., and Muehlbauer, F. J. (2000). Chromosomal location of *Fwf*, the *Fusarium* wilt race 5 resistance gene in *Pisum sativum*. *Pisum. Genet.* 32, 20–22.

Cross, H., Brick, M. A., Schwartz, H. F., Panella, L. W., and Byrne, P. F. (2000). Inheritance of resistance to *Fusarium* wilt in two common bean races. *Crop Sci.* 40, 954–958. doi: 10.2135/cropsci2000.404954x

Davis, R. M., and Frate, C. A. (2007). *UC IPM Pest Management Guide- Lines: Dry Beans Fusarium Wilt (Blackeyes/Cowpeas)*. Davis, CA: UC ANR Publication, 3446.

De Ascensao, A. R., and Dubery, I. A. (2000). Panama disease: Cell wall reinforcement in banana roots in response to elicitors from *Fusarium oxysporum* f. sp. cubense race four. *Phytopathology* 90, 1173–1180. doi: 10.1094/phyto.2000.90.10.1173

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., and Spanu, P. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13, 414–430. doi: 10.1111/j.1364-3703.2011.00783.x

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19, 592–601. doi: 10.1016/j.tplants.2014.05.006

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812

Doling, D. A. (1964). Glasshouse testing of pea varieties for resistance to *Fusarium* Wilt. *Plant Pathol.* 13, 65–68. doi: 10.1111/j.1365-3059.1964.tb00265.x

Dong, Y., Dong, K., Zheng, Y., Tang, L., and Yang, Z. X. (2014). Faba bean *Fusarium* wilt (*Fusarium oxysporum*) control and its mechanism in different wheat varieties and faba bean in- tercropping system. *J. Appl. Ecol.* 25, 1979–1987.

Dubey, S. C., Priyanka, K., Singh, V., and Singh, B. (2012). Race Profiling and Molecular Diversity Analysis of *Fusarium oxysporum* f.sp. ciceris Causing Wilt in Chickpea. *J. Phytopathol.* 160, 576–587. doi: 10.1111/j.1439-0434.2012.01954.x

Dubey, S. C., Singh, B., and srinivasa, N. (2017). Evaluation of chickpea genotypes against *Fusarium* wilt for resistant sources. *Indian Phytopath* 70, 254–255. doi: 10.24838/ip.2017.v70.i2.70612

Ehlers, J. D., Hall, A. E., Patel, P. N., Roberts, P. A., and Matthews, W. C. (2000). Registration of 'California Blackeye 27' cowpea. *Crop Sci.* 40, 854–855.

Ehlers, J. D., Sanden, B. L., Frate, C. A., Hall, A. E., and Roberts, P. A. (2009). Registration of 'California Blackeye 509cowpea. *J. Plant Registrat.* 3, 236–240. doi: 10.3198/jpr2009.01.0039crc

Elena, K., and Papas, A. C. (2002). Pathogenicity and vegetative compatibility of *Fusarium oxysporum* f.sp. phaseoli in Greece. *J. Phytopathol.* 150, 495–499. doi: 10.1046/j.1439-0434.2002.00786.x

Epstein, L., Kaur, S., Chang, P. L., Carrasquilla-Garcia, N., Lyu, G., Cook, D. R., et al. (2017). Races of the celery pathogen *Fusarium oxysporum* f. sp. apii are polyphyletic. *Phytopathology* 107, 463–473. doi: 10.1094/phyto-04-16-0174-r

Erskine, W., and Bayaa, B. (1996). Yield loss, incidence and inoculum density associated with vascular wilt of lentil. *Phytopathol. Mediterr.* 36, 24–32. doi: 10.2135/cropsci1996.0011183X0036000400055x

Erskine, W., Bayaa, B., and Dholli, M. (1990). The transmissibility of *Fusarium oxysporum* f. sp. lentis via seeds and the effect of some biotic and abiotic factors on its growth. *Arab. J. Plant Protect.* 8, 34–37. doi: 10.1111/j.1574-6941.2009.00777.x

Eujayl, I., Erskine, W., Bayaa, B., Baum, M., and Pehu, E. (1998). *Fusarium* vascular wilt in lentil: inheritance and identification of DNA markers for resistance. *Plant Breed.* 117, 497–499. doi: 10.1111/j.1439-0523.1998.tb01982.x

Eujayl, I., Erskine, W., Bayaa, B., Baum, M., and Pehu, E. (2006). *Fusarium* vascular wilt in lentil: inheritance and identification of DNA markers for resistance. *Plant Breed.* 117, 497–499. doi: 10.1111/j.1439-0523.1998.tb01982.x

Fall, A. L., Byrne, P. F., Jung, G., and Coyne, D. P. (2001). Detection and mapping of a major locus for *Fusarium* wilt resistance in common bean. *Crop Sci.* 41, 1494–1498. doi: 10.2135/cropsci2001.4151494x

Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137

Fourie, G., Steenkamp, E. T., Ploetz, R. C., Gordon, T. R., and Viljoen, A. (2011). Current status of the taxonomic position of *Fusarium oxysporum* formae specialis cubense within the *Fusarium oxysporum* complex. *Infec. Genet. Evol.* 11, 533–542. doi: 10.1016/j.meegid.2011.01.012

Gabe, H. L. (1975). Standardization of nomenclature for pathogenic races of *Fusarium oxysporum* f. sp. lycopersici. *Trans. Br. Mycol. Soc.* 64, 156–159. doi: 10.1016/s0007-1536(75)80089-1

Gao, C. (2018). The future of CRISPR technologies in agriculture. *Nat. Rev. Mol. Cell Biol.* 39, 1–2. doi: 10.1007/978-1-0716-0616-2_1

Garces-Fiallos, F. R., de Borba, M. C., Schmidt, E. C., Bouzon, Z. L., and Stadnik, M. J. (2017). Delayed upward colonization of xylem vessels is associated with resistance of common bean to *Fusarium oxysporum* f. Sp phaseoli. Eur. *J. Plant Pathol.* 149, 477–489. doi: 10.1007/s10658-017-1197-6

Garcia-Limones, C., Dorado, G., Navas-Cortes, J. A., Jimenez-Diaz, R. M., and Tena, M. (2009). Changes in the redox status of chickpea roots in response to infection by *Fusarium oxysporum* f. sp. ciceris: apoplastic antioxidant enzyme activities and expression of oxi- dative stress-related genes. *Plant Biol.* 11, 194–203. doi: 10.1111/j.1438-8677.2008.00095.x

Garcia-Limones, C., Hervás, A., Navas-Cortés, J. A., Jiménez-Díaz, R. M., and Tena, M. (2002). Induction of an antioxidant enzyme system and other oxidative stress markers associated with compatible and incompatible interactions between chickpea (*Cicer arietinum* L.) and *Fusarium oxysporum* f. sp. ciceris. *Physiol. Mol. Plant Pathol.* 61, 325–337. doi: 10.1006/pmpp.2003.0445

Garg, T., Mallikarjuna, B. P., Thudi, M., Samineni, S., Singh, S., Sandhu, J. S., et al. (2018). Identification of QTLs for resistance to *Fusarium* wilt and Ascochyta blight in a recombinant inbred population of chickpea (*Cicer arietinum* L.). *Euphytica* 214:45. doi: 10.1007/s10681-018-2125-3

Gaur, P. M., Pande, S., Upadhyaya, H. D., and Rao, B. V. (2006). Extra-large Kabuli chickpea with high resistance to *Fusarium* wilt. *SAT eJ* 2, 1–2. doi: 10.9734/cjast/2019/v38i630409

Gerlagh, M., and Blok, W. J. (1988). *Fusarium oxysporum* f. sp. cucurbitacearum n.f. embracing all formae speciales of *F. oxysporum* attacking cucurbitaceous crops. *Netherlands J Plant Pathol.* 94, 17–31. doi: 10.1007/bf01999804

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., and Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci U.S.A.* 115, 4613–4618. doi: 10.1073/pnas.1716999115

Ghosh, S., Watson, A., Gonzalez-Navarro, O. E., Ramirez-Gonzalez, R. H., Yanes, L., and Mendoza-Suárez, M. (2018). Speed breeding in growth cham- bers and glasshouses for crop breeding and model plant research. *Nat. Protoc.* 13, 2944–2963. doi: 10.1038/s41596-018-0072-z

Gilardi, G., Franco Ortega, S., Van Rijswick, P. C. J., Ortu, G., Gullino, M. L., and Garibaldi, A. (2017). A new race of *Fusarium oxysporum* f. sp. lactucae of lettuce. *Plant Pathol.* 66, 677–688. doi: 10.1111/ppa.12616

Gowda, S. J. M., Radhika, P., Kadoo, N., Mhase, L., and Gupta, V. (2009). Molecular mapping of wilt resistance genes in chickpea. *Mol. Breed.* 24, 177–184. doi: 10.1007/s11032-009-9282-y

Graham, P. H., and Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant Physiol.* 131, 872–877. doi: 10.1104/pp.017004

Grajal-Martin, M. J., and Muehlbauer, F. J. (2002). Genomic location of the *Fw* gene for resistance to *Fusarium* wilt race 1 in peas. *J. Hered.* 93, 291–293. doi: 10.1093/jhered/93.4.291

Gunnaiah, R., Kushalappa, A. C., Duggavathi, R., Fox, S., and Somers, D. J. (2012). Integrated metabolo-proteomic approach to decipher the mechanisms by which wheat QTL (Fhb1) contributes to resistance against *Fusarium graminearum*. *PLoS One* 7:e40695. doi: 10.1371/journal.pone.0040695

Gupta, S., Bhar, A., Chatterjee, M., and Das, S. (2013a). *Fusarium oxysporum* f. sp ciceri race 1 induced redox state alterations are coupled to downstream defense signaling in root tissues of chickpea (*Cicer arietinum* L.). *PLoS One* 8:e73163. doi: 10.1371/journal.pone.0073163

Gupta, S., Bhar, A., and Das, S. (2013b). Understanding the molecular defense responses of host during chickpea-*Fusarium* interplay: where do we stand? *Funct. Plant Biol.* 40, 1285–1297. doi: 10.1071/fp13063

Gupta, S., Chakraborti, D., Rangi, R. K., Basu, D., and Das, S. (2009). A molecular insight into the early events of chickpea (*Cicer arietinum*) and *Fusarium oxysporum* f. sp. ciceri (Race 1) interaction through cDNA-AFLP analysis. *Phytopathology* 99, 1245–1257. doi: 10.1094/phyto-99-11-1245

Gupta, S., Chakraborti, D., Sengupta, A., Basu, D., and Das, S. (2010). Primary metabolism of chickpea is the initial target of wound inducing early sensed *Fusarium oxysporum* f sp ciceri race 1. *PLos One* 5:e9030. doi: 10.1371/journal.pone.0009030

Gurjar, G., Barve, M., Giri, A., and Gupta, V. (2009). Identification of Indian pathogenic races of *Fusarium oxysporum* f. sp. ciceris with gene specific. ITS and random markers. *Mycologia* 101, 484–495. doi: 10.3852/08-085

Gurjar, G., Giri, A. P., and Gupta, V. S. (2012). Gene expression profiling during wilting in chickpea caused by *Fusarium oxysporum* f. sp. ciceri. *Am. J. Plant Sci.* 3, 190–201. doi: 10.4236/ajps.2012.32023

Haglund, W. A., and Kraft, J. M. (2001). "*Fusarium* wilt," in *Compendium of Pea Diseases*, 2nd Edn, eds J. M. Kraft and F. L. Peger, (St. Paul, MN: APS Press), 14–16.

Halila, I., Cobos, M. J., Rubio, J., Millaìn, T., Kharrat, M., Marrakchi, M., et al. (2009). Tagging and mapping a second resistance gene for fusarium wilt race 0 in chickpea. *Eur. J. Plant Pathol* 124, 87–92. doi: 10.1007/s10658-008-9395-x

Halila, M. H., and Harrabi, M. M. (1990). Breeding for dual resistance to Ascochyta and wilt diseases in chickpea. *Options Mediterr. Ser. Semin.* 9, 163–170.

Halila, M. H., and Strange, R. N. (1996). Identification of the casual agent of wilt of chickpea in Tunisia as *Fusarium oxysporum* f. sp. ciceri race 0. *Phytopathol. Mediterr.* 35, 67–74.

Hamwieh, A., Udupa, S. M., Choumane, W., Sarker, A., Dreyer, F., Jung, C., et al. (2005). A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. *Theor. Appl. Genet.* 110, 669–677. doi: 10.1007/s00122-004-1892-5

Hare, W. W. (1953). A new race of *Fusarium* causing wilt of cowpea. *Phytopathology* 43:291.

Hare, W. W., Walker, J. C., and Delwiche, E. J. (1949). Inheritance of a gene for near-wilt resistance in the garden pea. *J. Agric. Res.* 78, 239–251.

Harter, L. (1929). A Fusarium disease of beans. *Phytopathology* 19, 1–84.

Haware, M. P., Jimenez-Diaz, R. M., Amin, K. S., Phillips, J., and Halila, H. (1990). "Integrated management of wilt and root rots of chickpea," in *Chickpea in the Nineties: Proceedings of the Second International Work Shop on Chickpea Improvement*, Patancheru, 129–137.

Haware, M. P., and Nene, Y. L. (1982). Races of *Fusarium oxysporum* f. sp. ciceris. *Plant Dis.* 66, 809–810. doi: 10.1094/PD-66-809

Haware, M. P., Nene, Y. L., and Natrajan, M. (1986). "Survival of *Fusarium oxysporum* f. sp. ciceri in soil in the absence of chickpea," in *National Seminar on Management of Soil-Borne Diseases of Crop Plants (abstract)*, Coimbatore, 1.

Haware, M. P., Nene, Y. L., Pundir, R. P. S., and Rao, J. N. (1992). Screening of world chickpea germplasm for resistance to *Fusarium* wilt. *Field Crops Res.* 30, 147–154. doi: 10.1016/0378-4290(92)90063-f

Haware, M. P., Nene, Y. L., and Rajeshwari, R. (1978). Eradication of *Fusarium oxysporum* f. sp. ciceri transmitted in chickpea seed. *Phytopathology* 68, 1364–1367. doi: 10.1094/phyto-68-1364

Helms, D., Pamella, L., Buddenhagen, I. W., Workneh, F., Tucker, C. L., Foster, K. W., et al. (1992). Registration of 'UC 15' chickpea. *Crop Sci.* 32:500. doi: 10.2135/cropsci1992.0011183X003200020051x

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotech.* 37, 744–754. doi: 10.1038/s41587-019-0152-9

Hiremani, N., and Dubey, S. C. (2018). Race profiling of *Fusarium oxysporum* f. sp. lentis causing wilt in lentil. *Crop Protect.* 108, 23–30. doi: 10.1016/j.cropro.2018.02.010

Husaini, A. M., Sakina, A., and Cambay, S. R. (2018). Host–pathogen interaction in *Fusarium oxysporum* infections: where do we stand? *Mol. Plant Microbe Int.* 31, 889–898. doi: 10.1094/mpmi-12-17-0302-cr

Infantino, A., Kharrat, M., Riccioni, L., Coyne, C. J., McPhee, K. E., and Grunwald, N. J. (2006). Screening techniques and sources of resistance to root diseases in legumes. *Euphytica* 147, 201–221. doi: 10.1007/s10681-006-6963-z

Inoue, I., Namiki, F., and Tsuge, T. (2002). Plant colonization by the vascular wilt fungus *Fusarium oxysporum* requires FOW1, a gene encoding a mitochondrial protein. *Plant Cell* 14, 1869–1883. doi: 10.1105/tpc.002576

Jain, K. C., and Reddy, M. V. (1995). Inheritance of resistance to *Fusarium* wilt in pigeonpea (*Cajanus cajan* (L) Millsp.). *Indian J. Genet.* 55, 434–437.

Jain, S., Weeden, N. F., Kumar, A., Chittem, K., and McPhee, K. (2015). Functional codominant marker for selecting the Fw gene conferring resistance to *Fusarium* wilt race 1 in pea. *Crop Sci.* 55, 2639–2646. doi: 10.2135/cropsci2015.02.0102

Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9, 166–177. doi: 10.1093/bfgp/elq001

Jendoubi, W., Bouhadida, M., Millan, T., Kharrat, M., Gil, J., and Rubio, J. (2016). Identification of the target region including the Foc0 1 /foc0 1 gene and development of near isogenic lines for resistance to *Fusarium* Wilt race 0 in chickpea. *Euphytica* 210, 119–133. doi: 10.1007/s10681-016-1712-4

Jiang, S., Cheng, Q., Yan, J., Fu, R., and Wang, X. (2019). Genome optimization for improvement of maize breeding. *Theor. Appl. Genet.* 133, 1491–1502. doi: 10.1007/s00122-019-03493-z

Jimenez-Diaz, R. M., Alcala-Jimenez, A. R., Hervas, A., and Traperocasas, J. L. (1993). "Pathogenic variability and hosts resistance in the *Fusarium oxysporum* f. sp. ciceris/*Cicer arietinum* pathosys- tem,"in *Proceedings of the 3rd European Seminar on Fusarium Mycotoxins, Taxonomy, Pathogenicity and Host Resistance. Hodowsla Roslin Aklimatyazacja i Nasiennictwo*. Poland: Plant Breed- ing and Acclimatization Institute, 87–94.

Jiménez-Díaz, R. M., Castillo, P., Jimeìnez-Gasco, M. M., Landa, B. B., and Navas-Corteìs, J. A. (2015). *Fusarium* wilt of chickpeas: biology, ecology and management. *Crop Prot.* 73, 16–27. doi: 10.1016/j.cropro.2015.02.023

Jiménez-Gasco, M. M., and Jiménez-Díaz, R. M. (2003). Development of a specific polymerase chain reaction-based assay for the identification of *Fusarium oxysporum* f. sp. ciceris and its pathogenic races 0, 1A, 5, and 6. *Phytopathology* 93, 200–209. doi: 10.1094/phyto.2003.93.2.200

Jiménez-Gasco, M. M., Pérez-Artés, E., and Jiménez-Díaz, R. M. (2001). Identification of pathogenic races 0, 1B/C, 5, and 6 of *Fusarium oxysporum* f. sp. ciceris with random amplified polymorphic DNA (RAPD). *Eur. J. Plant Pathol.* 107, 237–248.

Jimeìnez-Diìaz, R. M., Singh, K. B., Trapero-Casas, A., and Trapero-Casas, J. L. (1991). Resistance in kabuli chickpeas to *Fusarium* wilt. *Plant Dis.* 75, 914–918. doi: 10.1094/pd-75-0914

Jingade, P., and Ravikumar, R. L. (2015). Development of molecular map and identification of QTLs linked to *Fusarium* wilt resistance in chickpea. *J. Genet.* 94, 723–729. doi: 10.1007/s12041-015-0589-7

Kaiser, W. J., Alcalá-Jiménez, A. R., Hervás-Vargas, A., Trapero-Casas, J. L., and Jiménez-Díaz, R. M. (1994). Screening wild species for resistance to races 0 and 5 of *Fusarium oxysporum* f. sp. ciceris. *Plant Dis.* 78, 962–967.

Kaiser, W. J., Ramsay, M. D., Makkouk, K. M., Bretag, T. W., Acikgoz, N., Kumar, J., et al. (2000). "Foliar diseases of cool season food legumes and their control," in *Linking Research and Marketing Opportunities for Pulses in the 21st Century*, ed. R. Knight (Dordrecht: Kluwer Academic Publishers), 437–455. doi: 10.1007/978-94-011-4385-1_41

Kamboj, R. K., Pandey, M. P., and Chaube, H. S. (1990). Inheritance of resistance to *Fusarium* wilt in Indian lentil germplasm (*Lens culinaris* Medik.). *Euphytica* 50, 113–117. doi: 10.1007/bf00023633

Kannaiyan, J., and Nene, Y. L. (1976). Reaction of lentil germplasm and cultivars against three root pathogens. *Indian J Agric. Sci.* 46, 165–167.

Karimi, R., Owuoche, J. O., and Silim, S. N. (2010). Inheritance of fusar- ium wilt resistance in pigeonpea (*Cajanus cajan* (L.) *Millspugh. Indian J. Genet.* 70, 271–276.

Kelly, A. G., Alcala-Jimenez, A. R., Bainbridge, B. W., Heale, J. B., Perez- Artes, E., and Jimenez-Diaz, R. M. (1994). Use of genetic fingerprinting and random amplified polymorphic DNA to characterize pathotypes of *Fusarium oxysporum* f. sp. ciceris infecting chickpea. *Phytopathology* 84, 1293–1298. doi: 10.1094/phyto-84-1293

Kistler, H. C. (1997). Genetic diversity in the plant-pathogenic fungus *Fusarium oxysporum*. *Phytopathology* 87, 474–479. doi: 10.1094/phyto.1997.87.4.474

Klessig, D. F., Durner, J., Shah, J., and Yang, Y. (1998). Salicylic acid-mediated signal transduction in plant disease resistance. *Phytochem. Signal Plant Microbe Interact.* 32, 119–137. doi: 10.1007/978-1-4615-5329-8_7

Kohli, D., Joshi, G., Deokar, A. A., Bhardwaj, A. R., Agarwal, M., Katiyar- Agarwal, S., et al. (2014). Identi cation and characterization of wilt and salt stress-responsive microRNAs in chickpea through high-throughput sequencing. *PLoS One* 9:e108851. doi: 10.1371/journal.pone.0108851

Kotresh, H., Fakrudin, B., Punnuri, S. M., Rajkumar, B. K., Thudi, M., Paramesh, H., et al. (2006). Identification of two RAPD markers genetically linked to a recessive allele of a *Fusarium* wilt resistance gene in pigeonpea (*Cajanus cajan* L. Millsp.). *Euphytica* 149, 113–120. doi: 10.1007/s10681-005-9059-2

Kraft, J. M. (1994). *Fusarium* wilt of peas a review. *Agronomie* 14, 561–567. doi: 10.1051/agro:19940901

Kraft, J. M., Burke, D. W., and Haglund, W. A. (1981). "*Fusarium* diseases of beans, peas and lentils," in *Fusarium: Diseases, Biology and Taxonomy*, eds P. E. Nelson, T. A. Toussoun, and R. J. Cook (University Park, PA: Pennsylvania State Univ. Press,), 142–156.

Kraft, J. M., Larsen, R. C., and Inglis, D. A. (1998). "Diseases of pea," in *The Pathology of Food and Pasture Legumes*, eds D. J. Allen and J. M. Lenneì (WALLINGFORD: . CAB International), 325–370.

Kreplak, J., Madoui, M. A., Capal, P., Novak, P., Labadie, K., Aubert, G., et al. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* 51, 1411–1422.

Kumar, J., Haware, M. P., and Smithson, J. B. (1985). Registration of four short duration *Fusarium* wilt-resistant kabuli (garbanzo) chickpea germplasms. *Crop Sci.* 25, 576–577. doi: 10.2135/cropsci1985.0011183x002500030047x

Kumar, S. (1998). Inheritance of resistance to *Fusarium* wilt (race 2) in chickpea. *Plant Breed.* 117, 139–142. doi: 10.1111/j.1439-0523.1998.tb01467.x

Kumar, S. (2006). Inheritance of resistance to *Fusarium* wilt (race 2) in chickpea. *Plant Breed.* 117, 139–142. doi: 10.1111/j.1439-0523.1998.tb01467.x

Kumar, Y., Dholakia, B. B., Panigrahi, P., Kadoo, N. Y., Giri, A. P., and Gupta, V. S. (2015). Metabolic profiling of chickpea-*Fusarium* interaction identifies differential modulation of disease resistance pathways. *Phytochemistry* 116, 120–129. doi: 10.1016/j.phytochem.2015.04.001

Kumar, Y., Zhang, L., Panigrahi, P., Dholakia, B. B., Dewangan, V., Chavan, S. G., et al. (2016). *Fusarium oxysporum* mediates systems metabolic reprogramming of chickpea roots as revealed by a combination of proteomics and metabolomics. *Plant Biotechnol. J.* 14, 1589–1603. doi: 10.1111/pbi.12522

Kunkel, B. N., and Brooks, D. M. (2002). Cross talk between signaling pathways in pathogen defense. *Curr. Opin. Plant Biol.* 5, 325–331. doi: 10.1016/s1369-5266(02)00275-3

Kwon, S. J., Smykal, P., Hu, J., Kim, S. J., McGee, R. J., McPhee, K., et al. (2013). User-friendly markers linked to *Fusarium* wilt race 1 resistance *Fw* gene for marker-assisted selection in pea. *Plant Breed.* 132, 642–648. doi: 10.1111/pbr.12085

Landa, B. B., Navas-Cortes, J. A., Jimenez-Gasco, M. M., Katan, R. B., and Jimenez-Diìaz, R. M. (2006). Temperature response of chickpea cultivars to races of Fusa- rium oxysporum f. sp. ciceris, the causal agent of *Fusarium* wilt. *Plant Dis.* 90, 365–374. doi: 10.1094/pd-90-0365

Langner, T., Kamoun, S., and Belhaj, K. (2018). CRISPR crops: plant genome editing toward disease resistance. *Annu. Rev. Phytopathol.* 56, 479–512. doi: 10.1146/annurev-phyto-080417-050158

Lanubile, A., Muppirala, U. K., Severin, A. J., Marocco, A., and Munkvold, G. P. (2015). Transcriptome profiling of soybean (*Glycine max*) roots challenged with pathogenic and non-pathogenic isolates of *Fusarium oxysporum*. *BMC Genomics* 16:1089. doi: 10.1186/s12864-015-2318-2

Leitão, S. T., Malosetti, M., Song, Q., van Eeuwijk, F., Rubiales, D., and Vaz Patto, M. C. (2020). natural variation in portuguese common bean germplasm reveals new sources of resistance Against *Fusarium oxysporum* f. sp. phaseoli and Resistance-Associated Candidate Genes. *Phytopathology* 110, 633–647. doi: 10.1094/phyto-06-19-0207-r

Leslie, J. F., and Summerell, B. A. (2006). *The Fusarium Laboratory Manual*. Ames, IA: Blackwell Publishing, doi: 10.1002/9780470278376

Li, C. Y., Deng, G. M., Yang, J., Viljoen, A., Jin, Y., Kuang, R. B., et al. (2012). Transcriptome profiling of resistant and susceptible Cavendish banana roots

following inoculation with *Fusarium oxysporum* f. sp. cubense tropical race 4. *BMC Genom.* 13:374. doi: 10.1186/1471-2164-13-374

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199

Li, W., Deng, Y. W., Ning, Y. S., Hu, Z. H., and Wang, G. L. (2020). Exploiting broad-spectrum disease resistance in crops: from molecular dissection to breeding. *Annu. Rev. Plant Biol.* 16, 1–25. doi: 10.1007/s11032-009-9323-6

Lievens, B., Rep, M., and Thomma, B. P. H. J. (2008). Recent development in the molecular discrimination of formae speciales of *Fusarium oxysporum*. *Pest Manage. Sci.* 64, 781–788.

Liu, S., Li, J., Zhang, Y., Liu, N., Viljoen, A., Mostert, D., et al. (2020). Fusaric acid instigates the invasion of banana by *Fusarium oxysporum* f. sp. cubense TR4. *New Phytol.* 225, 913–929. doi: 10.1111/nph.16193

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* 98, 767–782.

López-Díaz, C., Rahjoo, V., Sulyok, M., Ghionna, V., Logrieco, A. F., Franzoni, M., et al. (2017). Fusaric acid contributes to virulence of *Fusarium oxysporum* on plant and mammalian hosts. *Mol. Plant Pathol.* 19, 440–453. doi: 10.1111/mpp.12536

Loridon, K., McPhee, K., Morin, J., Dubreuil, P., Pilet-Nayel, M. L., Aubert, G., et al. (2005). Microsatellite marker polymor- phism and mapping in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 11, 1022–1031. doi: 10.1007/s00122-005-0014-3

Lowe, A., Harrison, N., and French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* 13:80.

Ma, K. W., and Ma, W. (2016). Phytohormone pathways as targets of pathogens to facilitate infection. *Plant Mol. Biol.* 91, 713–725. doi: 10.1007/s11103-016-0452-0

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1094/pdis-03-15-0340-fe

Mahlein, A. K., Oerke, E. C., Steiner, U., and Dehne, H. W. (2012). Recent advances in sensing plant diseases for precision crop protection. *Eur. J. Plant Pathol.* 133, 197–209. doi: 10.1007/s10658-011-9878-z

Mahmoud, A. F., and Abd El-Fatah, B. E. (2020). Genetic diversity studies and identification of molecular and biochemical markers associated with *Fusarium* wilt resistance in cultivated faba bean (*Vicia faba*). *Plant Pathol. J.* 36, 11–28.

Mannur, D. M., Babbar, A., Thudi, M., Sabbavarapu, M. M., Roorkiwal, M., Sharanabasappa, B. Y., et al. (2019). Super Annigeri 1 and improved JG 74: two *Fusarium* wilt-resistant introgression lines developed using marker-assisted backcrossing approach in chickpea (*Cicer arietinum* L.). *Mol. Breed.* 39:2.

Mayer, M., Tullu, A., Simon, C., Kumar, J., Kaiser, W., Kraft, J., et al. (1997). Development of a DNA marker for *Fusarium* wilt resistance in chickpea. *Crop Sci.* 37, 1625–1637. doi: 10.2135/cropsci1997.0011183x003700050036x

McClendon, M. T., Inglis, D. A., McPhee, K. E., and Coyne, C. J. (2002). DNA markers linked to *Fusarium* wilt race 1 resistance in pea. *J. Am. Soc. Hortic. Sci.* 127, 602–607. doi: 10.21273/jashs.127.4.602

McPhee, K. E., Inglis, D. A., Gundersen, B., and Coyne, C. J. (2012). Mapping QTL for *Fusarium* wilt race 2 partial resistance in pea (*Pisum sativum*). *Plant Breed.* 131, 300–306. doi: 10.1111/j.1439-0523.2011.01938.x

McPhee, K. E., Tullu, A., Kraft, J. M., and Muehlbauer, F. J. (1999). Resistance to *Fusarium* wilt race 2 in the *Pisum* core collection. *J. Amer. Soc. Hort. Sci.* 124, 28–31. doi: 10.21273/jashs.124.1.28

Mehta, A., Brasileiro, A. C. M., Souza, D. S. L., Romano, E., Campos, M. A., Grossi-De-Sa, M. F., et al. (2008). Plant-pathogen interactions: what is proteomics telling us? *FEBS J.* 275, 3731–3746. doi: 10.1111/j.1742-4658.2008.06528.x

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Michielse, C. B., van Wijk, R., Reijnen, L., Cornelissen, B. J. C., and Rep, M. (2009). Insight into the molecular requirements for pathogenicity of *Fusarium oxysporum* f. sp. lycopersici through large-scale insertional mutagenesis. *Genome Biol.* 10:R4.

Mohammadi, N., Puralibaba, H., Goltapeh, E. M., Ahari, A. B., and Sardrood, B. P. (2012). Advanced lentil lines screened for resistance to *Fusarium oxysporum* f. sp. lentis under greenhouse and field conditions. *Phytoparasitica* 40, 69–76. doi: 10.1007/s12600-011-0201-5

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419

Morales, G. J. A. (1986). Chickpea breeding programme in Sonora. *Mexico. Int. Chickpea Newsl.* 15, 11–12.

Morkunas, I., Kozłowska, M., Ratajczak, L., and Marczak, Ł (2007). Role of sucrose in the development of *Fusarium* wilt in lupine embryo axes. *Physiol. Mol. Plant Pathol.* 70, 25–37. doi: 10.1016/j.pmpp.2007.05.004

Morkunas, I., Marczak, Ł, Stachowiak, J., and Stobiecki, M. (2005). Sucrose-stimulated accumulation of isoflavonoids as a defense response of lupine to *Fusarium oxysporum*. *Plant Physiol. Biochem.* 43, 363–373. doi: 10.1016/j.plaphy.2005.02.011

Morkunas, I., and Ratajczak, L. (2014). The role of sugar signaling in plant defense responses against fungal pathogens. *Acta Physiol. Plant.* 36, 1607–1619. doi: 10.1007/s11738-014-1559-z

Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., et al. (2009). A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18159–18164. doi: 10.1073/pnas.0905886106

Mudryj, A. N., Yu, N., and Aukema, H. M. (2014). Nutritional and health benefits of pulses. *Appl. Physiol. Nutr. Metab.* 39, 1197–1204.

Naimuddin, and Chaudhary, R. G. (2009). Pathogenic variability in isolates of *Fusarium oxysporum* f. sp. lentis. *Trends Biosci.* 2, 50–52.

Nakagami, H., Pitzschke, A., and Hirt, H. (2005). Emerging MAP kinase pathways in plant stress signalling. *Trends Plant Sci.* 10, 339–346. doi: 10.1016/j.tplants.2005.05.009

Navas-Cortes, J. A., Hau, B., and Jimenez-Diìaz, R. M. (1998). Effect of sowing date, host cultivar, and race of *Fusarium oxysporum* f. sp. ciceris on development of *Fusarium* wilt of chickpea. *Phytopathology* 88, 1338–1346. doi: 10.1094/phyto.1998.88.12.1338

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* 7:482.

Nene, Y. L. (1980). Screening chickpea for reìsistance to wilt. *Plant Dis.* 64, 379–380. doi: 10.1094/pd-64-379

Nene, Y. L., and Haware, M. P. (1980). screening chickpea for resistance to Wilt. *Plant Dis.* 64, 379–380. doi: 10.1094/pd-64-379

Nene, Y. L., Haware, M. P., Reddy, N. M. V., Philps, J. P., Castro, E. L., Kotasthane, S. R., et al. (1989). Identification of broad based and stable resistance to wilt and root-rots in chickpea. *Indian Phytopathol.* 42, 499–505.

Neumann, S., and Xue, A. G. (2003). Reactions of field pea cultivars to four races of *Fusarium oxysporum* f. sp pisi. *Can. J. Plant Sci.* 83, 377–379. doi: 10.4141/p02-048

Nikraftar, F., Taheri, P., Rastegar, M. F., and Tarighi, S. (2013). Tomato partial resistance to Rhizoctonia solani involves antioxidative defense mechanisms. *Physiol. Mol. Plant Pathol.* 81, 74–83. doi: 10.1016/j.pmpp.2012.11.004

Ninþo-Saìnchez, J., Tello, V., Casado-del Castillo, V., Thon, M. R., Benito, E. P., and Diìaz-Miìnguez, J. M. (2015). Gene expression patterns and dynamics of the colonization of common bean (*Phaseolus vulgaris* L.) by highly virulent and weakly virulent strains of *Fusarium oxysporum*. *Front. Microbiol* 6:234. doi: 10.3389/fmicb.2015.00234

Nogales Moncada, A. N., Jimenez-Diìaz, R. M., and Perez-Artes, E. (2009). Vegetative compatibility groups in *Fusarium oxysporum* f. sp. ciceris and *F. oxysporum* nonpathogenic to chickpea. *J. Phytopathol.* 157, 729–735. doi: 10.1111/j.1439-0434.2009.01562.x

Odeny, D. A., Githiri, S. M., and Kimani, P. M. (2009). Inheritance of resistance to *Fusarium* wilt in pigeonpea. *J. Anim. Plant Sci.* 2, 89–95.

Okiror, M. A. (2002). Genetics of resistance to *Fusarium udum* in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Ind. J. Genet. Plant Breed.* 62, 218–220.

Okubara, P. A., Keller, K. E., Mcclendon, M. T., Inglis, D. A., Mcphee, K. E., and Coyne, C. J. (2005). Y15_999fw, a dominant scar marker linked to the *Fusarium* wilt race 1 (fw) resistance gene in pea. *Pisum Genet.* 37, 32–35.

Omoigui, L. O., Danmaigona, C. C., Kamara, A. Y., Ekefan, E. J., and Timko, M. P. (2018). Genetic analysis of *Fusarium* wilt resistance in cowpea (*Vigna unguiculata* Walp.). *Plant Breed.* 137, 773–781. doi: 10.1111/pbr.12628

Pal, B. P. (1934). Recent progress in plant breeding at Pusa. *Agric. Livest India* 4, 505–515.

Palomares-Rius, J. E., Castillo, P., Navas-Corteìs, J. A., Jimeìnez-Diìaz, R. M., and Tena, M. (2011). A proteomic study of in-root interactions between chickpea pathogens: the root-knot nematode Meloidogyne artiellia and the soil-borne fungus *Fusarium oxysporum* f. sp. ciceris race 5. *J. Proteom.* 74, 2034–2051. doi: 10.1016/j.jprot.2011.05.026

Pandey, R. N., Pawar, S. E., and Bhatia, C. R. (1996). Inheritance of wilt resistance in pigeonpea. *Indian J. Genet.* 56, 305–308.

Parihar, A. K., Basandrai, A. K., Saxena, D. R., Kushwaha, K. P. S., Chandra, S., Sharma, K., et al. (2017). Biplot evaluation of test environments and identification of lentil genotypes with durable resistance to *Fusarium* wilt in India. *Crop Pasture Sci.* 68, 1024–1030. doi: 10.1071/cp17258

Pastor-Corrales, M. A., and Abawi, G. S. (1987). Reactions of selected bean germ plasms to infection by *Fusarium oxysporumf.* sp. phase- oli. *Plant Dis.* 71, 990–993. doi: 10.1094/pd-71-0990

Patel, P. N. (1985). "Fungal, bacterial and viral diseases of cowpeas in the USA," in *Cowpea Research, Production and Utilization*, eds S. R. Singh and K. O. Rachie (Chichester: John Wiley and Sons), 205–213.

Pathak, G. N. (1970). "Red gram," in *Pulse Crops of India*, (New Delhi: Indian Council of Agricultural Research), 14–53.

Patil, B. S., Ravikumar, R. L., Bhat, J. S., and Soregaon, D. (2014). Molecular mapping of QTLs for resistance to early and late *Fusarium* wilt in chickpea. *Czech. J. Genet. Plant Breed.* 50, 171–176. doi: 10.17221/188/2013-cjgpb

Patil, P. G., Bohra, A., Dubey, J., Saabale, P. R., Mishra, R. K., Pandey, G., et al. (2017a). Genetic analysis and molecular resistance to race 2 of *Fusarium* wilt in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Crop Protect.* 100, 117–123. doi: 10.1016/j.cropro.2017.06.016

Patil, P. G., Dubey, J., Bohra, A., Mishra, R. K., Saabale, P. R., Das, A., et al. (2017b). Association mapping to discover significant marker-trait associations for resistance against *Fusarium* wilt variant 2 in pigeonpea [*Cajanus cajan* (L.) Millspaugh] using SSR markers. *J. Appl. Genet.* 58, 307–319. doi: 10.1007/s13353-017-0400-y

Pawar, N. B., and Mayee, C. D. (1986). Reaction of pigeonpea genotypes and their crosses to *Fusarium* wilt. *Indian Phytopathol.* 39, 70–74.

Pedley, K. F., and Martin, G. B. (2005). Role of mitogen-activated protein kinases in plant immunity. *Curr. Opin. Plant Biol.* 8, 541–547. doi: 10.1016/j.pbi.2005.07.006

Peng, A., Chen, S., Lei, T., Xu, L., He, Y., and Wu, L. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLob1 promoter in citrus. *Plant Biotechnol. J.* 15, 1509–1519. doi: 10.1111/pbi.12733

Poland, J., and Rutkoski, J. (2016). Advances and challenges in genomic selection for disease resistance. *Annu. Rev. Phytopathol.* 54, 79–98. doi: 10.1146/annurev-phyto-080615-100056

Pottorff, M., Wanamaker, S., Ma, Y. Q., Ehlers, J. D., Roberts, P. A., and Close, T. J. (2012). Genetic and physical mapping of candidate genes for resistance to *Fusarium oxysporum* f. sp. tracheiphilum race 3 in cowpea [*Vigna unguiculata* (L.) Walp]. *PLoS One* 7:e41600. doi: 10.1371/journal.pone.0041600

Pottorff, M. O., Li, G., Ehlers, J. D., Close, T. J., and Roberts, P. A. (2014). Genetic mapping, synteny, and physical location of two loci for *Fusarium oxysporum* f. sp. tracheiphilum race 4 resistance in cowpea [*Vignaunguiculata* (L.) Walp]. *Mol. Breed.* 33, 779–791. doi: 10.1007/s11032-013-9991-0

Pouralibaba, H. R., Rubiales, D., and Fondevilla, S. (2015). Identification of resistance to *Fusarium oxysporum* f. sp. lentis in Spanish lentil germplasm. *Eur. J. Plant Pathol.* 143, 399–405. doi: 10.1007/s10658-015-0692-x

Pouralibaba, H. R., Rubiales, D., and Fondevilla, S. (2016). Identification of pathotypes in *Fusarium oxysporum* f. sp. lentis. *Eur. J. Plant Pathol.* 144, 539–549. doi: 10.1007/s10658-015-0793-6

Pratap, A., Chaturvedi, S. K., Tomar, R., Rajan, N., Malviya, N., Thudi, M., et al. (2017). Marker-assisted introgression of resistance to *Fusarium* wilt race 2 in Pusa 256, an elite cultivar of desi chickpea. *Mol. Gen. Genomics* 293, 1237–1245. doi: 10.1007/s00438-017-1343-z

Ratnaparkhe, M., Santra, D., Tullu, A., and Muehlbauer, F. (1998). Inheritance of inter-simple-sequence-repeat polymorphisms and linkage with a *Fusarium* wilt resistance gene in chickpea. *Theor. Appl. Genet.* 96, 348–352. doi: 10.1007/s001220050747

Reddy, M. V., Nene, Y. L., Raju, T. N., Kannaiyan, J., Remanandan, P., Mengesha, M. H., et al. (1995). Registration of pigeonpea germplasm line ICP 9145 resistant to *Fusarium* wilt. *Crop Sci.* 35:1231. doi: 10.2135/cropsci1995.0011183x003500040092x

Reddy, M. V., Sharma, S. B., and Nene, Y. L. (1990). "Pigeonpea: disease management," in *The Pigeonpea*, eds Y. L. Nene, S. D. Hall, and V. K. Sheila (Wallingford: CAB International), 303–348.

Rep, M., Dekker, H. L., Vossen, J. H., de Boer, A. D., Houterman, P. M., Speijer, D., et al. (2002). Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected tomato. *Plant Physiol.* 130, 904–917. doi: 10.1104/pp.007427

Rep, M., and Kistler, H. C. (2010). The genomic organization of plant pathogenicity in *Fusarium* species. *Curr. Opin. Plant Biol.* 13, 420–426. doi: 10.1016/j.pbi.2010.04.004

Ribeiro, R. L. D., and Hagedorn, D. J. (1979). Inheritance and nature of resistance in beans to *Fusarium oxysporum* f. sp. phaseoli. *Phytopathology* 69, 859–861.

Rigert, K. S., and Foster, K. W. (1987). Inheritance of resistance to two races of *Fusarium* wilt in three cowpea cultivars. *Crop Sci.* 27, 220–224. doi: 10.2135/cropsci1987.0011183x002700020018x

Rispail, N., Bani, M., and Rubiales, D. (2015). Resistance reaction of Medicago truncatula genotypes to *Fusarium oxysporum*: effect of plant age, substrate and inoculation method. *Crop Pasture Sci.* 66, 506–515. doi: 10.1071/cp14216

Rispail, N., and Rubiales, D. (2015). Rapid and efficient estimation of pea resistance to the soil-borne pathogen *Fusarium oxysporum* by infrared imaging. *Sensors* 15, 3988–4000. doi: 10.3390/s150203988

Risser, G., Banihashemi, Z., and Davis, D. W. (1976). A proposed nomenclature of *Fusarium oxysporum* f. sp. melonis races and resistance genes in Cucumis melo. *Phytopathology* 66, 1105–1106. doi: 10.1094/phyto-66-1105

Roberts, P. A., Frate, C. A., Matthews, W. C., and Osterli, P. P. (1995). Interactions of virulent *Meloigyne incognita* and *Fusarium* wilt on resistant cowpea genotypes. *Phytopathology* 85, 1288–1295. doi: 10.1094/phyto-85-1288

Rodriguez, M. C., Petersen, M., and Mundy, J. (2010). Mitogen-activated protein kinase signalling in plants. *Annu Rev Plant Biol.* 61, 621–649.

Rolland, F., Moore, B., and Sheen, J. (2002). Sugar sensing and signaling in plants. *Plant Cell* 14, S185–S205.

Römer, C., Wahabzada, M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., et al. (2012). Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. *Funct. Plant Biol.* 39, 878–890. doi: 10.1071/fp12060

Rousseau, C., Belin, E., Bove, E., Rousseau, D., Fabre, F., and Berruyer, R. (2013). High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis. *Plant Methods* 9:17. doi: 10.1186/1746-4811-9-17

Rubab, A., Chaudhary, R. A., Farah, N., and Ghulam, S. (2014). Surveillance and morphological characterization of *Fusarium* isolates associated with lentil wilt. *Pakistan J. Phytopathol.* 26, 85–88.

Rubiales, D., Fondevilla, S., Chen, W., Gentzbittel, L., Higgins, T. J. V., and Castillejo, M. A. (2015). Achievements and challenges in legume breeding for pest and disease resistance. *Crit. Rev. Plant Sci.* 34, 195–236. doi: 10.1080/07352689.2014.898445

Rubio, J., Hajj-Moussa, E., Kharrat, M., Moreno, M. T., Millan, T., and Gil, J. (2003). Two genes and linked RAPD markers involved in resistance to *Fusarium oxysporum* f. sp. cicerisrace 0 in chickpea. *Plant Breed.* 122, 188–191. doi: 10.1046/j.1439-0523.2003.00814.x

Sabbavarapu, M. M., Sharma, M., Chamarthi, S. K., Swapna, N., Rathore, A., Thudi, M., et al. (2013). Molecular mapping of QTLs for resistance to *Fusarium* wilt (race 1) and Ascochytablight in chickpea (*Cicer arietinum* L.). *Euphytica* 193, 121–138. doi: 10.1007/s10681-013-0959-2

Salgado, M. O., Schwartz, H. F., and Brick, M. A. (1995). Inheritance of resistance to a Colorado race of *Fusarium oxysporum* f. sp. phaseoli in common beans. *Plant Dis.* 79, 279–281. doi: 10.1094/pd-79-0279

Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355.

Sarker, A., Bayaa, B., and Erskine, W. (2001). Registration of six lentil germplasm lines with resistance to vascular wilt. *Crop Sci.* 41:1655. doi: 10.2135/cropsci2001.4151655x

Saxena, K. B. (2008). Genetic improvement of pigeonpea-a review. *Trop. Plant Biol.* 1, 159–178.

Saxena, K. B., Kumar, R. V., Saxena, R. K., Sharma, M., Srivastava, R. K., Sultana, R., et al. (2012). Identification of dominant and recessive genes for resistance to *Fusarium* wilt in pigeonpea and their implication in breeding hybrids. *Euphytica* 188, 221–227. doi: 10.1007/s10681-012-0700-6

Saxena, R. K., Singh, V. K., Kale, S. M., Tathinen, S., Parupalli, S., Kumar, V., et al. (2017). Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for *Fusarium* wilt resistance in pigeonpea. *Sci. Rep.* 7:1911.

Scarafoni, A., Ronchi, A., Prinsi, B., Espen, L., Assante, G., Venturini, G., et al. (2013). The proteome of exudates from germinating *Lupinus albus* seeds is secreted through a selective dual-step process and contains proteins involved in plant defence. *FEBS J.* 280, 1443–1459. doi: 10.1111/febs.12140

Schäfer, W. (1994). Molecular mechanisms of fungal pathogenicity to plants. *Annu. Rev. Phytopathol.* 32, 461–477. doi: 10.1146/annurev.py.32.090194.002333

Schippers, B., and Van Eck, W. H. (1981). "Formation and survival of chlamydospores in *Fusarium*," in *Fusarium: Disease, Biology, and Taxonomy*, eds P. Nelson, T. A. Tousson, and R. J. Cook (University Park, CO: Penn State University Press), 250–260.

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008

Shah, T. M., Atta, B. M., Mirza, J. I., and Haq, M. A. (2009). Screening of chickpea (Cicer arietinum) induced mutants against Fusarium wilt. *Pak. J. Bot.* 41, 1945–1955.

Shakoor, N., Lee, S., and Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Curr. Opin. Plant Biol.* 38, 184–192. doi: 10.1016/j.pbi.2017.05.006

Shao, C., Xiang, D., Wei, H., Liu, S., Guo, L., Li, C., et al. (2020). Predicting Virulence of *Fusarium Oxysporum* F. Sp. cubense based on the production of mycotoxin using a linear regression model. *Toxins* 12:254. doi: 10.3390/toxins12040254

Sharma, K. D., Chen, W., and Muehlbauer, F. J. (2005). Genetics of chickpea resistance to five races of *Fusarium* wilt and a concise set of race differentials for *Fusarium oxysporum* f. sp. ciceris. *Plant Dis.* 89, 385–395. doi: 10.1094/pd-89-0385

Sharma, K. D., and Muehlbauer, F. J. (2007). *Fusarium* wilt of chickpea: physiological specialization, genetics of resistance and resistance gene tagging. *Euphytica* 157, 1–14. doi: 10.1007/s10681-007-9401-y

Sharma, K. D., Winter, P., Kahl, G., and Muehlbauer, F. J. (2004). Molecular mapping of *Fusarium oxysporumf.* sp. ciceris race 3 resistance gene in chickpea. *Theor. Appl. Genet.* 108, 1243–1256. doi: 10.1007/s00122-003-1561-0

Sharma, M., Ghosh, R., Tarafdar, A., Rathore, A., Kumar, A. V., Chobe, D. R., et al. (2019). Exploring the genetic cypher of chickpea (*Cicer arietinum* L.) through Identification and Multi-environment Validation of Resistant Sources against *Fusarium* wilt (*Fusarium oxysporum* f. sp. ciceris). *Front. Sust. Food Syst.* 3:78. doi: 10.3389/fsufs.2019.00078

Sharma, M., Ghosh, R., Telangre, R., Rathore, A., Saifulla, M., Mahalinga, D. M., et al. (2016a). Environmental influences on pigeon pea-*Fusarium* udum interactions and stability of genotypes to *Fusarium* wilt. *Front. Plant Sci.* 7:253. doi: 10.3389/fpls.2016.00253

Sharma, M., Sengupta, A., Ghosh, R., Agarwal, G., Tarafda, A., and Nagavardhini, N. (2016b). Genome wide transcriptome profiling of *Fusarium oxysporum* f sp. ciceris conidial germination reveals new insights into infection-related genes. *Sci. Rep.* 6:37353.

Sharma, M., Kiran Babu, T., Gaur, P. M., Ghosh, R., Rameshwar, T., Chaudhary, R. G., et al. (2012). Identification and multi-environment validation of resistance to *Fusarium oxysporum* f. spcicerisin chickpea. *Field Crop Res.* 135, 82–88. doi: 10.1016/j.fcr.2012.07.004

Sharma, M., Nagavardhini, A., Thudi, M., Ghosh, R., Pande, S., and Varshney, R. K. (2014). Development of DArT markers and assessment of diversity in *Fusarium oxysporum f.* sp. ciceris, wilt pathogen of chickpea (*Cicer arietinum* L.). *BMC Genomics* 15:454. doi: 10.1186/1471-2164-15-454

Sharma, M., and Pande, S. (2013). Unravelling effects of temperature and soil moisture stress response on development of dry root rot [*Rhizoctonia bataticola* (Taub.) Butler] in chick- pea. *Am. J. Plant Sci.* 4, 584–589. doi: 10.4236/ajps.2013.43076

Sharma, M., Pande, S., Mangala, U. N., Tripathi, S., and Gaur, P. M. (2010). Genetic resistance in desi and kabuli chickpea lines to *Fusarium* wilt caused by *Fusarium oxysporum* f. sp. *ciceris*. *Indian J. Plant Prot.* 38, 57–62.

Shubha, K., Dhar, S., Choudhary, H., Dubey, S. C., and Sharma, R. K. (2016). Identification of resistant sources and inheritance of *Fusarium* wilt resistance in garden pea (*Pisum sativum* ssp. hortense). *Indian J. Hort.* 73, 356–361. doi: 10.5958/0974-0112.2016.00077.3

Silvia Sebastiani, M., Bagnaresi, P., Sestili, S., Biselli, C., Zechini, A., OrruÌ, L., et al. (2017).). Transcriptome analysis of the melon-*Fusarium oxysporum* f. sp. melonis race 1.2 pathosystem in susceptible and resistant plants. *Front. Plant Sci.* 8:362. doi: 10.3389/fpls.2017.00362

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Singh, D., Sinha, B., Rai, V. P., Singh, M. N., Singh, D. K., Kumar, R., et al. (2016). Genetics of *Fusarium* wilt resistance in pigeonpea (*Cajanuscajan*) and efficacy of associated SSR markers. *Plant Pathol. J.* 32, 95–101. doi: 10.5423/ppj.oa.09.2015.0182

Singh, D. V., Lal, S., and Singh, S. N. (1974). Breeding gram(*Cicer arietinum* L.) for resistance to wilt. *Indian J. genet. Plant Breed.* 34, 267–270.

Singh, F., Singh, I. P., and Mazumder, N. D. (2011a). Identification of *Fusarium* wilt resistant sources of long-duration pigeonpea (*Cajanus cajan*). *Indian J. Agr. Sci.* 81, 1046–1051.

Singh, F., Singh, I. P., Majumder, N. D., and Katiyar, P. K. (2011b). IPA 16 F (IC0574574; INGR10021), a Pigeonpea (*Cajanuscajan*) Germplasm with Resistance to *Fusarium* Wilt, Donor for Resistance to Sterility Mosaic Disease of Pigeon Pea. *Indian J. Plant Genet. Resour.* 24.

Singh, H., Kumar, J., and Haware, M. P. (1987a). "Genetics of resistance to *Fusarium* wilt in chickpeas," in *Genetics and Plant Pathogenesis*, eds P. R. Day and G. J. Jellis (Oxford: Blackwell Scientific Publications), 339–342.

Singh, H., Kumar, J., Smithson, J. B., and Haware, M. P. (1987b). Complementation between genes for resistance to race 1 of *Fusarium oxyspomm* f. sp. ciceri in chickpea. *Plant Pathol.* 36, 539–543. doi: 10.1111/j.1365-3059.1987.tb02271.x

Singh, I. P., Bohra, A., and Singh, F. (2016). An overview of varietal development programme of pigeonpea in India. *Legume Perspect.* 11, 37–40.

Singh, K. B., Ocampo, B., and Robertson, L. D. (1998). Diversity for abiotic and biotic stress resistance in the wild annual *Cicer* species. *Genet. Resour. Crop Evolut.* 45, 9–17.

Singh, M., Rana, J. C., Singh, B., Kumar, S., Saxena, D. R., Saxena, A., et al. (2017). Comparative agronomic performance and reaction to *Fusarium* wilt of Lens culinaris × L. orientalis and L. culinaris × L. ervoides derivatives. *Front. Plant Sci.* 8:1162. doi: 10.3389/fpls.2017.01162

Singh, V. K., Khan, A. W., Saxena, R. K., Kumar, V., Kale, S. M., and Chitikineni, A. (2016). Next-generation sequencing for identification of candidate genes for *Fusarium* wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). *Plant Biotechnol. J.* 14, 1183–1194. doi: 10.1111/pbi.12470

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* 2016, 1–11. doi: 10.1155/2016/3289801

Smith, S. N., Helms, D. M., Temple, S. R., and Frate, C. (1999). The distribution of *Fusarium* wilt of blackeyed cowpeas within California caused by *Fusarium oxysporum* f. sp. tracheiphilum race 4. *Plant Dis.* 83, 694–694. doi: 10.1094/pdis.1999.83.7.694c

Srinivas, C., Devi, D. N., Murthy, K. N., Mohan, C. D., Lakshmeesha, T. R., and Singh, B. P. (2019). *Fusarium oxysporum* f. sp. lycopersici causal agent of vascular wilt disease of tomato: biology to diversity– A review. *Saud. J Biol. Sci.* 26, 1315–1324. doi: 10.1016/j.sjbs.2019.06.002

Srivastava, A. K., Kashyap, P. L., Chakdar, H., Kumar, M., Srivastava, A. K., Yadav, J., et al. (2018). First de novo draft genome sequence of the pathogenic fungus *Fusarium* udum F02845, associated with pigeonpea (*Cajanus cajan* L. Millspaugh) wilt. *Microbiol. Resour. Announc.* 7, e1001–e1018.

Summerell, B. A., Leslie, J. F., Liew, E. C. Y., Laurence, M. H., and Bullock, S. (2011). *Fusarium* species associated with plants in Australia. *Fungal Diversity* 46, 1–27. doi: 10.1007/s13225-010-0075-8

Swarupa, V., Ravishankar, K. V., and Rekha, A. (2014). Plant defense response against *Fusarium oxysporum* and strategies to develop tolerant genotypes in banana. *Planta* 239, 735–751. doi: 10.1007/s00425-013-2024-8

Tekeoglu, M., Tullu, A., Kaiser, W. J., and Muehlbauer, F. J. (2000). Inheritance and linkage of two genes that confer resis- tance to *Fusarium* wilt in chickpea. *Crop Sci.* 40, 1247–1256. doi: 10.2135/cropsci2000.4051247x

Thatcher, L. F., Williams, A. H., Garg, G., Buck, S. G., and Singh, K. B. (2016). Transcriptome analysis of the fungal pathogen *Fusarium oxysporum* f. sp. medicaginis during colonisation of resistant and susceptible Medicago truncatula hosts identifies differential pathogenicity profiles and novel candidate effectors. *BMC Genomics* 17:860. doi: 10.1186/s12864-016-3192-2

Thudi, M., Khan, A. W., Kumar, V., Gaur, P. M., Katta, K., Garg, V., et al. (2016). Whole genome resequencing reveals genome-wide variations among parental lines of 16 mapping populations in chickpea (*Cicer arietinum*). *BMC. Plant Biol.* 16:10. doi: 10.1186/s12870-015-0690-3

Tosi, L., and Cappelli, C. (2001). First report of *Fusarium oxysporum* F. Sp. lentis of lentil in Italy. *Plant Dis.* 85:562. doi: 10.1094/pdis.2001.85.5.562a

Tullu, A., Kaiser, W. J., Kraft, J. M., and Muehlbauer, F. J. (1999). A second gene for resistance to race 4 of *Fusarium* wilt in chickpea and linkage with a RAPD marker. *Euphytica* 109, 43–50.

Tullu, A., Muehlbauer, F. J., Simon, C. J., Mayer, M. S., Kumar, J., Kaiser, W. J., et al. (1998). Inheritance and linkage of a gene for resistance to race 4 of *Fusarium* wilt and RAPD markers in chickpea. *Euphytica* 102, 227–232.

Upadhyaya, H., Haware, M., Kumar, J., and Smithson, J. (1983a). Resistance to wilt in chickpea. I. Inheritance of late-wilting in response to race 1. *Euphytica* 32, 447–456. doi: 10.1007/bf00021454

Upadhyaya, H., Smithson, J., Haware, M., and Kumar, J. (1983b). Resistance to wilt in chickpea. Further evidence for two genes for resistance to race 1. *Euphytica* 32, 749–755. doi: 10.1007/bf00042155

Upasani, M. L., Gurjar, G. S., Kadoo, N. Y., and Gupta, V. S. (2016). Dynamics of colonization and expression of pathogenicity related genes in *Fusarium oxysporum* f.sp. ciceri during chickpea vascular wilt disease progression. *PLoS One* 11:e0156490. doi: 10.1371/journal.pone.0156490

Upasani, M. L., Limaye, B. M., Gurjar, G. S., Kasibhatla, S. M., Joshi, R. R., Kadoo, N. Y., et al. (2017). Chickpea-*Fusarium oxysporum* interaction transcriptome reveals di erential modulation of plant defense strategies. *Sci. Rep.* 7:7746.

Varshney, R., Mohan, S., Gaur, P., Chamarthi, S., Singh, V., Srinivasan, S., et al. (2014). Marker-assisted backcrossing to introgress resistance to *Fusarium* wilt race 1 and Ascochyta blight in C 214, an elite cultivar of chickpea. *Plant Genome* 7, 278–289.

Varshney, R. K., Chen, W., Li, Y., Bharati, A. K., Saxena, R. K., and Schlueter, J. A. (2012). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89. doi: 10.1038/nbt.2022

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., and Sharpe, A. G. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246.

Varshney, R. K., Kudapa, H., Pazhamala, L., Chitikineni, A., Thudi, M., Bohra, A., et al. (2015). Translational genomics in agriculture: some examples in grain legumes. *Crit. Rev. Plant Sci.* 34, 169–194. doi: 10.1080/07352689.2014.897909

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350. doi: 10.1146/annurev-arplant-042811-105552

Wade, B. L. (1929). *Inheritance of Fusarium Wilt Resistance in Canning Peas*. Madison, WI: Agricultural Experiment Station of the University of Wisconsin, 1–32.

Wallace, J. G., Rodgers-Melnick, E., and Buckler, E. S. (2018). On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu. Rev. Genet.* 52, 421–444. doi: 10.1146/annurev-genet-120116-024846

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., and Gao, Y. (2016). Enhanced rice blast resistance by CRISPR/Cas9- targeted mutagenesis of the ERF transcription factor gene OsERF922. *PLoS One* 11:e0154027. doi: 10.1371/journal.pone.0154027

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., and Goa, C. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969

Widmann, C., Gibson, S., Jarpe, M. B., and Johnson, G. L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79, 143–180. doi: 10.1152/physrev.1999.79.1.143

Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., et al. (2016). Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genom.* 17:191. doi: 10.1186/s12864-016-2486-8

Winter, P., Benko-Iseppon, A. M., Huttel, B., Ratnaparkhe, M., Tullu, A., Sonnante, G., et al. (2000). A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatumcross*: localization of resistance genes for *Fusarium* wilt races 4 and 5. *Theor. Appl. Genet.* 101, 1155–1167. doi: 10.1007/s001220051592

Wise, R. P., Moscou, M. J., Bogdanove, A. J., and Whitham, S. A. (2007). Transcript profiling in host-pathogen interactions. *Annu. Rev. Phytopathol.* 45, 329–369. doi: 10.1146/annurev.phyto.45.011107.143944

Wu, X., Wu, X., Xu, P., Wang, B., Lu, Z., and Li, G. (2015). Association mapping for *Fusarium* wilt resistance in Chinese asparagus bean germplasm. *Plant Genome* 8, 1–6. doi: 10.3835/plantgenome2014.11.008

Xue, R., Wu, J., Zhu, Z., Wang, L., Wang, X., Wang, S., et al. (2015). Differentially expressed genes in resistant and susceptible common bean (*Phaseolus vulgaris* L.) genotypes in response to *Fusarium oxysporum* f. sp. phaseoli. *PLoS One* 10:e0127698. doi: 10.1371/journal.pone.0127698

Yang, Y., Shah, J., and Klessig, D. F. (1997). Signal perception and transduction in plant defense responses. *Genes Dev.* 11, 1621–1639. doi: 10.1101/gad.11.13.1621

Yin, K., and Qiu, J. L. (2019). Genome editing for plant disease resistance: applications and perspectives. *Philos. Trans. R. Soc. B.* 374:20180322. doi: 10.1098/rstb.2018.0322

Zhang, C., Chen, W., and Sankaran, S. (2019). High-throughput field phenotyping of Ascochyta blight disease severity in chickpea. *Crop Protect.* 125:104885. doi: 10.1016/j.cropro.2019.104885

Zhang, Y., Bai, Y., Wu, G., Zou, S., Chen, Y., and Gao, C. (2017). Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. *Plant J.* 91, 714–724. doi: 10.1111/tpj.13599

Zipfel, C., and Robatzek, S. (2010). Pathogen-associated molecular pattern triggered immunity: veni, vidi…? *Plant Physiol.* 154, 551–554. doi: 10.1104/pp.110.161547

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Jha, Bohra, Pandey and Parida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 November 2020
doi: 10.3389/fgene.2020.584859

Identification and Validation of New Stable QTLs for Grain Weight and Size by Multiple Mapping Models in Common Wheat

Jiajia Cao[1†], Yaoyao Shang[1†], Dongmei Xu[1], Kangle Xu[1], Xinran Cheng[1], Xu Pan[1], Xue Liu[1], Mingli Liu[1], Chang Gao[1], Shengnan Yan[1], Hui Yao[1], Wei Gao[1], Jie Lu[1], Haiping Zhang[1], Cheng Chang[1]*, Xianchun Xia[2], Shihe Xiao[2] and Chuanxi Ma[1]*

[1] Key Laboratory of Wheat Biology and Genetic Improvement on Southern Yellow and Huai River Valley, Ministry of Agriculture and Rural Affairs, College of Agronomy, Anhui Agricultural University, Hefei, China, [2] Institute of Crop Sciences, National Wheat Improvement Center, Chinese Academy of Agricultural Sciences, Beijing, China

Improvement of grain weight and size is an important objective for high-yield wheat breeding. In this study, 174 recombinant inbred lines (RILs) derived from the cross between Jing 411 and Hongmangchun 21 were used to construct a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq). Three mapping methods, including inclusive composite interval mapping (ICIM), genome-wide composite interval mapping (GCIM), and a mixed linear model performed with forward–backward stepwise (NWIM), were used to identify QTLs for thousand grain weight (TGW), grain width (GW), and grain length (GL). In total, we identified 30, 15, and 18 putative QTLs for TGW, GW, and GL that explain 1.1–33.9%, 3.1%–34.2%, and 1.7%–22.8% of the phenotypic variances, respectively. Among these, 19 (63.3%) QTLs for TGW, 10 (66.7%) for GW, and 7 (38.9%) for GL were consistent with those identified by genome-wide association analysis in 192 wheat varieties. Five new stable QTLs, including 3 for TGW (*Qtgw.ahau-1B.1*, *Qtgw.ahau-4B.1*, and *Qtgw.ahau-4B.2*) and 2 for GL (*Qgl.ahau-2A.1* and *Qgl.ahau-7A.2*), were detected by the three aforementioned mapping methods across environments. Subsequently, five cleaved amplified polymorphic sequence (CAPS) markers corresponding to these QTLs were developed and validated in 180 Chinese mini-core wheat accessions. In addition, 19 potential candidate genes for *Qtgw.ahau-4B.2* in a 0.31-Mb physical interval were further annotated, of which *TraesCS4B02G376400* and *TraesCS4B02G376800* encode a plasma membrane H^+-ATPase and a serine/threonine-protein kinase, respectively. These new QTLs and CAPS markers will be useful for further marker-assisted selection and map-based cloning of target genes.

Keywords: common wheat, grain weight/size, QTL, SLAF-seq, GWAS, SNP

OPEN ACCESS

Edited by:
Deepmala Sehgal, International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Liu Jin Dong, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China
Himanshu Sharma, National Agri-Food Biotechnology Institute, India

***Correspondence:**
Haiping Zhang
zhhp20@163.com
Cheng Chang
changtgw@126.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *18 July 2020*
Accepted: *21 September 2020*
Published: *11 November 2020*

Citation:
Cao J, Shang Y, Xu D, Xu K, Cheng X, Pan X, Liu X, Liu M, Gao C, Yan S, Yao H, Gao W, Lu J, Zhang H, Chang C, Xia X, Xiao S and Ma C (2020) Identification and Validation of New Stable QTLs for Grain Weight and Size by Multiple Mapping Models in Common Wheat. Front. Genet. 11:584859. doi: 10.3389/fgene.2020.584859

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the most important cereal crops and provides approximately 20% of the dietary calories for humans worldwide. The demand for wheat is continually increasing due to rapid population growth (Organization of Food and Agriculture, 2009). Grain weight and size, including thousand grain weight (TGW), grain length (GL), and grain width (GW), are positively associated with wheat grain yield (Börner et al., 2002; Dholakia et al., 2003; Breseghello and Sorrells, 2007). Therefore, dissecting the genetic basis of these grain-related traits contributes to the improvement of wheat yield.

Numerous quantitative trait loci (QTLs) underlying grain weight and size were detected on almost all wheat chromosomes by linkage mapping and genome-wide association studies (GWAS) (Cui et al., 2014; Simmonds et al., 2014; Wu et al., 2015; Thorwarth et al., 2019). Several candidate genes for TGW, GW, and GL have been isolated in wheat. *TaCKX6-D1* encodes a cytokinin oxidase/dehydrogenase (CKX2) associated with grain number and grain weight (Ashikari et al., 2005; Li et al., 2011, 2013). *TaGW2-6A* encodes a RING-type E3 ubiquitin ligase and functions as a negative regulator of GW (Su et al., 2012; Yang et al., 2012; Vandana et al., 2015). *TaCwi-A1* encodes a cell wall invertase for carbon partitioning during early grain filling (Wang et al., 2008; Ma et al., 2012). *TaSAP-A1* belongs to a gene family of stress-associated proteins significantly related to TGW, number of grains per spike, and spike length (Chang et al., 2013). *TaGS1a* encodes a cytosolic glutamine synthetase underlying grain size (Ying et al., 2013). *TaGS-D1* is associated with TGW and GL (Zhang et al., 2014). *TaSus1* and *TaSus2* encode two isoforms of sucrose synthase (Jiang et al., 2011; Hou et al., 2014). *TaGS5-3A* regulates grain size (Ma L. et al., 2015). *6-SFT-A2* is involved in fructan biosynthesis and significantly associated with TGW under rainfed conditions (Yue et al., 2015). *TaSnRK2.3*, *TaSnRK2.9,* and *TaSnRK2.10*, three members of the SnRK2 family, are associated with TGW (Miao et al., 2017; Zhang Z. et al., 2017; Rehman et al., 2019). More recently, Cheng et al. (2020) isolated the *Tasg-D1* gene, which encodes serine/threonine protein kinase glycogen synthase kinase 3 (STKc_GSK3), leading to formation of round grains in Indian dwarf wheat (*Triticum sphaerococcum* Perc.). *TaSDIR1-4A*, which encodes a RING-type E3 ubiquitin ligase is associated with TGW in well-watered and heat-stress environments (Wang J. et al., 2020). The physical interaction of the proteins encoded by *TaDA1-A* and *TaGW2-6B* was significantly associated with grain size and weight (Liu H. et al., 2019; Liu J. et al., 2019). *TaGS3-7A*, a homologous gene of *OsGS3*, was shown to be associated with grain weight (Yang et al., 2019b). In addition, *TaTGW6* (Hu et al., 2016a), *Tabas1* (Zhu et al., 2016), *TaCKX4* (Chang et al., 2015), *TaFlo2* (Sajjad et al., 2017), *TaTGW-7A* (Hu et al., 2016b), *TaCYP78A3* (Ma M. et al., 2015), and *TaGL3* (Yang et al., 2019a) are also significantly associated with grain weight or size. All of the genes listed above are useful in genetic improvement of wheat yield. However, wheat grain weight and size are complex traits controlled by multiple genes. Identification and validation of more QTLs for grain-related traits will not only promote our understanding of the genetic basis of the two traits, but also accelerate the process of pyramiding favorable alleles in high-yield wheat breeding.

In the past decades, QTL mapping and GWAS were performed based on randomly amplified polymorphic DNA (RAPDs) (Qu and Hancock, 1997), restriction fragment length polymorphisms (RFLPs) (Keim et al., 1990), amplified fragment length polymorphisms (AFLPs) (Vos et al., 1995; Chagne et al., 2002), simple sequence repeats (SSRs) (Torada et al., 2005), and sequence-tagged site (STS) markers (Talbert et al., 1994). However, the number of these markers is limited, and this affects the accuracy and efficiency of QTL mapping and GWAS.

In recent years, several approaches for high-throughput molecular markers supported by next-generation sequencing technologies and SNP chips have developed rapidly, including DNA sequencing (RAD-seq) (Baird et al., 2008), genotyping-by-sequencing (GBS-seq) (Elshire et al., 2011), restriction site-associated specific locus amplified fragment sequencing (SLAF-seq) (Sun et al., 2013), and wheat 90K (Wang et al., 2014), 55K, and 660K SNP arrays (designed by the Chinese Academy of Agricultural Sciences) (Cui et al., 2017; Huang et al., 2019). These high-throughput genotyping methods obviously accelerate the identification of novel loci and candidate genes underlying wheat yield–related traits. Cui et al. (2017) constructed a genetic map with 4959 bin markers using a wheat 660K SNP array in a recombinant inbred line (RIL) population derived from Kenong 9204 × Jing 411 and detected a major QTL for kernel number per spike on chromosome 4A. Fiedler et al. (2017) identified a QTL for test weight on chromosome 1B by GBS-seq of 1184 lines collected from the North Dakota State University durum wheat breeding program. Wang F. et al. (2020) constructed a high-density genetic map (HDGM), including 3556 SLAF-based SNP and SSR markers, and detected 37 QTLs related to yield traits in an RIL population derived from cotton cultivars LMY 22 (high-yield) and LY 343 (superior fiber quality). Notably, SLAF-seq has been widely applied to construct a HDGM for QTL mapping as well as narrowing target regions in several crops, such as maize (Liu et al., 2016; Li et al., 2018), rice (Yang et al., 2017; Quan et al., 2018), soybeans (Li et al., 2017; Han et al., 2019), and peppers (Guo et al., 2017; Zhang et al., 2019). However, few HDGMs have been constructed by SLAF-seq for common wheat.

The aims of this study were to (1) identify a large number of SNP markers to construct a HDGM by SLAF-seq in an RIL population derived from a cross between high-TGW cultivar Jing 411 and low-TGW landrace Hongmangchun 21, (2) identify QTLs for grain weight and size using SNP markers, and (3) design cleaved amplified polymorphism sequences (CAPS) markers closely linked to new stable QTLs for grain weight and size and validate them in 192 wheat varieties and 180 Chinese mini-core wheat accessions.

MATERIALS AND METHODS

Plant Materials

A RIL population (174 lines) derived from a cross between Jing 411 (average TGW: 41.94 g) and Hongmangchun 21 (average TGW: 19.98 g) (designated JH-RILs) was used for QTL

mapping; it was planted during the 2005–2006, 2006–2007, 2007–2008, 2009–2010, 2010–2011, 2011–2012, 2013–2014, 2014–2015, 2015–2016, 2016–2017, and 2017–2018 cropping seasons. One hundred ninety-two wheat varieties (WVs) grown in the 2015–2016, 2016–2017, and 2017–2018 cropping seasons, containing 159 cultivars, 11 advanced breeding lines, and 22 landraces, were chosen for a GWAS panel to validate QTLs identified by linkage analysis. The origin of these varieties has been described in detail by Zhu et al. (2019). One hundred eighty Chinese mini-core collection accessions (CMCCs) planted in the 2014–2015, 2015–2016, and 2016–2017 cropping seasons comprising 74 cultivars, 11 advanced breeding lines, and 95 landraces were used to validate new stable loci for grain weight and size.

The above three populations were planted at the experimental station of Anhui Agricultural University in Hefei (31°58′N, 117°240′E), Anhui Province, China. Field trials were conducted in 2-m-long plots, each consisting of double rows spaced 0.25 m apart in randomized complete blocks with two replications.

TGW, GW, and GL Tests

Sixty spikes per plot were harvested and threshed at physiological maturity, which is characterized by yellow color on the whole plant including leaves, stems, and spikes (Kulwal et al., 2005). Three hundred threshed seeds were used to measure TGW, GL, and GW in triplicate using the SC-G wheat grain appearance quality image analysis system (Hangzhou WSeen Detection Technology Co., Ltd, Hangzhou, China) (Yin et al., 2015).

For JH-RILs, TGW was measured during the 2005–2006, 2006–2007, 2007–2008, 2009–2010, 2010–2011, 2011–2012, 2013–2014, 2014–2015, 2015–2016, 2016–2017, and 2017–2018 cropping seasons, and these measurements were designated 2006TGW-JH, 2007TGW-JH, 2008TGW-JH, 2010TGW-JH, 2011TGW-JH, 2012TGW-JH, 2014TGW-JH, 2015TGW-JH, 2016TGW-JH, 2017TGW-JH, and 2018TGW-JH, respectively. GW and GL were measured in part seasons, and these measurements were designated 2014GW-JH, 2015GW-JH, 2016GW-JH, 2017GW-JH, 2018GW-JH, 2014GL-JH, 2015GL-JH, 2016GL-JH, 2017GL-JH, and 2018GL-JH, respectively.

For 192 WVs, TGW was evaluated during the 2015–2016, 2016–2017, and 2017–2018 cropping seasons, designated 2016TGW-192, 2017TGW-192, and 2018TGW-192, respectively; GW and GL were also measured, designated 2016GW-192, 2017GW-192, 2018GW-192, 2016GL-192, 2017GL-192, and 2018GL-192, respectively.

For 180 CMCCs, TGW was evaluated during the 2014–2015, 2015–2016, and 2016–2017 cropping seasons, designated 2015TGW-CMCC, 2016TGW-CMCC, and 2017TGW-CMCC, respectively; GW and GL were measured and designated 2015GW-CMCC, 2016GW-CMCC, 2017GW-CMCC, 2015GL-CMCC, 2016GL-CMCC, and 2017GL-CMCC, respectively.

Statistical Analysis

Pearson's correlation analysis and Mann–Whitney *U*-tests were performed by the software SPSS 20.0 (IBM Corporation, Armonk, NY, United States).

Genomic DNA Isolation

Dry seeds were collected to extract genomic DNA using the SDS method (Kang et al., 1998). DNA quality was tested by ND5000 spectrophotometer (NanoDrop, Wilmington, DE, United States) and in 1% agarose gels.

Construction of SLAF Library

The JH-RILs were genotyped by a modified SLAF-seq strategy to develop genome-wide SNP markers (Sun et al., 2013). The Chinese Spring reference genome IWGSC RefSeq v1.0[1] was utilized to simulate *in silico* the number of markers digested by different restriction enzymes and design a pilot SLAF experiment (Sun et al., 2013). A restriction enzyme (*Rsa*I, New England Biolabs, United States) was used to digest the genomic DNA. The fragments ranging from 464 to 484 bp were selected and purified after agarose gel electrophoresis for 100-bp paired-end sequencing. To assess the accuracy of the experiment for SLAF library construction, *Oryza sativa* L. *japonica* was chosen as a control in comparison to mapping populations with the same experimental treatment and sequencing.

SLAF-Seq Data Analysis and Genotyping

In order to ensure the accuracy of data analysis, quality control for the raw sequencing data were performed according to the following criteria: (i) reads containing adaptor sequences were filtered out, (ii) reads with unknown bases exceeding 10% in length were filtered out, (iii) reads unaligned to the wheat reference genome IWGSC RefSeq v1.0 (see footnote 1) were filtered out, and (iv) reads with a single end aligned to the wheat reference genome IWGSC RefSeq v1.0 were also filtered out because of the inaccurate physical position.

SLAF paired-end mapped reads were clustered based on sequence similarity identified by alignments to the wheat reference genome IWGSC RefSeq v1.0 with BWA software (Chong et al., 2003). SNPs in all SLAF loci were identified between parents by the Genome Analysis Toolkit (McKenna et al., 2010)[2]. SLAFs with 8 or more SNPs, which are considered as high-frequency variable regions of wheat, were filtered out, and this influenced the accuracy of following steps. Minor allele frequency evaluation was used to define alleles in each SLAF. SLAF markers were generated following Sun et al. (2013). Polymorphic markers were selected by genotyping the parents and then classified into eight segregation types (ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc, and cc × ab). For the RIL population, SLAF markers in the segregation pattern of aa × bb were selected. The quality of SLAF markers used to build genetic maps was controlled by the following criteria: (i) average sequence depths must be >6-fold in parents, (ii) markers with more than 25% missing data were filtered out, (iii) markers with significant segregation distortion ($P < 0.001$) were rejected, and (iv) markers were required to have a logarithm of odds (MLOD) exceeding 3.

[1]https://urgi.versailles.inra.fr/blast_iwgsc/blast.php

[2]https://gatk.broadinstitute.org

Bulked Segregant Analysis of JH-RILs by 660K SNP Arrays

Two high-TGW bulks (HB1 and HB2) and two low-TGW bulks (LB1 and LB2) were made by the same amount of DNA from extremely high and low TGW lines (5 lines per bulk) of JH-RILs based on phenotypic values. The above four bulks and two parents were genotyped using the wheat 660K SNP array by the CapitalBio Corporation (Beijing, China). SNP genotyping and clustering were performed by Genome Studio Polyploid Clustering v1.0[3]. Consistent SNPs without missing data in HB1, HB2, and Jing 411 were classified as group I, and consistent SNPs without missing data in LB1, LB2, and Hongmangchun 21 were classified as group II. Different SNPs between groups I and II were selected for subsequent analysis and development of PCR-based molecular markers. Physical positions of these SNPs were searched by sequence BLAST with IWGSC RefSeq v1.0 (see footnote 1).

Development and Genotyping of PCR-Based Molecular Markers

Seventy-four SNPs with polymorphisms between groups I and II were converted to CAPS markers by Primer Premier 5.0[4] and used to genotype JH-RILs to construct genetic maps (**Supplementary Table 1**). Representative CAPS markers closely linked with new stable QTLs were used to genotype 180 CMCCs. A PCR mixture with a total volume of 10 μL comprised 1.0 μL of 10 × PCR buffer, 200 μM dNTPs, 4 pmol of each primer, 0.5 U *Taq* DNA polymerase, and 50–100 ng template DNA. The reaction was performed in a C1000 Thermal Cycler (Bio-Rad, United States) with the following program: denaturation at 94°C for 5 min followed by 40 cycles at 94°C for 30 s, touchdown starting at 62°C for 30 s (decreasing 0.3°C per cycle), and 72°C for 30 s that were followed by a final extension at 72°C for 8 min. The PCR products were digested with the corresponding restriction enzymes (**Supplementary Table 1**[5]) for 5 h and separated on 2.5% agarose gels.

Eleven functional markers specific to *TaSdr-2A* (Zhang Y. et al., 2017), *TaSus-2A* (Hou et al., 2014), *Tamyb10-3A* (Himi et al., 2011), *TaVp1B3-3B* (Yang Y. et al., 2007), *TaVp1-b2* (3B) (Chang et al., 2010), *Tamyb10-3D* (Himi et al., 2011), *TaTGW6-4A* (Hu et al., 2016a), *TaMKK3-4A* (Torada et al., 2016), *TaGW2-6A* (Su et al., 2012), *TaGASR7-A1* (7A) (Dong et al., 2014), and *TaPTF1-7B* (Zhu, 2016) were used for QTL analysis in JH-RILs in combination with SLAF and CAPS markers. PCR conditions and gel electrophoresis for functional markers were performed following the above reported studies (**Supplementary Table 2**).

Genetic Linkage Map Construction and QTL Mapping

SLAF markers distributed on 21 wheat chromosomes were selected by mapping using IWGSC RefSeq v1.0 (see footnote 1). To ensure the accuracy of genetic linkage maps, the MLOD scores between markers were estimated, and markers with MLOD scores less than 5 were removed. Then, SLAF, CAPS, and gene-specific markers were collectively used to construct a high-density genetic bin map (designated SLAF-map) without redundant markers by IciMapping v4.1 using the Kosambi mapping function in JH-RILs.

The inclusive composite interval mapping (ICIM) program of QTL IciMapping v4.1 (Li et al., 2007)[6] was used to detect QTLs for TGW, GL, and GW in JH-RILs with a walk speed of 1 cM and a window size of 10 cM. The genome-wide composite interval mapping (GCIM) of QTL.gCIM.GUI v1.0 (Wen et al., 2019)[7] in R version 3.6.0 software was applied to identify QTLs with a random model and a walk speed of 1 cM. For these two methods, an LOD score of 2.5 was used for claiming the presence of QTLs. The QTLnetwork v2.0 software (Yang J. et al., 2007; Yang et al., 2008), which is based on a mixed linear model (MLM), was used for forward–backward stepwise (NWIM) analysis with a threshold of $P = 0.05$ to select cofactors, multiple linear regression with a 1 cM walk speed, and a window size set at 10 cM.

Adjacent QTLs with the same sign of additive effects satisfying at least one of the following criteria were defined as the same QTL: (1) positions of QTL peaks within 10 cM (Kumar et al., 2016) and (2) QTLs with overlapped confidence intervals (Cui et al., 2017). QTLs for TGW detected in more than five environments and those for GW or GL in more than three environments were defined as stable loci (Cui et al., 2014).

Genotyping and Association Analysis of 192 WVs Using 90K SNP Array

The DNA samples of 192 WVs were genotyped by the Illumina iSelect 90K Infinium SNP array, including 81,587 SNPs (Wang et al., 2014), by the Beijing Compass Biotechnology Co., Ltd. Genotypic clusters for each SNP were determined by the Genome Studio version 2011.1 software (Illumina, see footnote 3). The physical positions of SNPs were obtained from IWGSC RefSeq v1.0 (see footnote 1). The SNPs with less than 20% missing data and a minor allele frequency exceeding 5% were used for association analysis.

Linkage disequilibrium and population structure were analyzed for 192 WVs following methods in our previous study (Zhu et al., 2019). The same K and Q matrix data were used in the present study to identify significant marker–trait associations (MTAs) for grain weight and size by the MLM. The MLM was performed by TASSEL 5.0 to detect MTAs at a significance level of $P < 0.001$ (Bradbury et al., 2007; Chen et al., 2016).

Gene Annotation

Databases from the IWGSC gene annotation[8] were used to perform gene function annotations by BLAST.

Cloning of *TraesCS4B02G376400*

Primer pair PMA-1 was designed to amplify the partial sequence of *TraesCS4B02G376400* based on the putative sequence from

[3]http://www.illumina.com

[4]https://www.PremierBiosoft.com

[5]https://www.neb-china.com

[6]http://www.isbreeding.net

[7]https://cran.r-project.org/web/packages/QTL.gCIMapping.GUI/index.html

[8]https://wheat-urgi.versailles.inra.fr/Tools/JBrowse

IWGSC RefSeq v1.0 (see footnote 1) (**Supplementary Table 3**). PCR amplifications were performed in 10-μL volumes that included 0.25 μM of each primer, 0.25 mM dNTPs, 100 ng genomic DNA, 0.5 unit *Fastpfu* polymerase, and 1 μL of 10 × *Fastpfu* PCR buffer (Beijing TransGen Biotech, Beijing, China). The amplification program consisted of an initial denaturation at 94°C for 5 min followed by 36 cycles of denaturation at 94°C for 45 s, annealing at 52°C for 50 s, and extension at 72°C for 1 min that were followed by a final extension at 72°C for 12 min. Amplified PCR fragments were separated on 1.5% agarose gels. Target fragments were recovered and cloned into a Blunt-zero vector and transformed into T1 competent cells (Beijing TransGen Biotech, Beijing, China).

Sequencing was carried out on an ABI 3500 genetic analyzer (Applied Biosystems, Shanghai, China). Sequence alignments were performed using DNAMAN 6.0[9].

Development of CAPS Marker for *TraesCS4B02G376400*

The SNP distinguished by restriction enzyme *Sty*I found in *TraesCS4B02G376400* between Jing 411 and Hongmangchun 21 was converted to a CAPS marker PMA-2 by Primer Premier 5.0 (see footnote 4) and used to genotype 180 CMCCs (**Supplementary Table 3**). The PCR conditions for the marker PMA-2 were consistent with those for 74 CAPS markers mentioned above.

RESULTS

Statistical Analysis of Phenotypic Data

Phenotypic data for TGW, GW, and GL exhibited extensive variations in JH-RILs and 192 WVs across environments. In JH-RILs, the ranges of TGW, GW, and GL were 13.71–68.28 g (mean 35.97 g), 2.25–4.38 mm (mean 3.01 mm), and 3.99–8.55 mm (mean 6.36 mm), respectively. In the 192 WVs, TGW, GW, and GL ranged from 19.58 to 54.21 g (mean 38.29 g), 1.21 to 4.68 mm (mean 3.01 mm), and 4.37 to 7.50 mm (mean 6.23 mm), respectively (**Supplementary Table 4**).

Highly significant correlations were observed among TGW, GW, and GL values in different environments. TGW showed positive correlations with GW and GL with correlation coefficients ranging from 0.39–0.97 (average 0.67) and 0.43–0.94 (average 0.76) in JH-RILs and from 0.28–0.84 (average 0.54) and 0.19–0.72 (average 0.47) in the 192 WVs (**Supplementary Tables 5**, **6**).

Construction of Genetic Linkage Maps

After SLAF library construction and high-throughput sequencing, 576.7 M clean reads were obtained with each read being 200 bp. The number of SLAF tags was 816,183 for Jing 411; 798,085 for Hongmangchun 21; and 462,529 for progenies. The average depths of the SLAF markers were 21.52 in Jing 411, 19.64 in Hongmangchun 21, and 4.16 in progenies. Based on the filtering criteria described above, 1529 SLAF bin markers containing 4820 high-quality polymorphic SLAF markers in combination with 74 CAPS and 11 functional markers were used for QTL mapping (**Supplementary Figure 1**). Finally, a high-density genetic bin map, including 1529 SLAF bin markers, 74 CAPS markers, and 11 gene-specific functional markers spanned 2014.43 cM in length with an average marker interval of 1.22 cM. The genetic distances of the A, B, and D genomes were 864.2 cM (608 markers), 900.0 cM (764 markers), and 250.2 cM (242 markers), respectively (**Supplementary Figure 1** and **Supplementary Table 7**).

QTLs for TGW, GW, and GL Identified by Linkage Analysis

A total of 30 QTLs for TGW were identified using three mapping methods (ICIM, GCIM, and NWIM) in JH-RILs and were distributed on chromosomes 1B (3), 2A (3), 2B (2), 2D, 3A (3), 3B, 4A (2), 4B (2), 4D, 5A (3), 6A, 6B (2), 7A (2), 7B (2), and 7D (2). These QTLs explained 1.1–33.9% of the phenotypic variance, especially five QTLs (*Qtgw.ahau-2A.3*, *Qtgw.ahau-5A.1*, *Qtgw.ahau-7A.2*, *Qtgw.ahau-7B.1*, and *Qtgw.ahau-7B.2*) that had relatively high phenotypic variance explained (PVE) ranging from 10.3 to 16.5%. In addition, 13 QTLs for TGW were detected by two methods across environments and were distributed on chromosomes 2A (2), 3A, 3B, 4A, 4D, 5A (2), 7A (2), 7B (2), and 7D. Five QTLs, including *Qtgw.ahau-1B.1*, *Qtgw.ahau-2D*, *Qtgw.ahau-4B.1*, *Qtgw.ahau-4B.2*, and *Qtgw.ahau-6B.1*, were collectively detected by all three methods across environments. Particularly, *Qtgw.ahau-4B.2* was identified in 10 environments, explaining 11.7% of the phenotypic variances on average and, thus, was considered a stable and major QTL (**Figure 1**, **Table 1**, and **Supplementary Table 8**).

Fifteen QTLs associated with GW were mapped on chromosomes 1B, 2D (3), 3B, 4A, 4B (3), 4D, 5A, 5B, 7A (2), and 7B using the aforementioned three mapping methods. Among them, 6 QTLs for GW on chromosomes 1B, 2D, 3B, 4A, 4B, and 4D were detected by two methods across environments. Notably, two QTLs (*Qgw.ahau-2D.3* and *Qgw.ahau-4B.3*), which explained 18.5 and 20.0% of the phenotypic variances on average, respectively, were jointly detected by all three methods (**Table 1** and **Supplementary Table 8**).

Eighteen QTLs for GL were identified on chromosomes 1B (3), 2A (3), 2B, 2D, 3A, 4B (4), 6B, 7A (3), and 7B with three mapping methods. Of these, two QTLs, *Qgl.ahau-1B.1* and *Qgl.ahau-4B.3*, were identified by two methods in more than one environment. Three QTLs, including *Qgl.ahau-2A.1*, *Qgl.ahau-6B*, and *Qgl.ahau-7A.2*, were consistently detected by all three methods in several environments. Of these, *Qgl.ahau-2A.1* and *Qgl.ahau-7A.2* explained 12.1% and 11.9% of the phenotypic variances, respectively, and were considered as stable and major QTLs (**Figure 1**, **Table 1**, and **Supplementary Table 8**).

Validation of QTLs for TGW, GW, and GL by GWAS

Association analysis was performed to validate QTLs for TGW, GW, and GL detected in JH-RILs using 192 WVs in three environments based on an MLM (**Supplementary Figure 2**).

[9]http://www.lynnon.com/

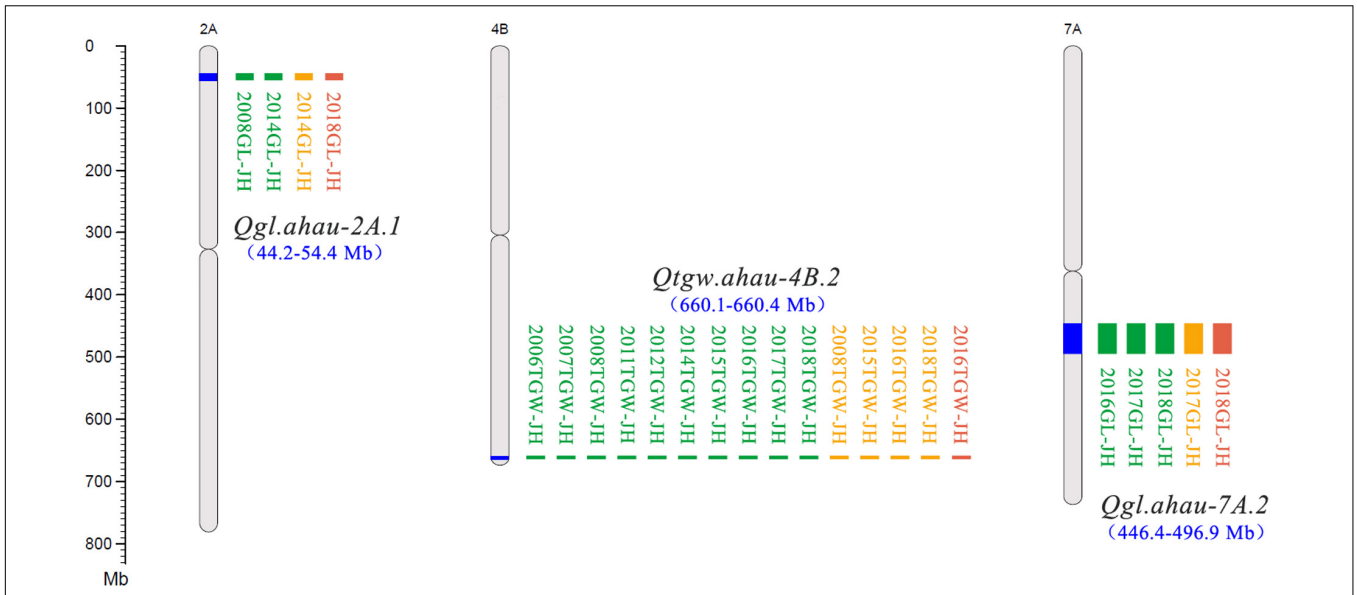


FIGURE 1 | New stable QTLs for TGW and GL in the Jing 411/Hongmangchun 21 population. Uniform million bp (Mb) scale is shown on the left. Green, orange, and red blocks indicate QTLs detected by ICIM performed by QTL IciMapping v4.1, GCIM using QTL.gCIM.GUI v1.0, and composite interval mapping based on mixed linear model using forward–backward stepwise (NWIM) in QTLNetwork v2.0, respectively.

The results indicated that 19 (63.3%) of 30 QTLs for TGW, 10 (66.7%) of 15 for GW, and 7 (38.9%) of 18 for GL detected in JH-RILs were consistent with those identified by GWAS in 192 varieties. Among them, two reported QTLs for GW (*Qgw.ahau-2D.1*, 272.5–420.5 Mb; *Qgw.ahau-4D*, 226.4–409.4 Mb) identified in JH-RILs were closely linked with two stable MTAs (*D_GB5Y7FA01EHPZX_186*, 353.1 Mb on chromosome 2D; *IACX65*, 352.4 Mb on chromosome 4D) as detected by GWAS in 192 WVs across all environments. Four reported intervals for TGW harboring *Qtgw.ahau-5A.3* (559.3–596.6 Mb), *Qtgw.ahau-6B.2* (679.6–698.3 Mb), *Qtgw.ahau-7A.1* (99.7–200.5 Mb), *Qtgw.ahau-7D.1* with positive addictive effect (256.1–383.1 Mb)/*Qtgw.ahau-7D.2* with negative addictive effect (270.6–391.9 Mb) identified in JH-RILs were near to four MTAs (*Excalibur_c56952_411*, 569.8 Mb on chromosome 5A; *RAC875_c17347_312*, 694.2 Mb on chromosome 6B; *RAC875_c5986_3670*, 200.1 Mb on chromosome 7A; *RAC875_c45846_454*, 324.5 Mb on chromosome 7D) detected in 192 WVs, explained 12.6, 11.0, 10.7, and 12.5% of the phenotypic variances on average, respectively. Notably, three novel and major QTLs, including *Qtgw.ahau-4B.2* (660.1–660.4 Mb), *Qgl.ahau-2A.1* (44.2–54.4 Mb), and *Qgl.ahau-7A.2* (446.4–496.9 Mb), were adjacent to three MTAs (*RAC875_c48025_483*, 666.1 Mb on chromosome 4B; *Tdurum_contig55610_784*, 42.5 Mb on chromosome 2A; *Kukri_c5101_2636*, 498.9 Mb on chromosome 7A) based on their physical positions in 192 wheat varieties, respectively (**Table 1** and **Supplementary Table 9**).

Validation of New Stable QTLs in 180 CMCCs

Four CAPS markers (1B-JHMfeI for *Qtgw.ahau-1B.1*, 2A-CAPSmin for *Qgl.ahau-2A.1*, 4B-8621 for *Qtgw.ahau-4B.1*, and 7A-3738 for *Qgl.ahau-7A.2*) were further developed to validate the associations of four new stable QTLs with grain weight and size in 180 CMCCs (**Table 1** and **Supplementary Tables 1**, **8**). Allelic variations of the CAPS markers were detected, designated as *1B-JHMfeI-C*/*1B-JHMfeI-T*, *2A-CAPSmin-T*/*2A-CAPSmin-C*, *4B-8621-C*/*4B-8621-G*, and *7A-3738-G*/*7A-3738-C*, respectively (**Supplementary Figure 3**). Compared with the Jing 411 genotypes (*1B-JHMfeI-C*, *2A-CAPSmin-T*, *4B-8621-C*, and *7A-3738-G*), the Hongmangchun 21 genotypes (*1B-JHMfeI-T*, *2A-CAPSmin-C*, *4B-8621-G*, and *7A-3738-C*) were significantly associated with lower TGW, GW, and GL values across environments ($P < 0.01$ or $P < 0.05$) (**Table 2**).

Candidate Gene Prediction for *Qtgw.ahau-4B.2*

The new stable and major QTL *Qtgw.ahau-4B.2* was flanked by two CAPS markers, 4B-2133 and 4B-8416, in a physical interval of 0.31 Mb (**Table 1**). According to IWGSC annotation databases, the target region of *Qtgw.ahau-4B.2* contains 19 genes. Among them, *TraesCS4B02G376400* and *TraesCS4B02G376800*, which encode a plasma membrane H^+-ATPase and a serine/threonine-protein kinase, respectively, were considered as potential candidate genes (**Supplementary Table 10**).

Cloning of *TraesCS4B02G376400* and Validation in 180 CMCCs

Sequence analysis indicated that *TraesCS4B02G376400* comprised 20 exons and 19 introns, with a complete sequence of 6024 bp (**Figure 2A**). After partial sequencing of *TraesCS4B02G376400* (1919–5084 bp), a SNP (A-G) was identified in the 12th intron between Jing 411 and

TABLE 1 | QTLs for grain weight and size identified by all three mapping methods in the Jing 411/Hongmangchun 21 RIL population.

Trait	QTL	Physical Interval (Mb)[b]	Representative Marker	No. of Environments	Method[c]	Peak Position (cM)	LOD or *P*-value[d]	PVE[e] (%)	Additive Effect	References
TGW	*Qtgw.ahau-1B.1*	648.1–652.1	1B-JHMfel	5	ICIM	184.0–185.0	3.27–4.38	5.4–8.0	1.46–2.23	**Novel** for TGW;
					GCIM	184.8–185.4	2.74–5.42	3.5–6.3	1.85–2.17	Cui et al., 2014
					NWIM	187.2	0.00	19.8	2.37	(**KNS**)
	Qtgw.ahau-2D	543.5–561.9	2D-5137	2	ICIM	60.0–62.0	3.27–6.45	8.2–8.5	2.03	Liu et al., 2018
					GCIM	61.5	3.52	5.0	1.94	
					NWIM	61.3	0.00	30.4	3.34	
	Qtgw.ahau-4B.1[a]	540.3–542.9	4B-8621	5	ICIM	34.0–44.0	5.05–14.74	9.6–29.3	1.97–4.77	**Novel**
					GCIM	43.5	2.57–3.70	1.9–3.5	1.93–2.09	
					NWIM	35.0	0.00	17.8	4.13	
	Qtgw.ahau-4B.2[a]	660.1–660.4	4B-9051	10	ICIM	115.0–116.0	3.37–10.25	5.2–23.2	1.47–3.26	**Novel**
					GCIM	115.2–116.2	2.80–8.63	2.6–15.0	1.63–3.10	
					NWIM	115.2	0.00	33.9	5.72	
	Qtgw.ahau-6B.1	677.7–679.6	6B-9077	3	ICIM	73.0	4.22	8.2	1.80	Li et al., 2019
					GCIM	73.3	2.86	3.7	1.96	
					NWIM	73.4	0.00	18.1	3.82	
GW	*Qgw.ahau-2D.3*[a]	543.5–561.9	2D-5137	1	ICIM	62.0	8.36	14.0	0.09	Liu et al., 2018
					GCIM	60.5	4.98	14.7	0.10	
					NWIM	60.6	0.00	26.7	0.12	
	Qgw.ahau-4B.3[a]	660.11–660.13	4B-9051	2	ICIM	116.0	4.21–14.27	11.0–20.7	0.06–0.10	**Novel**
					GCIM	116.2	3.44–7.31	8.1–26.0	0.09–0.17	
					NWIM	115.2	0.00	34.2	0.17	
GL	*Qgl.ahau-2A.1*[a]	44.2–54.4	2A-CAPSmin	3	ICIM	105.0–107.0	6.54–8.96	13.8–15.5	0.21–0.25	**Novel**
					GCIM	103.8	6.42	10.4	0.22	
					NWIM	211.0	0.01	8.5	0.25	
	Qgl.ahau-4B.1	31.3–38.4	Marker13172883	2	ICIM	6.0	5.16–16.88	6.5–18.0	0.14–0.22	Garcia et al., 2019
					GCIM	6.3–7.2	4.75–3.68	2.1–5.3	0.12–0.20	
					NWIM	6.4	0.00	11.9	0.19	
	Qgl.ahau-6B	676.8–678.2	6B-6004	4	ICIM	80.0–86.0	3.11–5.79	4.8–5.8	0.12–0.14	Li et al., 2019
					GCIM	88.0	3.17	1.7	0.11	
					NWIM	91.8	0.01	7.7	0.22	
	Qgl.ahau-7A.2[a]	446.4–496.9	7A-3738	3	ICIM	123.0	3.75–15.88	9.7–22.8	0.17–0.38	**Novel** for GL;
					GCIM	122.8	5.59	6.3	0.22	Wang et al., 2017
					NWIM	63.2	0.02	9.5	0.20	(**KNS**)

[a] *The same loci detected by linkage analysis in the JH RIL population and association analysis in 192 wheat varieties.*
[b] *The physical positions of SNP markers based on wheat genome sequences from the International Wheat Genome Sequencing Consortium (IWGSC Ref Seq 1.0, http://www.wheatgenome.org/).*
[c] *ICIM indicates the inclusive composite interval mapping method performed by QTL IciMapping v4.1, GCIM indicates the genome-wide composite interval mapping method of QTL. gCIM.GUI v1.0, NWIM indicates composite interval mapping based on mixed linear model performed using forward–backward stepwise in QTLNetwork v2.0.*
[d] *LOD values from ICIM and GCIM > 2.5, P-values from NWIM < 0.05.*
[e] *PVE indicates phenotypic variance explained.*

TABLE 2 | Association of allelic variations of four CAPS markers for four novel QTLs with TGW, GW, and GL in 180 CMCC.

	2015TGW-CMCC	2016TGW-CMCC	2017TGW-CMCC	2016GW-CMCC	2017GW-CMCC	2016GL-CMCC	2017GL-CMCC
1B-JHMfel-C[a] Mean ± SD	38.46 ± 8.28	37.92 ± 8.78	37.51 ± 8.42	3.02 ± 0.42	3.01 ± 0.23	6.38 ± 0.85	6.50 ± 0.78
1B-JHMfel-T Mean ± SD	30.61 ± 6.04	30.82 ± 6.72	29.71 ± 5.78	2.82 ± 0.40	2.79 ± 0.20	5.94 ± 0.74	6.07 ± 0.53
U-test	6.547**	6.040**	6.427**	4.988**	5.740**	5.184**	5.509**
2A-CAPSmin-T[a] Mean ± SD	36.20 ± 8.18	36.01 ± 7.95	35.08 ± 8.04	3.00 ± 0.36	2.94 ± 0.22	6.28 ± 0.72	6.38 ± 0.64
2A-CAPSmin-C Mean ± SD	27.61 ± 6.24	29.01 ± 7.70	27.88 ± 6.43	2.78 ± 0.39	2.72 ± 0.22	5.95 ± 0.77	6.06 ± 0.77
U-test	6.256**	5.056**	4.748**	4.877**	4.659**	3.013**	2.445*
4B-8621-C[a] Mean ± SD	38.40 ± 7.34	38.39 ± 7.17	37.93 ± 6.15	3.07 ± 0.35	3.02 ± 0.18	6.45 ± 0.67	6.62 ± 0.48
4B-8621-G Mean ± SD	31.57 ± 7.53	31.37 ± 8.17	30.50 ± 7.96	2.82 ± 0.43	2.82 ± 0.24	5.95 ± 0.85	6.05 ± 0.72
U-test	5.863**	6.039**	6.169**	5.714**	5.571**	5.588**	5.943**
7A-3738-G[a] Mean ± SD	39.26 ± 7.54	39.09 ± 7.41	37.92 ± 7.63	3.09 ± 0.32	3.04 ± 0.20	6.48 ± 0.70	6.55 ± 0.73
7A-3738-C Mean ± SD	29.73 ± 5.68	29.72 ± 6.83	29.12 ± 6.03	2.76 ± 0.44	2.77 ± 0.20	5.85 ± 0.80	6.01 ± 0.54
U-test	8.077**	7.772**	7.365**	7.084**	7.121**	7.255**	6.735**

** and ** represent significance at 0.05 and 0.01 probability level, respectively.*

[a]The favorable allele with Jing 411-type band and high grain weight/size.

Hongmangchun 21 (**Figure 2B**). Then, a gene-specific marker, PMA-2, was developed based on this SNP and used to detect allelic variations of *TraesCS4B02G376400* among 180 CMCCs (**Figure 2C**). Compared with the Jing 411-type (*PMA-2-A*), the Hongmangchun 21-type (*PMA-2-G*) was significantly associated with lower TGW, GW, and GL values across environments ($P < 0.01$) (**Table 3**).

DISCUSSION

Reliability of QTL Mapping

The reliability of QTL mapping is crucial for subsequent fine mapping and map-based cloning of target genes and is affected by the density and genotyping of markers used to construct genetic linkage maps (Cui et al., 2017; Han et al., 2019). Recently, RAD-seq (Baird et al., 2008), GBS-seq (Elshire et al., 2011), SLAF-seq (Sun et al., 2013), and SNP arrays (Wang et al., 2014)[10] were performed as low-cost and efficient strategies to develop abundant SNPs to increase the density of markers. In the present study, SLAF tags (4820), CAPS (74), and gene-specific markers (11) were collectively used to build a high-density genetic linkage map harboring 1614 bin markers spanning 2014.43 cM in length with an average marker interval of 1.22 cM (**Supplementary Figure 1**).

It is well known that different mapping methods are based on specific genetic models as well as their corresponding statistical hypothesis, from which the result is in fact a probability statement. Therefore, it is better to use multiple mapping methods to reduce the risks of identifying ghost QTLs or missing real QTLs (Su et al., 2010). Zhang et al. (2016) identified 63 and 16 additive QTLs for fiber strength of cotton using a composite interval mapping (CIM) method with WinQTLCartographer2.5 and an ICIM model by QTL IciMapping4.1, respectively, and four QTLs detected by both CIM and ICIM were considered as stable loci for the fiber strength of cotton. In our previous study, based on both single- and multi-locus MLMs, 23 and 39 MTAs for preharvest sprouting resistance were detected by GWAS, respectively, and 6 loci were jointly identified by the two models and were, thus, considered stable loci (Zhu et al., 2019).

In addition, a combination of QTL analysis and GWAS is also an important strategy to detect reliable loci. Wang et al. (2018) identified 41 QTLs for maize tocopherol content by linkage mapping in 6 RIL populations and 32 significant loci by GWAS in a diverse panel of 508 inbred lines, and a major QTL co-localized in both linkage analysis and GWAS was finely mapped and characterized as a non-tocopherol pathway gene involved in the modulation of natural tocopherol variation.

In this study, QTL analysis with three mapping methods (ICIM, GCIM, and NWIM) and GWAS with MLM were collectively performed to identify loci for grain weight and size. Notably, five QTLs for TGW, two for GW, and four for GL were identified by all three mapping methods in JH-RILs. Six of them, including *Qtgw.ahau-4B.1*, *Qtgw.ahau-4B.2*, *Qgw.ahau-2D*, *Qgw.ahau-4B.3*, *Qgl.ahau-2A.1*, and *Qgl.ahau-7A.2*, were

[10]https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php

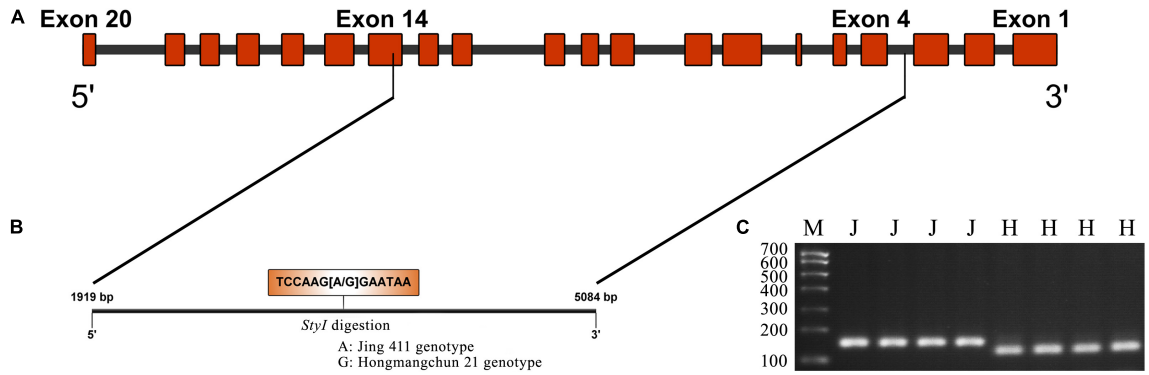


FIGURE 2 | CAPS marker development based on a SNP in *TraesCS4B01G376400*. **(A)** Diagram of the gene structure of *TraesCS4B01G376400*. **(B)** SNP (A–G) found in the partial sequence (1919–5084 bp) of *TraesCS4B01G376400*. **(C)** Electrophoresis patterns of the CAPS marker PMA-2 digested by *Sty*I between two parents Jing 411 and Hongmangchun 21. M indicates marker. "J" and "H" indicate electrophoresis patterns of Jing 411 genotype (undigested) and Hongmangchun 21 genotype (digested).

TABLE 3 | Association of allelic variations of the CAPS marker PMA-2 for *Qtgw.ahau-4B.2* with TGW, GW, and GL in 180 CMCC.

Trait	Allele	Mean ± SD	*U-test*
2015TGW-CMCC	PMA-2-A	36.35 ± 7.45	4.64**
	PMA-2-G	29.29 ± 5.45	
2016TGW-CMCC	PMA-2-A	36.21 ± 7.83	3.84**
	PMA-2-G	29.87 ± 6.61	
2017TGW-CMCC	PMA-2-A	35.50 ± 7.54	3.49**
	PMA-2-G	29.28 ± 6.66	
2016GW-CMCC	PMA-2-A	6.30 ± 0.73	3.34**
	PMA-2-G	5.94 ± 0.72	
2016GW-CMCC	PMA-2-A	6.45 ± 0.61	3.27**
	PMA-2-G	6.09 ± 0.29	
2016GL-CMCC	PMA-2-A	3.00 ± 0.37	3.57**
	PMA-2-G	2.76 ± 0.46	
2017GL-CMCC	PMA-2-A	2.95 ± 0.22	2.65**
	PMA-2-G	2.79 ± 0.24	

PMA-2-A indicates the Jing 411-type associated with higher TGW, GW, and GL values, PMA-2-G indicates the Hongmangchun 21-type associated with lower TGW, GW, and GL values.
***Represents significance at 0.01 probability level.*

validated by GWAS in 192 WVs. Thereby, these loci were considered to be reliable (**Table 1**, **Supplementary Table 9**, and **Supplementary Figure 2**).

Comparisons With Previous Studies

All QTLs for grain weight and size detected in the present study were compared with previously reported loci based on the wheat reference genome sequence (IWGSC RefSeq 1.0; IWGSC, 2018). In total, 23 (76.7%) of 30 QTLs for TGW, 11 (73.3%) of 15 for GW, and 13 (72.2%) of 18 for GL were completely or partially coincident with previous results (**Supplementary Table 8**). *Qtgw.ahau-2D* and *Qgw.ahau-2D.3* (543–561 Mb) were detected close to a major QTL for TGW (*Q.TKW.ui-2D-1*, which is flanked by cfd73 and IWB1093, 553 Mb) that was identified by Liu et al. (2018). *Qgl.ahau-4B.1* (31–38 Mb), identified in JH-RILs, was close to a major and stable locus for TGW (TKW-AX_110713957, 42 Mb) (Li et al., 2019). Moreover, *Qtgw.ahau-6B.1* and *Qgl.ahau-6B* (676–679 Mb) were detected near a stable QTL for TGW (TKW-AX_109917592, 675–677 Mb) that was identified by Li et al. (2019) (**Table 1**).

It is noteworthy that six QTL regions detected in this study contained eight reported genes for grain weight and size, including *TaFlo2-A1* (2A) (Sajjad et al., 2017), *TaSnRK2.9-5A* (Rehman et al., 2019), *TaGL3-5A* (Yang et al., 2019a), *TaGW2-6A* (Su et al., 2012; Yang et al., 2012; Hong et al., 2014; Qin et al., 2014; Vandana et al., 2015; Wang et al., 2017), *TaSus1-7A* (Hou et al., 2014), *TaGASR7-7A* (Dong et al., 2014), and *TaTGW-7A* (Hu et al., 2016b). This supports the reliability of QTL mapping performed in the current study. However, most of these reported genes had minor effects on grain weight and size traits in JH-RILs except for *Qtgw.ahau-7A.1* (co-located with *Qgw.ahau-7A.1* and *Qgl.ahau-7A.1*), indicating that control of grain weight and size is multigenic (**Supplementary Table 8**).

In particular, several putative, novel QTLs were identified in the present study compared to previous studies. For instance, *Qtgw.ahau-4B.1*, *Qtgw.ahau-4B.2*, and *Qgl.ahau-2A.2*, which explained 12.4, 11.7, and 12.1% of the phenotypic variances, on average, respectively, have not been reported by previous studies and are, thus, regarded as novel loci (**Table 1** and **Supplementary Table 8**). Interestingly, two loci for TGW and GL detected by all three mapping methods (*Qtgw.ahau-1B.1* and *Qgl.ahau-7A.2*) in the present study overlapped with the reported QTLs associated with kernel number per spike (KNS) (Cui et al., 2014; Wang et al., 2017), indicating a pleiotropy or close linkage between them (**Table 1** and **Supplementary Table 8**).

Pleiotropic or Co-localized QTLs

Thousand grain weight shows significant, positive correlations with GW and GL as shown previously and in the current study (Sun et al., 2009; Drikvand et al., 2013; Wang J. et al., 2020) (**Supplementary Tables 5**, **6**), suggesting that grain weight and size were simultaneously selected in high-yield wheat breeding. In this study, several QTL intervals were identified

for multiple yield-related traits. For example, two pairs of QTLs related to TGW and GW were detected in same intervals on chromosomes 2D and 4A; four QTL intervals related to both TGW and GL were identified on chromosomes 1B, 2A, 3A, and 6B; two QTL intervals for both GW and GL were detected on chromosomes 2D and 4B. Moreover, six QTL intervals associated with TGW, GW, and GL were identified on chromosomes 1B, 4B (2), 7A (2), and 7B, respectively. These include four novel QTL intervals, 648–652 Mb of chromosome 1B (*Qtgw.ahau-1B.1*, *Qgw.ahau-1B.1*, and *Qgl.ahau-1B.1*), 540–542 Mb of chromosome 4B (*Qtgw.ahau-4B.1*, *Qgw.ahau-4B.2*, and *Qgl.ahau-4B.2*), 660.11–660.42 Mb of chromosome 4B (*Qtgw.ahau-4B.2*, *Qgw.ahau-4B.3*, and *Qgl.ahau-4B.4*), and 446–496 Mb of chromosome 7A (*Qtgw.ahau-7A.2*, *Qgw.ahau-7A.2*, and *Qgl.ahau-7A.2*). Interestingly, four novel QTL intervals for grain weight and size identified in this study were co-located with QTLs for KNS detected in previous research, including *Qtgw.ahau-1B.1* (co-located with *Qgl.ahau-1B.1*) (Cui et al., 2014), *Qtgw.ahau-3A.2* (Li et al., 2019), *Qtgw.ahau-4D* (Cui et al., 2014), and *Qtgw.ahau-7A.2* (co-located with *Qgw.ahau-7A.2* and *Qgl.ahau-7A.2*) (Wang et al., 2017). In brief, these findings indicate that the above QTL intervals are either pleiotropic or tightly linked regions controlling multiple yield-related traits.

Candidate Gene Prediction of *Qtgw.ahau-4B.2*

The new stable locus *Qtgw.ahau-4B.2* spanning a physical interval of 0.31 Mb was detected in 10 of 11 environments in which 19 genes were annotated (Table S10). Among them, *TraesCS4B02G376400* and *TraesCS4B02G376800* encode a plasma membrane H^+-ATPase and a serine/threonine-protein kinase, respectively. The plasma membrane (PM) H^+-ATPase is an important ion pump in the plant cell membrane. By extruding protons from the cell and generating a membrane potential, this pump energizes the PM, which is a prerequisite for plant growth. Thus, the PM H^+-ATPase is regarded as a driver of growth (Falhof et al., 2016). Rober-Kleber et al. (2003) indicated that the PM H^+-ATPase is involved in auxin-mediated cell elongation during wheat embryo development. Auxin activates the proton pump, resulting in apoplastic acidification that contributes to cell wall loosening and elongation of the scutellum. Therefore, the PM H^+-ATPase is a component of the auxin-signaling cascade that may direct pattern formation in embryos. Moreover, several reported genes related to TGW belong to the serine/threonine-protein kinase family, such as *TaSnRK2.3* (Miao et al., 2017), *TaSnRK2.9* (Rehman et al., 2019), *TaSnRK2.10* (Zhang Z. et al., 2017), and *Tasg-D1* (Cheng et al., 2020), implying that the serine/threonine-protein kinase proteins play important roles in regulation of wheat grain development. Taken together, *TraesCS4B02G376400* and *TraesCS4B02G376800* are likely to be candidate genes of *Qtgw.ahau-4B.2*.

CONCLUSION

A total of 30, 15, and 18 putative additive QTLs for TGW, GW, and GL, respectively, were identified by SLAF-map in JH-RILs using three mapping methods. Particularly, five novel QTLs with stable and significant effects, including *Qtgw.ahau-1B.1*, *Qgl.ahau-2A.1*, *Qtgw.ahau-4B.1*, *Qtgw.ahau-4B.2*, and *Qgl.ahau-7A.2* were identified by all three mapping methods and further validated in a natural population. In addition, *TraesCS4B02G376400* and *TraesCS4B02G376800* were considered as potential candidate genes underlying *Qtgw.ahau-4B.2*. The novel QTLs and CAPS markers developed will be helpful for map-based cloning of the target regions and gene pyramiding in breeding for wheat PHS resistance.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JC and YS conceived the study, put into effect the main linkage analyses, and drafted the manuscript. DX, KX, and XC took part in the experiments and drafting of the manuscript. XP, XL, ML, CG, SY, HY, and WG processed the experimental data and helped to draft the manuscript. JL, HZ, CC, XX, SX, and CM conceived and guided the experiments, and helped in coordinating the project and drafting the manuscript. All the authors read and accepted the final manuscript.

FUNDING

Our research was supported by grants from the National Key Research and Development plan "Breeding new wheat varieties with high-yielding, high-quality and water-saving in the south of Huang-Huai River winter wheat area"—the breeding of new wheat germplasm and varieties with resistance to adversity (2017YFD0100703); The breeding and industrialization of new wheat varieties with high yield and multi-resistance (18030701182), The China Agriculture Research System (CARS-03); The National Key Research and Development Program of China (2016YFD0101802, 2017YFD0100804); Wheat Genetics and Breeding Research Platform Innovation Team of Anhui's University (2015-), Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP); The Agriculture Research System of Anhui Province (AHCYTX-02), and The University Synergy Innovation Program of Anhui Province (GXXT-2019-033).

ACKNOWLEDGMENTS

We thank Prof. Jizeng Jia for kindly providing Chinese micro-core wheat collections. We would like to thank TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.584859/full#supplementary-material

Supplementary Figure 1 | Genetic map containing 1614 bin markers for linkage mapping in the Jing 411/Hongmangchun 21 RILs population.

Supplementary Figure 2 | Manhattan plots and Quantile–quantile (Q-Q) plots for TGW, GW, and GL of 192 WVs (lines) by the MLM in Tassel v5.0.

Supplementary Figure 3 | Electrophoresis patterns of four CAPS markers 1B-JHMfeI, 2A-CAPSmin, 4B-8621 and 7A-3738 between two parents Jing 411 and Hongmangchun 21.

Supplementary Table 1 | Information of 74 CAPS markers developed in this study.

Supplementary Table 2 | Information on 11 gene-specific markers.

Supplementary Table 3 | Primers used for cloning and marker development of *TraesCS4B02G376400*.

Supplementary Table 4 | Phenotypic data of grain weight and size in the Jing 411/Hongmangchun 21 RIL population and 192 WVs.

Supplementary Table 5 | Pearson's correlations between TGW, GW, and GL in the Jing 411/Hongmangchun 21 RIL population.

Supplementary Table 6 | Pearson's correlations between TGW, GW, and GL in 192 WVs.

Supplementary Table 7 | Summary of the high-density genetic linkage bin maps constructed by SLAF-seq, CAPS, and gene-specific markers in the Jing 411/Hongmangchun 21 population.

Supplementary Table 8 | QTLs for grain weight and size identified in the Jing 411/Hongmangchun 21 RIL population.

Supplementary Table 9 | Thirty-one markers for grain weight and size in 192 WVs identified by GWAS.

Supplementary Table 10 | Gene annotation of the *Qtgw.ahau-4B.2* interval.

REFERENCES

Ashikari, M., Sakakibara, H., Lin, S. Y., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745. doi: 10.1126/science.1113373

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e03376. doi: 10.1371/journal.pone.0003376

Börner, A., Schumann, E., Fürste, A., Cöster, H., Leithold, B., Röder, M., et al. (2002). Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 105, 921–936. doi: 10.1007/s00122-002-0994-1

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Breseghello, F., and Sorrells, M. E. (2007). QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. *Field Crops Res.* 101, 172–179. doi: 10.1016/j.fcr.2006.11.008

Cao, S., Xu, D., Hanif, M., Xia, X., and He, Z. (2020). Genetic architecture underpinning yield component traits in wheat. *Theor. Appl. Genet.* 133, 1811–1823. doi: 10.1007/s00122-020-03562-8

Chagne, D., Lalanne, C., Madur, D., Kumar, S., Frigerio, J., Krier, C., et al. (2002). A high density genetic map of maritime pine based on AFLPs. *Ann. For. Sci.* 59, 627–636. doi: 10.1051/forest:2002048

Chang, C., Feng, J. M., Si, H. Q., Yin, B., Zhang, H. P., and Ma, C. X. (2010). Validating a novel allele of *viviparous-1* (*Vp-1Bf*) associated with high seed dormancy of Chinese wheat landrace, Wanxianbaimaizi. *Mol. Breeding* 25, 517–525. doi: 10.1007/s11032-009-9350-3

Chang, C., Lu, J., Zhang, H., Ma, C., and Sun, G. (2015). Copy number variation of cytokinin oxidase gene *Tackx4* associated with grain weight and chlorophyll content of flag leaf in common wheat. *PLoS One* 10:e0145970. doi: 10.1371/journal.pone.0145970

Chang, J., Zhang, J., Mao, X., Li, A., Jia, J., and Jing, R. (2013). Polymorphism of *TaSAP1-A1* and its association with agronomic traits in wheat. *Planta* 237, 1495–1508. doi: 10.1007/s00425-013-1860-x

Chen, G., Zhang, H., Deng, Z., Wu, R., and Tian, J. (2016). Genome-wide association study for kernel weight-related traits using SNPs in a Chinese winter wheat population. *Euphytica* 212, 1–13. doi: 10.1007/s10681-016-1750-y

Cheng, X., Xin, M., Xu, R., Chen, Z., Cai, W., Chai, L., et al. (2020). A single amino acid substitution in STKc_GSK3 kinase conferring semispherical grains and its implications for the origin of *Triticum sphaerococcum* Perc. *Plant Cell* 32, 923–934. doi: 10.1105/tpc.19.00580

Chong, C. F., Li, Y. C., Wang, T. L., and Hang, C. (2003). Stratification of adverse outcomes by preoperative risk factors in coronary artery bypass graft patients: an artificial neural network prediction model. *Amia Annu. Symp. Proc.* 2003, 160–164.

Cui, F., Zhang, N., Fan, X. L., Zhang, W., Zhao, C. H., Yang, L. J., et al. (2017). Utilization of a wheat 660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel numbe. *Sci. Rep.* 7:3788. doi: 10.1038/s41598-017-04028-6

Cui, F., Zhao, C., Ding, A., Li, J., Wang, L., Li, X., et al. (2014). Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theor. Appl. Genet.* 127, 659–675. doi: 10.1007/s00122-013-2249-8

Cuthbert, J. L., Somers, D. J., Brulebabel, A., Brown, P. D., and Crow, G. H. (2008). Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 117, 595–608. doi: 10.1007/s00122-008-0804-5

Dholakia, B. B., Ammiraju, J. S. S., Singh, H., Lagu, M. D., Röder, M. S., Rao, V. S., et al. (2003). Molecular marker analysis of kernel size and shape in bread wheat. *Plant Breeding* 122, 392–395. doi: 10.1046/j.1439-0523.2003.00896.x

Dong, L., Wang, F., Liu, T., Dong, Z., Li, A., Jing, R., et al. (2014). Natural variation of *TaGASR7-A1* affects grain length in common wheat under multiple cultivation conditions. *Mol. Breeding* 34, 937–947. doi: 10.1007/s11032-014-0087-2

Drikvand, R., Bihamta, M. R., Najafian, G., and Ebrahimi, A. (2013). Kernel quality association and path analysis in bread wheat. *Int. J. Biol.* 5, 73–79. doi: 10.5539/ijb.v5n3p73

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Falhof, J., Pedersen, J. T., Fuglsang, A. T., and Palmgren, M. (2016). Plasma membrane H^+-ATPase regulation in the center of plant physiology. *Mol. Plant* 9, 323–337. doi: 10.1016/j.molp.2015.11.002

Fiedler, J. D., Evan, S., Yuan, L., Monika, M. D. J., Hegstad, J. B., Chen, B., et al. (2017). Genome-wide association and prediction of grain and semolina quality traits in durum wheat breeding populations. *Plant Genome* 10, 1–12. doi: 10.3835/plantgenome2017.05.0038

Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese spring. *Front. Plant Sci.* 6:1099. doi: 10.1007/s10681-016-1682-6

Garcia, M., Eckermann, P., Haefele, S., Satija, S., Sznajder, B., Timmins, A., et al. (2019). Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS One* 14:e0211730. doi: 10.1371/journal.pone.0211730

Guo, G., Wang, S., Liu, J., Pan, B., Diao, W., Ge, W., et al. (2017). Rapid identification of QTLs underlying resistance to cucumber mosaic virusin

pepper (*Capsicum frutescens*). *Theor. Appl. Genet.* 130, 41–52. doi: 10.1007/s00122-016-2790-3

Han, J., Han, D., Guo, Y., Yan, H., and Qiu, L. (2019). QTL mapping pod dehiscence resistance in soybean (*Glycine max* L. Merr.) using specific-locus amplified fragment sequencing. *Theor. Appl. Genet.* 132, 2253–2272. doi: 10.1007/s00122-019-03352-x

Himi, E., Maekawa, M., Miura, H., and Noda, K. (2011). Development of PCR markers for *Tamyb10* related to *R-1*, red grain color gene in wheat. *Theor. Appl. Genet.* 122, 1561–1576. doi: 10.1007/s00122-011-1555-2

Hong, Y., Chen, L., Du, L. P., Su, Z., Wang, J., Ye, X., et al. (2014). Transcript suppression of *TaGW2* increased grain width and weight in bread wheat. *Funct. Integr. Genomics* 14, 341–349. doi: 10.1007/s10142-014-0380-5

Hou, J., Jiang, Q., Hao, C., Wang, Y., Zhang, H., and Zhang, X. (2014). Global selection on sucrose synthase haplotypes during a century of wheat breeding. *Plant Physiol.* 164, 1918–1929. doi: 10.1104/pp.113.232454

Hu, M. J., Zhang, H. P., Cao, J. J., Zhu, X. F., Wang, S. X., Jiang, H., et al. (2016a). Characterization of an IAA-glucose hydrolase gene *TaTGW6* associated with grain weight in common wheat (*Triticum aestivum* L.). *Mol. Breeding* 36:25. doi: 10.1007/s11032-016-0449-z

Hu, M. J., Zhang, H. P., Liu, K., Cao, J. J., Wang, S. X., Jiang, H., et al. (2016b). Cloning and characterization of *TaTGW-7A* gene associated with grain weight in wheat via SLAF-seq-BSA. *Front. Plant Sci.* 7:1902. doi: 10.3389/fpls.2016.01902

Huang, S., Wu, J., Wang, X., Mu, J., Xu, Z., Zeng, Q., et al. (2019). Utilization of the genomewide wheat 55K SNP array for genetic analysis of stripe rust resistance in common wheat line P9936. *Phytopathology* 109, 819–827. doi: 10.1094/phyto-10-18-0388-r

Huang, X. Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D. G., Noll, J. S., et al. (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 113, 753–766. doi: 10.1007/s00122-006-0346-7

Jiang, Q., Hou, J., Hao, C., Wang, L., Ge, H., Dong, Y., et al. (2011). The wheat (*T. aestivum*) *sucrose synthase 2* gene (*TaSus2*) active in endosperm development is associated with yield traits. *Funct. Integr. Genomics* 11, 49–61. doi: 10.1007/s10142-010-0188-x

Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., and Singh, R. P. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* 51, 1–10. doi: 10.1038/s41588-019-0496-6

Kang, H. W., Cho, Y. G., Yoon, U. H., and Eun, M. Y. (1998). A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed. *Plant Mol. Biol. Report* 16:90. doi: 10.1023/A:1007418606098

Keim, P., Diers, B. W., Olson, T. C., and Shoemaker, R. C. (1990). RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics* 126, 735–742. doi: 10.1007/BF00056365

Kulwal, P. L., Kumar, N., Gaur, A., Khurana, P., Khurana, J. P., Tyagi, A. K., et al. (2005). Mapping of a major QTL for pre-harvest sprouting tolerance on chromosome 3A in bread wheat. *Theor. Appl. Genet.* 111, 1052–1059. doi: 10.1007/s00122-005-0021-4

Kumar, A., Mantovani, E. E., Seetan, R., Soltani, A., Echeverry-Solarte, M., Jain, S., et al. (2016). Dissection of genetic factors underlying wheat kernel shape and size in an Elite × Nonadapted cross using a high density SNP linkage map. *Plant Genome* 9, 1–22. doi: 10.3835/plantgenome2015.09.0081

Li, B., Fan, S., Yu, F., Chen, Y., Zhang, S., Han, F., et al. (2017). High-resolution mapping of QTL for fatty acid composition in soybean using specific-locus amplified fragment sequencing. *Theor. Appl. Genet.* 130, 1467–1479. doi: 10.1007/s00122-017-2902-8

Li, F., Wen, W., Liu, J., Zhang, Y., and Cao, S. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* 19:168. doi: 10.1186/s12870-019-1781-3

Li, H., Ye, G., and Wang, J. (2007). A modified algorithm for the improvement of composite interval mapping. *Genetics* 175, 361–374. doi: 10.1534/genetics.106.066811

Li, M., Tang, D., Wang, K., Wu, X., Lu, L., Yu, H., et al. (2011). Mutations in the F-box gene *LARGER PANICLE* improve the panicle architecture and enhance the grain yield in rice. *Plant Biotech. J.* 9, 1002–1013. doi: 10.1111/j.1467-7652.2011.00610.x

Li, R., Song, W., Wang, B., Wang, J., Zhang, D., Zhang, Q., et al. (2018). Identification of a locus conferring dominant resistance to maize rough dwarf disease in maize. *Sci. Rep.* 8:3248. doi: 10.1038/s41598-018-21677-3

Li, S., Zhao, B., Yuan, D., Duan, M., Qian, Q., Tang, L., et al. (2013). Rice zinc finger protein DST enhances grain production through controlling *Gn1a/OsCKX2* expression. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3167–3172. doi: 10.1073/pnas.1300359110

Liu, C., Zhou, Q., Dong, L., Wang, H., Liu, F., Weng, J., et al. (2016). Genetic architecture of the maize kernel row number revealed by combining QTL mapping using a high-density genetic map and bulked segregant RNA sequencing. *BMC Genomics* 17:915. doi: 10.1186/s12864-016-3240-y

Liu, H., Li, H., Hao, C., Wang, K., Wang, Y., Qin, L., et al. (2019). *TaDA1*, a conserved negative regulator of kernel size, has an additive effect with *TaGW2* in common wheat (*Triticum aestivum* L.). *Plant Biotech. J.* 18, 1330–1342. doi: 10.1111/pbi.13298

Liu, J., Wu, B., Singh, R. P., and Velu, G. (2019). QTL mapping for mineral nutrient concentrations and agronomic-related traits in a hexaploid wheat mapping population. *J. Cereal Sci.* 88, 57–64. doi: 10.1016/j.jcs.2019.05.008

Liu, Y., Wang, R., Hu, Y. G., and Chen, J. (2018). Genome-wide linkage mapping of quantitative trait loci for late-season physiological and agronomic traits in spring wheat under irrigated conditions. *Agronomy* 8:60. doi: 10.3390/agronomy8050060

Ma, D., Yan, J., He, Z., Ling, W., and Xia, X. (2012). Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers. *Mol. Breeding* 29, 43–52. doi: 10.1007/s11032-010-9524-z

Ma, L., Li, T., Hao, C., Wang, Y., and Zhang, X. (2015). *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnol. J.* 14, 1269–1280. doi: 10.1111/pbi.12492

Ma, M., Wang, Q., Li, Z., Cheng, H., and Zhao, H. (2015). Expression of *TaCYP78A3*, a gene encoding cytochrome P450 CYP78A3 protein in wheat (*Triticum aestivum* L.), affects seed size. *Plant J.* 83, 312–325. doi: 10.1111/tpj.12896

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1016/S0002-9343(99)00454-4

Miao, L., Mao, X., Wang, J., Liu, Z., Zhang, B., Li, W., et al. (2017). Elite haplotypes of a protein kinase gene *TaSnRK2.3* associated with important agronomic traits in common wheat. *Front. Plant Sci.* 8:368. doi: 10.3389/fpls.2017.00368

Mir, R. R., Kumar, N., Jaiswal, V., Girdharwal, N., Prasad, M., Balyan, H. S., et al. (2012). Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. *Mol. Breeding* 29, 963–972. doi: 10.1007/s11032-011-9693-4

Mohler, V., Albrecht, T., Castell, A., Diethelm, M., Schweizer, G., and Hartl, L. (2016). Considering causal genes in the genetic dissection of kernel traits in common wheat. *J. Appl. Genet.* 57, 467–476. doi: 10.1007/s13353-016-0349-2

Organization of Food and Agriculture (2009). The state of food and agriculture. Livestock in the balance. *State Food Agric.* 79:410. doi: 10.1080/00369225508735601

Qin, L., Hao, C., Hou, J., Wang, Y., Li, T., Wang, L., et al. (2014). Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene *TaGW2*. *BMC Plant Biol.* 14:107. doi: 10.1186/1471-2229-14-107

Qu, L., and Hancock, J. F. (1997). Randomly amplified polymorphic DNA- (RAPD-) based genetic linkage map of blueberry derived from an interspecific cross between diploid *Vaccinium darrowi* and tetraploid *V. corymbosum*. *J. Am. Soc. Hortic. Sci.* 122, 69–73. doi: 10.1023/A:1008607226805

Quan, R., Wang, J., Hui, J., Bai, H., Lyu, X., Zhu, Y., et al. (2018). Improvement of salt tolerance using wild rice genes. *Front. Plant Sci.* 8:2269. doi: 10.3389/fpls.2017.02269

Quarrie, S. A., Steed, A., Calestani, C., Semikhodskii, A., Lebreton, C., Chinoy, C., et al. (2005). A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring×SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor. Appl. Genet.* 110, 865–880. doi: 10.1007/s00122-004-1902-7

Rehman, S. U., Wang, J., Chang, X., Zhang, X., Mao, X., and Jing, R. (2019). A wheat protein kinase gene *TaSnRK2.9-5A* associated with yield contributing traits. *Theor. Appl. Genet.* 132, 907–919. doi: 10.1007/s00122-018-3247-7

Rober-Kleber, N., Albrechtova, J. T. P., Fleig, S., Huck, N., Michalke, W., Wagner, E., et al. (2003). Plasma membrane H^+-ATPase is involved in auxin-mediated cell elongation during wheat embryo development. *Plant Physiol.* 131, 1302–1312. doi: 10.1104/pp.013466

Sajjad, M., Ma, X., Khan, S. H., Shoaib, M., Song, Y., Yang, W., et al. (2017). *TaFlo2-A1*, an ortholog of rice *Flo2*, is associated with thousand grain weight in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:164. doi: 10.1186/s12870-017-1114-3

Simmonds, J., Scott, P., Leveringtonwaite, M., Turner, A. S., Brinton, J., Korzun, V., et al. (2014). Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 14:191. doi: 10.1186/s12870-014-0191-9

Su, C., Lu, W., Zhao, T., and Gai, J. (2010). Verification and fine-mapping of QTLs conferring days to flowering in soybean using residual heterozygous lines. *Sci. Bull.* 55, 499–508. doi: 10.1007/s11434-010-0032-7

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2012). Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z

Sukumaran, S., Lopes, M., Dreisigacker, S., and Reynolds, M. (2017). Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. *Theor. Appl. Genet.* 131:999. doi: 10.1007/s00122-018-3066-x

Sun, X., Liu, D., Zhang, X., Li, W., Hui, L., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale De novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8:e58700. doi: 10.1371/journal.pone.0058700

Sun, X. Y., Wu, K., Zhao, Y., Kong, F. M., Han, G. Z., Jiang, H. M., et al. (2009). QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* 165, 615–624. doi: 10.1007/s10681-008-9794-2

Talbert, L. E., Blake, N. K., Chee, P. W., Blake, T. K., and Magyar, G. M. (1994). Evaluation of "sequence-tagged-site" PCR products as molecular markers in wheat. *Theor. Appl. Genet.* 87, 789–794. doi: 10.1007/BF00221130

Thorwarth, P., Liu, G., Ebmeyer, E., Schacht, J., Schachschneider, R., Kazman, E., et al. (2019). Dissecting the genetics underlying the relationship between protein content and grain yield in a large hybrid wheat population. *Theor. Appl. Genet.* 132, 489–500. doi: 10.1007/s00122-018-3236-x

Torada, A., Ikeguchi, S., and Koike, M. (2005). Mapping and validation of PCR-based markers associated with a major QTL for seed dormancy in wheat. *Euphytica* 143, 251–255. doi: 10.1007/s10681-005-7872-2

Torada, A., Koike, M., Ogawa, T., Takenouchi, Y., Tadamura, K., Wu, J., et al. (2016). A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. *Curr. Biol.* 26, 782–787. doi: 10.1016/j.cub.2016.01.063

Vandana, J., Vijay, G., Saloni, M., Priyanka, A., Kumar, K. M., Paul, K. J., et al. (2015). Identification of novel SNP in promoter sequence of *TaGW2-6A* associated with grain weight and other agronomic traits in wheat (*Triticum aestivum* L.). *PLoS One* 10:e0129400. doi: 10.1371/journal.pone.0129400

Vos, P., Hogers, R., Bleeker, M., Reijans, M., and Kuiper, M. (1995). AFLPs a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414. doi: 10.1093/nar/23.21.4407

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* 40, 1370–1374. doi: 10.1038/ng.220

Wang, F., Zhang, J., Chen, Y., Zhang, C., Gong, J., Song, Z., et al. (2020). Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G.barbadense* introgressions. *Plant Biotech. J.* 18, 707–720. doi: 10.1111/pbi.13237

Wang, H., Xu, S., Fan, Y., Liu, N., Zhan, W., Liu, H., et al. (2018). Beyond pathways: genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses. *Plant Biotech. J.* 16, 1464–1475. doi: 10.1111/pbi.12889

Wang, J., Wang, R., Mao, X., Zhang, J., Liu, Y., Xie, Q., et al. (2020). RING finger ubiquitin E3 ligase gene *TaSDIR1-4A* contributes to grain size in common wheat. *J. Exp. Bot.* 71:eraa271. doi: 10.1093/jxb/eraa271

Wang, L., Ge, H., Hao, C., Dong, Y., Zhang, X., and Hudson, M. E. (2012). Identifying loci influencing 1,000-kernel weight in wheat by microsatellite screening for evidence of selection during breeding. *PLoS One* 7:e29432. doi: 10.1371/journal.pone.0029432

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotech. J.* 12, 787–796.

Wang, S. X., Zhu, Y. L., Zhang, D. X., Shao, H., and Ma, C. X. (2017). Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. *PLoS One* 12:e0188662. doi: 10.1371/journal.pone.0188662

Wen, Y., Zhang, Y., Zhang, J., Feng, J., Dunwell, J. M., and Zhang, Y. (2019). An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F_2. *Brief. Bioinform.* 20, 1913–1924. doi: 10.1093/bib/bby058

Wu, Q., Chen, Y., Zhou, S., Fu, L., Chen, J., Xiao, Y., et al. (2015). High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population yanda1817 × beinong6. *PLoS One* 10:e0118144. doi: 10.1371/journal.pone.0118144

Xin, F., Zhu, T., Wei, S., Han, Y., Zhao, Y., Zhang, D., et al. (2020). QTL mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulked segregant analysis in wheat. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-019-56979-7

Xu, Y. F., Li, S. S., Li, L. H., Ma, F. F., Fu, X. Y., Shi, Z. L., et al. (2017). QTL mapping for yield and photosynthetic related traits under different water regimes in wheat. *Mol. Breeding* 37:34. doi: 10.1007/s11032-016-0583-7

Yang, J., Hu, C., Hu, H., Yu, R., Xia, Z., Ye, X., et al. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24, 721–723. doi: 10.1093/bioinformatics/btm494

Yang, J., Zhou, Y., Wu, Q., Chen, Y., Zhang, P., Zhang, Y., et al. (2019a). Molecular characterization of a novel *TaGL3-5A* allele and its association with grain length in wheat (*Triticum aestivum L.*). *Theor. Appl. Genet.* 132, 1799–1814. doi: 10.1007/s00122-019-03316-1

Yang, J., Zhou, Y., Zhang, Y., Hu, W., Wu, Q., Chen, Y., et al. (2019b). Cloning, characterization of *TaGS3* and identification of allelic variation associated with kernel traits in wheat (*Triticum aestivum* L.). *BMC Genetics* 20:1–11. doi: 10.1186/s12863-019-0800-6

Yang, J., Zhu, J., and Williams, R. W. (2007). Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 23, 1527–1536. doi: 10.1093/bioinformatics/btm143

Yang, X., Nong, B., Xia, X., Zhang, Z., Zeng, Y., Liu, K., et al. (2017). Rapid identification of a new gene influencing low amylose content in rice landraces (*Oryza sativa* L.) using genome-wide association study with specific-locus amplified fragment sequencing. *Genome* 60, 465–472. doi: 10.1139/gen-2016-0104

Yang, Y., Zhao, X. L., Xia, L. Q., Chen, X. M., Xia, X. C., Yu, Z., et al. (2007). Development and validation of a *Viviparous-1* STS marker for pre-harvest sprouting tolerance in Chinese wheats. *Theor. Appl. Genet.* 115, 971–980. doi: 10.1007/s00122-007-0624-z

Yang, Z., Bai, Z., Li, X., Wang, P., Wu, Q., Yang, L., et al. (2012). SNP identification and allelic-specific PCR markers development for *TaGW2*, a gene linked to wheat kernel weight. *Theor. Appl. Genet.* 125, 1057–1068. doi: 10.1007/s00122-012-1895-6

Yin, C., Li, H., Li, S., Xu, L., Zhao, Z., and Wang, J. (2015). Genetic dissection on rice grain shape by the two-dimensional image analysis in one *japonica* × *indica* population consisting of recombinant inbred lines. *Theor. Appl. Genet.* 128, 1969–1986. doi: 10.1007/s00122-015-2560-7

Ying, G., Sun, J., Zhang, G., Wang, Y., and Li, S. (2013). Haplotype, molecular marker and phenotype effects associated with mineral nutrient and grain size traits of *TaGS1a* in wheat. *Field Crops Res.* 154, 119–125. doi: 10.1016/j.fcr.2013.07.012

Yue, A., Li, A., Mao, X., Chang, X., Li, R., and Jing, R. (2015). Identification and development of a functional marker from *6-SFT-A2* associated with grain weight in wheat. *Mol. Breeding* 35:63. doi: 10.1007/s11032-015-0266-9

Zhang, J., Gizaw, S. A., Bossolini, E., Hegarty, J., Howell, T., Carter, A. H., et al. (2018). Identification and validation of QTL for grain yield and plant water status under contrasting water treatments in fall-sown spring wheats. *Theor. Appl. Genet.* 131, 1741–1759. doi: 10.1007/s00122-018-3111-9

Zhang, X., Wang, G., Dong, T., Chen, B., Du, H., Li, C., et al. (2019). High-density genetic map construction and QTL mapping of first flower node in pepper (*Capsicum annuum* L.). *BMC Plant Biol.* 19:167. doi: 10.1186/s12870-019-1753-7

Zhang, Y., Liu, J., Xia, X., and He, Z. (2014). *TaGS-D1*, an ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. *Mol. Breeding* 34, 1097–1107. doi: 10.1007/s11032-014-0102-7

Zhang, Y., Xia, X., and He, Z. (2017). The seed dormancy allele *TaSdr-A1a* associated with pre-harvest sprouting tolerance is mainly present in Chinese wheat landraces. *Theor. Appl. Genet.* 130, 81–89. doi: 10.1007/s00122-016-2793-0

Zhang, Z., Lv, G., Li, B., Wang, J., Zhao, Y., Kong, F., et al. (2017). Isolation and characterization of the *TaSnRK2.10* gene and its association with agronomic traits in wheat (*Triticum aestivum* L.). *PLoS One* 12:e0174425. doi: 10.1371/journal.pone.0174425

Zhang, Z., Shang, H., Shi, Y., Huang, L., Li, J., Ge, Q., et al. (2016). Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum.*). *BMC Plant Biol.* 16:79. doi: 10.1186/s12870-016-0741-4

Zhu, X. F. (2016). Cloning and expression analysis of *TaPTF1* and *Tabas1* genes associated with thousand grain weight in common wheat (*Triticum aestivum* L.) [D]. in Chinese with English abstract. *Anhui Agric. Univ.*

Zhu, X. F., Zhang, H. P., Hu, M. J., Wu, Z. Y., Jiang, H., Cao, J. J., et al. (2016). Cloning and characterization of *Tabas1-B1* gene associated with flag leaf chlorophyll content and thousand-grain weight and development of a gene-specific marker in wheat. *Mol. Breeding* 36:142. doi: 10.1007/s11032-016-0563-y

Zhu, Y., Wang, S., Wei, W., Xie, H., Liu, K., Zhang, C., et al. (2019). Genome-wide association study of pre-harvest sprouting tolerance using a 90K SNP array in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 2947–2963. doi: 10.1007/s00122-019-03398-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past collaboration with one of the authors SX.

Copyright © 2020 Cao, Shang, Xu, Xu, Cheng, Pan, Liu, Liu, Gao, Yan, Yao, Gao, Lu, Zhang, Chang, Xia, Xiao and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 12 November 2020
doi: 10.3389/fgene.2020.586462

Phytohormone-Mediated Molecular Mechanisms Involving Multiple Genes and QTL Govern Grain Number in Rice

*Priyanka Deveshwar, Ankita Prusty, Shivam Sharma and Akhilesh K. Tyagi**

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi, New Delhi, India

Increasing the grain number is the most direct route toward enhancing the grain yield in cereals. In rice, grain number can be amplified through increasing the shoot branching (tillering), panicle branching, panicle length, and seed set percentage. Phytohormones have been conclusively shown to control the above characteristics by regulating molecular factors and their cross-interactions. The dynamic equilibrium of cytokinin levels in both shoot and inflorescence meristems, maintained by the regulation of its biosynthesis, activation, and degradation, determines the tillering and panicle branching, respectively. Auxins and gibberellins are known broadly to repress the axillary meristems, while jasmonic acid is implicated in the determination of reproductive meristem formation. The balance of auxin, gibberellin, and cytokinin determines meristematic activities in the inflorescence. Strigolactones have been shown to repress the shoot branching but seem to regulate panicle branching positively. Ethylene, brassinosteroids, and gibberellins regulate spikelet abortion and grain filling. Further studies on the optimization of endogenous hormonal levels can help in the expansion of the grain yield potential of rice. This review focuses on the molecular machinery, involving several genes and quantitative trait loci (QTL), operational in the plant that governs hormonal control and, in turn, gets governed by the hormones to regulate grain number and yield in rice.

Keywords: grain number, panicle branching, phytohormones, rice, tiller, yield

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University, United States

Reviewed by:
Ying Wang,
University of Chinese Academy of Sciences, China
Sepehr Mohajeri Naraghi,
North Dakota State University, United States

***Correspondence:**
Akhilesh K. Tyagi
akhilesh@genomeindia.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *23 July 2020*
Accepted: *09 October 2020*
Published: *12 November 2020*

Citation:
Deveshwar P, Prusty A, Sharma S and Tyagi AK (2020) Phytohormone-Mediated Molecular Mechanisms Involving Multiple Genes and QTL Govern Grain Number in Rice. Front. Genet. 11:586462. doi: 10.3389/fgene.2020.586462

INTRODUCTION

The world's population is predicted to crest over 9 billion by 2050, with the current growth rate. Changes in the patterns of consumption and climate demand a holistic approach to increase food production and achieve global food security (Stamm et al., 2011; Cole et al., 2018; Tyczewska et al., 2018). Rice is the largest contributor to a daily per capita average calorie consumption (19%) in the world. Although current global rice production is in surplus, increasing population growth poses a threat to adequate supplies in the future (Elert, 2014). Hence, rice grain yield should be increased to fulfill future demands, and various strategies have been suggested to increase the yield potential (Khush, 2013). Grain number is a major trait of grain yield in rice and is a direct function of the number of productive tillers, numbers of panicles per plant, and panicle branching. Other traits like panicle size or panicle length and effective seed set also contribute to the grain number and hence grain yield in rice. Panicle length determines the number of primary branches,

whereas the length of the panicle branch determines the compactness of the inflorescence. The seed set rate is governed by grain filling and the ratio of seed abortion. In rice, many genes and quantitative trait loci (QTL) have been identified that determine the grain yield and are an indispensable genetic resource for crop improvement efforts worldwide (Jeon et al., 2011; Sreenivasulu and Schnurbusch, 2012; Gouda et al., 2020; Sakuma and Schnurbusch, 2020).

In crop plants, the branching pattern of lateral organs plays a central role in defining the plant architecture and the crop yield (Wang and Li, 2005). Shoot branches in rice produce tillers at the vegetative stage and panicles at the reproductive stage (McSteen and Leyser, 2005). Panicles are highly branched inflorescences that are produced on the tillers, and multiple axillary meristems initiate their development. The final panicle architecture is pre-established during development at the meristematic zones that define the branching arrangement and positioning of spikelets. The branch meristem produces primary and secondary branches, spikelet meristem forms spikelets, and floral meristem is responsible for floret and floral organ development. Phytohormones are the major internal factors that regulate these developmental events (Barazesh and McSteen, 2008; McSteen, 2009).

Phytohormones are small regulatory molecules that form an elaborate and sophisticated regulatory network in coordinating various developmental aspects of yield-related traits and thus control the yield potential of the plant (Zhang and Yuan, 2014). Auxins play a role in axillary meristem initiation, while cytokinins are involved in regulating meristem size and activity and thus affect panicle branching indirectly. Besides, cytokinin also promotes axillary bud outgrowth, whereas auxins and strigolactones inhibit axillary bud growth and affect panicle development (Ferguson and Beveridge, 2009; Shimizu-Sato et al., 2009; Dun et al., 2012). Cytokinins, auxins, and jasmonic acid are also involved in meristem fate determination (Zhang and Yuan, 2014). Other hormones like brassinosteroids, gibberellins, and ethylene are also involved in defining the features of panicle morphology and seed set. Gibberellins and brassinosteroids have been shown to regulate the spikelet abortion in addition to shoot branching (Ali et al., 2019). Ethylene is revealed to regulate grain filling and contribute to an effective seed set (Ali et al., 2019). Studies suggest that an intricate network of hormonal pathways regulates the panicle development and modulates panicle architecture (Zhang and Yuan, 2014).

Balancing panicle-related traits such as panicle length, panicle number, and the numbers of primary and secondary branches per panicle are key to improving the number of grains per plant. These traits can be manipulated by fine calibration of molecular component of hormonal signaling to enhance the grain yield efficiently. Thus, the information regarding the hormonal pathways, their homeostasis, and their complex networks is instrumental in designing high yielding crops for the future. The current review provides an overview of the recent progress of the genetics of phytohormone actions and their crosstalks in the context of grain number as a complex agronomic trait in rice.

CONTROL OF GRAIN NUMBER BY PHYTOHORMONES

Auxins, cytokinins, and strigolactones have been convincingly shown to have roles in shoot and panicle branching and also communicate among them to regulate it (Shimizu-Sato et al., 2009; Dun et al., 2012; Zha et al., 2019). Ethylene, brassinosteroids, and auxins interact to regulate panicle differentiation and degeneration (Ali et al., 2019). The role of each hormone in the regulation of grain number trait in rice will be discussed below. Recent elaborated studies that establish various cross-hormonal interactions involved in regulating grain number in rice are discussed in a separate section. **Table 1** is a compilation of major studies done so far to characterize genes involved in hormonal regulation of grain number.

Cytokinins

Cytokinins (CKs) are a class of adenine-derived compounds that are categorized as phytohormones involved primarily in cell divisions (Werner et al., 2001; Schaller et al., 2014). Natural CKs can be of two types depending upon the nature of the side chain attached to the adenine moiety. These side chains can be either an aromatic or an isoprene derivative (Sakakibara, 2006). The effect of side chain has been shown to affect shoot growth in *Arabidopsis* (Kiba et al., 2013); however, such effects are not confirmed in other plants. CK homeostasis is maintained in the meristems by a dynamic balance between its biosynthesis, activation, deactivation, reactivation, and degradation (Jameson and Song, 2016).

The first important QTL identified in rice for the increase in grain number was *Gn1a*, located on the short arm of chromosome 1 (Ashikari et al., 2005). The *Gn1a* allele of the high-yielding *indica* Habataki variety was held responsible for its 44% more number of grains per panicle compared to Koshihikari *japonica* variety. The *Gn1a* locus encodes for a cytokinin oxidase/dehydrogenase (OsCKX2), an enzyme that degrades the active form of CK irreversibly into adenine or adenosine and the side chains (**Figure 1**). Its loss-of-function mutation accumulated bioactive CK in the inflorescence meristems. Since CK regulates lateral meristem activity, more number of spikelets are formed that results in increased grain yield (Ashikari et al., 2005). Another QTL for grain number was identified on chromosome 4 and called as *GN4-1*. However, this QTL was fine mapped to a large 190-kb region consisting of 20 genes. Near-isogenic lines (NILs) carrying *GN4-1* also showed increased CK accumulation in young panicles and decreased transcript abundance of as many as eight CKX enzyme family members, including *OsCKX2* (Zhou et al., 2018).

A zinc-finger transcription factor, DROUGHT AND SALT TOLERANCE (DST), has been shown to negatively regulate the CK accumulation in the reproductive meristem by directly regulating the expression of *OsCKX2* (**Figure 1**). The mutant, *regulator of Gn1a* (*reg1*), was identified as a semidominant allele of *DST* called DST^{reg1} that showed increased activity in the inflorescence meristem due to higher CK accumulation, resulting in more panicle branching and greater grain yield.

TABLE 1 | Some genes/quantitative trait loci (QTL) regulating the grain number per plant in rice.

Gene/QTL	RGAP Locus ID	Hormone involved/ affected	Nature of protein encoded	Phenotype obtained for	Phenotype related to grain yield	References
AFB6	LOC_Os03g08850	Auxin and CK	Auxin receptor	Overexpression	Increase in spikelets per panicle and primary branch number; high photoperiodic sensitivity; delayed heading	He et al., 2018
ASP1	LOC_Os08g06480	Auxin	TOPLESS-related transcriptional co-repressor	Recessive mutant	Aberrations in spikelet morphology and branching pattern; reduction in rachis length and spikelet number with multiple phenotypes	Yoshida et al., 2012
BG1	LOC_Os03g07920	Auxin	Novel membrane localized protein	Dominant mutant/overexpression	Increase in grain size, panicle size, plant height along with other pleiotropic phenotypes	Liu L. et al., 2015
CYP71D8L	LOC_Os02g09220	GA	Cytochrome P450 monooxygenase	Gain-of-function mutant and overexpression	Reduced panicle length, grain number per panicle, and reduced plant height	Zhou et al., 2020
D2, SMG11	LOC_Os01g10040	BR	Cytochrome P450 (CYP90D2)	Recessive mutant	Dense and erect panicles; increase in secondary branching and number of grains per panicle; small grains	Fang et al., 2016
D3	LOC_Os06g06050	SL	F-box protein	Loss-of-function mutant	Reduction in panicle size and plant height; higher tillering	Ishikawa et al., 2005
D4	LOC_Os03g12660	BR	Cytochrome P450 (CYP90B2)	Overexpression	Restores clustered panicle branching phenotype in *pmm1-1* background	Li et al., 2018
D10	LOC_Os01g54270	SL	CCD8	Loss-of-function mutant	Reduction in panicle size and plant height; higher tillering	Ishikawa et al., 2005; Arite et al., 2007
D14, HTD2, D88	LOC_Os03g10620	SL	α/β-fold hydrolase (receptor)	Loss-of-function mutant	Reduction in panicle size and plant height; higher tillering; smaller seed size	Ishikawa et al., 2005; Wang et al., 2017; Liang et al., 2019
D17, HTD1	LOC_Os04g46470	SL	Carotenoid cleavage dioxygenase 7 (CCD7)	Loss-of-function mutant	Reduction in panicle size and plant height; higher tillering	Ishikawa et al., 2005; Zou et al., 2006
D27	LOC_Os11g37650	SL and Auxin	β-carotene isomerase	Loss-of-function mutant	Reduction in panicle size and plant height; higher tillering	Ishikawa et al., 2005; Hao et al., 2009
D53	LOC_Os11g01330	SL	Class I Clp ATPase	Gain-of-function mutant	Reduction in panicle size and plant height; higher tillering	Zhou et al., 2013
DCL3a	LOC_Os01g68120	GA and BR	RNase III-class Dicer-like 3	Knockdown	Small panicles, reduction in number of secondary branches, larger flag leaf angle and reduced plant height	Wei et al., 2014
DEP1	LOC_Os09g26999	CK	Phosphatidylethanolamine-binding protein like domain protein	Gain-of-function mutant	Increase in number of grains per panicle and reduction in length of inflorescence internode	Huang et al., 2009
Dof12	LOC_Os03g07360	BR	Dof transcription factor	Overexpression	Reduction in panicle size, primary branching, secondary branching, grain yield, plant height; aberrant leaf phenotypes	Wu et al., 2015
DOG, SAP11	LOC_Os08g39450	GA	A20/AN1 zinc-finger protein	Overexpression	Shorter panicle exsertion; reduced plant height	Liu et al., 2011
DST	LOC_Os03g57240	CK	Zinc finger protein	Frameshift mutation	Increased panicle branching, grain number and grain yield	Li et al., 2013
EATB	LOC_Os09g28440	Ethylene and GA	AP2/ERF transcription factor	Overexpression	Shortened panicle internodes; increase in the number of panicles per plant and spikelets per panicle; reduced plant height	Qi et al., 2011

(Continued)

TABLE 1 | Continued

Gene/QTL	RGAP Locus ID	Hormone involved/ affected	Nature of protein encoded	Phenotype obtained for	Phenotype related to grain yield	References
ETR2	LOC_Os04g08740	Ethylene	Ethylene receptor	Overexpression	Reduction in number of effective panicles and seed-setting rate; erect panicles; delayed flowering and other multiple phenotypes	Hada et al., 2009
EUI1	LOC_Os05g40384	GA	Cytochrome P450 monooxygenase	Loss-of-function mutant	Increased panicle exsertion; elongated uppermost internode; taller plants	Luo et al., 2006
FZP	LOC_Os07g47330	Ethylene	ERF transcription factor	Mutant	Defect in spikelet development; presence of sequential round of branches instead of florets/spikelets	Komatsu et al., 2003
Gn1a*, *OsCKX2	LOC_Os01g10110	CK	Cytokinin oxidase/dehydrogenase	Loss-of-function mutant	Increase in number of grains per panicle and grain yield	Ashikari et al., 2005
GNP1, GA20ox1	LOC_Os03g63970	GA and CK	GA20-oxidase 1	Near-isogenic line NIL-*GNP1*^{TQ}	Increased total grain number per panicle; filled grain number per panicle and secondary branch number; slight increase in plant height, slight decrease in grain length, width and 1,000-grain weight	Wu et al., 2016b
GSN1*, *OsMKP1	LOC_Os05g02500	CK and BR	Mitogen-activated protein kinase	Overexpression	Increased number of seeds per panicle and reduced seed size	Guo et al., 2018
IPA1, WFP, OsSPL14	LOC_Os08g39890	CK and SL	SQUAMOSA promoter binding protein -like 14	Overexpression	Increase in number of panicle branches and grain yield	Jiao et al., 2010; Miura et al., 2010
LAX1	LOC_Os01g61480	Auxin	bHLH transcription factor	Mutant	Abnormal spikelet meristem development and panicle architecture	Komatsu et al., 2001
LOG	LOC_Os01g40630	CK	Cytokinin riboside 5′-monophosphate phosphoribohydro-lase	Mutant	Reduction in panicle size and abnormal panicle branching	Kurakawa et al., 2007
LP, EP3	LOC_Os02g15950	CK	F-box protein	Recessive mutant	Increased panicle size, panicle branching, grain number, and grain yield	Li et al., 2011
MADS57	LOC_Os02g49840	GA and SL	MADS-box transcription factor	Knockdown	Reduced panicle exsertion, internode elongation, and plant height	Chu et al., 2019
NAL1	LOC_Os04g52479	Auxin	Novel protein	Overexpression	Increased panicle branching and spikelets number per panicle	Fujita et al., 2013
ONAC096	LOC_Os07g04560	CK	NAC domain containing protein	T-DNA insertion mutant	Increased tillering, number of panicles, and grain yield	Kang et al., 2019
NOG1	LOC_Os01g54860	JA	enoyl-CoA hydratase/isomerase	Overexpression	Enhanced grain number per panicle	Huo et al., 2017
OsAHP1 and OsAHP2	LOC_Os08g44350 and LOC_Os09g39400	CK	His-containing phosphotransfer proteins	Knockdown (RNAi)	Reduction in panicle size and seed set	Sun et al., 2014
OsBZR1	LOC_Os07g39220	BR and SL	Transcription factor	Overexpression	Promotes tillering	Fang et al., 2020
OsCCA1	LOC_Os08g06110	SL	MYB transcription factor	Overexpression	Negatively regulates tillering and positively regulates panicle development	Wang F. et al., 2020
OsER1	LOC_Os06g10230	CK and BR	Receptor like kinase	Mutant	Increased number of spikelets per panicle	Guo et al., 2020

(Continued)

TABLE 1 | Continued

Gene/QTL	RGAP Locus ID	Hormone involved/ affected	Nature of protein encoded	Phenotype obtained for	Phenotype related to grain yield	References
OsJAZ1	LOC_Os04g55920	JA	Zinc finger protein	Dominant mutant	Defects in spikelet development and morphogenesis	Cai et al., 2014
OsMED14_1	LOC_Os08g24400	Auxin	Mediator subunit	Knockdown (RNAi)	Reduced panicle branching, lesser seed set and pleiotropic phenotypes	Malik et al., 2020
OsPEX5	LOC_Os08g39080	JA	PTS1 receptor protein	Loss-of-function mutant	Aberrations in spikelet morphology	You et al., 2019
OsPRR1	LOC_Os02g40510	SL	Pseudo-Response Regulator	Overexpression	Positively regulates tillering and negatively regulates panicle development	Wang F. et al., 2020
OsTB1, FC1	LOC_Os03g49880	SL	TCP family transcription factor	Overexpression	Negatively regulates tillering by inhibiting the outgrowth of axillary buds	Takeda et al., 2003; Minakuchi et al., 2010
OsVIL2	LOC_Os12g34850	CK	PHD domain containing protein	Overexpression	Larger panicles, Increased primary and secondary branching and grain number	Yang et al., 2019
PAY1	LOC_Os08g31470	Auxin	Peptidase	Dominant mutant	Larger panicles with more secondary branching; increase in grain number per panicle, grain yield per plant, plant height; less number of tillers	Zhao et al., 2015
PIN1	LOC_Os02g50960	Auxin	Auxin efflux transporter	Knockout mutant	*pin1cpin1d* double mutant shows pin-like inflorescences with no panicle	Li et al., 2019
PIN2	LOC_Os06g44970	Auxin	Auxin efflux transporter	Overexpression	Reduction in panicle length, number of grains per panicle, grain weight per panicle, plant height; increase in tillers angle, number etc.	Chen et al., 2012
PIN5b	LOC_Os08g41720	Auxin	Auxin efflux carrier-like protein	Overexpression	Reduction in panicle length, number of seeds per panicle, seed setting rate, tiller number, plant height etc.	Lu et al., 2015
PMM1, D11	LOC_Os04g39430	BR	Cytochrome P450 (CYP724B1)	Knockout mutant	Clustered primary branching and small grains	Li et al., 2018
PRE	LOC_Os03g55800	JA	Allene oxide synthase	Mutant	Defects in seed setting	Hibara et al., 2016
Prl5	LOC_Os05g34854	GA	GA20-oxidase 4	Overexpression	Elongation of panicle rachis and lower primary branches of panicle	Agata et al., 2020
SP3	LOC_Os03g55610	CK	Dof transcriptional activator	T-DNA insertion mutant/knockdown mutation	Reduction in plant size, panicle size, number of secondary branches and spikelets	Huang et al., 2019
TBP1, OsBAK1	LOC_Os08g07760	BR	SERK protein	Mutant	Increased tillering, panicle branching and seed number but reduced panicle length and seed size	Lin et al., 2017

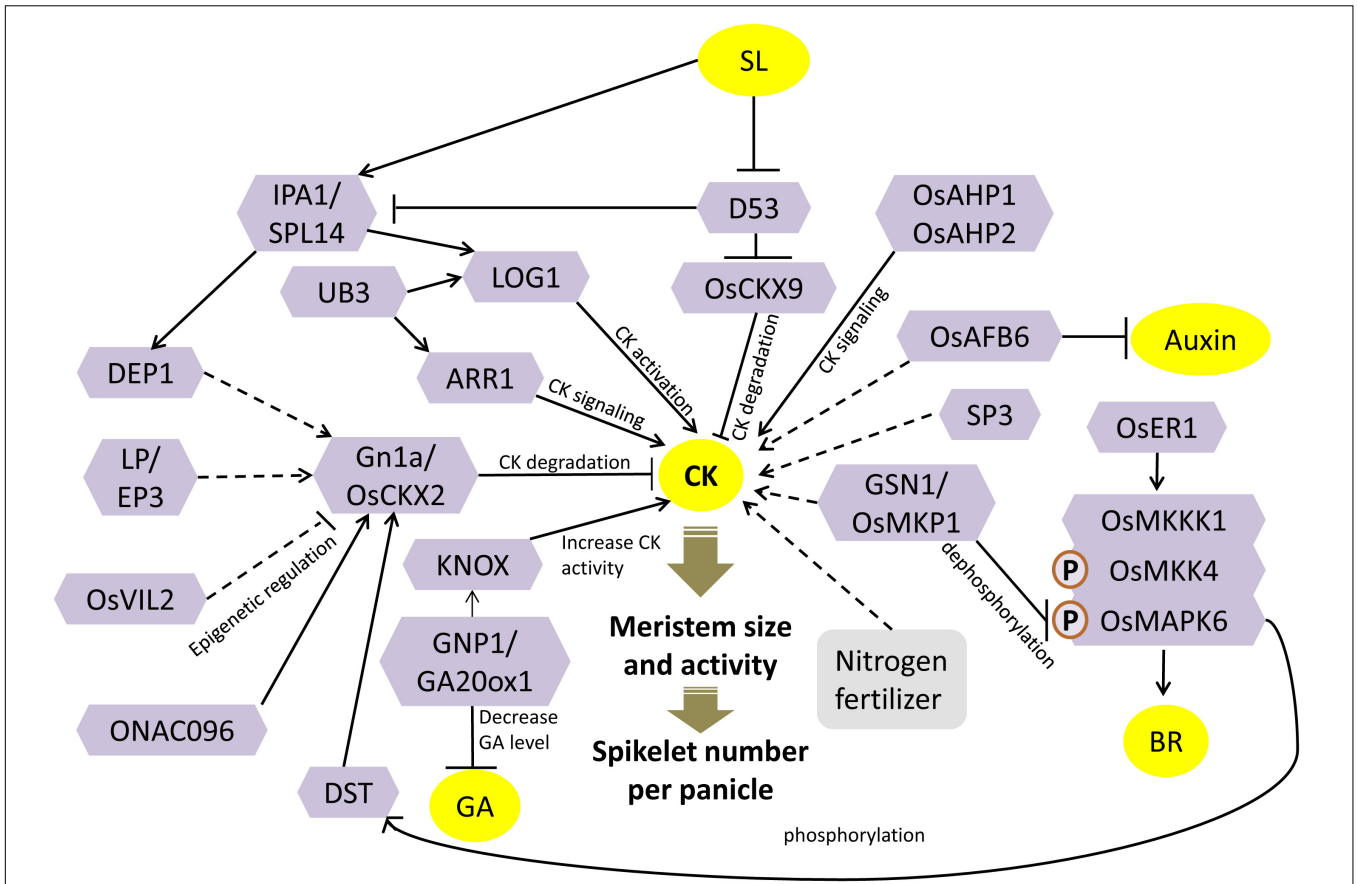


FIGURE 1 | Cytokinin (CK)-dependent meristem activity is central to panicle architecture. Enzymes like OsCKX2, OsCKX9, LOG1, and OsIPTs regulate the concentration of CK. These genes are further regulated by transcription factors, epigenetic regulators, and signaling factors. CK robustly interacts with other hormones, and a complex network determines the inflorescence meristem activity and hence panicle architecture. See the text for details. Solid lines represent direct regulation; dashed lines represent indirect regulation; arrows represent positive regulation; bar-headed lines represent negative regulation; block arrows represent the effect of a response. P represents phosphorylation.

The DST^{reg1} had a frameshift mutation that destroyed its C-terminal EAR motif. Although EAR motif is a well-known repression motif, but here, it showed a non-canonical activation activity, as its deletion destroyed the transcriptional activation activity of DST. Transcript levels of *OsCKX2*, along with other members of the *OsCKX* family, were significantly reduced in the mutant. Chromatin immunoprecipitation (ChIP) assay and electrophoretic mobility shift assay (EMSA) confirmed the binding of DST onto the promoter of *OsCKX2* (Li et al., 2013). DST has been shown to transactivate the OsCKX2 expression and hence positively regulate CK degradation and negatively regulate panicle branching and, consequently, grain number.

The transcription levels of *OsCKX2* were also severely downregulated in the young panicles of mutants of a kelch-repeat containing F-Box protein encoded by *Larger Panicle* (*LP*)/*Erect Panicle 3 (EP3)* (Piao et al., 2009; Li et al., 2011) (**Figure 1**). LP/EP3 subcellular localization in the endoplasmic reticulum, along with its molecular function as an E3 ubiquitin ligase subunit, proposed its involvement in ER-associated protein degradation. The *lp* mutants showed a significant increase in the size of the panicle, with a considerable increase in grain number and hence an 11% increase in the grain yield. The primary and secondary panicle branches were also significantly increased in the mutant that can be correlated to the *in situ* RNA accumulation of *LP* in the panicle branch primordial region. These phenotypes are similar to the *Gn1a* mutant and thus indicate that perhaps LP positively regulates the expression of *OsCKX2* by the degradation of some unknown proteins. Consequently, LP modulates the levels of CK by possibly indirectly controlling *OsCKX2* gene expression (Li et al., 2011). Although *EP3* and *LP* correspond to the same locus, their mutants showed different phenotypes. The absence of kelch motif in the *ep3* mutant does not increase grain yield; instead, it decreased grain yield and changed the panicle phenotype to a smaller and more compact one (Piao et al., 2009). The role of kelch motif in *LP*/*EP3* needs to be further investigated.

Cytokinin levels in the spikelet meristem are also synchronized by epigenetic regulation of *OsCKX2* gene expression by chromatin-modifying factors. OsVIL2 (*Oryza sativa* VIN3-LIKE 2) is a chromatin-modifying protein that contains a histone binding motif called plant homeodomain

(PHD) finger (Yang et al., 2013). Mutation in *OsVIL2* results in smaller plants with few grains per panicle. On the other hand, *OsVIL2* overexpression results in larger panicles with more primary and secondary branches and hence more grain number. The expression of *OsCKX2* was downregulated, whereas the CK levels were increased in $OsVIL2^{OX}$ plants. Enrichment of OsVIL2 in the transcript initiation region in the promoter of *OsCKX2*, and further enrichment of H3K27me3 around the same promoter region of *OsCKX2*, confirms the epigenetic repression of *OsCKX2* by OsVIL2 (**Figure 1**). Therefore, OsVIL2 represses *OsCKX2* gene expression by methylating its promoter and thereby increasing the CK accumulation and meristem activity (Yang et al., 2019).

OsCKX2 regulates the grain number by regulating both panicle branching and panicle number (tiller number) (Yeh et al., 2015). Mutation in *ONAC096*, a gene encoding NAC domain-containing transcription factor, results in a 16% increase in grain yield due to a 15% increase in the number of panicles (representation of increased tillering). The mutant *onac096* showed repression of *OsCKX2* gene expression, whereas the $ONAC096^{OX}$ showed an accumulation of *OsCKX2* transcript. Hence, *ONAC096* negatively regulates the panicle number, tillering, and CK levels by positively regulating *OsCKX2* gene expression (Kang et al., 2019) (**Figure 1**). Apart from panicle branching and shoot branching, CK has also been implicated in regulating panicle length. *Short Panicle 3* (*SP3*) encodes a DNA binding with one finger (Dof) transcription factor. A knockdown mutation created by T-DNA insertion in the promoter of *SP3* resulted in smaller plants with smaller panicle with significantly lesser number of secondary branches and spikelets. *SP3* is expressed at the branch primordia of the young panicles. The mutant *sp3* had an altered CK homeostasis wherein CK catabolism genes (four members of *OsCKX* gene family) were remarkably upregulated and CK biosynthesis genes (*OsIPT3* and *OsIPT7*) were drastically downregulated, thus resulting in an overall reduction in the CK levels (Huang et al., 2019) (**Figure 1**).

The activation of CK from its inactive forms, specifically in the meristematic cells, is another crucial metabolic step that regulates the maintenance of meristematic activity. *Lonely Guy* (*LOG*) encodes a novel enzyme called cytokinin riboside 5′-monophosphate phosphoribohydrolase that catalyzes the final step of direct CK bioactivation. It converts CK nucleotides into bioactive free base forms, specifically in the shoot meristem tips. A mutation in it causes a severe reduction in panicle size and abnormal panicle branching due to premature termination of the meristem (Kurakawa et al., 2007). Thus, this localized activation of CK by LOG in the shoot, branch, and spikelet meristems directly regulates grain number (**Figure 1**).

IPA1 (*IDEAL PLANT ARCHITECTURE*)/*WFP* (*Wealthy Farmer's Panicle*) is one of the most promising QTL identified for grain number enhancement (Jiao et al., 2010; Miura et al., 2010). The encoded protein, OsSPL14, is a member of the SQUAMOSA promoter binding protein-like (SPL). OsSPL14 has been shown to directly regulate CK biosynthesis by binding to the promoter of the CK-activating gene, *LOG* (Lu et al., 2013). Additionally, the ortholog of OsSPL14 in maize, UNBRANCHED3 (UB3), that is associated with kernel row number trait, is also shown to bind to the promoter of the rice *LOG* directly and regulates its expression. In addition, UB3 could also bind to the promoters of rice type-A response regulators (*ARRs*) that are regulators of CK signaling (Du et al., 2017). Further, it is known that OsSPL14 directly controls the expression of *DENSE AND ERECT PANICLE 1* (*DEP1*), a major regulator of grain number per panicle, by binding to its promoter (Lu et al., 2013). DEP1 encodes for a phosphatidylethanolamine-binding protein (PEBP)-like domain protein. A gain-of-function mutation, *dep1*, results in increased inflorescence meristem activity, decreased inflorescence internode length, and increased grains per panicle, to produce a dense and erect panicle. The expression of *OsCKX2* was evidently downregulated in NIL-*dep1* (Huang et al., 2009). Thus, *OsSPL14* regulates panicle branching and grain yield by regulating CK levels, by upregulating CK biosynthesis (*LOG* expression), and by downregulating the degradation of CK by positive regulation of *DEP1* and hence downregulation of *OsCKX2* (**Figure 1**).

The signaling response of CK is mediated by a relay of steps, including subsequent phosphorylation of CK receptors present on the membrane, to the histidine-containing phospho-transfer proteins (AHPs), followed by type-B response regulators in the nucleus. Disruption of CK signaling by simultaneous knockdown of *OsAHP1* and *OsAHP2* via RNA interference (RNAi) in rice plants showed a reduction in panicle size and low seed set along with other pleiotropic effects (Sun et al., 2014) (**Figure 1**).

Application of nitrogen fertilizers prior to panicle initiation is well known to increase grain yield by increasing the number of spikelets per panicle. Measurement of CK levels after nitrogen fertilizer application showed a local increment of CK levels in the panicles and no change in leaf and root. Analysis of expression profiles of CK metabolism genes highlighted the adenosine phosphate-isopentenyltransferase (*OsIPT*) gene family members that catalyze an initial rate-limiting step of CK biosynthesis. All the *OsIPT* genes showed considerable upregulation in the panicles by nitrogen application. Thus, localized CK accumulation in the panicle lateral meristems is the direct effect of nitrogen fertilizers that increases panicle branching (Ding et al., 2014) (**Figure 1**).

It seems that the fine tuning of bioactive CKs levels in the inflorescence meristem is a critical trait for engineering panicle architecture and grain number in rice. **Figure 1** illustrates the proposed model of CK-mediated regulation of grain number based on the discussed studies. Interactions of CK with other hormones are also depicted in **Figure 1**, and the details are discussed in section "Crosstalk of Phytohormones to Regulate Grain Number."

Auxins

Auxins regulate a plethora of responses both at the cellular and whole-plant level, imparting pleiotropic physiological effects (Thimann and Koepfli, 1935). The developmental module of a panicle involves the transition of the vegetative to the reproductive phase marked by the transformation of shoot apical meristem to axillary meristems and further its fate transition to spikelet meristems (Huijser and Schmid, 2011). Auxin has

a pivotal role in panicle development, as it is required for the initiation and maintenance of axillary meristems.

Auxin is produced mainly in growing shoot apices and is transported basipetally down the site along specific transport routes through polar transport machinery, and it indirectly inhibits the growth of axillary buds (Sieberer and Leyser, 2006; Zwiewka et al., 2019). Wild rice genotypes have short stature with thin stems, few grains, high tillering, and low yield. A gain-of-function mutation in wild rice introgression line YIL55 changed the plant architecture to that of an ideal crop with increased plant height, lesser number of tillers, thicker stems, larger panicles, and more secondary branches with a remarkable increase in grains per panicle as well as grain yield per plant (38%). The mutation was found in *PLANT ARCHITECTURE AND YIELD 1 (PAY1)* that encodes for a nuclear-localized peptidase and regulates plant architecture by affecting polar auxin transport and altering levels of endogenous indole 3-acetic acid (IAA). The *PAY1* mutation reduced the basipetal transport of IAA (**Figure 2**). This resulted in a trait wherein apical dominance was enhanced, leading to reduced tiller number. The introduction of the *PAY1* allele in cultivated high-yielding varieties further increased the grain number per panicle and grain yield per plant significantly (Zhao et al., 2015).

SPIKELET NUMBER (*SPIKE*) is an allele of *NARROW LEAF 1 (NAL1)* that affects panicle architecture pleiotropically to enhance the grain yield. *NAL1* encodes a novel protein that regulates polar auxin transport (Qi et al., 2008) (**Figure 2**). The overexpression of *NAL1* results in increased secondary panicle branching and greater number of spikelets per panicle. In addition, the source size and translocation capacity were increased with greater leaf area, robust root system, and vasculature, which contributed to better yield (Fujita et al., 2013).

The establishment of polar auxin transport is brought about by the asymmetrically localized PIN proteins that are instrumental in driving polar cell-to-cell transport of auxin. The PIN-FORMED (PIN) protein family is a group of auxin efflux transporters, having role in catalyzing transport of auxin from cells (Zazímalová et al., 2010). OsPIN5b is an endoplasmic reticulum localized protein that regulates cellular auxin homeostasis by facilitating conjugation-based auxin metabolism (Mravec et al., 2009; Barbez et al., 2012). Overexpression of *OsPIN5b* resulted in pleiotropic phenotypes with overall reduction in growth and more notable reduction in tiller number, biomass, seed setting rate, panicle length, number of seeds per panicle, and, thus, reduction in yield. On the other hand, RNAi plants with reduced expression of *OsPIN5b* presented longer panicles with more seeds. Auxin homeostasis is regulated by auxin conjugation with amino acids and sugars. The ratio of levels of auxin conjugates to that of free IAA was found to be low in $OsPIN5b^{OX}$ tissues. The enhanced level of free IAA in the overexpression lines positively correlated with higher transcript levels of *OsYUCCA1* (role in IAA biosynthesis).

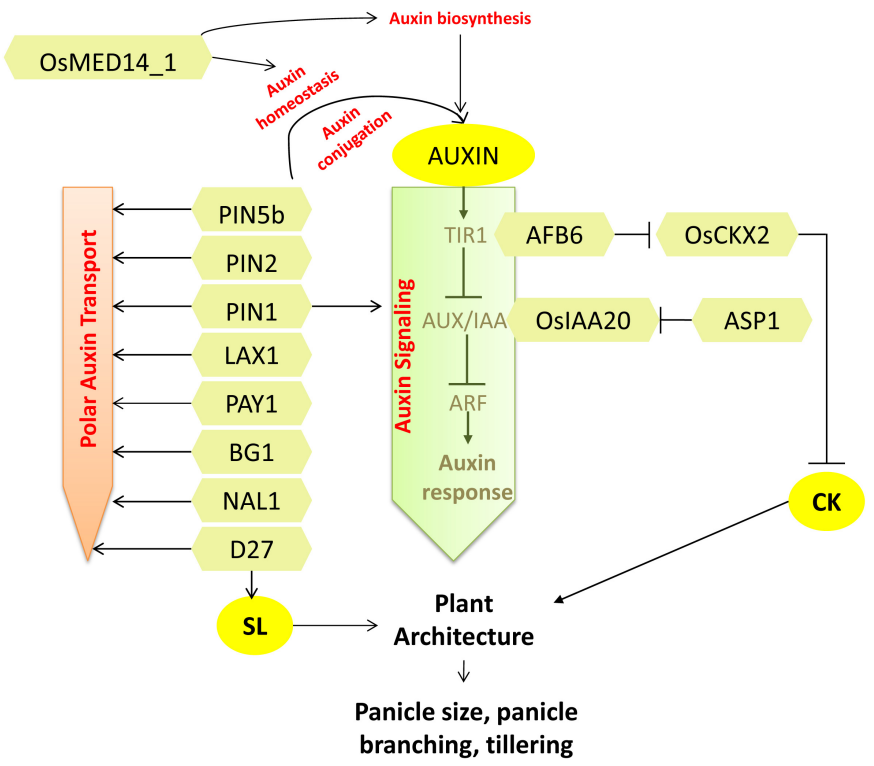


FIGURE 2 | Regulators of auxin signaling regulating plant architecture and affecting grain number. Factors affecting auxin biosynthesis, homeostasis, signaling, and transport identified in rice with functions known in affecting grain number are shown. Interconnections of auxin signaling with CK and SL also contribute to the plant architecture that affects grain number. Arrows represent positive regulation; bar-headed lines represent negative regulation.

Additionally, overexpression of OsPIN5b disrupted the polar auxin movement and auxin distribution. Conclusively, OsPIN5b is associated with auxin homeostasis, transport, and distribution to modulate plant architecture and plant yield (Lu et al., 2015) (**Figure 2**). Contrary to *OsPIN5b*, overexpression of another member of PIN family, *OsPIN2*, a functional auxin efflux transporter, showed phenotypes such as short plant height, a greater number of tillers, and larger tiller angle at the vegetative stage. At the reproductive stage, the plants exhibited short panicle length, lesser grains per panicle, lower grain length and breadth, and lower grain weight (Wang et al., 2009; Chen et al., 2012). Predominant auxin accumulation in root–shoot junction depicts its probable role in shoot architecture, but molecular mechanism describing its role in panicle development is yet to be revealed.

A recent report presents the functional divergence of four paralogous genes of *OsPIN1* (Li et al., 2019). Analysis of the mutants generated through CRISPR-Cas9 showed that the *pin1apin1b* double mutant had a role in determining root development, plant height, and tiller angle, whereas the *pin1cpin1d* double mutant developed abnormal, naked, pin-like inflorescences at the flowering stage as exactly shown by *Arabidopsis pin1* mutants (Gälweiler et al., 1998). However, the single mutants did not show any developmental abnormalities. Further investigation at the maturation stage revealed that *pin1cpin1d* mutants completely lost their secondary branches and spikelets, and thus, no panicle was formed. Furthermore, relatively low transcript levels of *OsYUCCAs*, *OsARF*, and *OsIAA* genes in these mutants define the functional role of *OsPIN1* in panicle growth and development through affecting auxin biosynthesis and signaling (Xu et al., 2005; Li et al., 2019) (**Figure 2**). Knockdown of a Mediator subunit protein encoding gene, *OsMED14_1*, showed various pleiotropic effects, including reduced panicle branching and lesser seed set. Evident alterations of auxin levels and transcript levels of auxin homeostasis genes in the *OsMED14_1*^{RNAi} plants suggested the role of auxins in *OsMED14_1*-mediated regulation of grain number (Malik et al., 2020). The *BIG GRAIN 1* (*BG1*) is another gene, which encodes a plasma membrane-localized protein that is shown to be a positive regulator of grain size. Plants with suitable overexpressed levels of *BG1* showed increased plant height, longer leaves, larger panicles, and increased grain size by regulating around 50% increased basipetal auxin distribution and transport (Liu L. et al., 2015) (**Figure 2**).

BARREN STALK1 (*BA1*) is a maize gene encoding a non-canonical bHLH transcription factor that is a determinant of inflorescence patterning. *BA1* regulates axillary meristem formation by generating an auxin response maxima via polar auxin transport that flanks the primordial inflorescence meristem (Gallavotti et al., 2004, 2008). *LAX PANICLE 1* (*LAX1*), the rice homolog of *BA1*, is also shown to regulate the initiation and maintenance of the axillary meristem. Accordingly, the *lax1* panicles were highly abnormal with an absolute absence of lateral spikelet but the presence of a terminal spikelet (Komatsu et al., 2001). An auxin maxima is required for both reproductive and vegetative axillary meristem in rice and maize (Oikawa and Kyozuka, 2009) in contrast to *Arabidopsis*, where an auxin minima is needed for vegetative axillary meristem initiation (Wang Q. et al., 2014; Wang Y. et al., 2014). LAX1 is shown to interact with LAX PANICLE2 (LAX2, encoding a novel nuclear protein), and their double mutant shows severe phenotype in reproductive and vegetative axillary meristem development (Tabuchi et al., 2011). Thus, *LAX1* functions in tiller and inflorescence meristem initiation independently or together with *LAX2* via regulating auxin signaling and transport.

The homolog of *Arabidopsis TOPLESS (TPL)* gene in rice is *ABERRANT SPIKELET AND PANICLE 1* (*ASP1*) that encodes a TPL-like transcriptional corepressor and regulates the panicle morphology. The *asp1* mutant displayed pleiotropic morphological abnormalities with around 80% reduction in rachis length with shorter primary branches and a reduction in the number of normal spikelets. In addition, the spikelet morphology was also severely affected. These characteristic phenotypes are due to aberrations in the fate of reproductive meristems, i.e., *asp1* fails to fine tune the proper initiation and maintenance of inflorescence meristem, branch meristem, and spikelet meristem in a stage-specific and time-dependent manner. On investigating the role of *asp1* in response to auxin, a marked upregulation in the expression of *OsIAA20* (a marker of auxin-dependent gene induction) in comparison to wild type was observed, suggesting that the auxin signaling was disrupted. Thus, *ASP1* upholds the auxin signaling by forming a repressor negative feedback complex with *OsIAA20* and thereby regulating panicle morphology (Yoshida et al., 2012) (**Figure 2**).

As discussed above, the combined effects of the core components of auxin machinery greatly affect the plant as well as panicle architecture. However, exploring the complexity of hormonal crosstalks will open up a broader network explaining each landmark events of panicle development (**Figure 2**).

Strigolactones

Strigolactones (SLs) are newly discovered carotenoid-derived plant hormones that play an inhibitory role in shoot branching in diverse species (Dun et al., 2009). In rice, mutant screening has revealed the involvement of various genes in SL biosynthesis and signaling pathways. Several of them control rice branching, which includes both shoot (tiller) and panicle branching (Kebrom et al., 2013). These mutants are marked by their short stature along with high tillering features. Hence, the genes involved in these mutants, i.e., *d3*, *d10*, *d14*, *d17* (or *htd1*), and *d27*, negatively regulate tiller bud activity in rice. Thus, they are named after their conspicuous dwarf phenotype as *DWARF* (*D*). Loss-of-function mutants of these genes show a subtle effect on panicle size and primary panicle branching. These mutants possess small panicles as compared to their respective wild types (Ishikawa et al., 2005; Zou et al., 2006). *D17*, *D10*, and *D27* genes are involved in the SL biosynthesis pathway, while *D3* and *D14* play an important role in SL signaling. *D17*/*HTD1 (HIGH-TILLERING DWARF1)* and *D10* encode CCD7 (CAROTENOID CLEAVAGE DIOXYGENASE 7) and CCD8, respectively (Arite et al., 2007; Umehara et al., 2008) (**Figure 3**). *D27* encodes a chloroplast localized iron-containing β -carotene isomerase enzyme that converts all *trans*-β-carotene into 9-*cis*-β-carotene, which is sequentially cleaved by the action of CCD7 and CCD8 to give SL precursors (Hao et al., 2009; Alder et al., 2012) (**Figure 3**).

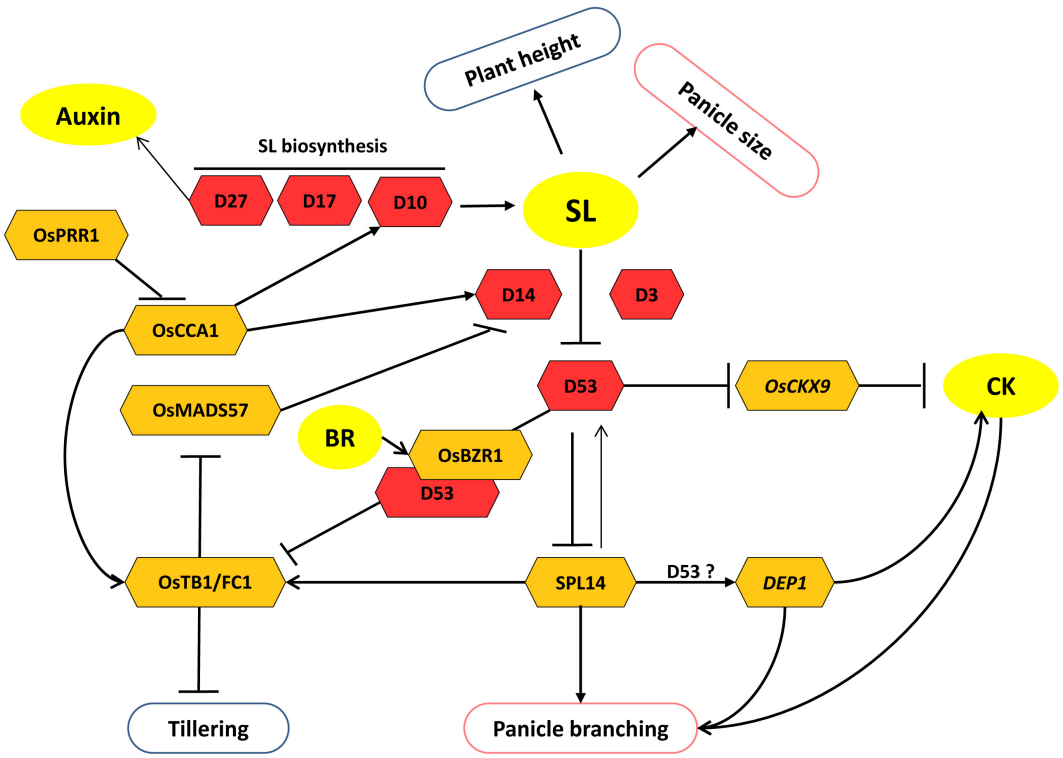


FIGURE 3 | Model of strigolactone (SL) regulated pathways specifically involved in the regulation of panicle and tiller development. SL inhibits tillering but promotes panicle size and branching. Crosstalk of SL with BR and CK regulates tillering and panicle branching, respectively. Core components of SL biosynthesis and signaling are shown as red hexagons. Other interacting factors are shown as yellow hexagons. Arrows represent positive regulation; bar-headed lines represent negative regulation.

D14/*D88*/*HTD2* encodes for the receptor for SL perception (Beveridge and Kyozuka, 2010) and negatively regulates the tiller bud outgrowth. Its loss of function results in a significant reduction in panicle size along with higher tillering, reduced height, and smaller seed size (Arite et al., 2009; Gao et al., 2009; Liu et al., 2009). Another loss-of-function allelic mutant of *D14*, *htd4*, bears a large number of effective panicles per plant. However, traits like panicle length, number of grains per panicle, primary branch numbers, seed-setting rate, and 1,000-grain weight are decreased in *htd4* plants (Wang et al., 2017). Recently, a new loss-of-function allele of *D14*, i.e., *dhta-34*, has been identified, whose plants also show similar phenotypes (Liang et al., 2019). Transcriptional regulation of *D14* has been shown to be mediated by OsMADS57. OsMADS57 negatively regulates *D14* expression and thus disrupts SL signaling and promotes tillering. However, its availability to regulate *D14* is limited by interaction with OsTB1 (TEOSINTE BRANCHED1) (Guo et al., 2013) (**Figure 3**). Although the role of *OsMADS57* in controlling panicle branching has not been observed, it affects panicle exsertion by regulating GA-deactivating genes (details in section "Gibberellic Acid"). *OsTB1*, also known as *FINE CULM1* (*FC1*), negatively regulates tillering in rice. It functions downstream of SL signaling and suppresses tillering by inhibiting the outgrowth of axillary buds (Minakuchi et al., 2010).

Other components involved in SL signaling are D3 and D53. D3 is an F-box protein that can make a complex with D14 (receptor) and D53 proteins. D53 is a molecular target of SL-induced degradation by a proteasome–ubiquitin pathway in a D14- and D3-dependent manner (Zhou et al., 2013) (**Figure 3**). It belongs to class I Clp ATPase proteins and contains a repression motif (EAR motif). In addition, its gain-of-function mutation results in small panicles, exaggerated tillering, and short stature. Thus, it acts as a repressor of the SL signaling pathway. Studies have suggested the involvement of D53 in the regulation of important panicle architecture governing genes. Like, in the absence of SL, D53 interacts with transcription factor SPL14/IPA1 (a key positive regulator of panicle branching) and corepressor TPL/TPR and inhibits IPA1 transcriptional activity. In the presence of SL, D53 undergoes proteasomal degradation, resulting in the release of repression of IPA1-regulated gene expression (Jiang et al., 2013; Song et al., 2017). Moreover, a negative feedback regulation exists because IPA1 directly binds to the promoter of *D53* and promotes its expression (**Figure 3**). *IPA1* expresses primarily during young panicle development. SPL14 promotes primary branch number in rice panicles that can result in increased grain productivity. SPL14 also promotes secondary branching in panicles (Xie et al., 2006; Jiao et al., 2010; Miura et al., 2010). Reports also indicate that SPL14 positively regulates the expression of *DEP1*, another important gene that determines panicle architecture and grain yield (Huang et al., 2009; Lu et al., 2013). However, it is still a matter of investigation whether D53 is also involved in the regulation of *DEP1* via SPL14 (**Figure 3**).

Recently, D53 has also been shown to inhibit the expression of *OsTB1* gene. This suggests a possible way to regulate tillering negatively by SL signaling. SL perception by D14 causes D53 degradation and thereby promotes the expression of *FC1*, which negatively regulates tillering in rice (Fang et al., 2020). Besides, *OsTB1* also facilitates SL signaling by promoting the expression of *D14* by limiting the availability of its negative regulator, i.e., OsMADS57, by interacting with it (Guo et al., 2013). Moreover, SPL14 also regulates tillering through *OsTB1* in rice. It positively regulates the expression of *OsTB1* and suppresses tillering (Lu et al., 2013) (**Figure 3**).

A new finding has revealed the role of circadian clock in the regulation of traits like tillering and panicle development through the SL pathway (Strable, 2020; Wang F. et al., 2020). Two important clock regulators in rice, *CIRCADIAN CLOCK ASSOCIATED1* (*OsCCA1*) and *PSEUDORESPONSE REGULATOR1* (*OsPRR1*), are antagonistic factors of the clock component. *OsCCA1* mediates the rhythmic expression of the clock and output genes during plant growth and development, whereas *OsPRR1* negatively regulates the expression of *OsCCA1* (Wang F. et al., 2020). The study showed that increment and reduction in *OsCCA1* expression in transgenic plants decreased and increased the axillary tiller bud formation, respectively. Further, $OsCCA1^{ox}$ plants had increased panicle size. Reverse phenotypes were observed for *OsPRR1* expression modified plants. The authors showed that OsCCA1 protein directly binds to the promoter of *D10*, *D14*, *OsTB1*, and *IPA1* and promotes their expression. D10 and D14 being SL biosynthetic enzyme and SL receptor, respectively, are effectors of SL concentration and perception. Thus, *OsCCA1* affects the SL pathway at both SL biosynthesis and signaling levels and negatively regulates tillering by repressing tiller-bud outgrowth and positively regulates panicle development by directly promoting the expression of *IPA1* (Wang F. et al., 2020) (**Figure 3**).

The exclusive role of SL in the regulation of panicle architecture has not been reported. Still, key regulators like SPL14 and DEP1, which control panicle branching and, thus, affect grain yield, seem to be regulated through SL signaling. Furthermore, the involvement of the SL pathway in the suppression of tiller development has been established. OsTB1 appears to be an integrator downstream of SL signaling that affects the suppression of tiller development. Taken together, it is evident that the SL pathway regulates both yield-determining traits, i.e., tiller and panicle development (**Figure 3**).

Gibberellic Acid

Gibberellic acid (GA) is a well-known class of phytohormones involved in the regulation of various processes of plant growth and development. The outcomes of manipulations of GA levels or responses with respect to stem elongation have substantiated its potential for generating high-yielding cultivars in cereal crops. However, GA has also been reported to regulate other yield-related traits like panicle exsertion and panicle branching (Gao and Chu, 2020).

Studies have suggested that panicle exsertion length (PEL) is regulated in a GA-dependent manner. PEL is the length of the peduncle that emerges from the flag leaf sheath. In some rice varieties, shorter PEL leads to a situation of panicle enclosure, i.e., panicles are partly or fully enclosed within the flag leaf sheath. It is mainly caused by the shortening of the uppermost internode (UI). Panicle enclosure is usually a problem associated with the cytoplasmic male sterile (CMS) lines used in hybrid rice seed production. It blocks normal pollination in hybrid rice and thus reduces seed production. It has been observed that panicles of CMS lines are deficient in IAA levels, which causes downregulation of GA biosynthesis gene *OsGA3ox2*. This results in low levels of bioactive GA (GA_1) in the UI leading to a reduction in the cell number and cell elongation. Therefore, UI is not long enough to push panicle out of flag leaf sheath, which leads to the panicle enclosure (Yin et al., 2007). Reverse phenotypes of increased panicle exsertion and UI elongation have been observed for loss-of-function *eui1* mutants. *EUI1* (*ELONGATED UPPERMOST INTERNODE1*) preferentially expresses in young panicles and encodes for a GA-deactivating enzyme (Luo et al., 2006; Zhu et al., 2006). Consequently, a higher bioactive GA level in the UI of *eui1* plants results in increased panicle exsertion. A recent report indicates that the expression of *EUI1* along with the expression of other GA deactivating gene *OsGA2ox3* (encodes GA 2-oxidase) is negatively regulated by the OsMADS57 transcription factor. Thereby, knockdown plants of *OsMADS57* contain low levels of bioactive GA due to elevated expression of *OsGA2ox3* and *EUI*, resulting in severe panicle enclosure and semi-dwarf phenotype (Chu et al., 2019). Another gene, *OsDOG* (*DWARF RICE WITH OVEREXPRESSION OF GIBBERELLIN-INDUCED GENE*), has been identified with a role in panicle exsertion, as its overexpression leads to a shorter PEL along with a dwarf phenotype. *OsDOG* (*or OsSAP11*) encodes A20/AN1 zinc-finger protein, which enhances the expression of GA catabolism related genes, whereas it downregulates the expression of GA biosynthesis gene (*GA3ox2*). Thus, unlike *MADS57*, *OsDOG* negatively regulates GA-mediated cell elongation and thus affects panicle exsertion (Vij and Tyagi, 2006; Liu et al., 2011) (**Figure 4**).

Gibberellic acid also affects other panicle-associated traits like panicle length, primary branching, secondary branching, and grain number per panicle. *OsCYP71D8L*, encoding a cytochrome P450 monooxygenase, controls panicle-related traits by affecting GA homeostasis. Gain-of-function of *OsCYP71D8L* leads to reduced panicle length, reduced grain number per panicle, and dwarfed plants (Zhou et al., 2020) (**Figure 4**). In a latest study, a gene, *Prl5* (*PANICLE RACHIS LENGTH5*), has been identified by QTL analysis by making a cross between two rice cultivars ST-1 and Koshihikari, which have distinct difference in panicle length. *Prl5* expresses mainly in young panicles and encodes a gibberellin biosynthesis enzyme, OsGA20ox4. It exclusively affects panicle architecture, as it promotes panicle rachis elongation and lowers primary branch elongation. Its expression in the vascular bundles, located near the lower primary branch meristem of young panicles, results in more accumulation of bioactive GA forms. This results in the elongation of both panicle rachis and lower primary branches of the panicle (Agata et al., 2020) (**Figure 4**).

These evidence clearly indicate the involvement of GA in the regulation of several panicle-related traits and show its potential

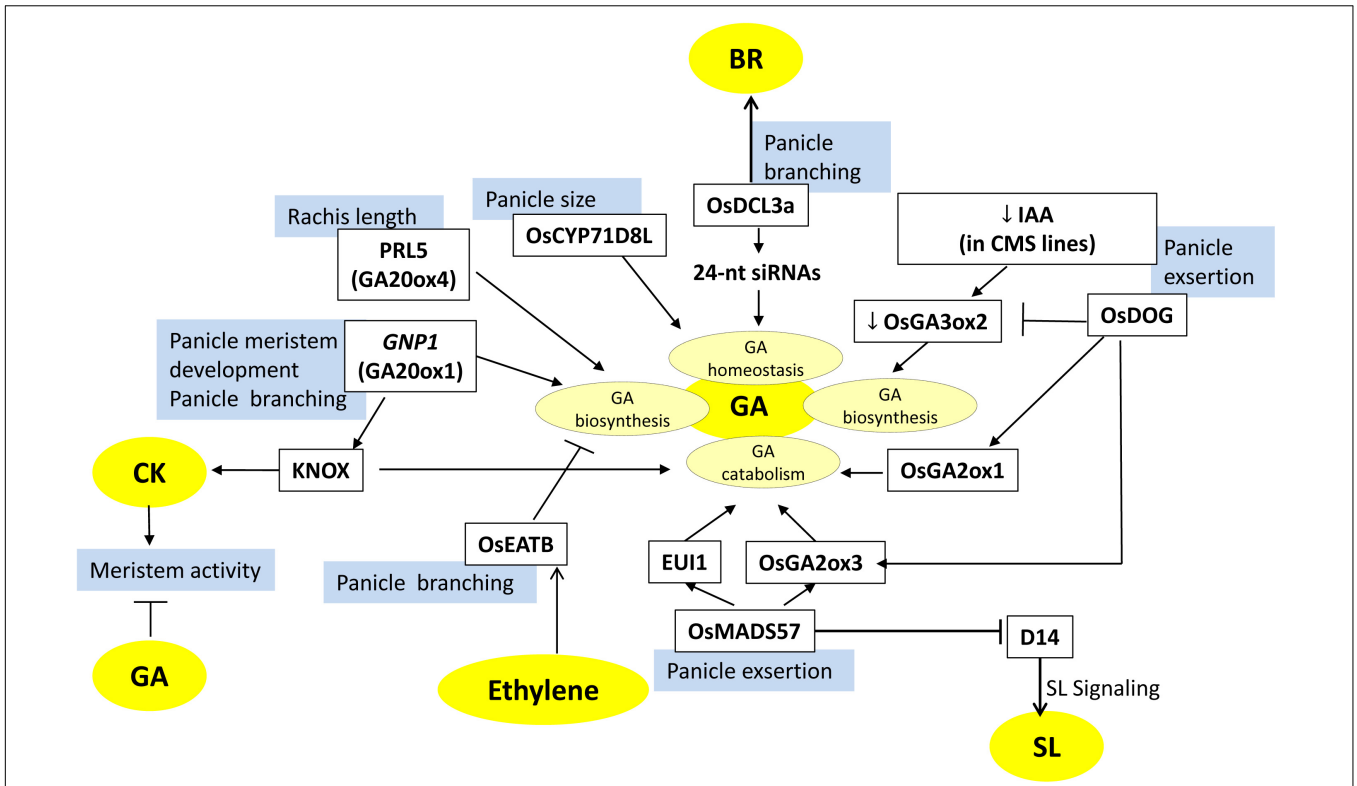


FIGURE 4 | Factors involved in gibberellin (GA)-mediated regulation of panicle architecture in rice. The model depicts the factors (white boxes) that are involved in GA biosynthesis, catabolism, or homeostasis and influence panicle development and morphology. Blue boxes represent the regulated panicle phenotypes. Arrows represent positive regulation; bar-headed lines represent negative regulation. Cytokinin (CK) and GA have antagonistic roles in regulating reproductive meristem activity. GA modulating factors also interact with other hormones (BR, SL, ethylene) mediated factors to make a complex meshwork.

to improve valuable agronomic traits in rice governing grain number per plant.

Brassinosteroids

Brassinosteroids (BRs) are a class of novel, naturally occurring, plant-specific steroidal hormones, which are featured by their polyhydroxylated sterol structure and play critical roles in mediating multiple biological processes like development and stress response (Yang et al., 2011; Saini et al., 2015).

Two BR biosynthetic genes, *DWARF4 (D4) and DWARF11* (*D11*), encoding cytochrome P450 (CYP90B2 and CYP724B1, respectively) with redundant functions are implicated in regulating plant architecture (Tanabe et al., 2005; Sakamoto et al., 2006a). *PANICLE MORPHOLOGY MUTANT 1 (PMM1)*, *GRAIN NUMBER AND SIZE ON CHROMOSOME 4* (*GNS4*), *CLUSTERED PRIMARY BRANCH 1* (*CPB1*), and *NOTCHED BELLY GRAIN 4* (*NBG4*) are all different alleles of *D11* (Tanabe et al., 2005; Wu et al., 2016a; Zhou et al., 2017; Li et al., 2018; Tong et al., 2018). Gene insertional mutant library screening resulted in the identification of a mutant *pmm1-1*, which showed morphological defects of clustered branch phenotype, i.e., each panicle branch clustered with two to three abnormal spikelets (Li et al., 2018). Enhancing the expression of *D4* (role in C-22 hydroxylation, a rate-limiting step in BR biosynthesis) in the *pmm1-1* background rescued the abnormal inflorescence phenotype of *pmm1-1* mutant, indicating that BR deficit in the mutant was complemented by higher expression of *D4* (Sakamoto et al., 2006a; Li et al., 2018). Preferential expression of *PMM1* in developing young panicles, specifically in branches and spikelet primordia, established the role of *PMM1* in determining inflorescence architecture.

Mutant screening identified another regulator of grain size, grain number, and grain yield named, *SMALL GRAIN 11* (*SMG11*), which is a novel allele of *DWARF2* and encodes a cytochrome P450 (CYP90D2) involved in BR biosynthesis (**Figure 5**). The morphological traits of the mutant, *smg11*, include erect, shorter, and denser panicles at the mature stage, which is due to decreased length of rachis and an increase in the number of secondary branches in addition to smaller grains. Overexpression of *SMG11* using *ACTIN* promoter produced different lines with different levels of *SMG11* expression. Those with very high *SMG11* expression had a reduction in yield due to large and heavy seeds but reduced panicle branching. However, in lines with only a little increase in *SMG11* expression, plants had greater yield due to an increase in grain size and normal panicle branching. Thus, the levels of BR as a function of SMG11 can be optimized to improve the grain yield (Fang et al., 2016).

Moreover, *OsBAK1* (*BRI1-ASSOCIATED RECEPTOR KINASE*)/*TBP1* (*TOP BENDING PANICLE 1*), encoding a somatic embryogenesis receptor kinase (SERK)

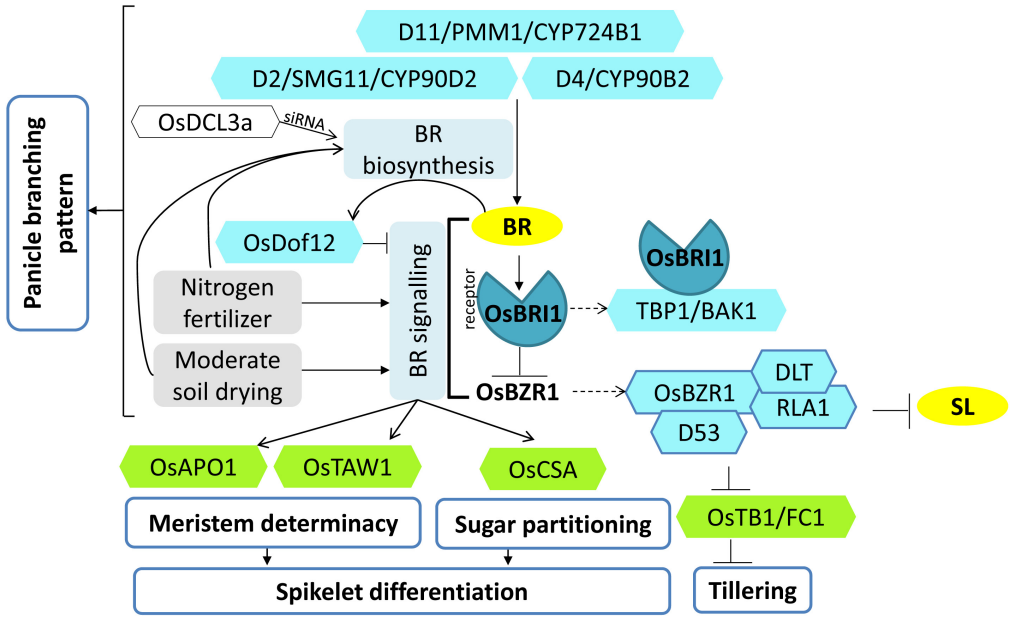


FIGURE 5 | Model of brassinosteroid (BR) regulation controlling panicle development, panicle architecture, and tillering. Various factors (blue hexagons) regulate BR biosynthesis and signaling. Their coordinated actions regulate the downstream targets (green hexagons) of BR signaling that control various aspects of spikelet differentiation, panicle development, and tillering. Nitrogen and moderate soil drying conditions serve as additional factors affecting panicle development by increasing BR levels and signaling. Arrows represent positive regulation; bar-headed lines represent negative regulation; dotted arrows represent the formation of protein complex.

domain-containing protein involved in BR signaling also regulates grain number and grain size (Lin et al., 2017; Yuan et al., 2017). The *TBP1* mutation resulted in an increased seed number due to increased tillering and greater primary and secondary panicle branching. However, the panicle length and the seed size were reduced; hence, an overall reduction in yield was observed. TBP1 interaction with the BR receptor (OsBRI1) was reduced in the *tbp1* mutant (**Figure 5**). Only a limited overexpression of TBP1 increased the yield by obtaining a balance between grain size and grain number; on the other hand, greater overexpression reduced the yield drastically (Lin et al., 2017). Thus, BR signaling components cannot be altered to greater levels, but careful fine tuning can help in higher yields.

A Dof transcription factor, OsDof12, when overexpressed in rice, displayed altered plant architecture with reduced plant height and noticeable reduction in the number of primary branches, secondary branches, number of spikelets per panicle, and overall size of the panicle. BR levels were found to be the same in the wild type and the overexpressor. Thus, the BR metabolic genes remain unaffected; however, the transcript levels of two BR signaling genes, *OsBRI1* and *OsBZR1*, were downregulated. Further, *Dof12* itself gets upregulated by BR treatment (Wu et al., 2015). Hence, *OsDof12* acts as a negative regulator of BR signaling in rice wherein it is a BR responsive gene (**Figure 5**). Thus, OsDof12 may be regulating the BR signaling homeostasis to regulate panicle architecture via a negative regulatory feedback loop.

The final panicle size is contributed by spikelet differentiation and spikelet degeneration. The phenomenon of young spikelet degeneration at the basal region of the panicle, often referred to as "preflowering floret abortion," causes severe yield loss in rice (Kato et al., 2008). The role of BR regulating spikelet abortion in young rice panicles has been recently studied wherein moderate soil drying treatments (soil water potential of -10 to -15 kPa) during panicle development enhanced BR biosynthesis in panicles. The higher expression level of *D11* led to increased BR contents [24-epiCS (24-epicastasterone) and 28-homoBL (28-homobrassinolide)] and ascorbic acid content in moderate drying-treated plants as compared to severe drying (soil water potential of -30 to -35 kPa) treated plants. Enhanced BR concentration suppressed the spikelet degeneration and enhanced spikelet differentiation, thus resulting in more spikelets per panicle. In addition, elevated expression of major determinants of spikelet meristem specification (*OsTAW1 and OsAPO2*), and sugar partitioning (*OsCSA*, a MYB domain protein), supports the notion that BR levels in young rice panicles promote spikelet differentiation condition by enhanced meristem activity as well as non-structural carbohydrate partitioning in moderate soil-drying treatment (Zhang et al., 2019a). Further, another study emphasized the role of BR in nitrogen-fertilizer-mediated enhancement of rice spikelet differentiation. Rice grown in different nitrogen fertilizer treatment, when supplied with exogenous BRs at spikelet primordium differentiation stage, resulted in elevated levels of endogenous BRs along with an upregulated expression of genes participating in BR biosynthesis (*OsD2*, *OsD11*) and BR signaling cascade (*OsBRI1*, *OsBAK1*) (**Figure 5**). These changes positively correlated with high H^+ATPase activity, high ATP

concentration, and high energy charge in panicles. These all together promoted spikelet differentiation and reduced spikelet degeneration (Zhang et al., 2019b).

The magnitude of actual crop productivity is determined by the level of spikelet sterility in rice. Hence, manipulating genes regulating BR biosynthesis and signaling during spikelet development will be a feasible approach for increasing grain yield (**Figure 5**).

Ethylene

Ethylene is a gaseous phytohormone known to be coordinating a vast array of developmental processes in plants. The dynamics of ethylene action is due to the equilibrium between its biosynthesis and its perception. In terms of contribution to important agronomic traits in rice, ethylene has a role in the regulation of panicle architecture, grain size control, and grain filling rate (Yin et al., 2017).

Rice panicle has asynchronous grain filling, i.e., the degree and rate of grain filling in individual spikelets depend on its position in the panicle (Mohapatra and Sahu, 1991). The superior spikelets located in apical branches flower early and fill faster to produce larger and heavy grains. The inferior spikelets present in proximal branches fill slowly and lack quality due to poor filling, making it unsuitable for human consumption. The poor grain filling can be attributed to the higher evolution rate of ethylene in inferior spikelets. These spikelets show significantly higher expression of *S*-adenosyl methionine synthase (SAM-synthase), an enzyme that produces *S*-adenosyl methionine (precursor of ethylene biosynthesis) from methionine. This high evolution rate of ethylene greatly affects the rates of cell division in the endosperm and, thus, the grain filling rate. Application of 1-methylcyclopropene (1-MCP), a potent ethylene action inhibitor, shows a pronounced inhibitory effect on inferior spikelets. 1-MCP treatment enhanced endosperm growth at the mid-grain filling stage by facilitating cytokinesis and endo-reduplication, through enhancing the expression of several cell cycle regulators like cyclins, cyclin-dependent kinase (CDK), and cyclin-dependent kinase inhibitor (CKI) (Panda et al., 2016).

The biosynthetic pathways of ethylene and polyamines (PAs) are interrelated through a key branch point intermediate, SAM (**Figure 6**). An increase in PA biosynthesis through SAM decarboxylase (SAMDC) activity greatly affects the ethylene biosynthesis rate (Walden et al., 1997). The potential metabolic interaction or competition between the two affects the spikelet development under various degrees of soil drought. Moderate drought at the early endosperm cell division stage and grain filling stage significantly increases the free PAs (spermidine and spermine) levels, which lead to a decrease in ACC (1-aminocylo-propane-1-carboxylic acid, ethylene precursor) content as well as ethylene evolution rate (Zhang W. et al., 2017). Moreover, the application of PAs to young panicles also affected the activity of various enzymes (sucrose synthase, ADP glucose pyrophosphorylase, soluble starch synthase) involved in sucrose–starch metabolic pathway. This, in turn, led to a higher cell division rate in the endosperm, a higher percentage of filled grains, and increased grain yield (Yang et al., 2008). Ethylene also negatively regulates postanthesis spikelet development and grain filling in coordination with abscisic acid (ABA). The lower ratio of ABA to ethylene and ACC contents in inferior spikelets directly correlates with the rate of cell division and grain filling rate and hence grain weight (Yang et al., 2006).

Transgenic and mutant studies have also conveyed the role of ethylene in grain filling. Overexpression of *ETHYLENE RESPONSE2* (*ETR2)*, an ethylene receptor, showed a reduction in the effective number of panicles, lowered seed setting rate, delayed floral transition, and enhanced starch accumulation in internodes. Further, *etr2* mutants and *ETR*-RNAi plants had considerably higher 1,000-grain weight due to better grain filling. Upregulation of flowering-related genes, *GIGANTEA (OsGI)* and *TERMINAL FLOWER1/CENTRORADIALIS homolog (RCN1)*, in $ETR2^{Ox}$ delayed the floral transition, whereas downregulation of *RAMY3D* (α-amylase gene) blocked sugar translocation from stems to grains leading to low seed setting rate and hence low yield (Hada et al., 2009) (**Figure 6**).

Ethylene response factors (ERFs) are transcription factors that are regulators of ethylene signaling and response. *FRIZZY PANICLE (FZP)*, an ERF transcription factor, is a rice ortholog of *BRANCHED SILKLESS1 (BD1)* in maize, having a key role in meristem transition and represses axillary meristem formation. The *fzp* mutant is characterized by the presence of sequential rounds of branching instead of florets. *FZP* is required to maintain the transition from spikelet meristem to floral meristem and to suppress the continued formation of axillary meristems from differentiated spikelet meristems (Komatsu et al., 2003). *FZP* regulates panicle branching by negatively regulating *APO2*/*RFL*, the gene responsible for regulating inflorescence meristem, and positively regulating several floral organ identity genes, especially the subset of B class (*OsMADS6*, *OsMADS17*) and E class (*OsMADS1*, *OsMADS7*, *OsMADS8*) (Bai et al., 2016). Recently, QTL *qSrn7* was identified, which is an allele of *FZP*. An increase in higher-order branching, especially in the upper regions or upper primary branches of panicles, was seen in lines carrying *qSrn7/FZP*. Thus, FZP regulates a unique branching pattern in rice, which is probably due to suppression of transition from branch meristem to spikelet meristem (Fujishiro et al., 2018) (**Figure 6**).

Investigating the changing concentrations of ethylene and dissecting the components of ethylene signaling machinery will be helpful in understanding the regulation of heterogeneous panicle architecture.

Jasmonic Acids

A healthy spikelet formation not only determines the reproductive success in cereal crops but is also essential for obtaining better grain yield. Jasmonic acids (JAs) play a key role in this process, as several mutant studies have revealed the role of JA in spikelet morphogenesis. The mutants like *extra glume 1*, *eg1* (a recessive mutant of JA biosynthesis-related gene), and *eg2* (a dominant mutant of *OsJAZ1*) show defects in spikelet formation with increased glume like structures, altered floral organ identity, and floral organ number along with defects in floral meristem determinacy. JA signaling specifies these processes by regulating the E-class gene, *OsMADS1*, which is known to play a role in spikelet development (Cai et al., 2014).

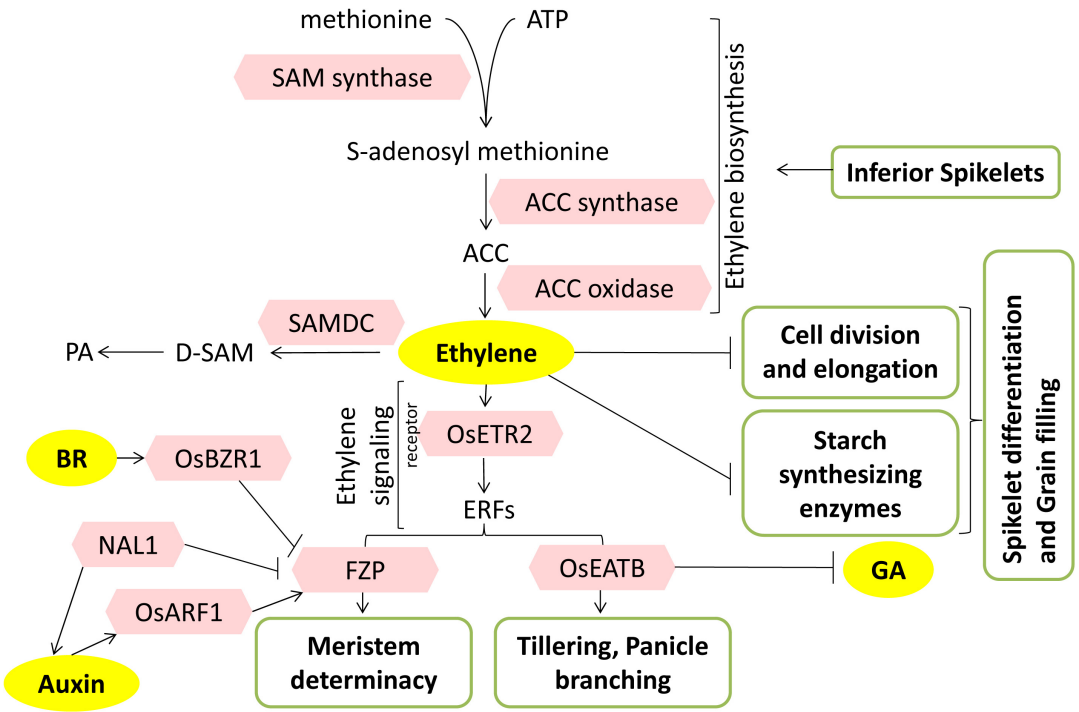


FIGURE 6 | Mode of action of ethylene during panicle development and grain filling. The model summarizes the role of ethylene-related factors (pink hexagons) in the regulation of panicle development and grain filling. Inferior spikelets accumulate ethylene by promoting its biosynthesis. Polyamines compete and reroute ethylene metabolism, thus antagonizing the ethylene response. Meristem transition determination by FZP greatly affects the branching potential of panicles. FZP interacts with BR and auxin pathway components. OsEATB coordinately with GA affects the tillering and panicle branching. PA, polyamines; D-SAM, decarboxylated SAM; SAMDC, SAM decarboxylase. Arrows represent positive regulation; bar-headed lines represent negative regulation.

Moreover, spikelet development is also shown to be regulated by another gene, *OsPEX5*, which facilitates JA biosynthesis. OsPEX5 is a peroxisomal targeting sequence 1 (PTS1) receptor protein and helps in the import of an enzyme (OsOPR7) involved in JA biosynthesis into the peroxisome. Its loss of function results in abnormal spikelet morphology like extra glumes, abnormalities in the lemma and palea, formation of the lateral floret, and altered number/aberration in the stamen and pistil. More severe spikelet-related defects are associated with the *Osmyc2* mutant. OsMYC2 is an activator of JA-responsive genes that regulates the expression of some MADS genes like *OsMADS1*, *OsMADS7*, and *OsMADS14* and plays roles in spikelet development (You et al., 2019). Thus, abnormal spikelet morphogenesis resulting from altered JA biosynthesis/signaling can affect the grain yield. For example, the heterologous overexpression of *Arabidopsis JMT* (encodes jasmonic acid carboxyl methyltransferase enzyme that converts JA into methyl Jasmonate, MeJA) in rice results in the reduction in grain yield due to decreased number of spikelets per panicle and filling rates. Low grain yield is obtained due to altered spikelet development resulting from the increased levels (sixfold) of MeJA in young panicles. Furthermore, wild-type young panicles show much higher levels of MeJA (∼19-fold) under drought stress resulting in loss of grain yield (Kim et al., 2009). In addition, another JA biosynthesis mutant, *pre* (*precocious*), which primarily shows a defect in juvenile-to-adult transition, also possesses defect in seed set. In normal conditions, swelled lodicules result in flower opening, while their withering results in flower closing soon after flowering. However, *pre* plants bear non-withering lodicules that resulted in persistent flower opening, and thus, plants are mostly devoid of seed set. *PRE* encodes for allene oxide synthase (OsAOS1) enzyme involved in JA biosynthesis (Hibara et al., 2016). *NOG1* (*NUMBER OF GRAINS 1*) also affects JA biosynthesis and controls grain yield. *NOG1* encodes an enoyl-CoA hydratase/isomerase enzyme, which is involved in fatty acid β-oxidation pathway. Its higher expression results in the downregulation of the genes associated with the JA biosynthesis pathway leading to low endogenous JA levels. This results in the enhancement of grain number per panicle, and thus, grain number per plant is increased (Huo et al., 2017). These studies suggest the involvement of JA in spikelet development, which in turn plays a vital role in determining the grain yield.

Crosstalk of Phytohormones to Regulate Grain Number

Signaling cascades and responses of several phytohormones are overlapping, and the molecular components are often shared among them. A complex network of effectors of multiple hormonal pathways collide and communicate to regulate critical agronomic traits. The following are a few well worked out examples reported so far.

A signaling module of *mitogen-activated protein kinases* (*MAPKs*), involving sequential phosphorylation of OsMKKK10-OsMKK4-OsMAPK6, is well established in regulating the seed size in rice by increasing cell proliferation (Xu et al., 2018). In this module, *OsMKK4 (smg1)* and *OsMAPK6* are known to regulate BR responses to affect grain size (Duan et al., 2014; Liu S. et al., 2015; Xu et al., 2018). Lately, a signaling link has been established between BR-mediated grain size regulation and CK-mediated grain number regulation in rice. *GRAIN SIZE AND NUMBER1* (*GSN1*) encodes another mitogen-activated protein kinase, OsMKP1, that regulates both grain size and grain number. The mutant *gsn1* had larger seeds but sparse panicle, whereas overexpression of *GSN1* results in an increased number of seeds per panicle but smaller seed size. Guo et al. (2018) showed that GSN1 negatively regulates the OsMKKK10–OsMKK4–OsMAPK6 cascade by directly dephosphorylating OsMAPK6/OsMPK6 (**Figure 1**). The CK levels were significantly reduced in *gsn1* with the increase in *OsCKX2* transcript and severe downregulation of CK-activating *LOG*. On the other hand, BR response was triggered in *gsn1* mutants by the upregulation of BR signaling genes (Guo et al., 2018). A recent study identified an upstream regulator, *ERECTA1* (*OsER1*), of the OsMKKK10–OsMKK4–OsMAPK6 cascade that encodes a receptor-like kinase protein. OsER1 activates the MAPK module either directly or indirectly, wherein OsMAPK6 directly interacts with DST to phosphorylate it (**Figure 1**). Phosphorylated DST transactivates the expression of *OsCKX2* that results in CK degradation. Interaction of GSN1 and OsMAPK6 maintains the homeostasis of CK levels (Guo et al., 2020) (**Figure 1**). Thus, CK and BR pathways contribute to a trade-off between grain number per panicle and grain size via MAP kinase signaling module (**Figure 1**).

Two recent studies have established that D53, a transcriptional repressor of the SL signaling, coordinates with components of SL, CK, and BR to regulate plant architecture that ultimately regulates grain number (Duan et al., 2019; Fang et al., 2020). In the first report, D53 has been shown to act as a link between SL and CK crosstalk, where D53 negatively regulates the expression of the CK catabolism gene, i.e., *OsCKX9* (*CYTOKININ OXIDASE/DEHYDROGENASE 9*) (Duan et al., 2019). *OsCKX9* encodes for a cytokinin oxidase enzyme involved in the degradation of CK, and it appears to be an SL-induced gene as a primary response of SL signaling. Its overexpression and loss of function display similar phenotypes of reduced panicle length, less primary and secondary branches per panicle, less grain number per panicle, along with other vegetative alterations like higher tillering and reduction in plant height. It has been evidenced that SL-induced degradation of D53 leads to an increase in the transcript and thus protein levels of *OsCKX9*, which catabolizes CK. In turn, low CK levels are responsible for reduced expression of a downstream CK-responsive gene *OsRR5* (type-A response regulator) (Duan et al., 2019). Therefore, SL regulates the rice branching, including panicle morphology, by activating CK catabolism (**Figures 1**, **3**). In the second study, SL and BR signaling pathways coordinate and regulate tillering in rice. Fang et al. (2020) showed that D53 interacts and makes a complex with another transcriptional repressor BZR1, which functions downstream of BR signaling to regulate the expression of genes involved in tillering. The D53–OsBZR1 complex inhibits the expression of *FC1/TB1* gene that encodes a TCP transcription factor, which has been reported to inhibit shoot branching in rice and is a target of SL signaling (Takeda et al., 2003; Fang et al., 2020). OsBZR1 directly binds to the *FC1* promoter and recruits D53 to inhibit the expression of *FC1* in the tiller bud. The presence of SL induces degradation of D53–OsBZR1 complex and leads to the upregulation of *FC1/TB1* expression, and thus, SL signaling inhibits rice tillering. On the contrary, BR signaling leads to the accumulation of BZR1 as well as the downstream components of the BR signaling pathway, RLA1, and DLT. The accumulated OsBZR1–RLA1–DLT complex interacts with D53 and inhibits *FC1/TB1* expression and hence promotes tillering (Song et al., 2017; Fang et al., 2020). Thus, antagonistic regulation of D53–OsBZR1 complex stability by SL and BR regulates tillering (**Figures 3**, **5**). SL and BR also regulate the panicle architecture independently, which raises the possibility of the existence of such coordination of SL and BR in controlling panicle-related traits also.

As another instance of crosstalk, two reports have shown that the upstream regulatory regions of ethylene-responsive factor *FZP*, which has a proven role in panicle branching, are occupied by BR and auxin-responsive transcription factors to regulate *FZP* expression. In the first report, a QTL named *Small Grain and Dense Panicle 7* (*SGPD7*) responsible for increased seed per panicle but decreased seed length was identified. *SGPD7* was identical to *FZP* with an 18-bp duplication in its ∼5.3 kb upstream region that turns out to be a silencer region. BR signaling transcriptional repressor, BZR1, was shown to bind to this region and represses the expression of *FZP* (**Figure 6**). The duplication of the silencer resulted in more number of secondary branches and more spikelets per panicle and an overall 15% increase in the yield (Bai et al., 2017). Second QTL, *CONTROL OF SECONDARY BRANCH 1* (*COS1*), also representing FZP, had a 4-bp deletion in its 2.7 kb upstream region. The deletion was shown to disrupt the binding of Auxin Response Factor6 (OsARF6) onto the FZP promoter and consequently results in decreased *FZP* expression, removal of branching repression, an increase in secondary branches, and more grain yield. Interestingly, this 4-bp deletion is shown to be strongly selected during domestication, as all 218 cultivated rice cultivars analyzed had this deletion (Huang et al., 2018). Moreover, FZP was shown to interact with NAL1 that encodes a trypsin-like serine and cysteine protease, which affects polar auxin transport. NAL1 negatively regulates FZP by enhancing its degradation (Huang et al., 2018) (**Figure 6**). Thus, BZR1- and OsARF6-mediated transcriptional control and NAL1-mediated posttranslational control of *FZP* makes a complex network of ethylene–BR–auxin pathways in regulating grain yield.

CKs and GAs play antagonistic roles in regulating reproductive meristem activity, with CKs having positive effects on meristem activity and maintenance, whereas GA is detrimental to meristem activity (Wu et al., 2016b). KNOTTED1-like homeobox (KNOX) proteins are homeodomain-containing transcription factors that accumulate in the cells around the meristem and regulate the accumulation of CK and GA in the

meristematic cells by directly targeting their biosynthesis and catabolism genes, respectively. KNOX proteins upregulate the expression of members of the *OsIPT* gene family that increase CK biosynthesis intermediates and CK response (**Figure 1**). On the other hand, it lowers the GA levels by downregulating *GA20Oxs* genes that catalyze the GA biosynthetic steps for bioactive GA. Accordingly, *KNOX* genes establish a high CK and low GA level balance in the meristem to maintain its activity (Sakamoto et al., 2006b). The QTL, *GNP1* (*GRAIN NUMBER PER PANICLE*), encoding a GA biosynthetic protein, GA20Ox1, shows increased accumulation of *KNOX* transcripts and high inflorescence meristem activity. The upregulation of *GNP1* (using NIL-$GNP1^{TQ}$) results in the increase in grain yield by 5.7–9.6% by enhancing the total grain number per panicle and secondary branching. Interestingly, NIL-$GNP1^{TQ}$ plants contain a low level of bioactive GA, despite the fact that the expression of GA biosynthesis genes increases. *GNP1* increases CK activity in rice panicles via negative feedback of CK biosynthesis mediated by the *KNOX* gene. Simultaneously, it reduces GA accumulation in the panicles via upregulating GA catabolism genes (Wu et al., 2016b). Hence, the crosstalk between KNOX-mediated CK activity and GA activity regulates panicle architecture (**Figures 1**, **4**).

Panicle development and branching are also regulated by the antagonistic interplay between CK and auxins. One of the core components of auxin signaling machinery is the F-BOX TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX PROTEIN (TIR1/AFB) receptor that directly links auxin perception to the degradation of Aux/IAA proteins (Dharmasiri et al., 2005) (**Figure 2**). *Auxin-signaling F-Box 6* (*OsAFB6*) is preferentially expressed in the meristematic tissues (shoot apical meristem and young inflorescences), and its overexpression resulted in larger panicles with increased spikelets per panicle, more primary branch number, and a marked increase in yield by 50%. Assessment of endogenous hormone levels in the young panicles showed a drastic reduction in IAA levels, whereas the concentration of bioactive CKs was found to be significantly high in $OsAFB6^{OX}$. Accordingly, the expression of *OsCKX2* was also downregulated in $OsAFB6^{OX}$ (**Figures 1**, **2**). Moreover, *WUSCHEL-LIKE HOMEOBOX (WOX3)*, which is known to positively regulate *KNOX* genes (promotes CK biosynthesis) via negative regulation of *YAB3*, was upregulated in $OsAFB6^{OX}$ (Dai et al., 2007). Consequently, CK accumulated in young panicles of $OsAFB6^{OX}$. Taken together, the opposite effects of CK and auxin in *OsAFB6*-mediated signaling provided robustness to the regulation of panicle development and branching and hence grain yield (He et al., 2018). Another report has shown that polar auxin transport required for axillary bud formation is affected by an SL biosynthetic enzyme, *Dwarf27* (*D27*), which encodes an iron-containing protein (**Figures 2**, **3**). The mutation *d27* augmented the polar auxin transport and exhibited the phenotype-like increased tillering and reduced plant height. The *d27* phenotype was also rescued by SL application; thus, SL and auxins are shown to interact together to control shoot branching in rice (Hao et al., 2009).

The crosstalk between GA and ethylene also seems to influence panicle branching and tillering through *ethylene response factor associated with tillering and panicle branching* (*OsEATB*). It is an AP2/ERF family transcription factor that is downregulated by ethylene treatment and possibly a negative regulator of the ethylene signaling pathway (Qi et al., 2011). Contrary to a general ethylene response, overexpression *OsEATB* reduced the plant height and panicle length at maturity. Downregulation of *OsCPS2* (ent-kaurene synthase A) in $OsEATB^{OX}$ transgenics restricted GA biosynthesis and decreased GA levels, leading to suppression in internode elongation and thus resulting in dwarf phenotype. Further, overexpression of OsEATB increased the number of panicles per plant, and spikelets per panicle, possibly because the reduction in plant height enhanced the branching potential of both tillers and panicles, thereby increasing yield. This intercommunication between GA and ethylene positively regulates the tillering and panicle branching in rice by a reduced responsiveness of GA during internode elongation via ethylene-induced decrease in GA biosynthesis (Qi et al., 2011) (**Figures 4**, **6**).

An interesting report has shown that small RNAs also regulate the homeostasis of multiple hormones to coordinate the features of agricultural traits. *OsDCL3a* encodes an RNase III-class Dicer-like 3 enzyme involved in the production of 24-nt small interfering RNAs (siRNAs), especially from miniature inverted repeat transposable elements (MITEs). Downregulation of *DCL3a* expression resulted in smaller panicles and reduced panicle branching with severe effects on secondary branching. Furthermore, other traits like flag leaf angle and plant height were also affected. Reduction in *OsDCL3a* function upregulated the expression of GA and BR biogenesis-related genes, whereas it downregulated the expression of GA-deactivating gene (*EUI1*). Thus, OsDCL3a generated 24-nt siRNAs directly target the expression of GA and BR homeostasis-related genes and, thereby, affect panicle morphology and hence grain yield (Wei et al., 2014) (**Figures 4**, **5**).

These interhormonal communications and coordinations are the core to the development of plant architecture. It is essential to understand them, to be able to fix the shortcomings of the available crop genotypes for the betterment of yield potential.

CONCLUSION AND FUTURE DIRECTIONS

Inflorescence architecture serves as a critical determinant of rice yield and contributes to securing reproductive success, and thus, it represents a major trait of agronomic importance. In cultivated varieties of rice, each spikelet is represented by a single floret and thus a single grain, in contrast to other grass crops. Increasing the number of florets per spikelet can be an innovative approach to increase the productivity of rice inflorescence. The terminal fertile floret is surrounded by lateral organs that are a pair of rudimentary glumes and a pair of sterile lemmas. Studies have suggested that sterile lemmas are rudimentary lateral florets, suggesting a three floret spikelet hypothesis in rice (Ren et al., 2020). Several reports have shown that hormonal signaling is a critical determinant of lateral organ development.

OsMADS1 controls the determination and differentiation of lateral organs by stimulating auxin transport, signaling, and response. Conversely, it suppresses the CK response by directly repressing the A-type response regulators (ARRs) (Khanday et al., 2013). MULTIFLORET SPIKELET1 (MFS1) is an ERF family protein that determines lateral organ identity by regulating the expression of other vital determinants of lateral organs (Ren et al., 2013). *ASP1*, as shown to be regulating auxin signaling, also determines the identity of the lateral organs (Zhang T. et al., 2017). JA-mediated signaling is also evidently shown to regulate lateral organ identity. An understanding of genes in the conception of lateral florets and the knowledge of precise hormonal equilibrium required for its differentiation may eventually materialize the generation of fertile lateral florets.

Combining genes and QTL with a known positive effect on grain number is a logical strategy to enhance the grain yield. *GNP1* and *NAL1* both are implicated in positively regulating grain number per panicle by adjusting CK–GA homeostasis and polar auxin transport, respectively. Moreover, both have pleiotropic effects on the leaf size (source tissue). Introgression lines generated by combining *GNP1* and *NAL1* significantly enhanced the yield potential further, in comparison to each allele independently, by increasing the source strength (broader leaves) and sink capacity (more grains) (Wang et al., 2020b). Although the grain yield increased owing to increased grain number per panicle, the pleiotropic effects of decreased grain weight remained unavoidable (Wang et al., 2020b; Zhai et al., 2020). The yield potential of the plant can only be fully exploited if the coordination of source strength and sink capacities is achieved in harmony with the assimilate flow (Smith et al., 2018; Zhai et al., 2020). Acquiring the ability to break the linkage between grain number and grain size would be fundamental to increase grain yields credibly, and, for this, hormone dynamics is indispensable (Panigrahi et al., 2019).

Further, phytohormone pathways were exploited to manipulate plant architecture to improve yields in Green Revolution varieties (Wen et al., 2018). A recent report showed that semidwarf and increased lodging resistance phenotype contributed by the GA pathway gene *SD1* was strongly coselected with an allele of *HTD1/D17* that contributed SL-induced increment of tillers in elite high-yielding Green Revolution rice varieties. The pyramiding of the GA and SL pathways by artificial co-selection led to a beneficial change in plant architecture during rice Green Revolution (Wang et al., 2020a). Another significant study demonstrated the fine tuning of panicle architecture by QTL pyramiding (Agata et al., 2020). QTL *Prl5* (encoding a GA biosynthetic enzyme) and QTL *PRIMARY BRANCH LENGTH6* (*Pbl6*) were shown to regulate non-overlapping components of panicle architecture and function independent of each other. *Pbl6/ABERRANT PANICLE ORGANIZATION1* (*APO1*) encoding an F-box-containing protein regulates primary branching and length of the upper primary branch. *Prl5*, on the other hand, governs the length of lower primary branches and rachis length. A combination of *Prl5* and *Pbl6* resulted in longer and more branched panicles with spatially arranged grains due to long primary branches, thus higher yield. Agata et al. (2020) showed that a range of panicle architecture patterns could be generated by regulating the expression of *Prl5* and *Pbl6* without a tradeoff relationship. These diversities of panicle architecture can be of great use to the breeders.

Clues from other plant systems are indispensable in understanding the dynamics of phytohormones in plants. The added information can be transferred or borrowed to the rice system for yield management strategies. A study in *Arabidopsis* showed that, in addition to the quantity of CK, the quality of CK also governs the physiological function in regulating shoot growth. The two bioactive forms of CK, viz., iP-type [N^6-(Δ^2-isopentenyl) adenine] and tZ-type (trans Zeatin), differ in their side-chain moieties. The iP-type CK can be converted to tZ type by trans-hydroxylation by specific cytochrome P450s. A decline in this enzymatic activity decreases the shoot growth without any quantitative reduction in CK levels (Gao et al., 2019). It would be of interest to find how far this mechanism affects the shoot and inflorescence branching in rice. A recent report showed that a maize serine/threonine-protein kinase, encoded by *KERNEL NUMBER PER ROW6* (*KNR6*), positively regulates grain yield. *KNR6* is involved in inflorescence meristem development via phosphorylation of an ARF GTPase-activating protein (AGAP) (Jia et al., 2020). Rice OsAGAP is well implicated in auxin transport and localization in root development (Zhuang et al., 2006), but its function is not established in inflorescence meristem maintenance. Moreover, the orthologous gene for *KNR6* is not yet characterized in rice. Assessing the function of *OsAGAP* and *OsKNR6* and their interaction in inflorescence development via an auxin-mediated pathway in rice can be a future endeavor.

Hormonal networks are complex, interlinked, and regulated at multiple levels. However, the existing understanding of molecular resources for hormonal engineering has tremendous potential for manipulation of plant architecture for crop yield improvement and, at the same time, have pleiotropic effects. Hence, mindful pyramiding of genes with an understanding of cross-hormone signaling dynamics can plausibly tap the plant's resources fully.

AUTHOR CONTRIBUTIONS

PD, AP, SS, and AKT compiled the literature, prepared the figures, and wrote the manuscript. AKT and PD conceptualized and designed the structure of the manuscript. All the authors read and approved the final manuscript.

FUNDING

The research in our lab is supported by the Department of Biotechnology (DBT) and Science and Engineering Research Board (SERB), Government of India. PD acknowledges the fund and fellowship received from DBT. AP and SS acknowledge the fellowships received from the Department of Science and Technology (DST) and Council of Scientific and Industrial Research (CSIR), Government of India, respectively.

REFERENCES

Agata, A., Ando, K., Ota, S., Kojima, M., Takebayashi, Y., Takehara, S., et al. (2020). Diverse panicle architecture results from various combinations of Prl5/GA20ox4 and Pbl6/APO1 alleles. *Commun. Biol.* 3:302. doi: 10.1038/s42003-020-1036-8

Alder, A., Jamil, M., Marzorati, M., Bruno, M., Vermathen, M., Bigler, P., et al. (2012). The path from β-carotene to carlactone, a strigolactone-like plant hormone. *Science* 335, 1348–1351. doi: 10.1126/science.1218094

Ali, A., Xu, P., Riaz, A., and Wu, X. (2019). Current advances in molecular mechanisms and physiological basis of panicle degeneration in rice. *Int. J. Mol. Sci.* 20:1613. doi: 10.3390/ijms20071613

Arite, T., Iwata, H., Ohshima, K., Maekawa, M., Nakajima, M., Kojima, M., et al. (2007). DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. *Plant J.* 51, 1019–1029. doi: 10.1111/j.1365-313X.2007.03210.x

Arite, T., Umehara, M., Ishikawa, S., Hanada, A., Maekawa, M., Yamaguchi, S., et al. (2009). D14, a strigolactone-insensitive mutant of rice, shows an accelerated outgrowth of tillers. *Plant Cell Physiol.* 50, 1416–1424. doi: 10.1093/pcp/pcp091

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745. doi: 10.1126/science.1113373

Bai, X., Huang, Y., Hu, Y., Liu, H., Zhang, B., Smaczniak, C., et al. (2017). Duplication of an upstream silencer of FZP increases grain yield in rice. *Nat. Plants* 3, 885–893. doi: 10.1038/s41477-017-0042-4

Bai, X., Huang, Y., Mao, D., Wen, M., Zhang, L., and Xing, Y. (2016). Regulatory role of FZP in the determination of panicle branching and spikelet formation in rice. *Sci. Rep.* 6:19022. doi: 10.1038/srep19022

Barazesh, S., and McSteen, P. (2008). Hormonal control of grass inflorescence development. *Trends Plant Sci.* 13, 656–662. doi: 10.1016/j.tplants.2008.09.007

Barbez, E., Kubeš, M., Rolčík, J., Béziat, C., Pěnčík, A., Wang, B., et al. (2012). A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. *Nature* 485, 119–122. doi: 10.1038/nature11001

Beveridge, C. A., and Kyozuka, J. (2010). New genes in the strigolactone-related shoot branching pathway. *Curr. Opin. Plant Biol.* 13, 34–39. doi: 10.1016/j.pbi.2009.10.003

Cai, Q., Yuan, Z., Chen, M., Yin, C., Luo, Z., Zhao, X., et al. (2014). Jasmonic acid regulates spikelet development in rice. *Nat. Commun.* 5:3476. doi: 10.1038/ncomms4476

Chen, Y., Fan, X., Song, W., Zhang, Y., and Xu, G. (2012). Over-expression of OsPIN2 leads to increased tiller numbers, angle and shorter plant height through suppression of OsLAZY1. *Plant Biotechnol. J.* 10, 139–149. doi: 10.1111/j.1467-7652.2011.00637.x

Chu, Y., Xu, N., Wu, Q., Yu, B., Li, X., Chen, R., et al. (2019). Rice transcription factor OsMADS57 regulates plant height by modulating gibberellin catabolism. *Rice* 12:38. doi: 10.1186/s12284-019-0298-6

Cole, M. B., Augustin, M. A., Robertson, M. J., and Manners, J. M. (2018). The science of food security. *Sci. Food* 2:14. doi: 10.1038/s41538-018-0021-9

Dai, M., Hu, Y., Zhao, Y., Liu, H., and Zhou, D. X. (2007). A WUSCHEL-LIKE HOMEOBOX gene represses a YABBY gene expression required for rice leaf development. *Plant Physiol.* 144, 380–390. doi: 10.1104/pp.107.095737

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. *Nature* 435, 441–445. doi: 10.1038/nature03543

Ding, C., You, J., Chen, L., Wang, S., and Ding, Y. (2014). Nitrogen fertilizer increases spikelet number per panicle by enhancing cytokinin synthesis in rice. *Plant Cell Rep.* 33, 363–371. doi: 10.1007/s00299-013-1536-9

Du, Y., Liu, L., Li, M., Fang, S., Shen, X., Chu, J., et al. (2017). UNBRANCHED3 regulates branching by modulating cytokinin biosynthesis and signaling in maize and rice. *New Phytol.* 214, 721–733. doi: 10.1111/nph.14391

Duan, J., Yu, H., Yuan, K., Liao, Z., Meng, X., Jing, Y., et al. (2019). Strigolactone promotes cytokinin degradation through transcriptional activation of CYTOKININ OXIDASE/ DEHYDROGENASE 9 in rice. *Proc. Natl. Acad. Sci. U.S.A.* 116, 14319–14324. doi: 10.1073/pnas.1810980116

Duan, P., Rao, Y., Zeng, D., Yang, Y., Xu, R., Zhang, B., et al. (2014). SMALL GRAIN 1, which encodes a mitogen-activated protein kinase kinase 4, influences grain size in rice. *Plant J.* 77, 547–557. doi: 10.1111/tpj.12405

Dun, E. A., Brewer, P. B., and Beveridge, C. A. (2009). Strigolactones: discovery of the elusive shoot branching hormone. *Trends Plant Sci.* 14, 364–372. doi: 10.1016/j.tplants.2009.04.003

Dun, E. A., Germain, A., de Saint Rameau, C., and Beveridge, C. A. (2012). Antagonistic action of strigolactone and cytokinin in bud outgrowth control. *Plant Physiol.* 158, 487–498. doi: 10.1104/pp.111.186783

Elert, E. (2014). Rice by the numbers: a good grain. *Nature* 514, S50–S51. doi: 10.1038/514S50a

Fang, N., Xu, R., Huang, L., Zhang, B., Duan, P., Li, N., et al. (2016). SMALL GRAIN 11 controls grain size, grain number and grain yield in rice. *Rice* 9:64. doi: 10.1186/s12284-016-0136-z

Fang, Z., Ji, Y., Hu, J., Guo, R., Sun, S., and Wang, X. (2020). Strigolactones and brassinosteroids antagonistically regulate the stability of the D53-OsBZR1 complex to determine FC1 expression in rice tillering. *Mol. Plant* 13, 586–597. doi: 10.1016/j.molp.2019.12.005

Ferguson, B. J., and Beveridge, C. A. (2009). Roles for auxin, cytokinin, and strigolactone in regulating shoot branching. *Plant Physiol.* 149, 1929–1944. doi: 10.1104/pp.109.135475

Fujishiro, Y., Agata, A., Ota, S., Ishihara, R., Takeda, Y., Kunishima, T., et al. (2018). Comprehensive panicle phenotyping reveals that qSrn7/FZP influences higher-order branching. *Sci. Rep.* 8:12511. doi: 10.1038/s41598-018-30395-9

Fujita, D., Trijatmiko, K. R., Tagle, A. G., Sapasap, M. V., Koide, Y., Sasaki, K., et al. (2013). NAL1 allele from a rice landrace greatly increases yield in modern indica cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20431–20436. doi: 10.1073/pnas.1310790110

Gallavotti, A., Yang, Y., Schmidt, R. J., and Jackson, D. (2008). The relationship between auxin transport and maize branching. *Plant Physiol.* 147, 1913–1923. doi: 10.1104/pp.108.121541

Gallavotti, A., Zhao, Q., Kyozuka, J., Meeley, R. B., Ritter, M. K., Doebley, J. F., et al. (2004). The role of barren stalk1 in the architecture of maize. *Nature* 432, 630–635. doi: 10.1038/nature03148

Gälweiler, L., Guan, C., Müller, A., Wisman, E., Mendgen, K., Yephremov, A., et al. (1998). Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. *Science* 282, 2226–2230. doi: 10.1126/science.282.5397.2226

Gao, S., and Chu, C. (2020). Gibberellin metabolism and signaling: targets for improving agronomic performance of crops. *Plant Cell Physiol.* 2020:pcaa104. doi: 10.1093/pcp/pcaa104

Gao, X. Q., Wang, N., Wang, X. L., and Zhang, X. S. (2019). Architecture of wheat inflorescence: insights from rice. *Trends Plant Sci.* 24, 802–809. doi: 10.1016/j.tplants.2019.06.002

Gao, Z., Qian, Q., Liu, X., Yan, M., Feng, Q., Dong, G., et al. (2009). Dwarf 88, a novel putative esterase gene affecting architecture of rice plant. *Plant Mol. Biol.* 71, 265–276. doi: 10.1007/s11103-009-9522-x

Gouda, G., Gupta, M. K., Donde, R., Mohapatra, T., Vadde, R., and Behera, L. (2020). Marker-assisted selection for grain number and yield-related traits of rice (*Oryza sativa* L.). *Physiol. Mol. Biol. Plants* 26, 885–898. doi: 10.1007/s12298-020-00773-7

Guo, S., Xu, Y., Liu, H., Mao, Z., Zhang, C., Ma, Y., et al. (2013). The interaction between OsMADS57 and OsTB1 modulates rice tillering via DWARF14. *Nat. Commun.* 4:2542. doi: 10.1038/ncomms2542

Guo, T., Chen, K., Dong, N. Q., Shi, C. L., Ye, W. W., Gao, J. P., et al. (2018). GRAIN SIZE AND NUMBER1 negatively regulates the OSMKKK10-OSMKK4-OSMPK6 cascade to coordinate the trade-off between grain number per panicle and grain size in rice. *Plant Cell* 30, 871–888. doi: 10.1105/tpc.17.00959

Guo, T., Lu, Z.-Q., Shan, J.-X., Ye, W.-W., Dong, N.-Q., and Lin, H.-X. (2020). ERECTA1 acts upstream of the OsMKKK10-OsMKK4-OsMPK6 cascade to control spikelet number by regulating cytokinin metabolism in rice. *Plant Cell* 2020:tpc.20.00351. doi: 10.1105/tpc.20.00351

Hada, W., Bo, Z., Cao, W. H., Biao, M., Gang, L., Liu, Y. F., et al. (2009). The ethylene receptor ETR2 delays floral transition and affects starch accumulation in rice. *Plant Cell* 21, 1473–1494. doi: 10.1105/tpc.108.065391

Hao, L., Renxiao, W., Qian, Q., Meixian, Y., Xiangbing, M., Zhiming, F., et al. (2009). DWARF27, an iron-containing protein required for the biosynthesis of strigolactones, regulates rice tiller bud outgrowth. *Plant Cell* 21, 1512–1525. doi: 10.1105/tpc.109.065987

He, Q., Yang, L., Hu, W., Zhang, J., and Xing, Y. (2018). Overexpression of an auxin receptor OsAFB6 significantly enhanced grain yield by increasing cytokinin and decreasing auxin concentrations in rice panicle. *Sci. Rep.* 8:14051. doi: 10.1038/s41598-018-32450-x

Hibara, K. I., Isono, M., Mimura, M., Sentoku, N., Kojima, M., Sakakibara, H., et al. (2016). Jasmonate regulates juvenile-to-adult phase transition in rice. *Development* 143, 3407–3446. doi: 10.1242/dev.138602

Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., et al. (2009). Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* 41, 494–497. doi: 10.1038/ng.352

Huang, Y., Bai, X., Luo, M., and Xing, Y. (2019). Short Panicle 3 controls panicle architecture by upregulating APO2/RFL and increasing cytokinin content in rice. *J. Integr. Plant Biol.* 61, 987–999. doi: 10.1111/jipb.12729

Huang, Y., Zhao, S., Fu, Y., Sun, H., Ma, X., Tan, L., et al. (2018). Variation in the regulatory region of FZP causes increases in secondary inflorescence branching and grain yield in rice domestication. *Plant J.* 96, 716–733. doi: 10.1111/tpj.14062

Huijser, P., and Schmid, M. (2011). The control of developmental phase transitions in plants. *Development* 138, 4117–4129. doi: 10.1242/dev.063511

Huo, X., Wu, S., Zhu, Z., Liu, F., Fu, Y., Cai, H., et al. (2017). NOG1 increases grain production in rice. *Nat. Commun.* 8:1497. doi: 10.1038/s41467-017-01501-8

Ishikawa, S., Maekawa, M., Arite, T., Onishi, K., Takamure, I., and Kyozuka, J. (2005). Suppression of tiller bud activity in tillering dwarf mutants of rice. *Plant Cell Physiol.* 46, 79–86. doi: 10.1093/pcp/pci022

Jameson, P. E., and Song, J. (2016). Cytokinin: a key driver of seed yield. *J. Exp. Bot.* 67, 593–606. doi: 10.1093/jxb/erv461

Jeon, J.-S., Jung, K.-H., Kim, H.-B., Suh, J.-P., and Khush, G. S. (2011). Genetic and molecular insights into the enhancement of rice yield potential. *J. Plant Biol.* 54, 1–9. doi: 10.1007/s12374-011-9144-0

Jia, H., Li, M., Li, W., Liu, L., Jian, Y., Yang, Z., et al. (2020). A serine/threonine protein kinase encoding gene KERNEL NUMBER PER ROW6 regulates maize grain yield. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-14746-7

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., et al. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* 504, 401–405. doi: 10.1038/nature12870

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591

Kang, K., Shim, Y., Gi, E., An, G., and Paek, N. C. (2019). Mutation of ONAC096 enhances grain yield by increasing panicle number and delaying leaf senescence during grain filling in rice. *Int. J. Mol. Sci.* 20:5241. doi: 10.3390/ijms20205241

Kato, Y., Kamoshita, A., and Yamagishi, J. (2008). Preflowering abortion reduces spikelet number in upland rice (*Oryza sativa* L.) under water stress. *Crop Sci.* 48, 2389–2395. doi: 10.2135/cropsci2007.11.0627

Kebrom, T. H., Spielmeyer, W., and Finnegan, E. J. (2013). Grasses provide new insights into regulation of shoot branching. *Trends Plant Sci.* 18, 41–48. doi: 10.1016/j.tplants.2012.07.001

Khanday, I., Ram Yadav, S., and Vijayraghavan, U. (2013). Rice LHS1/OsMADS1 controls floret meristem specification by coordinated regulation of transcription factors and hormone signaling pathways. *Plant Physiol.* 161, 1970–1983. doi: 10.1104/pp.112.212423

Khush, G. S. (2013). Strategies for increasing the yield potential of cereals: case of rice as an example. *Plant Breed.* 132, 433–436. doi: 10.1111/pbr.1991

Kiba, T., Takei, K., Kojima, M., and Sakakibara, H. (2013). Side-chain modification of cytokinins controls shoot growth in *Arabidopsis*. *Dev. Cell* 27, 452–461. doi: 10.1016/j.devcel.2013.10.004

Kim, E. H., Kim, Y. S., Park, S. H., Koo, Y. J., Choi, Y., Do Chung, Y. Y., et al. (2009). Methyl jasmonate reduces grain yield by mediating stress signals to alter spikelet development in rice. *Plant Physiol.* 149, 1751–1760. doi: 10.1104/pp.108.134684

Komatsu, M., Chujo, A., Nagato, Y., Shimamoto, K., and Kyozuka, J. (2003). Frizzy panicle is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. *Development* 130, 3841–3850. doi: 10.1242/dev.00564

Komatsu, M., Maekawa, M., Shimamoto, K., and Kyozuka, J. (2001). The LAX1 and FRIZZY PANICLE 2 genes determine the inflorescence architecture of rice by controlling rachis-branch and spikelet development. *Dev. Biol.* 231, 364–373. doi: 10.1006/dbio.2000.9988

Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., et al. (2007). Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445, 652–655. doi: 10.1038/nature05504

Li, M., Tang, D., Wang, K., Wu, X., Lu, L., Yu, H., et al. (2011). Mutations in the F-box gene LARGER PANICLE improve the panicle architecture and enhance the grain yield in rice. *Plant Biotechnol. J.* 9, 1002–1013. doi: 10.1111/j.1467-7652.2011.00610.x

Li, S., Zhao, B., Yuan, D., Duan, M., Qian, Q., Tang, L., et al. (2013). Rice zinc finger protein DST enhances grain production through controlling Gn1a/OsCKX2 expression. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3167–3172. doi: 10.1073/pnas.1300359110

Li, Y., Li, X., Fu, D., and Wu, C. (2018). Panicle morphology mutant 1 (PMM1) determines the inflorescence architecture of rice by controlling brassinosteroid biosynthesis. *BMC Plant Biol.* 18:348. doi: 10.1186/s12870-018-1577-x

Li, Y., Zhu, J., Wu, L., Shao, Y., Wu, Y., and Mao, C. (2019). Functional divergence of PIN1 paralogous genes in rice. *Plant Cell Physiol.* 60, 2720–2732. doi: 10.1093/pcp/pcz159

Liang, R., Qin, R., Yang, C., Zeng, D., Jin, X., and Shi, C. (2019). Identification and characterization of a novel strigolactone-insensitive mutant, Dwarfism with high tillering ability 34 (dhta-34) in rice (*Oryza sativa* L.). *Biochem. Genet.* 57, 403–420. doi: 10.1007/s10528-018-9896-z

Lin, Y., Zhao, Z., Zhou, S., Liu, L., Kong, W., Chen, H., et al. (2017). Top Bending Panicle1 is involved in brassinosteroid signaling and regulates the plant architecture in rice. *Plant Physiol. Biochem.* 121, 1–13. doi: 10.1016/j.plaphy.2017.10.001

Liu, L., Tong, H., Xiao, Y., Che, R., Xu, F., Hu, B., et al. (2015). Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11102–11107. doi: 10.1073/pnas.1512748112

Liu, S., Hua, L., Dong, S., Chen, H., Zhu, X., Jiang, J., et al. (2015). OsMAPK6, a mitogen-activated protein kinase, influences rice grain size and biomass production. *Plant J.* 84, 672–681. doi: 10.1111/tpj.13025

Liu, W., Wu, C., Fu, Y., Hu, G., Si, H., Zhu, L., et al. (2009). Identification and characterization of HTD2: a novel gene negatively regulating tiller bud outgrowth in rice. *Planta* 230, 649–658. doi: 10.1007/s00425-009-0975-6

Liu, Y., Xu, Y., Xiao, J., Ma, Q., Li, D., Xue, Z., et al. (2011). OsDOG, a gibberellin-induced A20/AN1 zinc-finger protein, negatively regulates gibberellin-mediated cell elongation in rice. *J. Plant Physiol.* 168, 1098–1105. doi: 10.1016/j.jplph.2010.12.013

Lu, G., Coneva, V., Casaretto, J. A., Ying, S., Mahmood, K., Liu, F., et al. (2015). OsPIN5b modulates rice (*Oryza sativa*) plant architecture and yield by changing auxin homeostasis, transport and distribution. *Plant J.* 83, 913–925. doi: 10.1111/tpj.12939

Lu, Z., Yu, H., Xiong, G., Wang, J., Jiao, Y., Liu, G., et al. (2013). Genome-wide binding analysis of the transcription activator IDEAL PLANT ARCHITECTURE1 reveals a complex network regulating rice plant architecture. *Plant Cell* 25, 3743–3759. doi: 10.1105/tpc.113.113639

Luo, A., Qian, Q., Yin, H., Liu, X., Yin, C., Lan, Y., et al. (2006). EUI1, encoding a putative cytochrome P450 monooxygenase, regulates internode elongation by modulating gibberellin responses in rice. *Plant Cell Physiol.* 47, 181–191. doi: 10.1093/pcp/pci233

Malik, N., Ranjan, R., Parida, S. K., Agarwal, P., and Tyagi, A. K. (2020). Mediator subunit OsMED14_1 plays an important role in rice development. *Plant J.* 101, 1411–1429. doi: 10.1111/tpj.14605

McSteen, P. (2009). Hormonal regulation of branching in grasses. *Plant Physiol.* 149, 46–55. doi: 10.1104/pp.108.129056

McSteen, P., and Leyser, O. (2005). Shoot branching. *Annu. Rev. Plant Biol.* 56, 353–374. doi: 10.1146/annurev.arplant.56.032604.144122

Minakuchi, K., Kameoka, H., Yasuno, N., Umehara, M., Luo, L., Kobayashi, K., et al. (2010). FINE CULM1 (FC1) works downstream of strigolactones to inhibit the outgrowth of axillary buds in rice. *Plant Cell Physiol.* 51, 1127–1135. doi: 10.1093/pcp/pcq083

Miura, K., Ikeda, M., Matsubara, A., Song, X. J., Ito, M., Asano, K., et al. (2010). OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* 42, 545–549. doi: 10.1038/ng.592

Mohapatra, P. K., and Sahu, S. K. (1991). Heterogeneity of primary branch development and spikelet survival in rice panicle in relation to assimilates of primary branches. *J. Exp. Bot.* 42, 871–879. doi: 10.1093/jxb/42.7.871

Mravec, J., Skùpa, P., Bailly, A., Hoyerová, K., Køeček, P., Bielach, A., et al. (2009). Subcellular homeostasis of phytohormone auxin is mediated by the ER-localized PIN5 transporter. *Nature* 459, 1136–1140. doi: 10.1038/nature08066

Oikawa, T., and Kyozuka, J. (2009). Two-step regulation of LAX PANICLE1 protein accumulation in axillary meristem formation in rice. *Plant Cell* 21, 1095–1108. doi: 10.1105/tpc.108.065425

Panda, B. B., Badoghar, A. K., Sekhar, S., Shaw, B. P., and Mohapatra, P. K. (2016). 1-MCP treatment enhanced expression of genes controlling endosperm cell division and starch biosynthesis for improvement of grain filling in a dense-panicle rice cultivar. *Plant Sci.* 246, 11–25. doi: 10.1016/j.plantsci.2016.02.004

Panigrahi, R., Kariali, E., Panda, B. B., Lafarge, T., and Mohapatra, P. K. (2019). Controlling the trade-off between spikelet number and grain filling: the hierarchy of starch synthesis in spikelets of rice panicle in relation to hormone dynamics. *Funct. Plant Biol.* 46, 507–523. doi: 10.1071/FP18153

Piao, R., Jiang, W., Ham, T. H., Choi, M. S., Qiao, Y., Chu, S. H., et al. (2009). Map-based cloning of the ERECT PANICLE 3 gene in rice. *Theor. Appl. Genet.* 119, 1497–1506. doi: 10.1007/s00122-009-1151-x

Qi, J., Qian, Q., Bu, Q., Li, S., Chen, Q., Sun, J., et al. (2008). Mutation of the rice narrow leaf1 gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiol.* 147, 1947–1959. doi: 10.1104/pp.108.118778

Qi, W., Sun, F., Wang, Q., Chen, M., Huang, Y., Feng, Y. Q., et al. (2011). Rice ethylene-response AP2/ERF factor OsEATB restricts internode elongation by down-regulating a gibberellin biosynthetic gene. *Plant Physiol.* 157, 216–228. doi: 10.1104/pp.111.179945

Ren, D., Li, Y., He, G., and Qian, Q. (2020). Multifloret spikelet improves rice yield. *New Phytol.* 225, 2301–2306. doi: 10.1111/nph.16303

Ren, D., Li, Y., Zhao, F., Sang, X., Shi, J., Wang, N., et al. (2013). MULTI-FLORET SPIKELET1, which encodes an AP2/ERF protein, determines spikelet meristem fate and sterile lemma identity in rice. *Plant Physiol.* 162, 872–884. doi: 10.1104/pp.113.216044

Saini, S., Sharma, I., and Pati, P. K. (2015). Versatile roles of brassinosteroid in plants in the context of its homoeostasis, signaling and crosstalks. *Front. Plant Sci.* 6:950. doi: 10.3389/fpls.2015.00950

Sakakibara, H. (2006). Cytokinins: activity, biosynthesis, and translocation. *Annu. Rev. Plant Biol.* 57, 431–449. doi: 10.1146/annurev.arplant.57.032905.105231

Sakamoto, T., Morinaka, Y., Ohnishi, T., Sunohara, H., Fujioka, S., Ueguchi-Tanaka, M., et al. (2006a). Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. *Nat. Biotechnol.* 24, 105–109. doi: 10.1038/nbt1173

Sakamoto, T., Sakakibara, H., Kojima, M., Yamamoto, Y., Nagasaki, H., Inukai, Y., et al. (2006b). Ectopic expression of KNOTTED1-like homeobox protein induces expression of cytokinin biosynthesis genes in rice. *Plant Physiol.* 142, 54–62. doi: 10.1104/pp.106.085811

Sakuma, S., and Schnurbusch, T. (2020). Of floral fortune: tinkering with the grain yield potential of cereal crops. *New Phytol.* 225, 1873–1882. doi: 10.1111/nph.16189

Schaller, G. E., Street, I. H., and Kieber, J. J. (2014). Cytokinin and the cell cycle. *Curr. Opin. Plant Biol.* 21, 7–15. doi: 10.1016/j.pbi.2014.05.015

Shimizu-Sato, S., Tanaka, M., and Mori, H. (2009). Auxin-cytokinin interactions in the control of shoot branching. *Plant Mol. Biol.* 69, 429–435. doi: 10.1007/s11103-008-9416-3

Sieberer, T., and Leyser, O. (2006). Auxin transport, but in which direction? *Science* 312, 858–860. doi: 10.1126/science.1127659

Smith, M. R., Rao, I. M., and Merchant, A. (2018). Source-sink relationships in crop plants and their influence on yield development and nutritional quality. *Front. Plant Sci.* 871:1889. doi: 10.3389/fpls.2018.01889

Song, X., Lu, Z., Yu, H., Shao, G., Xiong, J., Meng, X., et al. (2017). IPA1 functions as a downstream transcription factor repressed by D53 in strigolactone signaling in rice. *Cell Res.* 27, 1128–1141. doi: 10.1038/cr.2017.102

Sreenivasulu, N., and Schnurbusch, T. (2012). A genetic playground for enhancing grain number in cereals. *Trends Plant Sci.* 17, 91–101. doi: 10.1016/j.tplants.2011.11.003

Stamm, P., Ramamoorthy, R., and Kumar, P. P. (2011). Feeding the extra billions: strategies to improve crops and enhance future food security. *Plant Biotechnol. Rep.* 5, 107–120. doi: 10.1007/s11816-011-0169-0

Strable, J. (2020). Sugars inform the circadian clock how to shape rice shoots via the strigolactone pathway. *Plant Cell* 2020:tpc.20.00615. doi: 10.1105/tpc.20.00615

Sun, L., Zhang, Q., Wu, J., Zhang, L., Jiao, X., Zhang, S., et al. (2014). Two rice authentic histidine phosphotransfer proteins, OsAHP1 and OsAHP2, mediate cytokinin signaling and stress responses in rice. *Plant Physiol.* 165, 335–345. doi: 10.1104/pp.113.232629

Tabuchi, H., Zhang, Y., Hattori, S., Omae, M., Shimizu-Sato, S., Oikawa, T., et al. (2011). LAX PANICLE2 of rice encodes a novel nuclear protein and regulates the formation of axillary meristems. *Plant Cell* 23, 3276–3287. doi: 10.1105/tpc.111.088765

Takeda, T., Suwa, Y., Suzuki, M., Kitano, H., Ueguchi-Tanaka, M., Ashikari, M., et al. (2003). The OsTB1 gene negatively regulates lateral branching in rice. *Plant J.* 33, 513–520. doi: 10.1046/j.1365-313X.2003.01648.x

Tanabe, S., Ashikari, M., Fujioka, S., Takatsuto, S., Yoshida, S., Yano, M., et al. (2005). A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length. *Plant Cell* 17, 776–790. doi: 10.1105/tpc.104.024950

Thimann, K. V., and Koepfli, J. B. (1935). Identity of the growth-promoting and root-forming substances of plants. *Nature* 135, 101–102. doi: 10.1038/135101a0

Tong, X., Wang, Y., Sun, A., Bello, B., Ni, S., and Zhang, J. (2018). Notched Belly Grain 4, a novel allele of Dwarf 11, regulates grain shape and seed germination in rice (*Oryza sativa* L.). *Int. J. Mol. Sci.* 19:4069. doi: 10.3390/ijms19124069

Tyczewska, A., Woźniak, E., Gracz, J., Kuczyñski, J., and Twardowski, T. (2018). Towards food security: current state and future prospects of agrobiotechnology. *Trends Biotechnol.* 36, 1219–1229. doi: 10.1016/j.tibtech.2018.07.008

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., et al. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455, 195–200. doi: 10.1038/nature07272

Vij, S., and Tyagi, A. K. (2006). Genome-wide analysis of the stress associated protein (SAP) gene family containing A20/AN1 zinc-finger(s) in rice and their phylogenetic relationship with *Arabidopsis*. *Mol. Genet. Genom.* 276, 565–575. doi: 10.1007/s00438-006-0165-1

Walden, R., Cordeiro, A., and Tiburcio, A. F. (1997). Polyamines: small molecules triggering pathways in plant growth and development. *Plant Physiol.* 113, 1009–1013. doi: 10.1104/pp.113.4.1009

Wang, F., Han, T., Song, Q., Ye, W., Song, X., Chu, J., et al. (2020). Rice circadian clock regulates tiller growth and panicle development through strigolactone signaling and sugar sensing. *Plant Cell* 32, 3124–3138. doi: 10.1105/tpc.20.00289

Wang, Y., Shang, L., Yu, H., Zeng, L., Hu, J., Ni, S., et al. (2020a). A strigolactone biosynthesis gene contributed to the green revolution in rice. *Mol. Plant* 13, 923–932. doi: 10.1016/j.molp.2020.03.009

Wang, Y., Zhai, L., Chen, K., Shen, C., Liang, Y., Wang, C., et al. (2020b). Natural sequence variations and combinations of GNP1 and NAL1 determine the grain number per panicle in rice. *Rice* 13:14. doi: 10.1186/s12284-020-00374-8

Wang, J. R., Hu, H., Wang, G. H., Li, J., Chen, J. Y., and Wu, P. (2009). Expression of PIN genes in rice (*Oryza sativa* L.): tissue specificity and regulation by hormones. *Mol. Plant* 2, 823–831. doi: 10.1093/mp/ssp023

Wang, Q., Kohlen, W., Rossmann, S., Vernoux, T., and Theres, K. (2014). Auxin depletion from the leaf axil conditions competence for axillary meristem formation in *Arabidopsis* and tomato. *Plant Cell* 26, 2068–2079. doi: 10.1105/tpc.114.123059

Wang, Y., Wang, J., Shi, B., Yu, T., Qi, J., Meyerowitz, E. M., et al. (2014). The stem cell niche in leaf axils is established by auxin and cytokinin in *Arabidopsis*. *Plant Cell* 26, 2055–2067. doi: 10.1105/tpc.114.123083

Wang, Y., and Li, J. (2005). The plant architecture of rice (*Oryza sativa*). *Plant Mol. Biol.* 59, 75–84. doi: 10.1007/s11103-004-4038-x

Wang, Y. P., Tang, S. Q., Chen, H. Z., Wu, Z. F., Zhang, H., Duan, E. C., et al. (2017). Identification and molecular mapping of indica high-tillering dwarf mutant htd4, a mild phenotype allelic mutant of D14 in rice (*Oryza sativa* L.). *Plant Biol.* 19, 851–858. doi: 10.1111/plb.12612

Wei, L., Gu, L., Song, X., Cui, X., Lu, Z., Zhou, M., et al. (2014). Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3877–3882. doi: 10.1073/pnas.1318131111

Wen, C. K., Zhao, Y., and Ruan, Y. L. (2018). Editorial: hormonal control of important agronomic traits. *Front. Plant Sci.* 9:1504. doi: 10.3389/fpls.2018.01504

Werner, T., Motyka, V., Strnad, M., and Schmülling, T. (2001). Regulation of plant growth by cytokinin. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10487–10492. doi: 10.1073/pnas.171304098

Wu, Q., Li, D., Li, D., Liu, X., Zhao, X., Li, X., et al. (2015). Overexpression of OsDof12 affects plant architecture in rice (*Oryza sativa* L.). *Front. Plant Sci.* 6:833. doi: 10.3389/fpls.2015.00833

Wu, Y., Fu, Y., Zhao, S., Gu, P., Zhu, Z., Sun, C., et al. (2016a). CLUSTERED PRIMARY BRANCH 1, a new allele of DWARF11, controls panicle architecture and seed size in rice. *Plant Biotechnol. J.* 14, 377–386. doi: 10.1111/pbi.12391

Wu, Y., Wang, Y., Mi, X. F., Shan, J. X., Li, X. M., Xu, J. L., et al. (2016b). The QTL GNP1 encodes GA20ox1, which increases grain number and yield by increasing cytokinin activity in rice panicle meristems. *PLoS Genet.* 12:e006386. doi: 10.1371/journal.pgen.1006386

Xie, K., Wu, C., and Xiong, L. (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* 142, 280–293. doi: 10.1104/pp.106.084475

Xu, M., Zhu, L., Shou, H., and Wu, P. (2005). A PIN1 family gene, OsPIN1, involved in auxin-dependent adventitious root emergence and tillering in rice. *Plant Cell Physiol.* 46, 1674–1681. doi: 10.1093/PCP/PCI183

Xu, R., Duan, P., Yu, H., Zhou, Z., Zhang, B., Wang, R., et al. (2018). Control of grain size and weight by the OsMKKK10-OsMKK4-OsMAPK6 signaling pathway in rice. *Mol. Plant* 11, 860–873. doi: 10.1016/j.molp.2018.04.004

Yang, C. J., Zhang, C., Lu, Y. N., Jin, J. Q., and Wang, X. L. (2011). The mechanisms of brassinosteroids' action: from signal transduction to plant development. *Mol. Plant* 4, 588–600. doi: 10.1093/mp/ssr020

Yang, J., Cho, L.-H., Yoon, J., Yoon, H., Wai, A. H., Hong, W.-J., et al. (2019). Chromatin interacting factor OsVIL2 increases biomass and rice grain yield. *Plant Biotechnol. J.* 17, 178–187. doi: 10.1111/pbi.12956

Yang, J., Lee, S., Hang, R., Kim, S.-R., Lee, Y.-S., Cao, X., et al. (2013). OsVIL2 functions with PRC2 to induce flowering by repressing OsLFL1 in rice. *Plant J.* 73, 566–578. doi: 10.1111/tpj.12057

Yang, J., Yunying, C., Zhang, H., Liu, L., and Zhang, J. (2008). Involvement of polyamines in the post-anthesis development of inferior and superior spikelets in rice. *Planta* 228, 137–149. doi: 10.1007/s00425-008-0725-1

Yang, J., Zhang, J., Wang, Z., Liu, K., and Wang, P. (2006). Post-anthesis development of inferior and superior spikelets in rice in relation to abscisic acid and ethylene. *J. Exp. Bot.* 57, 149–160. doi: 10.1093/jxb/erj018

Yeh, S. Y., Chen, H. W., Ng, C. Y., Lin, C. Y., Tseng, T. H., Li, W. H., et al. (2015). Down-regulation of *Cytokinin Oxidase 2* expression increases tiller number and improves rice yield. *Rice* 8, 1–13. doi: 10.1186/s12284-015-0070-5

Yin, C., Gan, L., Ng, D., Zhou, X., and Xia, K. (2007). Decreased panicle-derived indole-3-acetic acid reduces gibberellin A1 level in the uppermost internode, causing panicle enclosure in male sterile rice Zhenshan 97A. *J. Exp. Bot.* 58, 2441–2449. doi: 10.1093/jxb/erm077

Yin, C. C., Zhao, H., Ma, B., Chen, S. Y., and Zhang, J. S. (2017). Diverse roles of ethylene in regulating agronomic traits in rice. *Front. Plant Sci.* 8:1676. doi: 10.3389/fpls.2017.01676

Yoshida, A., Ohmori, Y., Kitano, H., Taguchi-Shiobara, F., and Hirano, H. Y. (2012). ABERRANT SPIKELET and PANICLE1, encoding a TOPLESS-related transcriptional co-repressor, is involved in the regulation of meristem fate in rice. *Plant J.* 70, 327–339. doi: 10.1111/j.1365-313X.2011.04872.x

You, X., Zhu, S., Zhang, W., Zhang, J., Wang, C., Jing, R., et al. (2019). OsPEX5 regulates rice spikelet development through modulating jasmonic acid biosynthesis. *New Phytol.* 224, 712–724. doi: 10.1111/nph.16037

Yuan, H., Fan, S., Huang, J., Zhan, S., Wang, S., Gao, P., et al. (2017). 08SG2/OsBAK1 regulates grain size and number, and functions differently in Indica and Japonica backgrounds in rice. *Rice* 10:25. doi: 10.1186/s12284-017-0165-2

Zazímalová, E., Murphy, A. S., Yang, H., Hoyerová, K., and Hosek, P. (2010). Auxin transporters–why so many?. *Cold Spring Harb. Perspect. Biol.* 2:a001552. doi: 10.1101/cshperspect.a001552

Zha, M., Imran, M., Wang, Y., Xu, J., Ding, Y., and Wang, S. (2019). Transcriptome analysis revealed the interaction among strigolactones, auxin, and cytokinin in controlling the shoot branching of rice. *Plant Cell Rep.* 38, 279–293. doi: 10.1007/s00299-018-2361-y

Zhai, L., Wang, F., Yan, A., Liang, C., Wang, S., Wang, Y., et al. (2020). Pleiotropic effect of GNP1 underlying grain number per panicle on sink, source and flow in Rice. *Front. Plant Sci.* 11:933. doi: 10.3389/fpls.2020.00933

Zhang, D., and Yuan, Z. (2014). Molecular control of grass inflorescence development. *Annu. Rev. Plant Biol.* 65, 553–578. doi: 10.1146/annurev-arplant-050213-40104

Zhang, T., Li, Y., Ma, L., Sang, X., Ling, Y., Wang, Y., et al. (2017). LATERAL FLORET 1 induced the three-florets spikelet in rice. *Proc. Natl. Acad. Sci. U.S.A.* 114, 9984–9989. doi: 10.1073/pnas.1700504114

Zhang, W., Chen, Y., Wang, Z., and Yang, J. (2017). Polyamines and ethylene in rice young panicles in response to soil drought during panicle differentiation. *Plant Growth Regul.* 82, 491–503. doi: 10.1007/s10725-017-0275-2

Zhang, W., Sheng, J., Xu, Y., Xiong, F., Wu, Y., Wang, W., et al. (2019a). Role of brassinosteroids in rice spikelet differentiation and degeneration under soil-drying during panicle development. *BMC Plant Biol.* 19:409. doi: 10.1186/s12870-019-2025-2

Zhang, W., Zhu, K., Wang, Z., Zhang, H., Gu, J., Liu, L., et al. (2019b). Brassinosteroids function in spikelet differentiation and degeneration in rice. *J. Integr. Plant Biol.* 61, 943–963. doi: 10.1111/jipb.12722

Zhao, L., Tan, L., Zhu, Z., Xiao, L., Xie, D., and Sun, C. (2015). PAY1 improves plant architecture and enhances grain yield in rice. *Plant J.* 83, 528–536. doi: 10.1111/tpj.12905

Zhou, F., Lin, Q., Zhu, L., Ren, Y., Zhou, K., Shabek, N., et al. (2013). D14-SCF D3 -dependent degradation of D53 regulates strigolactone signalling. *Nature* 504, 406–410. doi: 10.1038/nature12878

Zhou, J., Li, Z., Xiao, G., Zhai, M., Pan, X., and Huang, R. (2020). CYP71D8L is a key regulator involved in growth and stress responses by mediating gibberellin homeostasis in rice. *Science* 71, 1160–1170. doi: 10.1093/jxb/erz491

Zhou, Y., Tao, Y., Yuan, Y., Zhang, Y., Miao, J., Zhang, R., et al. (2018). Characterisation of a novel quantitative trait locus, GN4-1, for grain number and yield in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 131, 637–648. doi: 10.1007/s00122-017-3025-y

Zhou, Y., Tao, Y., Zhu, J., Miao, J., Liu, J., Liu, Y., et al. (2017). GNS4, a novel allele of DWARF11, regulates grain number and grain size in a high-yield rice variety. *Rice* 10:34. doi: 10.1186/s12284-017-0171-4

Zhu, Y., Nomura, T., Xu, Y., Zhang, Y., Peng, Y., Mao, B., et al. (2006). ELONGATED UPPERMOST INTERNODE encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. *Plant Cell* 18, 442–456. doi: 10.1105/tpc.105.038455

Zhuang, X., Jiang, J., Li, J., Ma, Q., Xu, Y., Xue, Y., et al. (2006). Over-expression of OsAGAP, an ARF-GAP, interferes with auxin influx, vesicle trafficking and root development. *Plant J.* 48, 581–591. doi: 10.1111/j.1365-313X.2006.02898.x

Zou, J., Zhang, S., Zhang, W., Li, G., Chen, Z., Zhai, W., et al. (2006). The rice HIGH-TILLERING DWARF1 encoding an ortholog of *Arabidopsis* MAX3 is required for negative regulation of the outgrowth of axillary buds. *Plant J.* 48, 687–698. doi: 10.1111/j.1365-313X.2006.02916.x

Zwiewka, M., Bilanovičová, V., Seifu, Y. W., and Nodzyński, T. (2019). The nuts and bolts of PIN auxin efflux carriers. *Front. Plant Sci.* 10:985. doi: 10.3389/fpls.2019.00985

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Deveshwar, Prusty, Sharma and Tyagi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 November 2020
doi: 10.3389/fgene.2020.587968

Analysis and Identification of QTL for Resistance to *Sclerotinia sclerotiorum* in Pea (*Pisum sativum* L.)

Rahil Ashtari Mahini[1], *Ajay Kumar*[1], *Elias M. Elias*[1], *Jason D. Fiedler*[2], *Lyndon D. Porter*[3] *and Kevin E. McPhee*[4*]

[1] *Plant Science Department, North Dakota State University, Fargo, ND, United States,* [2] *USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND, United States,* [3] *USDA-ARS, Grain Legume Genetics and Physiology Research Unit, Prosser, WA, United States,* [4] *Plant Sciences and Plant Pathology Department, Montana State University, Bozeman, MT, United States*

OPEN ACCESS

Edited by:
Vijay Kumar Tiwari, University of Maryland, College Park, United States

Reviewed by:
Krishna Kishore Gali, University of Saskatchewan, Canada
Ze Peng, University of Florida, United States
Shyam Solanki, Washington State University, United States

***Correspondence:**
Kevin E. McPhee
kevin.mcphee@montana.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *27 July 2020*
Accepted: *23 October 2020*
Published: *19 November 2020*

Citation:
Ashtari Mahini R, Kumar A, Elias EM, Fiedler JD, Porter LD and McPhee KE (2020) Analysis and Identification of QTL for Resistance to Sclerotinia sclerotiorum in Pea (Pisum sativum L.). Front. Genet. 11:587968. doi: 10.3389/fgene.2020.587968

White mold caused by *Sclerotinia sclerotiorum* is an important constraint to field pea (*Pisum sativum* L.) production worldwide. To transfer white mold resistance into an adapted background, and study the genetics of the disease, two recombinant inbred line (RIL) populations (PRIL17 and PRIL19) were developed by crossing two partially resistant plant introductions with two susceptible pea cultivars. PRIL17 (Lifter × PI240515), and PRIL19 (PI169603 × Medora) were evaluated for resistance to white mold by measuring lesion expansion inhibition (LEI) and nodal transmission inhibition (NTI) at 3, 7, and 14 days post inoculation (dpi) under controlled environmental conditions. Lesion expansion inhibition percentage (LEIP), survival rate (SR), and area under disease progress curves (AUDPC) were also calculated accordingly. Because of a positive correlation between LEI and NTI with height, short and long internode individuals of each population were analyzed separately to avoid any confounding effect of height to pathogen response. A total of 22 short genotypes demonstrated partial resistance based on at least two Porter's resistance criteria. Only two pea genotypes with partial resistance to white mold (PRIL19-18 and PRIL19-124) had both semi-leafless (afila) and short internode traits. Both the RIL populations were genotyped using genotyping by sequencing (GBS). For PRIL17 and PRIL19, genetic maps were constructed from a total of 1,967 and 1,196 single nucleotide polymorphism (SNP) and spanned over 1,494 cM and 1,415 cM representing seven and nine linkage groups, respectively. A consensus map constructed using data from both populations, had 1,486 unique SNPs over 2,461 cM belonging to seven linkage groups. Inclusive composite interval mapping (ICIM) identified thirteen quantitative trait loci (QTL) associated with white mold resistance traits in both populations. Three of them were co-located with height genes (a morphological trait that reduces infection risk and acts as disease avoidance) and the other ten QTL were associated with two forms of physiological resistance (seven for LEI and three

for NTI) with LOD and r^2 ranging from 3.0 to 28.5 and 5.1 to 64.3, respectively. The development of resistance lines, genetic dissection and identification of markers associated will help accelerate breeding efforts for white mold resistance using molecular breeding approaches.

Keywords: *Sclerotinia sclerotiorum*, *Pisum sativum*, QTL mapping, lesion expansion inhibition, nodal transmission inhibition, genotyping by sequencing, white mold

INTRODUCTION

White mold caused by *Sclerotinia sclerotiorum* (Lib.) de Bary, causes significant yield losses in most dicotyledonous crops (Bolton et al., 2006; Shahoveisi and del Rio Mendoza, 2020). This hemi-biotrophic fungus has a host range of more than 400 plant species. White mold preferably develops in cool, cloudy, wet, and humid weather during flowering (Mueller et al., 2014). Narrow row spacing, and an early canopy closure typically by long-vine plants creates ideal conditions for mycelium germination and development of the disease (Kraft and Pfleger, 2001). The pathogen infects the stem, leaf or pod tissue of plants and produces water-soaked lesion with white cottony mass of mycelium on the surface (Mueller et al., 2014).

Annual losses from *S. sclerotiorum* in pulse crops was estimated to be as high as $12 million in the United States (USDA-ARS, 2016). White mold is a significant barrier to field pea production, not only in the Northwest and Midwest areas of the USA, but worldwide (Porter et al., 2009). The most economical and environment-friendly option for management of white mold is to develop resistant field pea varieties (Fernando et al., 2004; Jain et al., 2012; Peltier et al., 2012). However, breeding for white mold resistant varieties is difficult in most of the crops, because of its polygenic inheritance (Porter et al., 2009; Davar et al., 2013). Porter et al. (2009) stated that the expression of the partial quantitative white mold resistance in pea might be in two forms, lesion expansion inhibition (LEI) and nodal transmission inhibition (NTI These two forms are a measure of rate of lesion progression and ability of pathogen to spread in the stem through the nodes, respectively (Porter et al., 2009). Some morphological traits such as stem thickness, short internodes and leaf morphology (semi-leafless), can also influence white mold resistance (Porter et al., 2009). Plant height in soybean (Boland and Hall, 1986) and sunflower (Bazzalo et al., 1991) had shown positive correlation with resistance, resulting in disease escape. Limited attempts to screen pea germplasm for white mold resistance (Blanchette and Auld, 1978; Porter et al., 2009; Tashtemirov, 2012) identified only moderate levels of disease resistance. In this background, genetic dissection of the available disease resistance would be valuable to pyramid resistant genes from diverse sources.

Quantitative trait loci (QTL) mapping identification for *S. sclerotiorum* resistance has been reported in many crops including sunflower (Micic et al., 2005), soybean (Bastien et al., 2014; Iquira et al., 2015), bean (Kolkman and Kelly, 2003; Ender and Kelly, 2005; Miklas, 2007), rapeseed (*Brassica napus*) (Wu et al., 2013), and *Brassica oleracea* (Mei et al., 2013). However, only limited research on *S. sclerotiorum* resistance has been reported in pea (Tashtemirov, 2012). The only reported QTL mapping study for resistance to white mold in pea was conducted on F_2–derived F_3 family lines from a cross between Lifter and PI240515 (Tashtemirov, 2012). The recent availability of full genome sequence of the field pea (Kreplak et al., 2019) has contributed to a surge in identification of trait-associated markers and candidate genes for a number of agronomic, seed morphology and seed quality traits (Gali et al., 2019; Dissanayaka et al., 2020), and frost (Beji et al., 2020) and heat tolerance (Tafesse et al., 2020). Similarly, we expect that these genomic resources will also contribute to fine mapping of disease resistance QTLs including those associated with white mold resistance and a better understanding of the underlying genetics. Therefore, this study was conducted to (1) develop a reliable screening method for white mold resistance, (2) identify white mold resistant individuals in two RIL populations, and (3) detect QTL associated with white mold resistance in two mapping populations for marker-assisted selection in breeding programs.

MATERIALS AND METHODS

Plant Materials

Two mapping populations of pea were used for this study: (i) PRIL17 comprised of 192 F_7 recombinant inbred lines (RILs) derived from the cross of Lifter/ PI240515 and (ii) PRIL19 comprised of 324 F_7 RILs developed from the cross PI169603/Medora. Both populations were developed using single seed decent. Lifter and Medora are susceptible cultivars, while PI240515 and PI169603 have partial resistance to white mold (Porter et al., 2009). Lifter, developed by the USDA-ARS (Grain Legume Genetics Physiology Research unit located in Pullman, WA), has short internodes with normal leaf type, white flowers and green seed (McPhee and Muehlbauer, 2002). Medora, which was developed for Midwest production, has a short internode with afila leaf type, white flowers and smooth green seed color (Grain Legume Genetics and Physiology Research Unit, 2007). The two partially resistant plant introductions were selected based on a previous study (Porter et al., 2009). PI240515 (originated from India) and PI169603 (originated from Turkey) have long internodes, normal leaf morphology, white flowers and yellow cotyledon color (https://www.ars-grin.gov/).

Inoculum Preparation

Sclerotia of *Sclerotinia. sclerotiorum* isolate Sc102 were obtained from pea cultivar named "Snake" in 2003 by Dr. Lyndon Porter in Quincy, WA. Sclerotia were maintained at 4°C until used. To break sclerotia dormancy, they were placed in a 10% bleach solution for 20 min, followed by 3 rinses with sterile

distilled water. The rinsed sclerotia were surface sterilized in 95% ethanol for one minute, briefly flamed and cultured on sterile composite agar in dark at 21–23°C for three days until the mycelium colonized half the surface of a 15 mm diameter petri dish (Khan et al., 2020). The composite agar consisted of 18.5 g Difco TM potato dextrose agar and 8.75 g Difco TM oatmeal agar in 0.5 L distilled water and autoclaved at 121 C for 20 min (Tashtemirov, 2012). Actively growing mycelium from the leading edge of the colony was used for inoculation of mapping populations.

Phenotyping and Greenhouse Evaluation

Individual RILs grown in greenhouse were inoculated with *S. sclerotiorum* 14 days after sowing, using the jumbo agar plug technique (Porter et al., 2009). Plants were grown under natural sunlight supplemented with 600-Watt high pressure sodium lamps (P. L Light Systems, Inc., Beamsville, Ontario, Canada) in the greenhouse to maintain a 16:8 h photoperiod and temperatures from 20 to 25°C during the day and 20°C at night. The experiment was arranged in a randomized complete block design (RCBD) with 4 replicates and repeated three times. A total of 186 genotypes from PRIL17 and 140 individuals from PRIL19, in addition to 5 checks, were evaluated. The populations were divided indifferent experiment sets for logistic reasons (lack of space and labor).

The test plants were inoculated by placing 3 mm of actively growing mycelial plug at the 4th node using a jumbo pulpdent amalgam carrier and the media was pressed into the leaf axis. After inoculation, plants were transferred to a mist chamber with 100% humidity and 19 to 21°C temperature for three days in the dark. White mold lesions were measured 3 days post inoculation (dpi) with a digital caliper (mm). Plants were then transferred to a mist room for another 11 days with 80% humidity and 14:10 day: night photoperiod with supplemental lighting using 400-Watt high pressure sodium lamps (P. L. Light Systems, Inc., Beamsville, Ontario, Canada).

Lesion expansion inhibition (LEI) was measured by recording lesion size in millimeters at 3, 7, and 14 dpi where a smaller value indicated greater LEI. Nodal transmission inhibition (NTI) were measured at 7 and 14 dpi. NTI was scored on a scale of 0-4 based on lesion movement from the 4th node down as described in Porter et al. (2009). A larger value indicated greater NTI. Lesion expansion inhibition percentage (LEIP) was recorded 14 dpi by measuring lesion size (mm) and then dividing by plant height. Survival rate (SR) was calculated 14 dpi by measuring the uninfected portions of the main stems and dividing by plant height ([height at 14 dpi–lesion size at 14 dpi/ height in 14 dpi] × 100).

Means for LEI, NTI, LEIP and SR were calculated using ANOVA (analysis of variance) and PROC MIXED procedures in SAS Enterprise Guide 7.1 (SAS Institute Inc. USA). In the statistical analysis, genotypes were considered fixed effects, while experiments, replication within experiments, experiment by replication by genotype, and experiment by genotype were treated as random effects.

The height of each individual was measured at 3 dpi to identify the tall and short subsets and calculate the correlation between height and LEI (3, 7 and 14 dpi) and NTI (7 and 14 dpi). Also, correlation between LEI (7 and 14 dpi) and NTI (7 and 14 dpi) was calculated. Pearson's correlation coefficients were calculated using SAS Enterprise Guide 7.

Multiple observations of lesion size (mm) at 3, 7, and 14 dpi were used to calculate AUDPC in Excel for each genotype using the following formula, where y is disease level at time (t). The AUDPC is a quantitative measure of disease intensity with time (Simko and Piepho, 2012). A smaller value indicated a greater AUDPC.

$$AUDPC = \sum_{i=1}^{N_i-1} \frac{(y_i + y_{i+1})}{2} (t_{i+1} - t_i)$$

Since LEI, LEIP, AUDPC, and SR were calculated from lesion size (mm) values, they are considered as LEI subsets in the QTL analysis.

Resistant genotypes to white mold were identified based on having low disease based on one or more of the three criteria. Those criteria include mean LEI 3 dpi equal or <25 mm, mean NTI after two weeks equal or greater than one, and survival rate of 25% or greater (Porter et al., 2009).

Genotyping by Sequencing

DNA of PRIL17 and PRIL19 individuals were extracted using CTAB (Manfioletti and Schneider, 1988) and DArT (Diversity Arrays Technology, 2008) protocols. DNA of each individual was quantified using PicoGreen assay (Ahn et al., 1996) and restriction digested with *Ape*KI (Sonah et al., 2013). Ninety-six plex libraries of each population were constructed using the GBS protocol described by Elshire et al. (2011). GBS libraries of PRIL17 and PRIL19 were sequenced by single- end read sequencing using Illumina HiSeq 2000 at the McDermott Center (http://www.utsouthwesternedu/labs/dna-genotyping-core) and the Genomic Facility of Cornell University Biotechnology Resource Center (BRC), respectively. All sequencing datasets have been deposited with links to BioProject accession number PRJNA653945 in NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/653945/).

GBS Analysis and SNP Calling

Genotyping by sequencing raw data for PRIL17 and PRIL19 were analyzed using the GBS-SNP-CROP pipeline (Melo et al., 2016) by aligning the sequences of each population over the *Pisum_sativum*_v1a (Psat_v1a) reference genome (Kreplak et al., 2019). GBS-SNP-CROP pipeline is Perl scripts that implements parsing, filtering and SNP calling with bioinformatic tools such as Trimmomatic, and SAMtools. The GBS data was filtered with minimum coverage depth (minDP 3), maximum number of mismatches in alignment ($n = 1$), maximum missing data (max-missing 0.5), and minor allele frequency (maf 0.05). One hundred eighty-six individuals from each population and their parents were genotyped by GBS and analyzed using the GBS-SNP-CROP (GBS SNP-Calling Reference Optional) pipeline. Heterozygous and monomorphic parents were filtered out, and individuals with missing >60% SNPs were also excluded.

Linkage Group Mapping

Linkage groups for each population were constructed using JoinMap® 4 (Plant Research International B.V and Kyazma B.V) (Van Ooijen, 2006). Linkage group formation was based on linkage LOD scores (logarithm of the odds). The linkage LOD score was calculated by JoinMap to compare the estimated value of the pairwise recombination frequency with 0.5. We used the start value of 3.0, end value of 20.0, and step size of 1.0 which indicated the ranges and steps of significance level that are used for grouping. After linkage groups were determined, the linkage map was constructed using the *Monte Carlo maximum likelihood* (ML) mapping algorithm for each group. Four physiological markers (*Le* for height, *er-1* for powdery mildew, *Pl* for hilum color, and *I* for cotyledon color) in PRIL17 and five physiological markers (*Le, er-1*, *Pl, I,* and *af* or afila) in PRIL19 were used to anchor linkage groups to study co-localization and association with disease resistance loci.

Consensus Map Construction

A consensus map was created by QTL IciMapping V 4.2 software (Meng et al., 2015) from PRIL17 and PRIL19 linkage groups that shared common markers. The consensus map construction (CMP) function of QTL IciMapping has three steps to form an integrated map including grouping, ordering and rippling. All markers were assembled in one group and used the ordering algorithm of nearest neighbors with two-opt (nnTwoOp). The rippling criterion was SAD (Sum of Adjacent Distances), with the window size of five markers.

QTL Mapping

QTL analysis was performed using linkage maps of PRIL17 and PRIL19, and least square means of all disease resistance components across experiments for each population. Data from greenhouse evaluations from each population were separated into tall (height $>$ 150 mm) and short (height $\leq$ 150 mm) internode individuals and QTL analysis was conducted on each subset as well as the complete dataset. QTL analysis was conducted using the Inclusive Composite Interval Mapping (ICIM) method available in the software QTL IciMapping V 4.2 (Meng et al., 2015). Only unique loci from each linkage group were included in the analyses. Permutation tests based on 1,000 repeats were performed to determine significant LOD threshold values ($\alpha = 0.05$ Type I error). Only the QTL detected above threshold LOD score were included in this study. If any such significant QTL was identified with LOD below the threshold, but $>$ 2, the QTL were also included in the results as supporting information.

To compare the means of two group of RILs which differ for QTL alleles associated with a phenotypic trait, *t*-test (assuming unequal variances) function in Excel 2016 was used.

Identification of Candidate Genes Associated With Resistance to White Mold

The QTL region associated with the flanking markers were used to search for candidate genes. Sequences of flanking markers were used to search the pea sequence database *Pisum sativum* v1a JBrowse (https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData/Pea/Psat_v1a/data) to find candidate genes related to each QTL. These markers in QTL region are associated with few genes and can be utilized as a first step for resistance breeding purposes in pea.

RESULTS

Greenhouse Evaluation

Analysis of variance of checks across the experiments showed no significant difference ($P > 0.05$) (**Table S1**, **Table S2**), suggesting the data for different experiments could be combined for analysis for each population. Also, Levene's homogeneity test for LEI (3dpi) variance of checks for each population across different experiments showed no significant difference (PRIL17: $P = 0.042$, PRIL19: $P = 0.0001$), thus justifying a joining analysis across all experiments. Plant height showed a positive correlation with LEI and NTI (**Table S3**, **Table S4**). Survival rate of long internode individuals in both populations was noticeably higher than short internode individuals (**Table S6**). These may suggest that height might be a confounding factor and might obscure the real effect of *S. sclerotiorum* infection. Therefore, short and long internode individuals of each population were analyzed separately. A negative correlation coefficient was observed between LEI and NTI in the short and tall subsets of PRIL17 and PRIL19 (**Table S5**).

Genotypes with partial resistance were identified using the criteria of Porter et al. (2009). Analysis of the short internode subset of PRIL17 identified six individuals with LEI $\leq$ 25, thirty-three with SR $\geq$ 25% and 47 genotypes with NTI $\geq$ 1. The value of AUDPC ranged from 487 to 2270 in short genotypes of PRIL17. The AUDPC of PRIL17 parents, Lifter and PI240515, were 777 and 1119, respectively. Only PRIL17-181 among the short internode subset of PRIL17 showed all three of Porter's resistance criteria and low AUDPC (487). Seventeen short individuals of PRIL17 showed two of the resistant criteria together (**Table 1**).

Among long internode genotypes of PRIL17 only four genotypes restricted lesion expansion with low LEI (3 dpi). However, all long internode genotypes of PRIL17 showed NTI$\geq$ 1 and 101 genotypes met the survival rate criteria. Only four tall genotypes (PRIL17-16, PRIL17-28, PRIL17-125, and PRIL17-179) from PRIL17 showed all three criteria together for partial resistance to white mold). PRIL17-125 showed all three of Porter's resistance criteria as well as low AUDPC (412). AUDPC for tall individuals of PRIL17 ranged from 412 to 3594.

Analysis of the short internode subset of PRIL19 identified one individual with LEI$\leq$ 25, five individuals with SR$\geq$ 25% and five genotypes with NTI$\geq$ 1. The AUDPC ranged from 824 to 3048 in short varieties of PRIL19. The AUDPC of PRIL19 parents, Medora and PI169603 were 1062 and 990, respectively. Although, none of the short individuals of PRIL19 met all three criteria, five of them met two criteria for white mold partial resistance (**Table 1**).

Analysis of the long internode subset of PRIL19 revealed that none of the individuals had LEI equal or $<$25 mm. Fifty-six of the tall subset of PRIL19 demonstrated NTI$\geq$ 1 and forty-nine tall individuals with SR $\geq$ 25%. The AUDPC for tall individuals of PRIL19 ranged from 1107 to 3306.

TABLE 1 | Lesion expansion inhibition (LEI) 3 days post inoculation (dpi), nodal transmission inhibition (NTI), and survival rate (SR) after 14 dpi and area under disease progress curve (AUDPC) from greenhouse evaluation and leaf type of the 22 short pea genotypes from PRIL17 and PRIL19 with greatest partial resistance to white mold.

Short internode genotypes	LEI (3 dpi) mm	NTI (14 dpi)	SR	AUDPC	Leaf type
PRIL17					
PRIL17-11	24	0.2	27	651	Normal
PRIL17-54	50	1.4	44	2,270	Normal
PRIL17-56	24	1.5	16	654	Normal
PRIL17-58	41	1.5	33	1,232	Normal
PRIL17-71	43	1.2	34	1,294	Normal
PRIL17-97	40	1.4	30	1,038	Normal
PRIL17-127	41	1.4	52	1,302	Normal
PRIL17-128	30	1.0	25	764	Normal
PRIL17-129	27	2.3	32	834	Normal
PRIL17-139	34	1.2	31	883	Normal
PRIL17-141	36	2.9	30	991	Normal
PRIL17-145	27	2.5	54	581	Normal
PRIL17-149	46	1.3	45	1,666	Normal
PRIL17-158	27	1.8	25	728	Normal
PRIL17-166	25	0.3	34	655	Normal
PRIL17-180	31	1.6	35	837	Normal
PRIL17-181	23	1.5	58	487	Normal
PRIL19					
PRIL19-18	63	1.4	34	2,336	Afila
PRIL19-74	74	2.3	53	2,401	Normal
PRIL19-86	38	2.8	36	1,822	Normal
PRIL19-124	37	1.2	35	1,027	Afila
PRIL19-127	29	1.4	35	930	Normal

Overall, twenty-two individuals from short genotypes of PRIL17 and PRIL19 showed partial resistance and met at least two of the resistance criteria in the greenhouse evaluation. Two of those RILs (PRIL19-18 and PRIL19-124) had afila leaf type as well (**Table 1**).

Genetic Map Construction

For PRIL17, a total of 1985 polymorphic SNP markers were selected to construct the linkage map after filtering the genotypic data. Out of those markers, a total of 1967 SNPs were mapped onto seven linkage groups with a total map length of 1494 cM (**Figure S1**). The 1967 markers represented 899 unique loci with an average distance of 1.7 cM between two loci (**Table 2**). Height (*Le*), powdery mildew resistance (*er-1*), and hilum color (*Pl*) were anchored at 257-258 cM of chr5LG3, 65 cM of chr1LG6 and 170-171 cM chr1LG6, respectively, which aligned with the Weeden map (Bordat et al., 2011).

For PRIL19, a total of 1,207 polymorphic SNP markers were selected after filtering the genotypic data. Out of those, a total of 1196 markers were mapped to nine linkage groups with a total map distance of 1415 cM (**Figure S2**). The 1196 markers represented 724 unique loci with an average distance of 1.9 cM between loci (**Table 2**). Height (*Le*), powdery mildew resistance (*er-1*), leaf type (*af*) and hilum color (*Pl*) were anchored at 296-297 cM of chr5LG3(a), 60 cM of chr1LG6, 8–16 cM of chr2LG1(b), and 129–130 cM of chr1LG6, respectively, which aligned with the Weeden map (Bordat et al., 2011). Cotyledon color (*I*) was positioned at 95 cM of chr5LG3(a).

Consensus Map Construction

A total of 1486 SNPs were assembled into seven linkage groups. These markers spanned over 2461 cM with an average distance of 1.7 cM between two markers (**Figure S3**, **Table 2**).

Quantitative Trait Loci Mapping

PRIL17 (Lifter× PI240515)

This study identified five QTL for LEI, three QTL for NTI, three for AUDPC, one for SR, and one for LEIP in PRIL17 (**Table 3**). There were some common QTL between traits which means the total of seven unique QTL were identified in PRIL17 population. A significant QTL associated with LEI, NTI, LEIP, and SR was identified at 255–264 cM on chr5LG3 in all three datasets. This QTL was located at the same position as height (*Le*) gene (Weeden et al., 1998; Lee et al., 2017) (LOD = 22.8 and r^2 = 45.8%). The genotype Lifter contributed the alleles for increased value of LEI, and PI240515 alleles contributed to increased values of NTI, LEIP, and SR at all the identified loci (**Table 5**). Common QTL associated with LEI (*QLEI.17.ndsu.6*) and AUDPC (*QAUDPC.17.ndsu.6*) was identified at 88 cM of chr6LG2 in a short internode subset with LOD = 3.1, where Lifter contributed the allele for improved value of traits (**Table 5**). Another common QTL for LEI (*QLEI.17.ndsu.2*) and AUDPC (*QAUDPC.17.ndsu.2*) was identified at 157-158 cM of chr2LG1 in short internode individuals subset with the LOD = 3.7 and the allele for increased value of traits were conferred by Lifter. One more QTL for LEI was identified at 85-114 cM of chr7LG7 (*QLEI.17.ndsu.7*, LOD = 4.7) in complete datasets as well as short internode individuals' subset (**Figure 1A**). The separation of RILs based on parental alleles at this locus showed that the non-adapted PI240515 parent, improved LEI (**Table 5**). Also, one minor QTL for LEI (*QLEI.17.ndsu.4*), and two minor QTL for NTI (*QNTI.17.ndsu.2* and *QNTI.17.ndsu.1*) were identified in the complete dataset of PRIL17 (**Figure 2A**, **Table 3**).

Few RILs were identified with a combination of positive alleles at the identified QTL. PRIL17 line120 carried the alleles for increased trait value for *QLEI.17.ndsu.2*, *QNTI.17.ndsu.2*, *QLEI.17.ndsu.4*, *QLEI.17.ndsu.6*, and *QLEI.17.ndsu.7* loci (AA+ BB+ AA+ AA+ BB). The alleles for increased trait value for *QNTI.17.ndsu.2*, *QLEI.17.ndsu.4*, *QLEI.17.ndsu.6*, *QLEI.17.ndsu.7* and *QLEI.17.ndsu.5* loci (BB+ AA+ AA+ BB+ AA) were carried by PRIL17-83, PRIL17-157, PRIL17-164 and PRIL17-166. The desirable allele for *QLEI.17.ndsu.2*, *QNTI.17.ndsu.2*, *QLEI.17.ndsu.4*, *QLEI.17.ndsu.5*, and *QLEI.17.ndsu.7* and loci (AA+ BB+ AA+ AA+ BB) were carried by PRIL17-141. The PRIL17-145 also had positive allele to increase trait value for *QNTI.17.ndsu.2*, *QLEI.17.ndsu.6*, and *QLEI.17.ndsu.7* loci (BB+ AA+ BB).

TABLE 2 | Summary of genetic maps for PRIL17, PRIL19 and the consensus map.

PRIL17 map				PRIL19 map				Consensus map		
Chr*	**#Loci**	**#Unique loci**	**Map length cM**	**Chr**	**#Loci**	**#Unique loci**	**Map length cM**	**Chr**	**#Unique loci**	**Map length cM**
chr1LG6	209	114	173.8	chr1LG6	118	70	203.8	chr1LG6	275	250.6
chr2LG1	211	77	159.2	chr2LG1(a)	80	56	93.8	chr2LG1	150	247.1
ch3LG5	292	129	182.0	chr2LG1(b)	50	30	35.9	ch3LG5	206	353.8
ch4LG4	314	156	288.9	ch3LG5	178	109	203.8	ch4LG4	242	483.5
ch5LG3	323	152	271.4	ch4LG4	170	114	215.7	ch5LG3	206	353.8
ch6LG2	257	110	213.8	ch5LG3(a)	263	149	297.2	ch6LG2	160	336.3
ch7LG7	361	161	204.9	ch5LG3(b)	14	7	6.0	ch7LG7	247	436.2
				ch6LG2	145	72	165.7			
				ch7LG7	179	117	192.9			
Total	1967	899	1494	Total	1197	724	1415	Total	1486	2461

**Pea chromosome number, chr refers to chromosome number of pea genome assembly v1a and LG refers to linkage group of earlier genetic mapping studies (Tayeh et al., 2015).*

TABLE 3 | QTL for reaction to white mold infection identified in PRIL17 based on all individuals, short internode, and long internode genotype subsets.

QTL and trait	Dpi[a]	Chr[b]	Position (cM)	Other associated traits	Dataset type[c]	Left marker	Right marker	LOD	R^2 [e]	Add[f]
HEIGHT										
*Le***		chr5LG3	246-265	LEI, NTI, AUDPC SR, LEIP	All, short, tall	chr5LG3_557023279[d]	chr5LG3_569105948	25.5	48.7	100.7
LEI (cm)										
*QLEI.17.ndsu.2**	7	chr2LG1	157-158	AUDPC	Short	chr2LG1_425797878	chr2LG1_426258021	3.7	12.4	−154.3
QLEI.17.ndsu.4	3	chr4LG4	42		All	chr4LG4_25831962	scaffold01160_24125	2.8	5.1	−2.5
*QLEI.17.ndsu.5***	3,7,14	chr5LG3	255-264	*Le*, NTI, SR, LEIP, AUDPC	All, short	chr5LG3_562563492	chr5LG3_568430003	22.8	45.8	46.9
*QLEI.17.ndsu.6****	7	chr6LG2	88	AUDPC	Short	chr6LG2_163685348	chr6LG2_159992168	3.0	11.2	9.3
QLEI.17.ndsu.7	3,7	chr7LG7	85-114		All, short	chr7LG7_170686568	chr7LG7_334919307	4.7	8.9	3.3
NTI										
QNTI.17.ndsu.1	14	chr1LG6	0		All	chr1LG6_27014318	chr1LG6_27820673	2.3	3.4	0.2
QNTI.17.ndsu.2	14	chr2LG1	70		All	chr2LG1_115338387	scaffold01733_406502	2.3	3.5	−0.19
*QNTI.17.ndsu.5***	7,14	chr5LG3	254-260	*Le*, LEI, SR, LEIP	All, tall	chr5LG3_562563492	chr5LG3_570275258	16.3	32.2	0.6
AUDPC										
*QAUDPC.17.ndsu.2**		chr2LG1	157	LEI	All	chr2LG1_423525725	chr2LG1_424938982	3.7	12.4	−154.3
*QAUDPC.17.ndsu.5***		chr5LG3	257-260	*Le*, NTI, LEI, SR, LEIP	Short	chr5LG3_566156376	chr5LG3_570275258	3.0	11.1	171.0
*QAUDPC.17.ndsu.6****		chr6LG2	88	LEI	Short	chr6LG2_163685348	chr6LG2_159992168	3.1	12.1	151.9
SR										
*QSR.17.ndsu.5***		chr5LG3	261-262	*Le*, LEI, NTI, LEIP	All	chr5LG3_570780892	chr5LG3_569106574	14.9	31.0	−9.9
LEIP										
*QLEIP.17.ndsu.5***		chr5LG3	255-260	*Le*, LEI, NTI, SR	All	chr5LG3_562563492	chr5LG3_570275258	14.6	31.5	9.9

[a] Number of days relative QTL was found post inoculation of the lesion expansion inhibition or nodal transmission inhibition traits.
[b] Pea chromosome number, chr refers to chromosome number of pea genome assembly v1a and LG refers to linkage group of earlier genetic mapping studies (Kreplak et al., 2019).
[c] Type of datasets where relative QTL was found. all: All the genotypes in dataset were used for QTL analysis; short: short subset of the dataset was used for QTL analysis; tall: tall subsets of dataset were used for QTL analysis.
[d] In each SNP ID, chr refers to chromosome number of pea genome assembly v1a and LG refers to linkage group of earlier genetic mapping studies followed by the base pair position. For non-chromosomal SNPs, scaffold followed by the scaffold number and base pair position (Kreplak et al., 2019).
[e] Phenotypic variation explained by QTL.
[f] Additive effect.
, **, *: QTL with the same markings are the same.*

PRIL19 (PI169603× Medora)

In population PRIL19, CIM identified four QTL for LEI, two each for NTI and AUDPC, and one each for SR and LEIP (**Table 4**). Some of the QTL were common among different traits and six unique QTL associated with white mold resistance in PRIL19. Also, two QTL for height were identified in PRIL19; one of them

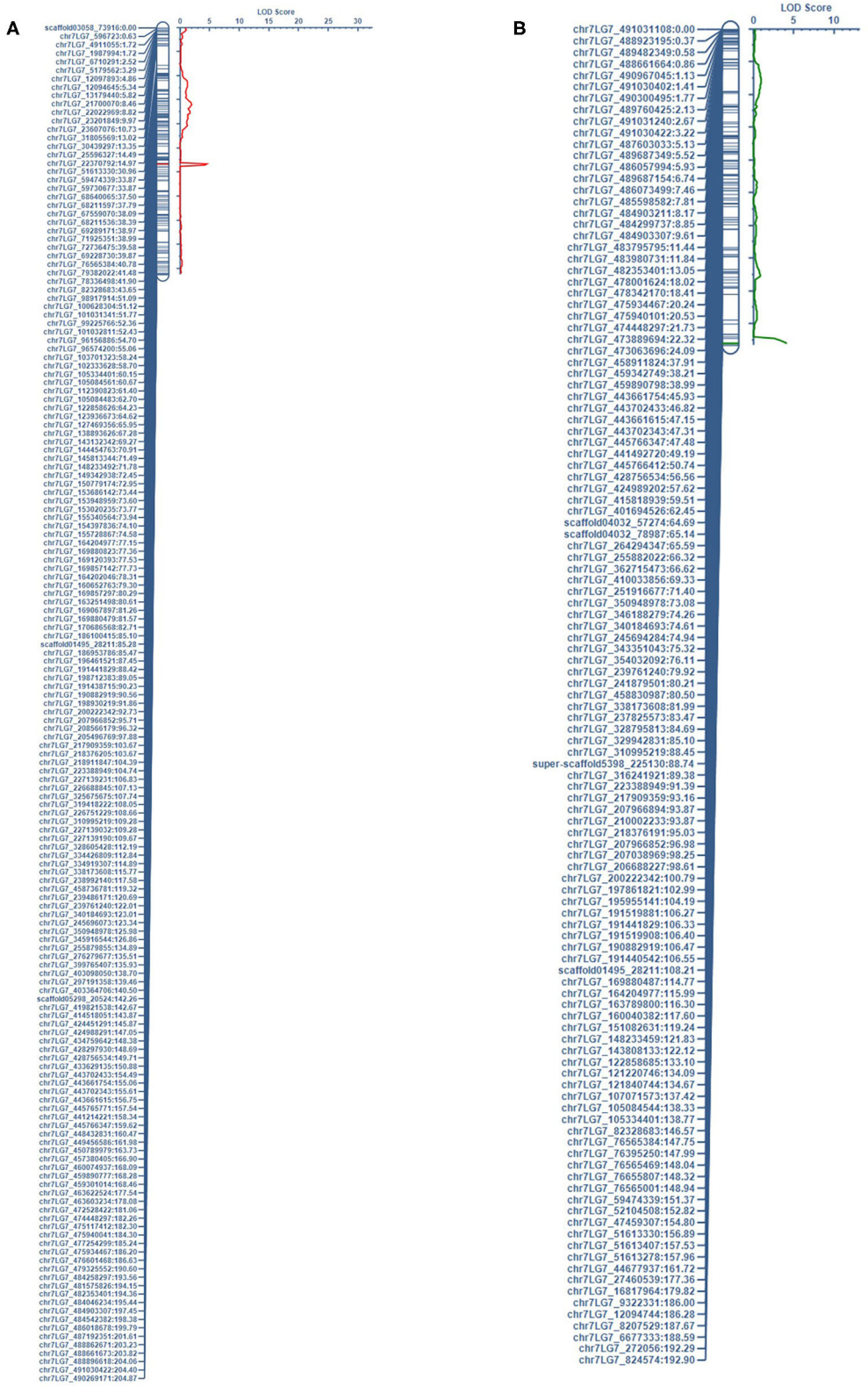


FIGURE 1 | The QTL associated with LEI on chr7LG7 was detected in the short internode subset of **(A)** PRIL17 (*QLEI.17.ndsu.7*) and **(B)** PRIL19 (*QLEI.19.ndsu.7*) populations.

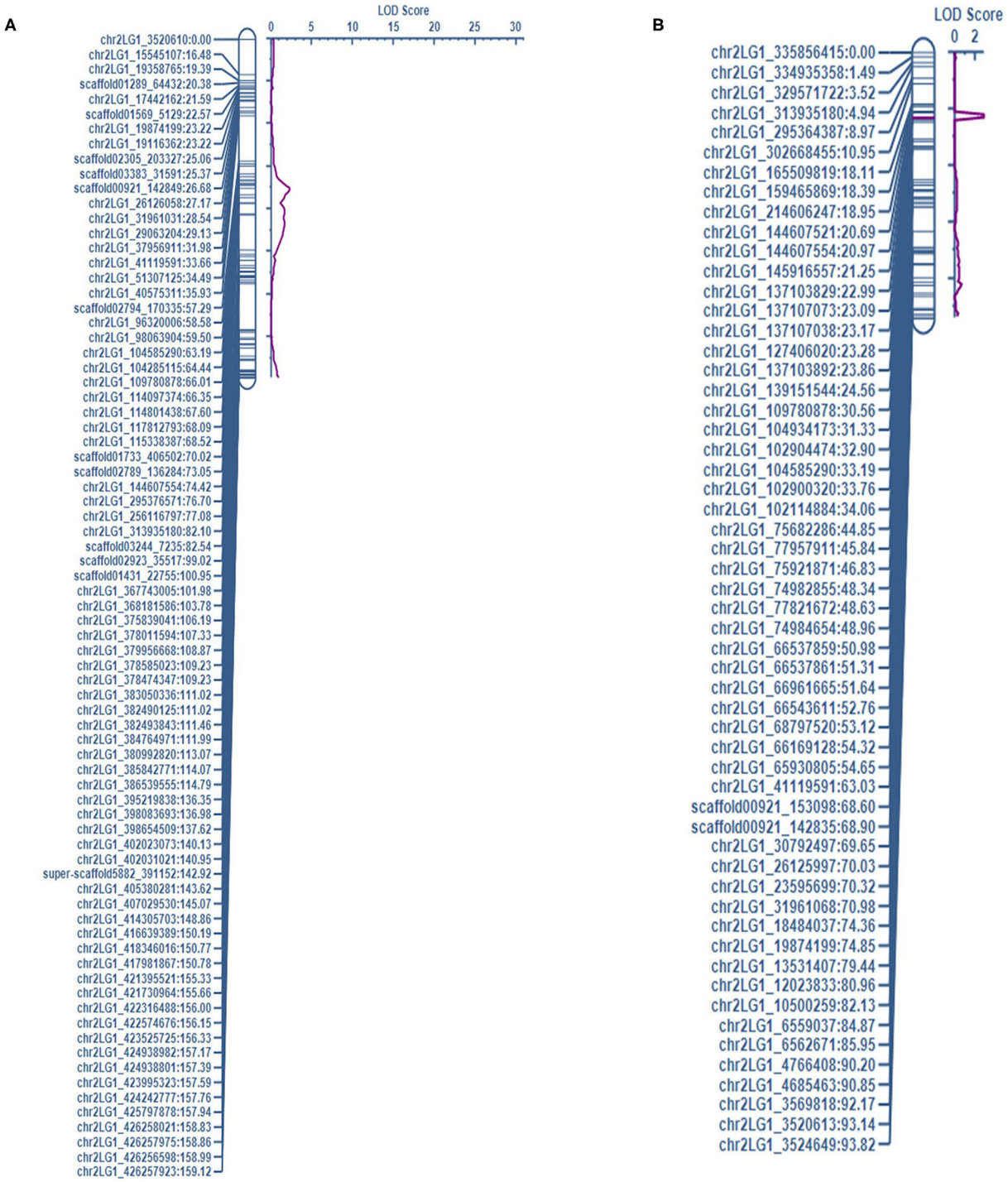


FIGURE 2 | The QTL associated with NTI on chr2LG1 were detected in the complete dataset of **(A)** PRIL17 (*QNTI.17.ndsu.2*) and, **(B)** PRIL19 (*QNTI.19.ndsu.2*) populations.

being located at the position of the Mendel internode length (*Le*) gene.

The most significant QTL identified at 296 cM on chr5LG3(a), was associated with NTI, LEIP and SR in the complete data set as well as the long internode subset. This QTL was located at the same position as the height (*Le*) gene with LOD= 20.4 which explained up to 50.4% of the phenotypic variance (Weeden et al., 1998; Lee et al., 2017). The PI169603 alleles contributed

TABLE 4 | QTL for reaction to white mold infection identified in PRIL19 based on all individuals, short internode, and long internode genotype subsets.

QTL and trait	Dpi[a]	Chr[b]	Position (cM)	Other associated traits	Dataset type[c]	Left marker	Right marker	LOD	R^2 [e]	Add[f]
HEIGHT										
*Le.ndsu**		chr5LG3(a)	0-5	LEI, AUDPC	All, short	chr5LG3_569106574[d]	chr5LG3_561561067	21.5	61.5	83.3
*Le.***		chr5LG3(a)	296	NTI, LEIP, SR	All	chr5LG3_566442038	chr5LG3_566847213	23.6	64.3	0.4
LEI (cm)										
*QLEI.19.ndsu.5.1**	3,7,14	chr5LG3(a)	0-5	*Le.ndsu*, AUDPC	All	chr5LG3_569106574	chr5LG3_561561067	9.3	34.1	37.7
*QLEI.19.ndsu5.2****	7	chr5LG3(a)	150-152	AUDPC	Short	chr5LG3_139674536	chr5LG3_138696064	3.2	22.4	−7.5
QLEI.19.ndsu.4	3	chr4LG4	174-184		Short	chr4LG4_52663612	scaffold01646_102677	2.3	19	4.6
QLEI.19.ndsu.7	7	chr7LG7	191-192		Short	chr7LG7_6677333	chr7LG7_272056	4.2	27.4	−8.3
NTI										
QNTI.19.ndsu.2	14	chr2LG1(a)	22-45		All, tall	chr2LG1_145916557	chr2LG1_77957911	2.8	13.0	0.3
*QNTI.19.ndsu.5***	7,14	chr5LG3(a)	296-297	*Le*, LEIP, SR	All, tall	chr5LG3_568430003	chr5LG3_569648908	20.3	50.4	0.7
AUDPC										
*QAUDPC.19.ndsu.5.1****		chr5LG3(a)	167	LEI	Short	chr5LG3_148081339	chr5LG3_167547321	2.7	21.8	−122.4
*QAUDPC.19.ndsu.5.2**		chr5LG3(a)	0-6	*Le.ndsu*	All	chr5LG3_569106574	chr5LG3_561561067	16.1	45.7	410.2
SR										
*QSR.19.ndsu.5***		chr5LG3(a)	297	*Le*, NTI, LEIP	All	chr5LG3_568430003	chr5LG3_569648908	14.7	42.6	−11.2
LEIP										
*QLEIP.17.ndsu.5***		chr5LG3(a)	297	*Le*, NTI, SR	All	chr5LG3_568430003	chr5LG3_569648908	14.0	40.9	11.1

[a] *Number of days relative QTL was found post inoculation of the lesion expansion inhibition or nodal transmission inhibition traits.*
[b] *Pea chromosome number, chr refers to chromosome number of pea genome assembly v1a and LG refers to linkage group of earlier genetic mapping studies (Kreplak et al., 2019).*
[c] *Type of datasets where relative QTL was found. all: All the genotypes in dataset were used for QTL analysis; short: short subset of the dataset was used for QTL analysis; tall: tall subsets of dataset were used for QTL analysis.*
[d] *In each SNP ID, chr refers to chromosome number of pea genome assembly v1a and LG refers to linkage group of earlier genetic mapping studies followed by the base pair position. For non-chromosomal SNPs, scaffold followed by the scaffold number and base pair position (Kreplak et al., 2019).*
[e] *Phenotypic variation explained by QTL.*
[f] *Additive effect.*
*, **, ***: *QTL with the same markings are the same.*

to increased value of NTI, LEIP and SR at all the identified loci (**Table 5**). Another significant common QTL between LEI and AUDPC was identified at 0-6 cM on chr5LG3(a) in the complete datasets. This QTL was located at the same position as the height QTL (*Le.ndsu*) and explained 45.7% of the phenotypic variation. Interestingly, non-adapted parent, PI169603 conferred alleles for increased height at *Le.ndsu* (**Table 5**).

For LEI, a QTL was identified in the short internode subset at 148-190 cM on chr5LG3(a) (*QLEI.19.ndsu5.2*) and explained 22.4% of phenotypic variation. The desirable allele for improving LEI were contributed through PI169603 (**Table 5**). Another QTL for LEI was found in short internode subset at 191–192 cM of chr7LG7 (*QLEI.19.ndsu.7*) (**Figure 1B**). This QTL explained 27.4% of the phenotypic variation with LOD = 4.2 which PI169603 provided alleles for improving LEI at all the identified loci. Also, one minor QTL for LEI (*QLEI.19.ndsu.4*) and one minor QTL for NTI (*QNTI.19.ndsu.2*) were identified in different datasets of PRIL19 (**Figure 2B**).

The PRIL19-16, PRIL19-101, PRIL19-124, PRIL19-127, PRIL19-135 possessed the desirable alleles for improving trait value for *QLEI.19.ndsu.5.1*, *QLEI.19.ndsu.5.2*, and *QLEI.19.ndsu.7* loci (BB+ AA+ AA). Also, the positive alleles for increased traits value for *QLEI.19.ndsu.5.1*, *QLEI.19.ndsu.5.2*, *QLEI.19.ndsu.7*, and *QLEI.19.ndsu.4* loci (BB+ AA+ AA+ BB) were carried by PRIL19-104.

Identification of Candidate Genes Associated With Resistance to White Mold

Genome browsing of flanking marker positions of QTL associated with white mold resistance with the *Pisum sativum* sequence revealed forty genes involved in plant defense mechanisms (**Table 6**).

DISCUSSION

Breeding white mold resistant pea varieties is challenging due to several reasons. There is limited insight into genes related to the pathogen-host interaction (pathogenicity and resistance genes) due to the absence of high levels of resistance in pea germplasm. The *S. sclerotiorum* pathogen infects a broad host range including important crops such as soybean, bean, canola, and sunflower, with different mechanisms and there is no report of complete resistance in any crop (Bolton et al., 2006). Therefore, in this study, we characterized two pea populations to identify partial resistance genotypes and dissect the genetics associated with white mold resistance.

Analyzing a detached stem assay to identify potential sources of white mold resistance in pea plants and comparision with the a whole plant assay in a previous study (Ashtari Mahini, 2018) indicated that pathogen might behave differently

TABLE 5 | Parental alleles related to increased and decreased trait values and observed significant mean differences in the individuals* for phenotypic trait values when a specific parental allele exists at QTL identified in this study.

QTL	Related trait	Allele for increased trait value	Trait value	Allele for decreased trait value	Trait value	*T*-test for mean difference (*P* value)
PRIL17						
Le	Height	BB	358.2	AA	125.6	1.12916E-27
QNTI.17.ndsu.1	NTI(14 dpi)	AA	2.5	BB	1.9	0.002264
QLEI.17.ndsu.2	LEI(7 dpi)	AA	52.6	BB	74.9	0.003389
QNTI.17.ndsu.2	NTI(14 dpi)	BB	2.4	AA	1.9	0.014825
QLEI.17.ndsu.4	LEI(3 dpi)	AA	36.1	BB	41.3	0.002767
QLEI.17.ndsu.5	LEI(3 dpi)	AA	34.6	BB	42.7	2.27648E-06
	LEI(7 dpi)	AA	62.3	BB	112.3	2.63E-15
	LEI(14 dpi)	AA	91.3	BB	176.7	7.32092E-14
	NTI(7 dpi)	BB	3.3	AA	2.4	1.95E-12
	NTI(14 dpi)	BB	2.8	AA	1.4	5.98E-16
	LEI%	BB	54.2	AA	75	5.69E-11
	SR%	BB	46.1	AA	27.2	3.43442E-10
	AUDPC	AA	1049.8	BB	1912.5	1.42E-19
QLEI.17.ndsu.6	LEI(7 dpi)	AA	55.5	BB	72.8	0.0171
	AUDPC	AA	946.4	BB	1217.6	0.019231
QLEI.17.ndsu.7	LEI(3 dpi)	BB	37.5	AA	43.3	0.002568
	LEI(7 dpi)	BB	84.2	AA	99.4	0.028224
QLEI.17.ndsu.4+ QLEI.17.ndsu.7	LEI(3 dpi)	AA+ BB	33.1	BB+ AA	46.3	0.001287
QLEI.17.ndsu.2+ QLEI.17.ndsu.6	LEI(3 dpi)	AA+ AA	48	BB+ BB	86.5	0.003701
	AUDPC	AA+ AA	792.7	BB+ BB	1435.3	0.002521
PRIL19						
Le	Height	AA	275.3	BB	148.6	1.37315E-06
Le.ndsu	Height	AA	298.1	BB	133.7	1.66032E-21
QNTI.19.ndsu.2	NTI(14 dpi)	AA	1.8	BB	1.1	0.031602
QLEI.19.ndsu.4	LEI(3 dpi)	BB	37.4	AA	43.8	0.053682
QLEI.19.ndsu.5.1	LEI(3 dpi)	BB	44.6	AA	58.1	1.53E-05
	LEI(7 dpi)	BB	86.5	AA	127	6.35E-08
	LEI(14 dpi)	BB	120.8	AA	193.5	2.76E-10
	AUDPC	BB	1390.3	AA	2176.2	2.04E-14
QLEI.19.ndsu.5.2	LEI(7 dpi)	AA	72.3	BB	83.9	0.004942
	AUDPC	AA	1215.9	BB	1419.7	0.002687
QNTI.19.ndsu.5	NTI(14 dpi)	AA	1.7	BB	0.3	4.57393E-24
	LEI%	AA	71.1	BB	94.2	2.26713E-18
	SR%	AA	28.6	BB	5.8	1.24479E-17
QLEI.19.ndsu.7	LEI(7 dpi)	AA	69.6	BB	87.8	0.002838

**Only the individuals showing a clear genotype for all the associated markers at the target QTL were used for the analysis.*
AA = Lifter allele; BB = PI240515 allele (PRIL17).
AA = PI169603 allele; BB = Medora allele (PRIL19).

in detached stems and the result might be associated with senescence other than defense pathways. We observed negative phenotypic correlation between LEI and NTI similar to Porter et al. (2009). This suggests different mechanisms are operating in the resistant reaction as reported in GWAS (genome-wide association study) analysis and RNA sequencing of the pea-*S. sclerotiorum* interaction (Chang et al., 2018). The RNA sequence analysis of Lifter and PI240515 at 12, 24, and 48 h post inoculation showed that more leucine rich-repeat containing transcripts and oxidoreductase transcripts were associated with greater lesion resistance, while VQ (Valine-glutamine) motif-containing proteins and a myo-inositol oxygenase increased for nodal resistance (Chang et al., 2018).

A small number of pea genotypes restricted lesion advancement in this study, but most of them did not survive after 14 days. This indicates that NTI should be combined with LEI to prevent lesion development through the stem (Porter et al., 2009). The lower AUDPC values indicate slower disease development and greater resistance to the disease compared to

TABLE 6 | Genes found from genome browsing of flanking markers associated with resistance to white mold in *Pisum sativum* v1a JBrowse (Kreplak et al., 2019).

QTL	SNP	Overlapping genes	Gene ID[b] *Pisum sativum* v1a	GO ID[c]
PRIL17				
QNTl.17.ndsu.1	chr1LG6_27014318[a]	Leucine rich repeats (2 copies)	Psat1g019120	0030275
	chr1LG6_27820673	Amino acid permease	Psat1g019680	0015171
QLEl.17.ndsu.2	chr2LG1_423525725	PH domain profile	Psat2g185280	0042731
	chr2LG1_424938982	KH domain	Psat2g187240	
	chr2LG1_425797878	Protein kinase domain profile	Psat2g189080	
	chr2LG1_426258021	DYW family of nucleic acid deaminases	Psat2g189520	
QNTl.17.ndsu.2	chr2LG1_115338387	Ethylene-responsive protein kinase Le-CTR1	Psat2g065120	
	scaffold02789_136284	Metal-dependent phosphohydrolase+ HD subdomain	Psat0s2789g0080	
QLEl.17.ndsu.4	chr4LG4_25831962	Ligase activity+ forming carbon-nitrogen bonds	Psat4g018040	0016879
	scaffold01160_24125	IQ calmodulin-binding motif	Psat0s1160g0040	0005516
QLEl.17.ndsu.5	chr5LG3_566156376	Hsp70 protein	Psat5g299000	0030544
	chr5LG3_570275258	C2 domain	Psat5g302040	
	chr5LG3_554216613	MULE transposase domain	Psat5g288200	
	chr5LG3_562563492	Taurine catabolism dioxygenase TauD+ TfdA family	Psat5g295640	
	chr5LG3_566847213	Cyclin+ N-terminal domain	Psat5g299560	
	chr5LG3_570780892	Hpt domain	Psat5g302360	0009927
	chr5LG3_569106574	Ribosomal protein S11	Psat5g300640	0003735
	chr5LG3_557023279	Rab-GTPase-TBC domain	Psat5g290800	0017137
	chr5LG3_579082448	Tub family	Psat5g308240	
QLEl.17.ndsu.6	chr6LG2_163685348	Zinc finger+ C3HC4 type (RING finger)	Psat6g099280	0071535
QLEl.17.ndsu.7	chr7LG7_334919307	UEV domain	Psat7g177600	
	chr7LG7_334426809	Enoyl-(Acyl carrier protein) reductase	Psat7g177080	0016631
	chr7LG7_186100415	Clathrin light chain	Psat7g112920	0071439
	chr7LG7_170686568	Reverse transcriptase-like	Psat7g105080	
PRIL19				
QNTl.19.ndsu.2	chr2LG1_75682286	Seven in absentia protein family	Psat2g042720	
	chr2LG1_77957911	Squalene/phytoene synthase	Psat2g044520	0051996
	chr2LG1_102114884	Ankyrin repeats (3 copies)	Psat2g058000	0005515
	chr2LG1_137103829	Cellular nucleobase+ nucleoside+ nucleotide and nucleic acid metabolic process	Psat2g073040	0006139
QLEl.19.ndsu.4	scaffold01646_102677	Protein of unknown function (DUF789)	Psat0s1646g0040	
QLEl.19.ndsu.5.1	chr5LG3_569648908	Polypeptide deformylase	Psat5g301200	0042586
	chr5LG3_568430003	Ion transport protein	Psat5g077720	0006811
	chr5LG3_566442038	ESCO1/2 acetyl-transferase	Psat5g299360	0007062
	chr5LG3_566847213	Cyclin+ N-terminal domain	Psat5g299560	
QLEl.19.ndsu5.2	chr5LG3_99955442	Hydrolase activity+ acting on acid anhydrides+ catalyzing transmembrane movement of substances	Psat5g055560	0016817
	chr5LG3_148081339	F-box-like	Psat5g082440	
	chr5LG3_167547321	GDP-fucose protein O-fucosyltransferase	Psat5g091800	0008417
QNTl.19.ndsu.5	chr5LG3_561561067	Helix-loop-helix DNA-binding domain	Psat5g294600	
QLEl.19.ndsu.7	chr7LG7_272056	Rhodanase C-terminal	Psat7g000320	
	chr7LG7_6677333	Elongation factor Tu GTP binding domain	Psat7g004080	

[a] *In each SNP ID, chr refers to chromosome number of pea genome assembly v1a and LG refers to linkage group of earlier genetic mapping studies (Tayeh et al., 2015) followed by the base pair position. For non-chromosomal SNPs, scaffold followed by the scaffold number and base pair position.*
[b] *Gene IDs were provided by (Kreplak et al., 2019).*
[c] *Gene Ontology (GO) ID refers to the statement about the molecular function of a particular gene (Binns et al., 2009).*

larger AUDPC values (Kull et al., 2004). However, this was not always true in our experiment.

Although, there is no direct relationship between stem strength and disease resistance to white mold (Porter, 2012; Ashtari Mahini, 2018), incorporating physical characters such as stem thickness, short internode and the afila leaf type can help develop genetically resistant genotypes which are commercially acceptable (Porter, 2012).

The inheritance of resistance to white mold is partial and quantitative as demonstrated by the distribution of phenotypic data for disease resistance and several observed putative QTL in this study. It seems that physiological resistance and morphological avoidance are the components of the partial resistance of pea to white mold. Plant height is most likely associated with disease escape. In soybean varieties, response to *S. sclerotiorum* also depends on disease escape due to height, maturity and reduced lodging (Boland and Hall, 1986). Overall, it may be better and more reasonable to select QTL for physiological resistance mechanisms which contribute to restrict the development of the pathogen in pea.

Although, there have been some reports of QTL identification for *S. sclerotiorum* resistance in sunflower (Micic et al., 2005), soybean (Bastien et al., 2014; Iquira et al., 2015), bean (Kolkman and Kelly, 2003; Ender and Kelly, 2005; Miklas, 2007), rapeseed (*Brassica napus*) (Wu et al., 2013), and *Brassica oleracea* (Mei et al., 2013), there is only one such report in pea (Tashtemirov, 2012). Until now, the only QTL mapping study for resistance to *S. sclerotiorum* in pea was conducted using F_2 lines from the cross between Lifter and PI240515. Two QTL (linked to SSR markers AA255 and AD73) were found on LG2 and LG3 related to NTI and LEI, respectively (Tashtemirov, 2012). We verified QTL associated with LEI on chr5LG3 in all three data type sets of each of the populations. Compared to earlier report (Tashtemirov, 2012), we observed more QTL in our study. Higher number of QTL identified in this study could be attributed to better experimental design and higher number of recombination events in RILs compared to F_2 families.

There were thirteen QTL identified which three of them were in the same loci as height gene (*Le*) and the other ten QTL associated with two forms of white mold resistance. Seven QTL contributed in reducing the rate of lesion development with LEI (LEIP, AUDPC, and SR considered as LEI subsets). Three QTL restricted development of the pathogen through the node (NTI) in the pea stem. The QTL associated with LEI on chr7LG7 was detected in the short internode subset of both populations (**Figure 1**). Likewise, the QTL associated with NTI on chr2LG1 were detected in the complete dataset of both populations (**Figure 2**).

The sequence analysis of the flanking markers of QTL associated with white mold resistance identified candidate genes that may be involved in resistance. We observed different candidate resistance genes for LEI and NTI. Candidate genes such as protein kinase, elongation factors, ion transport, C2 domain, heat shock protein 70, F box protein, and RING (Really Interesting New Gene) zinc finger were related to LEI, and genes such as ethylene-responsive protein, helix-loop-helix DNA-binding domain, ankyrin repeats, and leucine rich repeats (LRR) were associated to NTI. This further indicates the possibility of different genetic mechanisms underlying LEI and NTI as has been observed earlier (Chang et al., 2018). However, in contrast to the observations by Chang et al. (2018), we found that the LRR protein related to NTI and protein kinase was associated with LEI in our study.

The sequence analyses found that our QTL correspond to some candidate genes that have earlier been reported in transcriptome or proteomics analysis of various hosts during *S. sclerotiorum* infection. For example, *EF-Tu* (elongation factor thermo unstable) plays important roles in providing stress adaptation and is down regulated in the proteomics analysis of pea plants infected with *S. sclerotiorum* (Jain et al., 2015). The ion transport proteins provide a channel for Ca^{2+} flow across membrane and enable Ca^{2+} elevation in response to pathogen signals (Wang et al., 2019). On the other hand, C2 domain of *B. napus* interacts with the polygalacturonases (PGs) effector of *S. sclerotiorum* and involves in calcium signaling processes during plant infection (Wang et al., 2009). Ethylene-responsive protein and KH domain are involved in hormonal response ethylene and jasmonate pathway, respectively (Zhao et al., 2007; Thatcher et al., 2015). Protein kinase domain and genes encoding transcription factors such as RING zinc finger protein, and helix-loop-helix DNA-binding domain play a part in signal pathway (Zhao et al., 2007). Leucine rich repeats protein contribute to R-gene based resistance and differentially expressed during *S. sclerotiorum* and *B. napus* interaction (Zhao et al., 2007; Wei et al., 2016; Wu et al., 2016). Heat shock protein 70 (Hsp70), F box like protein, and ankyrin repeat involve in abiotic stress, protein degradation and metabolism, respectively and were induced during *S. sclerotiorum* invasion in *B. napus* (Zhao et al., 2007, 2009).

For pre-breeding purposes, we were able to find some partial resistance RILs with a combination of desirable alleles for important loci from both adapted and non-adapted parents. These lines include PRIL17-141, PRIL17-145, PRIL17-166, PRIL19-124, and PRIL19-127. These germplasm and markers developed in this study will facilitate marker-assisted selection for white mold resistant cultivars in pea.

CONCLUSION

In this study, two RIL populations were developed to study the genetics of white mold resistance in pea. Extensive phenotyping combined with genotyping by sequencing, identified a total of ten QTL associated with white mold resistance. Seven of those QTL were associated with LEI and three QTL were associated with NTI. About 50% of the beneficial QTL alleles for increased LEI and NTI were contributed by adapted parents in both populations which indicate the transgressive segregation due to combination of superior alleles. The non-native parents also contributed important loci for LEI and NTI and provide the opportunity to improve the adapted gene pool for white mold resistance. This further proved the importance of wild germplasm to broaden the gene pool and introgress lost beneficial genetic diversity to cultivated germplasm.

Twenty-two inbred lines with short plant stature were identified showing partial resistance to white mold and met at least two of the resistance criteria in the greenhouse evaluation as proposed by Porter et al. (2009). The partial resistant RILs having combinations of valuable alleles from both parents at identified loci could be important material for breeding programs to enhance resistance traits. To conclude, the pre-breeding material, genomic and genetic resources developed in this study could

be exploited through marker-assisted breeding programs for developing white mold resistant cultivars in pea.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

RA conceptualized and carried out the experiments on the material that KM and LP provided for the project. RA analyzed and wrote the manuscript with support from AK and JF who helped to analyse the data. KM and EE helped supervise and conceptualize the project. AK, JF, LP, EE, and KM also helped to review and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by National Sclerotinia Initiative, Grant/Award Number: 58-5442-9-241; USDA-ARS.

ACKNOWLEDGMENTS

This research was supported by National Sclerotinia Initiative-USDA-ARS. We would like to thank Dr. Seyed Mostafa Pirseyedi, Justin Hegstad, Dr. Xuehui Li, Dr. Cliff Hall, Dr. Francois Marais, and Dr. Thomas Walk for generously assisting us to complete this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.587968/full#supplementary-material

REFERENCES

Ahn, S. J., Costa, J., and Rettig Emanuel, J. (1996). PicoGreen quantitation of DNA: effective evaluation of samples pre-or psost-PCR. *Nucleic Acids Res.* 24, 2623–2625. doi: 10.1093/nar/24.13.2623

Ashtari Mahini, R. (2018). Analysis and identification of QTL for resistance to *Sclerotinia sclerotiorum* in pea *(Pisum sativum L.)*. [Doctoral dissertation]. North Dakota State University, Fargo, ND, United States.

Bastien, M., Sonah, H., and Belzile, F. (2014). Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. *Plant Genome.* 7:30. doi: 10.3835/plantgenome2013.10.0030

Bazzalo, M. E., Dimarco, P., Martinez, F., and Daleo, G. R. (1991). Indicators of resistance of sunflower plant to basal stalk rot (*Sclerotinia sclerotiorum*): symptomatological, biochemical, anatomical, and morphological characters of the host. *Euphytica*. 57, 195–205.

Beji, S., Fontaine, V., Devaux, R., Thomas, M., Negro, S. S., Bahrman, N., et al. (2020). Genome-wide association study identifies favorable SNP alleles and candidate genes for frost tolerance in pea. *BMC Gen.* 21, 1–21. doi: 10.1186/s12864-020-06928-w

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., and Apweiler, R. (2009). QuickGO: a web-based tool for gene ontology searching. *Bioinformatics*. 25, 3045–3046. doi: 10.1093/bioinformatics/btp536

Blanchette, B., and Auld, D. L. (1978). Screening field peas for resistance to white mold. *Crop Sci.* 18, 977–978.

Boland, G. J., and Hall, R. (1986). Growth room evaluation of soybean cultivars for resistance to *Sclerotinia sclerotiorum*. *Canadian J Plant Sci.* 66, 559–564.

Bolton, M. D., Thomma, B. P., and Nelson, B. D. (2006). *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. *Mol Plant Pathol.* 7, 1–16. doi: 10.1111/J.1364-3703.2005.00316. X

Bordat, A., Savois, V., Nicolas, M., Salse, J., Chauveau, A., Bourgeois, M., et al. (2011). Translational genomics in legumes allowed placing *in silico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *G3: Gene Genom Genet.* 1, 93–103. doi: 10.1534/g3.111.000349

Chang, H. X., Sang, H., Wang, J., McPhee, K. E., Zhuang, X., Porter, L. D., et al. (2018). Exploring the genetics of lesion and nodal resistance in pea (*Pisum sativum* L.) to *Sclerotinia sclerotiorum* using genome- wide association studies and RNA- Seq. *Plant Direct.* 2:e00064. doi: 10.1002/pld3.64

Davar, R., Darvishzadeh, R., and Majd, A. (2013). Changes in antioxidant systems in sunflower partial resistant and susceptible lines as affected by *Sclerotinia sclerotiorum*. *Biologia*. 68, 821–829. doi: 10.2478/s11756-013-0235

Dissanayaka, D. N., Gali, K. K., Jha, A. B., Lachagari, V. R., and Warkentin, T. D. (2020). Genome-wide association study to identify single nucleotide polymorphisms associated with Fe, Zn, and Se concentration in field pea. *Crop Sci.* 60, 2070–2084. doi: 10.1002/csc2.2016

Diversity Arrays Technology (2008). Plant DNA Extraction Protocol for DArT. Canberra, Australia, *DArT PL*.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379

Ender, M., and Kelly, J. D. (2005). Identification of QTL associated with white mold resistance in common bean. *Crop Sci.* 45, 2482–2490. doi: 10.2135/cropsci2005.0064

Fernando, W. G. D., Nakkeeran, S., and Zhang, Y. (2004). Ecofriendly methods in combating *Sclerotinia sclerotiorum* (Lib.) de Bary. *Recent Res. Devel. Environ. Bio.* 661, 329–347. Available online at: https://www.researchgate.net/publication/238111476

Gali, K. K., Sackville, A., Tafesse, E. G., Lachagari, V. B., McPhee, K. E., Hybl, M., et al. (2019). Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.). *Front Plant Sci.* 10:1538. doi: 10.3389/fpls.2019.01538

Grain Legume Genetics and Physiology Research Unit (2007). *Dry Pea, Lentil, Chickpea and Winter Legume Breeding 2007 Progress Report*. Pullman, WA: USDA. Available online at: http://sp.arsnet.usda.gov/ARSUserFiles/20902000/WinterLegumeProgressReport2007/WinterLegumeProgressReport2007.pdf.

Iquira, E., Humira, S., and François, B. (2015). Association mapping of QTLs for Sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biol.* 15:5. doi: 10.1186/s12870-014-0408-y

Jain, A., Singh, A., Singh, S., Singh, V., and Singh, H. B. (2015). Comparative proteomic analysis in pea reated with microbial consortia of beneficial microbes reveals changes in the protein network to enhance resistance against *Sclerotinia sclerotiorum*. *J. Plant Physiol.* 182, 79–94. doi: 10.1016/j.jplph.2015.05.004

Jain, A., Singh, S., Sarma, B. K., and Singh, H. B. (2012). Microbial consortium-mediated reprogramming of defence network in pea to enhance tolerance against *Sclerotinia sclerotiorum*. *J Appl Microbiol.* 112, 537–550. doi: 10.1111/j.1365-2672.2011. 05220.x

Khan, M. F., Bhuyian, M. Z. R., Chittem, K., Shahoveisi, F., Haque, M. E., Liu, Y., et al. (2020). First Report of *Sclerotinia sclerotiorum* causing Leaf Blight in Sugar Beet (*Beta vulgaris* L.) in North Dakota, USA. *Plant Dis*. 104, 1258–1258. doi: 10.1094/PDIS-11-19-2304-PDN

Kolkman, J. M., and Kelly, J. D. (2003). QTL conferring resistance and avoidance to white mold in common bean. *Crop Science.* 43, 539–548.

Kraft, J. M., and Pfleger, F. L. (2001). *Compendium of Pea Diseases and Pests, 2nd Ed*. St. Paul, MN: American Phytopathological Society Press (APS Press).

Kreplak, J., Madoui, M. A., Cápal, P., Novák, P., Labadie, K., Aubert, G., et al. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet*. 51, 1411–1422. doi: 10.1038/s41588-019-0480-1

Kull, L. S., Pedersen, W. L., Palmquist, D., and Hartman, G. L. (2004). Mycelial compatibility grouping and aggressiveness of *Sclerotinia sclerotiorum*. *Plant Dis*. 88, 325–332.

Lee, C., Yu, D., Choi, H. K., and Kim, R. W. (2017). Reconstruction of a composite comparative map composed of ten legume genomes. *Genes Genom*. 39, 111–119. doi: 10.1007/s13258-016-0481-8

Manfioletti, G., and Schneider, C. (1988). A new and fast method for prearing high quality lambda DNA suitable for sequencing. *Nucleic Acids Res.* 16, 2873–2884.

McPhee, K. E., and Muehlbauer, F. J. (2002). Registration of Lifter green dry pea. *Crop Sci.* 42:1377. doi: 10.2135/cropsci2002.1377

Mei, J., Ding, Y., Lu, K., Wei, D., Liu, Y., Disi, J. O., et al. (2013). Identification of genomic regions involved in resistance against *Sclerotinia sclerotiorum* from wild *Brassica oleracea*. *Theor. Appl. Genet.* 126:54956. doi: 10.1007/s00122-012-2000-x

Melo, A. T. O., Bartaula, R., and Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping- by-sequencing data. *BMC Bioinform*. 17:29. doi: 10.1186/s12859-016-0879-y

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Micic, Z., Hahn, V., Bauer, E., Schön, C. C., and Melchinger, A. E. (2005). QTL mapping of resistance to Sclerotinia midstalk rot in RIL of sunflower population NDBLOS $_{Sel}$×CM625. *Theor Appl Genet.* 110, 1490–1498. doi: 10.1007/s00122-005-1984-x

Miklas, P. N. (2007). Marker-assisted backcrossing QTL for partial resistance to Sclerotinia white mold in dry bean. *Crop Sci.* 47, 935–942. doi: 10.2135/cropsci2006.08.0525

Mueller, D. S., Wise, K., Smith, D., Chilvers, M., and Bradley, C. (2014). *Managing White Mold in Soybean.* Integrated Crop Management News. Iowa State University Extension and Outreach. Available online at: https://crops.extension.iastate.edu/cropnews/2014/07/managing-white-mold-soybean.

Peltier, A. J., Bradley, C. A., Chilvers, M. I., Malvick, D. K., Mueller, D. S., Wise, K. A., et al. (2012). Biology, yield loss and control of Sclerotinia stem rot of soybean. *J Integr Pest Manage.* 3, B1–B7. doi: 10.1603/IPM11033

Porter, L. D. (2012). Pea germplasm with partial resistance to extends the time required by the pathogen to infect host tissue. *Crop Sci.* 52, 1044–1050. doi: 10.2135/cropsci2011.09.0479

Porter, L. D., Hoheisel, G., and Coffman, V. A. (2009). Resistance of peas to *Sclerotinia sclerotiorum* in the *Pisum* core collection. *Plant Pathol.* 58, 52–60. doi: 10.1111/j.1365-3059.2008.01937.x

Shahoveisi, F., and del Rio Mendoza, L. E. (2020). Effect of wetness duration and incubation temperature on development of ascosporic infections by *Sclerotinia sclerotiorum*. *Plant Dis*. 104, 1811–1816. doi: 10.1094/PDIS-06-19-1304-RE

Simko, I., and Piepho, H. P. (2012). The area under the disease progress stairs: calculation, advantage, and application. *Phytopathology.* 102, 381–389. doi: 10.1094/PHYTO-07-11-0216

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE.* 8:e54603. doi: 10.1371/journal.pone.0054603

Tafesse, E. G., Gali, K. K., Lachagari, V. B., Bueckert, R., and Warkentin, T. D. (2020). Genome-wide association mapping for heat stress responsive traits in field pea. *Int J Mol Sci.* 21:2043. doi: 10.3390/ijms21062043

Tashtemirov, B. (2012). Inheritance of partial resistance to white mold in field pea *(Pisum sativum L.)*. [Master's Thesis]. North Dakota State University, Fargo, ND, United States.

Tayeh, N., Aubert, G., Pilet-Nayel, M. L., Lejeune-Hénaut, I., Warkentin, T. D., and Burstin, J. (2015). Genomic tools in pea breeding programs: status and perspectives. *Front. Plant. Sci.* 6:1037. doi: 10.3389/fpls.2015.01037

Thatcher, L. F., Kamphuis, L. G., Hane, J. K., Onate-Sanchez, L., and Singh, K. B. (2015). The *Arabidopsis* KH-domain RNA-binding protein ESR1 functions in components of jasmonate signalling, unlinking growth restraint and resistance to stress. *PLoS ONE.* 10:e0126978. doi: 10.1371/journal. pone.0126978

USDA-ARS (2016). *National Sclerotinia Research Initiative Strategic Plan for 2017 to 2021.* Available online at: https://www.ars.usda.gov/ARSUserFiles/30000000/WhiteMoldResearch/SI-Strategic-PLan_~2017-2021_v1_0_Jan16.pdf.

Van Ooijen, J. W. (2006). *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma BV.

Wang, X., Li, Q., Niu, X., Chen, H., Xu, L., and Qi, C. (2009). Characterization of a canola C2 domain gene that interacts with PG, an effector of the necrotrophic fungus *Sclerotinia sclerotiorum*. *J. Exp. Bot*. 60, 2613–2620. doi: 10.1093/jxb/erp104

Wang, Z., Ma, L. Y., Cao, J., Li, Y. L., Ding, L. N., Zhu, K. M., et al. (2019). Recent advances in mechanisms of plant defense to *Sclerotinia sclerotiorum*. *Front. Plant Sci.* 10:1314. doi: 10.3389/fpls.2019.01314

Weeden, N. F., Ellis, T. H. N., Timmerman-Vaughan, G. M., Swiecicki, W. K., Rozov, S. M., and Berdnikov, V. A. (1998). A consensus linkage map for *Pisum sativum*. *Pisum Genet*. 30, 1–3.

Wei, L., Jian, H., Lu, K., Filardo, F., Yin, N., Liu, L., et al. (2016). Genome-wide association analysis and differential expression analysis of resistance to Sclerotinia stem rot in *Brassica Napus*. *Plant Biotechnol J.* 14, 1368–1380. doi: 10.1111/pbi.12501

Wu, J., Cai, G., Tu, J., Li, L., Liu, S., Luo, X., et al. (2013). Identification of QTLs for resistance to Sclerotinia stem rot and *BnaC. IGMT5.a* as a candidate gene of the major resistant QTL *SRC6* in *Brassica Napus*. *PLoS ONE* 8:e67740. doi: 10.1371/journal.pone.0067740

Wu, J., Zhao, Q., Yang, Q., Liu, H., Li, Q., Yi, X., et al. (2016). Comparative transcriptomic analysis uncovers the complex genetic network for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. *Sci Rep.* 6:19007. doi: 10.1038/srep19007

Zhao, J., Buchwaldt, L., Rimmer, S. R., Sharpe, A., McGregor, L., Bekkaoui, D., et al. (2009). Patterns of differential gene expression in *Brassica napus* cultivars infected with *Sclerotinia sclerotiorum*. *Mol Plant Pathol.* 10, 635–649. doi: 10.1111/J.1364-3703.2009.00558.X

Zhao, J., Wang, J., An, L., Doerge, R. W.,Chen, Z. J., Grau, C. R., et al. (2007). Analysis of gene expression profiles in response to *Sclerotinia sclerotiorum* in *Brassica napus*. *Planta.* 227, 13–24. doi: 10.1007/s00425-007-0586-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KG declared a past collaboration with one of the authors KM to the handling editor.

Copyright © 2020 Ashtari Mahini, Kumar, Elias, Fiedler, Porter and McPhee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 November 2020
doi: 10.3389/fgene.2020.572975

Identification of QTLs/Defense Genes Effective at Seedling Stage Against Prevailing Races of Wheat Stripe Rust in India

Anjan Kumar Pradhan[1†], Sundeep Kumar[1†], Amit Kumar Singh[1†], Neeraj Budhlakoti[2], Dwijesh C. Mishra[2], Divya Chauhan[1], Shikha Mittal[1], Monendra Grover[2], Suneel Kumar[1], Om P. Gangwar[3], Subodh Kumar[3], Arun Gupta[4], Subhash C. Bhardwaj[3†], Anil Rai[2] and Kuldeep Singh[1]*

[1] Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources, New Delhi, India, [2] Indian Council of Agricultural Research-Indian Agricultural Statistics Research Institute, New Delhi, India, [3] Indian Council of Agricultural Research-Indian Institute of Wheat and Barley Research, Regional Station, Shimla, India, [4] Indian Council of Agricultural Research-Indian Institute of Wheat and Barley Research, Karnal, India

Resistance in modern wheat cultivars for stripe rust is not long lasting due to the narrow genetic base and periodical evolution of new pathogenic races. Though nearly 83 *Yr* genes conferring resistance to stripe rust have been cataloged so far, few of them have been mapped and utilized in breeding programs. Characterization of wheat germplasm for novel sources of resistance and their incorporation into elite cultivars is required to achieve durable resistance and thus to minimize the yield losses. Here, a genome-wide association study (GWAS) was performed on a set of 391 germplasm lines with the aim to identify quantitative trait loci (QTL) using 35K Axiom® array. Phenotypic evaluation disease severity against four stripe rust pathotypes, i.e., 46S119, 110S119, 238S119, and 47S103 (T) at the seedling stage in a greenhouse providing optimal conditions was carried out consecutively for 2 years (2018 and 2019 winter season). We identified, a total of 17 promising QTl which passed FDR criteria. Moreover these 17 QTL identified in the current study were mapped at different genomic locations i.e. 1B, 2A, 2B, 2D, 3A, 3B, 3D, 4B, 5B and 6B. These 17 QTLs identified in the present study might play a key role in marker-assisted breeding for developing stripe rust resistant wheat cultivars.

Keywords: stripe rust, wheat, SNP, GWAS, GAPIT

OPEN ACCESS

Edited by:
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Vijay Gahlaut, University of Delhi, India
Raj K. Pasam, AgriBio, La Trobe University, Australia

***Correspondence:**
Sundeep Kumar
sundeep.kumar@icar.gov.in

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *15 June 2020*
Accepted: *30 September 2020*
Published: *27 November 2020*

Citation:
Pradhan AK, Kumar S, Singh AK, Budhlakoti N, Mishra DC, Chauhan D, Mittal S, Grover M, Kumar S, Gangwar OP, Kumar S, Gupta A, Bhardwaj SC, Rai A and Singh K (2020) Identification of QTLs/Defense Genes Effective at Seedling Stage Against Prevailing Races of Wheat Stripe Rust in India. Front. Genet. 11:572975. doi: 10.3389/fgene.2020.572975

INTRODUCTION

In wheat, stripe rust caused by *Puccinia striiformis* f. sp. *tritici* (Pst) is the most damaging and widely prevalent disease. It causes significant yield losses in almost every part of the world where cool and humid conditions persist during crop season. In India, stripe (yellow) rust is the major disease in North West Plain Zones (NWPZ) especially sub-mountainous parts of Punjab, Haryana, and Western Uttar Pradesh, the major wheat growing regions of India. Occurrence of stripe rust has also been observed frequently in Jammu & Kashmir, Himachal Pradesh, and tarai regions of Uttarakhand (Chen, 2005; Indu and Saharan, 2011). However, a major outbreak of stripe rust was

observed in the North Western Plain Zone (NWPZ) and the Northern Hills Zone (NHZ) of India in 2006 and 2012–13 which caused heavy yield losses (Prashar et al., 2007; Saharan et al., 2013).

Yield losses due to this disease vary from 10 to 40% which depends upon various factors like severity level, susceptibility of cultivars, stage of infection, rate of disease development, and duration of disease. If infection takes place at the seedling stage and the conducive environment persists until maturity, then yield losses may go up to 100% (Afzal et al., 2007).

The presence of large number of pathotypes is the main reason for the epidemic of stripe rust. Out of 140 stripe rust pathotypes known globally, more than 28 pathotypes have been reported in India (Line and Qayoum, 1992; Chen et al., 2010; Bhardwaj et al., 2014; Tomar et al., 2014). As new pathotypes evolve fast, host plants require novel genes to encounter new pathotypes. Although there are more than 83 cataloged *Yr* resistance genes (McIntosh et al., 2017; Li et al., 2020) most of them are all-stage resistance genes with the exception of 22 genes, namely *Yr11-14, Yr16, Yr18/Lr34/Sr57/Pm38/Ltn1, Yr29/Lr46/Sr58/Pm39/Ltn2, Yr30/Lr27/Sr2, Yr36, Yr39, Yr46/Lr67/Sr55/Pm46/Ltn3, Yr52, Yr59, Yr62, Yr68, Yr71, Yr75, Yr77-80,* and *Yr82* (McIntosh et al., 1995; Chen and Kang, 2017; Feng et al., 2018; Nsabiyera et al., 2018; Pakeerathan et al., 2019). In addition, several temporary *Yr* resistance genes and QTL have been identified and reviewed (Rosewarne et al., 2013; Chen and Kang, 2017).

Although rust can be controlled by good agronomical management and fungicides, still cultivation of resistant cultivars is one of the best and most economical options (Wellings, 2011). To develop resistant cultivars at regular intervals, new sources of resistance and new genes/QTLs need to be identified from virgin germplasm lines. GWAS have been found effective for the identification of novel genes/QTLs in germplasm lines, i.e., leaf rust (Kertho et al., 2015), stem rust (Yu et al., 2012; Letta et al., 2013; Laidò et al., 2015), and stripe rust resistance to verify the effect of previously discovered *Yr* genes and QTLs (Tadesse et al., 2013; Zegeye et al., 2014; Naruoka et al., 2015; Bulli et al., 2016; Pasam et al., 2017). Recently, many studies have reported QTLs for stripe rust resistance in wheat using GWAS (Liu et al., 2017a; Muleta et al., 2017a; Yao et al., 2019; Cheng et al., 2020). However, very few GWAS attempts have been made on Indian wheat germplasm lines for identification of stripe rust genes. Most of the landraces used in this study are unexplored and not used in any breeding program. Therefore, the present study was conducted with the aim to identify QTLs/defense genes associated with stripe rust seedling stage resistance through GWAS.

Abbreviations: GWAS, Genome Wide Association Study; SNPs, Single Nucleotide Polymorphisms; MAS, Marker Assisted Selection; QTL, Quantitative Trait Locus; LD, Linkage Disequilibrium; QQ, Quantile-Quantile plot; PCA, Principal Component Analysis; GAPIT, Genomic Association and Prediction Integrated Tool; CMLM, Compressed Mixed Linear Model, MTA, Marker Trait Association; FDR, False Discovery Rate; YR, Yellow Rust; IWGSC, International Wheat Genome Sequencing Consortium; GO, Gene Ontology; BLAST, Basic Local Alignment Search Tool; ANOVA, Analysis of variance; REML, Restricted maximum Likelihood.

MATERIALS AND METHODS

Plant Material, Inoculation, and Phenotypic Scoring

An association mapping panel of 391 wheat germplasm lines which includes 290 Indian landraces/ indigenous germplasm, 24 exotic, and 77 other germplasm lines including advanced breeding lines and some Indian varieties released during 1960 (**Supplementary Table 1**). The association mapping panel was evaluated against four virulent and predominant pathotypes of *P. striiformis tritici* in India. The landraces, indigenous and exotic germplasm lines were collected from the National Gene bank of India located at ICAR-NBPGR, New Delhi. The advanced wheat breeding lines used in this analysis were developed at ICAR-Indian Institute of Wheat and Barley Research (IIWBR) Karnal. All the 391 wheat germplasm lines were evaluated at the seedling stage using a mixture of Pst pathotypes prevalent in India over 2 years (2018 and 2019 winter season) under greenhouse conditions at IIWBR Regional Station, Shimla. Reference lines known to possess specific *Yr* gene/s commonly occurring in wheat were used to confirm the purity of pathotypes (Bhardwaj et al., 2012). A mixture of fresh garden soil and FYM (1:1 ratio), autoclaved at 60°C for 4 h was used for growing wheat plants in plastic pots and aluminum bread pans/trays. In an aluminum bread pan tray, 4–5 seeds of each line were clump planted. Each tray contained 18 lines and a susceptible check (A-9-30-1). In plastic pots 10–12 seedlings were raised. After seeding, the trays/pots were kept in the greenhouse at 22 ± 2°C with proper labels and provided with optimum conditions to ensure normal germination and growth of the seedling.

After a week, wheat seedlings were inoculated with each pathotype separately by atomizing the uredosporic inoculum suspended in non-phytotoxic isoparaffinic oil (commercially known as Soltrol 170 produced by Chevron Phillips Chemicals Asia Pvt. Ltd., Singapore). The inoculated plants were fine sprayed with water and incubated for 24 h in water saturated glass chambers. The plants were then transferred to the greenhouse and dusted with elemental dust powder of Sulfur to avoid the occurrence of powdery mildew infection. All the optimum greenhouse conditions for good plant growth and proper stripe rust infection were provided. The greenhouse temperature was maintained at 15 ± 1°C. The response of host-pathogen interactions was recorded in the form of infection type after 16–18 days of inoculation following Nayar et al. (1997). Disease severity was recorded on 10 plants for each germplasm line. Wheat-rust infection responses (low or high) were recorded on experimental material and differentials (to ascertain the purity of pathotypes) by following McNeal et al. (1971) using a 0–9 scale where, 0 indicates Immune (no observed visible infection), 1 means highly resistant (necrotic/chlorotic flecks appears without sporulation), 2 means resistant (there are necrotic/chlorotic stripes without sporulation) 3 shows moderately resistant (trace sporulation, necrotic/chlorotic stripes are there), 4 represents moderately resistant (light sporulation, necrotic/chlorotic stripes appears), 5 suggests moderately susceptible (intermediate sporulation, necrotic/chlorotic stripes appears on leaves), 6 specify highly

moderate(moderate sporulation, necrotic/chlorotic stripes were termed as resistant), whereas 7 means moderately susceptible (abundant sporulation, necrotic/chlorotic stripes can be observed on the leaves), 8 means susceptible (abundant sporulation, with chlorosis), while 9 is highly susceptible (plant shows abundant sporulation, without chlorosis, were categorized as susceptible).

Phenotypic Data Analysis

To determine genotypic and year variances among pathotypes, Analysis of variance (ANOVA) was performed using SAS v9.3 (software). Frequency distribution of genotypes for different pathotypes under study has also been generated showing the performance of phenotypes. Heritability of the pathotypes infection was estimated using the restricted maximum likelihood (REML) method. For the analysis of 2-years data, pathotype mean, variance, standard deviation, and ranges of each germplasm line was calculated.

Genotyping

Leaves from 15-days old seedlings were collected and standard CTAB protocol was followed for genomic DNA extraction (Doyle, 1990). A total of 391 germplasm lines were genotyped using the 35K Axiom® array (Affymetrix product ID 550524) for wheat having 35,143 SNPs. SNPs having low-quality clustering and minor allele frequency (MAF $\leq$ 10%), across all the genotypes were excluded from the analysis. To assign an exact physical location to each SNP on wheat chromosomes, SNP probe sequences were subjected to BLAST against wheat reference genome RefSeq v1.0 (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies) following default parameters. A total of 19,090 polymorphic SNPs were assigned to an exact physical location on the wheat genome and used for further downstream analysis.

Analyses of Molecular Diversity and Population Structure

A set of 525 random SNP markers distributed across the 21 wheat chromosomes (25 markers per chromosome) were used to determine the population structure. The STRUCTURE v2.3.4 software based on admixture model with correlated allele frequency was applied to categorize sub-population in the current germplasm lines (Pritchard et al., 2000). For Structure, parameters like 20,000 iterations and 50,000 Monte Carlo Markov Chain (MCMC) replicates were set to determine K values in the range of 1–10. For each K value, ten independent structure runs were carried out and further the results were exported to Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) software for determining the most likely number of subpopulations in germplasm lines (Evanno et al., 2005). Further, fixation index (Fst) for subpopulations was estimated from various STRUCTURE runs. Principal components (PC) were also inferred using the Genomic Association and Prediction Integrated Tool (GAPIT) R package to further analyze population sub-structuring and a comparison was made from the results analyzed with STRUCTURE. To further determine the genetic structure of the lines, cluster analysis based on the neighbor joining (NJ) tree algorithm according to shared-allele distance was also performed in TASSELv5.0. The branching pattern in the NJ tree was assessed based on bootstrapping over loci with 1,000 replications, while the consensus bootstrap value was displayed with the help of ITOL program v5 (https://itol.embl.de/).

Linkage Disequilibrium Analysis

Linkage disequilibrium (LD) based on pairwise measures between SNP markers were estimated using TASSELv5.0 (Bradbury et al., 2007). LD can be estimated as squared allele frequency correlation (R^2) between pairs of intra-chromosomal SNPs with known chromosomal position. The background LD in the wheat AM panel was calculated to identify critical distance for LD decay. The average pattern of genome-wide LD decay over physical distance was determined by constructing a scatterplot of R^2 values against the corresponding physical distance among markers. Further, the extent of LD decay was also estimated using the Locally Weighted Scatter-plot Smoother (LOESS) model (Cleveland, 1979). The critical R^2 value that shows the area beyond which LD is due to true physical linkage was determined using 95th percentile of the square root of transformed R^2 data of unlinked markers (Breseghello and Sorrells, 2005). Further, the intersection of LD decay curve was observed at $R^2 = 0.156$ and at 2.5 Mb distance. Therefore, all the significantly associated SNPs (clustered SNPs) falling within this distance were designated as single QTLs.

Genome-Wide Association Analysis

GWAS was conducted using a panel of 19,090 high quality polymorphic SNP markers and disease severity data against four different pathotypes 47S103, 238S119, 46S119, and 110S119 at seedling stage. This analysis was done using the phenotyping data from both the year (2018 and 2019) separately as well as for pooled data. We have presented the GWAS results for the marker using pooled data, however we have also highlighted the markers which were consistent in both years. Marker-trait associations (MTAs) were identified using the compressed mixed linear model (CMLM) (Yu et al., 2006; Zhang et al., 2010) implemented in GAPIT R package (Lipka et al., 2012). CMLM uses the additional information like Principal component (usually three components) and Kinship of population (K), hence it is also called PC3 + K CMLM model. CMLM is compressed form of mixed linear model. The general equation of MLM can be written as follows

$$y = X\beta + Zu + e$$

TABLE 1 | Analysis of variance (ANOVA) for reaction to stripe rust bread wheat.

Stage	Sources	*df*	*F*	Sig
Seedling	Genotypes (G)	390	3.07	***
	Year (Yr)	1	9.44	0.0021
	Race	3	240.67	***
	G*Yr interaction	390	0.51	1.0000

****Significant difference at P < 0.0001.*

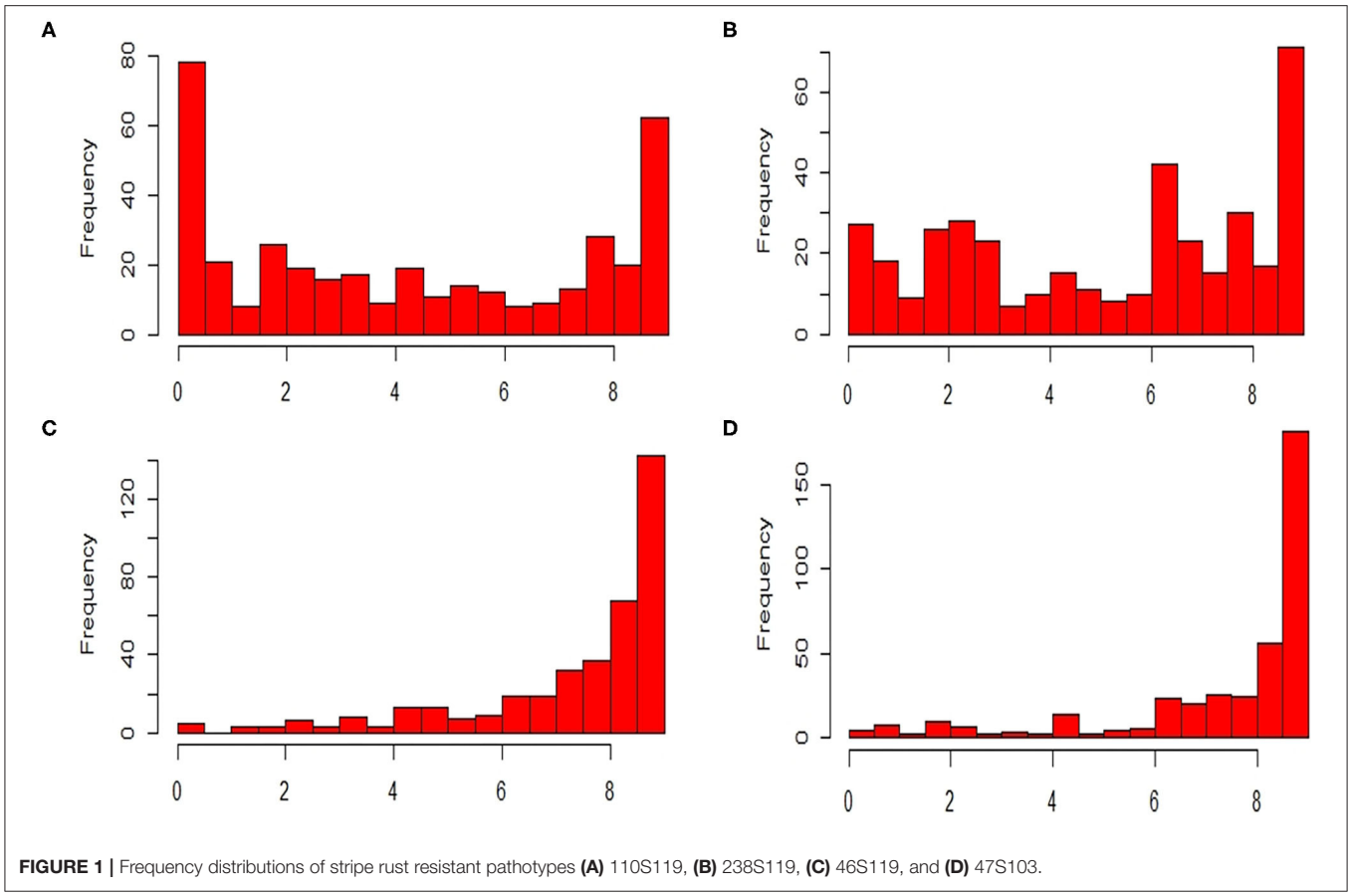


FIGURE 1 | Frequency distributions of stripe rust resistant pathotypes **(A)** 110S119, **(B)** 238S119, **(C)** 46S119, and **(D)** 47S103.

Where, y = observed phenotypic vector, β = vector of marker fixed effects; u = vector of random additive genetic effects from individual lines; e = vector of residuals; X and Z are known design matrices.

P-value and R^2 were used as parameter to identify significant marker-trait associations (MTAs). Significant MTAs were identified at the threshold of $P < 0.001$. In order to show the distribution of SNPs over the chromosome, Manhattan plots have also been generated.

Identification of Putative Candidate Gene and Their Annotation

Candidate genes and their corresponding molecular functions were retrieved from the wheat genome assembly IWGSC Ref-Sequence v1.1 using BLASTN function of BLAST program with default parameters. Associated SNPs were extracted along with their annotations for transcript located within the distance. The identified putative candidate genes were further validated using stripe rust disease resistance data from Sequence Read Archive (SRA) NCBI database (Bio Project—PRJNA613349).

RESULTS

Analysis of Phenotypic Variation for Disease Severity at Seedling Stage

The ANOVA results revealed highly significant variation ($P < 0.0001$) for germplasm lines and pathotypes (**Table 1**) while the genotype × year interaction effect was found to be non-significant. Further, the frequency distribution of the infection types (ITs) produced by the four different Pst races (a) 110S119, (b) 238S119, (c) 46S119, and (d) 47S103 in association mapping panel is shown in **Figures 1A–D**. Among 391 germplasm lines, 35 (10%), 40 (11%), 144 (37%), and 189 (49%) showed (**Supplementary Table 2**) seedling resistance to races 110S119, 238S119, 46S119, and 47S103, respectively in 1st year and 33 (9%), 37 (10%), 144 (37%), 181 (47%), respectively in 2nd year (**Figure 2B**). A higher percentage of germplasm lines were susceptible (IT = 8–9) to races 110S119 (60%), 238S119 (61%), 46S119 (29%), and 47S103 (23%) (**Figure 2A**) in first year as compared to the second year 110S119 (61%), 238S119 (47%), 46S119 (15%), and 47S103 (7%). Further, the heritability (h^2) value for stripe rust infection type (IT) was 0.58 (**Table 2**).

Marker Coverage and Population Structure Analysis

A total of 19,090 SNPs assigned to an exact physical location on the wheat genome were used for association mapping. Out of 19,090 SNPs, 6,083 were mapped on A sub-genome, 7,253 on B sub-genome and 5,754 on D sub-genome. The number of SNPs on individual chromosomes ranged from 378 on 4D to 1,315 on 2B. Chromosome level distribution of SNPs represented that A sub-genome possesses maximum SNPs on 2A (1,079), followed by 1A (1,053) and 7A (953). Whereas, in the B sub-genome, the maximum SNPs were on 2B (1,315), followed by 5B (1,257)

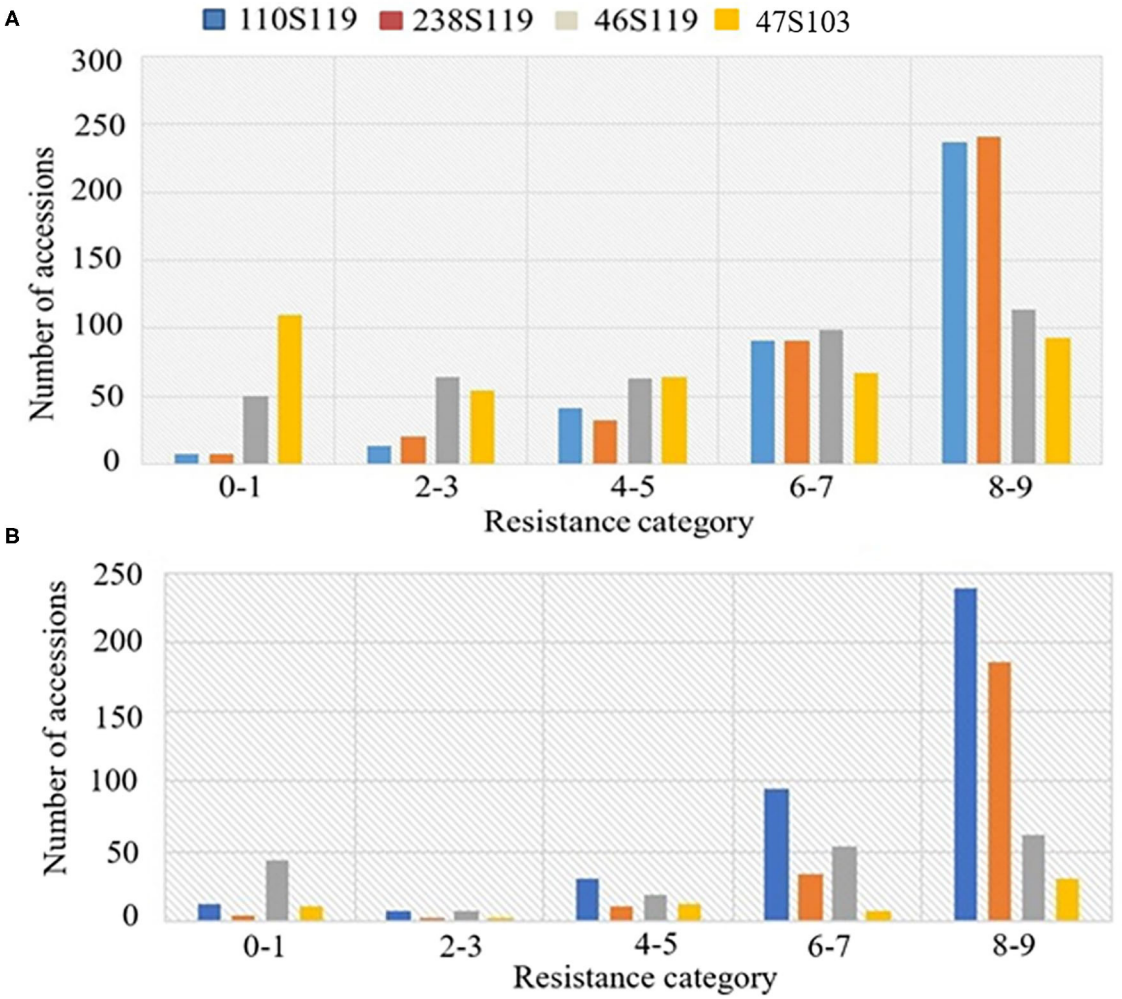


FIGURE 2 | Frequency distributions of stripe rust resistance types produced by seedlings of the bread wheat genotypes **(A)** 2018, **(B)** 2019.

and 1B (1,159). While, in the case of D sub-genome, maximum SNPs were on 2D (1,176) followed by 1D (1,039) and 5D (897) (**Table 3**).

Population structure analysis categorizes the 391 genotypes into four sub-populations (SP). Containing 205, 102, 10, and 74 genotypes, respectively. Sub-populations 1 (SP 1) was the largest among the four sub-populations, having 52.2% of the genotypes from the association mapping panel followed by sub-population 2 (SP 2), sub-population 4 (SP 4), and the smallest one sub-population 3 (SP 3), i.e., 2.5% (**Table 4**). Sub-populations 1 and 4 revealed the highest level of heterozygosity, i.e., 0.3438 and 0.3358, respectively. Individuals of each subpopulation were further categorized as pure and admixtures in type based on membership proportion. Genotypes that had a membership proportion of $\geq$0.8 were considered pure and genotypes $<$0.8 were considered admixtures. Based on this criterion, composition of four sub-populations was as follows; SP 1–60% pure and 40% admixtures; SP 2–36% pure and 64% admixtures; SP 3–40% pure and 60% admixtures and SP 4–37% pure and 63% admixtures

TABLE 2 | Descriptive statistics and heritability of races.

Races	Range	Mean ± SE	Variance	*SD*	Heritability
47S103	0.0-9	4.58±0.16	10.81	3.28	0.58
46S119	0.0-9	5.24±0.15	9.6	3.09	
110S119	0.0-9	7.48±0.09	3.78	1.94	
238S119	0.0-9	7.55±0.1	4.41	2.1	

(**Figure 3A**). Out of four subpopulations, SP 1 consists of a higher proportion of susceptible germplasm lines than others. In SP 1, most of the germplasm lines were susceptible to pathotype 238S119 and highly resistant to pathotype 46S119. On the other hand, in SP 2, most of the germplasm lines were susceptible to pathotype 110S119 and highly resistant to stripe rust pathotype 46S119. For the SP 4, maximum genotypes were resistant against 47S103 followed by 46S119, while susceptibility was for the 238S119 followed by 110S119.

TABLE 3 | Genetic clusters and their member genotypes, proportion of membership, expected heterozygosity, and the mean values of Fst observed from structure analysis of 391 Stripe rust resistance wheat genotypes.

Cluster	Genotypes	% Membership	Expected heterozygosity	Mean fixation index
1	205	0.496	0.3438	0.2958
2	102	0.240	0.1809	0.6170
3	10	0.070	0.0917	0.8336
4	74	0.194	0.3358	0.2385

Principal component analysis (PCA) was also performed using 391 genotypes to estimate population structure including 1st three PCs for further downstream analysis. PC1 and PC2 have explained 32.3 and 12.5% of the genetic variance, respectively. The PC analysis scatter plot (**Figure 3B**) also confirmed the results of population structure analysis as it showed that 1st and 2nd PCs were composed mainly by four sub-populations. Moreover, cluster analysis was carried out based on the neighbor-joining (NJ) algorithm that revealed four clusters in an association panel. Here, the neighbor-joining (NJ) tree of the stripe rust resistance wheat lines was evaluated by shared-allele genetic distance using high-density SNP markers (**Figure 4**).

Linkage Disequilibrium (LD)

The LD for each sub-genome, i.e., A, B, and D genome was estimated from all pairs of SNPs present over there. Individually the average R^2 of genome wide LD was 0.14 for A sub-genome, 0.16 for B sub-genome, and 0.15 for D sub-genome. SNP markers, whose map positions were assigned, were further used to estimate intra-chromosomal LD. A total of 31.46% intra-chromosomal pairs loci were in significant LD (i.e., R^2 of 0.2), while 18,842 SNP pairs were in a perfect LD (i.e., $R^2 = 1$). The decay of LD across the genome is an important parameter that determines the number of significant markers required for performing GWAS analysis. The extent of LD distribution was graphically demonstrated by plotting intra-chromosomal R^2 values for loci against their physical distance and a second-degree LOESS curve was also fitted for further exploration. The background LD in the analyzed AM panel was equal to 0.156 and taken as the threshold cut-off for estimating LD decay. In the selected wheat panel LD decayed the fastest in the A sub-genome comparison to B and D. In the A sub genome R^2 value for the marker pairs reached 0.156 (used as a threshold) at 1.9 Mb as compared to 2.3 Mb in B and 2.9 Mb in D sub-genomes (**Figures 5A–D**).

Genome-Wide Association Analysis

The association of SNP markers with stripe rust resistance against four different pathotypes at seedling stages were determined by CMLM analysis using kinship (K) matrix and population structure (Q matrix) as a covariate. A total of forty QTLs (51 MTAs) in 19 genomic regions (1A, 1B, 2A, 2B, 2D, 3A, 3B, 3D, 4A, 4B, 4D, 5A, 5B, 5D, 6A, 6B, 6D, 7A, and 7D) were linked with resistance to four pathotypes (at $P < 0.001$, and out of these 40 QTLs, 17 QTLs were significant at FDR adjusted $P < 0.30$).

TABLE 4 | Distribution of 19,090 SNPs in 21 chromosomes identified LD in 391 resistant of Stripe rust bread wheat genotypes.

Chr	Size (Mb)	No. of polymorphic SNP	Average number of SNPs per Mb	Chromosome (LD)	No. of marker pairs in perfect LD ($R^2 = 1$)
1A	594.1	1,053	1.77	0.212	1,803
1B	689.85	1,159	1.68	0.303	3,496
1D	495.45	1,039	2.10	0.325	4,523
2A	780.8	1,079	1.38	0.187	1,377
2B	801.26	1,315	1.64	0.160	746
2D	651.85	1,176	1.80	0.205	1,647
3A	750.84	783	1.04	0.122	244
3B	830.83	1,145	1.38	0.158	361
3D	615.55	692	1.12	0.105	118
4A	744.59	652	0.88	0.119	224
4B	673.62	550	0.82	0.108	191
4D	509.86	378	0.74	0.088	89
5A	709.77	828	1.17	0.126	298
5B	713.15	1,257	1.76	0.185	1,340
5D	566.08	897	1.58	0.152	776
6A	618.08	735	1.19	0.154	427
6B	720.99	975	1.35	0.148	300
6D	473.59	725	1.53	0.111	158
7A	736.71	953	1.29	0.129	352
7B	750.62	852	1.14	0.126	200
7D	638.69	847	1.33	0.102	172

These QTLs were distributed as follows; 12 QTLs for 47S103, 12 QTLs for 46S119, 7 QTLs for 110S119, and 9 QTLs for 238S119. Coefficient of determination (R^2) for each QTL was also determined. Phenotypic variance ranged from 6.6 to 28.8%. Ability of the GWAS model was tested using Manhattan plots and quantile-quantile-plots (Q-Q plots) between observed and expected *P*-values of association, that revealed a good fitting for the model with population structure and kinship (**Table 5**, **Figures 6A–D**), i.e., (a) 47S103, (b) 46S119, (c) 110S119, and (d) 238S119.

Twelve QTLs for 47S103 were mapped on 8 different chromosomes (i.e., 1B, 2A, 2B, 2D, 3B, 3D, 4B, and 6B) and most of them were on B sub-genome. Only 1 QTL was mapped on the A sub-genome (*Qyr.stripe-2AL* with *p*-value 2.52E-05 and $R^2 = 0.28$). Seven QTLs were mapped on B sub genome, out of which 2 QTLs, i.e., *Qyr.stripe-2BL.1* and *Qyr.stripe-2BL.2*, were found on the long arm of the chromosome 2B between 760 and 762 Mb (2.17 Mb region). Four QTLs were mapped on D sub-genome (*Qyr.stripe-2DL.1*, *Qyr.stripe-2DL.2*, *Qyr.stripe-3DS.1*, *Qyr.stripe-3DS.2*) at different positions.

For pathotype 46S119, a total of 12 QTLs were identified which mapped on nine different chromosomes (1B, 3A, 3B, 4A, 4D, 5A, 5B, 5D, and 6D), most of them were mapped on B sub-genome. In these QTLs, R^2 ranged from 11.7 to 13.5%. Four QTLs were mapped on A sub-genome (i.e., *Qyr.stripe-3AL*, *Qyr.stripe-3AS*, *Qyr.stripe-4AL*, *Qyr.stripe-5AS*) at various locations. Further, 5 QTLs were mapped on B sub-genome,

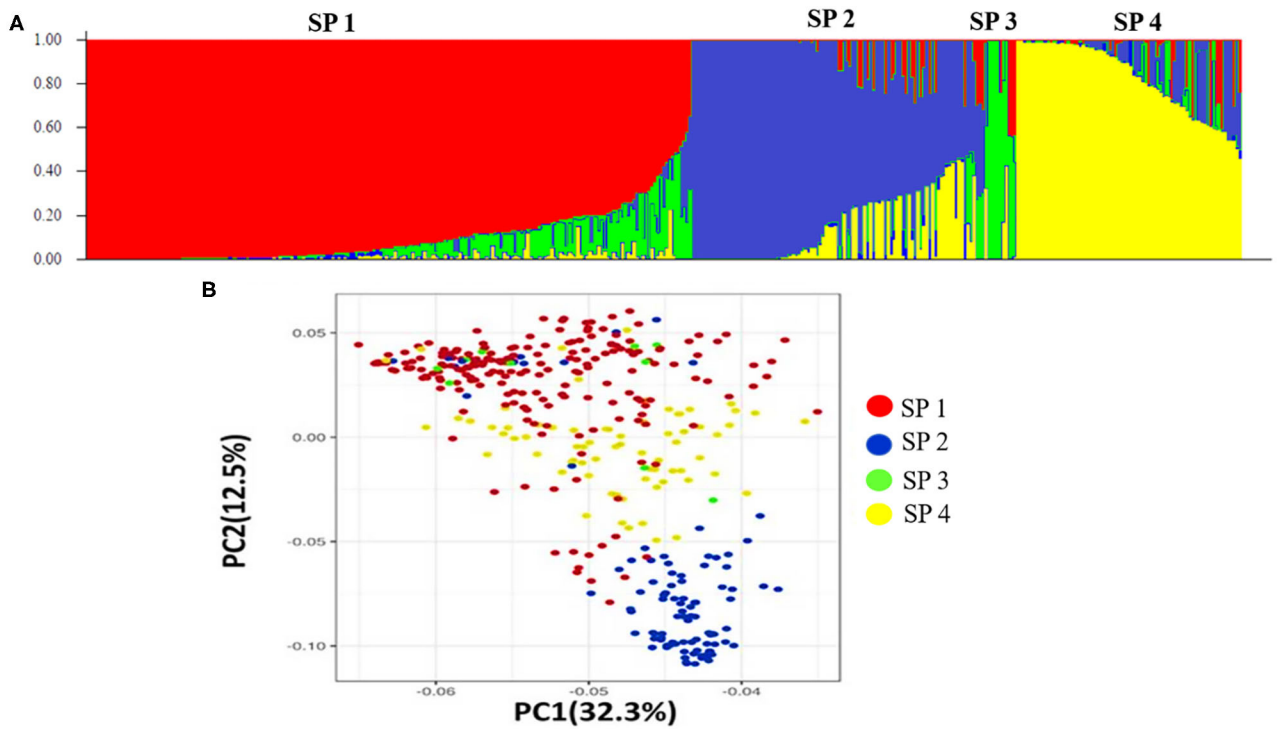


FIGURE 3 | Population structure of the 391 stripe rust germplasm lines combined with markers. **(A)** Population structure of the bread wheat among 19,090 SNP polymorphism markers, **(B)** Principal component analysis.

2 of them, i.e., *Qyr.stripe-3B.1*, *Qyr.stripe-3B.2* were found on 3B at 3 Mb and 542 Mb location as a close neighbor at 0.01 and 0.16 Mb apart, respectively. Other 3 QTLs were mapped on D sub-genome (*Qyr.stripe-4DL*, *Qyr.stripe-5DS*, *Qyr.stripe-6DL*).

In the case of pathotype 110S119, 7 significant QTLs were mapped on five different chromosomes (3B, 6A, 6D, 7A, and 7D). The highest number of QTLs were mapped on D sub-genome, i.e., 4 QTLs (*Qyr.stripe-6DL*, *Qyr.stripe-7DL*, *Qyr.stripe-7DS.1*, *Qyr.stripe-7DS.2*) and 2 QTLs were mapped on A sub-genome (*Qyr.stripe-6AS*, *Qyr.stripe-7AS*), whereas only 1 QTL was found on B sub-genome (*Qyr.stripe-3B* with *p*-value 9.76E-04 and $R^2 = 0.07$). Further for the pathotype 238S119, 9 QTLs were identified as significant. These QTLs were found on 6 different chromosomes at various locations (i.e., 1A, 4A, 1B, 5B, 4D, and 5D). Maximum number, i.e., 4 QTLs were found to be associated with B sub-genome (*Qyr.stripe-1BL.1*, *Qyr.stripe-1BL.2*, *Qyr.stripe-5BL*, *Qyr.stripe-5BS*) followed by D (*Qyr.stripe-4DL*, *Qyr.stripe-5DL.1*, *Qyr.stripe-5DL.2*) and A sub-genome (*Qyr.stripe-1AL*, *Qyr.stripe-4AS*).

Furthermore, in order to identify consistent MTAs across the years, GWAS was separately performed for 2018 and 2019. The results are presented in **Supplementary Tables 3**, **4**. A total of 9 QTLs were consistently present in 2018 and 2019, i.e., in agreement with the results of the pooled analysis. These include 3 QTLs for 47S103 (1BS: *Qyr.stripe-1BS*, 2BL: *Qyr.stripe-2BL.2* and 3B: *Qyr.stripe-3B*), 4 QTLs for 46S119 (3AL: *Qyr.stripe-3AL*, 3B: *Qyr.stripe-3B.2*, 4DL: *Qyr.stripe-4DL* and 5BS: *Qyr.stripe-5BS.2*), and 2 QTLs for 238S119 (1AL: *Qyr.stripe-1AL* and 5DL: *Qyr.stripe-5DL.1*).

Identification of Putative Candidate Genes

Candidate genes for pathotypes 47S103, 46S1119, 110S119, and 238S119 were identified by mapping the markers on wheat genome assembly. Seven prominent markers (47S103-2BL, i.e., for pathotype 47S103 on long arm of B sub-genome) mapped to 2.17 Mb interval ranging from 759.82 Mb (AX-94490490) to 761.99 Mb (AX-94868242) on chromosome 2BL. This interval contains 25 genes, of which 5 had a high enrichment score. Similarly, 238S119-1BL mapped to 1.69 Mb interval ranging from 606.84 Mb (AX-95243592) to 608.53 Mb (AX-94850928) on chromosome 1BL. This interval contains seven genes, of which two were had a high enrichment score. Interestingly, the two markers, i.e., AX-94877000 and AX-94904447 associated with 47S103 and 46S119 pathotypes, respectively were linked with a common gene (*TraesCS3B02G005900*) at 3.24 Mb on chromosome 3B. For pathotypes 46S119, 110S119, and 238S110 three of these genes, *TraesCS6D02G384800*, *TraesCS7A02G021700*, and *TraesCS1B02G376000* mapped on chromosome 6DL, 7AS, and 1BL, respectively. These genes were annotated as leucine-rich repeat receptor-like protein kinases (LRR) and serine-threonine/tyrosine-protein kinase (STPK) which have key roles in pathogen recognition and disease resistance.

The annotation of associated SNPs revealed potential candidate genes (**Supplementary Table 5** and **Table 6**).

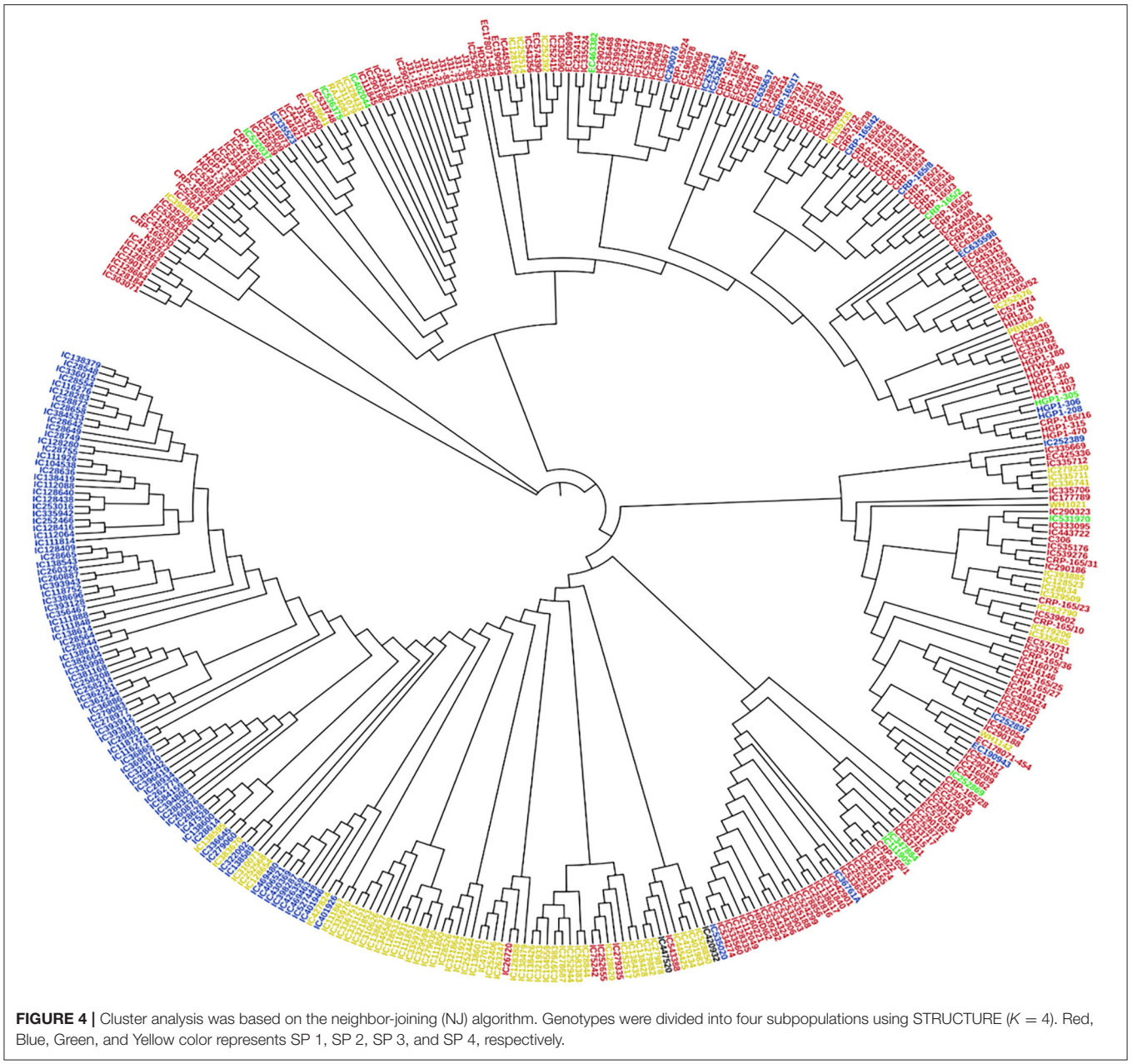


FIGURE 4 | Cluster analysis was based on the neighbor-joining (NJ) algorithm. Genotypes were divided into four subpopulations using STRUCTURE ($K = 4$). Red, Blue, Green, and Yellow color represents SP 1, SP 2, SP 3, and SP 4, respectively.

Maximum candidate genes were found for pathotype 47S103 which encodes various class of proteins and enzymes, such as Plant actin-related protein 8 (Arp complex) acts as a host response to pathogen infection (Qi et al., 2017), the heavy metal-associated domain plays an important role in the development of vascular plants and in plant responses to environmental changes. Phosphatidylinositol 3-/4-kinase (PI3K) acts as a catalytic domain and plays vital roles in the regulation of various cellular activities, including proliferation, differentiation, membrane ruffling and prevention of apoptosis (Cantley, 2002). Exocyst subunit Exo70 family protein (EXO70) describes the expression profiling of EXO70 genes from wheat (Zhao et al., 2019). Seven candidate genes were found for pathotype 46S119. Out of seven genes, *TraesCS6D02G384800* gene on 6DL was annotated leucine-rich repeat receptor-like protein kinases and serine-threonine/tyrosine-protein kinase (STPK) which has key roles in pathogen recognition and disease resistance. Further, six candidate genes (**Table 6**) identified against pathotype 110S119 were annotated, and they encoded Isopenicillin N synthase-like, Agent domain plant type, Isopenicillin N synthase-like, bifunctional inhibitor/plant lipid transfer protein, Mitochondrial substrate/solute carrier and leucine-rich repeat receptor-like protein kinases, respectively. The candidate genes annotated for 238S119 revealed functions, such as amino acid pathway regulation, transcription regulation, DNA repair and metabolite transfer.

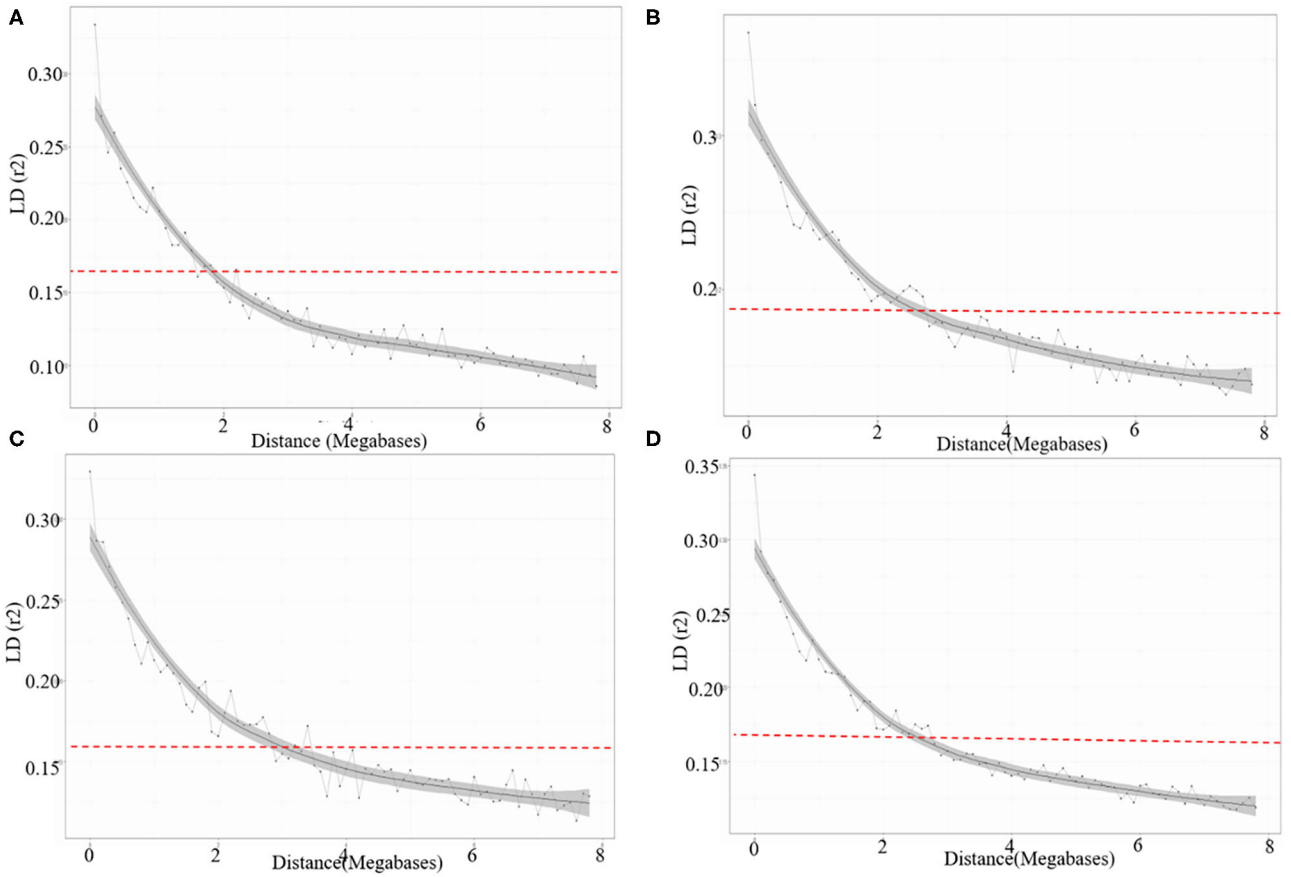


FIGURE 5 | LD decay across the **(A)** A, **(B)** B, **(C)** D sub-genomes, and **(D)** whole genome of wheat. The values on the Y-axis represent the squared correlation coefficient (R^2) and the values at X-axis represent physical distance (Mb).

The putative candidate genes were further validated using expression data available in the public domain. The expression of candidate genes was checked against susceptible (PBW343) and resistant (FLW29) wheat cultivars and their expression has been represented through heat map (**Figure 7** and **Supplementary Table 7**).

Effects of Favorable Alleles on Response to Different Pathotypes

A total of 51 MTAs (40 QTLs) in relation to different pathotypes were identified. The number of favorable alleles ranged from 1 to 41 for studied germplasm lines. Alleles associated with a reduction in disease response were considered as favorable alleles at each locus of the respective SNP. Broadly we have considered infection type 1–4 as resistant, infection type 5–6 and 7–9 were considered to be moderate and susceptible, respectively. Disease severity can also be understood as percentage of infection (0–9, i.e., on 10 scale). A significant negative correlation (at *p*-value <0.0001) was observed in case of all four pathotypes, i.e., 110S119 (−0.63), 238S119 (−0.70), 46S119 (−0.72), and 47S103 (−0.80), between the number of favorable alleles in each genotype to their respective disease severity value. The same fact has also been validated by fitting simple linear regression by considering disease severity as response and number of favorable alleles as an independent factor (**Figures 8A–D**). Model parameters were found to be significant (at *p*-value <0.0001) with a good range of model R^2 for all four pathotypes.

DISCUSSION

Identification of New Sources of Resistance

The evolution of new virulent races of *P. striiformis* has always been a big challenge before wheat scientists. The new virulent races make the existing ruling varieties susceptible as most of the varieties have race specific resistance that can be overcome easily by new virulent pathogens. Heavy yield losses have been reported due to evolution of new pathotypes (Burdon et al., 2014; Hulbert and Pumphrey, 2014). Hence, there is a need for a constant search for new resistance sources for developing resistant cultivars with durable resistance at regular interval. Global as well as National level efforts have been made to explore new sources of resistance by exploring wheat germplasm collection maintained in gene banks (Gurung et al., 2014; Sehgal et al., 2015; Kumar

TABLE 5 | Quantitative trait loci for stripe rust resistance identified at the seedling stage using pooled data.

Pathotype	QTL	SNP	Chr	Position (Mb)	Allele	MAF	*P*-value	FDR *P*-value	R^2 (%)	Allelic Effect
47S103	*Qyr.stripe-1BS*	AX-94839474*	1BS	92.87	T/C	0.194	4.95E-05	0.084702#	27.7	1.84
	Qyr.stripe-2AL	AX-95174135	2AL	591.62	C/T	0.363	2.52E-05	0.081406#	28.1	−1.05
	Qyr.stripe-2BL.1	AX-94848091	2BL	756.70	C/T	0.232	6.53E-05	0.081406#	27.6	1.02
	Qyr.stripe-2BL.2	AX-94490490	2BL	759.82	C/T	0.378	3.54E-05	0.081406#	27.9	−1.03
		AX-95175933	2BL	760.12	C/T	0.203	1.46E-05	0.081406#	28.3	−2
		AX-94608940	2BL	760.26	A/G	0.498	7.54E-05	0.081406#	27.5	1.12
		AX-95142963	2BL	760.89	A/C	0.387	1.79E-05	0.081406#	28.2	−1.07
		AX-95238626*	2BL	760.95	G/A	0.195	4.26E-05	0.084702#	27.8	1.86
		AX-94497849	2BL	761.98	C/T	0.31	5.47E-05	0.084702#	27.7	1.91
		AX-94868242*	2BL	761.99	A/C	0.475	6.33E-06	0.084702#	28.8	−1.03
	Qyr.stripe-2DL.1	AX-95007159	2DL	619.60	A/G	0.441	6.80E-05	0.081406#	27.6	1.16
	Qyr.stripe-2DL.2	AX-95166280	2DL	621.02	G/A	0.194	8.98E-06	0.084702#	28.6	1.97
	Qyr.stripe-3B	AX-94877000*	3B-	3.24	A/T	0.427	4.10E-05	0.081406#	27.8	−0.98
	Qyr.stripe-3DS.1	AX-95098367	3DS	3.98	A/G	0.256	7.54E-05	0.084702#	27.5	−0.96
	Qyr.stripe-3DS.2	AX-94417619	3DS	12.45	C/G	0.435	9.85E-05	0.104201#	27.4	0.94
	Qyr.stripe-4BL	AX-94853722	4BL	623.07	C/A	0.422	3.52E-05	0.081406#	27.9	−1.39
	Qyr.stripe-6BL.1	AX-94735973	6BL	117.99	C/T	0.413	6.56E-05	0.081406#	27.6	1.62
	Qyr.stripe-6BL.2	AX-94382687	6BL	408.65	G/A	0.162	3.05E-05	0.084702#	28	1.34
46S119	*Qyr.stripe-1BL*	AX-94482117	1BL	486.48	T/G	0.134	6.69E-04	0.97075	11.8	1.14
	Qyr.stripe-3AL	AX-94419836*	3AL	374.73	G/T	0.182	8.77E-05	0.250823#	13	1.02
	Qyr.stripe-3AS	AX-95193648	3AS	0.56	A/G	0.172	5.19E-05	0.278927#	13.3	−1.58
	Qyr.stripe-3B.1	AX-94486149	3B-	3.23	G/A	0.221	6.66E-05	0.250823#	13.2	1.24
		AX-94904447	3B-	3.24	G/T	0.197	3.86E-05	0.250823#	13.5	−1.53
	Qyr.stripe-3B.2	AX-94702815*	3B-	542.74	A/T	0.199	4.12E-05	0.254159#	13.5	−1.94
		AX-94701179*	3B-	542.90	A/G	0.373	9.24E-04	0.97075	11.7	−0.86
	Qyr.stripe-4AL	AX-95630255	4AL	602.25	G/A	0.199	4.55E-04	0.97075	12.1	−0.95
	Qyr.stripe-4DL	AX-94541083*	4DL	226.26	T/C	0.142	5.52E-04	0.97075	11.9	1.47
	Qyr.stripe-5AS	AX-94797468	5AS	253.60	C/T	0.454	8.48E-04	0.97075	11.7	−0.82
	Qyr.stripe-5BS.1	AX-94706157	5BS	38.36	C/A	0.486	5.26E-05	0.250823#	13.3	0.85
	Qyr.stripe-5BS.2	AX-95145565*	5BS	45.24	C/T	0.154	9.01E-04	0.97075	11.7	−1.22
	Qyr.stripe-5DS	AX-94429414	5DS	98.26	C/T	0.019	5.36E-04	0.97075	12	−3.97
	Qyr.stripe-6DL	AX-94859969	6DL	462.86	G/A	0.306	4.11E-04	0.97075	12.1	0.85
110S119	*Qyr.stripe-3B*	AX-94883935	3B-	816.28	A/C	0.291	9.76E-04	1	6.6	0.59
	Qyr.stripe-6AS	AX-94989376	6AS	221.88	T/G	0.177	2.70E-04	1	7.4	−1.01
	Qyr.stripe-6DL	AX-94633926	6DL	369.66	C/G	0.279	2.95E-04	1	7.4	−0.48
	Qyr.stripe-7AS	AX-94844913	7AS	8.78	A/G	0.5	6.54E-04	1	6.9	0.4
	Qyr.stripe-7DL	AX-95227592	7DL	518.05	A/G	0.205	7.70E-04	1	6.8	0.48
	Qyr.stripe-7DS.1	AX-94943939	7DS	14.17	A/G	0.234	4.39E-04	1	7.1	0.5
	Qyr.stripe-7DS.2	AX-94878859	7DS	16.70	A/G	0.393	3.24E-04	1	7.3	0.51
238S119	*Qyr.stripe-1AL*	AX-95153545*	1AL	541.98	G/A	0.195	5.01E-04	0.993302	8.2	0.64
		AX-94381637	1AL	542.87	C/T	0.14	3.37E-04	0.993302	8.4	−0.82
	Qyr.stripe-1BL.1	AX-94717933	1BL	326.02	C/T	0.414	5.28E-04	0.993302	8.1	0.75
	Qyr.stripe-1BL.2	AX-95243592	1BL	606.84	G/T	0.152	5.36E-04	0.993302	8.1	−0.7
		AX-95229302	1BL	608.47	C/T	0.171	2.83E-04	0.993302	8.5	−0.69
		AX-94850928	1BL	608.53	C/T	0.168	6.11E-04	0.993302	8.1	−0.66
	Qyr.stripe-4AS	AX-94398581	4AS	176.88	C/T	0.229	8.25E-04	0.993302	7.9	−0.55
	Qyr.stripe-4DL	AX-94948233	4DL	211.03	T/C	0.371	5.13E-04	0.993302	8.2	−0.63
	Qyr.stripe-5BL	AX-95258852	5BL	499.12	T/C	0.422	3.79E-04	0.993302	8.3	−0.82
	Qyr.stripe-5BS	AX-94809513	5BS	26.45	T/C	0.317	9.81E-04	0.993302	7.8	0.69
	Qyr.stripe-5DL.1	AX-95244380*	5DL	344.50	C/G	0.146	1.44E-04	0.993302	8.9	−0.8
	Qyr.stripe-5DL.2	AX-95202253	5DL	428.83	G/A	0.114	6.15E-04	0.993302	8.1	−0.97

*Here * denotes MTAs which were also found to be significant to both year (i.e. 2018 and 2019) association mapping results and # represents FDR adjusted $p < 0.30$.*

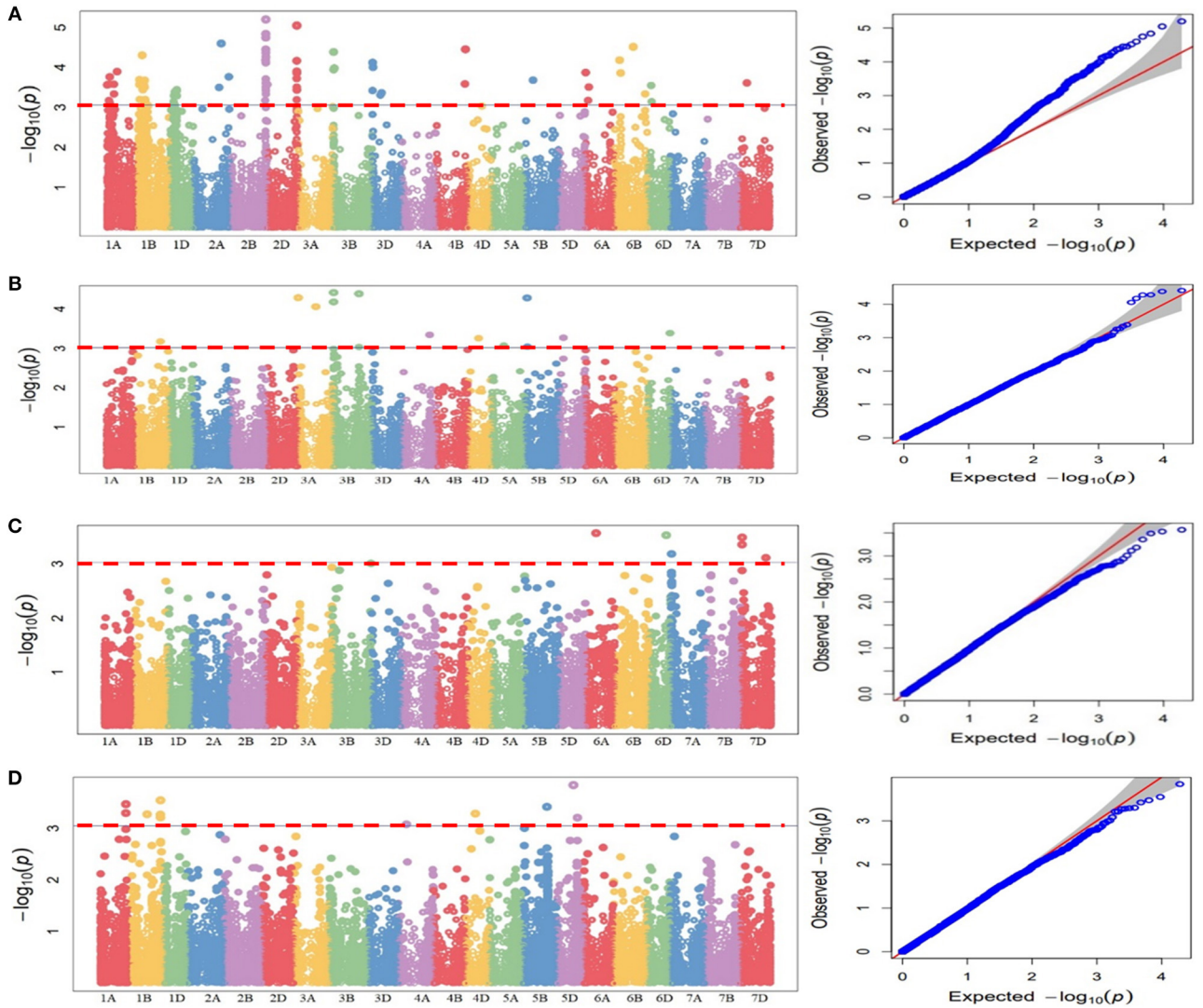


FIGURE 6 | Manhattan plots for statistical significance *P*-values across 21 wheat chromosomes for SNP markers associated with stripe rust four pathotypes. **(A)** 47S103, **(B)** 46S119, **(C)** 110S119, **(D)** 238S119 and Quantile-Quantile plots indicating the normality of data four pathotypes. Horizontal line represents significant MTAs at threshold of $P < 0.001$.

et al., 2016). Most of the germplasm lines conserved in Gene banks have not been utilized in any breeding program therefore, chances of getting novel sources of resistance is comparatively higher in comparison to breeding lines which are already in use in breeding programs. Further, with the advancement of genotyping technology, these germplasm lines can be effectively characterized and utilized by breeding programs (Cavanagh et al., 2013; Wang et al., 2014). In India, prime focus is to evaluate the wheat lines at the seedling stage for resistance against three new, i.e., 110S119, 238S119, and 47S103 and two already known *Yr9*-virulent pathotypes, i.e., 46S119 and 78S84 (Gangwar et al., 2019).

Germplasm lines including landraces are the primary source of genes and a valuable source of resistance, until now very few landraces have been used in wheat breeding programs (Bajgain et al., 2015; Naruoka et al., 2015; Turner et al., 2017). In our study, significant phenotypic variation among the genotypes has been observed as depicted by ANOVA results. On the basis of IT scores, our results suggested 110S119 and 238S119 were the most virulent pathotypes in 2018 whereas 110S119 was the most virulent in 2019. Recently, a study of diverse spring wheat panel for stripe rust also showed similar results, i.e., 110S119, the most virulent pathotype (Kumar et al., 2020).

We identified novel resistance sources for all the studied pathotypes. Some of them were resistant against two of the pathotypes which includes IC111888, IC290156. In indigenous wheat germplasm lines including landraces for 2 consecutive years. Germplasm lines with higher levels of resistance against prevailing races of stripe rust can be utilized in breeding rust resistant cultivars in the future.

TABLE 6 | Candidate genes identified in significant associated SNPs with studied pathotypes.

Pathotype	Marker	Chr	Position (Mb)	Gene	Gene description
47S103	AX-95166280	2DL	621.02	*TraesCS2D02G542500*	Exocyst subunit Exo70 family protein
	AX-95007159	2DL	619.6	*TraesCS2D02G538600*	Sulfotransferase
	AX-94877000	3B-	3.24	*TraesCS3B02G005900*	Transcriptional regulatory protein Sin3-like
	AX-95098367	3DS	3.98	*TraesCS3D02G010600*	integral component of membrane
	AX-94417619	3DS	12.45	*TraesCS3D02G034000*	Proteinase inhibitor I12, Bowman-Birk
46S119	AX-94904447	3B-	3.24	*TraesCS3B02G005900*	Transcriptional regulatory protein Sin3-like
	AX-94702815	3B-	542.74	*TraesCS3B02G336700*	integral component of membrane
	AX-94797468	5AS	253.6	*TraesCS5A02G119800*	Inositol monophosphatase-like
	AX-94706157	5BS	38.36	*TraesCS5B02G035500*	FAS1 domain superfamily
	AX-95145565	5BS	45.24	*TraesCS5B02G040800*	Bidirectional sugar transporter SWEET
	AX-94429414	5DS	98.26	*TraesCS5D02G091100*	Amino acid transporter, transmembrane domain
	AX-94859969	6DL	462.86	*TraesCS6D02G384800*	Leucine-rich repeat domain superfamily/Leucine-rich repeat-containing N-terminal, plant-type
110S119	AX-94883935	3B-	816.28	*TraesCS3B02G592000*	Isopenicillin N synthase-like
	AX-94989376	6AS	221.88	*TraesCS6A02G186500*	Agenet domain, plant type
	AX-94633926	6DL	369.66	*TraesCS6D02G261300*	Isopenicillin N synthase-like
	AX-94844913	7AS	8.78	*TraesCS7A02G021700*	Leucine-rich repeat domain superfamily/P-loop containing nucleoside triphosphate hydrolase
	AX-94878859	7DS	16.7	*TraesCS7D02G031800*	Bifunctional inhibitor/plant lipid transfer protein
	AX-95227592	7DL	518.05	*TraesCS7D02G400800*	Mitochondrial substrate/solute carrier
238S119	AX-94381637	1AL	542.87	*TraesCS1A02G362200*	F-box-like domain superfamily
	AX-95229302	1BL	608.47	*TraesCS1B02G376800*	Phosphoribosylformylglycinamidine synthase PurL
	AX-95258852	5BL	499.12	*TraesCS5B02G316700*	Domain unknown function DUF295
	AX-95202253	5DL	428.83	*TraesCS5D02G340400*	Thioredoxin-like superfamily

Most of the breeding lines and varieties included in our study showed resistance against the pathotype 47S103 and 46S119 and it could be because these pathotypes evolved in 1991 and 1996, respectively and most of the breeding programs were focusing on these pathotypes only (Bhardwaj et al., 2014). The other two pathotypes 110S119 and 238S119 came into existence during 2014.

Marker Coverage and Population Structure

Genetic diversity is the basic requirement of any breeding program. The determination of extent of genetic diversity and population structure are the foremost requirement for initiating and utilizing plant genetic resources in breeding programs and also for genetic studies (Atwell et al., 2010). We found the marker density of 19,090 SNPs on different chromosomes and identified that Genome B showed the highest marker density as compared to sub-genome A and D. Previous studies also revealed highest marker density of polymorphic SNPs on the B sub-genome (Kumar et al., 2020).

Population structure is an important factor which influences LD (Flint-Garcia et al., 2003). Our AM panel contains the genotypes from different geographic regions, out of 391 genotypes, 247 were from northern part of India, 86 were from other parts of India like the central zone (Rajasthan, Madhya Pradesh and Gujrat), or Southern part of India (Karnataka, Tamilnadu, Andhra Pradesh, Maharashtra) while 58 genotypes were exotic. Approximately 50% of genotypes were land races/or locally collected germplasm lines. Population structure determined by both STRUCTURE program and PCA based approach suggested four sub-populations (Kumar et al., 2020). A significant level of admixture may be due to the sharing of the germplasm across breeding programs. However, some collection regions have strong enrichment in specific genetic-based sub-populations, like SP 1 contains 76% of the germplasm lines of Indian origin either landraces or breeding lines and released varieties while 24% breeding lines were procured from the CIMMYT Mexico. Out of 102 genotypes from SP 2, only five lines were from CIMMYT Mexico, the rest were local land races (65) and old varieties (32). SP 3 contained only 10 genotypes, of which two were from CIMMYT and eight were from North Indian Origin. SP 4 contains 51 local land races and 22 breeding lines of Indian origin.

LD is one of the most important factors that determine the power of association analysis. We estimated LD for all the chromosomes of three sub-genomes, i.e., A, B, and D. The critical R^2 values for genome A, B, and D were 0.14, 0.16, and 0.15, respectively. We observed faster decay of LD in the sub-genome A than the other two sub-genomes (B and D). Many other studies have also showed rapid decay of LD in sub-genome A (Chao et al., 2010; Voss-Fels et al., 2015). Further, as in this analysis, the longest LD decay distance was observed for the sub-genome D, the same has been reported in previous studies (Chao et al., 2010; Voss-Fels et al., 2015; Liu et al., 2017b; Maulana et al., 2018; Qaseem et al., 2019). Since the

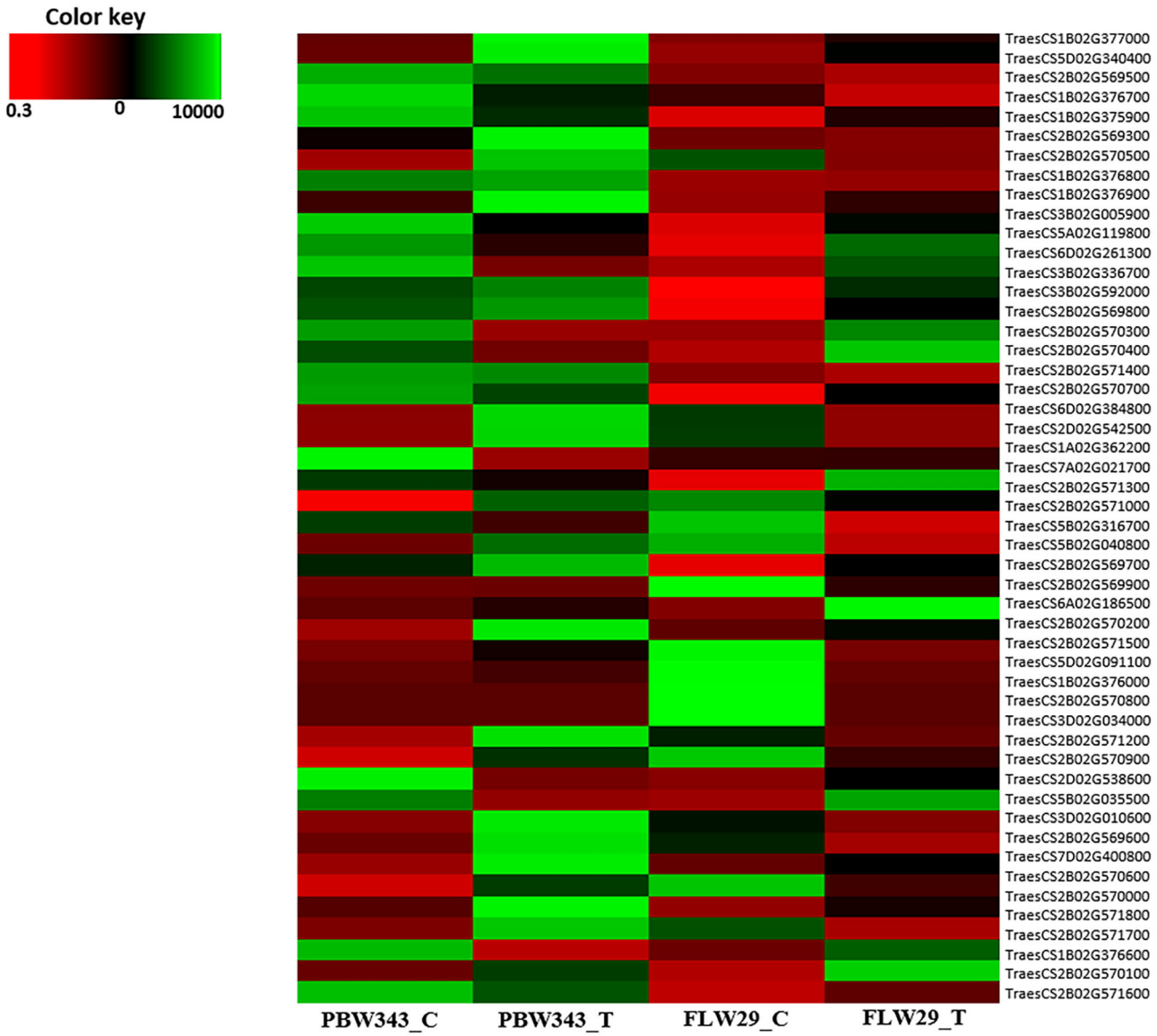


FIGURE 7 | Heat map representing expression of putative candidate genes.

LD decay is influenced by population composition, it thus may vary in different populations, but broadly the sub-genome D generally has a longer LD decay distance as compared to the other two sub-genomes.

Further, although GWAS enables high resolution mapping of traits, often it might also reveal false positive associations if the confounding factors like population structure and genetic-relatedness among genotypes of the association panel are not accounted for. Therefore, in our analysis, a CMLM based method which accounts for both of these factors was used (Wei et al., 2017). The population structure analysis using model based approach, N-J based phylogeny and PCA revealed four sub-populations in the AM panel. However, the clustering pattern observed in our analysis was not explainable on the basis of geographical origin/source of included germplasm lines. One of the possible reasons for this could be the extensive sharing of germplasm within the wheat breeding program of India in the past six decades.

Comparison of Identified QTLs With Previously Published *Yr* Genes or QTLs

Our study has demonstrated the power of the GWAS approach in uncovering genomic regions associated with stripe rust resistance. In total, 51 MTAs were identified for four studied stripe rust pathotypes. Based on LD decay distance, we classified these 51 MTAs into 40 QTLs (23 suggestive QTLs and 17 significantly associated QTLs) distributed on all the wheat chromosomes except Chr1D and Chr7B. These QTLs further

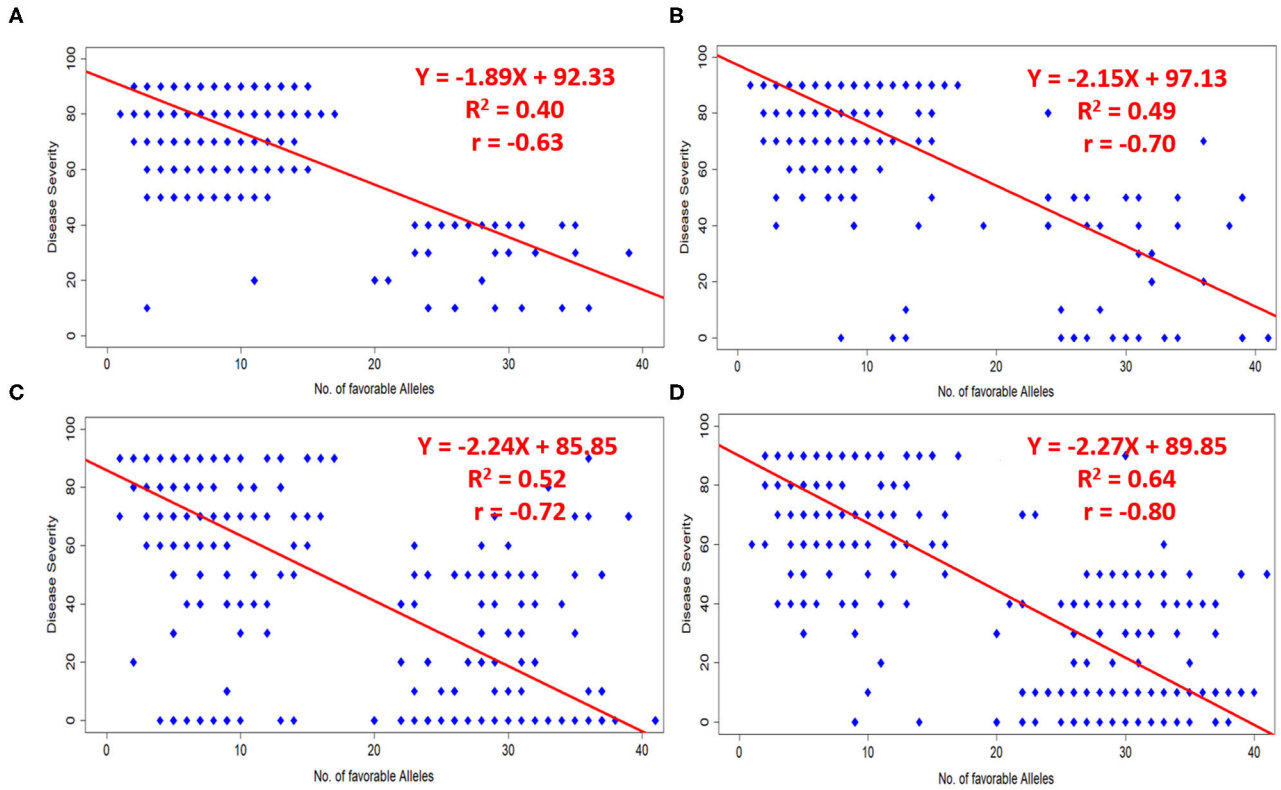


FIGURE 8 | Effects of favorable alleles on response to four pathotypes **(A)** 110S119, **(B)** 238S119, **(C)** 46S119, and **(D)** 47S103.

need to be validated before it can be used in any future breeding programs. Previous studies have also identified *Yr* genes/QTLs distributed on almost all the wheat chromosomes as both major and minor genes are known to be responsible for conferring resistance against stripe rust pathogens (Zegeye et al., 2014; Liu et al., 2017b; Muleta et al., 2017a; Ye et al., 2019; Muhammad et al., 2020). Zegeye et al. (2014) identified stripe rust resistance QTLs on 1B, 2A, 2B, 3B, 3D, 5A, 5B, 6D, and 7A. Muleta et al. (2017b) identified stripe rust resistance QTLs on 1B, 2A, 2B, 2D, 5B, and 7B. Whereas, Ye et al. (2019) identified 12 QTLs on the long arms of 1B, 3D, 5A, 5B, and 7B and the short arms of chromosomes 1A, 5A, 6A, 6B, and 7A.

Further, we have compared the location of identified QTLs in this study with that of the previous studies. For some of the QTLs, exact comparison across different studies was difficult due to the difference in the number of markers, mapping populations and the genotyping techniques (SSR or SNP) used in these studies. For such QTLs, chromosome arm location (short/long) seems to be a good criterion and was used in our analysis. The details of the previously identified QTLs are provided in **Supplementary Table 6**. Important novel and previously identified QTLs for stripe rust resistance identified in our analysis are discussed below.

For pathotype 47S103 (T), our analysis identified 2 QTLs (*Qyr.stripe-2BL.1, Qyr.stripe2BL.2*) on the long arm of 2B. Of these, *Qyr.stripe-2BL.1* at 756.7 Mb explained 27.6% of phenotypic variation for disease severity, and coincided to Yr5 on 2BL, close to marker wsnp_Ex_c2153_4043746 (Zegeye et al., 2014). In fact, another stripe rust resistance gene is Yr7 located at 685Mb on 2BL was also very near to *Qyr.stripe-2BL.1* identified in our analysis (Zhang et al., 2009). Another QTL on 2B, *Qyr.stripe-2BL.2* QTL was associated with 7 markers (AX-94490490, AX-95175933, AX-94608940, AX-95142963, AX-95238626, AX-94497849, AX-94868242) and located in the interval 759.82-761.99 Mb and was near to SSR marker Xwmc361 that is close to stripe rust resistance QTL Naxos *(QYr.cass-2BL)* (Ren et al., 2012). Muhammad et al. (2020) had also reported two QTLs on 2BL (*QYr.uaf.2BL.1* and *QYr.uaf.2BL.2)* spanning the genomic region containing three important genes *Yr5, Yr43* and *Yr54.* Further, QTL identified on the small arm of 3BS, *Qyr.stripe-3B* (at 3.24 Mb) coincided with *Qyr.ramp-3B.1* which was associated with infection caused by two stripe rust pathotypes Yr-47S103 and Yr-46S119 (Kumar et al., 2020). The QTL on 6BL, *Qyr.stripe-6BL.2* at 408 Mb was close to Xwmc397 and Xwmc105b SSR markers associated with stripe rust resistance (Christiansen et al., 2006).

For pathotype 46S119, 12 QTLs were identified. Among these, six QTLs (*Qyr.stripe-1BL, Qyr.stripe-3AS, Qyr.stripe-4AL, Qyr.stripe-5AS Qyr.stripe-5BS.2, Qyr.stripe-5DS)* were located in the previously located regions and remaining were novel (*Qyr.stripe-3AL, Qyr.stripe-3B.1 Qyr.stripe-3B.2, Qyr.stripe-4DL, Qyr.stripe-5BS.1* and *Qyr.stripe-6DL).* The QTL, *Qyr.stripe-1BL*

identified on 1BL explained 11.8% of phenotypic variance for disease severity and coincided to the genomic region containing *Yr29,* independently identified by Lan et al. (2014) and Muhammad et al. (2020). The long arm of chromosome 1BL, is considered to be rich in *Yr* resistance genes (Muhammad et al., 2020) which was clearly evident from the detection of 3 QTLs in our analysis, one QTL for pathotype 46S119 and two QTLs for pathotype 238S119. The *Qyr.stripe-4AL* at 602.2 Mb explained 12.1% of phenotypic variance and was very near to *Yr51* and *Yr60* (Randhawa et al., 2014; Herrera-Foessel et al., 2015).

For pathotype 238S119, a total of nine genome regions were identified in our analysis. Among these, three QTLs (*Qyr.stripe-1BL.1,* Qyr.stripe-1BL.2 Qyr.stripe-5BL) were located in the previously identified genomic regions for stripe rust resistance and other six were novel (*Qyr.stripe-1AL, Qyr.stripe-4AS, Qyr.stripe-4DL, Qyr.stripe-5BS, Qyr.stripe-5DL.1 and Qyr.stripe-5DL.2). Qyrstripe-1BL* on long arm of 1BL at 326.58 Mb, defined up to 8.2% of phenotypic variance for disease severity and coincided to the genomic region containing *Yr64* and *Yr65* genes (Cheng et al., 2014). Besides this, there were other *Yr* genes, such as *Yr10* and *Yr29* in the vicinity of the identified QTL on Chr1BL (Liu et al., 2008; Lan et al., 2014; Muhammad et al., 2020). Moreover, we have also identified a haplotype block of 3 SNPs (AX-95243592, AX-95229302 and AX-94850928) on 1B in the interval 606–608 Mb. Since these 3 SNPs were in high LD, this region was considered just one locus, QTL *Qyr.stripe-1BL.2*. A previous study on stripe rust resistance in a durum wheat cultivar Wollaroi revealed *Yr29* (Xgwm818, Xgwm259) on 1BL at 670 Mb, which was close to *Qyrstripe-1BL.2* identified in our study (McIntosh et al., 1995, Bansal et al., 2014). The QTL *Qyr.stripe-5BL* was located at 499.12 Mb on chromosome 5BL. This QTL was found to be in the vicinity of QTL *Qyr.sun-5B*, that provided resistance to adult stage stripe resistance (Bariana et al., 2010). Besides the known QTLs, our study has revealed 5 novel QTLs for stripe rust resistance (pathotype 238S119) distributed on 4AS, 4DL, 5BS, and 5DL. These QTLs were at a large distance from the previously identified QTL on these chromosomes.

Similarly, for 110S119, 7 QTLs were identified which included two (*Qyr.stripe-6AS* and *Qyr.stripe-6DL*) that coincided with previously reported genomic regions and five were novel (*Qyr.stripe-3B, Qyr.stripe-7AS, Qyr.stripe-7DL Qyr.stripe-7DS.1,* and *Qyr.stripe-7DS.2*) (Muhammad et al., 2020). Further, all of them explained <10% phenotypic variance suggesting they were minor QTLs and could be important for providing durable resistance.

Identification of favorable alleles for stripe rust resistance is necessary to enhance the cultivars resistance. In the present study, the correlation between stripe rust resistance for different pathotypes and favorable alleles was highly significant and biologically meaningful. In our data, we identified some indigenous lines which were immune (i.e., having 0 disease severity index) to stripe rust resistance. For example, IC111888 (local germplasm) and IC290156 (local germplasm) from NBPGR, New Delhi was found to be immune in 2 races i.e., 238S119 and 46S119; 46S119 and 47S103, respectively. IC111888 and IC290156 had 41 and 37 favorable alleles. It could be concluded that these germplasm lines have relatively few or low identified resistance-associated favorable alleles and therefore showed high disease severity index. This finding will further provide insights for wheat breeders when choosing the diverse parents as a source of rust resistance to breed wheat for the 12 million hectares prone to this disease.

CONCLUSION

The results of the present investigation showed the value of our diverse genetic resources conserved in Indian National Gene bank. The germplasm lines/landraces found resistant in the present investigation are valuable sources of resistance and can be used to achieve durable and diverse resistance against stripe rust. In the present study, out of the 40 identified QTLs, 20 QTLs were potentially novel for stripe rust resistance. Further, three putative candidate genes associated with these QTLs are expected to play major roles in marker-assisted wheat breeding for stripe rust resistance in wheat. Further, Genomic regions identified in the present investigations have significant associations with stripe rust resistance in Indian wheat germplasm. However, identified QTLs need to be examined for favorable SNP alleles associated with resistance genes so that they can be used in breeding programs.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in Zenodo, doi: 10.5281/zenodo.4058520.

AUTHOR CONTRIBUTIONS

AKP: investigation, data recording, data analysis, writing—original draft, and writing-review and editing. SundK: conceptualization, data curation, formal analysis, investigation, writing—original draft, and writing-review and editing. AS: conceptualization, investigation, and writing-review and editing. NB: contributed in bioinformatics analysis. DM: contributed in bioinformatics analysis. DC: investigation and data recording. SM and MG: review and editing. SuneK: formal analysis, assisted in writing original draft, review, and editing. OG: phenotypic evaluation of association mapping panel against four different pathotypes of stripe rust. SubK: prepared the inoculum and created artificial epiphytotic conditions for all four pathotypes for 2 years in greenhouse. AG: germplasm curation and reviewing manuscript. SB: phenotypic evaluation of association mapping panel against four different pathotypes of stripe rust. AR and KS: writing—review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The financial support provided by IASRI-CABin for conducting this study was highly acknowledged. All the authors were thankful to Division of Germplasm Conservation and Division of Germplasm Evaluation for providing germplasm lines for this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.572975/full#supplementary-material

Supplementary Table 1 | Details of wheat germplasm lines and distribution of favorable alleles to different pathotypes among 391 wheat germplasm lines used in the present study.

Supplementary Table 2 | Disease severity data of 391 germplasm lines against four different pathotypes at seedling stage in 2 consecutive years.

Supplementary Table 3 | List of significant MTAs associated to four pathotypes for year 2018.

Supplementary Table 4 | List of significant MTAs associated to four pathotypes for year 2019.

Supplementary Table 5 | Candidate genes identified within the interval region for chromosome 2BL (759.82–761.99 Mb) and 1BL (606.84–608.53 Mb).

Supplementary Table 6 | Comparison of identified QTLs with previously published *Yr* genes or QTLs.

Supplementary Table 7 | Expression of putative candidate genes against susceptible and resistant variety.

REFERENCES

Afzal, S. N., Haque, M. I., Ahmedani, M. S., Bashir, S., and Rattu, A.R. (2007). Assessment of yield losses caused by *Puccinia striiformis* triggering stripe rust in the most common wheat varieties. *Pak. J. Bot.* 39, 2127–2134.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800

Bajgain, P., Rouse, M. N., Bulli, S., Bhavani, S., Gordon, T., Wanyera, R., et al. (2015). Association mapping of North American spring wheat breeding germplasm reveals loci conferring resistance to Ug99 and other African stem rust races. *BMC Plant Biol.* 15:249. doi: 10.1186/s12870-015-0628-9

Bansal, U. K., Kazi, A. G., Singh, B., Hare, R. A., and Bariana, H. S. (2014). Mapping of durable stripe rust resistance in a durum wheat cultivar Wollaroi. *Mol. Breed.* 33, 51–59. doi: 10.1007/s11032-013-9933-x

Bariana, H. S., Bansal, U. K., Schmidt, A., Lehmensiek, A., Kaur, J., Miah, H., et al. (2010). Molecular mapping of adult plant stripe rust resistance in wheat and identification of pyramided QTL genotypes. *Euphytica* 176, 251–260. doi: 10.1007/s10681-010-0240-x

Bhardwaj, S. C., Gangwar, O., Prasad, P., and Khan, H. (2014). Newsletter "Mehtaensis" (Flowerdale Shimla) of Indian Institute of Wheat and Barley Research. *Mehtaensis* 34, 1–23.

Bhardwaj, S. C., Gangwar, O., Singh, S. B., Saharan, M. S., and Sharma, S. (2012). Rust situation and pathotypes of *Puccinia* species in Leh Ladakh in relation to recurrence of wheat rusts in India. *Indian Phytopath.* 65, 230–232.

Bradbury, J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Breseghello, F., and Sorrells, F. M. (2005). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Bulli, P., Zhang, J., Chao, S., Chen, X., and Pumphrey, M. (2016). Genetic architecture of resistance to stripe rust in a global winter wheat germplasm collection. *G3 Genes Genomes Genet.* 116:028407. doi: 10.1534/g3.116.028407

Burdon, J. J., Barrett, L. G., Rebetzke, G., and Thrall, H. (2014). Guiding deployment of resistance in cereals using evolutionary principles. *Evol. Appl.* 7, 609–624. doi: 10.1111/eva.12175

Cantley, L. C. (2002). The phosphoinositide 3-kinase pathway. *Science* 296, 1655–1657. doi: 10.1126/science.296.5573.1655

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8057–8062. doi: 10.1073/pnas.1217133110

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M. C., Baenziger, S., Matnyazov, R., et al. (2010). Population-and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics* 11:727 doi: 10.1186/1471-2164-11-727

Chen, X. and Kang, Z. (eds.). (2017). *Stripe Rust*. Dordrecht: Springer Netherlands.

Chen, X. M. (2005). Epidemiology and control of stripe rust [*Puccinia striiformis* f. s *tritici*] on wheat. *Can. J. Plant Pathol.* 27, 314–337. doi: 10.1080/07060660509507230

Chen, X. M., Penman, L., Wan, A. M., and Cheng, P. (2010). Virulence races of Puccinia striiformis f. s tritici in 2006 and 2007 and development of wheat stripe rust and distributions, dynamics, and evolutionary relationships of races from 2000 to 2007 in the United States. *Can. J. Plant Pathol.* 32, 315–323. doi: 10.1080/07060661.2010.499271

Cheng, B., Gao, X., Cao, N., Ding, Y., Gao, Y., Chen, T., et al. (2020). Genome-wide association analysis of stripe rust resistance loci in wheat genotypes from southwestern China. *J. Appl. Genet.* 61, 37–50. doi: 10.1007/s13353-019-00533-8

Cheng, P., Xu, L. S., Wang, M. N., See, D. R., and Chen, X. M. (2014). Molecular mapping of genes Yr64 and Yr65 for stripe rust resistance in hexaploidderivatives of durum wheat accessions PI 331260 and PI 480016. *Theor. Appl. Genet.* 127, 2267–2277. doi: 10.1007/s00122-014-2378-8

Christiansen, M. J., Feenstra, B., Skovgaard, I. M., and Andersen, S. B. (2006). Genetic analysis of resistance to yellow rust in hexaploid wheat using a mixture model for multiple crosses. *Theor. Appl. Genet.* 112, 581–591. doi: 10.1007/s00122-005-0128-7

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* 74, 829–836. doi: 10.1080/01621459.1979.10481038

Doyle, J. J. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Feng, J., Wang, M., See, D. R., Chao, S., Zheng, Y., and Chen, X. (2018). Characterization of novel gene Yr79 and four additional quantitative trait loci for all-stage and high-temperature adult-plant resistance to stripe rust in spring wheat PI 182103. *Phytopathology* 108, 737–747. doi: 10.1094/PHYTO-11-17-0375-R

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Ann. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Gangwar, O., Kumar, S., Bhardwaj, S. C., Kashyap, L., Prasad, P., and Khan, H. (2019). Characterization of three new Yr9-virulences and identification of sources of resistance among recently developed Indian bread wheat germplasm. *J. Plant Pathol.* 101, 955–963. doi: 10.1007/s42161-019-00302-w

Gurung, S., Mamidi, S., Bonman, J. M., Xiong, M., Brown-Guedira, G., and Adhikari, T. B. (2014). Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS ONE* 9:e0108179. doi: 10.1371/journal.pone.0108179

Herrera-Foessel, S. A., Singh, R. P., Lan, C. X., Huerta-Espino, J., Calvo-Salazar, V., Bansal, U. K., et al. (2015). Yr60, a gene conferring moderate resistance to stripe rust in wheat. *Plant Dis.* 99, 508–511. doi: 10.1094/PDIS-08-14-0796-RE

Hulbert, S., and Pumphrey, M. (2014). A time for more booms and fewer busts? Unraveling cereal-rust interactions. *Mol. Plant Microbe Interact.* 27, 207–214. doi: 10.1094/MPMI-09-13-0295-FI

Indu, S., and Saharan, M. S. (2011). Status of wheat diseases in India with a special reference to stripe rust. *Plant Dis. Res.* 26:156.

Kertho, A., Mamidi, S., Bonman, J. M., McClean, P. E., and Acevedo, M. (2015). Genome-wide association mapping for resistance to leaf and stripe

rust in winter-habit hexaploid wheat landraces. *PLoS ONE* 10:e0129580. doi: 10.1371/journal.pone.0129580

Kumar, D., Kumar, A., Chhokar, V., Gangwar, O., Bhardwaj, S. C., Sivasamy, M., et al. (2020). Genome-wide association studies in diverse spring wheat panel for stripe, stem and leaf rust resistance. *Front. Plant Sci.* 11:748. doi: 10.3389/fpls.2020.00748

Kumar, S., Archak, S., Tyagi, R. K., Kumar, J., Jacob, S. R., Srinivasan, K., et al. (2016). Evaluation of 19,460 wheat genotypes conserved in the Indian national genebank to identify new sources of resistance to rust and spot blotch diseases. *PLoS ONE* 11:e0167702. doi: 10.1371/journal.pone.0167702

Laidò, G., Panio, G., Marone, D., Russo, M. A., Ficco, D. B., Giovanniello, V., et al. (2015). Identification of new resistance loci to African stem rust race TTKSK in tetraploid wheats based on linkage and genome-wide association mapping. *Front. Plant. Sci.* 6:1033. doi: 10.3389/fpls.2015.01033

Lan, C., Singh, R., Huerta-Espino, J., Calvo-Salazar, V., and Herrera-Foessel, S. A. (2014). QTL characterization of resistance to leaf rust and stripe rust in the spring wheat line Francolin-1. *Mol. Breed.* 34, 789–803. doi: 10.1007/s11032-014-0075-6

Letta, T., Maccaferri, M., Badebo, A., Ammar, K., Ricci, A., Crossa, J., et al. (2013). Searching for novel sources of field resistance to Ug99 and Ethiopian stem rust races in durum wheat via association mapping. *Theor. Appl. Genet.* 126, 1237–1256. doi: 10.1007/s00122-013-2050-8

Li, J., Dundas, I., Dong, C., Li, G., Trethowan, R., Yang, Z., et al. (2020). Identification and characterization of a new stripe rust resistance gene *Yr*83 on rye chromosome 6R in wheat. *Theor. Appl. Genet.* 133, 1095–1107. doi: 10.1007/s00122-020-03534-y

Line, R. F., and Qayoum, A. (1992). Virulence, aggressiveness, evolution and distribution of races of Puccinia striiformis (the cause of stripe rust of wheat) in North America, 1968–87. *US Dept. Agric. Tech. Bull.* 1788, 1–44.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., and Bradbury, P. J. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, D., Xia, X. C., He, Z. H., and Xu, S. C. (2008). A novel homeobox-like gene associated with reaction to stripe rust and powdery mildew in common wheat. *Phytopathology* 98, 1291–1296. doi: 10.1094/PHYTO-98-12-1291

Liu, W., Maccaferri, M., Chen, X., Laghetti, G., Pignone, D., Pumphrey, M., et al. (2017a). Genome-wide association mapping reveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum ss dicoccum*). *Theor. Appl. Genet.* 130, 2249–2270. doi: 10.1007/s00122-017-2957-6

Liu, W., Maccaferri, M., Rynearson, S., Letta, T., Zegeye, H., Tuberosa, R., et al. (2017b). Novel sources of stripe rust resistance identified by genome-wide association mapping in Ethiopian durum wheat (*Triticum turgidum* ssp. *durum*). *Front. Plant Sci.* 8:774. doi: 10.3389/fpls.2017.00774

Maulana, F., Ayalew, H., Anderson, J. D., Kumssa, T. T., Huang, W., and Ma, X. F. (2018). Genome-wide association mapping of seedling heat tolerance in winter wheat. *Front. Plant Sci.* 9:1272 doi: 10.3389/fpls.2018.01272

McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Morris, C., and Xia, X. C. (2017). *Catalogue of Gene Symbols for Wheat*.

McIntosh, R. A., Wellings, C. R., and Park, R. F. (1995). *Wheat Rusts: An Atlas of Resistance Genes*. CSIRO Publishing.

McNeal, F. H., Konzak, C. F., Smith, E., Tate, W. S., and Russell, T. S. (1971). A uniform system for recording and processing cereal research data. *US Dept. Agric. Res. Serv.* 34, 121–143.

Muhammad, S., Sajjad, M., Khan, S. H., Shahid, M., Zubair, M., Awan, F.S., et al. (2020). Genome-wide association analysis for stripe rust resistance in spring wheat (Triticum aestivum L.) germplasm. *J. Integr. Agric.* 19, 2035–2043. doi: 10.1016/S2095-3119(19)62841-8

Muleta, K. T., Bulli, P., Rynearson, S., Chen, X., and Pumphrey, M. (2017a). Loci associated with resistance to stripe rust (*Puccinia striiformis* f. s *tritici*) in a core collection of spring wheat (*Triticum aestivum*). *PLoS ONE* 12:e0179087. doi: 10.1371/journal.pone.0179087

Muleta, K. T., Rouse, M. N., Rynearson, S., Chen, X., Buta, B. G., and Pumphrey, M. O. (2017b). Characterization of molecular diversity and genome-wide mapping of loci associated with resistance to stripe rust and stem rust in Ethiopian bread wheat genotypes. *BMC Plant Biol.* 17:134. doi: 10.1186/s12870-017-1082-7

Naruoka, Y., Garland-Campbell, K., and Carter, A. (2015). Genome-wide association mapping for stripe rust (*Puccinia striiformis* f. s *tritici*) in US Pacific Northwest winter wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 128, 1083–1101. doi: 10.1007/s00122-015-2492-2

Nayar, S. K., Prashar, M., and Bhardwaj, S. C. (1997). *Manual of Current Techniques in Wheat Rusts*. Research Bulletin No. 2: 32 Regional Station, Directorate of Wheat Research Flowerdale, Shimla.

Nsabiyera, V., Bariana, H. S., Qureshi, N., Wong, D., Hayden, M. J., and Bansal, U. K. (2018). Characterisation and mapping of adult plant stripe rust resistance in wheat line Aus27284. *Theor. Appl. Genet.* 131, 1459–1467. doi: 10.1007/s00122-018-3090-x

Pakeerathan, K., Bariana, H., Qureshi, N., Wong, D., Hayden, M., and Bansal, U. (2019). Identification of a new source of stripe rust resistance Yr82 in wheat. *Theor. Appl. Genet.* 132, 3169–3176. doi: 10.1007/s00122-019-03416-y

Pasam, R. K., Bansal, U., Daetwyler, H. D., Forrest, K. L., Wong, D., and Petkowski, J. (2017). Detection and validation of genomic regions associated with resistance to rust diseases in a worldwide hexaploid wheat landrace collection using BayesR and mixed linear model approaches. *Theor. Appl. Genet.* 130, 777–793. doi: 10.1007/s00122-016-2851-7

Prashar, M., Bhardwaj, S. C., Jain, S. K., and Datta, D. (2007). Pathotypic evolution in Puccinia striiformis in India during 1995–2004. *Aust. J. Agric. Res.* 58, 602–604. doi: 10.1071/AR07126

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genet. Soc. Am.* 155, 945–959.

Qaseem, M. F., Qureshi, R., Shaheen, H., and Shafqat, N. (2019). Genome-wide association analyses for yield and yield-related traits in bread wheat (*Triticum aestivum* L.) under pre-anthesis combined heat and drought stress in field conditions. *PLoS ONE* 14:e0213407. doi: 10.1371/journal.pone.0213407

Qi, T., Wang, J., Sun, Q., Day, B., Guo, J., and Ma, Q. (2017). TaARPC3, contributes to wheat resistance against the stripe rust fungus. *Front. Plant Sci.* 8:1245. doi: 10.3389/fpls.2017.01245

Randhawa, M., Bansal, U., Valárik, M., Klocová, B., Doležel, J., and Bariana, H. (2014). Molecular mapping of stripe rust resistance gene Yr51 in chromosome 4AL of wheat. *Theor. Appl. Genet.* 127, 317–324. doi: 10.1007/s00122-013-2220-8

Ren, Y., He, Z., Li, J., Lillemo, M., Wu, L., Bai, B., et al. (2012). QTL mapping of adult-plant resistance to stripe rust in a population derived from common wheat cultivars Naxos and Shanghai 3/Catbird. *Theor. Appl. Genet.* 125, 1211–1221. doi: 10.1007/s00122-012-1907-6

Rosewarne, G. M., Herrera-Foessel, S. A., Singh, R., Huerta-Espino, J., Lan, C. X., and He, Z. H. (2013). Quantitative trait loci of stripe rust resistance in wheat. *Theor. Appl. Genet.* 126, 2427–2449. doi: 10.1007/s00122-013-2159-9

Saharan, M. S., Selvakumar, R., and Sharma, I. (2013). Wheat crop health. *Newsletter* 18, 1–8.

Sehgal, D., Vikram, P., Sansaloni, C., Ortiz, C., Saint Pierre, C., Payne, T., et al. (2015). Exploring and mobilizing the gene bank biodiversity for wheat improvement. *PLoS ONE* 10:e0132112. doi: 10.1371/journal.pone.0132112

Tadesse, W., Ogbonnaya, F., Jighly, A., Nazari, K., Rajaram, S., and Baum, M. (2013). Association mapping of resistance to yellow rust in winter wheat cultivars and elite genotypes. *Crop Sci.* 54, 607–616. doi: 10.2135/cropsci2013.05.0289

Tomar, S. M. S., Singh, S. K., Sivasamy, M., and Vinod, P. (2014). Wheat rusts in India: resistance breeding and gene deployment–a review. *Indian J. Genet. Plant Breed.* 74, 129–156. doi: 10.5958/0975-6906.2014.00150.3

Turner, M. K., Kolmer, J. A., Pumphrey, M. O., Bulli, P., Chao, S., and Anderson, J. A. (2017). Association mapping of leaf rust resistance loci in a spring wheat core collection. *Theor. Appl. Genet.* 130, 345–361. doi: 10.1007/s00122-016-2815-y

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S., et al. (2015). Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome* 8, 1–13. doi: 10.3835/plantgenome2015.03.0013

Wang, X., Vignjevic, M., Jiang, D., Jacobsen, S., and Wollenweber, B. (2014). Improved tolerance to drought stress after anthesis due to priming before anthesis in wheat (*Triticum aestivum* L.) var. Vinjett. *J. Exp. Bot.* 65, 6441–6456. doi: 10.1093/jxb/eru362

Wei, W., Mesquita, A. C. O., Figueiró, A. D. A., Wu, X., Manjunatha, S., Wickland, D., et al. (2017). Genome-wide association mapping of

resistance to a Brazilian isolate of *Sclerotinia sclerotiorum* in soybean genotypes mostly from Brazil. *BMC Genomics* 18:849. doi: 10.1186/s12864-017-4160-1

Wellings, C. R. (2011). Global status of stripe rust: a review of historical and current threats. *Euphytica* 179, 129–141. doi: 10.1007/s10681-011-0360-y

Yao, F., Zhang, X., Ye, X., Li, J., Long, L., Yu, C., et al. (2019). Characterization of molecular diversity and genome-wide association study of stripe rust resistance at the adult plant stage in Northern Chinese wheat landraces. *BMC Genet.* 20:38. doi: 10.1186/s12863-019-0736-x

Ye, X., Li, J., Cheng, Y., Yao, F., Long, L., Yu, C., et al. (2019). Genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. tritici) in Sichuan wheat. *BMC Plant Biol.* 19:147. doi: 10.1186/s12870-019-1764-4

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., and Doebley, J. F. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Yu, L. X., Morgounov, A., Wanyera, R., Keser, M., Singh, S., and Sorrells, M. (2012). Identification of Ug99 stem rust resistance loci in winter wheat germplasm using genome-wide association analysis. *Theor. Appl. Genet.* 125, 749–758. doi: 10.1007/s00122-012-1867-x

Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., and Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS ONE* 9:e105593. doi: 10.1371/journal.pone.0105593

Zhang, P., McIntosh, R. A., Hoxha, S., and Dong, C. (2009). Wheat stripe rust resistance genes Yr5 and Yr7 are allelic. *Theor. Appl. Genet.* 120, 25–29. doi: 10.1007/s00122-009-1156-5

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., and Gore, M. A. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–60. doi: 10.1038/ng.546

Zhao, J., Zhang, X., Wan, W., Zhang, H., Liu, J., Li, M., et al. (2019). Identification and characterization of the EXO70 gene family in polyploid wheat and related species. *Int. J. Mol. Sci.* 20:60. doi: 10.3390/ijms20010060

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Pradhan, Kumar, Singh, Budhlakoti, Mishra, Chauhan, Mittal, Grover, Kumar, Gangwar, Kumar, Gupta, Bhardwaj, Rai and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fgene.2020.584678

Gene Expression Dynamics in Rice Peduncles at the Heading Stage

*Manu Kandpal[1], Chandrapal Vishwakarma[2], Kushagra Krishnan[1], Viswanathan Chinnusamy[2], Ashwani Pareek[3], Manoj K. Sharma[4] and Rita Sharma[1]**

[1] Grass Genetics and Informatics Group, School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, [2] Division of Plant Physiology, ICAR-Indian Agricultural Research Institute, New Delhi, India, [3] School of Life Sciences, Jawaharlal Nehru University, New Delhi, India, [4] Grass Genetics and Informatics Group, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India

Improving grain yield in the staple food crop rice has been long sought goal of plant biotechnology. One of the traits with significant impact on rice breeding programs is peduncle elongation at the time of heading failing which leads to significant reduction in grain yield due to incomplete panicle exsertion. To decipher transcriptional dynamics and molecular players underlying peduncle elongation, we performed RNA sequencing analysis of elongating and non-elongating peduncles in two Indian cultivars, Swarna and Pokkali, at the time of heading. Along with genes associated with cell division and cell wall biosynthesis, we observed significant enrichment of genes associated with auxins, gibberellins, and brassinosteroid biosynthesis/signaling in the elongating peduncles before heading in both the genotypes. Similarly, genes associated with carbohydrate metabolism and mobilization, abiotic stress response along with cytokinin, abscisic acid, jasmonic acid, and ethylene biosynthesis/signaling were enriched in non-elongating peduncles post heading. Significant enrichment of genes belonging to key transcription factor families highlights their specialized roles in peduncle elongation and grain filling before and after heading, respectively. A comparison with anther/pollen development-related genes provided 76 candidates with overlapping roles in anther/pollen development and peduncle elongation. Some of these are important for carbohydrate remobilization to the developing grains. These can be engineered to combat with incomplete panicle exsertion in male sterile lines and manipulate carbohydrate dynamics in grasses. Overall, this study provides baseline information about potential target genes for engineering peduncle elongation with implications on plant height, biomass composition and grain yields in rice.

Keywords: heading, rice, peduncle, stem, anther, panicle exsertion, internode

OPEN ACCESS

Edited by:
Pinky Agarwal, National Institute of Plant Genome Research (NIPGR), India

Reviewed by:
Hai Zhou, South China Agricultural University, China
Vijay Gahlaut, University of Delhi, India

***Correspondence:**
Rita Sharma
rita.genomics@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *17 July 2020*
Accepted: *09 November 2020*
Published: *03 December 2020*

Citation:
Kandpal M, Vishwakarma C, Krishnan K, Chinnusamy V, Pareek A, Sharma MK and Sharma R (2020) Gene Expression Dynamics in Rice Peduncles at the Heading Stage. Front. Genet. 11:584678. doi: 10.3389/fgene.2020.584678

INTRODUCTION

With huge diversity in morphological architectures of flowering plants, a thorough understanding of plant morphology is not only essential to appreciate the diverse shapes and forms exhibited by the plant world but also vital to engineering crop plants with improved agricultural traits. A slight modification in plant architecture can have a profound effect on the agronomic performance of crop plants. This was clearly demonstrated during the green revolution, where breeding for semi-dwarf

genotypes led to higher yields and lodging resistance in rice (Coyne, 1980). Subsequently, several architectural features such as shape, size and angle of the leaf, number of tillers, and branching, were found to dramatically impact grain yield and stress responses (Peng et al., 2008).

One of the crucial morphological traits that significantly influences grain yield in rice is the elongation of the uppermost internode, also known as the peduncle. Elongation of peduncle at the time of heading facilitates the emergence of panicle from the flag leaf sheath, commonly known as panicle exsertion, which is also crucial for anther dehiscence and pollination (Bardenas, 1965). Incomplete panicle exsertion dramatically reduces seed set in rice and thus adversely affects agronomic yield (Yin et al., 2007). Moreover, peduncle elongation is highly sensitive to environmental factors. Several abiotic stress factors, particularly extremes of temperature and drought stress at the time of heading, result in dramatic yield losses by inhibiting peduncle elongation and, subsequent panicle exsertion and anther dehiscence in rice (Muthurajan et al., 2011; He and Serraj, 2012; Wu et al., 2016). Incomplete panicle exsertion is also observed in cytoplasmic male sterile lines with about 30–40% of panicles remaining enclosed in the flag leaf sheath, therefore making the spikelets unavailable for cross-pollination (Devi et al., 2011). Thus, incomplete panicle exsertion is one of the major impediments to obtain high yields in rice breeding programs. One of the common strategies used to combat this challenge involve spray of gibberellins but it significantly enhances the cost of production (Yin et al., 2007).

Furthermore, peduncle also contributes to grain filling. The excess photoassimilates, accumulated in the form of sucrose and starch in rice stems before heading, are mobilized to the panicles after heading to supplement grain filling (Chen and Wang, 2008; Wang et al., 2016). The transport of these non-structural carbohydrates from flag leaf to the panicle is facilitated through a continuous mature transport phloem network in leaf sheath, peduncle, and the rachis that supports spikelets on panicle (Slewinski, 2012). Since peduncle goes through a sink to source transition during the heading stage, elucidating transcriptional dynamics of peduncle at the time of heading is of fundamental importance to understand the mechanism of non-structural carbohydrate remobilization as well.

Till date, a handful of genes have been experimentally characterized to affect internode elongation through forward and/or reverse genetic strategies (Rutger and Carnahan, 1981; Luo et al., 2006; Zhu et al., 2011; Duan et al., 2012; Gao et al., 2012; Magome et al., 2013; Ji et al., 2014, 2019; Wang et al., 2017; Xie et al., 2018; Chu et al., 2019). However, lack of a comprehensive study to understand the overall gene expression dynamics during panicle elongation at the time heading is still limiting our understanding of panicle exsertion process in rice.

To fill this gap in our understanding, we investigated the transcriptional dynamics during peduncle elongation at the time of heading in two *indica* genotypes of rice, Swarna and Pokkali, so that genotype-specific effects could be minimized. These genotypes exhibit contrasting physiological and morphological traits in terms of cell wall composition, plant height, and grain yields (Jahn et al., 2011). Swarna, an elite *indica* cultivar, developed by crossing Vasista and Mahsuri rice varieties in 1982, has low glycemic index and is the most widely grown rice cultivar in Southern India (Rathinasabapathi et al., 2015). Pokkali, on the other hand, is a salinity tolerant *indica* landrace, mainly grown in coastal region of Southern India, for its salinity tolerance, high protein content, extra-large grains, peculiar taste, and medicinal properties (Hoang et al., 2016; Mishra et al., 2020). Expression profiling of elongating and non-elongating peduncles collected before and after heading, respectively, from both the genotypes, revealed conserved genetic elements and pathways underlying peduncle elongation at the time of heading in rice. These genes can be targeted using reverse genetic approaches to increase yield stability and stress tolerance in rice (Slewinski, 2012). Those implicated in cell wall biosynthesis would be important candidates for engineering stem biomass composition to enhance biofuel production. At the same time, genes regulating anther development as well as peduncle elongation can be leveraged to devise strategies to combat panicle enclosure in male sterile lines.

MATERIALS AND METHODS

Plant Material and Sample Preparation

Oryza sativa ssp. *indica* cultivars Swarna and Pokkali plants were grown in the fields under puddle transplanted conditions at ICAR-Indian Agricultural Research Institute (ICAR-IARI), Pusa, New Delhi, India. The phenotypic data on plant height, peduncle length, days to heading, and maturity were recorded from the field-grown plants. Elongating peduncles (EP) about 2–4 days before heading, and non-elongating peduncles (NP), about 2–4 days after heading, were separately sampled for each cultivar. Before heading, samples were collected when the maximum bulge could be observed in the flag leaf with the panicle still concealed beneath the leaf sheath. Conversely, heading stage samples were collected once the panicle had fully emerged out from the flag leaf sheath accompanied by flowering (**Figure 1A**). The samples were frozen in liquid nitrogen and stored at −80°C until further processing.

RNA Isolation, Library Preparation, and Sequencing

Total RNA from three biological replicates of peduncle samples, collected from each cultivar at both elongating and non-elongating stages, was extracted using TRIzol^{TM} reagent (Invitrogen) as per the manufacturer's instructions. The quantity and quality of RNA samples were determined using Nanodrop 2000 and Agilent 2100 Bioanalyzer. RNA samples with RIN value >7 were processed for library preparation using TruSeq® Stranded Total RNA Sample Preparation Kit. Sequencing was performed on Hisequation 2000 with a read length of 100 bp (Illumina). The data has been submitted to NCBI GEO (Accession No. GSE157727).

Data Analysis

After sequencing, raw reads were filtered to remove low-quality and adapter-contaminated reads using NGSQC Toolkit with

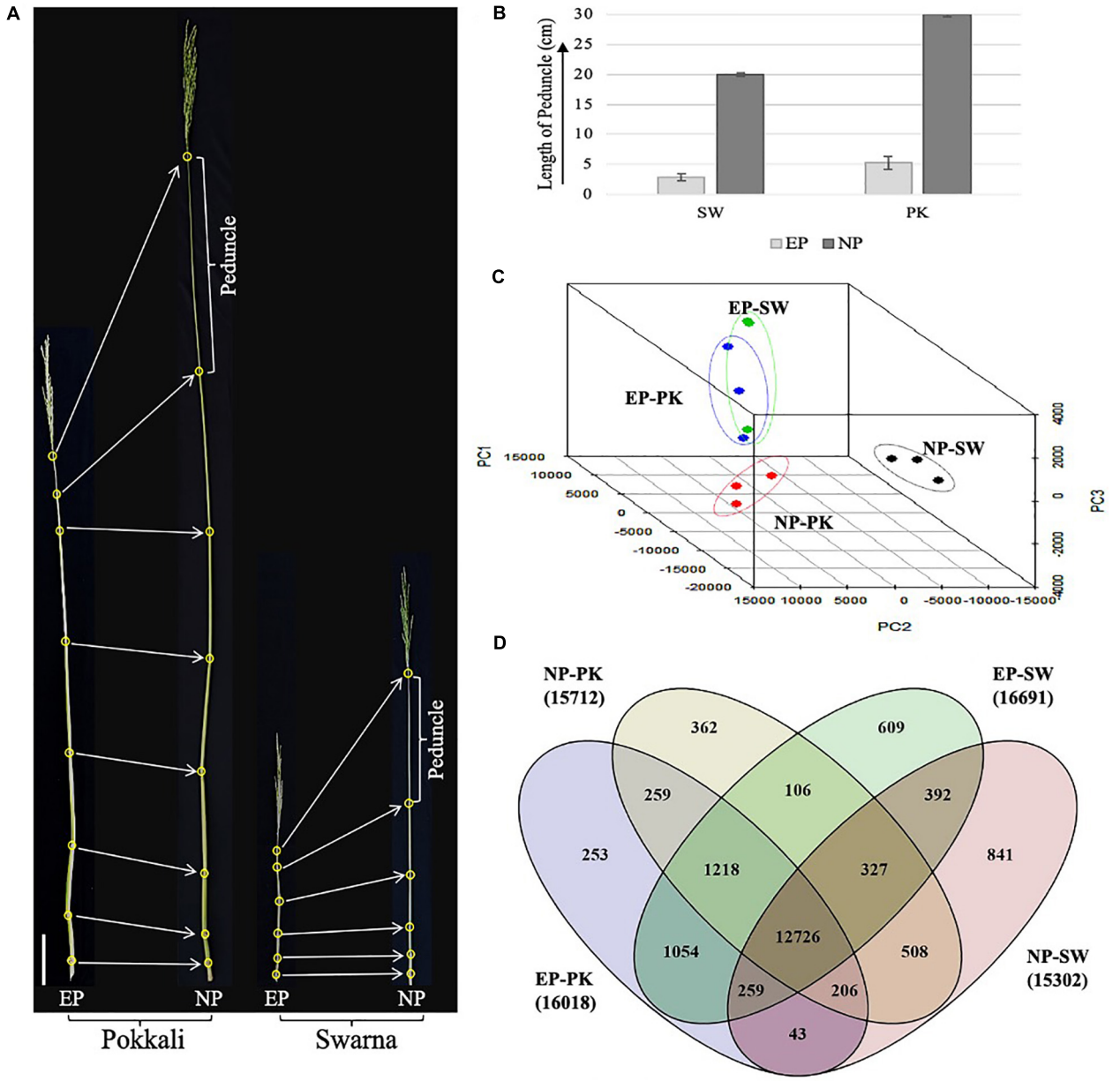


FIGURE 1 | Phenotypic assessment and transcriptomic analysis of rice peduncles. **(A)** The topmost internode, also known as peduncle, is marked in both before heading and after heading Pokkali and Swarna plants. All leaves were removed to clearly show the stem internodes. **(B)** Graph showing comparative length of elongating and non-elongating peduncles in Swarna and Pokkali plants. **(C)** Principal component analysis of all three biological replicates of each sample. Dots of the same color represent biological replicates from same sample. **(D)** Venn diagram showing overlap between number of genes expressing with FPKM ≥ 1 in each sample. The total number of genes expressed in each sample is given in brackets with sample names. SW, Swarna; PK, Pokkali; EP, elongating peduncles before heading, and NP, non-elongating peduncles after heading.

default parameters (Patel and Jain, 2012). The high-quality reads with Phred score ≥30 were aligned with the rice genome available at Rice Genome Annotation Project Database version 7[1], using Hisat version 2.1.0 with the option – rna-strandness RF, specific for stranded reads (Kim et al., 2015). Assembly was performed using Stringtie with −rf option (Pertea et al., 2015). The final assembly for all samples was obtained by Cuffmerge (Trapnell et al., 2012). Differential gene expression between both the samples was determined by Cuffdiff using the – fr option (Trapnell et al., 2012). The transcripts with $\log_2$ fold change ≥1 (upregulated genes) or ≤−1 (downregulated genes) with *P*-value cutoff ≤0.05 were considered differentially expressed. The expression levels of novel and annotated genes were quantified in terms of Fragments Per Kilobase per Million (FPKM) using

[1]http://rice.plantbiology.msu.edu/

Cufflinks (Trapnell et al., 2010). Principal component analysis (PCA) was performed on normalized FPKM values using prcomp function (center = TRUE, scale = FALSE) of the R package stats[2]. The 3D diagram of PCA for all 12 samples was plotted using the R package scatterplot3d (version 0.3–37[3]). Gene ontology (GO) enrichment for differentially expressed transcripts was performed using PlantGSEA[4]; (Yi et al., 2013), and *P*-values for enrichment analysis of each GO term were calculated using F-test. Only GO terms exhibiting *P*-value $\leq$0.05 were considered significant.

For metabolic pathway analysis, differentially expressed genes were also mapped on Mapman software using mapping file available for IRGSP loci (Thimm et al., 2004;[5]). The information about previously characterized genes for their roles in internode elongation, anther/pollen development, and grain filling was collected from OGRO (Overview of functionally characterized Genes in Rice Online database; Yamamoto et al., 2012;[6]) and FunRice genes (Yao et al., 2018;[7]) databases supplemented with manual literature survey. The list of differentially expressed genes in nine rice male sterile mutants was extracted from supplementary data provided by Lin et al. (2017).

RESULTS

Plant Growth and Phenotypic Analysis

The plant height and developmental stages of *Oryza sativa spp. indica* genotypes, Swarna and Pokkali, grown in natural fields, were monitored over time especially closer to the heading stage. Both heading and maturity take longer in Pokkali as compared to Swarna plants as Pokkali plants took on an average of 123 and 151 days after sowing for heading and maturity, respectively, whereas, in Swarna plants, the average numbers of days for heading and maturity were 112 and 136 days after sowing, respectively. Phenotypic data collected from field-grown plants revealed striking contrast in plant height and peduncle length in both the genotypes (**Figure 1A**). The average height of Pokkali was almost double (158 cm) compared to Swarna (82 cm) plants. A similar trend was observed in the peduncle length. The average length of elongating peduncles 2–4 days before heading was 2.9 and 5.3 cm in Swarna and Pokkali, respectively. However, with prolific elongation observed closer to heading, peduncle lengths increased to an average of ∼20 cm in Swarna and ∼30 cm in Pokkali plants within 2–4 days post heading (**Figure 1B**). No significant increase in peduncle length was observed 4 days post heading, suggesting that the peduncle ceases to elongate as soon as the whole panicle has emerged out of the flag leaf.

[2]https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html

[3]https://cran.r-roject.org/web/packages/scatterplot3d/index.html

[4]http://systemsbiology.cau.edu.cn/PlantGSEA/

[5]https://mapman.gabipd.org/

[6]http://qtaro.abr.affrc.go.jp/ogro

[7]https://funricegenes.github.io/

Sequencing Statistics and Principal Component Analysis

RNA sequencing of 12 libraries representing elongating (EP) and non-elongating peduncles (NP) of both the genotypes yielded a total of 690 million paired-end reads with an average read length of 100 bp. After removing low-quality and adapter contaminated reads, >616 million high-quality reads with an average Phred quality score (Q) ≥30 were used for reference-based transcriptome assembly. More than 90% of the high-quality reads from each replicate could be mapped onto the rice genome (**Supplementary Table 1**), and a total of 54,966 transcripts were annotated from both the cultivars.

To emphasize variation and correlation among EP and NP stages collected from both the genotypes, we performed PCA. As expected, elongating peduncles exhibited a distinct separation from the non-elongating ones (**Figure 1C**). Furthermore, the results revealed a higher correlation among EP samples of both the cultivars. In contrast, non-elongating peduncles from both the cultivars were well-separated indicating genotypic differences in the transcriptional repertoire of non-elongating peduncles (**Figure 1C**).

Further, genes expressed with FPKM ≥ 1 in all four samples were compared (**Figure 1D**). The total number of genes expressing in EP and NP stages in both the genotypes were comparable with more than 16,000 genes expressing at EP stage and over 15,000 genes expressing in NP samples from both the genotypes. A total of 12,726 transcripts were common to both EP and NP samples. In both the genotypes, Swarna had higher number of genes expressing exclusively in elongating (609) as well as non-elongating (841) peduncles (**Figure 1D**).

In silico Validation of Gene Expression Patterns

Several genes in rice have been previously implicated in peduncle elongation and panicle exsertion at the time of heading using forward and/or reverse genetic approaches. To investigate if we could capture their differential accumulation during peduncle elongation, we checked the expression patterns of nine previously characterized genes in our data (**Figure 2**). Among these, *OsPK1* encoding a pyruvate kinase has been previously implicated in regulating plant height, panicle exsertion and carbohydrate transport during grain filling (Zhang et al., 2012). The mutant exhibits dwarfism, panicle enclosure, and reduced seed set. Consistent with its demonstrated role, *OsPK1* expressed at high levels in elongating peduncles before heading which escalated further in non-elongating peduncles post heading in both the genotypes (**Figure 2**). Another gene, *OsSUT1*, encoding the sucrose transporter, has been previously shown to express in rice peduncle with a critical role in facilitating the transport of assimilates from the flag leaf blade to the base of filling grains (Scofield et al., 2002, 2007). We observed a significant increase in *OsSUT1* expression post-heading in peduncles of both the genotypes (**Figure 2**).

A cytochrome P450 gene, *CYP714B1*, that encodes gibberellin 13-oxidase has been shown to reduce gibberellin (GA) activity in rice (Magome et al., 2013). Double mutation in CYP714B1

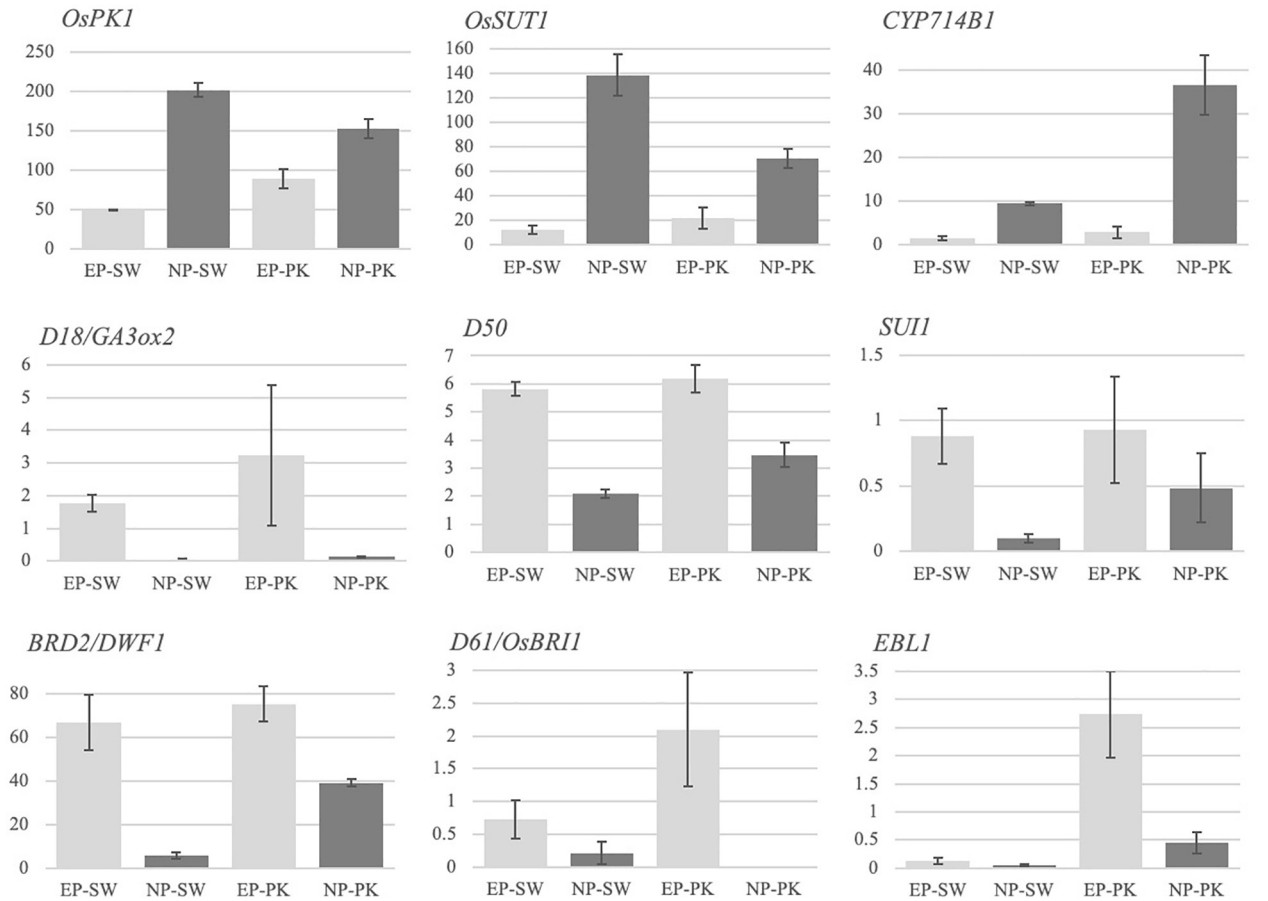


FIGURE 2 | Expression profiles of previously characterized genes for involvement in peduncle elongation and/or panicle exsertion. The *Y*-axis represents FPKM values, and the *X*-axis represents samples. Error bars represent standard error between biological replicates. SW, Swarna; PK, Pokkali; EP, elongating peduncles, and NP, non-elongating peduncles. The names of the genes are given in italics on the top left of each bar graph.

and *CYP714B2* resulted in elongation of uppermost internode in rice at the time of heading indicating their role in suppressing internode elongation post heading by downregulating GA activity. Conforming to these observations, *CYP714B1*, with negligible expression before heading, showed high expression in rice peduncles after heading in both the rice genotypes (**Figure 2**), whereas expression of a GA biosynthetic gene, *GA3Ox2* (*Gibberellin 3 beta-dioxygenase 2*) that plays a decisive role in internode elongation in rice, was detected in elongating peduncles pre-heading in both the genotypes (Iwamoto et al., 2011; **Figure 2**).

Dwarfism 50 (*D50*) encoding inositol polyphosphate 5-phosphatase is essential for proper formation of intercalary meristem by regulating the direction of cell division, deposition of cell wall pectin, and actin organization (Sato-Izawa et al., 2012). As expected, *D50* exhibited higher expression in rice peduncles before heading with a significant decline observed post heading in both the genotypes (**Figure 2**). Similarly, another gene, *SUI1* (*Shortened Uppermost Internode 1*) encoding phosphatidylserine synthase, mediates cell expansion during elongation of uppermost internode in rice by regulating the secretion of cell wall components (Zhu et al., 2011; Yin et al., 2013; Ma et al., 2016). The *sui1* mutants exhibit immensely shortened uppermost internode and a sheathed panicle indicating its role in peduncle elongation and panicle exsertion at the time of heading. *SUI1*, though expressed at very low levels, exhibited a significant decline in expression post heading in both the genotypes (**Figure 2**).

Furthermore, rice genes associated with brassinosteroid biosynthesis and signaling have also been shown to affect internode elongation. Loss of function of *BRD2* (*Brassinosteroid-deficient dwarf 2*) involved in the brassinosteroid biosynthetic pathway and a putative BR receptor kinase, *BRI1* (*Brassinosteroid Insensitive 1*), leads to dwarf phenotypes confirming their role in internode elongation (Yamamuro et al., 2000; Hong et al., 2005). Both the genes exhibited a higher expression in elongating peduncles before heading in rice genotypes with a significant decline post heading in our data. Similarly, phytochrome-regulated *OsEREBP1*-like transcription factor *EBL1* which regulates internode elongation at the heading stage in rice by upregulating the expression of *ACO1* (*aminocyclopropane-1-carboxylate oxidase*), an enzyme associated with ethylene

biosynthesis (Iwamoto and Takano, 2011), exhibited a significant decline in expression in Pokkali plants post heading (**Figure 2**). On the contrary, *EBL1* expression in Swarna plants was negligible both before and after heading. The differential expression of *EBL1* could be due to differential photoperiod sensitivity or response to ethylene levels in both the genotypes. Overall, we obtained the expected expression patterns of the previously characterized genes conforming to their demonstrated functions.

Differential Gene Expression Analysis and Pathway Enrichment

To identify genes exhibiting differential accumulation during peduncle elongation at the time of heading in both the genotypes, we used a cutoff ≥2-fold change with *P*-value ≤0.05. Expression levels of 8,894 and 5,018 genes were altered in Swarna and Pokkali, respectively. A total of 4,280 and 2,348 genes were expressed at higher levels in elongating peduncles of Swarna and Pokkali, respectively (**Figure 3A**), whereas 4,614 and 2,670 genes expressed at higher levels in non-elongating peduncles of Swarna and Pokkali, respectively. Since our study aimed to identify the genes with conserved roles in peduncle elongation instead of genotype-specific differences, we compared the differentially expressed gene sets from both the genotypes. A total of 1,500 genes were expressed at higher levels in EPs of both the genotypes compared to NPs (**Figure 3B**). Similarly, 1,723 genes expressed at higher levels in non-elongating peduncles. A large number of genes exhibiting contrasting transcript profiles imply genotype-specific variations likely responsible for variation in cell wall, biomass composition, stress tolerance, and other physiological parameters (**Figure 3B**).

To identify critical genes and pathways underlying peduncle elongation, irrespective of the genotype, we performed pathway enrichment analysis with 1,500 and 1,723 genes exhibiting differential expression in peduncles in both the genotypes (**Supplementary Table 2**). Among 1,500 genes with higher expression in elongating peduncles, those involved in transcriptional regulation, posttranslational modifications, cell organization, DNA synthesis, cell wall biosynthesis, lipid metabolism, stress response, cell cycle, and auxin metabolism were particularly enriched (**Figure 3C**). On the other hand, the enriched categories among 1,723 genes exhibiting higher expression in non-elongating peduncles were transcriptional regulation, protein degradation, posttranslational modification, development, calcium signaling, stress response, amino acid metabolism, sugar transport, and ethylene metabolism (**Figure 3C**). These observations conform with the developmental events involving active cell division in the elongating peduncles and mobilization of carbohydrates in non-elongating peduncles after heading. Among those exclusively upregulated in elongating peduncles of Swarna, protein synthesis, RNA processing, development, and photosynthetic light reaction categories were enriched. On the contrary, genes associated with vesicular transport, signaling, and cell wall biosynthesis were exclusively enriched in elongating peduncles of Pokkali (**Supplementary Table 5**). Interestingly, genes associated with photosynthesis were exclusively upregulated in non-elongating peduncles of Pokkali, whereas those associated with protein degradation, posttranslational modification, signaling, lipid metabolism, and protein targeting were exclusively enriched in non-elongating peduncles of Swarna plants (**Supplementary Table 5**).

In-Depth Analysis of Differentially Expressed Genes Associated With Transcriptional Regulation, Cell Wall, Carbohydrate, and Hormone Metabolism

Taking cues from the pathways enriched during peduncle elongation, we performed an in-depth profiling of differentially expressed genes associated with cell wall biosynthesis and modification, carbohydrate metabolism and transport, hormone biosynthesis, signaling, and response, and transcriptional regulation.

Cell Wall Biosynthesis and Modification

A total of 24 cellulose synthases (*CESA*) and cellulose synthase-like (*CSL*) genes were differentially expressed in both the genotypes with 22 of them exhibiting higher expression in elongating peduncles compared to non-elongating ones (**Figure 4A**; **Supplementary Table 3A**). The CSL genes exhibiting higher expression in elongating peduncles belong to the CSLA, CSLC, CSLC, CSLD, and CSLF families (**Figure 4A**). *CSLE6* and *CSLH1*, on the contrary, exhibited higher expression post heading (**Figure 4A**). Several genes implicated in secondary metabolite biosynthesis also exhibited higher expression in post heading peduncles (**Figure 4B**).

However, genes associated with lignin biosynthesis showed a mixed pattern with *PAL* (phenylalanine ammonia-lyase) and *CCR* (cinnamoyl-CoA reductase) genes expressing at higher levels in EPs while *4CL* (4-coumarate: coenzyme A ligase) and *CCOMT* (caffeoyl CoA 3-O-methyltransferase) exhibited higher expression in non-elongating peduncles (**Figure 4C**; **Supplementary Table 3B**). Further, three genes encoding 4CLs (4-coumarate: coenzyme A ligases) and COMT (caffeic acid 3-O-methyltransferase) exhibited contrasting patterns in both the genotypes with higher expression in elongating peduncles of Pokkali and non-elongating peduncles of Swarna rice. These results may be explained by varying levels of lignification of peduncles after heading in both the genotypes.

Carbohydrate Metabolism and Transport

Since the peduncle comprises an essential component of the transport system facilitating carbohydrate mobilization during grain filling post heading, we investigated the expression of rice genes implicated in sucrose and starch metabolism as well as transport in our data (**Supplementary Table 3C**). Several genes regulating sucrose and starch degradation, as well as sugar transport, exhibited differential expression in peduncles at the time of heading (**Figure 5**). Notable among those were sucrose synthases (Huber and Akazawa, 1986; Stein and Granot, 2019). Rice *SUS4* (*sucrose synthase 4*), which has previously been shown to be associated with grain filling exhibited higher expression in post heading peduncles, whereas *SUS1* (*sucrose synthase 1*) having a more ubiquitous role was downregulated after heading

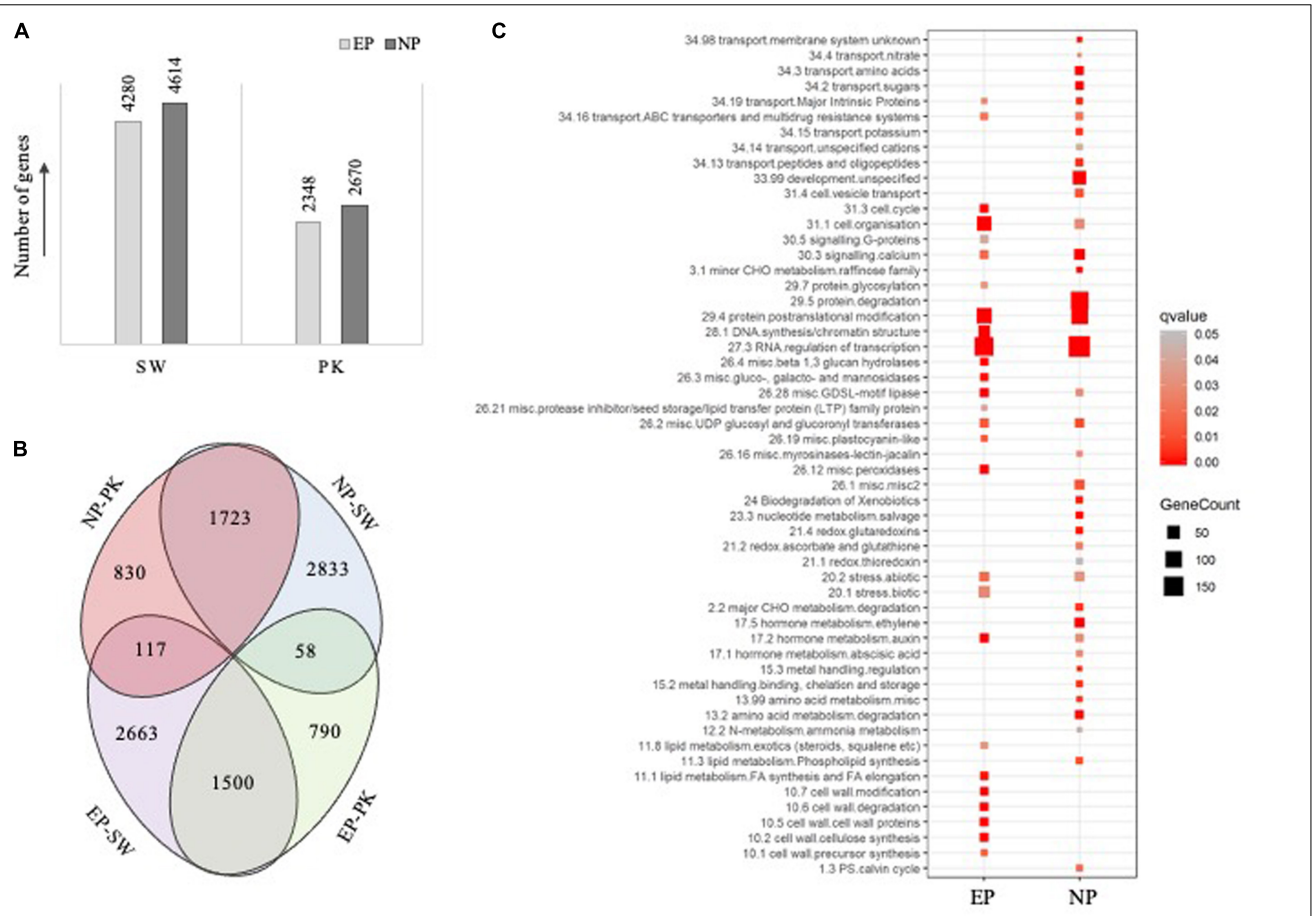


FIGURE 3 | Results of differential expression analysis and pathway mapping. **(A)** Bar diagram showing the number of genes differentially expressed (fold change ≥ 2 and *P*-value ≤ 0.05) at the time of heading with EP representing the number of genes exhibiting higher expression in elongating peduncles before heading and NP representing number of genes exhibiting higher expression post heading. SW, Swarna; PK, Pokkali. **(B)** Venn diagrams showing the overlap between differentially expressed genes in both the genotypes. **(C)** Pathway enrichment analysis of the differentially expressed genes, with EP representing genes exhibiting higher expression in elongating peduncles before heading and NP representing genes exhibiting higher expression in non-elongating peduncles post heading, in both the genotypes. The size of square correlates with number of genes in each category while color signifies *P*-values as shown by legend on the right.

in peduncles (**Figure 5A**). Antagonistic patterns of these genes are likely due to their differential activity in different tissues. *SUS1* is implicated in the regulation of vegetative growth, while *SUS4* is involved in assimilate portioning to the caryopsis during grain filling (Sparks(ed.), 2012).

Invertases modulate the hexose to sucrose ratio by hydrolyzing sucrose into glucose and fructose. Most of the invertases exhibited a higher expression in post heading peduncles (**Figure 5A**) Both *cytosolic invertase 1* (*CIN1*) and vacuolar invertase (*INV3*) were upregulated after heading (Morey et al., 2018). A similar trend was observed for other starch metabolism and sugar transport-related genes with significant upregulation after heading (**Figure 5B**).

Before loading into the sieve element–companion cell complex through sucrose transporters, sucrose is effluxed from phloem parenchyma cells by *SWEET* (*Sugars Will Eventually Be Exported Transporters*) genes (Ayre, 2011; Chen et al., 2012). We observed significant upregulation of several *SWEET* genes in rice peduncles after heading (**Figure 5C**). Besides, the genes encoding ABC, potassium, calcium, peptides, amino acids, nucleotides, phosphate, sulfate, ammonium, nitrate, and *P*- and v-ATPase transporters also exhibited a higher expression post heading. Several genes associated with carbohydrate partitioning such as *SUT1* and *GWD1* also exhibited higher expression in peduncles after heading (**Figure 5C**). The sucrose transporter-encoding gene *OsSUT1* has previously been shown to play a role in long-distance transport of assimilates from the flag leaf blade to the base of filling grains through the flag leaf sheath and uppermost internode, referred to as peduncle in this study (Scofield et al., 2002, 2007). *GWD1*, encoding alpha-glucan water dikinase, has been implicated in carbohydrate partitioning after heading in rice (Wada et al., 2017). Loss of function of *GWD1* led to hyperaccumulation of starch in leaves (Hirose et al., 2013).

Hormone biosynthesis, signaling, and response

Among differentially expressed genes, 68 genes have been previously implicated in plant hormone biosynthesis/signaling (**Figure 6**; **Supplementary Table 3D**). Although some of

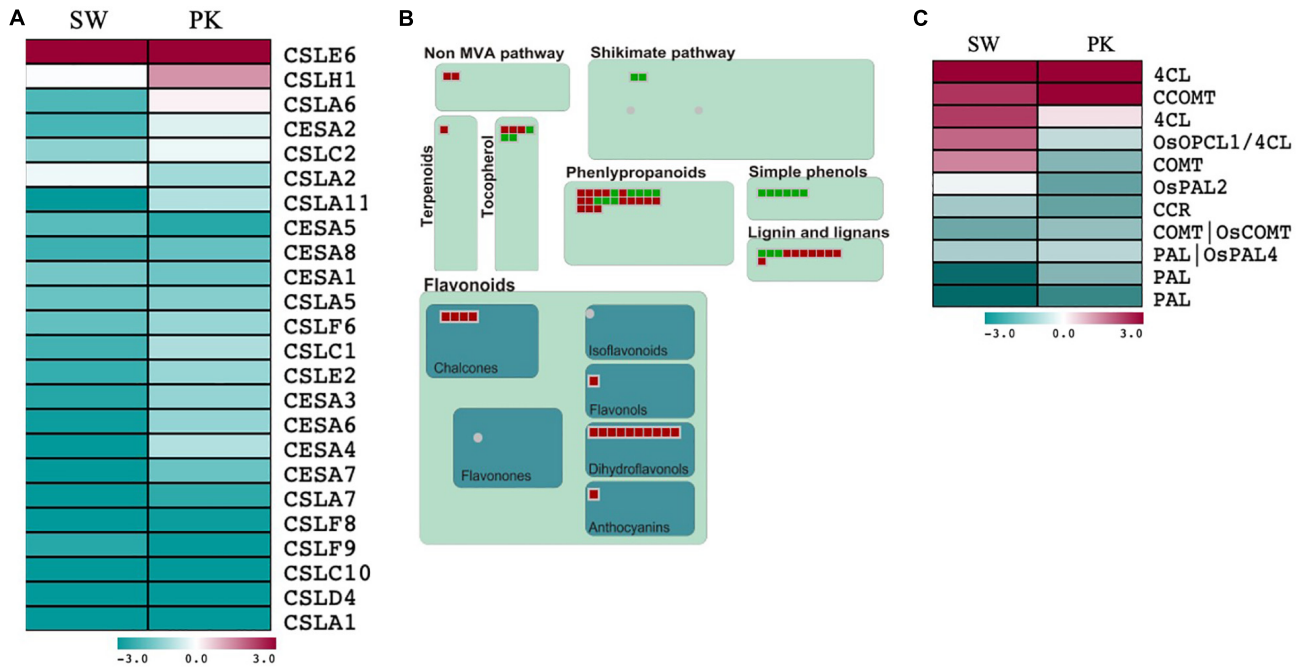


FIGURE 4 | Expression profiling of genes associated with cell wall biosynthesis and secondary metabolism. **(A)** Heatmap showing differential expression of cell wall biosynthesis-related genes during heading in Swarna (SW) and Pokkali (PK) plants. Teal blue color represents genes exhibiting ≥2-fold expression in elongating peduncles while cayenne red indicates ≥2-fold expression in non-elongating peduncles post heading. Gene names are given on the right of the heatmap. **(B)** MapMan secondary metabolism overview with green boxes presenting genes with higher expression in elongating peduncles and red boxes representing genes exhibiting higher expression in non-elongating peduncles. **(C)** Heatmap showing differential expression of lignin biosynthesis-related genes. Teal blue color represents genes exhibiting ≥2-fold expression in elongating peduncles while cayenne red indicates ≥2-fold expression in non-elongating peduncles post heading. Gene names are given on the right of the heatmap.

them showed a contrasting pattern, by and large genes associated with auxins and brassinosteroids (BR) exhibited higher expression in elongating peduncles before heading, whereas those regulating abscisic acid (ABA), jasmonic acid (JA), and cytokinin metabolism were predominantly upregulated post heading (**Figure 6**). The genes associated with GA and ethylene biosynthesis/signaling and response exhibited a mixed pattern though overall trend suggested higher expression of larger number of genes post heading (**Figure 6**).

Similarly, although majority of ABA-biosynthesis/signaling genes expressed at higher levels in post-heading peduncles, some exhibited contrasting expression patterns in both genotypes. Several of the ABA signaling-related genes which exhibited a higher expression after heading in peduncles of Swarna, expressed at higher levels before heading in Pokkali plants (**Figure 6**). Conversely, some of the genes associated with ABA deactivation that expressed at higher levels before heading in Swarna were upregulated in peduncles of Pokkali. These observations point to ABA activation in Swarna while ABA deactivation in Pokkali after heading.

Transcription Factor-Encoding Genes

A total of 836 transcription factors (TFs), belonging to more than twenty families, were differentially expressed in rice peduncles after heading (**Supplementary Table 3E**). Albeit members of the same transcription factor family do not necessarily perform the same function, we observed a large number of genes belonging to the same family exhibiting similar expression patterns, thereby indicating similar/coordinated functions of the gene family members. Genes belonging to ARF, AUX-IAA, GRF, HMG, OFP, Trihelix, PHD, and TCP families, mostly implicated in cell division and elongation, exhibited a higher expression in elongating peduncles (**Figure 7**), whereas those belonging to ARR-B, C2C2-CO like C2C2-Dof, MADS, TRAF, AP2-EREBP, bZIP, NAC, G2-like, HSF, OPHANS, WRKY, TIFFY, REP-RK, and PLATZ families expressed at higher levels post heading. Members of C3H, SNF2, C2H2, bHLH, GRAS, and HB exhibited mixed patterns (**Figure 7**).

Differential Expression of Anther/Pollen-Associated Genes During Peduncle Elongation

Incomplete peduncle elongation and panicle enclosure are frequently observed in male sterile lines. To investigate if the genes associated with peduncle elongation have anything to do with anther development, we compared the differentially expressed genes during peduncle elongation with those associated with anther/pollen development. Previously, Lin et al. (2017) had generated a rice gene co-expression network for anther development (RiceAntherNet) from 57 rice anther tissue microarrays. They mapped differentially expressed genes

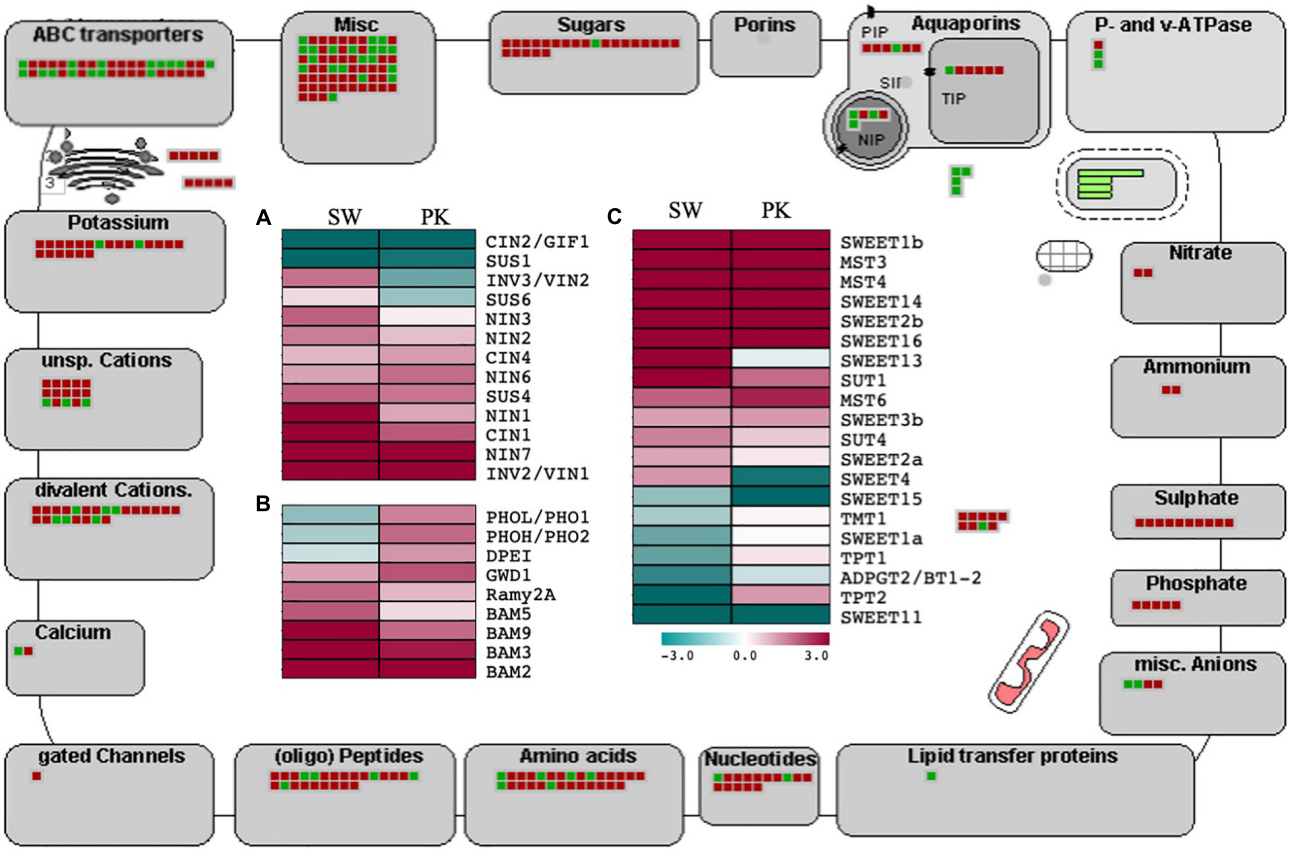


FIGURE 5 | Expression profiling of genes associated with carbohydrate metabolism and transport. The outer illustration is the MapMan transport overview with green boxes presenting genes with higher expression in elongating peduncles and red boxes representing genes exhibiting higher expression in non-elongating peduncles. The inner portion presents heatmaps showing differential regulation of genes associated with sucrose metabolism **(A)**, starch metabolism **(B)**, and sugar transport **(C)** Swarna (SW) and Pokkali (PK) plants. Teal blue color represents ≥2 folds expression in elongating peduncles and cayenne red indicates ≥2 folds expression in non-elongating peduncles post heading. Gene names are given on the right of each heatmap.

from nine rice male-sterile mutants onto this network and shortlisted a set of 286 genes associated with pollen formation. We appended this list with anther/pollen-associated genes extracted from Funrice and OGRO databases of rice. Finally, a list of 493 genes implicated in anther/pollen development in rice was compiled. Out of these 493 genes, 76 (15%) were differentially expressed in peduncles at the time of heading in both the genotypes (**Figure 8**; **Supplementary Table 4**). Out of these, 50 overlapped with genes exhibiting ≥2-fold expression in EPs, whereas 26 overlapped with genes exhibiting ≥2-fold expression in NPs (**Figure 8**). Most of the anther development-related genes that exhibited higher expression in elongating peduncles before heading (50 in number) encode glycoside hydrolases, serine carboxypeptidases, kinases, and other cell wall-related proteins. Some of these genes have already been characterized for their roles in regulating plant height, grain filling and/or anther/pollen development. Conversely, those overlapping with high-expressing genes in non-elongating peduncles (26 in number) as well as anther development have been implicated in abiotic stress and grain filling.

DISCUSSION

Peduncle elongation at the time of heading not only is crucial for panicle exsertion but also exerts significant impact on grain yield. To identify molecular players regulating peduncle elongation and associated agronomically important traits, we have performed RNA sequencing with peduncles before and after heading collected from two contrasting genotypes of rice. Although the total number of expressed genes (FPKM ≥1) was similar in both the genotypes, those exhibiting differential accumulation between EP and NP stages in Swarna were almost double in number compared to differentially expressed genes in Pokkali. This could possibly be because Swarna has to make more transcriptional adjustments to cope up with the abiotic stress post heading compared to the relatively stress-tolerant Pokkali. Further, most of the genes exclusively differentially expressed in Swarna are involved in the basic metabolic functions including protein synthesis, degradation and posttranslational modifications, signaling, RNA processing, lipid metabolism, transport, development, and abiotic stress response. PCA analysis further highlighted noticeable distance

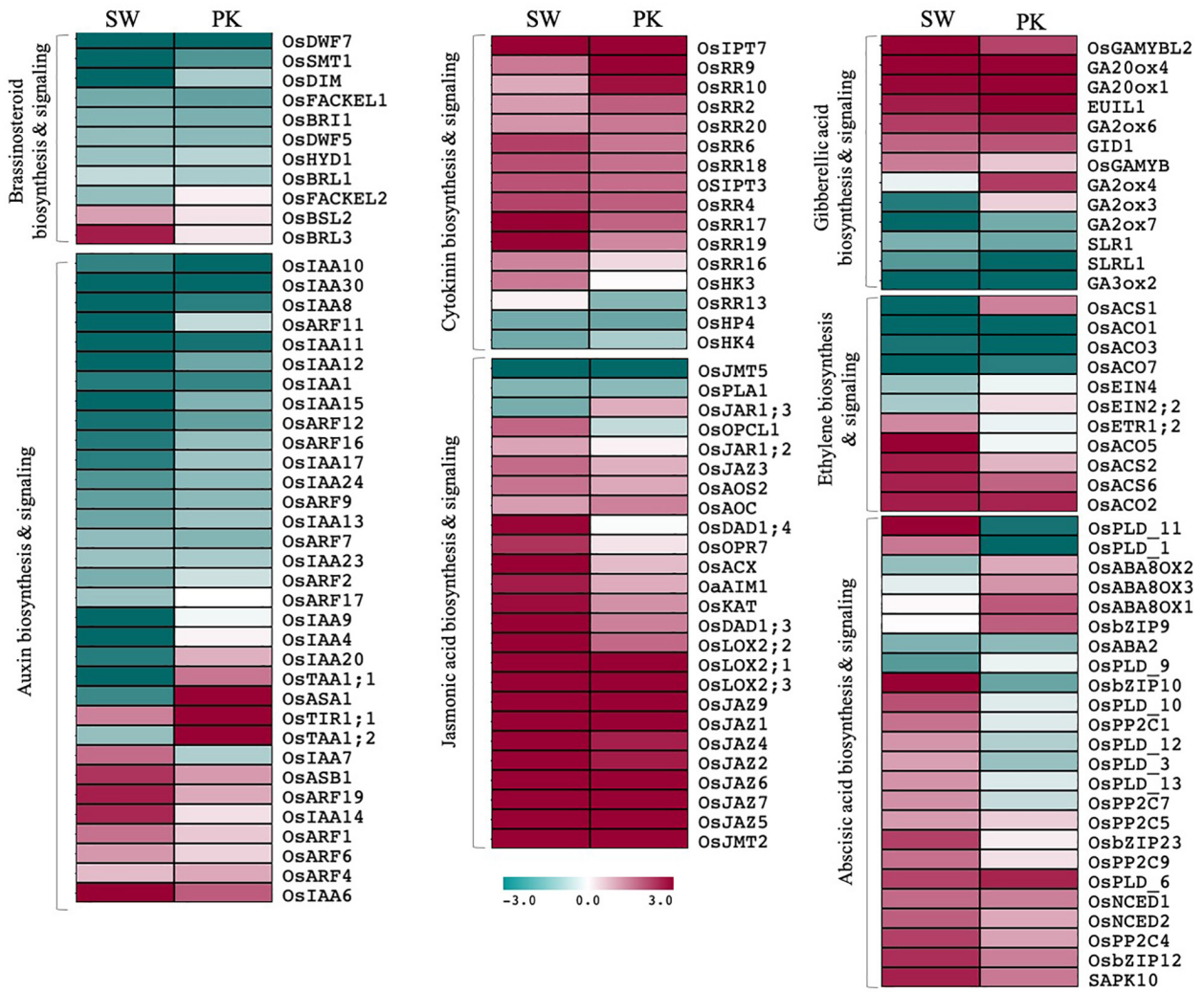


FIGURE 6 | Expression profiling of genes associated with hormone biosynthesis and signaling. Heatmaps showing differential expression of genes associated with hormone biosynthesis and signaling in Swarna (SW) and Pokkali (PK). Teal blue color represents ≥2 folds expression in elongating peduncles and cayenne red indicates ≥2 folds expression in non-elongating peduncles post heading. Gene names are given on the right of the heatmaps and hormone categories are given on the left.

between non-elongating peduncles compared to elongating peduncles likely due to differences in post heading stem biomass composition and stress response of these genotypes (Jahn et al., 2011). However, when we compared the expression patterns of previously characterized genes which have been experimentally shown to regulate peduncle elongation and grain filling in rice, all of them exhibited a similar expression pattern in both the genotypes though the amplitude of expression and fold change varied, indicating that the genetic factors underlying peduncle elongation are conserved across genotypes. Furthermore, changes observed at the transcriptional level may not reflect in protein abundance. Earlier, comparative analysis of transcript and protein abundance of Pokkali and IR29 revealed a much higher number of genes differentially expressed in Pokkali at the protein level with little difference at the transcript level in response to salt stress (Li et al., 2018). In fact, 75.5% of genes exhibiting high protein abundance did not show relatively high transcriptional abundance in Pokkali. The authors suggested that higher stability and efficient loading of mRNAs in Pokkali could be the possible explanation for lower number of differentially expressed transcripts in Pokkali.

Candidate Genes Associated With Peduncle Elongation and Grain Filling

Elongation of peduncles is accompanied by active cell division and cell wall biosynthesis. This is obvious from enrichment of glycosyltransferase (GTs) in elongating peduncles of both the genotypes. Glycosyltransferases comprise a large gene family with more than 600 genes in rice which have been classified into subfamilies based on the presence of distinct sequence domains (Cao et al., 2008). Among these, CESAs belonging to the GT2

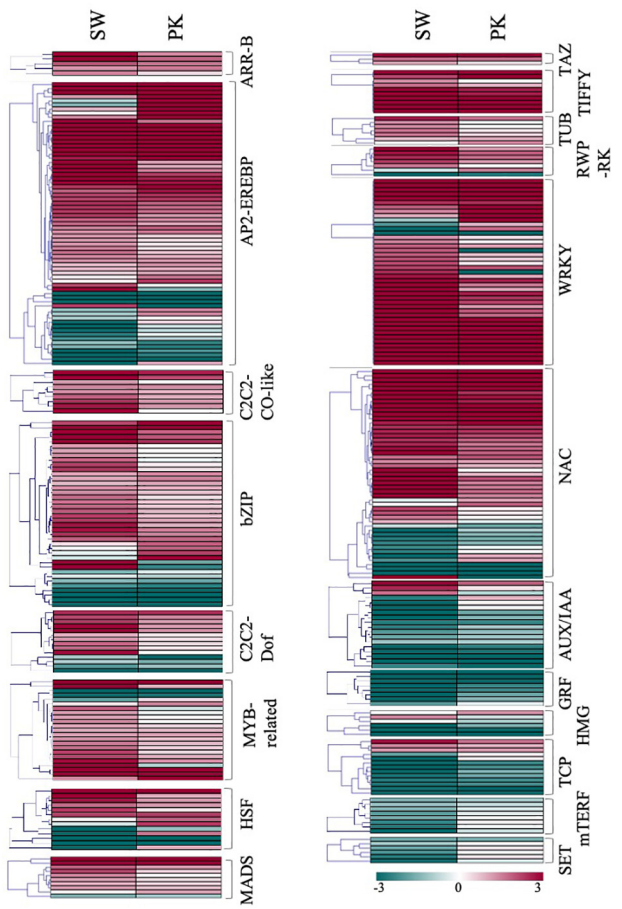


FIGURE 7 | Expression profiling of transcription factor-encoding genes. Heatmaps showing differential expression of transcription factor-encoding genes in Swarna (SW) and Pokkali (PK). Members of the same gene family are clubbed together in one heatmap. Teal blue color represents ≥2 fold expression in elongating peduncles while cayenne red indicates ≥2 fold expression in non-elongating peduncles post heading. Names of the transcription factor families are given on the right of each heatmap.

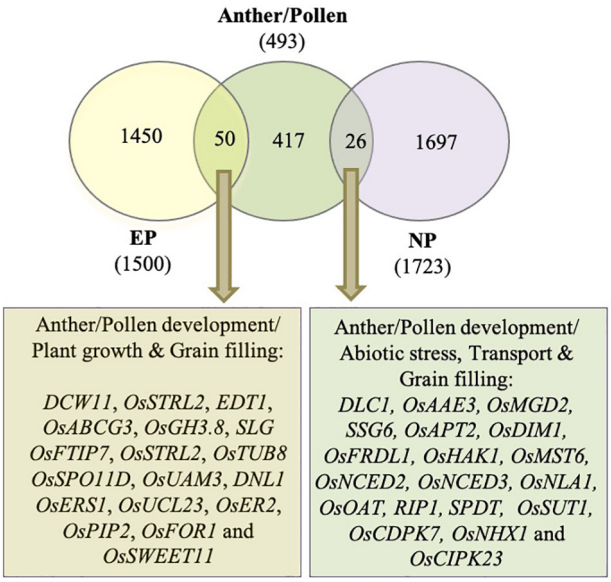


FIGURE 8 | Overlap between differentially expressed genes during peduncle elongation and anther development. The Venn diagram shows the overlap between genes exhibiting ≥2 fold expression in elongating peduncles (EP), non-elongating peduncles (NP), and anther/pollen development. The key pathways and characterized genes highlighted in the gene set common between EP/NP and anther/pollen associated genes are given.

subfamily are mainly involved in cellulose biosynthesis, while cellulose synthase-like (CSL) genes play key roles in hemicellulose biosynthesis (Richmond and Somerville, 2001). CSLAs form the β-1,4-mannan backbone of galactomannan and glucomannan, CSLDs determine cellulose and xylose content, whereas, CSLFs are essential for grass-specific mixed linkage glucan biosynthesis (Li et al., 2009; Luan et al., 2011; Vega-Sánchez et al., 2012; Kaur et al., 2017). Many of these have been shown to exhibit tissue and developmental stage-specific expression. In-depth analysis of these genes in our data provided 22 candidates with significant accumulation in EPs relative to NPs implying their potential role in regulating cell wall biosynthesis and biomass composition in elongating rice peduncles.

Also, a shift from primary to secondary cell wall biosynthesis is anticipated at the time of heading for providing mechanical strength required to support increase in panicle weight due to grain filling. *OsMYB103* has earlier been reported to play a role in secondary cell wall in sclerenchyma by activating *OsCESA4*, *OsCESA7*, *OsCESA9*, and *OsBC1* (Hirano et al., 2013; Yang et al., 2014; Ye et al., 2015). All of these genes exhibit higher accumulation in EPs indicating that molecular players responsible for secondary cell wall biosynthesis are initiated before heading. On the other hand, higher accumulation of genes associated with biosynthesis of secondary metabolites in NPs indicates their role in post-heading resistance to pathogens though some of them may also act as precursors for secondary cell wall components. Genes involved in lignin biosynthesis exhibited contrasting pattern of expression in Swarna and Pokkali likely due to varied lignin composition in these genotypes (Jahn et al., 2011).

Among the differentially expressed transcription factor families, members of MYB, NAC, WRKY, bZIP, PLATZ, and HSF transcription factors have been previously associated with secondary growth, abiotic stress response, and carbon mobilization (Oh et al., 2003; Agarwal et al., 2006; Zhang et al., 2010; Golldack et al., 2014). Further characterization of these genes would not only help to decipher the molecular mechanism underlying peduncle elongation and biomass composition but will also help in optimizing carbon remobilization in rice.

Interplay of Phytohormones During Peduncle Elongation in Rice

Phytohormones play crucial roles in plant growth and development. Previous studies have shown the involvement of auxins, gibberellins, and brassinosteroids in promoting cell division and growth during internode elongation in rice (Chen,

2001; Yin et al., 2002; Tong et al., 2014; Oh et al., 2020). For instance, BR receptor kinase *OsBRI1* of rice positively regulates internode elongation by inducing the formation of the intercalary meristems (Yamamuro et al., 2000). Loss of function of this gene prevents internode elongation and bending of the lamina joint. We noticed a high expression of *OsBRI1* before heading in both the genotypes with an apparent decline after heading. Except for *OsBRL3*, which is known to play a role in BR perception in roots (Nakamura et al., 2006), all other brassinosteroid-related genes also exhibited a higher expression in elongating peduncles before heading. Similarly, auxins also act as a signal for cell elongation (Chen, 2001). Developing panicles in grasses export auxin to stem for promoting elongation (Wolbang et al., 2004). Except for a few, most of the auxin-related genes were expressed at higher levels in rice peduncles before heading. Among the genes exhibiting higher expression post heading, *OsARF19* is a negative regulator of cell division as its overexpression in rice leads to dwarf phenotype. Furthermore, *OsARF19* directly regulates the expression of BR receptor *OsBRI1* by binding to its regulatory region (Zhang et al., 2015). Upregulation of *OsARF19* and downregulation of *OsBRI1* after heading therefore suggest that *OsARF19* likely suppresses peduncle elongation after heading by downregulating *OsBRI1*.

Gibberellic acid (GA) has also been shown to act as a key determining factor in peduncle elongation and panicle exsertion (Yin et al., 2007; Chen et al., 2013). We observed a mixed pattern of the genes associated with GA and ethylene biosynthesis/signaling likely due to multilevel cross-talk or negative feedback regulation of some of them. GA biosynthetic gene, *gibberellin 3-beta-hydroxylase* (*GA3Ox*), which catalyzes the final step of GA biosynthesis in plants by converting GA20 to GA1, exhibited a higher expression before heading with notable downregulation after heading (Iwamoto et al., 2011; **Figure 6**). *OsGA3Ox2* corresponds to the *D18* mutant that exhibits severe dwarfism due to decrease in level of active GA1 (Itoh et al., 2002). Antisense plants with reduced *OsGA3Ox2* expression exhibited a higher expression of another gene *OsGA20Ox1* that regulates the synthesis of GA21 from GA53 (Itoh et al., 2002). Interestingly, both the genes in our study exhibited antagonistic patterns with *GA3Ox2* exhibiting downregulation and *GA20Ox1* showing upregulation after heading. Conversely, GA receptor *Gibberellin Insensitive Dwarf 1* (*GID1*) which makes a complex with GA and induces degradation of GA repressing DELLA proteins is upregulated (Ueguchi-Tanaka et al., 2007; Sun, 2008) while DELLA protein-encoding genes, *SLR1* (*SLENDER RICE 1*) and *SLR1 Like-1* (*SLRL1*), are downregulated after heading, indicating that suppression of GA activity after heading is not mediated through DELLA proteins. It has been shown that a high concentration of BR triggers BR-associated inactivation of GA. The BR signaling component involved in this inactivation process, *BRI1-GSK2-DLT*, is upregulated before heading in our data (Tong et al., 2014). The effect of this can be observed with increased expression of GA deactivation genes after heading. For instance, *CYP714B1* encoding gibberellin 13-oxidase which reduces gibberellin activity in rice as well as *GA2ox4* and *GA2ox6*, implicated in endogenous GA deactivation (Chu et al., 2019), expressed at higher levels in peduncles after heading (Lo et al., 2008; Magome et al., 2013). Overall, regulation of *OsBRI1* by *OsARF19* and its (*OsBRI1*) role in deactivation of GA indicates interaction of auxin, brassinosteroids, and gibberellins in maintaining the hormonal homeostasis during peduncle elongation (Zhang et al., 2015).

Cytokinins, on the other hand, have been associated with post-anthesis grain filling in rice (Yang et al., 2002; Zhang et al., 2009). Since mostly cytokinins are transported from roots to shoots and other aerial organs *via* xylem sap, elevated levels of cytokinins in rice peduncles after heading suggest their role in grain filling (Yang et al., 2002; Zhang et al., 2009). Panicle enclosure has been shown to adversely affect spikelet fertility and grain yield (He and Serraj, 2012). A genome-wide association study using 205 rice cultivars clearly drew a positive correlation between panicle exsertion and 1,000-Grain Weight (Zhan et al., 2019). The grain filling is facilitated by transport of carbohydrate resources stored in elongating stem in rice in the form of starch and sucrose (Scofield et al., 2007, 2009; Slewinski, 2012). The carbon remobilization from the stem reserves to grains possibly occurs by an apoplasmic mechanism through sugar transporters (Slewinski, 2012). Many of the sugar transport genes were upregulated in the peduncles suggesting their role in grain filling. For example, *OsSUT1*, a member of the sucrose transporter family, is required for actively pumping sucrose into phloem against the concentration gradient (Scofield et al., 2007). In the present study, its upregulation before heading conforms with its role in grain filling. Monosaccharide transporter genes, *OsMST4* and *6*, implicated in supplying monosaccharides for seed development during grain filling, are also upregulated in NPs after heading in our study (Wang et al., 2007, 2008). *OsSUS4*, involved in portioning of the carbon assimilates to the caryopsis during grain filling (Sparks(ed.), 2012), is also upregulated in peduncles after heading. This suggests that peduncle plays a critical role in transporting carbohydrate reserves stored during stem elongation to the developing grains, and hence, rice genes exhibiting increased expression in NPs after heading may be leveraged to optimize carbohydrate mobilization in rice.

Jasmonic acid (JA) is another important phytohormone associated with biotic and abiotic stress response and is known to inhibit shoot growth in rice explaining the upregulation of JA biosynthetic genes after heading (Liu et al., 2015). The level of JAs has also been associated with availability of carbon and soluble sugars, indicating the role of JA in carbon remobilization (Huang et al., 2017). Upregulation of ABA-related genes after heading has been associated with abiotic stress response during reproductive development (Baron et al., 2012; Singh et al., 2012). Some genes implicated in ABA biosynthesis and signaling exhibited contrasting patterns of expression in both the genotypes. Further investigations would be required to understand if this can be explained by differential stress tolerance in Swarna and Pokkali post heading. Overall, our data shows a clear trend with higher expression of growth hormones, viz., AUX, GA, and BR, in elongating peduncles before heading, whereas JA, ABA, and cytokinins, associated with stress response, sugar signaling, and remobilization, exhibit higher expression after heading.

Common Genetic Elements Underlying Peduncle Elongation and Male Reproductive Organ Development

Crop improvement in rice to optimize agronomic yields and stress tolerance largely relies on breeding programs. However, panicle enclosure due to impeded peduncle elongation in male sterile lines is a major challenge in achieving these goals. A number of studies have associated enclosed spikelets with inhibition of anther dehiscence, decreased pollen viability, slower pollen tube growth, and abnormal ovary development (Ekanayake et al., 1989, 1990; Jagadish et al., 2011, 2015). A comparative study of cytoplasmic male sterile line with its maintainer lines during panicle development revealed increased levels of ABA and reduced levels of GA and IAA in the cytoplasmic male sterile line (Chang'en et al., 1998). Deficiency of IAA and GAs and excess of ABA hinder anther development and induce pollen abortion (Shimizu and Kuno, 1967; Nakajima et al., 1991; Sunohara et al., 2009). Later, Yin et al. (2007) showed that deficiency of indole-3-acetic acid in panicle of male sterile lines downregulates GA oxidase (*OsGA3ox2*) resulting in decreased active GA levels, thereby hampering peduncle elongation manifested in the form of enclosed panicles and poor seed set (Yin et al., 2007). In the present study, several genes related to anther and pollen development were differentially expressed in both elongating and non-elongating peduncles. Although it is a well-known fact that male sterile lines exhibit varying degrees of panicle enclosure, the genetic basis of the relationship between anther development and peduncle elongation has not been established till date. To identify the common genetic elements regulating peduncle elongation and anther/pollen development, we compiled a list of 493 genes implicated in anther/pollen development and checked their expression in peduncles.

Most of the anther development-related genes, also exhibiting higher expression in elongating peduncles before heading (50 in number), encode glycoside hydrolases, serine carboxypeptidases, and kinases. Some of these genes have already been characterized for their roles in anther development and/or internode elongation. Among these, *DCW11* (Fujii and Toriyama, 2008), *OsSTRL2* [strictosidine synthase; (Zou et al., 2017)], *EDT1* [*Earlier Degraded Tapetum 1*; (Bai et al., 2019)], *OsABCG3* [ATP-binding cassette transporter; (Chang et al., 2018)], *OsGH3.8* (Yadav et al., 2011), *OsSPO11D* (Shingu et al., 2012), *OsUAM3* (Sumiyoshi et al., 2015), and *OsUCL23* (Zhang et al., 2020) have been shown to regulate diverse aspects of anther and pollen development. *OsSTRL2* expression was detected only in tapetal cells and microspores, and knockout mutant resulted in male sterility mainly due to defects in anther wall and pollen exine formation (Zou et al., 2017). Several others in this category have been shown to regulate both vegetative and reproductive organ development. For example, *DNL1* (*Dwarf and Narrow Leafed 1*) encoding cellulose synthase like D4 is the major QTL regulating plant height and leaf width in rice. The mutant exhibits defects in anther dehiscence and pollen formation as well (Yoshikawa et al., 2013). Similarly, antisense transgenic plants of *OsTUB8* encoding β-tubulin exhibit shorter plant height as well as reduced seed set (Yang et al., 2007). Loss of function of *OsER2* (*ERECTA*) encoding phytosulfokine receptor led to reduced plant height and panicle size by affecting cell proliferation and growth in rice (Zhang et al., 2018). Furthermore, *SLG* (*Slender Grain*) from this group encodes BAHD acyltransferase and has been shown to regulate leaf angle, grain length, and plant height by modulating brassinosteroid levels (Liu et al., 2016). Interestingly, *OsSWEET11*, involved in sucrose transport and maintaining the sucrose concentration in the embryo-sac, also exhibited higher expression in EPs before heading (Ma et al., 2017).

Similarly, some of the genes exhibiting high-expressing genes in non-elongating peduncles as well as anther development have been experimentally shown to play role in grain filling and abiotic stress response. Among these, *OsMGD2* (monogalactosyldiacylglycerol synthase) affects plant height, anther development, and overall rice grain yield (Basnet et al., 2019). *SSG6* (*Substandard Starch Grain 6*) has previously been shown to regulate starch grain morphology and size in pollen and seeds (Matsushima et al., 2016). The *ssg6* mutant affected duration to heading, culm length, number of panicles, and seed weight. *OsSUT1* an important sucrose transporter is also upregulated after heading. Disruption of *OsSUT1* does not affect pollen maturation but their function gets impaired (Hirose et al., 2010), whereas *OsMST6* (Monfared et al., 2020) and *OsNHX1* (Almeida et al., 2017) regulate abiotic stress response. Interestingly, *OsCIPK23* (calcineurin B-like interacting protein kinase) has been implicated in mediating common signaling pathways between pollination and drought stress response (Yang et al., 2008).

The pivotal roles played by these genes in regulating anther development, plant height, grain filling, and abiotic stress response at the time of heading suggest that these as crucial candidates for alleviating the problem of panicle enclosure in male sterile lines and related agronomically important traits.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157727.

AUTHOR CONTRIBUTIONS

RS conceived and designed the study. MK performed all the experiments and the data analysis. MK and RS drafted the manuscript. CV, VC, KK, AP, and MKS participated in experiment design, sampling as well as implementation of the project. All authors read and approved the manuscript.

FUNDING

This study was carried out with the financial support from JNU through UPE-II and DST (Department of Science and Technology)-Purse-II grants. RS and MKS acknowledge Department of Biotechnology, Government of India, for

Ramalingaswami fellowships. MK thanks JNU and UGC for non-NET fellowship. KK acknowledges University Grants Commission (UGC), Government of India, for Rajiv Gandhi National Fellowship. Financial assistance from Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India (ECR/2016/001581), to RS and MKS is also duly acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.584678/full#supplementary-material

Supplementary Table 1 | Summary statistics of RNA sequencing.

Supplementary Table 2A | List of genes exhibiting higher expression in non-elongating peduncles (NP) compared to elongating peduncles (EP) in both Swarna and Pokkali.

Supplementary Table 2B | List of genes exhibiting higher expression in elongating peduncles (EP) compared to non-elongating peduncles (NP) in both Swarna and Pokkali.

Supplementary Table 3A | List of differentially expressed cellulose synthase and synthase like genes in rice peduncles.

Supplementary Table 3B | List of differentially expressed lignin biosynthesis genes in rice peduncles.

Supplementary Table 3C | List of differentially expressed carbohydrate metabolism and transport-related genes in rice peduncles.

Supplementary Table 3D | List of differentially expressed hormone-related genes in rice peduncles.

Supplementary Table 3E | List of differentially expressed transcription factor encoding genes in rice peduncles.

Supplementary Table 4 | List of differentially expressed genes in rice peduncles overlapping with anther/pollen associated genes extracted from RiceAntherNet, funRiceGenes and OGRO databases.

Supplementary Table 5A | Enriched MapMan pathways among upregulated genes in elongating peduncles (EP) of both Pokkali and Swarna plants.

Supplementary Table 5B | Enriched MapMan pathways among the genes exclusively upregulated in elongating peduncles (EP) of Pokkali.

Supplementary Table 5C | Enriched MapMan pathways among exclusively upregulated genes in elongating peduncles (EP) of Swarna.

Supplementary Table 5D | Enriched MapMan pathways among upregulated genes in non-elongating (NP) peduncles of both Pokkali and Swarna plants.

Supplementary Table 5E | Enriched MapMan pathways among exclusively upregulated genes in non-elongating (NP) peduncles of Pokkali.

Supplementary Table 5F | Enriched MapMan pathways among exclusively upregulated genes in non-elongating (NP) peduncles of Swarna.

REFERENCES

Agarwal, P. K., Agarwal, P., Reddy, M. K., and Sopory, S. K. (2006). Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.* 25, 1263–1274. doi: 10.1007/s00299-006-0204-8

Almeida, D. M., Gregorio, G. B., Oliveira, M. M., and Saibo, N. J. M. (2017). Five novel transcription factors as potential regulators of OsNHX1 gene expression in a salt tolerant rice genotype. *Plant Mol. Biol.* 93, 61–77. doi: 10.1007/s11103-016-0547-7

Ayre, B. G. (2011). Membrane-transport systems for sucrose in relation to whole-plant carbon partitioning. *Mol. Plant* 4, 377–394. doi: 10.1093/mp/ssr014

Bai, W., Wang, P., Hong, J., Kong, W., Xiao, Y., Yu, X., et al. (2019). Earlier degraded tapetum1 (EDT1) encodes an ATP-citrate lyase required for tapetum programmed cell death. *Plant Physiol.* 181, 1223–1238. doi: 10.1104/pp.19.00202

Bardenas, E. A. (1965). *Morphology and Varietal Characteristics of the Rice Plant, The. Int. Rice Res. Inst. The Morphology and Varietal Characteristics of the Rice Plant. Technical Bulletin.* Manila: International Rice Research Institute.

Baron, K. N., Schroeder, D. F., and Stasolla, C. (2012). Transcriptional response of abscisic acid (ABA) metabolism and transport to cold and heat stress applied at the reproductive stage of development in *Arabidopsis thaliana*. *Plant Sci.* 188, 48–59. doi: 10.1016/j.plantsci.2012.03.001

Basnet, R., Zhang, J., Hussain, N., and Shu, Q. (2019). Characterization and mutational analysis of a monogalactosyldiacylglycerol synthase Gene OsMGD2 in rice. *Front. Plant Sci.* 10:992. doi: 10.3389/fpls.2019.00992

Cao, P. J., Bartley, L. E., Jung, K. H., and Ronald, P. C. (2008). Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. *Mol. Plant* 1, 858–877. doi: 10.1093/mp/ssn052

Chang, Z., Jin, M., Yan, W., Chen, H., Qiu, S., Fu, S., et al. (2018). The ATP-binding cassette (ABC) transporter OsABCG3 is essential for pollen development in rice. *Rice* 11:58. doi: 10.1186/s12284-018-0248-8

Chang'en, T., Jun, D., Chengye, L., Yuwen, H., and Hong-clan, L. (1998). Changes in phytohormones in cytoplasmic male sterile rice and its maintainer line during panicle development. *J. Trop. Subtrop. Bot.* 2, 137–143.

Chen, H., Jiang, S., Zheng, J., and Lin, Y. (2013). Improving panicle exsertion of rice cytoplasmic male sterile line by combination of artificial microRNA and artificial target mimic. *Plant Biotechnol. J.* 11, 336–343. doi: 10.1111/pbi.12019

Chen, H.-J., and Wang, S.-J. (2008). Molecular regulation of sink–source transition in rice leaf sheaths during the heading period. *Acta Physiol. Plant* 30, 639–649. doi: 10.1007/s11738-008-0160-8

Chen, J.-G. (2001). Dual auxin signaling pathways control cell elongation and division. *J. Plant Growth Regul.* 20, 255–264. doi: 10.1007/s003440010028

Chen, L.-Q., Qu, X.-Q., Hou, B.-H., Sosso, D., Osorio, S., Fernie, A. R., et al. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* 335, 207–211. doi: 10.1126/science.1213351

Chu, Y., Xu, N., Wu, Q., Yu, B., Li, X., Chen, R., et al. (2019). Rice transcription factor OsMADS57 regulates plant height by modulating gibberellin catabolism. *Rice* 12:38. doi: 10.1186/s12284-019-0298-6

Coyne, D. P. (1980). Modification of plant architecture and crop yield by breeding. *Hortscience* 15, 244–247.

Devi, A. S., Krishnasamy, V., Raveendran, M., Senthil, N., and Manonmani, S. (2011). Molecular and genetic dissection of peduncle elongation in rice. *Electron. J. Plant Breed.* 2, 1–7.

Duan, Y. L., Guan, H. Z., Zhuo, M., Chen, Z. W., Li, W. T., Pan, R. S., et al. (2012). Genetic analysis and mapping of an enclosed panicle mutant locus esp1 in rice (*Oryza sativa* L.). *J. Integr. Agric.* 11, 1933–1939. doi: 10.1016/S2095-3119(12)60449-3

Ekanayake, I. J., Datta, S. K. D., and Steponkus, P. L. (1989). Spikelet sterility and flowering response of rice to water stress at anthesis. *Ann. Bot.* 63, 257–264. doi: 10.1093/oxfordjournals.aob.a087740

Ekanayake, I. J., Steponkus, P. L., and De Datta, S. K. (1990). Sensitivity of pollination to water deficits at anthesis in Upland Rice. *Crop Sci.* 30, 310–315. doi: 10.2135/cropsci1990.0011183x003000020013x

Fujii, S., and Toriyama, K. (2008). DCW11, down-regulated gene 11 in CW-type cytoplasmic male sterile rice, encoding mitochondrial protein phosphatase 2C is related to cytoplasmic male sterility. *Plant Cell Physiol.* 49, 633–640. doi: 10.1093/pcp/pcn036

Gao, X., Chen, Z., Zhang, J., Li, X., Chen, G., Li, X., et al. (2012). OsLIS-L1 encoding a lissencephaly type-1-like protein with WD40 repeats is required

for plant height and male gametophyte formation in rice. *Planta* 235, 713–727. doi: 10.1007/s00425-011-1532-7

Golldack, D., Li, C., Mohan, H., and Probst, N. (2014). Tolerance to drought and salt stress in plants: unraveling the signaling networks. *Front. Plant Sci.* 5:151. doi: 10.3389/fpls.2014.00151

He, H., and Serraj, R. (2012). Involvement of peduncle elongation, anther dehiscence and spikelet sterility in upland rice response to reproductive-stage drought stress. *Environ. Exp. Bot.* 75, 120–127. doi: 10.1016/j.envexpbot.2011.09.004

Hirano, K., Kondo, M., Aya, K., Miyao, A., Sato, Y., Antonio, B. A., et al. (2013). Identification of transcription factors involved in rice secondary cell wall formation. *Plant Cell Physiol.* 54, 1791–1802. doi: 10.1093/pcp/pct122

Hirose, T., Aoki, N., Harada, Y., Okamura, M., Hashida, Y., Ohsugi, R., et al. (2013). Disruption of a rice gene for α-glucan water dikinase, OsGWD1, leads to hyperaccumulation of starch in leaves but exhibits limited effects on growth. *Front. Plant Sci.* 4:147. doi: 10.3389/fpls.2013.00147

Hirose, T., Zhang, Z., Miyao, A., Hirochika, H., Ohsugi, R., and Terao, T. (2010). Disruption of a gene for rice sucrose transporter, OsSUT1, impairs pollen function but pollen maturation is unaffected. *J. Exp. Bot.* 61, 3639–3646. doi: 10.1093/jxb/erq175

Hoang, T. M. L., Tran, T. N., Nguyen, T. K. T., Williams, B., Wurm, P., Bellairs, S., et al. (2016). Improvement of salinity stress tolerance in rice: challenges and opportunities. *Agronomy* 6:54. doi: 10.3390/agronomy6040054

Hong, Z., Ueguchi-Tanaka, M., Fujioka, S., Takatsuto, S., Yoshida, S., Hasegawa, Y., et al. (2005). The rice brassinosteroid-deficient dwarf2 mutant, defective in the rice homolog of arabidopsis DIMINUTO/DWARF1, is rescued by the endogenously accumulated alternative bioactive brassinosteroid, dolichosterone. *Plant Cell* 17, 2243–2254. doi: 10.1105/tpc.105.030973

Huang, J., Reichelt, M., Chowdhury, S., Hammerbacher, A., and Hartmann, H. (2017). Increasing carbon availability stimulates growth and secondary metabolites via modulation of phytohormones in winter wheat. *J. Exp. Bot.* 68, 1251–1263. doi: 10.1093/jxb/erx008

Huber, S. C., and Akazawa, T. (1986). A novel sucrose synthase pathway for sucrose degradation in cultured sycamore cells. *Plant Physiol.* 81, 1008–1013. doi: 10.1104/pp.81.4.1008

Itoh, H., Ueguchi-Tanaka, M., Sakamoto, T., Kayano, T., Tanaka, H., Ashikari, M., et al. (2002). Modification of rice plant height by suppressing the height-controlling gene, D18, in rice. *Breed. Sci.* 52, 215–218. doi: 10.1270/jsbbs.52.215

Iwamoto, M., Kiyota, S., Hanada, A., Yamaguchi, S., and Takano, M. (2011). The multiple contributions of phytochromes to the control of internode elongation in rice. *Plant Physiol.* 157, 1187–1195. doi: 10.1104/pp.111.184861

Iwamoto, M., and Takano, M. (2011). Phytochrome-regulated EBL1 contributes to ACO1 upregulation in rice. *Biotechnol. Lett.* 33, 173–178. doi: 10.1007/s10529-010-0405-7

Jagadish, S. V. K., Murty, M. V. R., and Quick, W. P. (2015). Rice responses to rising temperatures - challenges, perspectives and future directions. *Plant Cell Environ.* 38, 1686–1698. doi: 10.1111/pce.12430

Jagadish, S. V. K., Raveendran, M., Rang, Z. W., Malo, R., Heuer, S., Bennett, J., et al. (2011). Spikelet proteomic response to combined water deficit and heat stress in rice (*Oryza sativa* cv. N22). *Rice* 4, 1–11. doi: 10.1007/s12284-011-9059-x

Jahn, C. E., Mckay, J. K., Mauleon, R., Stephens, J., McNally, K. L., Bush, D. R., et al. (2011). Genetic variation in biomass traits among 20 diverse rice varieties. *Plant Physiol.* 155, 157–168. doi: 10.1104/pp.110.165654

Ji, H., Han, C. D., Lee, G. S., Jung, K. H., Kang, D. Y., Oh, J., et al. (2019). Mutations in the microRNA172 binding site of SUPERNUMERARY BRACT (SNB) suppress internode elongation in rice. *Rice* 12:62. doi: 10.1186/s12284-019-0324-8

Ji, H., Kim, H., Yun, D. W., Yoon, U. H., Kim, T. H., Eun, M. Y., et al. (2014). Characterization and fine mapping of a shortened uppermost internode mutant in rice. *Plant Biotechnol. Rep.* 8, 125–134. doi: 10.1007/s11816-013-0280-5

Kaur, S., Dhugga, K. S., Beech, R., and Singh, J. (2017). Genome-wide analysis of the cellulose synthase-like (Csl) gene family in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:193. doi: 10.1186/s12870-017-1142-z

Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Li, M., Xiong, G., Li, R., Cui, J., Tang, D., Zhang, B., et al. (2009). Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. *Plant J.* 60, 1055–1069. doi: 10.1111/j.1365-313x.2009.04022.x

Li, Y. F., Zheng, Y., Vemireddy, L. R., Panda, S. K., Jose, S., Ranjan, A., et al. (2018). Comparative transcriptome and translatome analysis in contrasting rice genotypes reveals differential mRNA translation in salt-tolerant Pokkali under salt stress. *BMC Genomics* 19(Suppl. 10):935. doi: 10.1186/s12864-018-5279-4

Lin, H., Yu, J., Pearce, S. P., Zhang, D., and Wilson, Z. A. (2017). RiceAntherNet: a gene co-expression network for identifying anther and pollen development genes. *Plant J.* 92, 1076–1091. doi: 10.1111/tpj.13744

Liu, X., Feng, Z. M., Zhou, C. L., Ren, Y. K., Mou, C. L., Wu, T., et al. (2016). Brassinosteroid (BR) biosynthetic gene lhdd10 controls late heading and plant height in rice (*Oryza sativa* L.). *Plant Cell Rep.* 35, 357–368. doi: 10.1007/s00299-015-1889-3

Liu, Z., Zhang, S., Sun, N., Liu, H., Zhao, Y., Liang, Y., et al. (2015). Functional diversity of jasmonates in rice. *Rice* 8:42.

Lo, S.-F., Yang, S.-Y., Chen, K.-T., Hsing, Y.-I., Zeevaart, J. A. D., Chen, L.-J., et al. (2008). A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *Plant Cell* 20, 2603–2618. doi: 10.1105/tpc.108.060913

Luan, W., Liu, Y., Zhang, F., Song, Y., Wang, Z., Peng, Y., et al. (2011). OsCD1 encodes a putative member of the cellulose synthase-like D sub-family and is essential for rice plant architecture and growth. *Plant Biotechnol. J.* 9, 513–524. doi: 10.1111/j.1467-7652.2010.00570.x

Luo, A., Qian, Q., Yin, H., Liu, X., Yin, C., Lan, Y., et al. (2006). EUI1, encoding a putative cytochrome P450 monooxygenase, regulates internode elongation by modulating gibberellin responses in rice. *Plant Cell Physiol.* 47, 181–191. doi: 10.1093/pcp/pci233

Ma, J., Cheng, Z., Chen, J., Shen, J., Zhang, B., Ren, Y., et al. (2016). Phosphatidylserine synthase controls cell elongation especially in the uppermost internode in rice by regulation of exocytosis. *PLoS One* 11:e0153119. doi: 10.1371/journal.pone.0153119

Ma, L., Zhang, D., Miao, Q., Yang, J., Xuan, Y., and Hu, Y. (2017). Essential role of sugar transporter OsSWEET11 during the early stage of rice grain filling. *Plant Cell Physiol.* 58, 863–873. doi: 10.1093/pcp/pcx040

Magome, H., Nomura, T., Hanada, A., Takeda-Kamiya, N., Ohnishi, T., Shinma, Y., et al. (2013). CYP714B1 and CYP714B2 encode gibberellin 13-oxidases that reduce gibberellin activity in rice. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1947–1952. doi: 10.1073/pnas.1215788110

Matsushima, R., Maekawa, M., Kusano, M., Tomita, K., Kondo, H., Nishimura, H., et al. (2016). Amyloplast membrane protein SUBSTANDARD STARCH GRAIN6 controls starch grain size in rice endosperm. *Plant Physiol.* 170, 1445–1459. doi: 10.1104/pp.15.01811

Mishra, M., Wungrampha, S., Kumar, G., Singla-Pareek, S. L., and Pareek, A. (2020). How do rice seedlings of landrace Pokkali survive in saline fields after transplantation? Physiology, biochemistry, and photosynthesis. *Photosynth. Res.* doi: 10.1007/s11120-020-00771-6 [Epub ahead of print].

Monfared, H. H., Chew, J. K., Azizi, P., Xue, G. P., Ee, S. F., Kadkhodaei, S., et al. (2020). Overexpression of a rice monosaccharide transporter gene (OsMST6) confers enhanced tolerance to drought and salinity stress in *Arabidopsis thaliana*. *Plant Mol. Biol. Rep.* 38, 151–164. doi: 10.1007/s11105-019-01186-x

Morey, S. R., Hirose, T., Hashida, Y., Miyao, A., Hirochika, H., Ohsugi, R., et al. (2018). Genetic evidence for the role of a rice vacuolar invertase as a molecular sink strength determinant. *Rice* 11:6.

Muthurajan, R., Shobbar, Z.-S., Jagadish, S. V. K., Bruskiewich, R., Ismail, A., Leung, H., et al. (2011). Physiological and proteomic responses of rice peduncles to drought stress. *Mol. Biotechnol.* 48, 173–182. doi: 10.1007/s12033-010-9358-2

Nakajima, M., Yamaguchi, I., Kizawa, S., Murofushi, N., and Takahashi, N. (1991). Semi-quantification of GAX and GA4 in male-sterile anthers of rice by radioimmunoassay. *Plant Cell Physiol.* 32, 511–513.

Nakamura, A., Fujioka, S., Sunohara, H., Kamiya, N., Hong, Z., Inukai, Y., et al. (2006). The role of OsBRI1 and its homologous genes, OsBRL1 and OsBRL3, in rice. *Plant Physiol.* 140, 580–590. doi: 10.1104/pp.105.072330

Oh, M.-H., Honey, S. H., and Tax, F. E. (2020). The control of cell expansion, cell division, and vascular development by brassinosteroids: a historical perspective. *Int. J. Mol. Sci.* 21:1743. doi: 10.3390/ijms21051743

Oh, S., Park, S., and Han, K.-H. (2003). Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. *J. Exp. Bot.* 54, 2709–2722. doi: 10.1093/jxb/erg304

Patel, R. K., and Jain, M. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. doi: 10.1371/journal.pone.0030619

Peng, S., Khush, G. S., Virk, P., Tang, Q., and Zou, Y. (2008). Progress in ideotype breeding to increase rice yield potential. *Field Crops Res.* 108, 32–38. doi: 10.1016/j.fcr.2008.04.001

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., Salzberg, S. L., et al. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads HHS Public Access Author manuscript. *Nat Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Rathinasabapathi, P., Purushothaman, N., Ramprasad, V. L., and Parani, M. (2015). Whole genome sequencing and analysis of Swarna, a widely cultivated indica rice variety with low glycemic index. *Sci. Rep.* 5:11303.

Richmond, T. A., and Somerville, C. R. (2001). "Integrative approaches to determining Csl function," in *Plant Cell Walls*, eds N. C. Carpita, M. Campbell, and M. Tierney, (Dordrecht: Springer), 131–143. doi: 10.1007/978-94-010-0668-2_8

Rutger, J. N., and Carnahan, H. L. (1981). A fourth genetic element to facilitate hybrid cereal production—A recessive tall in rice 1. *Crop Sci.* 21, 373–376. doi: 10.2135/cropsci1981.0011183x002100030005x

Sato-Izawa, K., Nakaba, S., Tamura, K., Yamagishi, Y., Nakano, Y., Nishikubo, N., et al. (2012). DWARF50 (D50), a rice (*Oryza sativa* L.) gene encoding inositol polyphosphate 5-phosphatase, is required for proper development of intercalary meristem. *Plant Cell Environ.* 35, 2031–2044. doi: 10.1111/j.1365-3040.2012.02534.x

Scofield, G. N., Hirose, T., Aoki, N., and Furbank, R. T. (2007). Involvement of the sucrose transporter, OsSUT1, in the long-distance pathway for assimilate transport in rice. *J. Exp. Bot.* 58, 3155–3169. doi: 10.1093/jxb/erm153

Scofield, G. N., Hirose, T., Gaudron, J. A., Furbank, R. T., Upadhyaya, N. M., and Ohsugi, R. (2002). Antisense suppression of the rice transporter gene, OsSUT1, leads to impaired grain filling and germination but does not affect photosynthesis. *Funct. Plant Biol.* 29, 815–826. doi: 10.1071/pp01204

Scofield, G. N., Ruuska, S. A., Aoki, N., Lewis, D. C., Tabe, L. M., and Jenkins, C. L. D. (2009). Starch storage in the stems of wheat plants: localization and temporal changes. *Ann. Bot.* 103, 859–868. doi: 10.1093/aob/mcp010

Shimizu, M., and Kuno, K. (1967). Some cyto-histological observations on the morphogenetically abnormal rice spikelets caused by a low temperature. *Jpn. J. Crop Sci.* 36, 489–502. doi: 10.1626/jcs.36.4_489

Shingu, Y., Tokai, T., Agawa, Y., Toyota, K., Ahamed, S., Kawagishi-Kobayashi, M., et al. (2012). The double-stranded break-forming activity of plant SPO11s and a novel rice SPO11 revealed by a Drosophila bioassay. *BMC Mol. Biol.* 13:1. doi: 10.1186/1471-2199-13-1

Singh, A., Pandey, A., Baranwal, V., Kapoor, S., and Pandey, G. K. (2012). Comprehensive expression analysis of rice phospholipase D gene family during abiotic stresses and development. *Plant Signal. Behav.* 7, 847–855. doi: 10.4161/psb.20385

Slewinski, T. L. (2012). Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. *J. Exp. Bot.* 63, 4647–4670. doi: 10.1093/jxb/ers124

Sparks, D. L. (ed.) (2012). *Advances in Agronomy*. Cambridge, MA: Academic Press.

Stein, O., and Granot, D. (2019). An overview of sucrose synthases in plants. *Front. Plant Sci.* 10:95. doi: 10.3389/fpls.2019.00095

Sumiyoshi, M., Inamura, T., Nakamura, A., Aohara, T., Ishii, T., Satoh, S., et al. (2015). UDP-Arabinopyranose mutase 3 is required for pollen wall morphogenesis in Rice (*Oryza sativa*). *Plant Cell Physiol.* 56, 232–241. doi: 10.1093/pcp/pcu132

Sun, T.-P. (2008). Gibberellin metabolism, perception and signaling pathways in Arabidopsis. *Arabidopsis Book* 6:e0103. doi: 10.1199/tab.0103

Sunohara, H., Kawai, T., Shimizu-Sato, S., Sato, Y., Sato, K., and Kitano, H. (2009). A dominant mutation of TWISTED DWARF 1 encoding an α-tubulin protein causes severe dwarfism and right helical growth in rice. *Genes Genet. Syst.* 84, 209–218. doi: 10.1266/ggs.84.209

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-313X.2004.02016.x

Tong, H., Xiao, Y., Liu, D., Gao, S., Liu, L., Yin, Y., et al. (2014). Brassinosteroid regulates cell elongation by modulating gibberellin metabolism in rice. *Plant Cell* 26, 4376–4393. doi: 10.1105/tpc.114.132092

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Ueguchi-Tanaka, M., Nakajima, M., Katoh, E., Ohmiya, H., Asano, K., Saji, S., et al. (2007). Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. *Plant Cell* 19, 2140–2155. doi: 10.1105/tpc.106.043729

Vega-Sánchez, M. E., Verhertbruggen, Y., Christensen, U., Chen, X., Sharma, V., Varanasi, P., et al. (2012). Loss of cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. *Plant Physiol.* 159, 56–69. doi: 10.1104/pp.112.195495

Wada, H., Masumoto-Kubo, C., Tsutsumi, K., Nonami, H., Tanaka, F., Okada, H., et al. (2017). Turgor-responsive starch phosphorylation in *Oryza sativa* stems: a primary event of starch degradation associated with grain-filling ability. *PLoS One* 12:e0181272. doi: 10.1371/journal.pone.0181272

Wang, D. R., Wolfrum, E. J., Virk, P., Ismail, A., Greenberg, A. J., and McCouch, S. R. (2016). Robust phenotyping strategies for evaluation of stem non-structural carbohydrates (NSC) in rice. *J. Exp. Bot.* 67, 6125–6138. doi: 10.1093/jxb/erw375

Wang, J., Wang, R., Wang, Y., Zhang, L., Zhang, L., Xu, Y., et al. (2017). Short and Solid Culm/RFL/APO 2 for culm development in rice. *Plant J.* 91, 85–96. doi: 10.1111/tpj.13548

Wang, Y., Xiao, Y., Zhang, Y., Chai, C., Wei, G., Wei, X., et al. (2008). Molecular cloning, functional characterization and expression analysis of a novel monosaccharide transporter gene OsMST6 from rice (*Oryza sativa* L.). *Planta* 228, 525–535. doi: 10.1007/s00425-008-0755-8

Wang, Y., Xu, H., Wei, X., Chai, C., Xiao, Y., Zhang, Y., et al. (2007). Molecular cloning and expression analysis of a monosaccharide transporter gene OsMST4 from rice (*Oryza sativa* L.). *Plant Mol. Biol.* 65, 439–451. doi: 10.1007/s11103-007-9228-x

Wolbang, C. M., Chandler, P. M., Smith, J. J., and Ross, J. J. (2004). Auxin from the developing inflorescence is required for the biosynthesis of active gibberellins in barley stems. *Plant Physiol.* 134, 769–776. doi: 10.1104/pp.103.030460

Wu, C., Cui, K., Wang, W., Li, Q., Fahad, S., Hu, Q., et al. (2016). Heat-induced phytohormone changes are associated with disrupted early reproductive development and reduced yield in rice. *Sci. Rep.* 6:34978.

Xie, Y., Zhang, Y., Han, J., Luo, J., Li, G., Huang, J., et al. (2018). The intronic *cis* element *SE1* recruits trans-acting repressor complexes to repress the expression of *ELONGATED UPPERMOST INTERNODE1* in rice. *Mol. Plant* 11, 720–735. doi: 10.1016/j.molp.2018.03.001

Yadav, S. R., Khanday, I., Majhi, B. B., Veluthambi, K., and Vijayraghavan, U. (2011). Auxin-responsive OsMGH3, a common downstream target of OsMADS1 and OsMADS6, controls rice floret fertility. *Plant Cell Physiol.* 52, 2123–2135. doi: 10.1093/pcp/pcr142

Yamamoto, E., Yonemaru, J. I., Yamamoto, T., and Yano, M. (2012). OGRO: the overview of functionally characterized genes in rice online database. *Rice* 5:26. doi: 10.1186/1939-8433-5-26

Yamamuro, C., Ihara, Y., Wu, X., Noguchi, T., Fujioka, S., Takatsuto, S., et al. (2000). Loss of function of a rice brassinosteroid insensitive1 homolog prevents internode elongation and bending of the lamina joint. *Plant Cell* 12, 1591–1606. doi: 10.1105/tpc.12.9.1591

Yang, C., Li, D., Liu, X., Ji, C., Hao, L., Zhao, X., et al. (2014). OsMYB103L, an R2R3-MYB transcription factor, influences leaf rolling and mechanical strength in rice (*Oryza sativa* L.). *BMC Plant Biol.* 14:158. doi: 10.1186/1471-2229-14-158

Yang, G., Jan, A., and Komatsu, S. (2007). Functional analysis of OsTUB8, an anther-specific β-tubulin in rice. *Plant Sci.* 172, 832–838. doi: 10.1016/j.plantsci.2006.12.019

Yang, J., Zhang, J., Huang, Z., Wang, Z., Zhu, Q., and Liu, L. (2002). Correlation of cytokinin levels in the endosperms and roots with cell number and cell division activity during endosperm development in rice. *Ann. Bot.* 90, 369–377. doi: 10.1093/aob/mcf198

Yang, W., Kong, Z., Omo-Ikerodah, E., Xu, W., Li, Q., and Xue, Y. (2008). Calcineurin B-like interacting protein kinase OsCIPK23 functions in pollination and drought stress responses in rice (*Oryza sativa* L.). *J. Genet. Genomics* 35, S1–S2. doi: 10.1016/S1673-8527(08)60073-9

Yao, W., Li, G., Yu, Y., and Ouyang, Y. (2018). funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *Gigascience* 7:gix119. doi: 10.1093/gigascience/gix119

Ye, Y., Liu, B., Zhao, M., Wu, K., Cheng, W., Chen, X., et al. (2015). CEF1/OsMYB103L is involved in GA-mediated regulation of secondary wall biosynthesis in rice. *Plant Mol. Biol.* 89, 385–401. doi: 10.1007/s11103-015-03760

Yi, X., Du, Z., and Su, Z. (2013).). PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Res.* 41, W98–W103. doi: 10.1093/nar/gkt281

Yin, C., Gan, L., Ng, D., Zhou, X., and Xia, K. (2007). Decreased panicle-derived indole-3-acetic acid reduces gibberellin A 1 level in the uppermost internode, causing panicle enclosure in male sterile rice Zhenshan 97A. *J. Exp. Bot.* 58, 2441–2449. doi: 10.1093/jxb/erm077

Yin, H., Gao, P., Liu, C., Yang, J., Liu, Z., and Luo, D. (2013). SUI-family genes encode phosphatidylserine synthases and regulate stem development in rice. *Planta* 237, 15–27. doi: 10.1007/s00425-012-1736-5

Yin, Y., Wang, Z.-Y., Mora-Garcia, S., Li, J., Yoshida, S., Asami, T., et al. (2002). BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* 109, 181–191. doi: 10.1016/s0092-8674(02)00721-3

Yoshikawa, T., Eiguchi, M., Hibara, K. I., Ito, J. I., and Nagato, Y. (2013). Rice SLENDER LEAF 1 gene encodes cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell proliferation. *J. Exp. Bot.* 64, 2049–2061. doi: 10.1093/jxb/ert060

Zhan, C., Hu, J., Pang, Q., Yang, B., Cheng, Y., Xu, E., et al. (2019). Genome-wide association analysis of panicle exsertion and uppermost internode in rice (*Oryza sativa* L.). *Rice* 12:72. doi: 10.1186/s12284-019-0330-x

Zhang, H., Liang, W., Yang, X., Luo, X., Jiang, N., Ma, H., et al. (2010). Carbon starved anther encodes a MYB domain protein that regulates sugar partitioning required for rice pollen development. *Plant Cell* 22, 672–689. doi: 10.1105/tpc.109.073668

Zhang, H., Tan, G., Yang, L., Yang, J., Zhang, J., and Zhao, B. (2009). Hormones in the grains and roots in relation to post-anthesis development of inferior and superior spikelets in japonica/indica hybrid rice. *Plant Physiol. Biochem.* 47, 195–204. doi: 10.1016/j.plaphy.2008.11.012

Zhang, S., Wang, S., Xu, Y., Yu, C., Shen, C., Qian, Q., et al. (2015). The auxin response factor, OsARF19, controls rice leaf angles through positively regulating OsGH3-5 and OsBRI1. *Plant Cell Environ.* 38, 638–654. doi: 10.1111/pce.12397

Zhang, Y., Li, S., Xue, S., Yang, S., Huang, J., and Wang, L. (2018). Phylogenetic and CRISPR/Cas9 studies in deciphering the evolutionary trajectory and phenotypic impacts of rice ERECTA genes. *Front. Plant Sci.* 9:473. doi: 10.3389/fpls.2018.00473

Zhang, Y., Xiao, W., Luo, L., Pang, J., Rong, W., and He, C. (2012). Downregulation of OsPK1, a cytosolic pyruvate kinase, by T-DNA insertion causes dwarfism and panicle enclosure in rice. *Planta* 235, 25–38. doi: 10.1007/s00425-011-1471-3

Zhang, Y. C., He, R. R., Lian, J. P., Zhou, Y. F., Zhang, F., Li, Q. F., et al. (2020). OsmiR528 regulates rice-pollen intine formation by targeting an uclacyanin to influence flavonoid metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 117, 727–732. doi: 10.1073/pnas.1810968117

Zhu, L., Hu, J., Zhu, K., Fang, Y., Gao, Z., He, Y., et al. (2011). Identification and characterization of SHORTENED UPPERMOST INTERNODE 1, a gene negatively regulating uppermost internode elongation in rice. *Plant Mol. Biol.* 77, 475–487. doi: 10.1007/s11103-011-9825-6

Zou, T., Li, S., Liu, M., Wang, T., Xiao, Q., Chen, D., et al. (2017). An atypical strictosidine synthase, OsSTRL2, plays key roles in anther development and pollen wall formation in rice. *Sci. Rep.* 7:6863.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kandpal, Vishwakarma, Krishnan, Chinnusamy, Pareek, Sharma and Sharma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fgene.2020.589490

Haplotype-Based, Genome-Wide Association Study Reveals Stable Genomic Regions for Grain Yield in CIMMYT Spring Bread Wheat

*Deepmala Sehgal[1], Suchismita Mondal[1], Leonardo Crespo-Herrera[1], Govindan Velu[1], Philomin Juliana[1], Julio Huerta-Espino[2], Sandesh Shrestha[3], Jesse Poland[3], Ravi Singh[1] and Susanne Dreisigacker[1]**

[1] International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [2] Campo Experimental Valle de México, INIFAP, México, Mexico, [3] Kansas State University, Manhattan, KS, United States

OPEN ACCESS

Edited by:
Santosh B. Satbhai,
Indian Institute of Science Education and Research Mohali, India

Reviewed by:
Madhav Bhatta,
Bayer Crop Science, United States
Liu Jin Dong,
Agricultural Genomics Institute at Shenzhen (CAAS), China
Vandana Jaiswal,
Institute of Himalayan Bioresource Technology (CSIR), India

***Correspondence:**
Susanne Dreisigacker
s.dreisigacker@cgiar.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 30 July 2020
Accepted: 21 October 2020
Published: 03 December 2020

Citation:
Sehgal D, Mondal S, Crespo-Herrera L, Velu G, Juliana P, Huerta-Espino J, Shrestha S, Poland J, Singh R and Dreisigacker S (2020) Haplotype-Based, Genome-Wide Association Study Reveals Stable Genomic Regions for Grain Yield in CIMMYT Spring Bread Wheat. Front. Genet. 11:589490. doi: 10.3389/fgene.2020.589490

We untangled key regions of the genetic architecture of grain yield (GY) in CIMMYT spring bread wheat by conducting a haplotype-based, genome-wide association study (GWAS), together with an investigation of epistatic interactions using seven large sets of elite yield trials (EYTs) consisting of a total of 6,461 advanced breeding lines. These lines were phenotyped under irrigated and stress environments in seven growing seasons (2011–2018) and genotyped with genotyping-by-sequencing markers. Genome-wide 519 haplotype blocks were constructed, using a linkage disequilibrium-based approach covering 14,036 Mb in the wheat genome. Haplotype-based GWAS identified 7, 4, 10, and 15 stable (significant in three or more EYTs) associations in irrigated (I), mild drought (MD), severe drought (SD), and heat stress (HS) testing environments, respectively. Considering all EYTs and the four testing environments together, 30 stable associations were deciphered with seven hotspots identified on chromosomes 1A, 1B, 2B, 4A, 5B, 6B, and 7B, where multiple haplotype blocks were associated with GY. Epistatic interactions contributed significantly to the genetic architecture of GY, explaining variation of 3.5–21.1%, 3.7–14.7%, 3.5–20.6%, and 4.4– 23.1% in I, MD, SD, and HS environments, respectively. Our results revealed the intricate genetic architecture of GY, controlled by both main and epistatic effects. The importance of these results for practical applications in the CIMMYT breeding program is discussed.

Keywords: haplotype blocks, haplotype-based GWAS, GBS, EYT, heat map

INTRODUCTION

Bread wheat (*Triticum aestivum* L., 2n = 6x = 42, AABBDD), with global production of 761.5 million tons, is a staple food source for over 2.5 billion people worldwide and an important crop for food security (FAO, 2020). Climate change and population growth will make attainment of food security a challenging task over the coming decades. Development of high-yielding, climate-resilient wheat varieties has therefore become imperative for wheat breeders. Improvement of grain yield (GY) is an arduous task for the global plant-breeding community due to low heritability and intractable "genotype × environment" interactions associated with it, particularly under stress

environments (Quarrie et al., 2006; Sehgal et al., 2017, 2020). Nevertheless, wheat breeders have revealed genetic gains up to 1% for GY annually, but further efforts are required to cope with an estimated 2% yearly increase in world population (Tadesse et al., 2019).

Advances in next-generation sequencing technologies have revolutionized the field of plant genomics. Low-cost genotyping platforms that generate thousands to millions of data points are now available for all agronomically important crops, providing effective means for crop genetic research studies (Ganal et al., 2012). For wheat, where marker number and density were major lacunae in conducting in-depth genetic analyses, the availability of dense sets of single-nucleotide polymorphisms (SNPs) from different genotyping platforms has made a powerful step change in the marker tool kit (Poland et al., 2012; Cavanagh et al., 2013; Wang et al., 2014). The resulting high-density genomic data have opened up new possibilities for untangling the genetic architecture of complex traits by genome-wide association study (GWAS) and to perform other genomic studies, for instance, the analysis of selective sweeps within or across species (Afzal et al., 2019; Liu et al., 2019). Additionally, the recent availability of the high-quality reference genome of bread wheat (IWGSC, 2018) has enhanced our understanding of the regulation of genome organization, gene expression, and evolutionary mechanisms shaping its genome (Alaux et al., 2018; Ramírez-González et al., 2018; Wicker et al., 2018). With genome resolution reaching megabase-scale level in wheat, it is envisioned that genomics-assisted breeding can be escalated to a scale that was not possible previously (Keeble-Gagnère et al., 2018).

Although high-density markers, such as genotyping-by-sequencing (GBS) or SNP arrays, have been used extensively in wheat to explore the genetic architecture of GY and yield components using GWAS (Neumann et al., 2011; Zhang et al., 2013; Edae et al., 2014; Ain et al., 2015; Azadi et al., 2015; Lopes et al., 2015; Sukumaran et al., 2015; Sehgal et al., 2016; Qaseem et al., 2018; Garcia et al., 2019; Li et al., 2019, 2020; Ward et al., 2019; Shokat et al., 2020), panel sizes have been relatively small to dissect such a complex trait, and results therefore were quite variable, identifying hundreds of small-effect QTL. GWAS reports in larger germplasm panels are still rare (Sehgal et al., 2017, 2020; Juliana et al., 2019). Small panel sizes have also hindered scientists from exploring epistatic interactions due to lack of reasonable statistical power (Mackay, 2014).

To boost the power of single-marker GWAS, meta-GWAS has emerged as a leading approach to dissect traits (Evangelou and Ioannidis, 2013). In this approach, summary statistics of multiple trials are analyzed in a single frame to determine the most effective stable loci over space and time while simultaneously reducing false positives. In wheat, this approach has been used successfully to identify important loci associated with quality traits in unbalanced datasets (Battenfield et al., 2018). However, this GWAS approach fails to address the issue of "missing heritability," which is common in single marker–based GWAS. The alternative approach to boost the power of GWAS is by constructing haplotypes between neighboring SNPs on a chromosome. As specific sets of alleles observed on a single chromosome, haplotypes are inherited together with little chance of contemporary recombination. Recent studies on wheat and other crops have shown that GWAS analysis with haplotypes can be superior to single marker–based analysis in terms of statistical significance (better *p*-values) and in estimating allelic effects (Hao et al., 2012; Lu et al., 2012; N'Diaye et al., 2017; Ledesma-Ramírez et al., 2019; Li et al., 2019; Sehgal et al., 2020; Shokat et al., 2020).

In the present study, we targeted exploration of stable regions in the genome that define the backbone of the genetic architecture of GY in CIMMYT spring bread wheat germplasm using a haplotype-based GWAS and investigating the interactions among haplotypes. We used seven large cohorts of advanced breeding lines from different breeding cycles phenotyped under well-managed multiple testing environments (irrigated and stress conditions) and genotyped with GBS markers. The specific objectives were to (i) construct haplotypes using GBS data across 6,461 lines distributed in seven elite yield trials (EYTs); (ii) conduct haplotype-based GWAS in each EYT using phenotyping data derived from the four testing environments; (iii) identify stable haplotypes associated with GY under individual testing environments and across multiple testing environments; and (iv) investigate the contribution of epistatic interactions to the genetic architecture of GY.

MATERIALS AND METHODS

Plant Materials, Phenotyping, and Statistical Analysis

A total of 6,461 spring bread wheat lines, which formed the entries of seven EYTs during 7 consecutive years, were used in this study (**Supplementary Table 1**). EYT2011-12, EYT2012-13, EYT2013-14, EYT2014-15, EYT2015-16, EYT2016-17, and EYT2017-18 consisted of 643, 998, 983, 942, 829, 1,086, and 980 non-overlapping lines, respectively. Each trial year the breeding program selects 1,092 advanced lines for second-year yield testing, which is the source for the lines above. The 1,092 lines in each year were divided into 39 experiments, each with 28 entries and 2 checks in an alpha lattice design with 3 replications. All EYT were phenotyped at the Norman E. Borlaug Experimental Research Station (CENEB) in Ciudad Obregon, Mexico, under multiple contrasting environments by modulating planting dates and irrigation. All trials were sown in bed planting. The plot size was 2.8 m × 1.6 m (2 beds of 0.8 m with 3 rows each).

The multiple environments included optimum irrigated (I) and three stress environments [mild drought stress (MD), severe drought (SD) stress, and heat stress (HS)]. In environment I, five irrigations were applied (at germination and 40, 70, 95, and 115 days after the first irrigation) with a total water supply of maximum 500 mm distributed through five irrigation events across the crop cycle, while in MD environment two irrigations were applied, one at germination and the other after 50 days (using furrow irrigation; the total water supply was 280 mm). In SD environment, drip irrigation was applied at germination and after 50 days with a total water supply of 180 mm available for the plant. In HS environment, planting was delayed by 3 months (end of February) and around 500 mm of water was applied across the crop cycle through five to six irrigation events. Ciudad Obregon

station has little to no rainfall during the crop growing season (November to April). It has a desert-type climate with rains concentrated during the months of August to October (Mondal et al., 2020). However, if there is rain, the irrigation in stressed environments is adjusted to maintain the amounts. Trials were phenotyped for GY, days to heading (DH), and plant height (PH) in each year, as detailed in Sehgal et al. (2020).

The phenotypic data of GY collected for each genotype were adjusted for block effects within each of three replications per trial (incomplete blocks considered as random effects) using the PROC MIXED function in SAS 9. For DH and PH, the adjusted means were calculated by the formula $Y = (Y_{ij} - Y_i) + Y_{\text{all trials}}$, where Y_{ij} is the value of the entry for a trial, Y_i is the mean of checks of that trial, and $Y_{\text{all trials}}$ is the mean of checks of all trials. The summary statistics function in GenStat 14th ed. was used to obtain the minimum and maximum values of each trait in each trial. ANOVA was performed using a customized script in R version 3.4 (**Supplementary Datasheet 1**).

Genomic DNA Extraction and Genotyping

Genomic DNA was extracted from dried leaves collected from five plants per line using a modified CTAB method described in CIMMYT laboratory protocols (Dreisigacker et al., 2016). All lines were genotyped using GBS Kansas State University using 192-plexing on Illumina HiSeq2000. SNP calling was done using TASSEL 5 pipeline as described in Rutkoski et al. (2016). To obtain physical positions of SNPs, sequence reads of the SNPs were blasted to the wheat reference genome RefSeq V.1.1 (IWGSC, 2018).

Population Structure, Linkage Disequilibrium (LD), and Haplotype Blocks

The population structure was assessed through principal component analysis (PCA) using the rgl package in R (Adler and Murdoch, 2013). GAPIT version 2.0 was used to obtain correlation estimates of the frequency of the squared allele of LD (r^2) for all pairwise comparisons. LD decay was visualized by plotting pairwise r^2 values against the physical distance (Mb) for the whole genome, separately for each EYT, and using combined data from the 6,333 lines. A smooth line was fit to the data using second-degree, locally weighted scatterplot smoothing (Breseghello and Sorrells, 2006). For the LOESS estimation of LD decay, genetic distance was estimated as the point where the LOESS curve first crosses the baseline r^2 of 0.1.

To avoid obtaining different haplotype blocks in each of the seven EYTs due to different minor allele frequencies (MAF) of the markers, the GBS data of all seven EYTs were considered together to generate the haplotype blocks. The MAF threshold was set to 0.15 instead of the usual 0.05 so that a 0.05 MAF could be achieved in each EYT. The haplotype blocks were constructed in R, based on the confidence interval algorithm developed by Gabriel et al. (2002) and detailed in Sehgal et al. (2019, 2020). Briefly, D' 95% confidence intervals between SNPs was calculated, and comparisons were divided into categories of "strong LD," "inconclusive," or "strong recombination." If 95% of the comparisons in one block were in "strong LD," a haplotype block was created. The minimum lower and upper confidence interval values were set to 0.6 and 0.95, respectively. The constructed blocks transformed into multiallelic markers, considering the allelic combinations within each block to be independent alleles.

Haplotype-Based GWAS

GWAS was performed in each individual EYT using a mixed linear model (MLM) using Plink version 1.07 (Purcell et al., 2007) with PCA as fixed variate and kinship as random. PCA was conducted using the rgl package in R, and the appropriate number of principal components to be used in MLM was assessed based on Bayesian information criteria (Schwarz, 1978). The kinship matrix was calculated by the VanRaden algorithm (VanRaden, 2008).

A haplotype was considered stable for a testing environment when it showed P value $< 10^{-4}$ in one EYT and at least $P < 10^{-3}$ across three or more EYTs (Sehgal et al., 2020). Similarly, a haplotype was categorized as stable for multiple testing environments when it showed significance of at least $P < 10^{-3}$ in two or more testing environments across two or more EYTs. The allelic effect of the associated haplotypes was estimated as the difference between the mean value of lines with and without favorable allele and was presented as box plots.

Epistatic Interactions

A linear regression model was used to calculate P values and percentage variation as R^2 for two- and three-locus haplotype interactions using an in-house designed script in R (**Supplementary Datasheet 1**). A significant threshold of $P < 0.0001$ was used to declare significant interactions.

RESULTS

Phenotypic Trait Variation in EYT Under Contrasting Environments

Phenotypic traits revealed a wide distribution in all EYTs in all environments (**Supplementary Figures 1a,b**). GY showed significant ($P < 0.001$) and positive correlations with PH in 26 EYTs × environment combinations, while the correlations with DH were positive in irrigated environments and negative in stress environments (MD, SD, and HS) across years.

Mean GY across all trials and environments ranged from 1622 kg/ha (EYT2015-16 in SD) to 8622 kg/ha (EYT2011-12 in B-5IR) (**Supplementary Table 2**). In general, SD and HS were the least yielding environments, except in EYT2011-12 and EYT2015-16 (**Figure 1**). ANOVA showed highly significant effects ($P < 0.001$) of genotypes, environments, and genotype × environment interactions for GY in the seven EYTs (**Supplementary Table 2**). Broad sense heritability estimates ranged from 0.31 (EYT2015-16) to 0.63 (EYT2011-12) (**Supplementary Table 2**).

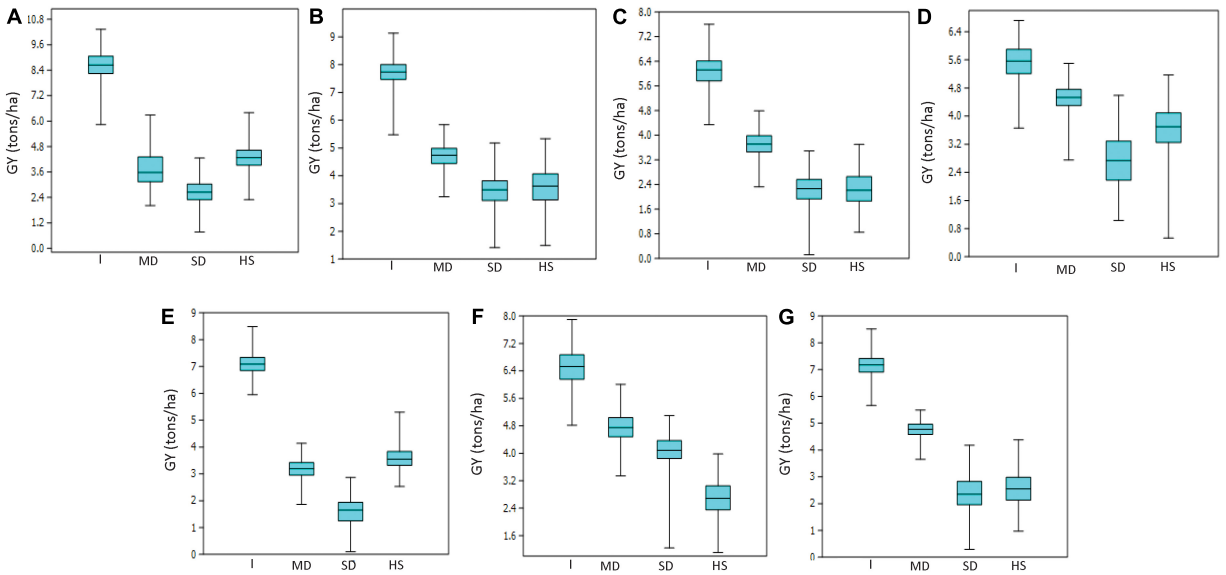


FIGURE 1 | Grain yield variation under different environments in EYT2011-12 **(A)**, EYT2012-13 **(B)**, EYT2013-14 **(C)**, EYT2014-15 **(D)**, EYT2015-16 **(E)**, EYT2016-17 **(F),** and EYT2017-18 **(G)**.

Haplotype Blocks; Genome-Wide Coverage and Distribution

An initial set of 50,058 SNP markers was obtained on 6,461 lines. Of these, a filtered set of 14,027 SNP with maximum 30% missing data and a minor allele frequency (MAF) $\geq$ 0.15 was extracted without imputation. Lines showing more than 60% missing data were also culled with 6,333 genotypes remaining for further analysis.

A total of 519 haplotype blocks were established across the genomes. The haplotype blocks covered a total genome length of 14,036 Mb with 4,925, 5,170, and 3,941 Mb covered in the A, B, and D genomes, respectively (**Table 1**). The blocks were distributed according to the length of each chromosome, and the density of the markers with the highest numbers were obtained in A and B genomes (231 and 239, respectively) and the lowest in the D genome (49). The highest number was obtained on chromosomes 7A (51), followed by chromosomes 2B and 7B (46 each), whereas the lowest number of haplotype blocks was obtained on chromosome 4D (1).

Population Structure, Linkage Disequilibrium Decay, and Significant Associations

All seven EYTs showed a moderate structure with two to three subgroups deciphered by PCA (**Supplementary Figure 2**). Whole genome linkage disequilibrium (LD) decay in the individual EYT and combined EYTs is shown in **Supplementary Figure 3**, which revealed that LD decay varied from 1.8 Mb in EYT2014-15 to 2.3 Mb in EYT2016-17, with an average LD decay of approximately 2 Mb.

For environment I, seven stable associations were identified across EYTs (**Supplementary Table 3**) on chromosomes 2A (1), 3A (1), 4A (1), 4B (1), 5B (1), and 6B (2). Of these, favorable alleles of two associated blocks on chromosomes 3A and 4B showed GY advantage of >100 kg/ha across EYTs. For haplotype block HB3A.1, the favorable haplotype ACGA resulted in GY increase of 215 to 525 kg/ha in three EYTs (**Figure 2**). Similarly, the favorable haplotype TC in haplotype block HB4B.12 resulted in an increase of 168 to 429 kg/ha in GY across EYTs. The HB5B.29 linked to flowering time gene *Vrn-B1* had two favorable alleles (AC and GT) and showed allelic effects of 47 to 568 kg/ha across EYTs (**Supplementary Table 3**).

In environment MD, four haplotype blocks on chromosomes 1A (1), 1B (1), 2A (1), and 3B (1) showed association with GY. Of these, HB1A.13 showed GY advantage of >100 kg/ha across, while HB1B.19 showed effects up to 553 kg/ha in EYT2011-12 (**Supplementary Table 3**). In the SD environment, 10 blocks showed association with GY. These were identified on chromosomes 1A (1), 1B (1), 4A (2), 4D (1), 5B (2), 6A (1), and 7B (2). Of these, three associations on chromosomes 1B, 4D, and 6A showed the largest allele effects. For HB1B.3, two favorable alleles (GG and AG) resulted in a 126–359 kg/ha increase in GY across EYTs, while favorable alleles at HB4D.1 (TCTG) and HB6A.6 (GA) resulted in an increase of 151–362 kg/ha and 203–248 kg/ha, respectively (**Figure 2**). The block HB5B.29 linked to the major vernalization gene *Vrn-B1* showed association with GY across five EYTs, with the favorable allele resulting in an increase of 263–430 kg/ha (**Supplementary Table 3**). Since both I and SD environments showed significant association with this block, both phenological traits (DH and PH) were used as covariates in GWAS to test the significance of this locus. However, when DH and PH were used as covariates, HB5B.29 locus either disappeared or became less significant.

For HS, 15 haplotype blocks were associated on chromosomes 2B (1), 3A (1), 3B (2), 4A (1), 5B (1), 6A (1), 6B (2), 7A (5),

TABLE 1 | Summary of haplotype blocks (HB) in 6,333 lines of seven elite yield trials.

Chromosome	No. of SNPs	Physical position (bp)- First and Last SNP across chromosome*	Number of HB	Total number of alleles	Number of SNPs in an HB (min–max)
1A	585	114,5442–593,790,090	32	85	2–5
1B	804	143,0915–688,983,196	23	60	2–5
1D	444	78,777–492,641,722	8	21	2–5
2A	784	138,1013–779,674,593	29	82	2–6
2B	1,042	19,097–800,780,364	46	128	2–8
2D	550	152,5890–651,419,179	13	34	2–6
3A	731	711,587–750,500,626	31	86	2–7
3B	903	198,860–829,535,127	33	92	2–6
3D	576	344,069–615,458,508	6	16	2–6
4A	639	881,423–744,304,645	29	79	2–5
4B	423	622,455–672,953,537	12	30	2–6
4D	171	739,257–509,849,696	1	3	4
5A	669	217,5212–709,755,448	29	72	2–6
5B	793	16,637–712,600,907	41	103	2–6
5D	337	93,2817–565,975,773	5	13	2–4
6A	609	684,328–617,838,620	30	81	2–5
6B	880	195,536–720,519,123	38	100	2–6
6D	379	687,775–473,287,249	10	25	2–6
7A	1,055	289,461–736,572,283	51	143	2–6
7B	899	259,901–747,616,899	46	131	2–6
7D	754	114,2643–638,541,382	6	16	2–5

**Physical positions of the first and the last SNP on individual chromosomes.*

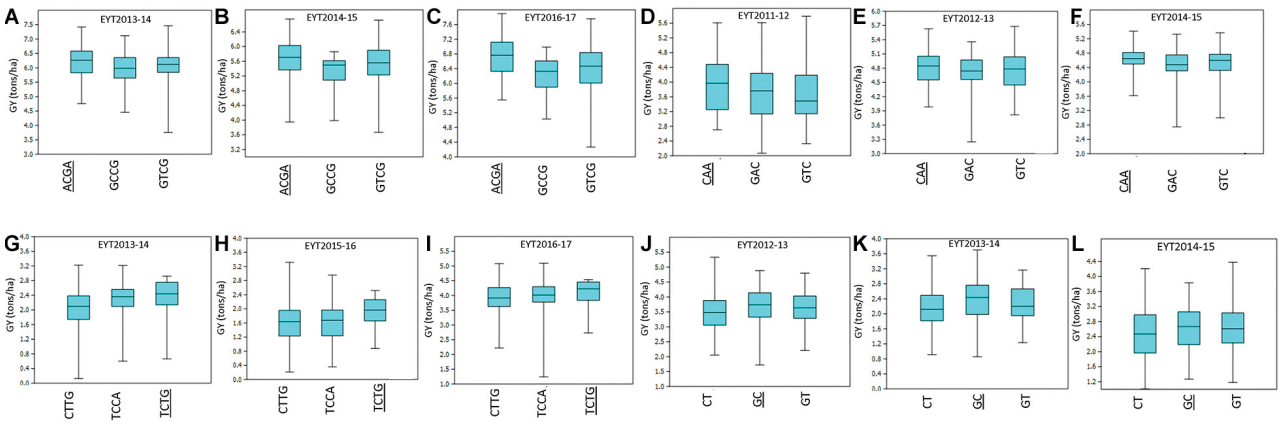


FIGURE 2 | Allelic effects of haplotype blocks associated with GY under I **(A–C)**, MD **(D–F)**, SD **(G–I)**, and HS **(J–L)** environments. The favorable haplotype in each part of the figure is shown as underscored. The parts **(A–L)** show allelic effects of HB3A.1, HB2A.29, HB 4D.1, and HB7A.2, respectively.

and 7B (1). The associations on chromosomes 2B (HB2B.12) and 4A (HB4A.24) and all associations on chromosome 7A (HB7A.2, HB7A.3, HB7A.20, HB7A.28, and HB7A.32) showed large allelic effects compared to other blocks (**Supplementary Table 3**).

Thirty stable haplotypes that are favorable in multiple environments and across EYTs were identified (**Table 2** and **Supplementary Table 4**), including seven hotspots on chromosomes 1A, 1B, 2B, 4A, 5B, 6B, and 7B, where multiple haplotype blocks on same chromosome were associated with GY. The associations on chromosomes 2B (HB2B.10), 3B (HB3B.2), 4B (HB4B.12), 5D (HB5D.5), and 7B (HB7B.18) resulted in a GY increase of 177 to 357, 148 to 449, 168 to 429, 116 to 496, and 122 to 470 kg/ha in different environments and EYT, respectively (**Supplementary Table 4**). **Figure 3** shows all stable haplotypes on chromosome maps, and **Figure 4** shows the frequencies of the favorable haplotypes of each of the 30 associated blocks in all seven EYTs. The frequencies of 23 haplotypes varied from 11 to up to 78% across EYTs, while the frequency of seven haplotypes (HB1B.20, HB2B.10, HB3B.2, HB4A.20, HB4A.27, HB5A.15, and HB6B.38) remained low across EYTs (**Supplementary Table 5**). Potential candidate genes underlying 28 out of 30 haplotype blocks were identified and are listed in **Supplementary Table 6**.

TABLE 2 | Thirty stable associations identified for grain yield considering all elite yield trials and testing environments (I: Irrigated; MD: Moderate drought; SD: Severe Drought; HS: Heat Stress) together.

Haplotype block	Markers in blocks	Chromosome	Interval; First-Last SNP (bp)	Elite yield trials; Testing environment
HB1A.12	S1A_494392059, S1A_494393037	1A	978	EYT2011-12; DS, EYT2012-13; HS, EYT2013-14; MD, EYT2013-14; SD, EYT2016-17; MD
HB1A.13	S1A_497201550, S1A_497201682	1A	132	EYT2012-13; MD, EYT2013-14; MD, EYT2013-14; SD, EYT2016-17; MD, EYT2016-17; HS
HB1A.14	S1A_499864157, S1A_499864420, S1A_499864432, S1A_500074551	1A	210394	EYT2011-12; MD, EYT2012-13; HS, EYT2015-16; MD, EYT2016-17; MD, EYT2016-17; HS
HB1B.3	S1B_18569448, S1B_18570787	1B	1339	EYT2011-12; I, EYT2013-14; SD, EYT2014-15; MD, EYT2014-15; HS, EYT2015-16; SD, EYT2017-18; SD
HB1B.19	S1B_639415604, S1B_639415692, S1B_639426265	1B	10661	EYT2011-12; MD, EYT2012-13; HS, EYT2013-14; MD, EYT2014-15; I, EYT2014-15; MD, EYT2014-15; HS, EYT2015-16; MD
HB1B.20	S1B_642616640, S1B_642616658	1B	18	EYT2014-15; SD, EYT2016-17; MD, EYT2016-17; HS, EYT2017-18; SD
HB2B.10	S2B_24883552, S2B_24887574, S2B_24899507, S2B_24899536, S2B_24899572	2B	16020	EYT2012-13; HS, EYT2016-17; I, EYT2017-18; MD, EYT2017-18; HS
HB2B.15	S2B_75832544, S2B_75848057	2B	15513	EYT2012-13; DS, EYT2012-13; HS, EYT2015-16; DS, EYT2016-17; HS
HB2B.42	S2B_784544719, S2B_784774250, S2B_784905791, S2B_784905811	2B	361092	EYT2011-12; MD, EYT2011-12; HS, EYT2012-13; I, EYT2014-15; SD, EYT2015-16; SD
HB3B.2	S3B_7240747, S3B_7240753	3B	6	EYT2011-12; MD, EYT2011-12; HS, EYT2012-13; SD, EYT2017-18; HS
HB3B.23	S3B_758391015, S3B_758438223, S3B_758464620, S3B_758612003, S3B_758612056, S3B_758613386	3B	222371	EYT2011-12; MD, EYT2011-12; HS, EYT2013-14; MD, EYT2016-17; I, EYT2016-17; SD, EYT2017-18; HS
HB4A.20	S4A_713064971, S4A_713506269, S4A_713517340, S4A_713522176	4A	457205	EYT2011-12; SD, EYT2013-14; I, EYT2013-14; SD, EYT2014-15; HS, EYT2017-18; SD
HB4A.25	S4A_721406670, S4A_721406696, S4A_721826603, S4A_721826636	4A	419966	EYT2012-13; MD, EYT2013-14; SD, EYT2015-16; I, EYT2015-16; MD
HB4A.27	S4A_730188545, S4A_730188899	4A	354	EYT2011-12; SD, EYT2012-13; SD, EYT2012-13; HS, EYT2014-15; I, EYT2014-15; MD
HB4B.8	S4B_644330895, S4B_644330917	4B	22	EYT2011-12; I, EYT2012-13; SD, EYT2012-13; HS, EYT2013-14; I, EYT2013-14; MD, EYT2013-14; SD, EYT2016-17; HS
HB4B.12	S4B_663621978, S4B_663622013	4B	35	EYT2013-14; I, EYT2014-15; I, EYT2014-15; SD, EYT2016-17; I
HB5A.15	S5A_548234618, S5A_548234636, S5A_548387200, S5A_548422588	5A	187970	EYT2012-13; HS, EYT2014-15; MD, EYT2016-17; I, EYT2017-18; I
HB5B.3	S5B_24292046, S5B_24537970, S5B_24648800, S5B_24677091	5B	385045	EYT2011-12; HS, EYT2012-13; I, EYT2013-14; HS, EYT2017-18; MD
HB5B.6	S5B_47584429, S5B_47592949	5B	8520	EYT2011-12; HS, EYT2012-13; SD, EYT2012-13; HS, EYT2013-14; SD, EYT2014-15; SD
HB5B.21	S5B_513712393, S5B_513713184	5B	791	EYT2011-12; SD, EYT2012-13; MD, EYT2014-15; I, EYT2015-16; HS
HB5D.5	S5D_550192169, S5D_550192174	5B	5	EYT2012-13; HS, EYT2014-15; I, EYT2016-17; I, EYT2016-17; MD

(Continued)

TABLE 2 | Continued

Haplotype block	Markers in blocks	Chromosome	Interval; First-Last SNP (bp)	Elite yield trials; Testing environment
HB6B.6	S6B_17686703, S6B_17701765	6B	15062	EYT2011-12; I, EYT2014-15; I, EYT2016-17; I, EYT2017-18; I
HB6B.20	S6B_459374225, S6B_459374299	6B	74	EYT2011-12; HS, EYT2012-13; SD, EYT2012-13; HS, EYT2017-18; HS
HB6B.38	S6B_708712113, S6B_708712131	6B	18	EYT2011-12; SD, EYT2011-12; HS, EYT2012-13; HS, EYT2014-15; HS
HB7A.2	S7A_7938818, S7A_7938819	7A	1	EYT2012-13; HS, EYT2013-14; HS, EYT2014-15; HS, EYT2017-18; SD, EYT2017-18; HS
HB7A.3	S7A_12011058, S7A_12011069	7A	11	EYT2011-12; SD, EYT2011-12; HS, EYT2012-13; HS, EYT2013-14; HS, EYT2015-16; HS, EYT2017-18; SD, EYT2017-18; HS
HB7B.11	S7B_124548883, S7B_124549059	7B	176	EYT2011-12; SD, EYT2011-12; HS, EYT2012-13; HS, EYT2013-14; SD, EYT2014-15; HS, EYT2017-18; HS
HB7B.18	S7B_576927863, S7B_576927877	7B	14	EYT2012-13; HS, EYT2014-15; SD, EYT2015-16; HS, EYT2016-17; MD
HB7B.21	S7B_605313385, S7B_605313397	7B	12	EYT2011-12; SD, EYT2012-13; SD, EYT2013-14; HS, EYT2016-17; HS, EYT2017-18; SD
HB7B.45	S7B_733461150, S7B_733461162	7B	12	EYT2011-12; SD, EYT2012-13; HS, EYT2013-14; I, EYT2014-15; MD, EYT2016-17; I

We constructed heat maps for all seven EYTs to visualize the series of favorable haplotypes accumulated in individual genotypes. **Figures 5**, **6** show heat maps of selected lines in EYT2015-16 displaying having none to maximum number of favorable haplotypes under HS environment and across all environments and EYTs, respectively (15 and 30 haplotype blocks). Heat maps shown here revealed that the maximum number of favorable haplotypes accumulated in lines from EYT2015-16 were 11 and 23 from the total 15 and 30 haplotypes identified under HS environment, and across environments and EYTs, respectively. We further estimated the additive effects of the favorable haplotypes on GY for (a) the environment-specific haplotypes and (b) all 30 stable multi-environmental haplotypes. The trend showed that with an increasing number of haplotypes, GY increases in all EYTs in all environments. **Figure 7** shows the additive effects with an increasing number of haplotypes on GY in the two stress environments, SD and HS (**Figures 7A,B**) and across all environments (**Figure 7C**). The increase in GY ranged from 2.5 to 14.1% and 4.3 to 17.7% across EYTs in SD and HS environments, respectively (**Figures 7A,B**). When stable associations from all environments were tested, GY increase was on average 8% (**Figure 7C**).

Epistatic Interactions

Except in MD environment, epistatic interactions were observed in all environments among associated loci (**Supplementary Figures 4–6**). Most importantly, in both I and SD environments, *Vrn-B1*-linked locus HB5B.29 contributed significantly to epistatic interactions. In environment I, HB5B.29 interacted with HB4B.12 and HB6B.6 more frequently than others, while in environment SD, interactions between HB5B.29 and HB6A.6 were frequent. In environment HS, four associated haplotype blocks from chromosome 7A (HB7A.2, HB7A.3, HB7A.28, and HB7A.32) were mainly involved in interactions among themselves and with other loci. The percent variation explained ranged from 1.5 to 7.5%, 3.6 to 12.9%, and 3.4 to 10.9% in I, SD, and HS environments, respectively.

Genome-wide epistatic interactions were observed in all four environments (**Supplementary Figures 7–10**). In environments SD and HS, interactions were observed in all EYTs. The percent variation explained ranged from 3.5 to 21.1%, 3.7 to 14.7%, 3.5 to 20.6%, and 4.4 to 23.1% in I, MD, SD, and HS environments, respectively.

DISCUSSION

Although much research exploring the genetic architecture of yield and yield-associated traits has been reported in wheat using GWAS, the identification of more stable key determinants of GY remain relatively unexplored, largely due to the complexity of the trait and small panel sizes used in previous studies leading to the so-called "large p small n" or "short-fat data" problem (Diao and Vidyashankar, 2013). Additionally, use of bi-allelic SNPs accentuated "missing heritability" issues and therefore reported markers had limited impact in breeding. In the present study, we performed haplotype-based GWAS using 519 haplotype blocks

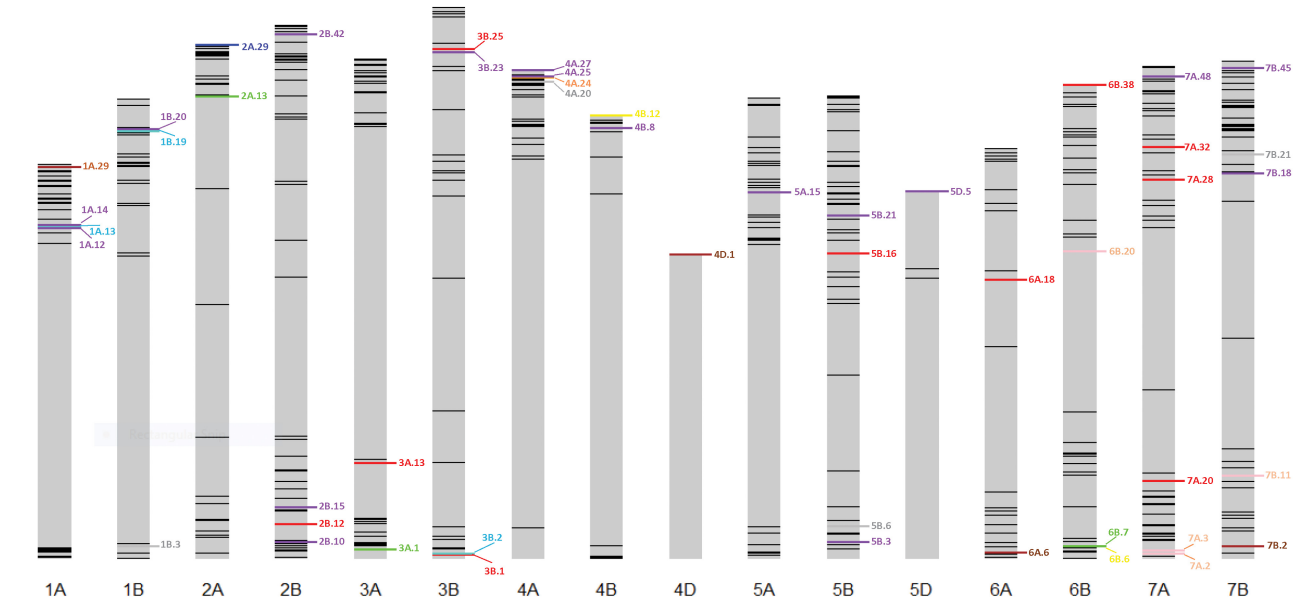


FIGURE 3 | Environment-specific stable haplotypes and haplotypes identified to be significant across multiple environments and EYTs. Green, blue, brown, and red colors show environment-specific associations with GY under I, MD, SD, and HS environments, respectively. Purple color shows stable (S) associations significant across multiple environments and EYTS, while turquoise, pink, yellow, orange, and gray show associations that were identified under two categories; turquoise (MD, S), pink (HS, S), yellow (I, S), orange (SD, HS), and gray (SD, S).

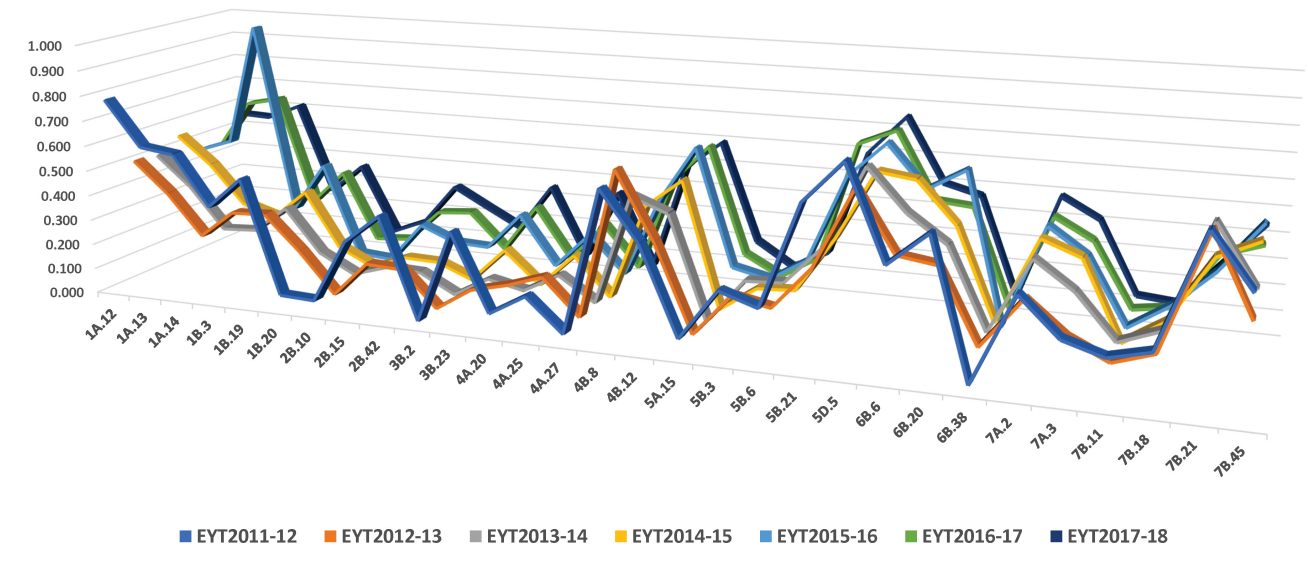


FIGURE 4 | Frequency of stable haplotypes over 7 years (EYT2011-12 to EYT2017-18).

on seven large cohorts of advanced CIMMYT spring bread wheat lines consisting of 6,333 genotypes overall. In addition, epistatic interactions among the genome-wide haplotypes were investigated, an important aspect that has not yet been fully explored in wheat GWAS in order to address the missing heritability (Zuk et al., 2014; Sehgal and Dreisigacker, 2019).

Three approaches are generally used to construct haplotype blocks: (1) user-defined length, (2) sliding-window, and (3) LD. The user-defined fixed length of haplotype blocks (2–15 bp) is the easiest approach; however, generated haplotypes do not reflect genetic principles such as recombination or LD (Gabriel et al., 2002) or a shared evolutionary history (Templeton et al., 2005). The sliding-window approach is the most widely used for building haplotypes in GWAS (Braz et al., 2019). This approach is easy to use and handle; however, when adjacent SNPs are in strong LD, it provides redundant information, making it no more informative than SNPs. Similarly, when LD patterns vary over large genomic regions, it is difficult to determine the

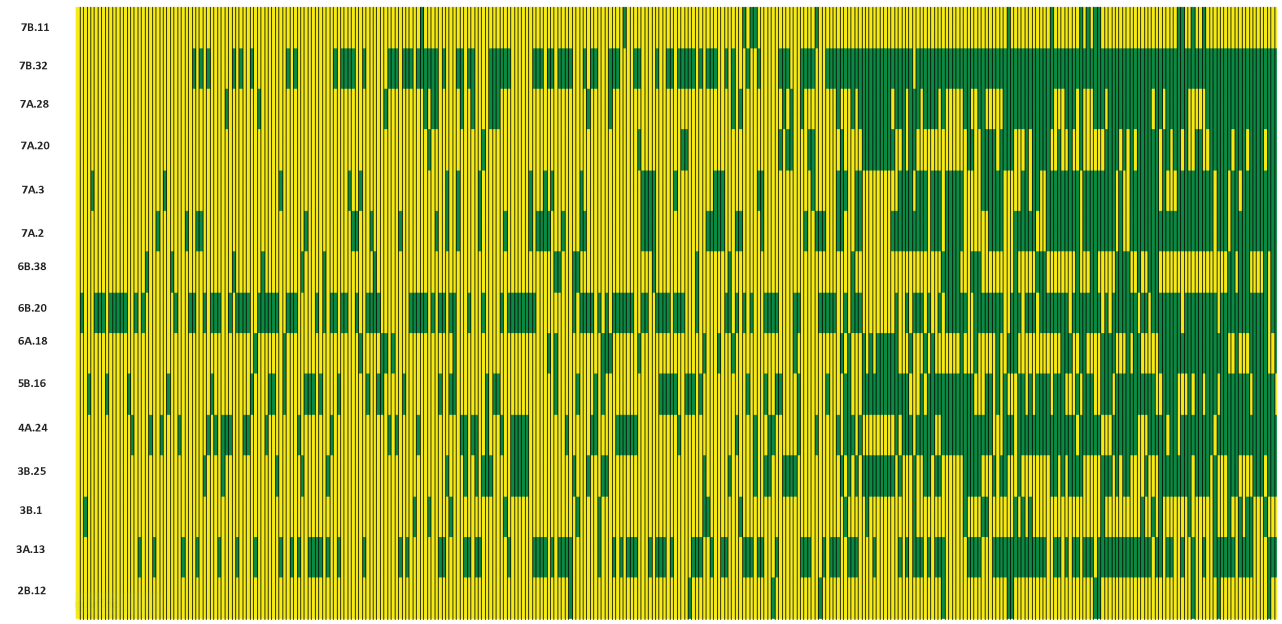


FIGURE 5 | Heat map showing contrasting lines from EYT2015-16 showing all 15 haplotype blocks identified in HS environment. Each yellow vertical line represents a genotype, and each green vertical green rectangle represents a favorable haplotype of an associated block from a chromosome. The name of the haplotype block is shown on the left.

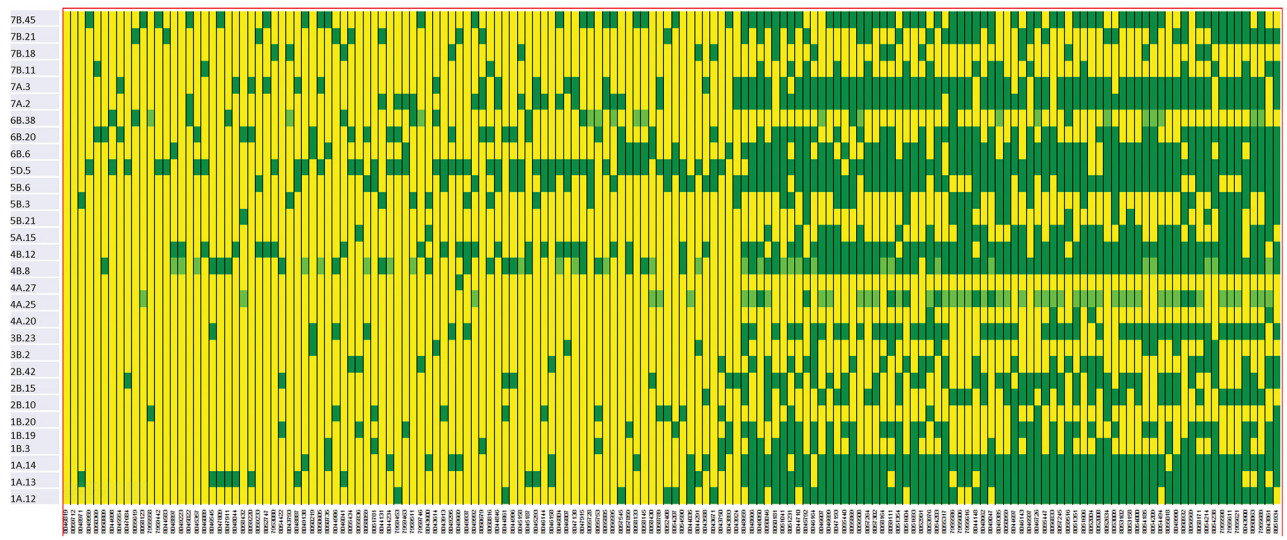


FIGURE 6 | Heat map showing contrasting lines from EYT2015-16 showing favorable alleles of all 30 haplotype blocks identified to be stable across environments and EYTs. Each yellow vertical line represents a genotype. Each dark green rectangle represents the first favorable haplotype of an associated block from a chromosome while light green color represents the second favorable haplotypes identified in a few haplotype blocks. The name of the haplotype block is shown on the left.

appropriate window size for a genome-wide scan. The LD-based approach is the most advantageous because it focuses directly on the detection of historical recombination in the test population (Qian et al., 2017).

We constructed haplotypes using an LD-based approach and conducted a haplotype-based GWAS and epistatic scan to dissect the genetic architecture of GY under contrasting sets of environments and across seven EYTs. The total number of genome-wide haplotype blocks obtained was in a similar range as reported in the recent studies using same marker platform (Singh et al., 2018; Ledesma-Ramírez et al., 2019; Shokat et al., 2020). Li et al. (2019) used a much higher density of markers from two platforms (wheat 90K and 660K Illumina SNP arrays) and thus were able to obtain much higher numbers of haplotype blocks

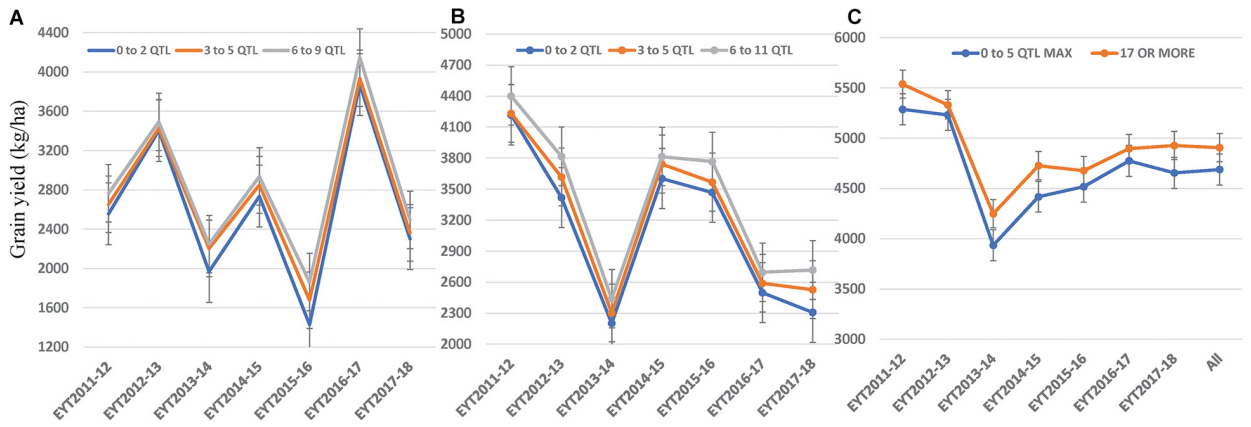


FIGURE 7 | Average grain yield per trial observed by simulating numbers of favorable haplotypes identified in SD **(A)** and HS **(B)** environments and all environments **(C)**.

per chromosome and across the genome. However, panel size remained small (166 lines) in their study. The average LD decay in the seven EYTs in the present study was observed at ∼2 Mb. Comparison of LD decay with previous studies in wheat in which physical distance was used for estimating LD decay (Liu et al., 2017; Ladejobi et al., 2019; Li et al., 2019) revealed a faster decay in the CIMMYT germplasm (2 Mb in CIMMYT germplasm vs. 4–8 Mb in the above-mentioned studies). This suggests high levels of genetic diversity in the current CIMMYT breeding germplasm, which consists of lines selected from a wide range of genetic backgrounds. The higher diversity of CIMMYT germplasm vis-à-vis other wheat germplasm sets has also been observed in previous studies (Warburton et al., 2006; Dreisigacker et al., 2008; Sehgal et al., 2015).

We compared the stable haplotypes identified in the our study with GWAS peaks for GY and yield-related traits identified in various other panels using the GrainGenes genome browser[1]. Additionally, we investigated overlaps of the stable haplotypes against the meta-QTL (MQTL) reported by Acuña-Galindo et al. (2015) associated with adaptation to drought and heat stress (**Supplementary Table 3**). Furthermore, we compared our results with those reported by Li et al. (2019), who located 12 stable QTL for GY on the wheat reference genome using haplotype-based GWAS (**Supplementary Table 4**). Of the 7, 4, 10, and 15 environment-specific associations identified in the I, MD, SD, and HS environments, respectively, the four associations identified in the MD environment corresponded to MQTL 2, 6, 13, and 27. One (HB5.6) and three (HB5B.16, HB7A.20, HB7A.32) haplotype blocks identified in the SD and HS environments, respectively, corresponded to MQTL 44, 58, and 59 of Acuña-Galindo et al. (2015). Further, two (HB3A.1 and HB6B.7), three (HB4A.20, HB5B.6, and HB6A.6), and two (HB3B.25 and HB7A.3) haplotype blocks identified in the I, SD, and HS environments, respectively, overlapped with known GY QTL in GrainGenes (**Supplementary Table 3**). Juliana et al. (2019) used single marker–based GWAS on a smaller subset (3,485 lines) of the same EYT investigated here. The authors reported QTL within 0.2–2.2 Mb of the stable haplotype blocks reported on chromosome 3B (HB3B.25) in the HS environment and on chromosome 4A (HB4A.23 and HB4A.24) in the I, SD, and HS environments. Other QTL reported by Juliana et al. (2019) were on the same chromosomes as the present study; however, these were 58–510 Mb apart. For instance, haplotype blocks identified on chromosomes 5B (HB 5B.21) and 6B (HB 6B.20) were 58 and 146 Mb apart, whereas the four haplotype blocks identified on chromosome 7B (HB7B.11, HB7B.18, HB7B.21, and HB7B.45) were 98, 353, 382, and 510 Mb apart, respectively. Most significantly in our study, a haplotype block hotspot region was identified on chromosome 7A for heat tolerance, which was not detected in previous studies.

Of the 30 stable haplotype blocks identified in multiple environments and across EYTs, six corresponded with GWAS peaks identified in GrainGenes, while eight blocks corresponded to five MQTL (MQTL2 covered by HB1A.12, HB1A.13, and HB1A.14; MQTL6 covered by HB1B.19 and 1B.20; MQTL27 covered by HB3B.2; MQTL44 covered by HB5B.21; and MQTL51 covered by HB6B.6) of Acuña-Galindo et al. (2015). When comparisons were made with the 12 stable QTL reported by Li et al. (2019) for GY and yield components, only two were found in close vicinity from 5 to 20 Mb (**Supplementary Table 4**).

The frequencies of the favorable haplotypes of the 30 stable multi-environmental haplotypes blocks revealed that eight favorable haplotypes in blocks HB1A.12, HB1B.3, HB1B.19, HB2B.42, HB4B.8, HB5B.3, HB6B.20, and HB7B.18 decreased slightly by 10–18% over the 7 years, and only one favorable haplotype in the block HB7B.21 showed a sharp decrease of 36% in the seventh year (EYT2017-18). Eleven favorable haplotypes were maintained in moderate (30–50%) frequencies. Intriguingly, favorable haplotype in the block HB5D.5, with an allelic effect of +116–496 kg/ha across environments, was maintained at the highest frequency (up to 77%) in all seven EYTs, whereas the frequencies of the seven favorable haplotypes in blocks HB1B.20, HB2B.10, HB3B.2, HB4A.20, HB4A.27, HB5A.15, and HB6B.38 remained consistently low (2–15%) across EYTs.

[1]https://wheat.pw.usda.gov/GG3/genome_browser

These low-frequency haplotypes were significantly associated with GY in three or all four environments and showed moderate to high allelic effects varying from +85–233 kg/ha to +148–449 kg/ha across EYTs and hence are important targets for future validation.

Despite the awareness that epistasis contributes significantly to the genetic architecture of most quantitative traits, epistatic interactions are usually not explored in GWAS studies (Sehgal et al., 2017, 2020; Assefa et al., 2019). The most important reason is that it is time consuming and computationally exhaustive to estimate genome-wide interactions in large datasets. Further, unlike in bi-parental populations, ready-to-use models are not available to estimate marker interaction effects along with main additive effects in GWAS panels (Rio et al., 2020). Additionally, the lack of sufficiently large experimental datasets has been a limiting factor to obtain reasonable statistical power when screening the genome for multi-locus epistasis. The size of our GWAS panel (6,333 lines) in the present study, along with the comprehensive phenotypic datasets generated in multiple environments (irrigated and stress environments), in combination with the fact that a large single SNP dataset was reduced to a set with fewer haplotype blocks, made the study of multi-locus epistatic interactions feasible with reasonable statistical power. We observed significant interactions among stable haplotypes. Most importantly, the haplotype block HB5B.29 linked to the vernalization locus *Vrn-B1* seemed to contribute significantly to interactions in both irrigated and drought-stressed environments, explaining up to 12.9% additional variation (**Supplementary Figures 4, 5**). This reinforces that major flowering genes can contribute to yield advantage in both irrigated and drought-stressed environments by both additive and epistatic effects (Cockram et al., 2007; Sehgal et al., 2017; He et al., 2019).

Likewise, significant epistatic interactions were obtained among genome-wide haplotypes for GY, explaining a higher percentage of variation in severely stressed environments (SD and HS) compared to the I environment in all EYTs (**Supplementary Figures 9, 10**). Our results are in contrast to Reif et al. (2011), who reported that main effects dominated the genetic architecture of GY and epistatic interactions contributed only little. We attribute these discrepancies to a narrower panel of elite breeding lines (455 lines, derivatives from a few parents) used in Reif et al. (2011) that probably did not retain enough power to reveal epistasis among loci. Further, Reif et al. (2011) studied GY only in irrigated environments whereas in the current study multiple environments were analyzed.

To be able to utilize stable QTL in a breeding program, we constructed heat maps for all environments in all EYTs. This approach led us to recognize different sets of lines with contrasting haplotype composition. CIMMYT and other breeding programs start to routinely genotype all lines that enter yield trials. Therefore, lines with higher numbers of favorable haplotypes and complementary haplotypes can be identified and re-incorporated as parents in breeding programs to maintain and further accumulate the favorable haplotypes in subsequent breeding cycles. The results can also be exploited in multiple-trait integration or line-conversion pipelines by using the lines carrying high numbers of favorable haplotypes as elite parents in crosses with donor parents selected for additional target traits (e.g., disease resistance) to be able to reveal a comprehensive performance package. The latest studies have shown that integrating haplotypes and epistatic interactions as fixed effects in genome-wide prediction models can improve prediction abilities for GY by about 10% (Spindel et al., 2016; Sehgal et al., 2020). This approach that attempts to boost prediction abilities with the contribution of GWAS peaks has yet to be further tested. Further, Mérida-García et al. (2020) reported candidate genes underpinning metaQTL reported by Acuña-Galindo et al. (2015) on chromosomes 3B and 4A. The candidate genes reported here (**Supplementary Table 6**) with proven roles in abiotic stress tolerance in model crops or having expression evidences in wheat under various stress conditions expand opportunities for future validation studies (Mérida-García et al., 2020).

DATA AVAILABILITY STATEMENT

The data related to manuscript has been provided in **Supplementary Material**. The genotypic and phenotypic data are available at link: https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10548504.

AUTHOR CONTRIBUTIONS

DS and SD conceptualized the manuscript. SD designed the research. DS analyzed the data and wrote the manuscript. SM, LC-H, GV, JH-E, and RS generated phenotypic data. PJ, SS, and JP provided allele called GBS data. All authors reviewed the manuscript.

FUNDING

The work was supported by funding from CRP WHEAT, the Durable Rust Resistance in Wheat project (OPPGD1389) led by Cornell University and funded by the Bill & Melinda Gates Foundation (BMGF), the UK Department for International Development (DFID), and the US Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.589490/full#supplementary-material

REFERENCES

Acuña-Galindo, M. A., Mason, R. E., Subramanian, N. K., and Hays, D. B. (2015). Meta-analysis of wheat Qtl regions associated with adaptation to drought and heat stress. *Crop Sci.* 55, 477–492. doi: 10.2135/cropsci2013.11.0793

Adler, D., and Murdoch, D. (2013). *rgl: 3D Visualization Device System (opengl)*. Available online at: http://CRAN.R-project.org/package=rgl (accessed April 14, 2020).

Afzal, F., Li, H., Gul Kazi, A., Subhani, A., Ahmad, A., Mujeeb-Kazi, A., et al. (2019). Genome-wide analyses reveal footprints of divergent selection and drought adaptive traits in synthetic-derived wheats. *G3 Genes Gen. Genet.* 9, 1957–1973. doi: 10.1534/g3.119.400010

Ain, Q. U., Rasheed, A., Anwar, A., Mahmood, T., Imtiaz, M., Mahmood, T., et al. (2015). Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. *Front. Plant Sci.* 6:743. doi: 10.3389/fpls.2015.00743

Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., et al. (2018). Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.* 19:111. doi: 10.1186/s13059-018-1491-4

Assefa, T., Otyama, P. I., Brown, A. V., Kalberer, S. R., Kulkarni, R. S., and Cannon, S. B. (2019). Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genom.* 20:527. doi: 10.1186/s12864-019-5907-7

Azadi, A., Mardi, M., Hervan, E. M., Mohammadi, S. A., Moradi, F., Tabatabaee, M. T., et al. (2015). QTL mapping of yield and yield components under normal and salt-stress conditions in bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol. Rep.* 33, 102–120. doi: 10.1007/s11105-014-0726-0

Battenfield, S. D., Sheridan, J. L., Silva, L. D. C. E., Miclaus, K. J., Dreisigacker, S., Wilfinger, R. D., et al. (2018). Breeding-assisted genomics: applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS One* 3:e0204757. doi: 10.1371/journal.pone.0204757

Braz, C. U., Taylor, J. F., Bresolin, T., Espigolan, R., Feitosa, F. L. B., Carvalheiro, R., et al. (2019). Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. *BMC Genet.* 20:8. doi: 10.1186/s12863-019-0713-4

Breseghello, F., and Sorrells, M. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8057–8062. doi: 10.1073/pnas.1217133110

Cockram, J., Chiapparino, E., Taylor, S. A., Stamati, K., Donini, P., Laurie, D. A., et al. (2007). Haplotype analysis of vernalization loci in European barley germplasm reveals novel *VRN-H1* alleles and a predominant winter VRN-H1/VRN-H2 multi-locus haplotype. *Theor. Appl. Genet.* 115, 993–1001. doi: 10.1007/s00122-007-0626-x

Diao, G., and Vidyashankar, A. N. (2013). Assessing genome-wide statistical significance for large p small n problems. *Genetics* 194, 781–783. doi: 10.1534/genetics.113.150896

Dreisigacker, S., Kishii, M., Lage, J., and Warburton, M. (2008). Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. *Aust. J. Agric. Res.* 59, 413–420. doi: 10.1071/AR07225

Dreisigacker, S., Sehgal, D., Reyes Jaimez, A. E., Luna Garrido, B., Muñoz Zavala, S., Núñez Ríos, C., et al. (2016). *Laboratory Protocols and Applications to Wheat Breeding, Version 1*. México City: CIMMYT.

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* 127, 791–807. doi: 10.1007/s00122-013-2257-8

Evangelou, E., and Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 14, 379–389. doi: 10.1038/nrg3472

FAO (2020). *Crop Prospects and Food Situation–Quarterly Global Report No. 2, Rome, July 2020*. Rome: FAO.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229. doi: 10.1126/science.1069424

Ganal, M. W., Polley, A., Graner, E. M., Plieske, J., Wieseke, R., Luerssen, H., et al. (2012). Large SNP arrays for genotyping in crop plants. *J. Biosci.* 37, 821–828. doi: 10.1007/s12038-012-9225-3

Garcia, M., Eckermann, P., Haefele, S., Satija, S., Sznazdar, B., Timmins, A., et al. (2019). Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS One* 14:e0211730. doi: 10.1371/journal.pone.0211730

Hao, D., Cheng, H., Yin, Z., Cui, S., Zhang, D., Wang, H., et al. (2012). Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. *Theor. Appl. Genet.* 124, 447–458. doi: 10.1007/s00122-011-1719-0

He, T., Hill, C. B., Angessa, T. T., Zhang, X. Q., Chen, K., Moody, D., et al. (2019). Gene-set association and epistatic analyses reveal complex gene interaction networks affecting flowering time in a worldwide barley collection. *J. Exp. Bot.* 70, 5603–5616. doi: 10.1093/jxb/erz332

IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191.

Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., et al. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* 51, 1530–1539. doi: 10.1038/s41588-019-0496-6

Keeble-Gagnère, G., Rigault, P., Tibbits, J., Pasam, R., Hayden, M., Forrest, K., et al. (2018). Optical and physical mapping with local finishing enables megabase-scale resolution of agronomically important regions in the wheat genome. *Genome Biol.* 19:112. doi: 10.1186/s13059-018-1475-4

Ladejobi, O., Mackay, I. J., Poland, J., Praud, S., Hibberd, J. M., and Bentley, A. R. (2019). Reference genome anchoring of high-density markers for association mapping and genomic prediction in European winter wheat. *Front. Plant Sci.* 10:1278. doi: 10.3389/fpls.2019.01278

Ledesma-Ramírez, L., Solís-Moya, E., Iturriaga, G., Sehgal, D., Reyes-Valdes, M. H., Montero-Tavera, V., et al. (2019). GWAS to identify genetic loci for resistance to yellow rust in wheat pre-breeding lines derived from diverse exotic crosses. *Front. Plant Sci.* 10:1390. doi: 10.3389/fpls.2019.01390

Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., et al. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* 19:168. doi: 10.1186/s12870-019-1781-3

Li, X., Xu, X., Liu, W., Li, X., Yang, X., and Ru, Z. (2020). Dissection of superior alleles for yield-related traits and their distribution in important cultivars of wheat by association mapping. *Front. Plant Sci.* 11:175. doi: 10.3389/fpls.2020.00175

Liu, J., He, Z., Rasheed, A., Wen, W., Yan, J., Zhang, P., et al. (2017). Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:220. doi: 10.1186/s12870-017-1167-3

Liu, J., Rasheed, A., He, Z., Imtiaz, M., Arif, A., Mahmood, T., et al. (2019). Genome-wide variation patterns between landraces and cultivars uncover divergent selection during modern wheat breeding. *Theor. Appl. Genet.* 132, 2509–2523. doi: 10.1007/s00122-019-03367-4

Lopes, M. S., Dreisigacker, S., Peña, R. J., Sukumaran, S., and Reynolds, M. P. (2015). Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. *Theor. Appl. Genet.* 128, 453–464. doi: 10.1007/s00122-014-2444-2

Lu, Y., Xu, J., Yuan, Z., Hao, Z., Xie, C., Li, X., et al. (2012). Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. *Mol. Breed.* 30, 407–418. doi: 10.1007/s11032-011-9631-5

Mackay, T. F. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* 15, 22–33. doi: 10.1038/nrg3627

Mérida-García, R., Gálvez, S., Paux, E., Dorado, G., Pascual, L., and Giraldo, P. (2020). High resolution melting and insertion site-based polymorphism markers for wheat variability analysis and candidate genes selection at drought and heat MQTL loci. *Agronomy* 10:1294. doi: 10.3390/agronomy10091294

Mondal, S., Dutta, S., Crespo-Herrera, L., Huerta-Espino, J., Braun, H., and Singh, R. P. (2020). Fifty years of semi-dwarf spring wheat breeding at CIMMYT: grain yield progress in optimum, drought and heat stress environments. *Field Crops Res.* 250, 107757. doi: 10.1016/j.fcr.2020.107757

N'Diaye, A., Haile, J. K., Cory, A. T., Clarke, F. R., Clarke, J. M., Knox, R. E., et al. (2017). Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. *PLoS One* 12:e0187178. doi: 10.1371/journal.pone.0187178

Neumann, K., Kobiljski, B., Denčić, S., Varshney, R. K., and Börner, A. (2011). Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* 27, 37–58. doi: 10.1007/s11032-010-9411-7

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7:e32253. doi: 10.1371/journal.pone.0032253

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qaseem, M. F., Qureshi, R., Muqaddasi, Q. H., Shaheen, H., Kousar, R., and Röder, M. S. (2018). Genome-wide association mapping in bread wheat subjected to independent and combined high temperature and drought stress. *PLoS One* 13:e0199121. doi: 10.1371/journal.pone.0199121

Qian, L., Hickey, L. T., Stahl, A., Werner, C. R., Hayes, B., Snowdon, R. J., et al. (2017). Exploring and harnessing haplotype diversity to improve yield stability in crops. *Front. Plant Sci.* 8:1534. doi: 10.3389/fpls.2017.01534

Quarrie, S., Quarrie, S. P., Radosevic, R., Rancic, D., Kaminska, A., Barnes, J. D., et al. (2006). Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. *J. Exp. Bot.* 57, 2627–2637. doi: 10.1093/jxb/erl026

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:eaar6089. doi: 10.1126/science.aar6089

Reif, J. C., Maurer, H. P., Korzun, V., Ebmeyer, E., Miedaner, T., Wurschum, T., et al. (2011). Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. *Theor. Appl. Genet.* 123, 283–292. doi: 10.1007/s00122-011-1583-y

Rio, S., Mary-Huard, T., Moreau, L., Bauland, C., Palaffre, C., and Madur, D. (2020). Disentangling group specific QTL allele effects from genetic background epistasis using admixed individuals in GWAS: an application to maize flowering. *PLoS Genet.* 16:e1008241. doi: 10.1371/journal.pgen.1008241

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Perez, L. G., Crossa, J., et al. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3* 6, 2799–2808. doi: 10.1534/g3.116.032888

Schwarz, E. G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Sehgal, D., Autrique, E., Singh, R. P., Ellis, M., Singh, S., Dreisigacker, S., et al. (2017). Identification of genomic regions for grain yield and yield stability and their epistatic interactions. *Sci. Rep.* 7:41578. doi: 10.1038/srep41578

Sehgal, D., and Dreisigacker, S. (2019). Haplotypes-based genetic analysis: benefits and challenges. *Vav. J. Genet. Breed.* 23, 803–808. doi: 10.18699/VJ19.37-o

Sehgal, D., Dreisigacker, S., Belen, S., Küçüközdemir, Ü, Mert, Z., Özer, E., et al. (2016). Mining centuries old *in situ* conserved Turkish wheat landraces for grain yield and stripe rust resistance genes. *Front. Genet.* 7:201. doi: 10.3389/fgene.2016.00201

Sehgal, D., Mondal, S., Guzman, C., Barrios, G. G., Franco, C., Singh, R., et al. (2019). Validation of candidate gene-based markers and identification of novel loci for thousand -grain weight in spring bread wheat. *Front. Plant Sci.* 10:1189. doi: 10.3389/fpls.2019.01189

Sehgal, D., Rosyara, U., Mondal, S., Singh, R., Poland, J., and Dreisigacker, S. (2020). Incorporating genome-wide association mapping results into genomic prediction models for grain yield and yield stability in CIMMYT spring bread wheat. *Front. Plant Sci.* 11:197. doi: 10.3389/fpls.2020.00197

Sehgal, D., Vikram, P., Sansaloni, C. P., Ortiz, C., Saint Pierre, C., Payne, T., et al. (2015). Exploring and mobilizing the gene bank biodiversity for wheat improvement. *PLoS One* 10:e0132112. doi: 10.1371/journal.pone.0132112

Shokat, S., Sehgal, D., Fulai, L., and Sukhwinder, S. (2020). GWAS analysis of wheat pre-breeding germplasm for terminal drought stress using next generation sequencing technology. *Int. J. Mol. Sci.* 21:3156. doi: 10.20944/preprints202002.0272.v1

Singh, S., Vikram, P., Sehgal, D., Burguano, J., Sharma, A., Singh, S. K., et al. (2018). Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Nat. Sci. Rep.* 8:12527. doi: 10.1038/s41598-018-30667-4

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redona, E., Jannink, J.-L., et al. (2016). Genome-wide prediction models that incorporate *de novo* GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116, 395–408. doi: 10.1038/hdy.2015.113

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments *Theor. Appl. Genet.* 128, 353–363. doi: 10.1007/s00122-014-2435-3

Tadesse, W., Sanchez-Garcia, M. A., Solomon, A., Ahmed, B., Zewdie, O., and Francis Baum, M. (2019). Genetic gains in wheat breeding and its role in feeding the world. *Crop Breed. Genet. Genom.* 1:e190005.

Templeton, A. R., Maxwell, T., Posada, D., Stengard, J. H., Boerwinkle, E., and Sing, C. F. (2005). Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. *Genetics* 169, 441–453. doi: 10.1534/genetics.104.030080

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Wang, S. C., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Warburton, M. L., Crossa, J., Franco, J., Kazi, M., Trethowan, R., Rajaram, S., et al. (2006). Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica* 149, 289–301. doi: 10.1007/s10681-005-9077-0

Ward, B. P., Brown-Guedira, G., Kolb, F. L., Van Sanford, D. A., Tyagi, P., Sneller, C. H., et al. (2019). Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. *PLoS One* 4:e0208217. doi: 10.1371/journal.pone.0208217

Wicker, T., Gundlach, H., Spannagl, M., Cristobal, U., Borrill, P., Ramirez-Gonzalez, R. H., et al. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol.* 19:103. doi: 10.1186/s13059-018-1479-0

Zhang, K., Wang, J., Zhang, L., Rong, C., Zhao, F., Peng, T., et al. (2013). Association analysis of genomic loci important for grain weight control in elite common wheat varieties cultivated with variable water and fertilizer supply. *PLoS One* 8:e57853. doi: 10.1371/journal.pone.0057853

Zuk, O., Schaffner, S. F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., et al. (2014). Searching for missing heritability: designing rare variant association studies. *Proc. Acad. Natl. Sci. U.S.A.* 111, 455–464. doi: 10.1073/pnas.1322563111

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sehgal, Mondal, Crespo-Herrera, Velu, Juliana, Huerta-Espino, Shrestha, Poland, Singh and Dreisigacker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 December 2020
doi: 10.3389/fgene.2020.578123

Strategies Using Genomic Selection to Increase Genetic Gain in Breeding Programs for Wheat

Biructawit Bekele Tessema[1], Huiming Liu[1], Anders Christian Sørensen[1], Jeppe Reitan Andersen[2] and Just Jensen[1]*

[1] Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark, [2] Nordic Seed A/S, Odder, Denmark

Conventional wheat-breeding programs involve crossing parental lines and subsequent selfing of the offspring for several generations to obtain inbred lines. Such a breeding program takes more than 8 years to develop a variety. Although wheat-breeding programs have been running for many years, genetic gain has been limited. However, the use of genomic information as selection criterion can increase selection accuracy and that would contribute to increased genetic gain. The main objective of this study was to quantify the increase in genetic gain by implementing genomic selection in traditional wheat-breeding programs. In addition, we investigated the effect of genetic correlation between different traits on genetic gain. A stochastic simulation was used to evaluate wheat-breeding programs that run simultaneously for 25 years with phenotypic or genomic selection. Genetic gain and genetic variance of wheat-breeding program based on phenotypes was compared to the one with genomic selection. Genetic gain from the wheat-breeding program based on genomic estimated breeding values (GEBVs) has tripled compared to phenotypic selection. Genomic selection is a promising strategy for improving genetic gain in wheat-breeding programs.

Keywords: wheat, genetic gain, genomic selection, stochastic simulation, genetic correlation

OPEN ACCESS

Edited by:
Andrew H. Paterson, University of Georgia, United States

Reviewed by:
Arif Ashraf, University of Massachusetts Amherst, United States
Ashish Kumar Srivastava, Bhabha Atomic Research Centre (BARC), India

****Correspondence:***
Biructawit Bekele Tessema
biructawit.tessema@qgg.au.dk

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *30 June 2020*
Accepted: *11 November 2020*
Published: *04 December 2020*

Citation:
Tessema BB, Liu H, Sørensen AC, Andersen JR and Jensen J (2020) Strategies Using Genomic Selection to Increase Genetic Gain in Breeding Programs for Wheat. Front. Genet. 11:578123. doi: 10.3389/fgene.2020.578123

INTRODUCTION

Conventional wheat-breeding programs use phenotypic values for selection of best individuals. Such programs are reported to yield yearly genetic gains that are lower than 1% (Godin et al., 2012). The procedure in phenotype-based selection involves creating genetic variation by crossing two parents followed by several rounds of selfing to create inbred lines. Resulting inbred lines are tested for a range of phenotypic parameters. Besides the most important trait, grain yield, breeders evaluate traits such as disease resistance, lodging, quality parameters, and a range of agronomical traits. In the early generations, breeders "visual preference" which is based on previous experience can also influence selection decisions. The aim of a breeding programs is to develop superior cultivars and the phenotypes of all the traits of interest are used for all selection decisions. However, it is the selection of individuals based on their breeding values, which would influence the response to selection in the next generation of a breeding cycle (Akdemir et al., 2019). In addition, selection of parental lines based on their breeding values will influence genetic gain in the subsequent breeding cycles. With the advent of molecular markers, selection decision was made by integrating information from both molecular markers and phenotypic data through marker-assisted

selection (MAS) (Fujino et al., 2019). However, many complex traits including yield is under the control of many genes with small effects where MAS will be of a limited use due to low statistical power to detect individual genes (Bernardo, 2008). For many major crop plants, including wheat, many QTLs have been identified for many different traits, but the practical application of MAS faces many limitations (Bernardo, 2008) mainly because the QTL identified only account for a limited fraction of the genetic variance.

In contrast to MAS, genomic selection (GS) uses genome-wide markers to capture both large and small effect QTL to predict breeding values for complex traits (Meuwissen et al., 2001). Genomic prediction of expected breeding values will have advantages over phenotypic selection mainly because the accuracy in estimating breeding value is higher when genomic information is included for selection decision (Daetwyler et al., 2013). Genomic estimated breeding values (GEBVs) are calculated as the sum of effects related to genetic markers in linkage disequilibrium (LD) with one or more QTLs across the entire genome (Goddard and Hayes, 2007). Genomic selection uses a prediction model that is first trained using a population that contains both genotyped and phenotyped individuals. The trained model is then used to predict true breeding values of selection candidates. Such selection candidates may have no phenotypes and then their performance will be based on genomic information only. However, for selection candidates that are phenotyped and genotyped accuracy of prediction will improve due to optimal combination of genomic and phenotypic information.

A number of studies have explored application of genomic information in breeding programs for different plant species. Bernardo and Yu (2007) reported a 43% increase in genetic gain in a simulation study by integrating genomic information in maize breeding program compared to a program based on marker-assisted selection. A simulation study of Gaynor et al. (2017) compared wheat-breeding programs with and without genomic information and showed breeding program with genomic information outperformed phenotype based breeding program. A breeding program where selection decisions are only based on phenotypes aims to select best lines from a large segregating early generation and to evaluate fewer lines with greater replication in advanced generations. Integrating genomic selection in conventional wheat-breeding programs can increase genetic gain by selecting superior inbred line with higher selection accuracy (Bassi et al., 2015; Gaynor et al., 2017). In this way, only few changes are required in an already on-going phenotype-based wheat-breeding program. Generally, conventional wheat breeding starts by crossing selected inbred lines followed by several generations of selfing to get stable inbred lines for yield evaluations.

The motivation of this study comes from the observation that some of the simulation studies on genomic selection in wheat-breeding program does not mimic the complexity of actual wheat-breeding programs (Gaynor et al., 2017). In this simulation, we used real wheat genome data and the performance of genomic selection using real haplotype data would be closer to what happens in actual wheat-breeding program. Heritability of traits was based on real data study (Cericola et al., 2017). Plot size and density, yield plot at preliminary yield trial and advanced yield trial mimicked real wheat-breeding setup. In addition, the different actions taken at each breeding cycle stage reflected actual wheat-breeding programs. This was possible due to the use of the complex stochastic simulation program ADAM (Liu et al., 2019). The results of this study will contribute to the research on plant breeding as well as serve as a clear guideline for the plant breeding companies for implementation of genomic selection.

We applied genomic selection in conventional wheat-breeding programs to investigate the expected change in genetic gain. Stochastic simulation was used to quantify expected genetic gain in conventional wheat-breeding programs when incorporating genomic information in making selection decision. Stochastic simulation allow to model breeding schemes that mimics an actual breeding program on a very detailed level (Liu et al., 2019). For instance, stochastic simulation can be used to simulate an entire population of individual plants and it can accounts for the change in allele frequencies caused by selection. It also allows simulation of many rounds of selection and results in a more accurate estimate of genetic variance and response to selection. This makes stochastic simulation very precise in predicting consequences of alternative breeding schemes. Studies have also showed the benefit of stochastic simulation (Wensch-Dorendorf et al., 2011; Gaynor et al., 2017).

In the present simulation work, we compared phenotype-based selection programs with genomic-based selection programs, where both selection criteria applied in conventional wheat-breeding programs. To achieve this, phenotype and genomic-based breeding schemes were simulated following a standard commercial wheat-breeding program, where the number of crosses, families, and single plants mimic real life wheat-breeding program. The aims of this study are to (1) investigate if incorporation of genomic information in wheat breeding program increases genetic gain, (2) study the change in genetic variance between phenotypic and genomic selection, and (3) investigate accuracy of selection in phenotypic and genomic selection.

MATERIALS AND METHODS

Simulation Design

A stochastic simulation model was used to simulate a conventional wheat breeding strategy that run for 25 breeding cycles. Simulations were carried out using ADAM software (Liu et al., 2019). Two different wheat breeding strategies were simulated, (1) phenotypic selection and (2) genomic selection. Phenotypes and underlying genotypes were simulated including both QTL and markers. Based on phenotypic (and marker) information (G)EBV were predicted using a linear model.

A new breeding cycle was initiated every year and the breeding cycles, therefore, were overlapping (**Figure 1**). A breeding cycle represent eight generations from initial crossing until final elite line selection. A breeding cycle starts with parental lines (P), followed by generation F1 up to F8, where the number represent the generation in which they are generated (**Figure 1**). Since the

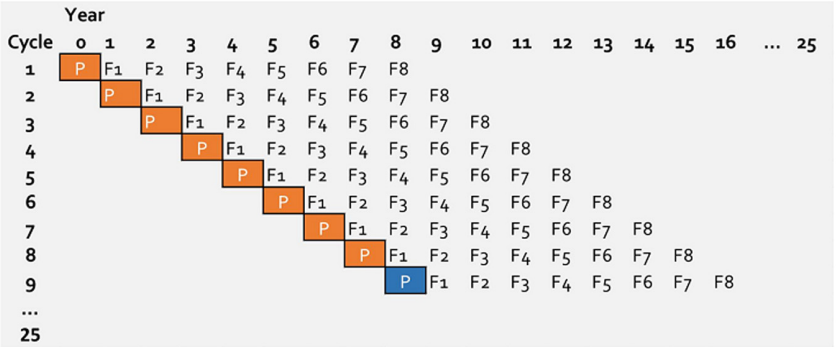


FIGURE 1 | Structure of simulated wheat breeding program running over 25 years. Every year a new breeding cycle is initiated. For the first 7 years parents (P with orange box) are selected from the base population and after year 7 parents (P with blue box) is selected from previous cycles. For genomic selection the first 8 years are burn-in period.

breeding cycles are overlapping, in a given year, events (selection or mating) on eight different cycles were simulated. This allowed movement of information between cycles every year.

In a practical genomic wheat-breeding program, genomic selection would be implemented in an on-going phenotypic selection program. To mimic this, from the simulated 25 years of breeding program for genomic selection, the selection in the first 8 years was based on phenotypes. This means genomic information was introduced on the ninth year of the breeding program. This burn-in stage has also served to create differences between breeding cycles.

Simulation of Founder Population

The genome for the founder population was generated from a realized wheat genome data set (Cericola et al., 2017) that were read into ADAM software. The genome data set represents commercial wheat lines obtained from three breeding cycles and includes 988 F6 lines. In the first generation of founder population, all SNPs are evenly distributed across the total (21) chromosomes and every *N*th locus harbors a QTL that code for a trait under selection and the remaining loci are genetic markers. Thus from the total 9582 markers, 1039 loci were randomly chosen across the genome was assumed as QTLs while the remaining 8543 loci were assumed as anonymous markers.

The extent of linkage disequilibrium in the genome data set is explained in Cericola et al. (2017). The average r^2 within chromosomes was 0.05, indicating the presence of low LD decay. The average r^2 for the genome A, B, and D was 0.05, 0.05, and 0.11, respectively. The average distance of markers with $r^2 > 0.5$ was 9.47, 8.38, and 7.73 cM for genome A, B, and D, respectively.

Simulation of Base Population

A base population of 480 lines were generated from the founder population. The genotype of each line was sampled from a pool of chromosomes of the founder population. Each line was generated by, for each chromosome 1 to 21, randomly sampling one chromosome without replacement from the pool of chromosomes. Then the second chromosome is set to be identical to the first one to generate a fully inbred line.

Parental lines were chosen from the base population for the first seven breeding cycles. The breeding program was run in parallel, meaning that a new breeding cycle was started every year. After this stage, parents were selected randomly from F6, F7, and F8 of the previous breeding cycles.

Simulation of Phenotypes

Three traits for selections were simulated and these were breeder's visual preference (BVP), yield at preliminary yield trial (PYT) and yield at advanced yield trial (AYT). The observation of the traits were realized at the different stages of a breeding cycle, breeder's visual preference (BVP) at F2 and F4 generation, yield at preliminary yield trial (PYT) at F5 and yield at advanced yield trial (AYT) at F6 and F7 generations. AYT was also applied for family selection at F3. The preliminary yield trial represents un-replicated plot with limited amount of seeds sown sparsely. Whereas advanced yield trial is a standard yield plot with normal seed density and plots replicated three times in three different locations. The phenotype, y, was calculated as, $y = g + e$, where g is the true additive-genetic value and e is residual value. True breeding values (TBVs) of the traits were determined by summing the allelic effects of its QTL. The QTL effects $\left(a_j^{'}\right)$ was sampled from multivariate normal distribution, that is $a_j = a_j^i \times \sqrt{\frac{\sigma^2_{qtl}}{\sum_k^n 2pk(1-pk)a_k^{'2}}}$, where subscripts k and j denote the QTL k and QTL j, p_k and p_j are the minor allele frequencies of QTL k and QTL j, $a_k^{'}$ and $a_j^{'}$ are the substitution effect of QTL k and QTL j before being scaled. When simulating correlation between traits, additive genetic variance and heritability of each trait was specified along with the genetic and residual correlations between traits. The additive genetic covariance matrix was then derived from the additive genetic variances and the additive genetic correlation matrix (Faux et al., 2016). The genetic variance for each trait was set to 1 (standardized unit) in the base population. The residual effect (e) for each individual is sampled from a normal distribution $N \sim \left(O,\sigma_e^2\right)$. The residual was simulated independently assuming phenotypic correlation from genetic

correlation. Heritability (h^2) for BVP was set 0.1 and calculated $h^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2}$. Plot heritability $\left(h_{plot}^2\right)$ for yield at PYT and AYT were 0.2 and 0.3, respectively, based on results of Cericola et al. (2017) which was based on results from a large commercial wheat population. In order to achieve targeted plot heritability for yield trait, the value for residual variance $\left(\sigma_e^2\right)$ was calibrated following this equation derived from Falconer and Mackay (1996) under the assumption of Hardy-Weinberg Equilibrium.

$$\sigma_e^2 = \frac{2N_p H\sigma_g^2\left(1-h_{plot}^2\right)}{h_{plot}^2} - (1-H)\,\sigma_g^2$$

Where H is expected frequency of homozygosity in the generation, σ_g^2 is genetic variance, σ_e^2 is residual variance.

Statistical Model

Breeding value estimation was done using multi-trait BLUP model in DMU software (Madsen and Jensen, 2013). For genomic selection a GBLUP model was implemented.

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \boldsymbol{y_3} \end{bmatrix} = \begin{bmatrix} \mathbf{1_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1_3} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \boldsymbol{e_3} \end{bmatrix}$$

Where $\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \boldsymbol{y_3} \end{bmatrix}$ is the vector of traits BVP, PYT and AYT, and $\mathbf{1}_1$, $\mathbf{1}_2$ **and** $\mathbf{1}_3$ are the identity matrices, $\begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_1} \\ \boldsymbol{\mu_1} \end{bmatrix}$ is the vector of population means of BVP, PYT and AYT, $\begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{bmatrix}$ is the vector of additive genetic effects of the three traits. $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ are the design matrices that associate breeding values with BVP, PYT and AYT and $\begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \boldsymbol{e_3} \end{bmatrix}$ is the vector of residual errors of BVP, yield at PYT and AYT. It is assumed that $\begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{bmatrix} \sim N\,(\mathrm{O},\, G \otimes \boldsymbol{G}_0)$ where $\boldsymbol{G_O}$ is $\begin{bmatrix} \sigma_{a1}^2 & \sigma_{a12} & \sigma_{a13} \\ \sigma_{a21} & \sigma_{a2}^2 & \sigma_{a23} \\ \sigma_{a21} & \sigma_{a32} & \sigma_{a3}^2 \end{bmatrix}$ when genomic information is used and $\begin{bmatrix} \boldsymbol{p_1} \\ \boldsymbol{p_2} \\ \boldsymbol{p_3} \end{bmatrix} \sim N\,(\mathrm{O},\, I \otimes \mathbf{I_0})$ where $I_0 \begin{bmatrix} \sigma_{p1}^2 & \sigma_{p12} & \sigma_{p13} \\ \sigma_{p21} & \sigma_{p2}^2 & \sigma_{p23} \\ \sigma_{p31} & \sigma_{p32} & \sigma_{p3}^2 \end{bmatrix}$ when phenotypic information is used. $\begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \boldsymbol{e_3} \end{bmatrix} \sim N\,(\mathrm{O},\, I \otimes \boldsymbol{R})$, where $\boldsymbol{R}$ is the residual variance covariance matrix of the three traits. The three traits are measured on different lines and different years meaning the residual variance is independent and have value of zero.

Breeding Schemes

Phenotypic Selection (PS) Breeding Scheme

A breeding program running for 25 years was simulated following a conventional winter wheat breeding structure (**Figure 1**). Every year in the breeding program a new breeding cycle starts.

Parents (P): For the first seven breeding cycles, 60 parental lines were chosen randomly from the base population while for the rest of 18 breeding cycles parents were selected based on their breeding values from F6, F7, and F8 of the previous breeding cycles and this allowed elite lines to be used for crossing (**Figure 2**). Each parental line was used for a maximum of six crosses and from the total 1770 possible crosses, 100 crosses were randomly chosen.

F1: One hundred F1 plants were generated and each F1 was allowed to produce 30 (F2) seeds. Families from F2 to F4 share a common ancestor (F1), thus families from F2 to F4 were referred as F1 families.

F2: In total, there were 100 plots and each F2 plot had 30 plants. With-in each F1 family eight highest-ranking single plants were selected based on their breeding value for trait 1 (BVP). Heritability of 0.1 was assigned for the trait BVP. Selected eight single plants from each family were advanced to the next stage in the breeding program.

F3: Each F1 family was planted in three plots and evaluated based on yield performance. The plot set up was similar as in AYT. Out of the 100 families, 75 highest ranking families based on their breeding values were advanced to the next stage.

F4: The advanced 75 F1 families were planted in un-replicated field trials. Ten single plants per family were selected based of BVP in order to form individual lines based on single seed descent. Selected 750 lines were advanced to preliminary yield trial (PYT).

F5: F5 lines were created by advancing selected F4 individuals through single seed descent (SSD). A single plot was simulated to generate 50 plants that were genetically identical. This preliminary yield trial (PYT) is un-replicated and this step is mainly done for the purpose of seed multiplication. Yield recording was made on each plot. Plot mean was used to select 150 best lines from the available 750 lines. Selected lines were stored to be potentially used as parents in the proceeding breeding cycles.

F6: The 150 lines selected from F5 were evaluated in advanced yield trail (AYT), where plots and trials were replicated three times across three different locations. However, yield evaluation here was made based on the assumption that there is no genotype by environment interaction and heritability is chosen to reflect the part of additive genetic variance that is common to different environments (Cericola et al., 2017). The

AYT trait was set as a selection criteria. Thirty best lines were selected from the 150 lines to be advanced to the next stage. The selected lines will be added to a pool of potential parents.

F7: The selected 30 best lines were evaluated for yield and the field trial setup was similar to the AYT in F6. Five best lines are selected based on yield and were selfed to produce F8. Besides, the selected lines will be stored to potentially become parents. The best F8 lines are sent to an official yield trial test to be released as varieties.

Genomic Selection (GS) Breeding Scheme

The proposed genomic wheat-breeding program was designed to introduce genomic selection starting from the preliminary yield trial stage, F5 (**Figure 2**). The genomic breeding program combines genotype and phenotype information to predict GEBVs for yield. The first 8 years of the breeding program was a burn-in phase which had similar set up as phenotypic selection described earlier, which means genomic information was introduced in the breeding program at year 9. The genomic breeding program from year 9 to 25 is described as follows.

F4: For genomic selection breeding strategy, all selected 750 single plants were genotyped in each cycle. F4 genotypes together with phenotypes for the targeted traits were added yearly for building the reference population. The initial reference population at year 16 (cycle 16) contained 750 genotypes. Every year the reference populations increased by 750 new genotypes.

F5: Each F5 plot had 50 plants descended from the single seed and only one replicate was simulated. Yield was recorded on all F5 plots in each F4-line. Line selection was done based on

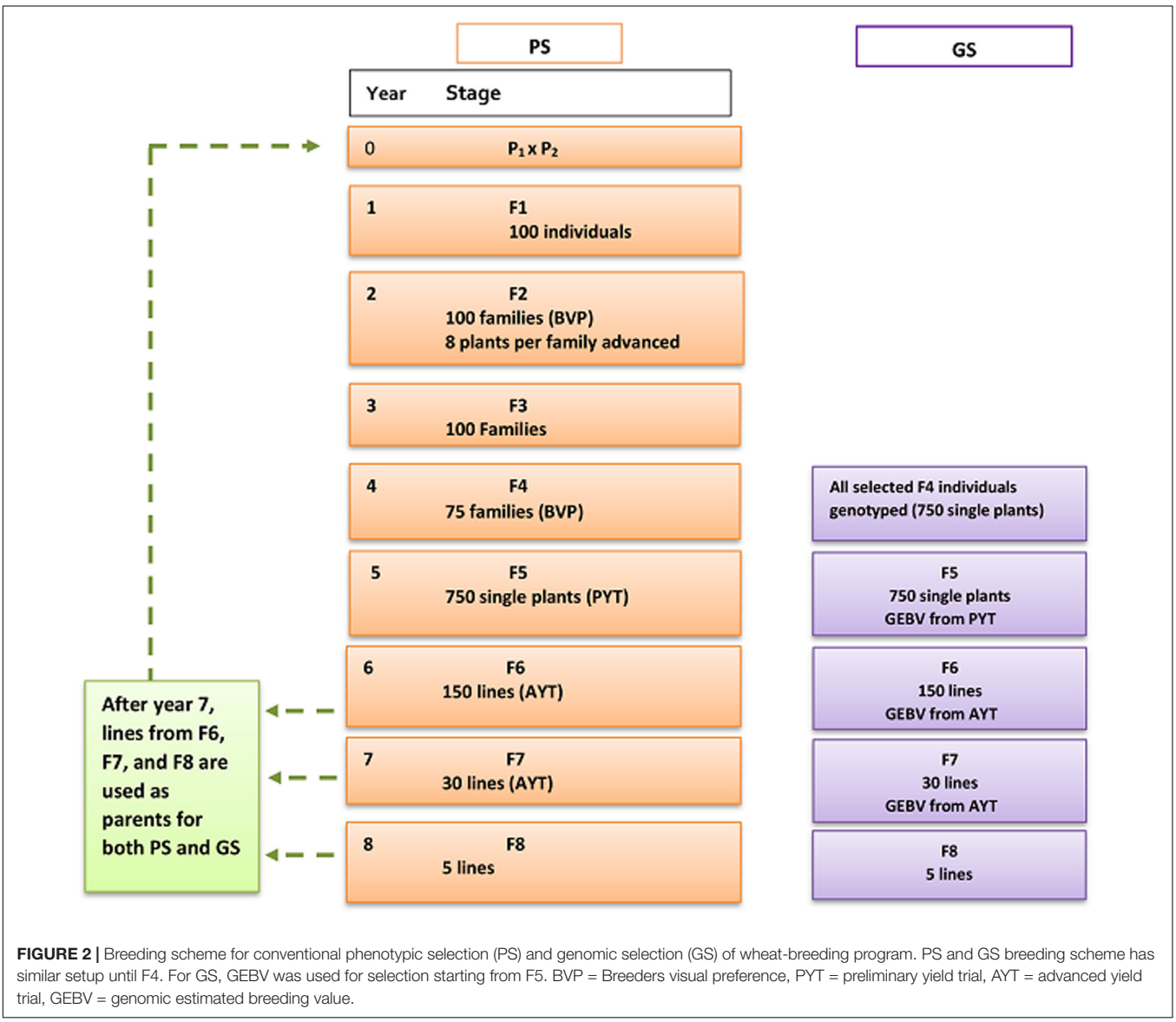


FIGURE 2 | Breeding scheme for conventional phenotypic selection (PS) and genomic selection (GS) of wheat-breeding program. PS and GS breeding scheme has similar setup until F4. For GS, GEBV was used for selection starting from F5. BVP = Breeders visual preference, PYT = preliminary yield trial, AYT = advanced yield trial, GEBV = genomic estimated breeding value.

plot yield performance of 750 F4-lines. The phenotypes and genotypes of 750 F4-lines together with the existing reference population were used to estimate breeding values of each lines using GBLUP model (Madsen and Jensen, 2013). Thus, GEBVs of PYT were used to select the highest ranking 150 lines to be advanced to the next stage. The information of phenotypes and genotypes in the current breeding cycle was stored to be used for predicting breeding values in the next cycles. The germplasm of the 150 selected lines were stored and potentially become parental lines for the proceeding cycles.

F6: For each F6 line, nine replicates of plot was simulated. 150 F6 lines are phenotyped for yield for all the nine replicates in an advanced yield trial. Each F6 plot had 1500 plants. The phenotype of 150 lines combined with the genotype information of their corresponding F4 genotypes were used for prediction of breeding values. Based on GEBV of AYT, the 30 highest-ranking lines were selected from 150 lines. The selected lines were stored to be used as parents in the proceeding breeding cycles.

F7: The advanced 30 lines were evaluated for yield on all the nine replicates similar to F6. Each plot had 1500 plants. The average yield performance of each plot was recorded. The phenotype and genotype of the 30 lines together with the existing reference population was used to estimate breeding values. Five lines out of 30 lines were selected based on their GEBVs.

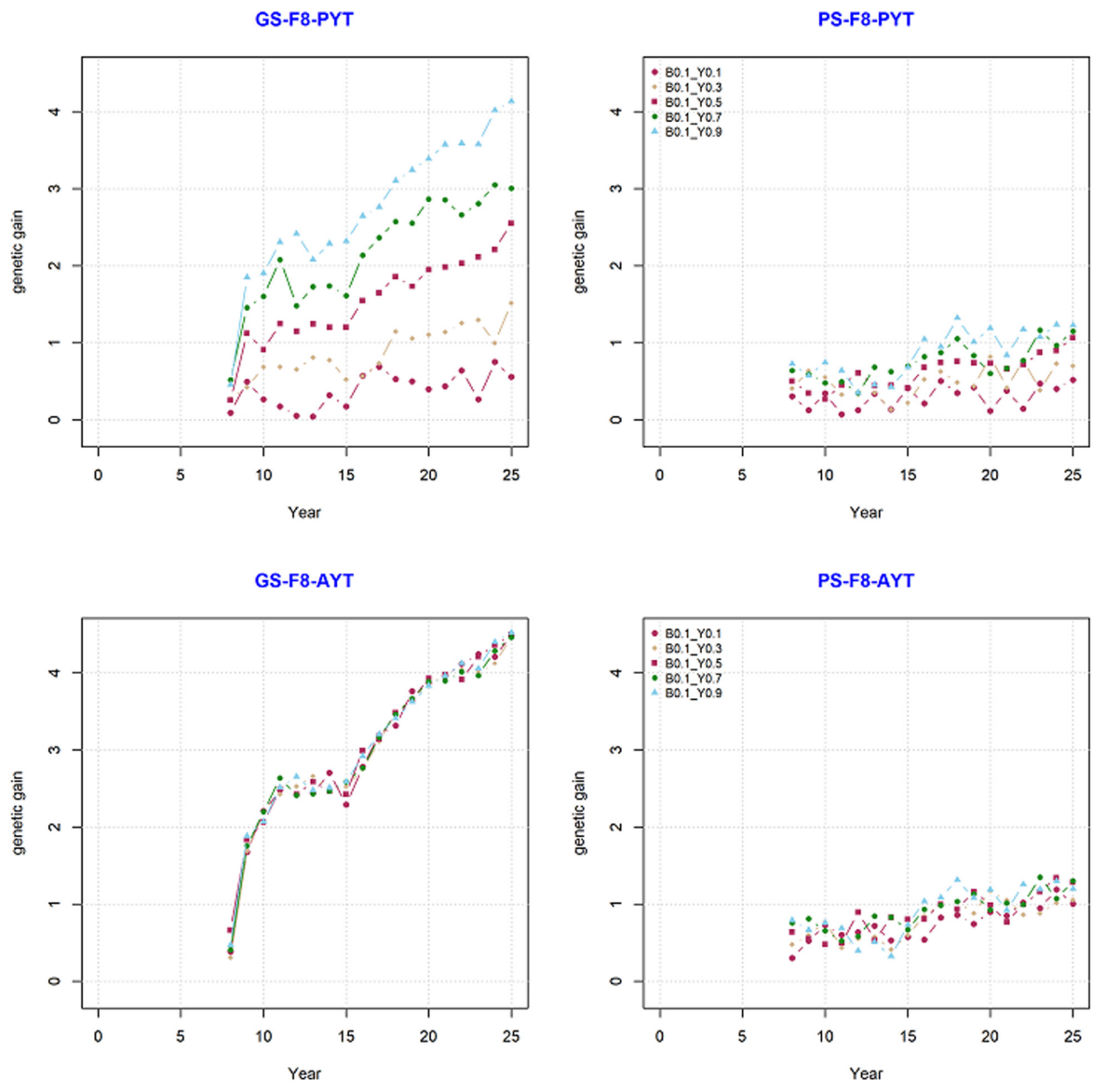


FIGURE 3 | Genetic gain for the final product (F8) for PYT and AYT for genomic and phenotypic selection. The five levels of correlation (Y0.1, Y0.3, Y0.5, Y0.7, and Y0.9) between PYT and AYT, and correlation of BVP with PYT and AYT (B0.1) is shown.

For both GS and PS, five scenarios were simulated with five different levels of genetic correlation between PYT and AYT. The correlation levels are 0.1, 0.3, 0.5, 0.7, and 0.9. The correlation of BVP with PYT and AYT was 0.1 for all the five scenarios. Each scenario in the simulation was replicated 10 times.

Additive Genetic Variance

Additive genetic variance was computed in generation F5 as the variance of mean TBVs of all the individuals in each breeding cycle. Generation F5 is the breeding cycle stage where there is highest selection intensity.

Prediction Accuracy

Selection accuracy for phenotypic selection and genomic selection was evaluated as a correlation between predicted breeding values and TBVs in each cycle. For phenotypic selection, accuracy was calculated as a correlation between plot phenotype and TBVs while for genomic selection it was a correlation between plot GEBVs and TBV.

Comparison of Breeding Strategies

The expected genetic gain was quantified from each simulated breeding strategy. Genetic gain comparison on F8 generation was done by plotting the mean of true breeding values (TBVs) of F8 individuals against time. For the comparison of conventional and genomic breeding strategies, data from years 9 to 25 was used in the analysis. The annual genetic gain was computed for each scenario by regressing TBVs of AYT on time with the assumption of linear response from years 9 to 25 within 10 replicates. The standard deviation of genetic response for AYT is estimated as a measure of uncertainty of the breeding program. The development of genetic variance over several rounds of selection was compared between genomic and phenotypic selection. Genetic variance per breeding cycle (F1–F8) is plotted over time. The change in genetic variance for all scenarios at the end of breeding cycle is presented. Plotting and calculation was done using R statistical programming language and environment (R Core Team, 2019).

RESULTS

Genetic Gain

Genetic gain from phenotypic selection and genomic selection was compared based on their mean breeding values over the period of 25 years. The comparison was done for trait PYT and AYT of the F8 generations since they were the end product in the current simulation. Our simulation result showed that the breeding program that uses genomic information generated significantly higher genetic gain than the breeding program with only phenotypic selection (**Figure 3**). The change in genetic gain when GS is introduced at year nine to an already ongoing PS can be clearly seen in **Figure 3**. There is a high jump in genetic gain at year nine in GS compared to PS. The trend in genetic gain was increasing at higher rate over time in GS compared to PS. The difference in genetic trends can also be seen among different level of genetic correlations between PYT and AYT within both PS and GS (**Figure 3**).

Genetic gain within PS and GS was also compared based on the level of genetic correlation between PYT and AYT and there were five level of genetic correlations (0.1, 0.3, 0.5, 0.7, and 0.9). According to the level of genetic correlations between traits there were a difference in mean breeding value with-in each selection criteria (PS and GS). The difference in genetic gain for the different levels of correlation can be clearly seen for PYT than AYT. The reason for this is that AYT was considered as the breeding goal and thus economic weight was assigned to AYT. This means genetic responses for PYT depends on the level of correlation with AYT for GS, the higher the correlation with AYT the higher genetic gain for PYT. This happened because of the genetic correlation between PYT and AYT as well as the genomic information available from the previous cycles are used for prediction. In addition, selection intensity at PYT is higher than AYT. However, the highest genetic gain is realized for AYT. The different levels of genetic correlation for AYT did not show significant differences among the different correlations. The standard deviation (SD) of the different scenarios for PS and GS breeding schemes tells the uncertainty of the breeding program. The SD of the individual estimates for AYT ranged 0.310–0.378 for PS whereas for GS the range is 1.127–1.143 (**Table 1**).

Annual Genetic Gain

Annual genetic gain represents how much genetic gain have been obtained from each breeding cycle. Annual genetic gain for PS and GS was compared and the result for the five tested scenarios of PS and GS is presented in **Table 1**. Breeding programs that use genomic information has higher genetic annual genetic gain in all the tested scenarios than breeding program with phenotypic information only. For PS, annual genetic gain range from 0.035 to 0.049 while for GS the range was from 0.183 to 0.186 genetic standard deviation. Compared to PS, GS breeding scheme has tripled genetic gain. This increase in genetic gain when using GS was seen across the different levels of genetic correlations between traits. However, there was no significant differences of among the different levels of genetic correlations with in GS as well as PS.

TABLE 1 | Rate of genetic gain (ΔG) for AYT in genetic standard deviation with standard errors in brackets from year 9 to 25 and for 10 replicates and standard deviation (SD) of AYT for PS (phenotypic selection) and GS (genomic selection).

	ΔG		SD	
Scenario	**PS**	**GS**	**GS**	**PS**
B0.1_Y0.1	0.035(0.001)	0.186(0.004)	1.141	0.310
B0.1_Y0.3	0.041(0.001)	0.185(0.002)	1.138	0.345
B0.1_Y0.5	0.044(0.001)	0.186(0.002)	1.143	0.356
B0.1_Y0.7	0.045(0.001)	0.183(0.002)	1.127	0.348
B0.1_Y0.9	0.049(0.001)	0.183(0.003)	1.127	0.378

The five level of correlation (Y0.1, Y0.3, Y0.5, Y0.7, and Y0.9) between PYT and AYT, and correlation of BVP with PYT and AYT (B0.1).

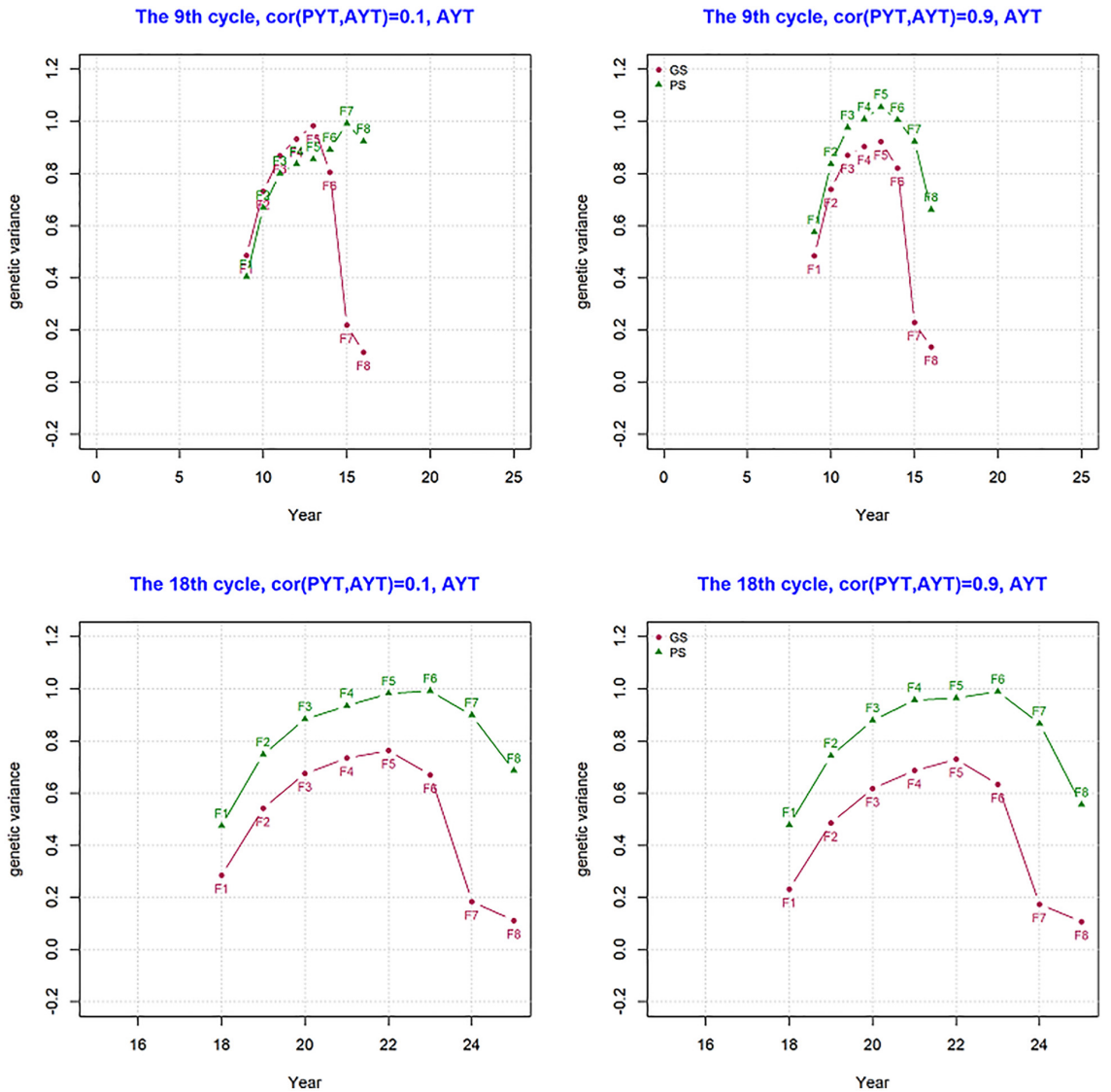


FIGURE 4 | Genetic variance for AYT (advanced yield trial) in a singe breeding cycle from F1 to F8 at year 9 and 18 for genetic correlation of 0.1 and 0.9 between PYT and AYT.

Genetic Variance per Cycle

Genetic variance per cycle was measured and this shows how much of the genetic variance created by crossing and recombination is reduced along the different stages of selection within a breeding cycle. Genetic variance for trait AYT was measured within a breeding cycle for PS and GS breeding schemes and the result is shown in **Figure 4**. The figure shows what happens in a single breeding cycle (F1 to F8) that started in the year 9 and 18 of the breeding program as an example for both PS and GS breeding schemes. In both year 9 and 18 of the breeding programs, an increase in genetic variance happened from F1 to F5 due to recombination. This trend is seen for both PS and GS breeding schemes. Genetic variance at F1 is 0.5 which is half of parent's variance and this is expected since the parents are inbred and Mendelian sampling is 0. The decline in genetic variance after F5 comes as a result of selection. After selection only the best lines are kept to proceed to the next breeding stage. In GS, the change in variance after F5 shows a sharp decline due to selection when compared to PS due to more accurate predicted breeding values.

Genetic Variance Across Breeding Cycles

The development of genetic variance over the period of 25 years was measured and the result is present in **Figure 5**. The figure shows mean genetic variance for the trait AYT for the genetic correlation 0.1 and 0.9 over the period of 25 breeding cycles for generation F5. The variance at F5 shows how much variance is available for subsequent selection. The change in genetic variance for AYT was different between genomic and phenotypic breeding

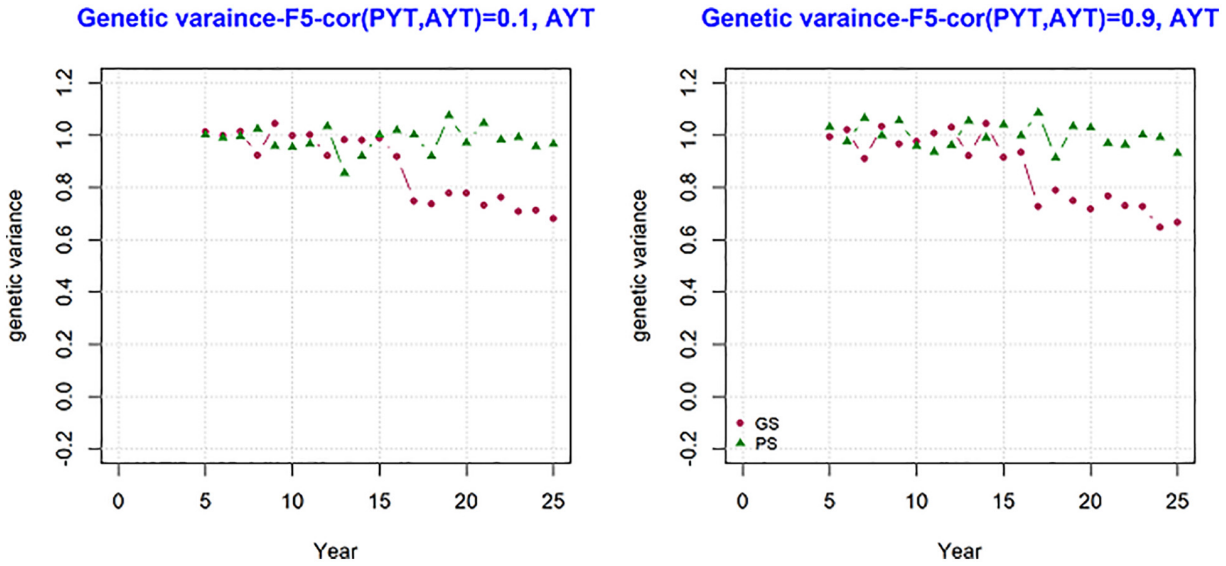


FIGURE 5 | The change in genetic variance over the period of 25 years measured at generation F5 for AYT for GS (Genomic selection) and PS (Phenotypic selection) for genetic correlation of 0.1 and 0.9 between PYT (preliminary yield trial) and AYT (advance yield trial). The first 8 years for GS is burn-in and similar to PS.

programs. In both PS and GS variance is 1 at year 5, however the change in the level of genetic variance is higher for GS than PS in the subsequent period of selection. This shows the loss in genetic variance is higher for GS than PS. By the end of the breeding program, PS has genetic variance about 40% more than GS when the correlation was 0.9.

At the end of the breeding program, the change in genetic variance for both genomic and phenotypic selection for grain yield measured for AYT for F5, F6, F7, and F8 at the end of the breeding cycle (25th year) together with the corresponding breeding values is shown in **Table 2**. The table shows mean genetic gain for PS and GS and it is significantly different between PS and GS. By the end of the breeding program GS has produced about 300% more gentic gain than PS. The loss in genetic variance was higher in genomic selection than phenotypic selection in all teseted scenarios.

Predicition Accuracy

Predicition accuracy of breeding value was computed as correlation between TBV and predicted breeding values. The mean predicition accuray of plot yield for PYT and AYT from year 9 to 25 for generation F6, F7, and F8 is presented in **Table 3**. In all the three generations GS had higher predicition accuracy than PS. For PS, accuracy was 0.269, 0.76, and 0.755 for F6, F7, and F8, respectively, while for GS the accuracy was 0.397 for F6, 0.992 for F7 and 0.966 F8. Among the generations, F7 and F8 has the highest accuracy for both PS and GS.

DISCUSSION

The current wheat breeding simulation study was done to test a hypothesis that genomic selection can increase genetic gain compared to phenotypic selection. For testing this hypothesis, a wheat breeding program was simulated for both genomic and phenotypic selection. The breeding program ran for 25 years. Within PS and GS breeding programs, different level of genetic correlation between PYT and AYT was tested. Our result confirmed the hypothesis that wheat-breeding program that used genomic information has tripled genetic gain compared to the conventional phenotypic selection.

Genetic Gain

Genomic selection breeding schemes has outperformed phenotypic selection for genetic gain. The increase in genetic gain that was brought by adding genomic information in conventional wheat breeding program was 3-fold. The main advantage of using genomic selection was increasing selection accuracy which inturn increased genetic gain, which was cosistent with the finding of Gaynor et al. (2017). Our result have shown that selection accuracy was higher in GS than PS. The use of genomic information to select parental lines has been shown to contribute to an increase in genetic gain through enhancing selection accuracy (Gaynor et al., 2017). In our simulation, lines that were selected based on their breeding values were stored to be used potentially as parents in the preceding cycles. However, parents were selected randomly from the stored line for the subsequent cycle. In crop breeding, selection of parents are one of the most important steps as it determines the direction of change in the genetic improvement (to bring genetic progress) (Akdemir et al., 2019).

In the current study, the increase in genetic gain was about 300% more when genomic information was used for selection. Previous study (Gaynor et al., 2017) has reported about 1.21 times more genetic gain when using genomic information in wheat breeding program. However, in our study GS had produced more genetic gain than reported by Gaynor et al. (2017). This huge difference was because genetic gain reported for

TABLE 2 | Genetic gain and genetic variance for AYT for generation F5, F6, F7, and F8 at the end of the breeding program (year 25) for phenotypic (PS) and genomic selection (GS) of each scenario.

		Genetic gain		Genetic variance	
Scenario	Generation	GS mean(se)	PS mean(se)	GS mean(se)	GS mean(se)
	F5	3.07(0.03)	0.74(0.07)	0.68(0.03)	0.97(0.04)
	F6	3.35(0.07)	0.61(0.05)	0.62(0.04)	0.96(0.05)
	F7	4.11(0.07)	0.84(0.06)	0.15(0.02)	0.94(0.07)
	F8	4.48(0.09)	1.00(0.18)	0.11(0.01)	0.69(0.14)
B0.1_Y0.3	F5	3.14(0.03)	0.74(0.06)	0.61(0.01)	0.96(0.03)
	F6	3.40(0.04)	0.84(0.10)	0.57(0.04)	1.01(0.05)
	F7	4.16(0.05)	1.04(0.10)	0.17(0.01)	0.99(0.05)
	F8	4.46(0.09)	1.05(0.08)	0.08(0.01)	0.77(0.16)
B0.1_Y0.5	F5	3.15(0.03)	0.87(0.06)	0.68(0.01)	0.98(0.06)
	F6	3.41(0.08)	0.96(0.09)	0.51(0.02)	0.95(0.05)
	F7	4.15(0.05)	1.15(0.09)	0.16(0.01)	0.96(0.10)
	F8	4.50(0.09)	1.28(0.10)	0.12(0.02)	0.82(0.17)
B0.1_Y0.7	F5	3.15(0.03)	0.89(0.07)	0.61(0.01)	0.93(0.02)
	F6	3.47(0.05)	0.88(0.05)	0.60(0.03)	1.01(0.04)
	F7	4.18(0.04)	1.10(0.10)	0.18(0.02)	0.88(0.08)
	F8	4.46(0.08)	1.30(0.12)	0.12(0.02)	0.75(0.14)
B0.1_Y0.9	F5	3.14(0.04)	1.03(0.08)	0.67(0.01)	0.93(0.04)
	F6	3.36(0.04)	1.08(0.07)	0.55(0.02)	0.97(0.03)
	F7	4.24(0.04)	1.44(0.10)	0.19(0.01)	0.99(0.07)
	F8	4.52(0.07)	1.19(0.15)	0.11(0.01)	0.56(0.09)

The five levels of correlation (Y0.1, Y0.3, Y0.5, Y0.7, and Y0.9) between PYT and AYT, and correlation of BVP with PYT and AYT (B0.1).

phenotypic selection in the current study was very much lower than what was reported by Gaynor et al. (2017) while for GS we did not observe huge difference in reported genetic gain with their studies.

Rate of genetic gain is used to compare an outcome from different breeding schemes that will help in designing a new breeding program (Rutkoski, 2019). Our study showed that the increase in genetic gain that came by adding genomic information at the preliminary selection stage could help a wheat breeder for practical decision making to switch to GS breeding program. This study can provide a guideline on how to apply genomic selection starting from the preliminary yield trial. Preliminary yield trial (PYT) is un-replicated and limited amount of seed is available for each selection candidate. However, this selection step strongly influences the subsequent advanced yield trial (AYT), which are commonly tested in multiple locations (Poland et al., 2018). In addition, GS at PYT allows for early elite parental selection for the next breeding cycles. In the current study, we assumed no genotype by environmental interactions in the advanced yield trial and this might overestimate the advantage of GS over PS.

Additive Genetic Variance

Selection causes changes in variances, allele frequencies, and LD relationships between markers and QTL (Bulmer, 1971; Muir, 2007). In our simulation study, the change in genetic varaince over time had a different trend for genomic selection and phenotypic selection. The loss in genetic variance over time was higher for genomic selection than for phenotypic selection. A similar result where genomic selection decreases genetic variance in wheat breeding prgram more than phenotypic selection was reported by Gaynor et al. (2017). By the end of the breeding program (year 25), about 66 % of genetic varaince was available for GS when the correlation between PYT and AYT is 0.9 while for PS it was about 90%. However, Gaynor et al. (2017) showed that about 33% genetic varaince was available for conventional genomic selection program. This difference could be related to the number of years of the breeding program, the current study is 25 years while in Gaynor et al. (2017) breeding programs runs for 40 years with 20 years of burn-in period. In addition, the number of lines for AYT is higher (150) in the current study while in Gaynor et al. (2017) which was 50 lines.

The change in genetic variance within a breeding cycle at the eary stage (year 9) and later stage (year 18) of the breeding program clearly showed differences in genetic varaince between

TABLE 3 | The mean (se) accuracy of F6, F7, and F8 across the entire breeding cycle for cor(PYT,AYT) = 0.9 scenario.

	Accuracy	
Generation	PS	GS
F6	0.269(< 0.001)	0.397(< 0.001)
F7	0.763(< 0.001)	0.992(< 0.001)
F8	0.755(< 0.001)	0.966(< 0.001)

GS and PS. Genetic varaince within a breeding cycle is produced through recombination until F5, after this stage varaince starts to decline faster because of selection. At the early stage of the breeding program (year 9), more genetic varaince was available for GS, however, as selection continues variance starts to decline for as seen in year 18. It also showed that the loss in genetic varaince is accelerated in GS than PS.

In our study, the change in genetic variance depending on the genetic correlation between PYT and AYT was not significantly different from each other. Additive genetic variance plays an important role in predicting the change in the population mean due to selection (Bernardo, 2010). That means if there are more variance present in a population, more genetic improvement in a population is possible. The current simulation assumes closed breeding program, which might exacerbate the loss of genetic variance. In real breeding program, breeders exchange breeding materials that will help introduce genetic variance to the on-going breeding program. Thus introducing breeding materials from outside can help in reducing the faster decline of genetic variance. Furthermore, the increased rate of decline in genetic variance when applying genomic selection calls for more measures to ensure sufficient genetic variance in future generations of parents. Including methods of optimum contribution selection should be investigated as means of controlling the loss of genetic variance in future generations of parents (Cowling et al., 2017; De Beukelaer et al., 2017; Gorjanc et al., 2018).

This study has shown considerable advantage of genomic selection over conventional phenotypic selection for winter wheat-breeding programs by drastically improving genetic gain. Generating more genetic gain implies additional revenue, although generating additional revenues may add additional costs such as genotyping. Accuracy in GS, dependes on the size of the training population which are required to be genotyped which may add extra cost. Thus, it is important to determine the optimum training population size without compromizing the selection accuracy. However, genotyping costs are decreasing and this makes GS a more efficient tool to bring genetic improvement in cereal breeding than PS (Robertsen et al., 2019).

This simulation was done following commercial wheat breeding schemes, including the number of families and single plants meaning the result easily can be translated into practical wheat-breeding programs. Genomic selection has increased response to selection with better estimates of breeding values and more accurate selection of future parents. We believe this study will help in desicion making process for a wheat breeder to switch to genomic selection. In addition to implementing genomic selection to increase genetic gain, accelerating the breeding cycle through speed breeding is also a promising approach to increase genetic gain that needs further investigations.

CONCLUSION

The current study shows that incorporation genomic information in conventional wheat-breeding program can increase genetic gain. Besides, using genomic selection for a conventional wheat-breeding program requires a minimal change in an already existing breeding program. The increase in genetic gain in GS was mainly due to an increase in selection accuracy. Individuals selected based on their GEBVs were stored to be used as a parents. Genetic variance was reduced higher in GS than in PS, indicating it is necessary to incorporated optimum contribution selection in order to conserve genetic variation in the breeding program.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

BBT performed the analyses and prepared the manuscript. All authors helped in designing breeding scenarios, interpreting the results, and reviewing the manuscript. All authors read and approved the final manuscript.

FUNDING

This project was funded by The Danish Council for Strategic Research, Center for Genomic Selection in Animals and Plants (GenSAP), grant no. 12-132452 and Green Development Program (GUDP) grant nos. 34009-13-0607 and 34009-15-0951.

REFERENCES

Akdemir, D., Beavis, W., Fritsche-Neto, R., Singh, A. K., and Isidro-Sánchez, J. (2019). Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity* 122, 672–683. doi: 10.1038/s41437-018-0147-1

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2015). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131

Bernardo, R. (2010). "Phenotypic and genetic variances," in *Breeding for Quantitative Traits in Plants*, ed. R. Bernardo (Woodbury: Stemma Press), 177–203.

Bernardo, R., and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690

Bulmer, M. G. (1971). The effect of selection on genetic variability. *Am. Nat.* 105, 201–211. doi: 10.1086/282718

Cericola, F., Jahoor, A., Orabi, J., Andersen, J., and Janss, L. (2017). Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in KASP. *PLoS One* 12:e0169606. doi: 10.1371/journal.pone.0169606

Cowling, W. A., Li, L., Siddique, K. H. M., Henryon, M., Berg, P., Banks, R. G., et al. (2017). Evolving gene banks: improving diverse populations of crop and exotic germplasm with optimal contribution selection. *J. Exp. Bot.* 68, 1927–1939. doi: 10.1093/jxb/erw406

Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G., and Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193, 347–365. doi: 10.1534/genetics.112.147983

De Beukelaer, H., Badke, Y., Fack, V., and De Meyer, G. (2017). Moving beyond managing realized genomic relationship in long-term genomic selection. *Genetics* 206, 1127–1138. doi: 10.1534/genetics.116.194449

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics,* 4th Edn. Prentice Hall.

Faux, A.-M., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., et al. (2016). AlphaSim: software for breeding program simulation. *Plant Genome* 9, 1–14. doi: 10.3835/plantgenome2016.02.0013

Fujino, K., Hirayama, Y., and Kaji, R. (2019). Marker-assisted selection in rice breeding programs in Hokkaido. *Breed. Sci.* 69, 383–392. doi: 10.1270/jsbbs.19062

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., et al. (2017). A two-part strategy for using genomic selection to develop inbred lines. *Crop Sci.* 57, 2372–2386. doi: 10.2135/cropsci2016.09.0742

Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330.

Godin, C., Mailliard, A., Chassin, A., Gardet, O., Giraud, A., Heumez, E., et al. (2012). A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *Eur. J. Agron.* 40, 28–38. doi: 10.1016/j.eja.2012.02.007

Gorjanc, G., Gaynor, R. C., and Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor. Appl. Genet.* 131, 1953–1966. doi: 10.1007/s00122-018-3125-3

Liu, H., Tessema, B. B., Jensen, J., Cericola, F., Andersen, J. R., and Sørensen, A. C. (2019). ADAM-Plant: a software for stochastic simulations of plant breeding from molecular to phenotypic level and from simple selection to complex speed breeding programs. *Front. Plant Sci.* 9:1926. doi: 10.3389/fpls.2018.01926

Madsen, P., and Jensen, J. (2013). *DMU: A User's Guide. A Package for Analysing Multivariate Mixed Models*. Version 6, Release 5.2. Denmark: Aarhus University.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Muir, W. M. (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124, 342–355. doi: 10.1111/j.1439-0388.2007.00700

Poland, J., Belamkar, V., Lorenz, A. J., El-basyoni, I., Baenziger, P. S., Jarquín, D., et al. (2018). Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3 Genes Genom. Genet.* 8, 2735–2747. doi: 10.1534/g3.118.200415

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Robertsen, C., Hjortshøj, R., and Janss, L. (2019). Genomic selection in cereal breeding. *Agronomy* 9:95. doi: 10.3390/agronomy9020095

Rutkoski, J. E. (2019). Estimation of realized rates of genetic gain and indicators for breeding program assessment. *Crop Sci.* 59, 981–993. doi: 10.2135/cropsci2018.09.0537

Wensch-Dorendorf, M., Yin, T., Swalve, H. H., and König, S. (2011). Optimal strategies for the use of genomic selection in dairy cattle breeding programs. *J. Dairy Sci.* 94, 4140–4151. doi: 10.3168/jds.2010-4101

Conflict of Interest: JA was employed by company Nordic seed.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Tessema, Liu, Sørensen, Andersen and Jensen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 December 2020
doi: 10.3389/fgene.2020.586687

Genome-Wide Association Mapping and Genomic Prediction of Anther Extrusion in CIMMYT Hybrid Wheat Breeding Program via Modeling Pedigree, Genomic Relationship, and Interaction With the Environment

Anil Adhikari[1,2], Bhoja Raj Basnet[3], Jose Crossa[3], Susanne Dreisigacker[3], Fatima Camarillo[3], Pradeep Kumar Bhati[4], Diego Jarquin[5], Yann Manes[6] and Amir M. H. Ibrahim[1]*

[1] Texas A&M University, College Station, TX, United States, [2] Department of Horticulture, University of Wisconsin, Madison, WI, United States, [3] International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [4] Borlaug Institute for South Asia (BISA), Ludhiana, India, [5] Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE, United States, [6] Syngenta Seeds, Saint-Sauveur, France

OPEN ACCESS

Edited by:
Arron Carter,
Washington State University,
United States

Reviewed by:
Dev Paudel,
University of Florida, United States
Sreepriya Pramod,
Altria, United States

***Correspondence:**
Bhoja Raj Basnet
b.r.basnet@cgiar.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 23 July 2020
Accepted: 13 October 2020
Published: 08 December 2020

Citation:
Adhikari A, Basnet BR, Crossa J, Dreisigacker S, Camarillo F, Bhati PK, Jarquin D, Manes Y and Ibrahim AMH (2020) Genome-Wide Association Mapping and Genomic Prediction of Anther Extrusion in CIMMYT Hybrid Wheat Breeding Program via Modeling Pedigree, Genomic Relationship, and Interaction With the Environment. Front. Genet. 11:586687. doi: 10.3389/fgene.2020.586687

Anther extrusion (AE) is the most important male floral trait for hybrid wheat seed production. AE is a complex quantitative trait that is difficult to phenotype reliably in field experiments not only due to high genotype-by-environment effects but also due to the short expression window in the field condition. In this study, we conducted a genome-wide association scan (GWAS) and explored the possibility of applying genomic prediction (GP) for AE in the CIMMYT hybrid wheat breeding program. An elite set of male lines ($n = 603$) were phenotype for anther count (AC) and anther visual score (VS) across three field experiments in 2017–2019 and genotyped with the 20K Infinitum is elect SNP array. GWAS produced five marker trait associations with small effects. For GP, the main effects of lines (L), environment (E), genomic (G) and pedigree relationships (A), and their interaction effects with environments were used to develop seven statistical models of incremental complexity. The base model used only L and E, whereas the most complex model included L, E, G, A, and G × E and A × E. These models were evaluated in three cross-validation scenarios (CV0, CV1, and CV2). In cross-validation CV0, data from two environments were used to predict an untested environment; in random cross-validation CV1, the test set was never evaluated in any environment; and in CV2, the genotypes in the test set were evaluated in only a subset of environments. The prediction accuracies ranged from −0.03 to 0.74 for AC and −0.01 to 0.54 for VS across different models and CV schemes. For both traits, the highest prediction accuracies with low variance were observed in CV2, and inclusion of the interaction effects increased prediction accuracy for AC only. In CV0, the prediction accuracy was 0.73 and 0.45 for AC and VS, respectively, indicating the high reliability of across

environment prediction. Genomic prediction appears to be a very reliable tool for AE in hybrid wheat breeding. Moreover, high prediction accuracy in CV0 demonstrates the possibility of implementing genomic selection across breeding cycles in related germplasm, aiding the rapid breeding cycle.

Keywords: GWAS, anther extrusion, floral traits, hybrid wheat, genome-wide prediction

INTRODUCTION

Hybrid wheat offers great promise in terms of higher grain yield and stability across a wide range of wheat-producing environments globally (Gowda et al., 2010, 2012; Mühleisen et al., 2014; Zhao et al., 2015; Basnet et al., 2019; Adhikari et al., 2020a; Easterly et al., 2020). Hybrid wheat has gotten attention in private and public sector breeding programs since the 1950s, but it has not led to any considerable commercial success (Virmani and Edwards, 1983; Longin et al., 2012; Adhikari et al., 2020a). There has been a growing interest in hybrid wheat research again in the last decade in North America (Dreisigacker et al., 2005; Basnet et al., 2019; Adhikari et al., 2020a,b; Easterly et al., 2020) and Europe (Gowda et al., 2012; Longin et al., 2012; Zhao et al., 2015). However, hybrid wheat varieties are currently commercially marketed only in Europe and some parts of India and China, and acreages are fairly low in all three areas (Longin et al., 2012; Gupta et al., 2019).

The most vital limitation for the commercial success of hybrid wheat has always been the complexity and cost of hybrid seed production even though chemical-based sterility, cytoplasmic male sterility, and genetic male sterility systems with varying levels of efficiency are available (Virmani and Edwards, 1983; Longin et al., 2012). Irrespective of the method of hybrid seed production, one of the major factors that determines hybrid seed set is the outcrossing ability, which is determined by pollen load released by male parents upon flowering outside the floral structure (De Vries, 1971; Whitford et al., 2013) and opening of flowers in the female parents (Longin et al., 2014). Wheat flowers are cleistogamous, and extrusion of anthers outside the floral structure is a highly correlated proxy of pollen mass release that determines hybrid seed production (Whitford et al., 2013; Langer et al., 2014; Boeven et al., 2016, 2018). Hence, inheritance of anther extrusion (AE) in the context of hybrid seed production plays a key role and has, therefore, been extensively studied in Europe and North America (Langer et al., 2014; Boeven et al., 2016, 2018; Muqaddasi et al., 2017a,c,b, 2019).

AE was previously thought to be inherited in an oligogenic manner (Sage and De Isturiz, 1974). However, recent studies on the inheritance of AE with the use of high-density molecular markers have found the trait to be complex and quantitative (Boeven et al., 2016). Several groups have studied the inheritance of AE in the context of Fusarium head blight (caused by *Fusarium* species, including *Fusarium graminearum*, *F. culmorum*, *F. avenaceum*, *F. poae*, and *Microdochium nivale*) resistance and found similar results (Skinnes et al., 2010; Lu et al., 2013; Buerstmayr and Buerstmayr, 2015; Steiner et al., 2019). Genome-wide association studies (GWAS) and bi-parental mapping have identified dwarfing genes/alleles, also referred to as "reduced height loci" (*Rht*), present in modern wheat germplasm as the major loci or co-localized with major loci governing AE and pollen mass in wheat (Lu et al., 2013; Boeven et al., 2016; He et al., 2016). The *Rht* genes (*Rht-B1* and *Rh-tD1*) that cause plant height reduction also shorten the anther filaments and negatively impact AE (Boeven et al., 2016). Another height reduction loci *Rht24*, which does not reduce AE, has been suggested as an alternative to the *Rht1* loci for use in male lines in hybrid breeding (Würschum et al., 2018). Other than the *Rht* genes, previous studies reported the presence of several minor effect loci positively and negatively associated with AE in European winter wheat (Muqaddasi et al., 2017a,b) and CIMMYT spring wheat (Muqaddasi et al., 2019). Despite the identification of marker-trait associations and QTL, their application in hybrid wheat breeding via marker-assisted selection (MAS) for extending AE has limited scope since they are of very modest effects. Moreover, the *Rht* alleles have a very important role in height reduction of the wheat plant that is paramount for maintaining grain yield via prevention of lodging. *Rht* alleles have been widely deployed since the green revolution and are ubiquitous in elite CIMMYT germplasm (Ogihara et al., 2013; Aisawi et al., 2015; Basnet et al., 2019). These genes cannot be excluded for the sole purpose of increasing AE. In this context, genome-wide prediction or genomic selection (GS) appears to be the best strategy to breed for AE by exploiting the cumulative effect of many effect loci scattered throughout the genome (Meuwissen et al., 2001).

GS via whole-genome regression methods uses the information from thousands of molecular markers to capture not only major-effect genes but also the contribution of genomic regions with small effects (Meuwissen et al., 2001). GS has been utilized extensively in animal breeding and plant breeding to predict traits with complex genetic architecture using the information from molecular markers and pedigree information (Hayes et al., 2009; Crossa et al., 2017). In the case of multi-environment trials (METs) in plant breeding, GS models did not explicitly model G × E information since the phenotypic data from METs were analyzed separately to derive single phenotypic estimates, and they were used as single trait in a genomic estimated best linear unbiased predictor (G-BLUP) model (Burgueño et al., 2012). G-BLUP models have been extended for a multi-environment setting by Burgueño et al. (2012), where genetic correlations are used to explicitly model G × E. Jarquín et al. (2014) extended the G × E model to include environmental covariates both as main effects and interaction effects with genotypes and locations in a reaction norm model. The reaction norm model has been used extensively to predict complex traits with multi-environment data in wheat and cotton and was found to reveal higher prediction accuracies than single-trait G-BLUP and multi-trait G-BLUP models (Jarquín et al., 2014;

Pérez-Rodríguez et al., 2015; Sukumaran et al., 2017, 2018). GS has been used to predict AE using single-trait G-BLUP (Boeven et al., 2016; Muqaddasi et al., 2017c). However, G × E and reaction norm models have not been tested thus far, despite AE showing high levels of G × E.

This study aims to (i) explore the wheat genome for major effect QTL associated with AE via GWAS and (ii) apply reaction norm G × E models to predict AE in a multi-environment setting with the goal of driving genetic gain for AE in the CIMMYT hybrid wheat breeding program.

MATERIALS AND METHODS

Plant Materials and Field Experiments

The study consisted of 603 advanced parental lines from the CIMMYT hybrid wheat breeding program. The lines were planted in 2-m-long, double-row linear plots, with 20-cm inter-row spacing at El Batan, Mexico (20.83° N, 100.83° W) in 2017 and 2019, and at Obregon, Mexico (27.48° N, 109.93° W) in 2018 growing cycles. The trials within a location were unreplicated, and plants/spikes per plant were used as biological replicates (i.e., two to three individual plants per genotype and one to three spikes per plant). In each experiment, four or five random spikes from different plants in each plot were tagged prior to flowering, and two male floral traits, as they relate to AE, AE visual scores (VS), and extruded anther count (AC), were taken from the plots at flowering and post-flowering stages, respectively. Trapped anthers in two lateral florets (first and second florets) from six to eight middle spikelets were counted in each randomly tagged spike, and then the deduced AC was expressed as a percentage by using the formula by Boeven et al. (2016):

$$\text{Anther extrusion (\%)} = \frac{\text{Total extruded anthers}}{\text{Total number of anthers}} \times 100\% \quad (1)$$

For VS, a score of 1 to 10 was assigned to each genotype during flowering based on visual observations, where 1 indicates no anther extrusion and 10 indicates maximum anther extrusion. AC data were collected from all three environments, whereas VS was collected from only 2018 and 2019 experiments. In 2017, we missed the critical time during flowering to collect reliable VS data.

Statistical Analysis of Phenotypic Data

Phenotypic data of AC from each trial were analyzed separately using software package META-R (Alvarado et al., 2015). Best linear unbiased estimates (BLUEs) were calculated for each genotype, considering the effects of genotypes as fixed and the effects of environments as random. In addition to the BLUEs, variance components were calculated considering all the factors as random. Variance components were estimated for the combined analysis of AC data from all three environments to get an estimate of G × E. For VS, since there were no repeated measurements, the environments were considered replications for the purpose of variance components estimation. The genomic prediction models were run using BLUEs from each experiment for AC and raw phenotypic data for VS.

Broad-sense heritability (H^2) within environment was estimated as:

$$H^2 = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_e/r} \quad (2)$$

H^2 across environments was estimated as:

$$H^2 = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_{GE}/l + \sigma^2_e/lr} \quad (3)$$

where σ^2_G is the variance due to genotype, σ^2_{GE} is the variance due to genotype × environment, σ^2_e is the error variance, l is the number of environments, and r is the number of replications using multi-environment trial analysis.

DNA Extraction and Genotyping

Genomic DNA was extracted from freeze-dried leaves collected from five individual plants per line using a modified CTAB (cetyltrimethylammonium bromide) method described in CIMMYT laboratory protocols (Dreisigacker et al., 2016) and quantified using a NanoDrop 8000 spectrophotometer V 2.1.0.

The population was genotyped with the 20K Infinium is elect SNP array by TraitGenetics (Gatersleben, Germany). The marker dataset was filtered for polymorphism and minor allele frequency (<5%) and >50% missing data were removed. In addition to these filtering steps, markers with known genetic positions in the genetic map developed by Wang et al. (2014) were extracted for use in GWAS.

Genome-Wide Association Study

BLUEs from each individual environment were used for GWAS using the mixed linear model (MLM) model implemented in the Genome Association and Prediction Integrated Tool (GAPIT) (Tang et al., 2016). The population structure was assessed via principal component analysis, and the first three components were used as covariates in the population structure (Q) defined by the kinship (K) (Q + K) model. Linkage disequilibrium (LD) decay was assessed by plotting pairwise LD between marker pairs and their genetic distance. Bonferroni correction for multiple testing implemented in GAPIT was used to identify significant associations, which corresponded to a $-\log_{10}(P)$ value of ≥ 5.

Statistical Models for Genomic Prediction

We used the conventional Genomic Best Linear Unbiased Prediction (GBLUP) extended by the genotype-by-environment interaction term using molecular markers (G × E) and pedigree information (A × E) via the reaction norm model (Jarquín et al., 2014).

Baseline Line Model

Consider that the trait performance (y_{ij}) of the ith line observed in the jth environment can be described as the sum of an overall mean common to all genotypes in all environments μ, plus random deviations as follows:

$$y_{ij} = \mu + E_j + L_i + LE_{ij} + \varepsilon_{ij} \quad (4)$$

where L_i is the random effect of the ith line, E_j is the random effect of the jth environment, LE_{ij} is the interaction between the ith line and the jth environment, and e_{ij} is the random error term accounting for non-explained variability. The effects are assumed to be independent and identically distributed outcomes following normal densities, such that $E_i \overset{iid}{\sim} N(0, \sigma^2_E)$, $L_i \overset{iid}{\sim} N(0, \sigma^2_L)$, and $e_{ij} \overset{iid}{\sim} N(0, \sigma^2_e)$, while the interaction term from properties of the multivariate density is distributed as follows: $LE_{ij} \overset{iid}{\sim} N(0, Z_E Z'_E \# Z_L Z'_L \sigma^2_{LE})$ where, σ^2_E, σ^2_L, σ^2_{LE}, and σ^2_e are the associated variance components, and Z_E and Z_L are the incidence matrices that connect phenotypes with environments and lines, and # represents the Hadamar product (cell-by-cell product between two matrices). Z'_E and Z'_L are the transpose of the respective incidence matrices.

In the model above, the random effect of the line (L_i) can be replaced by g_i, which is an approximation of the genetic value of the ith line from the genomic relationship matrix. Also the effects of the line (L_i) can be replaced by a_i, which is the additive effect obtained from the pedigree information. In the models described below, we used either g_i or a_i or both g_i and a_i as well as their interactions with environment E_j (gE_{ij}, or aE_{ij}).

G × E Models for AC and VS Measured in Environments

We applied a sequence of reaction norm models similar to that used by Jarquín et al. (2014), with genomic-based relationship matrices, and by Pérez-Rodriguez et al. (2015), with pedigree-based relationship matrices. Model 1 included only the main effects of environment (E) and lines (L), whereas model 2 added genomic (G) genomic information to model 1. Model 3 included all three main effects of L, G, and E, and the genomic × environment interactions (G × E). Model 4 included main effects of environment (E), lines (L), and pedigree (A), whereas model 5 added the pedigree × environment (A × E) to the main effect terms of model 4. Model 6 included the main effects of environment (E), lines (L), genomic (G), and pedigree (A). Finally, we fitted model 7, which included all main effects and the two interactions G × E and A × E. A description of the seven models considered in this study is given below.

Main Effect Model 1

This simple main effect model considers the response of the ith wheat male line and the jth environment (y_{ij}) as a function of a random effect model that accounts for only the effect of the environment (E_i) and the accession (L_j), plus a residual (ε_{ij}):

$$y_{ij} = \mu + E_i + L_j + \varepsilon_{ij} \tag{5}$$

where μ is an intercept, and the random terms remaining are described as in the baseline model. The main effect of environment (E_i) models the environment information via the incidence matrix of genotypes (Z_L) observed in different environments. In this model, the effects of the lines are regarded as independent; therefore, there is no borrowing of information between untested and tested landrace accessions.

Main Effect Model 2

The next main effect model adds in Eq. 5, the random effect of genomic relationship g_i, which is an approximation of the true genetic value of the ith male wheat line. This approximation is given by the jointly regression on marker covariates $g_i = \sum_{m=1}^{p} x_{im} b_m$, where x_{im} is the genotype of the ith line at the mth marker, and b_m is the corresponding effect with the assumption that $b_m \overset{iid}{\sim} N(0, \sigma^2_b)$ (m = 1, . . ., p) and σ^2_b is the variance of the marker effects. The vector $g = (g_1, \ldots, g_I)'$ contains the genomic values of all the lines, and it is assumed to follow a multivariate normal density with zero mean and covariance matrix $Cov(g) = G\sigma^2_g$, where **G** is the genomic relationship matrix that describes the genomic similarities between pairs of lines and, σ^2_g which is proportional to σ^2_b ($\sigma^2_g \propto \sigma^2_b$), the genomic variance. Therefore, main effect model 2 becomes

$$y_{ij} = \mu + E_j + L_i + g_i + \varepsilon_{ij} \tag{6}$$

where the vector of random effects is assumed $g \sim N(0, G\sigma^2_g)$ to be, and the other random effects remain as described The random effects $g = (g_1, \ldots, g_J)'$ are correlated such that model 2 allows borrowing of information across L_i tested and untested lines. The genomic matrix **G** given by $G \propto (XX')/(2\sum_{m=1}^{p} 2p_m(1 - p_m))$, where p_m is the estimated frequency of the allele whose number of copies at the ith accession is counted in x_{im}. Centering (i.e., subtracting $2p_m$ from the genotype codes) and standardization (i.e., dividing by $\sum_{m=1}^{p} 2p_m(1 - p_m)$) allows interpreting $\sigma^2_g = \sigma^2_b \sum_{m=1}^{p} 2p_m(1 - p_m)$ as a genomic variance. This model does not allow specific genomic effects for each environmental condition but rather a common effect for same lines across environments. Thus, the interaction between markers and environments is introduced in the next model.

Main Effect and Interaction Model 3

This model is obtained by extending the main effect model 2 (Eq. 6) to include interaction effects (gE_{ij}) between each marker SNP and each environment. This model can be written as

$$y_{ij} = \mu + E_j + L_i + g_i + gE_{ij} + \varepsilon_{ij} \tag{7}$$

where E_j, L_i, and g_i have already been defined, $gE \sim N(0, (Z_L G Z'_L) \# (Z_E Z'_E)\sigma^2_{Eg})$ is the interaction of the genome with environment, with σ^2_{gE} as the variance component of gE, and the other model terms are as defined previously.

Main Effect and Interaction Model 4

Model 4 is similar to model 2 (Eq. 6), but instead of including the random effect of genomic g_i, it includes the random effect accounted for by the pedigree a_i. This model adds the random effect that incorporates pedigree information by means of the additive relationship matrix (***A***) to model 1 (Eq. 5),

$$y_{ij} = \mu + E_j + L_i + a_i + \varepsilon_{ij} \tag{8}$$

where a_i is a random additive effect of the line, which in this case accounts for pedigree relationships, where $a = (a, \ldots, a_I)'$ contains the pedigree values of all the lines and is assumed to follow a multivariate normal density with zero mean and

covariance matrix $Cov(a) = A\sigma_a^2$, where $\boldsymbol{A}$ is the additive relationship matrix, and σ_a^2 is the additive genetic variance. The random effects are correlated such that model 4 allows borrowing between tested and untested lines based on the numerical relationship matrix ($\boldsymbol{A}$). Similarly, this model does not allow specific responses to each environment but instead common effects across environments. Thus, the interaction between lines and environments is introduced in the following model via pedigree information instead of marker data.

Main Effect and Interaction Model 5

This model is obtained by extending model 4 (Eq. 8) to include interaction effects (aE_{ij}). Thus, where,E_j, L_i, and g_i have already been defined, $aE \sim N\left(0, \left(Z_L A Z_L'\right) \# \left(Z_E Z_E'\right) \sigma_{aE}^2\right)$ is the interaction of the genome with environment, with d σ_{aE}^2 as the variance component of aE:

$$y_{ij} = \mu + E_j + L_i + a_i + aE_{ij} + \varepsilon_{ij} \tag{9}$$

Main Effect and Interaction Model 6

Model 6 is similar to model 4 but adds the genomic relationship g_i:

$$y_{ij} = \mu + E_j + L_i + a_i + g_i + \varepsilon_{ij} \tag{10}$$

Main Effect and Interaction Model 7

This is the complete model with all main effects and interactions:

$$y_{ij} = \mu + E_j + L_j + a_i + g_i + aE_{ij} + gE_{ij} + \varepsilon_{ij} \tag{11}$$

where the terms are already defined.

Assessing Model Prediction Accuracy by Random Cross-Validation

The described models were fitted in various validation settings to estimate prediction accuracy within an environment (i.e., despite how training and testing set were configured, the correlation between predicted and observed values was computed within environments). For both traits measured, three different validation schemes were studied. We repeated the random cross-validations from Burgueño et al. (2012) and Jarquín et al. (2014) and considered three prediction problems: (1) (CV1) prediction of 20% of wheat lines that have not been evaluated in any environment; (2) (CV2) prediction in incomplete field trials i.e., prediction of performance of lines that have been evaluated in some environments but not in others; and (3) prediction of performance of all lines in an untested environment, using performance data of those lines from correlated environments. CV1 was obtained by assigning accessions to folds; hence, when the phenotype of an accession is predicted, the corresponding training set contains no record of this accession. CV2 was obtained by assigning individual records of each accession to folds; hence, when one is predicting the ith line, there are records for the same accession that were part of the training set but observed in a different environment. In both prediction problems, a fivefold cross-validation was performed, where 80% of the accessions formed the training set and 20% of the accessions comprised the testing set for each partition. The assignation of training and testing sets was repeated 20 times ($5 \times 2 = 100$ random partitions) for each one of these cross-validation schemes (CV1 and CV2).

We also evaluated cross-validation CV0, where all lines in one environment were fully predicted by the other environments.

Data Repository

The data repository can be found at http://hdl.handle.net/11529/10548495.

RESULTS

Phenotypic Variation and Correlation Between Traits

The AC in 2017 ranged from 14.3 to 95.8 with an average of 47.2 and a standard deviation (SD) of 18.1 (**Figure 1** and **Table 1**). AC in 2018 ranged from 15.0 to 99.2 with an average of 68.2 and SD of 17.4. The VS in 2018 ranged from 3 to 7 with an average of 5.77 and SD of 0.75. In the 2019 experiment, AC data ranged from 7.61 to 100.00 with an average of 67.58 and SD of 19.44 whereas the VS ranged from 1 to 8 with an average of 4.94 and SD of 1.47.

The genotypic variance was significant for both traits in all three trials (**Table 1**). The broad-sense heritability estimates for AC ranged from 0.78 to 0.93 across the three environments, whereas the broad-sense heritability for combined VS was 0.32.

The BLUEs of AC were significantly correlated across the three experiments (0.6 –0.67, $p \leq 0.01$) (**Supplementary Figure 1**). The visual scores across the two environments (2018 and 2019) were also significantly correlated (0.44, $p \leq 0.01$). The VS data were also highly correlated with AC data from all three trials (**Supplementary Figure 1**).

Markers Retained After Quality Control

After filtering for polymorphism, minor allele frequency, and missing data, 10,534 markers were retained. The whole marker dataset was used to create a genomic relationship matrix used for GP. For GWAS, the marker dataset was additionally filtered for presence of known genetic positions in the consensus map by Wang et al. (2014). The number of markers retained for GWAS was 7,649.

Population Structure and Linkage Disequilibrium

Population structure was assessed via principal component analysis. Population structure observed was not very strong since the first three principal components (PCs) only explained 16% of the cumulative variance (**Figure 2**). Linkage disequilibrium (LD) was assessed by calculating pairwise LD decay over genetic distance (Figure S2). Considering $r^2 = 0.2$ to be the extent of average intrachromosomal LD, in this population, LD calculated on a sliding window of 100 adjacent markers showed the LD blocks extended up to 25 cM (**Figure 2**).

Marker Trait Associations From GWAS

Five marker-trait associations (MTAs) were observed in the GWAS across two environments (2017 and 2019). Of the five

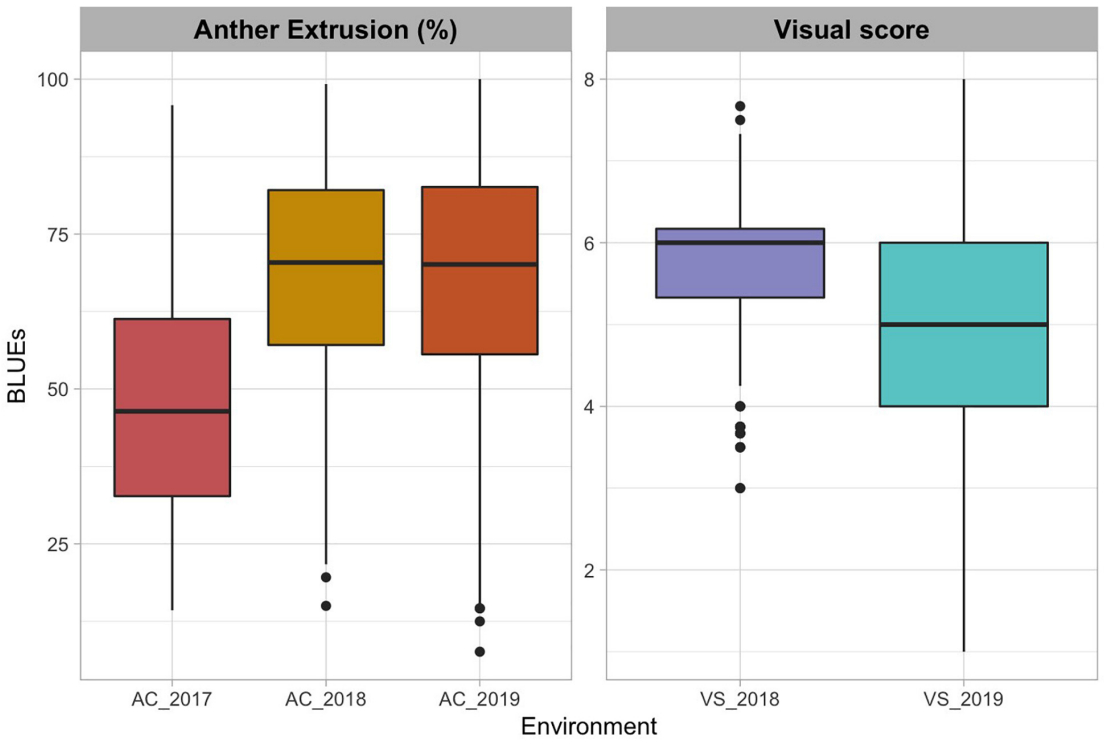


FIGURE 1 | Boxplots of best linear unbiased estimates of anther count (AC) expressed in percentage and visual score (VS) across the three environments (El Batan 2017, Obregon 2018 and El Batan 2019). VS data were collected from only two environments (2018 Obregon and 2019 El Batan). Visual score scale ranges from 0 (0% extruded anthers) to 10 (100% extruded anthers), assessed visually during flowering. Visual score data were not collected in 2017.

MTAs in 2017 and 2019 for VS and AC, two were located in chromosome 1B, two in chromosome 5A, and one in chromosome 5B (**Table 2**). For AC, three MTAs were observed, two of which were the same and were linked to marker Tdurum_contig75938_1546 (60.62 cM, 62.49 Mb) in 2017 and 2019. This MTA on chromosome 1B had a positive effect on AC, with an allele substitution effect of 5.68–6.84. The two MTAs on chromosome 5A (one each for AC and VS) had a negative effect on AE. SNP marker Ku_c69633_1873 (26.51 cM, 98.94 Mb) decreased the AC by 6.16%, whereas BS00065313_51 decreased the VS by 0.3. The MTA detected on chromosome 5B by the significant SNP marker RAC875_rep_c104919_902 increased the VS by 0.39. For three out of the four markers significantly associated with AC and VS, the favorable allele has a higher frequency in the elite male population we studied (**Table 2**).

TABLE 1 | Variance components and heritabilities for anther count (%) and visual score across three environments (2017, 2018, and 2019).

Trait	Anther count (%)			Visual score$^{\varphi}$	
Env	2017	2018	2019	Combined	2018–2019
Genotype	254.34**	281.92**	330.91**	211.09**	0.29**
Environment	–	–	–	182.70**	–
G × E	–	–	–	80.45**	–
Error	290.82	101.51	191.70	185.03	1.25
Replications	4	5	4	5	2.00
H^2	0.78	0.93	0.87	0.83	0.32
Grand mean	47.16	68.21	67.57	57.31	5.36

*$^{\varphi}$Visual score data did not have replications within one environment, and data from environments were used as replication for estimating variance components. **$p < 0.001$.*

Genome-Wide Prediction

We used seven models of increasing complexity for the genome-wide predictions. The prediction accuracy was assessed using three cross-validation scenarios.

Cross-Validation Scenario 1 (CV1)

For CV1, phenotypes of lines that have never been evaluated in the field were predicted using line information, environment information, genomic information, pedigree information, and interaction terms. Prediction accuracies of model 1 (E + L) were negative for both traits under CV1 (**Table 3**). Upon the addition of genomic information (model 2, E + L + G), the prediction accuracy increased and was in the range of 0.42–0.44 across environments for AC and 0.33–0.46 for VS (**Table 3**). Inclusion of pedigree information without genomic relationship information decreased the prediction accuracy for both traits. Model 4 (E + L + A) had prediction accuracies in the range of 0.31–0.34 for AC and 0.22–0.34 for VS, which is lower than the prediction accuracies obtained by model 2 for both traits (**Table 3**). When pedigree information was added in the

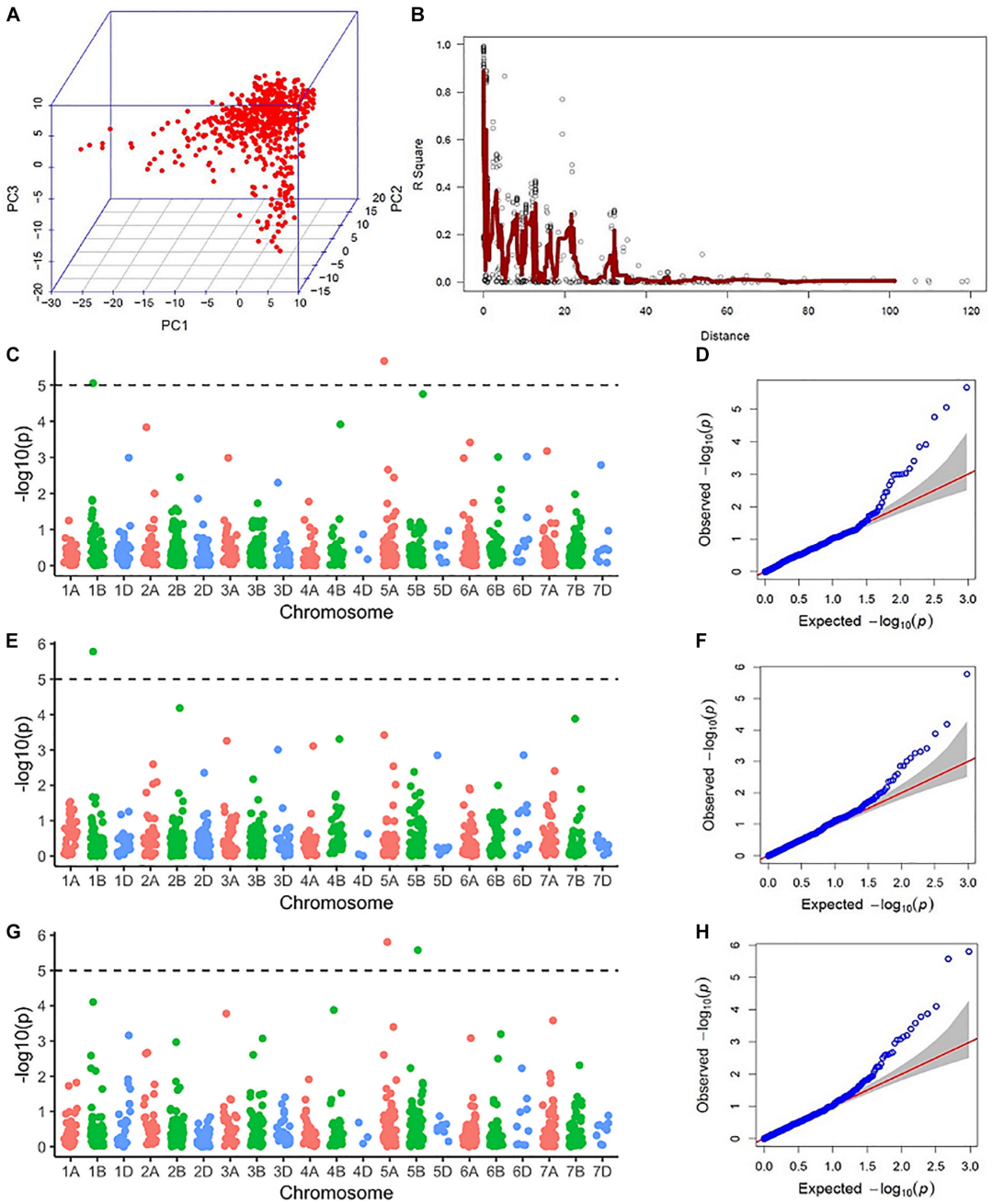


FIGURE 2 | Results from the genome-wide association scan: **(A)** Three-dimensional scatterplot showing the relationship between the first three principal components (PC1, PC2, and PC3) from molecular marker data. **(B)** A pairwise linkage disequilibrium (LD) decay plot with pairwise marker LD in a sliding window of 100 adjacent markers in the y-axis and genetic distance in centimorgans from the genetic map by Wang et al. (2014) in the *x*-axis. The red line indicates a moving average of r^2 values for 10 adjacent markers. **(C)** A Manhattan plot showing $-\log_{10}$(P) values of marker trait association (MTA) for anther count (AC) in El Batan (2017) across the genome. The dotted line represents the Bonferroni significance threshold for MTA. **(D)** A quantile-quantile (QQ) plot showing the distribution of expected vs actual −log10(P) values of GWAS using AC from El Batan (2017). **(E)** A Manhattan plot showing $-\log_{10}$(P) values of MTA for anther count (AC) in El Batan (2019) across the genome. **(F)** A QQ plot for GWAS using AC data from El Batan (2019). **(G)** A Manhattan plot showing $-\log_{10}$(P) values of MTA for visual score (VS) in El Batan (2019) across the genome. **(H)** A QQ plot for GWAS using VS data from El Batan (2019).

presence of genomic information, it improved the prediction accuracies. The prediction accuracies for AC were higher for model 6 (E + L + G + A) than those for model 4. In the case of VS, the inclusion of pedigree information along with genomic relationship slightly increased the prediction accuracies in 2018 (**Table 3**).

Models with interaction terms generally have slightly higher accuracies as compared to models with main effects only. For

TABLE 2 | Marker trait associations in the genome-wide association scan (GWAS) for anther count percentage (AC) and visual score (VS) from a field experiment evaluating anther extrusion across three environments (El Batan 2017, Obregon 2018 and El Batan 2019).

Env	Trait	SNP	CHR[1]	Genetic position (cM)	Physical position[2] (Mb)	$-\log_{10}(P)$	$(R^2)^3$ (%)	Allele[4]	MAF[5]	Allele effects
2017	AC	Ku_c69633_1873	5A	26.51	98.94	5.67	3.2	**A**/G	0.08	−6.16
2017	AC	Tdurum_contig75938_1546	1B	60.62	62.49	5.06	4.6	**A**/G	0.06	5.68
2019	AC	Tdurum_contig75938_1546	1B	60.62	62.49	5.78	3.7	**A**/G	0.06	6.84
2019	VS	BS00065313_51	5A	62.72	492.79	5.81	2.5	**G**/A	0.31	−0.30
2019	VS	RAC875_rep_c104919_902	5B	117.84	594.60	5.58	2.7	A/**C**	0.17	0.39

A consensus genetic map from Wang et al. (2014) and SNP data from 20K Infinitum SNP assay were used for GWAS.
[1]*Chromosome.*
[2]*Physical position of SNP marker from reference genome of Chinese Spring (Refseq_v1).*
[3]*Phenotypic variance explained by the SNP in marker trait association. It was calculated by subtracting R^2 of model without SNP from R^2 of model with SNP.*
[4]*Favorable allele is highlighted.*
[5]*Minor allele frequency (>5%) (Major allele/minor allele).*

example, for both traits model 3 (E + L + G + G × E) had slightly higher prediction accuracy compared to model 2 (E + L + G), and model 5 (E + L + A + A × E) had slightly higher prediction accuracy compared to model 4 (E + L + A). Model 7 (E + L + G + A + G × E + A × E), which is the most complex model, had comparable prediction accuracies for both traits with model 3 (E + L + G + G × E). Compared to model 3, model 7 had higher prediction accuracy only in 2018 for AC and 2019 for VS (**Table 3**). In most cases, model 3 (E + L + G + G × E) had the highest prediction accuracies among the seven models tested in CV1 for both traits.

Cross-Validation Scenario 2 (CV2)

CV2 represents the case of incomplete field trials, where some lines are tested in one environment but are missing in other environments. The phenotypic record of one or more environments is used in conjunction with genomic, pedigree, line, and environment information to predict the missing phenotypic record of lines. Prediction accuracies were higher in CV2 for all seven models across two traits compared to CV1 (**Table 3**). In CV2, model 1, which had negative prediction accuracy, had comparable prediction accuracies with model 3. Very nominal increments in prediction accuracies were observed with increasing model complexities across both traits. Some variation was observed in prediction across environments. For example, prediction accuracies for AC were higher in 2018 and 2019 compared to 2017. Similarly, prediction accuracies for VS were higher in 2018 compared to 2019. As in CV1, the models predicted AC better than VS in CV2. In addition to higher prediction accuracies, the standard deviation (SD) of prediction accuracies was much smaller in CV2 compared to CV1 (**Table 3**).

Predicting an Untested Environment (CV0)

CV0 is a prediction scenario, where a full dataset from a correlated environment is used to predict the performance of lines in an untested environment. Under CV0, all seven prediction models performed very well across all three environments and two traits (**Table 4**). The prediction accuracies for AC ranged from 0.68 to 0.69 in 2017, 0.70 to 0.71 in 2018, and 0.71 to 0.73 in 2019 (**Table 4**). The prediction accuracies for VS ranged from 0.43 to 0.47 in 2018 and 0.43 to 0.48 in 2019 (**Table 4**). The differences that were apparent between model 2 and model 4 via the inclusion of pedigree vs. genomic relationship in CV1 were not apparent in CV0. Similarly, the inclusion of interaction effects also did not make as much difference as it did in CV1.

DISCUSSION

The success of hybrid wheat breeding depends on reduced costs for hybrid seed production and grain yield heterosis. The presence of heterosis of grain yield in hybrid wheat has already been established over several decades, while newer studies have suggested, in addition, that the development of heterotic pools could increase the level of heterosis (Longin et al., 2013; Zhao et al., 2015; Rembe et al., 2019). Reducing the cost of hybrid seed production appears to be a more complex challenge. Methods of hybrid seed production, such as cytoplasmic male sterility, genetic male sterility, and chemical hybridization methods, need to be optimized. In addition, the floral biology of wheat needs to be redesigned to favor cross-pollination (Whitford et al., 2013; Boeven et al., 2016). The most important factor to facilitate cross-pollination in wheat is higher AE, while ensuring that the male parent does not lose its ability to contribute to higher grain yield in the subsequent hybrid crosses. A recurrent selection scheme needs to be implemented within the male germplasm pool to develop superior male lines with desirable floral traits favoring cross-pollination along with other attributes for superior yield and quality. In the absence of large-effect loci, high G × E variance and labor-intensive phenotyping, MAS, and visual selection are inadequate. Hence, in this study, we demonstrate the utility of genome-wide prediction for AE via modeling for G × E and environmental covariates as a more reliable substitute for MAS and/or visual selection.

Phenotypic Evaluation of AE

AC and VS appear to be reliable measurements for AE, as demonstrated by their high heritability estimates. The heritability of AC in this study ranged from 0.79 to 0.93 which is comparable to heritability reported in similar previous studies (Boeven et al., 2016; Muqaddasi et al., 2017a). The heritability of VS in this

TABLE 3 | Mean and standard deviation of genomic prediction accuracies from the reaction norm models (Jarquín et al., 2014) under the two cross-validation scenarios (CV1 and CV2) for two traits representing anther extrusion; anther count (%) and visual score were collected from a field experiment spanning three environments (El Batan 2017, Obregon 2018 and El Batan 2019).

		Test environments					
		2017		2018		2019	
Traits	**Models**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
				CV1			
Anther count (%)	E + L	−0.032	0.036	−0.040	0.034	−0.038	0.028
	E + L + G	0.424	0.014	0.428	0.012	0.441	0.011
	E + L + G + G × E	0.429	0.017	0.435	0.011	0.454	0.012
	E + L + A	0.342	0.013	0.329	0.014	0.302	0.015
	E + L + A + A × E	0.355	0.014	0.333	0.016	0.307	0.016
	E + L + G + A	0.431	0.016	0.427	0.014	0.440	0.014
	E + L + G + A + G × E + A × E	0.441	0.017	0.431	0.015	0.452	0.015
Visual score	E + L			−0.055	0.054	−0.067	0.039
	E + L + G			0.332	0.012	0.459	0.016
	E + L + G + G × E			0.325	0.016	0.461	0.016
	E + L + A			0.217	0.014	0.335	0.017
	E + L + A + A × E			0.216	0.021	0.336	0.018
	E + L + G + A			0.328	0.012	0.458	0.016
	E + L + G + A + G × E + A × E			0.330	0.015	0.448	0.017
				CV2			
Anther count (%)	E + L	0.674	0.001	0.704	0.001	0.716	0.002
	E + L + G	0.687	0.001	0.710	0.001	0.726	0.002
	E + L + G + G × E	0.690	0.004	0.722	0.003	0.733	0.004
	E + L + A	0.686	0.001	0.708	0.001	0.715	0.002
	E + L + A + A × E	0.690	0.003	0.718	0.004	0.724	0.003
	E + L + G + A	0.689	0.001	0.708	0.001	0.726	0.002
	E + L + G + A + G × E + A × E	0.694	0.004	0.722	0.004	0.736	0.005
Visual score	E + L			0.424	0.005	0.417	0.007
	E + L + G			0.452	0.007	0.540	0.009
	E + L + G + G × E			0.451	0.011	0.522	0.012
	E + L + A			0.427	0.006	0.496	0.010
	E + L + A + A × E			0.434	0.009	0.454	0.015
	E + L + G + A			0.453	0.006	0.541	0.008
	E + L + G + A + G × E + A × E			0.455	0.011	0.509	0.013

CV1 is when 20% of lines in the test set do not have any phenotypic record, and CV2 is when the lines in the test set have a phenotypic record from a different environment but not in the environment being tested.

The models tested in increasing order of complexity are (E + L): Environment and line main effects; (E + L + G): Environment, line and genomic relationship main effects; (E + L + G + G × E): Environment, line, genomic relationship main effects, and genomic relationship by environment effects; (E + L + A): Environment, line, and pedigree relationship main effects; (E + L + A + A × E): Environment, line, pedigree relationship main effects, and pedigree by environment interaction effects; (E + L + G + A): Environment, line, genomic and pedigree relationship main effects; and (E + L + G + A + G × E + A × E): Environment, line, pedigree relationship, genomic relationship main effects, and genomic relationship by environment and pedigree by environment interaction effects.

study was quite modest compared to similar previous studies. Previous studies have reported heritabilities in the range of 0.5 to 0.8 for VS (Boeven et al., 2016; Muqaddasi et al., 2019). However, it should be noted that the error variance for VS was confounded with G × E due to the lack of replications within an environment. High positive correlations between VS and AC within the same environment indicated that VS could be a reliable trait for measuring AE. VS has been found to be an adequate trait to measure AE in several previous studies (Boeven et al., 2016, 2018; Muqaddasi et al., 2017c,a). The continuous distribution of AC indicated that AE is a complex trait governed by cumulative effects of numerous minor effect loci, making it suitable for GS. Moreover, significant genotypic variances for the two traits measured indicated that these traits can be improved by breeding efforts (**Table 1**; Boeven et al., 2016, 2018).

Marker Trait Associations for AE

Height-reducing loci, such as *Rht-B1* and *Rht-D1*, have been shown to reduce AE in several previous studies (Boeven et al., 2016; Würschum et al., 2018). None of the large effect *Rht* loci were identified in our analysis. CIMMYT spring wheat germplasm has been subjected previously to GWAS for AE and, similarly, *Rht* loci were not identified (Muqaddasi et al., 2017c).

TABLE 4 | Mean and standard deviation of genomic prediction accuracies from the reaction norm models (Jarqin et al., 2014) under the cross-validation scenarios (CV0) for two traits representing anther extrusion; anther count (%) and visual score were collected from a field experiment spanning three environments (2017, 2018, and 2019).

		Test environments		
		2017	2018	2019
Traits	**Models**	**Mean**	**Mean**	**Mean**
			CV0	
Anther extrusion (%)	E + L	0.676	0.706	0.718
	E + L + G	0.686	0.703	0.722
	E + L + G + GE	0.683	0.705	0.724
	E + L + A	0.684	0.705	0.710
	E + L + A + AE	0.677	0.707	0.716
	E + L + G + A	0.687	0.701	0.720
	E + L + G + A + GE + AE	0.681	0.701	0.725
Visual score	E + L		0.432	0.433
	E + L + G		0.465	0.470
	E + L + G + GE		0.453	0.450
	E + L + A		0.434	0.464
	E + L + A + AE		0.434	0.435
	E + L + G + A		0.463	0.482
	E + L + G + A + GE + AE		0.462	0.470

CV0 is prediction of performance of lines in a new environment using a phenotypic record from other environments.

The models tested in increasing order of complexity are: (E + L): Environment and line main effects; (E + L + G): Environment, line, and genomic relationship main effects; (E + L + G + G × E): Environment, line, genomic relationship main effects, and genomic relationship by environment effects; (E + L + A): Environment, line, and pedigree relationship main effects; (E + L + A + A × E): Environment, line, pedigree relationship main effects, and pedigree by environment interaction effects; (E + L + G + A): Environment, line, genomic, and pedigree relationship main effects; and (E + L + G + A + G × E + A × E): Environment, line, pedigree relationship, genomic relationship main effects, and genomic relationship by environment and pedigree by environment interaction effects.

This is most likely due to the fact that *Rht* loci, in particular *Rht-B1*, have been largely deployed in CIMMYT since the 1970s, and Rht-B1 is almost fixed in recent elite germplasm (Aisawi et al., 2015; Basnet et al., 2019).

Two MTAs were identified in chromosome 5A, one on the distal and one on the proximal end of the chromosome, based on physical positions. Previous studies have reported QTL for AE with minor to moderate effects on chromosome 5A via linkage mapping in biparental populations (Lu et al., 2013; Buerstmayr and Buerstmayr, 2015; Muqaddasi et al., 2019). The previously reported QTL for AE are spread throughout chromosome 5A, and the MTAs in 5A share close proximity with several of these previously reported QTLs. For example, the MTAs identified in this study in the short arm of 5A is at 26.51 cM (98.94 Mb), which is in between the AE QTL identified by Buerstmayr and Buerstmayr (2015) at 20 cM and Lu et al. (2013) at 33 cM. Similarly, Muqaddasi et al. (2019) have reported AE QTLs in CIMMYT germplasm at 59 cM, which lies very close to the other MTA identified in this study at 62.72 cM (492.79 Mb). Several previous studies have also reported QTL for AE on 1B and 5B (Skinnes et al., 2010; Boeven et al., 2016; Muqaddasi et al., 2017b,a). Boeven et al. (2016) have reported an MTA for AE in 1B at 74.4 cM; Muqaddasi et al. (2017b) have reported an MTA at 56.4 cM; Muqaddasi et al. (2017a) have reported an MTA at 70.08 cM; and Skinnes et al. (2010) have reported a QTL with confidence interval of 86–102 cM. The MTA detected in 1B in this study at 60.62 cM lies close enough to these previously reported loci. Similarly, in 5B Muqaddasi et al. (2017b) have reported an MTA at 108.7 cM, whereas we identified an MTA at 117.84 cM in this study.

Since the phenotypic effects of these MTAs across studies were low, there is limited scope to use these MTAs in MAS. The only option might be to use these QTL together with the *Rht* loci in MAS, in the event that they can be successfully validated and concurrently do not show any negative effect on other traits, e.g., lodging or grain yield.

Despite having high heritabilities for VS and AC, the MTAs explain a very low amount of phenotypic variance. This signifies the highly polygenic nature of inheritance of AE and suggests that MAS is not the best strategy for driving genetic gains in AE. In this context, genomic prediction/selection can be an excellent strategy for making selection gains in breeding for AE.

Genome-Wide Predictions

Genome-wide prediction for AE has been previously conducted on an unrelated smaller subset of CIMMYT spring wheat germplasm (Muqaddasi et al., 2017c) and winter wheat germplasm from Western Europe (Boeven et al., 2016). Both studies used single-trait models without explicitly modeling G × E. Incorporating G × E effects and environmental covariates has previously shown higher prediction accuracy for grain yield (Burgueño et al., 2012; Jarquín et al., 2014; Sukumaran et al., 2017, 2018; Krause et al., 2019), micronutrient concentration (Velu et al., 2016), and lint yield in cotton (Pérez-Rodríguez et al., 2015). In this study, we attempted to implement the same approach for predicting AE.

Prediction of Performance of Untested Lines

In CV1, lines that were never tested before were predicted by borrowing information from closely related individuals, which is akin to the previous GS studies for AE (Boeven et al., 2016; Muqaddasi et al., 2017c). Boeven et al. (2016) reported prediction accuracies of 0.3 for VS and 0.6 for AC implementing ridge regression BLUP (RR-BLUP), whereas inclusion of weighted effects of *Rht* loci via weighted ridge regression BLUP (wRR-BLUP) increased the prediction accuracies to 0.5 for VS and 0.7 for AC. Muqaddasi et al. (2017c) reported a prediction accuracy of 0.6 for VS in a CIMMYT population using RR-BLUP. However, the prediction accuracy was standardized with the square root of heritability in the Muqaddasi et al. (2017c) study. The prediction accuracies of the main effect model (model 2) in CV1 for both VS and AC are comparable to these previous studies. When interaction effects were added in models 3 and 7, the prediction accuracies tended to increase slightly (1–3%). In CV1, models using the G matrix (models 2 and 3) always had higher prediction accuracies than models using pedigree-based A matrix (models 4 and 5), which has

been reported previously by Campos et al. (2009, 2010) and Crossa et al. (2010, 2011).

In model 1, where only the main effect of environments and lines are used for prediction, negative prediction accuracy is observed. Since neither pedigree information nor genomic information is included in this model, it does a very poor job of predicting performance. This is expected since genomic prediction is based on using information borrowed from related individuals via pedigree and genomic relationship, and in model 1 only incidence matrices for lines and environments are included. Once we start including pedigree and genomic information the prediction accuracies are positive and higher.

Prediction of Performance of Previously Tested Lines in Untested Environments

CV2 and CV0 were the scenarios where phenotypic information of the same line from one environment was used to predict VS and AC in another environment. CV2 is similar to what is also called sparse testing, where some of the lines are missing, whereas CV0 is the prediction of the whole population in a previously untested environment. In cases when a phenotypic record of the line being tested is used to train the model (CV2 and CV0), the prediction accuracy is higher compared to the case where the line has never been tested previously (CV1) (Sukumaran et al., 2017, 2018; Basnet et al., 2019). We found similar results in the study, as expected. These CV scenarios can be very useful in the case of traits that are difficult to phenotype due to cost or the labor-intensive nature of phenotyping. The results for CV0 and CV2 indicate that untested sites, environments, and years can be predicted with high reliability. CV0 and CV2 scenarios can supplement the field evaluation efforts of breeding programs. In particular, sparse testing, i.e., the CV2 scenario, is already in practice in hybrid breeding for the development of heterotic pools in wheat (Zhao et al., 2015).

Inclusion of interaction effects such as G × E and A × E produced mixed results. For AC, the inclusion of interaction effects produced a very nominal increase in prediction accuracy (1–3%), whereas for VS it decreased nominally in most cases.

Implications for Hybrid Wheat Breeding

GS is promising for driving the genetic gain of AE. However, the prediction accuracies are also dependent on trait heritability values (Velu et al., 2016; Acosta-Pech et al., 2017). AC had higher heritability compared to VS in our study. VS is easier to phenotype than AC. Here, VS data were unreplicated, and the error variance was confounded with G × E. It is easier and cost-effective to increase replications for VS than to collect data for AC routinely in the breeding program. Hence, the use of replicated trials can help increase the prediction accuracy for VS.

Inclusion of interaction terms had a very nominal advantage in prediction accuracy for AC, whereas it was sometimes counterproductive in VS, most likely due to increasing model complexity. Based on observations from this study, it is possible to predict AE with reasonable accuracy using pedigree data only, but the inclusion of genomic data should always be preferred. Inclusion of both pedigree and genomic data appears to work best but not in all cases.

Reciprocal recurrent selection is a promising strategy in hybrid wheat breeding (Rembe et al., 2019). For VS, sparse testing appears to be a good strategy, where a subset of lines would be tested in some environments but not all. Information on tested relatives in an environment can be used to predict untested lines. For AC, which is labor-intensive and expensive to phenotype, data from a subset of highly maintained trials can be used to predict performance in an untested environment.

DATA AVAILABILITY STATEMENT

The datasets for this study can be found in the CIMMYT Research Data & Software Repository Network (http://hdl.handle.net/11529/10548495).

AUTHOR CONTRIBUTIONS

AA collected and analyzed the data, and wrote the manuscript. BB conceived and designed the experiment, collected and analyzed phenotypic data, and wrote the manuscript. JC and DJ analyzed the data and wrote the manuscript. PB and FC did field experiments, collected phenotypic data, and reviewed the manuscript. SD generated and curated genotypic data, analyzed data, and reviewed the manuscript. YM and AI developed the concept and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors thank the CGIAR Research Program in Wheat (CRP-WHEAT https://wheat.org/) for funding this work. The authors also thank Syngenta for generous support in SNP genotyping of wheat lines used in this study under the CIMMYT-Syngenta partnership on hybrid wheat research.

ACKNOWLEDGMENTS

The authors thank their field research assistants Roman Delgado, Jose Luis Tapia, and Eira Gomez for their dedication and hard work, which were vital to the success of the field experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.586687/full#supplementary-material

Supplementary Figure 1 | Plot showing pairwise correlations between traits measuring anther extrusion from three environments (El Batan 2017, Obregon 2018 and El Batan 2019). Trait abbreviations: AE_2017: Anther count (%) 2017; AE_2018: Anther count (%) 2018; AE_2019: Anther count (%) 2019; VS_2018: Visual score of anther extrusion 2018; and VS_2019: Visual score of anther extrusion 2019. Visual score data were not collected in 2017.

REFERENCES

Acosta-Pech, R., Crossa, J., de los Campos, G., Teyssèdre, S., Claustres, B., Pérez-Elizalde, S., et al. (2017). Genomic models with genotype×environment interaction for predicting hybrid performance: an application in maize hybrids. *Theor. Appl. Genet.* 130, 1431–1440. doi: 10.1007/s00122-017-2898-2890

Adhikari, A., Ibrahim, A. M., Rudd, J. C., Baenziger, P. S., and Sarazin, J. (2020a). Estimation of heterosis and combining abilities of US winter wheat germplasm for hybrid development in Texas. *Crop Sci.* 60, 788–803. doi: 10.1002/csc2.20020

Adhikari, A., Ibrahim, A. M. H., Rudd, J. C., Baenziger, P. S., Easterly, A., Garst, N., et al. (2020b). Supplementing selection decisions in a hybrid wheat breeding program by using F2 yield as a proxy of F1 performance. *Euphytica* 216:130. doi: 10.1007/s10681-020-02664-0

Aisawi, K. A. B., Reynolds, M. P., Singh, R. P., and Foulkes, M. J. (2015). The physiological basis of the genetic progress in yield potential of CIMMYT spring wheat cultivars from 1966 to 2009. *Crop Sci.* 55, 1749–1764. doi: 10.2135/cropsci2014.09.0601

Alvarado, G., López, M., Vargas, M., Pacheco, Á, Rodríguez, F., Burgueño, J., et al. (2015). *META-R (Multi Environment Trail Analysis with R for Windows) Version 5.0*. Mexico: International Maize and Wheat Improvement Center (CIMMYT).

Basnet, B. R., Crossa, J., Dreisigacker, S., Pérez-Rodríguez, P., Manes, Y., Singh, R. P., et al. (2019). Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. *Plant Genome* 12, 1–13. doi: 10.3835/plantgenome2018.07.0051

Boeven, P. H., Longin, C. F. H., Leiser, W. L., Kollers, S., Ebmeyer, E., and Würschum, T. (2016). Genetic architecture of male floral traits required for hybrid wheat breeding. *Theor. Appl. Genet.* 129, 2343–2357. doi: 10.1007/s00122-016-2771-6

Boeven, P. H. G., Würschum, T., Rudloff, J., Ebmeyer, E., and Longin, C. F. H. (2018). Hybrid seed set in wheat is a complex trait but can be improved indirectly by selection for male floral traits. *Euphytica* 214:110. doi: 10.1007/s10681-018-2188-2181

Buerstmayr, M., and Buerstmayr, H. (2015). Comparative mapping of quantitative trait loci for Fusarium head blight resistance and anther retention in the winter wheat population Capo × Arina. *Theor. Appl. Genet.* 128, 1519–1530. doi: 10.1007/s00122-015-2527-2528

Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop. Sci.* 52, 707–719. doi: 10.2135/cropsci2011.06.0299

Campos, G. D. L., Gianola, D., Rosa, G. J. M., Weigel, K. A., and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res.* 92, 295–308. doi: 10.1017/S0016672310000285

Campos, G. D. L., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501

Crossa, J., Campos, G. D. L., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186, 713–724. doi: 10.1534/genetics.110.118521

Crossa, J., Pérez, P., Campos, G., de los, Mahuku, G., Dreisigacker, S., and Magorokosho, C. (2011). Genomic selection and prediction in plant breeding. *J. Crop Improv.* 25, 239–261. doi: 10.1080/15427528.2011.558767

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011

De Vries, A. P. (1971). Flowering biology of wheat, particularly in view of hybrid seed production—a review. *Euphytica* 20, 152–170. doi: 10.1007/bf00056076

Dreisigacker, S., Melchinger, A., Zhang, P., Ammar, K., Flachenecker, C., Hoisington, D., et al. (2005). Hybrid performance and heterosis in spring bread wheat, and their relations to SSR-based genetic distances and coefficients of parentage. *Euphytica* 144, 51–59. doi: 10.1007/s10681-005-4053-2

Dreisigacker, S., Sehgal, D., Reyes Jaimez, A. E., Luna Garrido, B., Muñoz Zavala, S., Núñez Ríos, C., (eds.) et al. (2016). *CIMMYT Wheat Molecular Genetics: Laboratory Protocols and Applications to Wheat Breeding*. Mexico, DF: CIMMYT.

Easterly, A. C., Garst, N., Belamkar, V., Ibrahim, A. M., Rudd, J. C., Sarazin, J., et al. (2020). Evaluation of hybrid wheat yield in Nebraska. *Crop Sci.* 60, 1210–1222. doi: 10.1002/csc2.20019

Gowda, M., Kling, C., Würschum, T., Liu, W., Maurer, H., Hahn, V., et al. (2010). Hybrid breeding in durum wheat: heterosis and combining ability. *Crop Sci.* 50, 2224–2230. doi: 10.2135/cropsci2009.10.0637

Gowda, M., Longin, C. F. H., Lein, V., and Reif, J. C. (2012). Relevance of specific versus general combining ability in winter wheat. *Crop Sci.* 52, 2494–2500. doi: 10.2135/cropsci2012.04.0245

Gupta, P. K., Balyan, H. S., Gahlaut, V., Saripalli, G., Pal, B., Basnet, B. R., et al. (2019). Hybrid wheat: past, present and future. *Theor. Appl. Genet.* 132, 2463–2483.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92, 433–443. doi: 10.3168/jds.2008-1646

He, X., Singh, P. K., Dreisigacker, S., Singh, S., Lillemo, M., and Duveiller, E. (2016). Dwarfing Genes Rht-B1b and Rht-D1b are associated with both Type I FHB susceptibility and low anther extrusion in two bread wheat populations. *PLoS One* 11:e0162499. doi: 10.1371/journal.pone.0162499

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127, 595–607. doi: 10.1007/s00122-013-2243-2241

Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., et al. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes Genomes Genet.* 9, 1231–1247. doi: 10.1534/g3.118.200856

Langer, S. M., Longin, C. F. H., and Würschum, T. (2014). Phenotypic evaluation of floral and flowering traits with relevance for hybrid breeding in wheat (T riticum aestivum L.). *Plant Breed* 133, 433–441. doi: 10.1111/pbr.12192

Longin, C. F. H., Gowda, M., Mühleisen, J., Ebmeyer, E., Kazman, E., Schachschneider, R., et al. (2013). Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. *Theor. Appl. Genet.* 126, 2791–2801. doi: 10.1007/s00122-013-2172-z

Longin, C. F. H., Mühleisen, J., Maurer, H. P., Zhang, H., Gowda, M., and Reif, J. C. (2012). Hybrid breeding in autogamous cereals. *Theor. Appl. Genet.* 125, 1087–1096. doi: 10.1007/s00122-012-1967-7

Longin, C. F. H., Reif, J. C., and Würschum, T. (2014). Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory. *Theor. Appl. Genet.* 127, 1635–1641. doi: 10.1007/s00122-014-2325-8

Lu, Q., Lillemo, M., Skinnes, H., He, X., Shi, J., Ji, F., et al. (2013). Anther extrusion and plant height are associated with Type I resistance to Fusarium head blight in bread wheat line 'Shanghai-3/Catbird.'. *Theor. Appl. Genet.* 126, 317–334. doi: 10.1007/s00122-012-1981-1989

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 154, 1819–1829.

Mühleisen, J., Piepho, H.-P., Maurer, H. P., Longin, C. F. H., and Reif, J. C. (2014). Yield stability of hybrids versus lines in wheat, barley, and triticale. *Theor. Appl. Genet.* 127, 309–316. doi: 10.1007/s00122-013-2219-1

Muqaddasi, Q. H., Brassac, J., Börner, A., Pillen, K., and Röder, M. S. (2017a). Genetic architecture of anther extrusion in spring and winter wheat. *Front. Plant Sci.* 8:754. doi: 10.3389/fpls.2017.00754

Muqaddasi, Q. H., Pillen, K., Plieske, J., Ganal, M. W., and Röder, M. S. (2017b). Genetic and physical mapping of anther extrusion in elite European winter wheat. *PLoS One* 12:e0187744. doi: 10.1371/journal.pone.0187744

Muqaddasi, Q. H., Reif, J. C., Li, Z., Basnet, B. R., Dreisigacker, S., and Röder, M. S. (2017c). Genome-wide association mapping and genome-wide prediction of anther extrusion in CIMMYT spring wheat. *Euphytica* 213:73. doi: 10.1007/s10681-017-1863-y

Muqaddasi, Q. H., Reif, J. C., Röder, M. S., Basnet, B. R., and Dreisigacker, S. (2019). Genetic mapping reveals large-effect QTL for anther extrusion in CIMMYT spring wheat. *Agronomy* 9:407. doi: 10.3390/agronomy9070407

Ogihara, Y., Takumi, S., and Handa, H. (2013). *Advances in Wheat Genetics: From Genome to Field*. Berlin: Springer.

Pérez-Rodríguez, P., Crossa, J., Bondalapati, K., Meyer, G. D., Pita, F., and Campos, G. D. L. (2015). A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. *Crop Sci.* 55, 1143–1151. doi: 10.2135/cropsci2014.08.0577

Rembe, M., Zhao, Y., Jiang, Y., and Reif, J. C. (2019). Reciprocal recurrent genomic selection: an attractive tool to leverage hybrid wheat breeding. *Theor. Appl. Genet.* 132, 687–698. doi: 10.1007/s00122-018-3244-x

Sage, G. C. M., and De Isturiz, M. J. (1974). The inheritance of anther extrusion in two spring wheat varieties. *Theor. Appl. Genet.* 45, 126–133. doi: 10.1007/BF00291142

Skinnes, H., Semagn, K., Tarkegne, Y., Marøy, A. G., and Bjørnstad, Å (2010). The inheritance of anther extrusion in hexaploid wheat and its relationship to Fusarium head blight resistance and deoxynivalenol content. *Plant Breed* 129, 149–155. doi: 10.1111/j.1439-0523.2009.01731.x

Steiner, B., Buerstmayr, M., Wagner, C., Danler, A., Eshonkulov, B., Ehn, M., et al. (2019). Fine-mapping of the Fusarium head blight resistance QTL Qfhs.ifa-5A identifies two resistance QTL associated with anther extrusion. *Theor. Appl. Genet.* 132, 2039–2053. doi: 10.1007/s00122-019-03336-x

Sukumaran, S., Crossa, J., Jarquin, D., Lopes, M., and Reynolds, M. P. (2017). Genomic Prediction with pedigree and genotype × environment interaction in spring wheat grown in South and West Asia. North Africa, and Mexico. *G3 Genes Genomes Genet.* 7, 481–495. doi: 10.1534/g3.116.036251

Sukumaran, S., Jarquin, D., Crossa, J., and Reynolds, M. (2018). Genomic-enabled prediction accuracies increased by modeling genotype × environment interaction in durum wheat. *Plant Genome* 11:170112. doi: 10.3835/plantgenome2017.12.0112

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). GAPIT Version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9:lantgenome2015.11.0120. doi: 10.3835/plantgenome2015.11.0120

Velu, G., Crossa, J., Singh, R. P., Hao, Y., Dreisigacker, S., Perez-Rodriguez, P., et al. (2016). Genomic prediction for grain zinc and iron concentrations in spring wheat. *Theor. Appl. Genet.* 129, 1595–1605. doi: 10.1007/s00122-016-2726-y

Virmani, S., and Edwards, I. B. (1983). Current status and future prospects for breeding hybrid rice and wheat. *Adv. Agronomy* 36, 145–214. doi: 10.1016/s0065-2113(08)60354-5

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Whitford, R., Fleury, D., Reif, J. C., Garcia, M., Okada, T., Korzun, V., et al. (2013). Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *J. Exp. Bot.* 64, 5411–5428. doi: 10.1093/jxb/ert333

Würschum, T., Liu, G., Boeven, P. H. G., Longin, C. F. H., Mirdita, V., Kazman, E., et al. (2018). Exploiting the Rht portfolio for hybrid wheat breeding. *Theor. Appl. Genet.* 131, 1433–1442. doi: 10.1007/s00122-018-3088-3084

Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Würschum, T., et al. (2015). Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci. U S A.* 112, 15624–15629. doi: 10.1073/pnas.1514547112

Conflict of Interest: YM was employed by company Syngenta France S.A.S, France.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Adhikari, Basnet, Crossa, Dreisigacker, Camarillo, Bhati, Jarquin, Manes and Ibrahim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 December 2020
doi: 10.3389/fgene.2020.584527

Identification of Allelic Variation in Drought Responsive Dehydrin Gene Based on Sequence Similarity in Chickpea (*Cicer arietinum* L.)

Tapan Kumar[1,2], *Neha Tiwari*[2], *Chellapilla Bharadwaj*[1]*, *Ashutosh Sarker*[2], *Sneha Priya Reddy Pappula*[1], *Sarvjeet Singh*[3] and *Mohar Singh*[4]

[1] Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India, [2] International Center for Agricultural Research in the Dry Areas, Bhopal, India, [3] Department of Plant Breeding & Genetics, Punjab Agricultural University, Ludhiana, India, [4] Department of Plant Breeding & Genetics, ICAR-NBPGR Regional Station, Shimla, India

OPEN ACCESS

Edited by:
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Cheng Qin,
Zunyi Vocational and Technical College, China
Manish Kumar Pandey,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India
S. K. Jain,
Sri Karan Narendra Agriculture University, India

***Correspondence:**
Chellapilla Bharadwaj
drchbharadwaj@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 17 July 2020
Accepted: 18 November 2020
Published: 14 December 2020

Citation:
Kumar T, Tiwari N, Bharadwaj C, Sarker A, Pappula SPR, Singh S and Singh M (2020) Identification of Allelic Variation in Drought Responsive Dehydrin Gene Based on Sequence Similarity in Chickpea (Cicer arietinum L.). Front. Genet. 11:584527. doi: 10.3389/fgene.2020.584527

Chickpea (*Cicer arietinum* L.) is an economically important food legume grown in arid and semi-arid regions of the world. Chickpea is cultivated mainly in the rainfed, residual moisture, and restricted irrigation condition. The crop is always prone to drought stress which is resulting in flower drop, unfilled pods, and is a major yield reducer in many parts of the world. The present study elucidates the association between candidate gene and morpho-physiological traits for the screening of drought tolerance in chickpea. Abiotic stress-responsive gene Dehydrin (*DHN*) was identified in some of the chickpea genotypes based on the sequence similarity approach to play a major role in drought tolerance. Analysis of variance revealed a significant effect of drought on relative water content, membrane stability index, plant height, and yield traits. The genotypes Pusa1103, Pusa362, and ICC4958 were found most promising genotypes for drought tolerance as they maintained the higher value of osmotic regulations and yield characters. The results were further supported by a sequence similarity approach for the dehydrin gene when analyzed for the presence of single nucleotide polymorphisms (SNPs) and indels. Homozygous indels and single nucleotide polymorphisms were found after the sequencing in some of the selected genotypes.

Keywords: chickpea, drought tolerance, dehydrin (DHN), sequence similarity, morphological characterization

INTRODUCTION

Drought is an environmental condition that arises due to the water scarcity and is a result of very low rainfall or water supply. The severity of the drought is determined by its timing and duration (Toker et al., 2007). It is estimated that fifty per cent yield losses are caused by drought and heat stress (Gaur et al., 2012a). Discovering the genotypic variation between the chickpea genotypes for drought tolerance is most important for the execution of breeding programs for chickpea (Kumar et al., 2018). Chickpea (*Cicer arietinum* L.; *Fabaceae* family) is a diploid plant, containing chromosome number (2n = 16), self-pollinated and cool-season pulse. The genome size of chickpea is approx. ∼738 Mb and reported to be having an estimated 28,269 genes (Varshney et al., 2013). Chickpea is being grown in more than fifty countries across the globe (Upadhyaya et al., 2011; Gaur et al., 2012b). It is the most important food legume crop, grown in tropical, subtropical, and

temperate regions (Mohammed et al., 2017). Chickpea is cultivated mostly in the rainfed condition (Kumar and Abbo, 2001) and drought is a major constraint for chickpea production (Toker et al., 2007).

Chickpea crop responds variably to the drought stress depending upon the variety, growth stage, and stress duration (Maqbool et al., 2017). Considerable variation exists for the morphological and physiological traits for drought resilience at the different developmental stages. Various studies have established several morpho-physiological parameters for the screening of the drought-like days to 50% flowering (DTF), maturity in days (DTM), relative water content (RWC), membrane stability index (MSI), yield components, etc (Bharadwaj et al., 2011; Maqbool et al., 2017; Shah et al., 2020). A rigorous phenotypic screening technique is required for a better understanding of the crop responses under the stress conditions.

Dehydrin (*DHN*) proteins are the stress-responsive proteins observed under low temperature or in dehydration (Hanin et al., 2011). In the total seed, protein dehydrin is present up to 4% and is assumed to be involved in protecting the embryo and seed tissues from osmotic disturbances when available water in the mature seed is very low (Wise and Tunnacliffe, 2004). Transgenic plants overexpressing *DHN* showed better growth and tolerance when exposed to the drought and freezing stress compared to the wild-type plants (Puhakainen et al., 2004a). *DHNs* are one of the many proteins that have been precisely related to qualitative and quantitative changes in the cold hardiness (Close, 1996). It is also found that plants engineered for *DHN* over-expression, displayed better endurance when exposed to the low temperature in the Arabidopsis (Puhakainen et al., 2004b).

Drought stress shortens the growing season, which affects the yield components viz., seed weight, total biomass, pod number, yield per plant, and seed number of the plants (Toker et al., 2007). The unavailability of water negatively affects the various physiological processes associated with crop developments including stomatal regulation, photosynthesis, and transpiration. It also affects cell growth, hormonal, and enzyme concentration (Hsiao, 1973; Boyer and McPherson, 1975; Begg and Turner, 1976). Improved sequencing technology is a quick and low-cost method through which enormous sequence data can be generated and is eventually helpful for the identification of genes responsible for stress tolerance (Castro et al., 2012). Some of the candidate genes known for the abiotic stress tolerance are Snf-1 related kinase (AKIN), DREB2A gene, dehydrin (*DHN*), CAP2 gene, and Myb transcription factor (MYB) (Roorkiwal et al., 2014). The sequences of these abiotic stress-related candidate genes can be used as a reference sequence for crop improvement via molecular breeding, especially for complex traits (Deshmukh et al., 2014). For the present investigation, fifty chickpea genotypes consisting of released varieties, germplasm collections, landraces, and wild derivatives were used for the study of morpho-physiological characters and the sequence similarity identification of *DHN* gene.

MATERIALS AND METHODS

Fifty chickpea genotypes consisting of released varieties, germplasm collections, landraces, and wild derivatives were evaluated in two replications under the normal and rainfed conditions at the experimental farm of IARI, NeDelhi (28.6377° N and 77.1571° E) with altitude 228.61 m over mean sea level) during 2015–16 and 2016–17. Field experiments were performed in the randomized block design with two replications for all the genotypes including tolerant and susceptible checks. Genotypes under investigation were grown in two meters and two rows with a spacing of 45 cm between rows and 10 cm within the rows (**Supplementary Table 1**).

Morphological Characterization of Chickpea Genotypes by Agronomic Data

Data for the morpho-physiological characters were recorded for days to 50% flowering, maturity in days, hundred seed weight, plant yield, relative water content, and stability index of the membrane. Mean values were used for analysis in CROPSTAT (version 7.2) statistical package: https://cropstat.software.informer.com/7.2/. Pearson's Correlation matrix among the traits under control and rainfed conditions had been generated by employing GenSTAT version 16.1: www.vsni.co.uk/software/Genstat (**Table 1**).

Factorial and clusters analysis for drought based on morpho-physiological traits has been done by using DARwin 5 software 5.0.158 (Perrier and Jacquemoud-Collet, 2006).

Relative Water Content (RWC)

Three leaflets on top, middle, and lower part of the plant (0.5 g) were taken for measuring relative water content (%) at 50% podding stage. The calculation was done by the following formula given by Blum and Ebercon (1981).

$$RWC\ (\%) = F_w - D_w/T_w - D_w\ x100$$

Where, F_w = Fresh weight, T_w = Turgid weight, D_w = Dry weight.

Membrane Stability Index (MSI)

Two-gram fresh weight of leaf samples were taken to record the membrane stability index at 50% flowering stage. MSI calculations were done by the following formula given by Blum and Ebercon (1981).

$$MSI = (1 - C_1/C_2)\ x\ 100$$

Where, C_1 = Electrical conductivity at 40°C for 30 min
C_2 = Electrical conductivity at 100°C for 10 min.

Identification of Candidate Gene Related to Abiotic Stress Tolerance

Based on morpho-physiological characterization data obtained, a subset of genotypes that were found tolerant to drought was selected for the validation of candidate gene linked to drought and for the identification of their allelic variation for the *DHN* gene.

Gel Extraction of DNA Fragment

The anticipated fragments were cut and scooped out from the gel. Purification of the gel was done as per the recommended protocol by using a gel extraction kit (PureLinkTM Quick gel extraction

TABLE 1 | Mean, standard error, coefficient of variation, range, heritability in broad sense, genetic advance and percentage decrease of traits in normal and rainfed condition.

Traits	Normal					Drought					Percent decrease of traits (%)
	Mean ± SE	CV (%)	Range	h^2	GA	Mean ± SE	CV (%)	Range	h^2	GA	
Days to 50% Flowering	78 ± 2.32	20.97	57 –103	0.979	36.518	68 ± 1.79	18.77	49–87	0.984	36.258	13.97
Days to maturity	134 ± 0.32	1.68	124 –136	0.586	4.667	116 ± 0.43	2.61	112–121	0.149	1.073	13.57
Plant height	57 ± 0.97	11.95	44 –75	0.959	15.772	50 ± 0.76	10.76	37–65	0.916	15.405	12.09
Pods per plant	38 ± 1.31	24.49	21 –56	0.718	17.368	24 ± 1.06	30.79	11–44	0.708	17.108	36.63
Seed per pod	2 ± 0.07	30.93	1 –2	0.627	1.699	2 ± 0.07	30.93	1–2	0.488	1.026	0.00
Relative water content	67 ± 1.46	15.48	52 –83	0.808	13.796	66 ± 1.67	17.79	51–90	0.906	18.180	0.35
Membrane stability index	60 ± 1.67	19.60	41 –78	0.921	19.211	59 ± 1.78	21.21	40–80	0.885	16.164	1.08
Yield	79 ± 3.09	27.59	40 –139	0.766	57.805	52 ± 2.44	33.39	24–85	0.737	30.635	34.41

TABLE 2 | Two way ANOVA for the morphological traits under normal vs. rainfed conditions.

Source of variation	Mean sum of square							
	DTF	MSI	DTM	PH	PPP	RWC	SPP	YLD
ANOVA table for control vs. drought								
Genotype	1211.98**	878.8**	24.694**	205.50**	289.48**	713.7**	1.46939	2067.93**
Treatment	8769.61**	37.45**	25208.33**	3508.92**	13981.01**	8.00**	0.0	56115.36**
Genotype treatment	70.02**	15.3**	17.81**	23.24**	138.20**	23.3**	0.0	258.82**
Residual	1.85	0.08	1.874	0.4436	1.672	0.08	0.0	1.599

*, **Significance at 5 and 1% respectively; DTF, Days to 50% flowering; MSI, Membrane Stability Index; DTM, Days to Maturity; PH, Plant Height; PPP, Pods per Plant; RWC, Relative Water Content; SPP, Seeds per Pod; YLD, Plant Yield.*

and PCR purification Combo kit, Invitrogen, Carlsbad, CA). Gel solubilization buffer was added to the gel weight (w/v), which contained DNA, and was dissolved by heating to 50°C for 10 min. The solution containing DNA was loaded to the Pure Link TM Clean-up spin column and a short spin for 1 min at 10,000 RPM was given. Flow-through was discarded and the columns were washed with wash buffer. Purified DNA was eluted with 50 μL of elution buffer and rechecked by electrophoresis and stored at −20°C for further use.

DNA Sequencing and Sequences Analysis

The PCR amplified products were further confirmed by sequencing. The sequencing of the selected amplicons was determined using an ABI automated sequencer (Chromous Biotech Pvt. Ltd., Bangalore, India).

The raw sequences of the desired candidate gene were aligned by the forward and reverse sequences of each genotype and gene identities were confirmed using BLAST against chickpea reference genome assembly. Similarity searches for the nucleotides were performed using BLAST at NCBI (www.ncbi.nlm.nih). Using the ORF finder, open reading frames (ORFs) were identified in NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The DNA sequences were aligned using the BioEdit version 7.2.5. This software was also used to detect single nucleotide polymorphisms (SNPs) and mutation/deletion.

RESULTS

Screening of Fifty Chickpea Genotypes for Drought Tolerance Under Irrigated and Rainfed Conditions

Two-way analysis of variance was carried out for all the characters under the normal and rainfed conditions (**Table 2**). Significant variation was observed under normal and drought environments (**Figure 1**, **Table 3**). Days to 50% flowering was positively and significantly correlated with days to maturity ($r =$ 0.417). Membrane stability index was negatively correlated with days to flowering ($r = -0.006$). Plant height was positively and significantly correlated with the days to maturity ($r = 0.47$). Pods per plant were positively and significantly correlated with the days to maturity ($r = 0.602$). Relative water content was positively and significantly correlated with the membrane stability index ($r = 0.912$). Seeds per pod was showing a significant but negative correlation with the days to flowering (—0.265). Yield is positively and significantly correlated with days to maturity ($r = 0.549$), membrane stability index (0.585), plant height (0.265) pods per plant (0.588) and relative water content (0.590) (**Figure 2**, **Table 2**).

In the present investigation, high heritability values coupled with high genetic advance were recorded for days to flowering, plant height, pods per plant, relative water content and

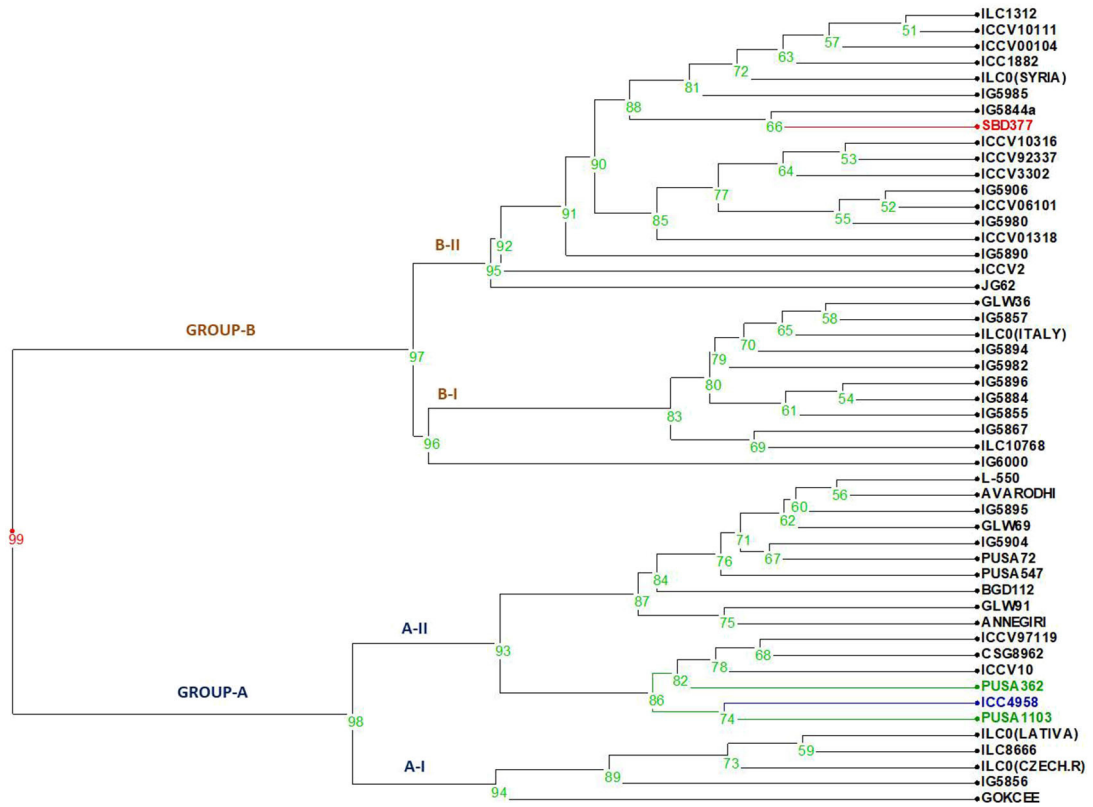


FIGURE 1 | Dendrogram generated from an unweighted pair group method analysis (UPGMA) cluster analysis based on all the stressed morphological characters for drought. The first two clusters form Group A showing all tolerant to moderate tolerant genotypes.

TABLE 3 | Pearson's Correlation matrix among the traits under control vs. rainfed conditions.

	DTF	DTM	MSI	PH	PPP	RWC	SPP	YLD
DTF	1**							
DTM	0.417**	1**						
MSI	– 0.006	0.033	1**					
PH	– 0.08	0.47**	– 0.075	1**				
PPP	0.173	0.602**	0.303**	0.231**	1**			
RWC	0.041	0.026	0.912**	– 0.076	0.274**	1**		
SPP	– 0.265**	– 0.043	0.16	0.033	0.266**	0.227**	1**	
YLD	0.182	0.549**	0.585**	0.265**	0.588**	0.59**	0.184	1**

*, **Significance at 5 and 1% respectively; DTF, Days to 50% Flowering; DTM, Days to Maturity; MSI, Membrane Stability Index; PH, Plant Height; PPP, Pod Per Plant; RWC, Relative Water Content; SPP, Seeds Per Pod; YLD, Plant Yield.

membrane stability index under irrigated and drought condition. Genotypes retaining early flowering, good plant height, pods per plant, water retention capacity and membrane stability under moisture stress are likely to be more productive under stress environment.

Moderate heritability was accompanied by low genetic advance for days to maturity and seeds per pods under control and drought condition. The comparison of heritability for all the traits was done under irrigated and drought stress conditions (**Table 3**). All the morpho-physiological data were analyzed and Euclidean distances were calculated for the stress condition and the genotypes grouped as per their characters. Two distinct groups (A and B) were formed (**Figure 3**). Tolerant lines were grouped in A and the susceptible lines were in group B. The tolerant genotypes were further divided into two groups, highly tolerant and moderately tolerant. Similarly, the susceptible genotypes were divided into two groups highly susceptible to moderately susceptible. According to all morpho-physiological characters, Pusa1103 and Pusa362 were found the most tolerant genotypes. Also, these tolerant genotypes were grouped with

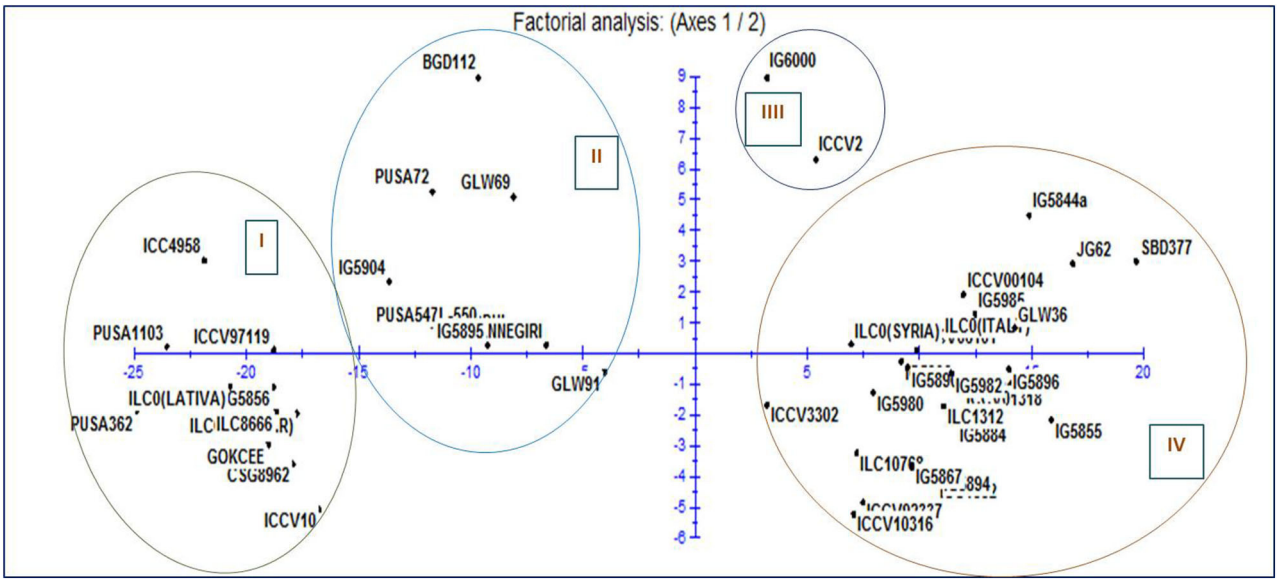


FIGURE 2 | Representation of the 1-2 plane of factorial analysis based on drought stress morphological traits for fifty chickpea genotypes.

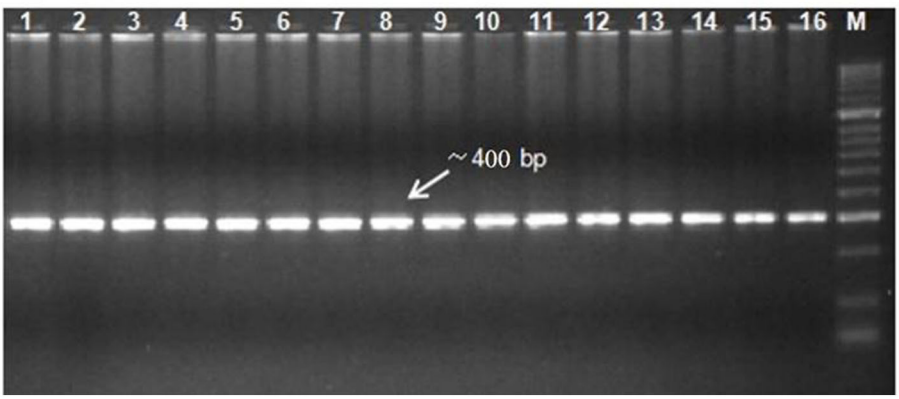


FIGURE 3 | Gel electrophoresis of *DHN* amplicons from different chickpea genotypes (lane: 1-16). Lane M: 2 kb DNA ladder.

ICC4958 which is a well-known donor for the drought tolerance. A data matrix plot based on the morphological characters had been subjected to Principal Component Analysis (PCoA) for estimating genetic differentiation among the fifty genotypes of chickpea. The scatter plot based on these components disclosed a pattern of mainly two groups. The tolerant genotypes formed a separate group with ICC4958 while the susceptible genotypes formed a group with SBD377. Most of the genotypes were scattered between tolerance and susceptible genotypes. The distribution of genotypes according to geographical origin was lacking in the matrix plot (**Figure 4**).

Sequence Similarity and Allelic Variation of the *DHN* Gene

Based on the morpho-physiological data, a subset of genotypes were selected for the sequence similarity and identification of allelic variation through sequencing. The *DHN* gene sequence was partially amplified through the genomic primer (Roorkiwal and Sharma, 2012). The amplification of the *DHN* gene primer generated a PCR product of ~400 bp in length (**Figure 5**). The amplified product has been further purified by using a gel purification technique as described in the materials and method section. Purified PCR products were subjected to sequencing (Chromous Biotech Pvt. Ltd., Bangalore, India). Analysis of the sequences and nucleotide identity searches had been done by BLASTN and BLASTX in NCBI (www.ncbi.nlm.nih) and found that the sequences are showing the highest identity with the homologous gene. The nucleotide sequence analysis indicated that the *DHN* gene from different genotypes had diversity among the sequences. The amino acid sequences homology of the *DHN* were found highest with the reference *DHN* gene (XM_004512880).

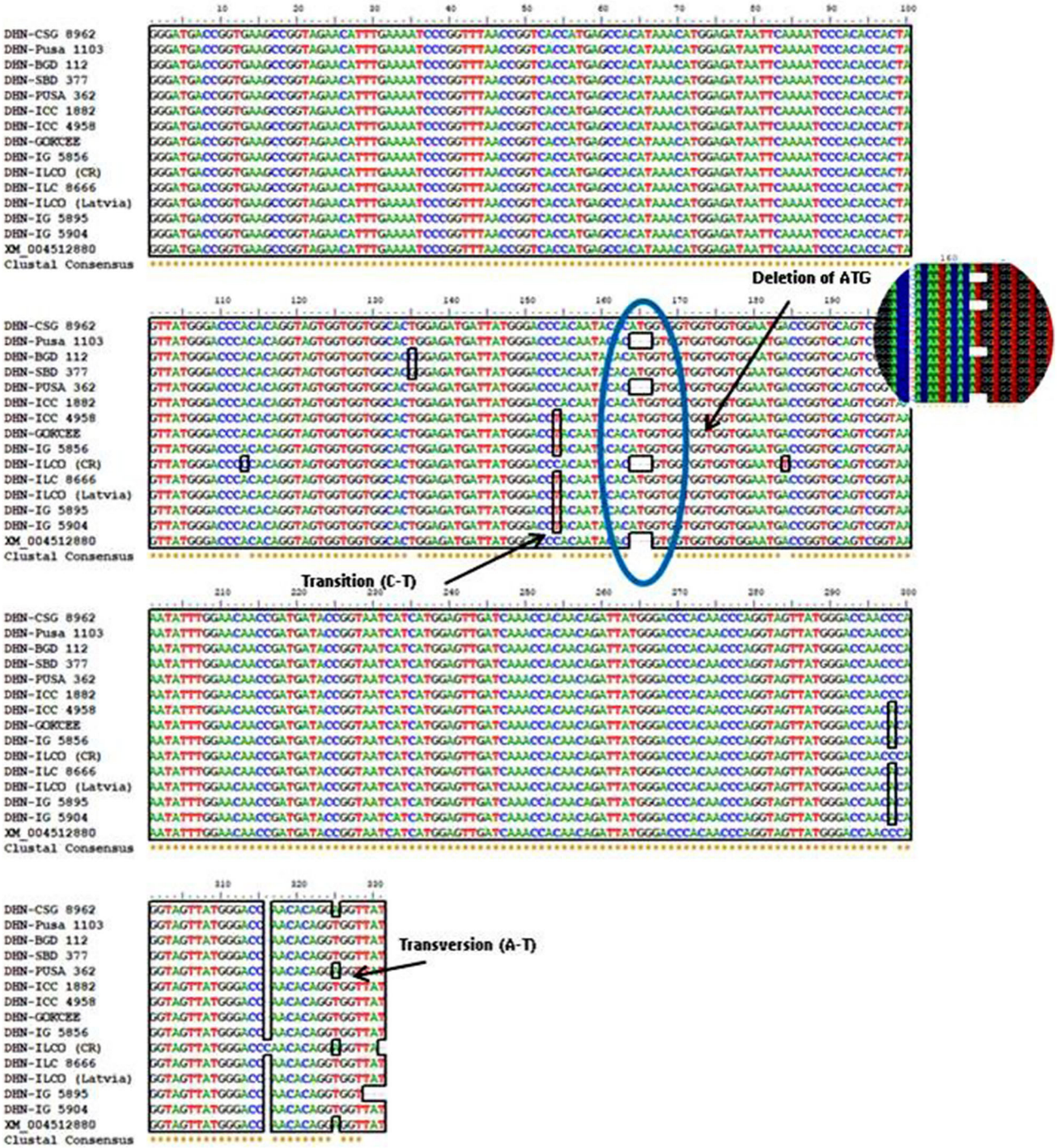


FIGURE 4 | Multiple alignments of nucleotide sequences of the *DHN* gene from different chickpea genotypes were done by using the BioEdit version 7.0.9. The presence of SNP in the *DHN* gene is been indicated.

The sequencing alignment revealed a different number of SNPs in the candidate gene among genotypes. The indels were only present in the candidate gene from genotypes Pusa1103, Pusa362, ILC0 (Latvia), and the reference *DHN* gene (XM_004512880). Also, a single nucleotide polymorphism was present in genotype Pusa362. The base by base comparison revealed that the homozygous alleles of SNPs in the position 154 and 298 are present in the drought-tolerant genotypes (ICC4958, GOKCEE, IG5856, ILC8666, ILC0 (Latvia), IG5895 and IG5904). But, in the case of genotype BGD112 and SBD377 heterozygous alleles of SNPs were found in position 135 (**Figure 6**,

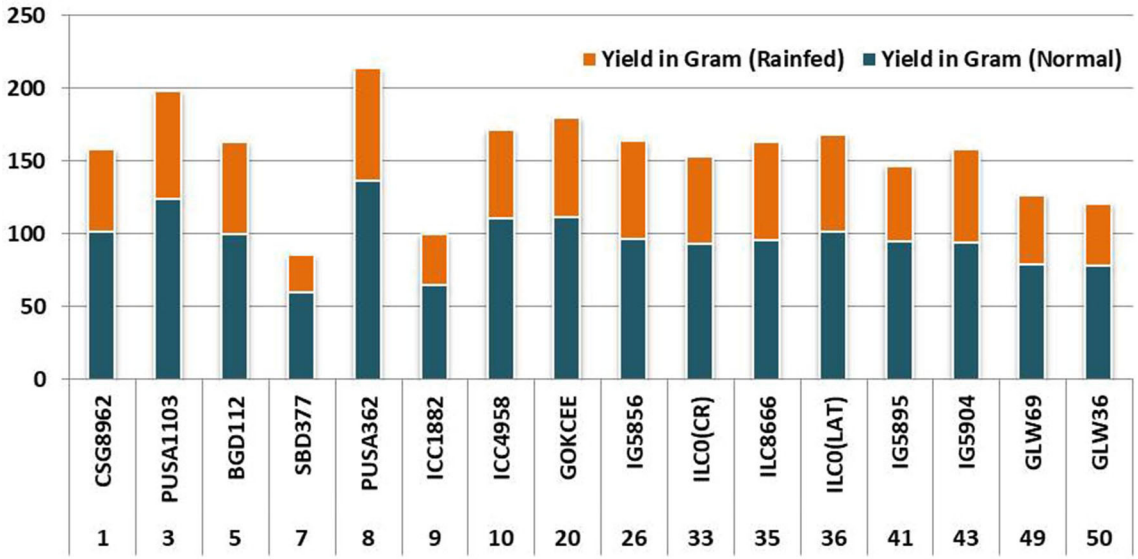


FIGURE 5 | Comparison of the yield of selected genotypes under rainfed and normal conditions.

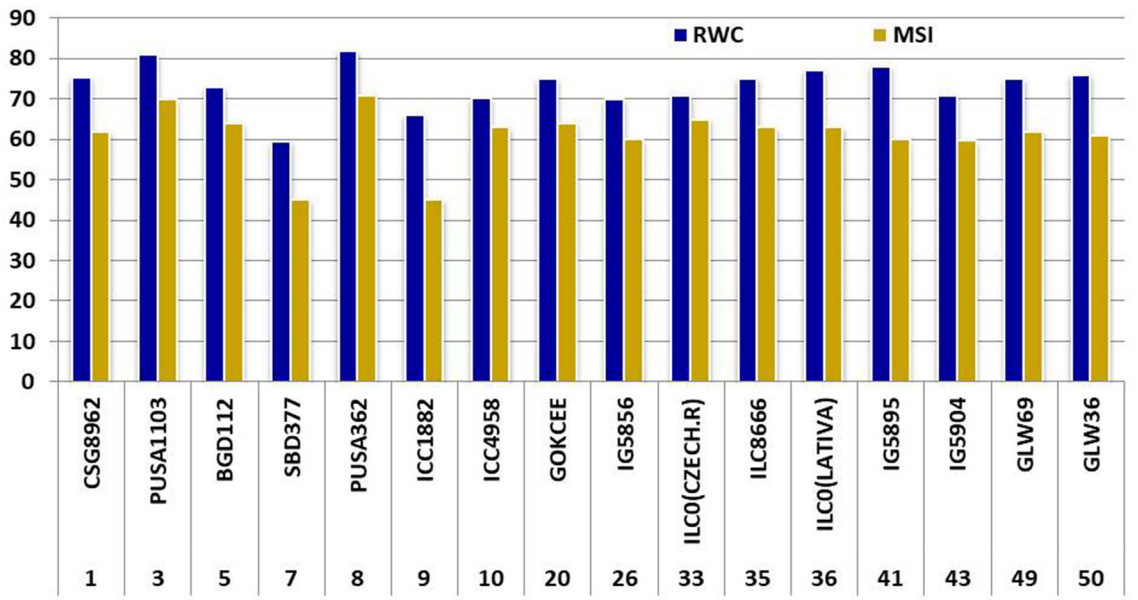


FIGURE 6 | Comparison of the RWC and MSI of selected genotypes.

Supplementary Table 2). The results suggest that the genotypes Pusa1103 and Pusa362 having homozygous indels and SNPs showed a significantly higher value of relative water content, membrane stability index, and yield in comparison to other genotypes. The gene sequences were deposited in NCBI with ID's CSG8962 (KY542275), PUSA1103 (MF469826), BGD112 (MF469827), SBD377 (MF469828), PUSA362 (KY542276), ICC1882 (MF469829), ICC4958 (MF469830), GOKCEE (MF469831), IG5856 (MF469832), ILC8666 (MF469833), ILC0 (Latvia) (MF469834), IG5895 (MF469835), IG5904 (MF469836).

DISCUSSION

The acclimation of plants to drought stress conditions is dynamic and complex, which involves hundreds of genes and their interactions with different environmental factors throughout the plant development (Kumari et al., 2009). Drought has become one of the most important constraints for chickpea production. In recent years, significant improvement by way of breeding for chickpea adaptation to the drought was achieved (Devasirvatham et al., 2012; Kumar et al., 2015). However, a gap still exists in the understanding of the physiological

and molecular mechanisms under water stress conditions. It is imperative to study the plant physiological responses under water stress conditions and to develop drought-tolerant chickpea cultivars by utilizing improved screening techniques and various modern genetic approaches. The severity of drought can be estimated by morphological features and physiological processes of plants during its growth and development like days to 50% flowering, days to maturity, plant height, pods per plant, seed per pod, relative water content, membrane stability index, yield (Toker and Cagirgan, 1998; Jaleel et al., 2009). Relative water content and membrane stability index are the best indices that can accurately indicate the balance between water absorbed by the plant and the amount consumed through transpiration. In wheat, Schonfeld et al. (1985) disclosed that the cultivars having high relative water content were more tolerant to drought stress. Ramos et al. (2003) identified significant differences in relative water content in bean leaves and the values were lesser in drought conditions than normal. Many researchers also reported characters like membrane thermostability, canopy temperature depression to be highly effective in screening for drought conditions (Leport et al., 1999).

All the tolerant genotypes had high values of relative water content and membrane stability index. The lower the difference between them, the greater the genotype has tolerance to drought. The genotypes Pusa1103 and Pusa362 not only had lower variation in the relative water content and membrane stability index values but they were also high yielding and thus are promising under normal and rainfed conditions. It is essential to explore the variation for drought indicatory parameters in crops for their effective utilization (Ali et al., 2011 and Dhanda et al., 2004). Relative water content is a function of water uptake by the roots as well as water loss by transpiration and is also considered as a pivotal index for dehydration tolerance. Drought susceptibility is a result of low relative water content in a wide variety of plants including chickpea (Jain and Chattopadhyay, 2010; Yucel and Anlarsal, 2010; Rahbarian et al., 2011). Water stress impairs both membrane structure and function of the plant cells/tissue (Cave, 1981). The cell membrane is one of the first targets of many plant stresses like drought and affects cell membrane integrity and stability (Lyevitt, 1972). Hence, the maintenance of cell membranes integrity and stability under water stress is also one of the measures for tolerance to drought (Vieira da Silva et al., 1974).

Better heritability values of traits were having more possibilities of improvement (Ahmed and Khaliq, 2007; Songsri et al., 2008). High heritability accompanied by low genetic advance for days to maturity and seeds per pods suggested that high heritability may not necessarily lead to increased genetic gain until variability present for the trait (Sardana et al., 2007). It has been accentuated that without hereditary development, the heritability esteems would not be of useful significance in choice dependent on phenotypic appearance. Thus, hereditary development should be considered along with heritability in coherent selection breeding program.

During dehydration conditions, plant cells lose water and membrane-associated complexes and proteins undergo an undesirable denaturation process. Several studies proposed that *DHN* acts as "space-filler" and it gathers relatively in larger amounts in various compartments inside the cells during dehydration (Battaglia et al., 2008). Close (1996) and Allagulova et al. (2003) have shown a positive correlation between the accumulation of *DHN* proteins and tolerance to freezing, drought, and salinity. Thus under dehydration condition dehydrin gene can help in keeping the original cell volume, preventing cellular collapse (Hanin et al., 2011). Various workers also reported the correlated responses of Dehydrin with stress tolerance viz., in chickpea (Gao et al., 2008); oat (Maqbool et al., 2002); rice (Moons et al., 1997) and tobacco (Kim et al., 2005).

DHN genes can be explored in developing superior chickpea varieties with improved yield under abiotic stress conditions Roorkiwal and Sharma, 2012. The effect of drought and the response to morpho-physiological changes can be used to identify tolerant genotypes with increased yield (Nam et al., 2001; Martinez et al., 2007). It is important to know the molecular as well as phenotypic responses under the restricted condition and to identify the suitable genotypes that respond in rainfed conditions (Upadhyaya et al., 2012). Knowledge of candidate genes for stress resilience is limiting in chickpea (Lata and Prasad, 2011). Modernization and cost-effectiveness of sequencing technology provide a rapid method for the generation of remarkable sequence data which helps to identify the genes responsible for various stress tolerance. This information would greatly aid in crop improvement through SNPs linked with the preferred trait or directly through a transgenic approach (Ray and Satya, 2011; He et al., 2014).

Sequencing-based allele mining involves PCR-based amplification of alleles of a gene in diverse genotypes, followed by DNA sequencing to identify the nucleotide variance in the alleles. Multiple alleles among the cultivars can be identified through this approach. The method enables us to analyze individuals for haplotype structure and diversity to infer genetic association studies in plants. This allows us to recognize the effect of mutations on gene structure and the location of point mutations or SNPs and insertions or deletions (InDels) to construct haplotypes. Sequencing-based allele mining is found to be an efficient approach to expand the rice blast R gene source and manage damaging blast disease (Vasudevan et al., 2015). The presence of homozygous indels and SNPs in the *DHN* gene in Pusa1103 and Pusa362 genotypes suggests that such changes can be highly associated with drought tolerance response. Furthermore, these genotypes are identified as tolerant as they maintained the higher value of relative water content, membrane stability index, and yield in comparison to other genotypes. This study will help us to identify and characterize the drought-tolerant genotypes by utilizing the morphological traits and allelic variation of the *DHN* gene by sequencing techniques and also to discover the allelic variation of the gene.

Data from the coding regions are regularly in use for the identification of genes responsible for stress tolerance from different plant species like Arabidopsis thaliana, Medicago truncatula, and many more. Identification of allelic variations in the drought-responsive candidate genes from diverse genotypes can provide genomic resources with different alleles to develop improved genotypes for drought tolerance.

The present study provides a comparative study of the candidate gene and morphological traits for drought tolerance in chickpea, which can be used in improving drought tolerance in chickpea.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, KY542276; https://www.ncbi.nlm.nih.gov/, MF469826; https://www.ncbi.nlm.nih.gov/, MF469830; https://www.ncbi.nlm.nih.gov/, MF469827; https://www.ncbi.nlm.nih.gov/, KY542275; https://www.ncbi.nlm.nih.gov/, MF469828; https://www.ncbi.nlm.nih.gov/, MF469829; https://www.ncbi.nlm.nih.gov/, MF469831; https://www.ncbi.nlm.nih.gov/, MF469832; https://www.ncbi.nlm.nih.gov/, MF469833; https://www.ncbi.nlm.nih.gov/, MF469834; https://www.ncbi.nlm.nih.gov/, MF469835; https://www.ncbi.nlm.nih.gov/, MF469836.

AUTHOR CONTRIBUTIONS

TK contributed to the analysis, interpretation of data, prepared the manuscript, and conceived the study. NT contributed to the analysis and interpretation of data. CB designed the study and reviewed and edited the manuscript. AS provided feedback and edited the manuscript. SP contributed to the analysis and interpretation of data. SS reviewed and edited the manuscript. MS provided guidance and edited the manuscript. All the authors read and approved the manuscript.

FUNDING

This study was funded by ICAR-ICARDA (Code: 24-445) collaborative project and by ICAR-NTPC (Code: 21-21) project Functional genomics for drought tolerance in chickpea and the institutional fund from ICAR-IARI.

ACKNOWLEDGMENTS

The authors are greatly acknowledged ICAR-IARI, DAC, and ICARDA for providing financial support for this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.584527/full#supplementary-material

REFERENCES

Ahmed, N. C. M., and Khaliq, I. M. M. (2007). The inheritance of yield and yield components of five wheat hybrid populations under drought conditions. *Indones. J. Agric. Sci.* 8, 53–59. doi: 10.21082/ijas.v8n2.2007.p53-59

Ali, M. A., Jabran, K., Awan, S. I., Abbas, A., Ehsanullah, Z. M., and Tuba, A. (2011). Morpho-physiological diversity and its implications for improving drought tolerance in grain sorghum at different growth stages. *Aust. J. Crop Sci.* 5, 308–317.

Allagulova, C. R., Gimalov, F. R., Shakirova, F. M., and Vakhitov, V. A. (2003). The plant dehydrins: structure and putative functions. *Biochemistry* 68, 945–951. doi: 10.1023/A:1026077825584

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., and Covarrubias, A. A. (2008). Review: the enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* 148, 6–24. doi: 10.1104/pp.108.120725

Begg, J. E., and Turner, N. C. (1976). Crop water deficits. *Adv. Agron.* 28, 161–217. doi: 10.1016/S0065-2113(08)60555-6

Bharadwaj, C., Srivastava, R., Chauhan, S. K., Satyavathi, C. T., Kumar, J., Faruqui, A., et al. (2011). Molecular diversity and phylogeny in geographical collection of chickpea (*Cicer* sp.) accessions. *J. Genet.* 90, 94–100. doi: 10.1007/s12041-011-0114-6

Blum, A., and Ebercon, A. (1981). Cell membrane stability as a measure of drought and heat tolerance in wheat. *Crop Sci.* 21, 43–47. doi: 10.2135/cropsci1981.0011183X002100010013x

Boyer, J. S., and McPherson, H. G. (1975). Physiology of water deficits in cereal crops. *Adv. Agron.* 27, 1–23. doi: 10.1016/S0065-2113(08)70006-3

Castro, A. M., Vilanova, S., Canizares, J., Pascual, L., Blanca, J. M., Diez, M. J., et al. (2012). Application of genomic tools in plant breeding. *Curr. Genomics* 13, 179–195. doi: 10.2174/138920212800543084

Cave, G. (1981). Water and membranes: the interdependence of their physico-chemical properties in the case of phospholipids head groups. *Studiabiophysica* 91, 41–46.

Close, T. J. (1996). Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiol. Plant.* 97, 795–803. doi: 10.1111/j.1399-3054.1996.tb00546.x

Deshmukh, R., Sonah, H., Patil, G., Chen, W., Prince, S., Mutava, R., et al. (2014). Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* 5:244. doi: 10.3389/fpls.2014.00244

Devasirvatham, V., Gaur, P. M., Mallikarjuna, N., Raju, T. N., Trethowan, R. M., and Tan, D. K. Y. (2012). Effect of high temperature on the reproductive development of chickpea genotypes under controlled environments. *Funct. Plant Biol.* 39, 1009–1018. doi: 10.1071/FP12033

Dhanda, S. S., Sethi, G. S., and Behl, R. K. (2004). Indices of drought tolerance in wheat genotypes at early stages of plant growth. *J. Agron. Crop Sci.* 190, 6–12. doi: 10.1111/j.1439-037X.2004.00592.x

Gao, S., Zhang, H., Tian, Y., Li, F., Zhang, Z., and Lu, X. (2008). Expression of TERF1 in rice regulates expression of stress-responsive genes and enhances tolerance to drought and high-salinity. *Plant Cell Rep.* 27, 1787–1795. doi: 10.1007/s00299-008-0602-1

Gaur, P. M., Jukanti, A. K., Samineni, S., Chaturvedi, S. K., Basu, P. S., Babbar, A., et al. (2012a). "Climate change and heat stress tolerance in chickpea" in *Climate Change and Plant Abiotic Stress Tolerance*, eds N. Tuteja and S. S. Gill (Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA), 839–855. doi: 10.1002/9783527675265.ch31

Gaur, P. M., Jukanti, A. K., and Varshney, R. K. (2012b). Impact of genomic technologies on chickpea breeding strategies. *Agronomy* 2, 199–221. doi: 10.3390/agronomy2030199

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal Behav.* 6, 1503–1509. doi: 10.4161/psb.6.10.17088

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H. K., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5:484. doi: 10.3389/fpls.2014.00484

Hsiao, T. C. (1973). Plant Responses to water stress. *Ann. Rev. Pl. Physiol.* 24, 519–570. doi: 10.1146/annurev.pp.24.060173.002511

Jain, D., and Chattopadhyay, D. (2010). Analysis of gene expression in response to water deficit of chickpea (*Cicer arietinum* L.) varieties differing in drought tolerance. *BMC Plant Biol.* 10:24. doi: 10.1186/1471-2229-10-24

Jaleel, C. A., Manivannan, P., Wahid, A., Farooq, M., Al-Juburi, H. J., Somasundaram, R., et al. (2009). Drought stress in plants: a review on morphological characteristics and pigments composition. *Int. J. Agric. Biol.* 11, 100–105.

Kim, H. S., Lee, J. H., Kim, J. J., Kim, C. H., Jun, S. S., and Hong, Y. N. (2005). Molecular and functional characterization of CaLEA6, the gene for a hydrophobic LEA protein from *Capsicum annuum*. *Gene* 344, 115–123. doi: 10.1016/j.gene.2004.09.012

Kumar, J., and Abbo, S. (2001). Genetics of flowering time in chickpea and its bearing on productivity in semiarid environments. *Adv. Agron.* 72, 107–13810. doi: 10.1016/S0065-2113(01)72012-3

Kumar, T., Bharadwaj, C., Rizvi, A. H., Sarker, A., Tripathi, S., Alam, A., et al. (2015). Chickpea landraces: a valuable and divergent source for drought tolerance. *Int. J. Tropi. Agri.* 33, 633–638.

Kumar, T., Bharadwaj, C., Tiwari, N., Satyavathi, C. T., Patil, B. S., Sarker, A., et al. (2018). Morphological characterization and grouping of chickpea genotypes for drought tolerance. *Indian J. Agric. Sci.* 88, 1740–1745.

Kumari, S., Sabharwal, V. P., Kushwaha, H. R., Sopory, S. K., Singla-Pareek, S. L., and Pareek, A. (2009). Transcriptome map for seedling stage specific salinity stress response indicates a specific set of genes as candidate for saline tolerance in *Oryza sativa* L. *Funct. Integr. Genomics* 9, 109–123. doi: 10.1007/s10142-008-0088-5

Lata, C., and Prasad, M. (2011). Role of DREBs in regulation of abiotic stress responses in plants. *J. Exp. Bot.* 62, 4731–4748. doi: 10.1093/jxb/err210

Leport, L., Turner, N. C., French, R. J., Barr, M. D., Duda, R., Davies, S. L., et al. (1999). Physiological responses of chickpea genotypes to terminal drought in a Mediterranean-type environment. *Eur. J. Agron.* 11, 279–291. doi: 10.1016/S1161-0301(99)00039-8

Lyevitt, J. (1972). *Responses of Plants to Environmental stresses*. New York, NY: Academic Press.

Maqbool, B., Zhong, H., El-Maghraby, Y., Ahmad, A., Chai, B., Wang, W., et al. (2002). Competence of oat (*Avena sativa* L.) shoot apical meristems for integrative transformation, inherited expression, and osmotic tolerance of transgenic lines containing hva1. *Theor. Appl. Genet.* 105, 201–208. doi: 10.1007/s00122-002-0984-3

Maqbool, M. A., Aslam, M., and Ali, H. (2017). Breeding for improved drought tolerance in chickpea (*Cicer arietinum* L.). *Plant Breed.* 136, 300–318. doi: 10.1111/pbr.12477

Martinez, J. P., Silva, H., Ledent, J. F., and Pinto, M. (2007). Effect of drought stress on the osmotic adjustment, cell wall elasticity and cell volume of six cultivars of common beans (*Phaseolus vulgaris* L.). *Eur. J. Agron.* 26, 30–38. doi: 10.1016/j.eja.2006.08.003

Mohammed, A., Tana, T., Singh, P., Korecha, D., and Mollad, A. (2017). Management options for rainfed chickpea (*Cicer arietinum* L.) in northeast Ethiopia under climate change condition. *Clim. Risk. Manag.* 16, 222–233. doi: 10.1016/j.crm.2016.12.003

Moons, A., De Keyser, A., and Van Montagu, M. (1997). A group 3 LEA cDNA of rice, responsive to abscisic acid, but not to jasmonic acid, shows variety-specific differences in salt stress response. *Gene* 191, 197–204. doi: 10.1016/S0378-1119(97)00059-0

Nam, N. H., Chauhan, Y. S., and Johansen, C. (2001). Effect of timing of drought stress on growth and grain yield of extra-short-duration pigeonpea lines. *J. Agric. Sci.* 136, 179–189. doi: 10.1017/S0021859601008607

Perrier, X. and Jacquemoud-Collet, J. P. (2006) DARwin Software. Available online at: http://darwin.cirad.fr/darwin.

Puhakainen, T., Hess, M. W., Makela, P., Svensson, J., Heino, P., and Palva, E. T. (2004a). Over expression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. *Plant Mol. Biol.* 54, 743–753. doi: 10.1023/B:PLAN.0000040903.66496.a4

Puhakainen, T., Li, C., Boije-Malm, M., Kangasjarvi, J., Heino, P., and Palva, E. T. (2004b). Short-day potentiation of low temperature-induced gene expression of a C-repeat-binding factor-controlled gene during cold acclimation in silver birch. *Plant Physiol.* 136, 4299–4307. doi: 10.1104/pp.104.047258

Rahbarian, R., Khavari-Nejad, R., Ganjeali, A., Bagheri, A., and Najafi, F. (2011). Drought stress effects on photosynthesis, chlorophyll fluorescence and water relations in tolerant and susceptible Chickpea (*Cicer Arietinum* L.) genotypes. *Acta Biol. Cracov. Ser. Bot.* 53, 47–56. doi: 10.2478/v10182-011-0007-2

Ramos, G., Parsons, R., Sprent, J., and James, E. K. (2003). Effect of water stress on nitrogen fixation and nodule structure of common bean. *Pesqui. Agropecuaria Bras. Brasilia* 38, 339–347. doi: 10.1590/S0100-204X2003000300002

Ray, S., and Satya, P. (2011). Next generation sequencing technologies for next generation plant breeding. *Front. Plant Sci.* 5:367. doi: 10.3389/fpls.2014.00367

Roorkiwal, M., Nayak, S. N., Thudi, M., Upadhyaya, H. D., Brunel, D., Mournet, P., et al. (2014). Allele diversity for abiotic stress responsive candidate genes in chickpea reference set using gene based SNP markers. *Front. Plant Sci.* 5:248. doi: 10.3389/fpls.2014.00248

Roorkiwal, M., and Sharma, P. C. (2012). Sequence similarity based identification of abiotic stress responsive genes in chickpea. *Bioinformation* 8, 92–97. doi: 10.6026/97320630008092

Sardana, S., Mahjan, R., Gautam, N., and Ram, B. (2007). Genetic variability in pea (*Pisum sativum* L.) germplasm for utilization. *SABRAO. J. Breed. Genet.* 39, 31–41.

Schonfeld, M. A., Johnson, R. C., Carver, B. F., and Mornhinweg, D. W. (1985). Water relations in winter wheat as drought resistance indicator. *Crop Sci.* 28, 526–531. doi: 10.2135/cropsci1988.0011183X002800030021x

Shah, T. M., Imran, M., Atta, B. M., Ashraf, M. Y., Hameed, A., Waqar, I., et al. (2020). Selection and screening of drought tolerant high yielding chickpea genotypes based on physio-biochemical indices and multi-environmental yield trials. *BMC Plant Biol.* 20:171. doi: 10.1186/s12870-020-02381-9

Songsri, P., Joglloy, S., Kesmala, T., Vorasoot, N., Akkasaeng, C. P. A., and Holbrook, C. (2008). Heritability of drought resistance traits and correlation of drought resistance and agronomic traits in peanut. *Crop Sci.* 48, 2245–2253. doi: 10.2135/cropsci2008.04.0228

Toker, C., and Cagirgan, M. (1998). Assessment of response to drought stress of chickpea (*Cicer arietinum* L.) lines under rain field conditions. *Turk. J. Agri. Forestry* 22, 615–621.

Toker, C., Lluch, C., Tejera, N. A., Serraj, R., and Siddique, K. H. M. (2007). "Abiotic Stresses," in *Chickpea Breeding and Management*, eds S. S. Yadav, R. Redden, W. Chen, and B. Sharma (Wallingford, CT: CAB International), 474–496. doi: 10.1079/9781845932138.023

Upadhyaya, H. D., Kashiwagi, J., Varshney, R. K., Gaur, P. M., Saxena, K. B., Krishnamurthy, L., et al. (2012). Phenotyping chickpeas and pigeonpeas for adaptation to drought. *Front. Physiol.* 3:179. doi: 10.3389/fphys.2012.00179

Upadhyaya, H. D., Thudi, M., Dronavalli, N., Gujaria, N., Singh, S., Sharma, S., et al. (2011). Genomic tools and germplasm diversity for chickpea improvement. *Plant Genet. Resour.* 9, 45–58. doi: 10.1017/S1479262110000468

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., and Sharpe, A. G. (2013). Draft 500 genome sequence of chickpea (*Cicer arietinum* L.) provides a resource for trait 501 improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491

Vasudevan, K., Gruissem, W., and Bhullar, N. (2015). Identification of novel alleles of the rice blast resistance gene Pi54. *Sci. Rep.* 5, 15678–15688. doi: 10.1038/srep15678

Vieira da Silva, J., Naylor, A. W., and Kramer, P. J. (1974). Some ultrastructural and enzymatic effects of drought stress in cotton (*Gossypiumhirsutum* L.) leaves. *Proc. Natl. Acad. Sci. U.S.A.* 71, 3243–3324. doi: 10.1073/pnas.71.8.3243

Wise, M. J., and Tunnacliffe, A. (2004). POPP the question: what do LEA protein do? *Trends Plant Sci.* 9, 13–17. doi: 10.1016/j.tplants.2003.10.012

Yucel, D. O., and Anlarsal, A. E. (2010). Determination of selection criteria with path coefficient analysis in chickpea (*Cicer arietinum* L.) breeding. *Bulgarian J. Agric. Sci.* 16, 42–48.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kumar, Tiwari, Bharadwaj, Sarker, Pappula, Singh and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 December 2020
doi: 10.3389/fgene.2020.598183

Genome-Wide Identification and Capsaicinoid Biosynthesis-Related Expression Analysis of the *R2R3-MYB* Gene Family in *Capsicum annuum* L.

Jin Wang[1,2], *Yi Liu*[3], *Bingqian Tang*[3], *Xiongze Dai*[2,4], *Lingling Xie*[5], *Feng Liu*[4,5]* and *Xuexiao Zou*[2,4]*

[1] College of Horticulture, Nanjing Agricultural University, Nanjing, China, [2] Engineering Research Center for Horticultural Crop Germplasm Creation and New Variety Breeding, Ministry of Education, Changsha, China, [3] Longping Branch, Graduate School of Hunan University, Changsha, China, [4] College of Horticulture, Hunan Agricultural University, Changsha, China, [5] Hunan Vegetable Research Institute, Changsha, China

OPEN ACCESS

Edited by:
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Satinder Kaur,
Punjab Agricultural University, India
Douglas S. Domingues,
São Paulo State University, Brazil

***Correspondence:**
Feng Liu
liufengrich@126.com
Xuexiao Zou
zouxuexiao428@163.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 24 August 2020
Accepted: 30 November 2020
Published: 21 December 2020

Citation:
Wang J, Liu Y, Tang B, Dai X, Xie L, Liu F and Zou X (2020) Genome-Wide Identification and Capsaicinoid Biosynthesis-Related Expression Analysis of the R2R3-MYB Gene Family in Capsicum annuum L. Front. Genet. 11:598183. doi: 10.3389/fgene.2020.598183

Capsaicinoids are naturally specialized metabolites in pepper and are the main reason that Capsicum fruits have a pungent smell. During the synthesis of capsaicin, MYB transcription factors play key regulatory roles. In particular, *R2R3-MYB* subfamily genes are the most important members of the MYB family and are critical candidate factors in capsaicinoid biosynthesis. The 108 *R2R3-MYB* genes in pepper were identified in this study and all are shown to have two highly conserved MYB binding domains. Phylogenetic and structural analyses clustered *CaR2R3-MYB* genes into seven groups. Interspecies collinearity analysis found that the *R2R3-MYB* family contains 16 duplicated gene pairs and the highest gene density is on chromosome 00 and 03. The expression levels of *CaR2R3-MYB* differentially expressed genes (DEGs) and capsaicinoid-biosynthetic genes (CBGs) in fruit development stages were obtained *via* RNA-seq and quantitative polymerase chain reaction (qRT-PCR). Co-expression analyses reveal that highly expressed *CaR2R3-MYB* genes are co-expressed with CBGs during early stages of pericarp and placenta development processes. It is speculated that six candidate *CaR2R3-MYB* genes are involved in regulating the synthesis of capsaicin and dihydrocapsaicin. This study is the first systematic analysis of the *CaR2R3-MYB* gene family and provided references for studying their molecular functions. At the same time, these results also laid the foundation for further research on the capsaicin characteristics of *CaR2R3-MYB* genes in pepper.

Keywords: Capsicum, *CaR2R3-MYB* family, capsaicinoid biosynthesis, expression analysis, co-expression

INTRODUCTION

Secondary metabolites are important compounds in plants to resist stress, deter herbivores, and prevent attack from some pathogens (Tewksbury et al., 2008). It is clear that plants have evolved their own special secondary metabolites on the basis of adaptations to the surrounding environment. Amongst these, the unique defensive chemical compounds produced by

Capsicum, including capsaicin (CAP), dihydrocapsaicin (DhCAP), and several analogs, collectively known as capsaicinoids (CAPDs), are the most widely involved (Stewart et al., 2007). It is also known that CAPDs, unique flavoring substances in chili peppers, make peppers spicy but also influence the synthesis and accumulation of volatile aroma substances (Aza-González et al., 2011). In addition to self-protection, CPADs are also widely applied across industries including food, pharmaceuticals, and medical areas (Naves et al., 2019).

Thus, regarding CPAD biosynthetic pathways, CAP and DhCAP account for nearly 90% in pepper species, divided into phenylpropanoid and branched chain fatty acid pathways (Choi et al., 2006). One specific approach is to synthesize capsaicin by condensing vanillylamine molecules and to derive this compound from phenylalanine *via* branched chain fatty acids (between 9 and 11 carbon atoms), themselves synthesized from either valine or leucine (Arce-Rodriguez and Ochoa-Alejo, 2019). Indeed, as sequencing technology has developed, studies have revealed that Capsicum fruit biosynthesis is strongly influenced by genotype-environment interactions (Qin et al., 2014). Capsaicinoid-biosynthetic genes (CBGs) are expressed preferentially as typical response factors, specifically in the pericarp and placenta, during pepper fruit development processes (Liu et al., 2013). Studies have identified several structural CBGs [such as, *CoMT*, *C4H*, *AT3*, *KAS*, putative aminotransferase (*pAMT*), and *Acl*] involved in capsaicinoid biosynthesis (Von Wettsteinknowles et al., 2000); these accumulate in epidermal cell vesicles in placental tissue and start accumulating between 10 and 20 days post anthesis (DPA), increasing between 20 and 40 DPA (Arce-Rodriguez and Ochoa-Alejo, 2017). Orthologous genes in the pathways of other solanaceous plants (e.g., tomato and potato) are rarely expressed at this stage (Kim et al., 2014). Genetic studies have revealed that two leaky *pAMT* alleles (*pamtL1* and *pamtL2*) as well as a loss-of-function *pAMT* allele reduce capsaicinoid levels (Tanaka et al., 2019), while mutations in acyltransferase (*Pun1*) and *pAMT* lead to disruption of the capsaicinoid biosynthesis putative gene ketoacyl-ACP reductase (*CaKR1*) and a loss of pungency (Koeda et al., 2019). It is also clear that *Pun1* encodes an acyltransferase necessary to biosynthesize capsaicinoid (Stewart et al., 2005), and silenced *AT3* negatively influences the transcription of CBGs (Arce-Rodriguez and Ochoa-Alejo, 2015). The bulk of CBGs exhibit tissue- and stage-specific expressions accompanying the gradual accumulation of capsaicinoids. The transcription factors *Erf* and *Jerf* within the complex ERF family are expressed early in fruit development and participate in regulation of the pungency phenotype in chili (Keyhaninejad et al., 2014). These observations show that transcription factors also participate and play key regulatory roles in capsaicin pathway synthesis and metabolism.

Myeloblastosis (MYB) is one of the most important and the largest transcription factor gene families (Dubos et al., 2002). The MYB gene is divided into four subfamilies based on incomplete MYB domain repeats (R), each containing about 52 amino acid residues. This group includes the 4R-MYB, 3R-MYB, R2R3-MYB, and MYB-related subfamilies which each contains a single or partial MYB-related repeat, respectively (Jia et al., 2004). Specifically, R2R3-MYB is the dominant subfamily, occurring in the largest numbers in most plants (Rosinski and Atchley, 1998). Different MYB-type family members have been identified in many species, including in *Arabidopsis thaliana* (196 members) (Dubos et al., 2002), and watermelon (*Citrullus lanatus*) (162 members, of which 89 are *R2R3-MYB* type genes) (Wang et al., 2020). Similarly, 559 *R2R3-MYB*s have been identified in Solanaceae, including 119 complete sequences in tomato (*Lycopersicon esculentum* Mill.) (Gates et al., 2016). These genes have a wide range of functions and play pivotal regulatory roles in the synthesis of capsaicin. Methyl jasmonate induced *CaMYB108* is also involved in the regulation of capsaicin biosynthesis and stamen development (Sun et al., 2019), while the silencing of this gene significantly reduces the expression of CBGs and capsaicinoid content. These observations showed that *MYB* genes are widely involved in the regulation of capsaicinoid biosynthetic pathway structural genes (Arce-Rodriguez and Ochoa-Alejo, 2017). Natural variations *MYB31* and its elite allele *WRKY9* can served as transcription regulation direct targets for pepper pungency levels. These pathways have determined the evolution of extremely pungent peppers (Zhu et al., 2019).

Currently, it remains unclear whether, or not, the members of the *R2R3-MYB* family have more genes involved in the capsaicinoid biosynthesis process in pepper (*Capsicum annuum* L.) and its regulatory network. Thus, 108 *CaR2R3-MYB* genes were identified in this study in both CM334 pepper and "Zunla-1" pepper genomes. Expression profiles in the pericarp and placenta were determined during fruit development, and co-expression networks of *CaR2R3-MYB* genes and CBGs were associated with gene structures, phylogenetic relationships, interspecies synteny, and *cis*-element compositions. The outcomes of this analysis imply that *Capana01g000495*, *Capana02g000906*, *Capana02g003351*, *Capana07g001604*, *Capana08g000900*, and *Capana08g001690* are candidate *CaR2R3-MYB* genes involved in capsaicin biosynthesis.

MATERIALS AND METHODS

The Identification of *R2R3-MYB* Transcription Factors in Pepper

A high-quality draft genome sequence of both hot pepper *C. annuum* cv. CM334 (Criollo de Morelos 334) (*C. annuum* Cultivars in Mexico) and a Chinese inbred derivative "Zunla-1" (*C. annuum* Cultivars in China) were used as reference genomes in this study. A HMM profile of Myb_DNA-binding domain (PF00249) was downloaded from the Pfam database (Elgebali et al., 2019), while HMMER 3.0 was applied to identify *MYB* family members with *E*-values $\leq$ 0.01 threshold (Finn et al., 2011). Protein domains of R2R3-MYBs were validated *via* SMART-Normal online software (Letunic and Bork, 2018). Protein modeling was predicted using the SWISS-MODEL online tool (Schwede et al., 2003). Theoretical the isoelectric points (PI) and molecular weights (Mw) values were computed using the ExPaSy online tool (Gasteiger et al., 2003), and subcellular localization values were predicted using the Softberry service

platform-ProtComp 9.0 (Predict the sub-cellular localization for Plant proteins) online tool[1].

Gene Structure, Motifs, and Phylogenetic Analysis

The MEME v5.1.0 online tool (National Institutes of Health, Bethesda, MD, United States) was used to investigate conserved domains. Gene structures were analyzed using the Gene Structure Display Server (Hu et al., 2015). Full-length protein sequences of CaR2R3-MYB from *C. annuum* were aligned by ClustalW method, and used Gblocks[2] online website to extract the gaps. Using unrooted neighbor-joining phylogenetic tree method of MEGA-X with the bootstrap test replicated 1,000 times (Kumar et al., 2018). The genome of Chinese inbred derivative "Zunla-1" acquired from pepper databases was used as the reference genome[3].

Chromosomal Location and Synteny Analysis

MCScanX was used to perform gene synteny and collinearity analysis, with match score of 50, gap score of -3, match size of 5, and *E*-value of $1e^{-10}$ parameters to analyze and calculate in-species duplicated genes (Wang et al., 2012). The Circos based Perl approach shows both gene chromosome positions and the synteny relationship of the pepper *R2R3-MYB* family (Krzywinski et al., 2009). KaKs_Calculator 1.2 were used to estimate the synonymous (Ks) and non-synonymous (Ka) substitution rates (Zhang et al., 2006).

Cis-Elements Analysis in Promoter Regions

The Bedtools software was used to select the length of 2.0 kb upstream sequence for each gene CDS sequence from its promoter region (Quinlan and Hall, 2010), and to examine *cis*-regulatory elements of promoter sequences by PlantCARE-Search for CARE website[4]. Plots are presented using Tbtools (Chen et al., 2020).

Materials and Transcriptome Data Analysis

A high-generation inbred Capsicum line 6421 was used for pepper development experiments. The Pericarp between 10 and 60 DAP (numbered G1–G11), the placenta and seed between 10 and 15 DAP (numbered ST1 and ST2), and the placenta between 20 and 60 DAP (numbered T3–T10) were taken from pepper fruit. The raw data for the transcriptome analysis used in this study were downloaded from Pepper Hub (Liu et al., 2017). The quality of sequencing data was controlled by Fastqc (Brown et al., 2017), and Trimmomatic-0.36 was used to filter the quality of the test data and remove low-quality sequences (Bolger et al., 2014). HISAT2 was used to compare two terminal sequencing reads to the reference genome of "Zunla-1" (Kim et al., 2014). The number of counts was calculated by using FeatureCounts (Yang et al., 2013). The R v3.6.1 language package DESeq2 was used to standardize counts data (Varet et al., 2016). FPKM (fragments per kilobase of transcript per million mapped reads) values were calculated to represent the gene expression.

Co-expression Analysis Based on RNA-Seq Data

The weighted Gene Co-Expression Network Analysis (WGCNA) package was used in R v3.6.1 language[5]. RNA-seq data were used to perform WGCNA analysis. The weighted gene correlation network analysis (WGCNA) method was used to construct a co-expression network. WGCNA analyzes the gene expression patterns of multiple samples through gene expression data (Langfelder and Horvath, 2008). By calculating the adjacent order function formed by the gene network and the difference coefficients of different nodes, the TOM similarity algorithm calculates the co-expression correlation matrix to express the gene correlation in the network. The correlation network diagram is drawn by extracting the non-weight coefficients (weight) of related *CaR2R3-MYB* and *CBGs* in the matrix. Cytoscape v3.6.0 was used to reveal a co-expression plot (Shannon et al., 2003).

Real Time Fluorescence Quantitative PCR (qRT-PCR)

Total RNA extraction was carried out using the TransZol kit (TransGen Biotech, Inc., Beijing, China). cDNA reverse transcription refers to use of the HiScript®IIQ RT SuperMix for qPCR (+gDNA wiper) vazyme kit (Vazyme, Piscataway, NJ, United States). quantitative polymerase chain reaction (qRT-PCR) was carried out in LightCycle * 96 Real-Time PCR System (Roche, Basel, Switzerland) with 25 μL reaction system. Three biological repeats and three technical repeats were used to calculate the relative quantification according to the Ct values collected by the instrument. The formula is: $2^{-\Delta\Delta Ct} = 2^{-[(Target\ gene\ control\ Ct - Target\ gene\ sample\ Ct) - (Reference\ gene\ control\ Ct - Reference\ gene\ sample\ Ct)]}$.

The actin gene *Capana04g001698* was used as reference gene which was selected from pepper. The primers of six *CaR2R3-MYB* DEGs and four CBGs were developed by GenScript Real-time PCR (TaqMan) Primer and Probes Design Tool[6], which were listed in **Supplementary Table 3**.

RESULTS

Genome-Wide Identification of CaR2R3-MYB Genes

On the basis of a Hidden Markov Model (HMM) MYB profile, there were 216 *CaMYB* genes in both CM334 and Zunla-1

[1] Softberry: http://www.softberry.com

[2] Gblocks: http://molevol.cmima.csic.es/castresana/Gblocks_server.html

[3] "Zunla-1" Reference Genome: http://peppersequence.genomics.cn/

[4] PlantCARE: http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

[5] WGCNA-R package: https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/index.html

[6] GenScript: https://www.genscript.com/tools/real-time-pcr-taqman-primer-design-tool

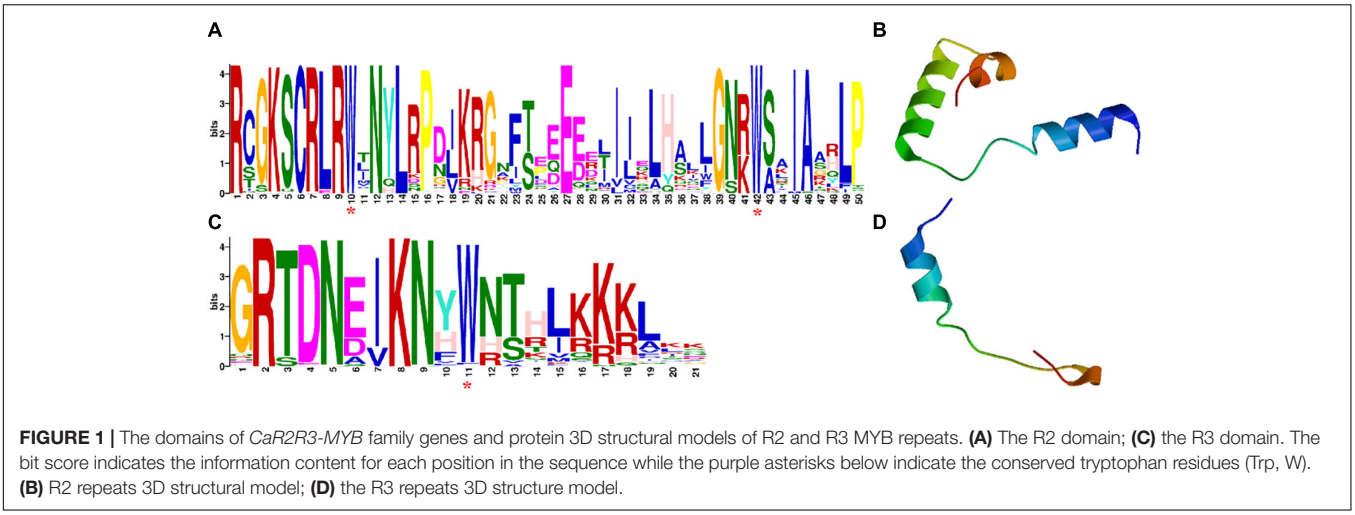


FIGURE 1 | The domains of *CaR2R3-MYB* family genes and protein 3D structural models of R2 and R3 MYB repeats. **(A)** The R2 domain; **(C)** the R3 domain. The bit score indicates the information content for each position in the sequence while the purple asterisks below indicate the conserved tryptophan residues (Trp, W). **(B)** R2 repeats 3D structural model; **(D)** the R3 repeats 3D structure model.

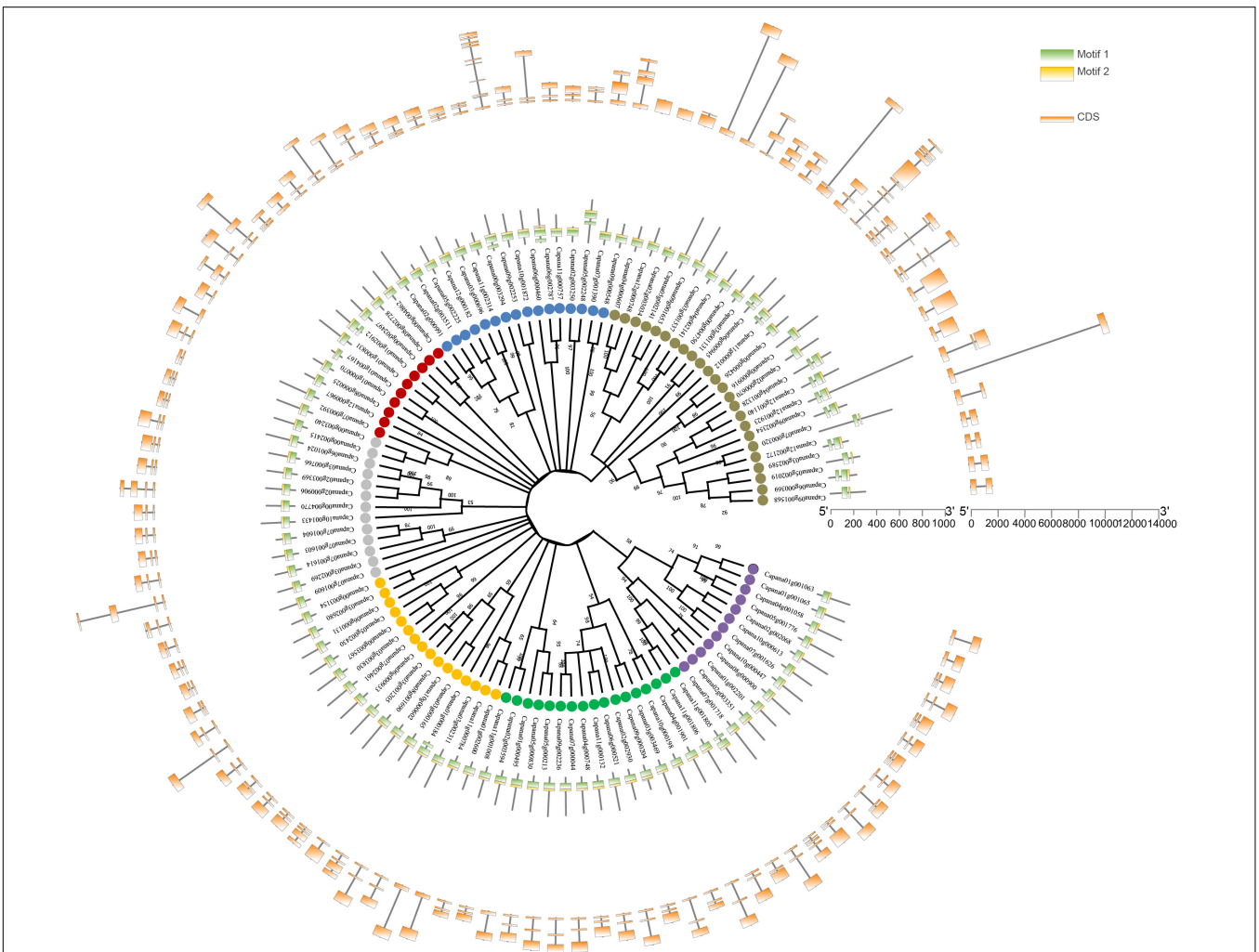


FIGURE 2 | Phylogenetic relationships, conserved motifs, and gene structural analysis of *CaR2R3-MYBs*. Phylogenetic tree in the middle of 108 *CaR2R3-MYBs*. This is an unrooted phylogenetic tree constructed via the neighbor-joining (NJ) method with 1,000 bootstrap replicates. Colored dots distinguish seven groups, the distributions of conserved motifs in *CaR2R3-MYBs*. Two conserved putative motifs are indicated with green and yellow boxes. Exon/intron organization in *CaR2R3-MYBs* is in the outermost of the circle. Orange boxes represent coding sequence (CDS) regions while black lines show intron regions. Exon length can be inferred via the scale on the right side.

genomic databases. Amongst *CaMYB* genes, 108 *R2R3* type, two *R3* type, one *R4* type, and 105 *MYB-related* types were further classified by searching both Pfam and SMART databases within the pepper genome. Further, comparing *CaR2R3-MYB*s between Zunla-1 and CM334, 32 homologous gene pairs exhibited different chromosomal annotation information. *R2R3* type genes were selected for further analysis and dubbed *CaR2R3-MYB*. Thus, Zunla-1 *CaR2R3-MYB*s were mainly used for the remaining analysis of this study.

All *CaR2R3-MYB*s contain two highly conserved MYB binding domains. The motif logo of *CaR2R3-MYB*s has 50 amino acid residues in the R2 repeat and 21 amino acid residues in the R3 repeat, respectively (**Figures 1A,C**). The HTH structure of these two domains as revealed by three-dimensional (3D) protein structural models showed that *CaR2R3-MYB*s matches the typical characteristics of the *R2R3-MYB* family (**Figures 1B,D**). **Supplementary Table 1** showed that the PI of *CaR2R3-MYB*s range between 4.76 (*Capana06g000131*) and 10.18 (*Capana05g002248*), while the Mw range between 12.6148 KD (*Capana07g000392*) and 109.76643 KD (*Capana11g000012*). Subcellular localization prediction revealed that 101 *CaR2R3-MYB*s are located in the nuclear while three *CaR2R3-MYB*s are located in the cytoplasmic, three *CaR2R3-MYB*s are located in the mitochondrial region, and one *CaR2R3-MYB* (*Capana01g002201*) is located in the extracellular zone. There were 93.5% *CaR2R3-MYB*s are transcription factors, the functions of 5.6% *CaR2R3-MYB*s were related to the cytoplasm and mitochondrial organelles.

Gene Structure, Motifs, and Phylogenetic Relationships of the *CaR2R3-MYB* Family

On the basis of statistically of high bootstrapping values, 108 *CaR2R3-MYB*s were separated into seven main groups in the unrooted phylogenetic tree based on protein sequences. All *CaR2R3-MYB* genes contained highly conserved MYB binding domains with two typical motifs, R1 and R2. **Figure 2** showed

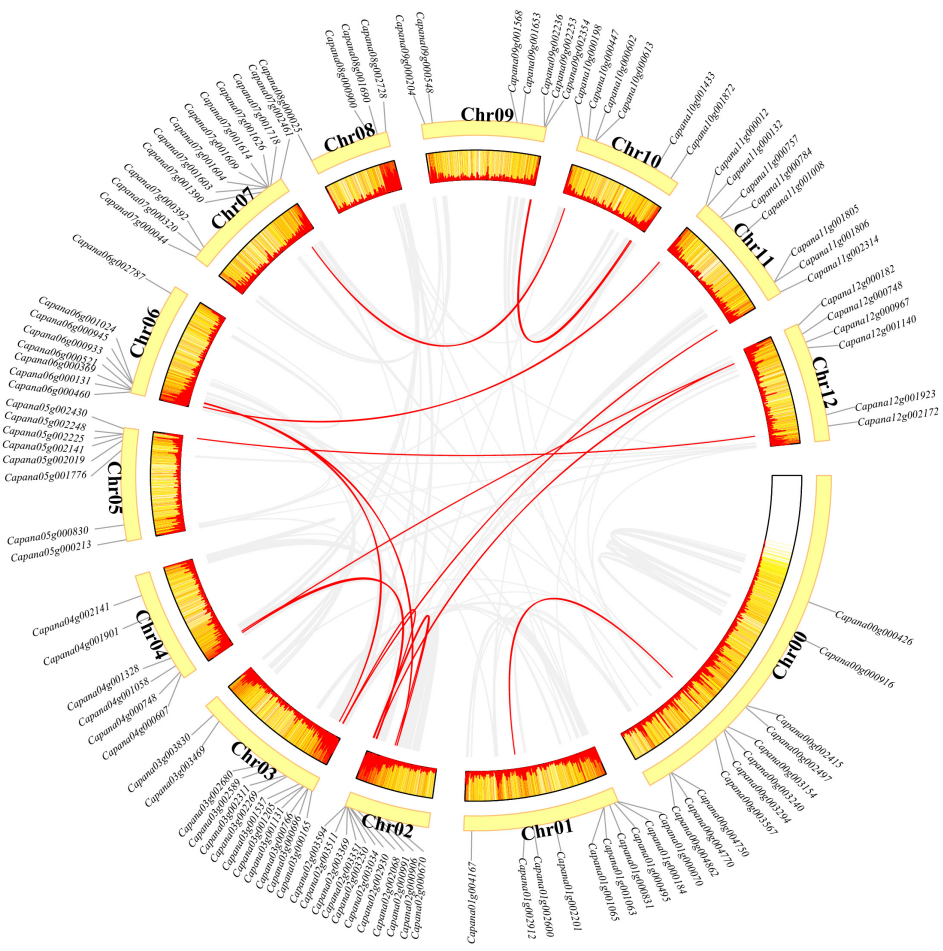


FIGURE 3 | Chromosomal location, gene density, and interspecies synteny of *CaR2R3-MYB*s. The positions of *CaR2R3-MYB* genes in the pepper genome are marked on chromosomes. A heatmap shows the gene density of each chromosome. Red lines in the middle indicate duplication gene pairs of *CaR2R3-MYB*s, while grey lines indicate genome duplication gene pairs, and Chr refers to chromosome.

TABLE 1 | Duplication models for *CaR2R3-MYB* gene pairs in pepper.

Duplicate gene pair	Ka	Ks	Ka/Ks	AverageS-sites	AverageN-sites
Capana00g003276-Capana01g002912	0.191183	0.719625	0.26567	72.58333	245.4167
Capana02g000906-Capana02g003369	0.232616	0.816092	0.285037	189.6667	683.3333
Capana02g000991-Capana02g003511	0.124785	0.635704	0.196295	120.3333	464.6667
Capana02g003369-Capana03g000766	0.370969	2.304326	0.160988	193.75	688.25
Capana02g003034-Capana04g000607	0.275051	1.776355	0.15484	199.0833	610.9167
Capana02g002930-Capana04g000748	0.494677	1.560429	0.317013	196.6667	709.3333
Capana02g000906-Capana06g001024	0.390622	1.592982	0.245214	214.9167	772.0833
Capana02g003034-Capana12g000748	0.453205	1.669887	0.271399	161.6667	546.3333
Capana03g001205-Capana06g000933	0.236095	0.46796	0.50452	178.3333	700.6667
Capana03g000696-Capana11g002314	0.469324	NaN	NaN	164.25	654.75
Capana04g000607-Capana12g000748	0.251881	1.535706	0.164016	163.1667	562.8333
Capana05g002444-Capana12g002172	0.20509	1.593561	0.128699	124.1667	445.8333
Capana06g000521-Capana11g000132	0.359847	0.854725	0.421009	174.1667	689.8333
Capana07g001626-Capana10g000447	0.15329	0.858551	0.178545	199.6667	742.3333
Capana09g002253-Capana10g001872	0.143859	0.747546	0.192441	79.25	301.75
Capana09g002354-Capana10g001956	0.30921	1.488141	0.207783	220.4167	889.5833

Ks, Ka, and Ka/Ks values are shown.

the gene structure of exon-intron compositions on the outermost side of the circle. The numbers of exons range between 1 and 10 in *CaR2R3-MYB*s. Among them, 71 (65.7%) *CaR2R3-MYB*s have three exons, 22 (20.4%) *CaR2R3-MYB*s have two exons, seven *CaR2R3-MYB*s have four exons, four *CaR2R3-MYB*s have five exons, two *CaR2R3-MYB*s have five exons, and 10 exons and 11 exons have one *CaR2R3-MYB* each, respectively. These results revealed a high degree of sequence diversity which indicated that *CaR2R3-MYBs* may be related to formation mechanisms and evolutionary processes.

Chromosomal Location and Interspecies Synteny Analysis

All *CaR2R3-MYB* genes were mapped onto the 12 different chromosomes of the pepper genome including the unclear information "00g" chromosome and "01g" to "12g" chromosomes. Chromosome 00 and chromosome 03 contained most *CaR2R3-MYBs* (12 genes), while chromosome 01 had 10 *CaR2R3-MYBs*, chromosome 02 and chromosome 07 harbored 11 *CaR2R3-MYBs*, chromosome 04, 10, and 12 contained six *CaR2R3-MYB* genes, chromosome 05, 06, and 11 had eight *CaR2R3-MYB*s, chromosome 08 harbored four *CaR2R3-MYBs*, and chromosome 09 contained seven *CaR2R3-MYBs* (**Figure 3**). A heatmap showed that gene density on chromosome 03 is the highest and no annotation genes are present at the front of chromosome 00.

The circos plot also revealed that 16 *CaR2R3-MYB* duplicated gene pairs are present. Non-synonymous mutation (Ka), synonymous mutation (Ks), and their ratios (Ka/Ks) were calculated to estimate selection pressure in duplicated genes. Ks values ranged between 0.46 and 2.3. In particular, *Capana03g000696-Capana11g002314* had no Ks value (NaN), indicating that duplication caused mutation at the nucleic acid level but that the amino acid sequence remained unchanged. The Ka/Ks values of the *CaR2R3-MYB* duplicated gene pairs ranged between 0.128 and 0.5 (**Table 1**). Indeed, all Ka/Ks values were less than 1.00 suggesting that *CaR2R3-MYB* duplicated genes have undergone purifying selection during the evolutionary process. Minimum Ks and maximum Ka/Ks values were observed between the duplicated gene pair *Capana03g001205-Capana06g000933*, indicating that these two genes might have experienced more purified selection.

CaR2R3-MYB Putative *Cis*-Elements in Promoter Regions

The 2,000 base pairs (bp) upstream *CaR2R3-MYB* genes and actin gene sequences of the coding region were used to predict *cis*-regulatory elements *via* the PlantCARE online tool (**Figure 4**). A total of eight cellular development related *cis*-regulatory elements of *CaR2R3-MYB* genes were predicted on this basis, including meristem and endosperm expression, palisade mesophyll cells, flavonoid biosynthetic genes regulation, cell cycle regulation, and seed-specific regulation. There were 13 hormone-related *cis*-regulatory elements are also present, including abscisic acid, auxin, MeJA-, gibberellin-, and salicylic acid responsiveness as well as zein metabolism regulation. Similarly, 19 stress related *cis*-elements were also identified including light responsive elements, anaerobic induction, circadian control, anoxic specific inducibility, low-temperature responsiveness, defense and stress responsiveness, and wound-responsiveness. MBS and MRE are specifically MYB binding sites involved in drought-inducibility and light responsiveness (**Supplementary Table 2**). G-Box, ABRE, GT1-motif, MSA-like, and CCAAT-box were also present in the actin gene promoter region indicating that *cis*-elements are conserved in the promoter region of pepper genes.

Expression of Capsaicinoid-Biosynthetic Genes and *CaR2R3-MYB* DEGs

RNA-seq data from the pericarp between 10 days after germination (DAP) and 60 DAP, as well as from the placenta

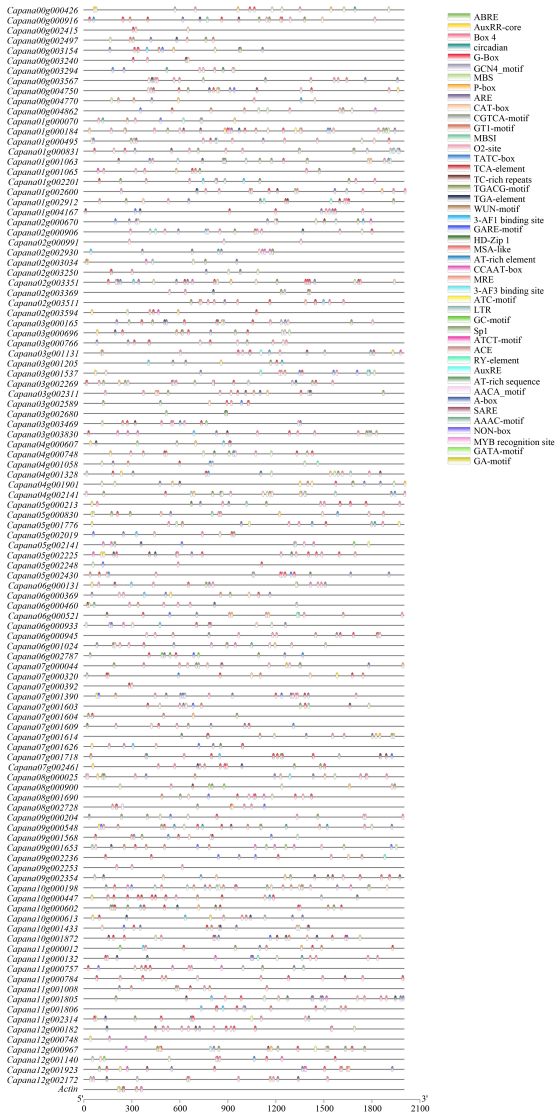


FIGURE 4 | Predicted cis-elements in *CaR2R3-MYBs* promoters. Promoter sequences (−2,000 bp) of *CaR2R3-MYBs* and one actin gene were analyzed using PlantCARE. Different shapes and colors represent different elements. Annotations of cis-elements are listed in **Supplementary Table S2**.

and seed between 10 and 15 DAP, and the placenta between 20 and 60 DAP were used to determine expression levels of *CaR2R3-MYB* DEGs and CBGs. A total of 35 (32.4%) *CaR2R3-MYB* DEGs of *CaR2R3-MYB* family genes from the pericarp between 10 and 60 DAP (adjust *P*-value < 0.01, $|\text{Log2foldchange}| > 1$) were identified. Nine DEGs were down-regulated while 26 genes were up-regulated (**Supplementary Figure 1**). The expression of *CaR2R3-MYB* DEGs can be separated into two groups (**Figure 5A**); the first part of expression levels was higher in late stage pericarp and placenta. There were eight *CaR2R3-MYB* DEGs in this part, of which seven were down-regulated genes. The expression of the other part was higher in the early stage of the pericarp, placenta and seed, and placenta, which contained 27 *CaR2R3-MYB* DEGs. Indeed, data suggested that *CaR2R3-MYB* genes are widely involved in the regulation of pepper fruit development process. The expression levels of CBGs in the capsaicinoid biosynthetic pathway were also identified (**Figure 5B**); *C3H*, *COMT*, *KAS*, *FAT*, *KR*, *DH*, *ENRa*, and *ACS1* were highly expressed in the early stage of the pericarp. Throughout the placenta and seed development process, *C4H*, *C3H*, *FAT*, and *DH* were significantly expressed between 10 and 15 DAP. Data showed that *4CL*, *HCT*, *ACL*, and *ENRa* were highly expressed in the early stage of the placenta, while *CCoAOMT*, *HCHL*, *pAMT*, *BCAT*, *BCKDH*, *KAS*, and *AT3* were highly expressed in the late stage of the placenta.

Co-expression Analysis of Capsaicinoid-Biosynthetic Genes and *CaR2R3-MYBs*

The expression levels of 108 *CaR2R3-MYB*, 35 *CaR2R3-MYB* DEGs, and 20 CBGs were used to predict candidate *CaR2R3-MYB* genes related to capsaicin synthesis. As shown in **Figure 6**, **19** *CaR2R3-MYB* genes are co-expressed with four CBGs. Indeed, in the phenylpropanoid pathway, *C4H* was co-expressed with *Capana08g001690*, while *COMT* was highly co-expressed with *Capana07g002461*, *Capana03g000165*, *Capana11g000784*, *Capana03g001537*, and *Capana00g004750*. Similarly, *4CL* was co-expressed with *Capana11g000784*, and *Capana03g001537*. In a branched chain fatty acid pathway, *ACL* was co-expressed with *Capana02g003351*, *Capana02g000906*, *Capana07g002461*, *Capana03g000165*, *Capana11g000784*, *Capana03g001537*, *Capana01g000495*, *Capana08g000025*, *Capana03g001131*, and *Capana00g004750*. Six *CaR2R3-MYB* genes (*Capana01g000495*, *Capana02g000906*, *Capana02g003351*, *Capana07g001604*, *Capana08g000900*, and *Capana08g001690*) in the co-expression network were DEGs in the pericarp between 10 and 60 DAP. Data showed that these six *CaR2R3-MYB* DEGs are the candidate genes involved in capsaicin or capsanthin synthesis processes.

The real-time qRT-PCR was performed to analyze the transcription levels of *Capana01g000495*, *Capana02g000906*, *Capana02g003351*, *Capana07g001604*, *Capana08g000900*, and *Capana08g001690* (**Figure 7**). Results illustrated that qRT-PCR and RNA-seq expression levels were similar for all six *CaR2R3-MYB* DEGs (**Supplementary Figure 2**). In terms of progressing fruit development, expression levels of *Capana02g003351* and *Capana08g000900* kept decreasing, while expression levels of *Capana01g000495*, *Capana07g001604*, *Capana02g000906*, and *Capana08g001690* first increased and then decreased. It is revealed that these six *CaR2R3-MYB* genes play key roles as candidate genes in capsaicin synthesis.

DISCUSSION

Identification and Characterization of *CaR2R3-MYB* Genes

Pepper (*Capsicum*) is famous for its spiciness and is an economically important Solanaceae family crop cultivated

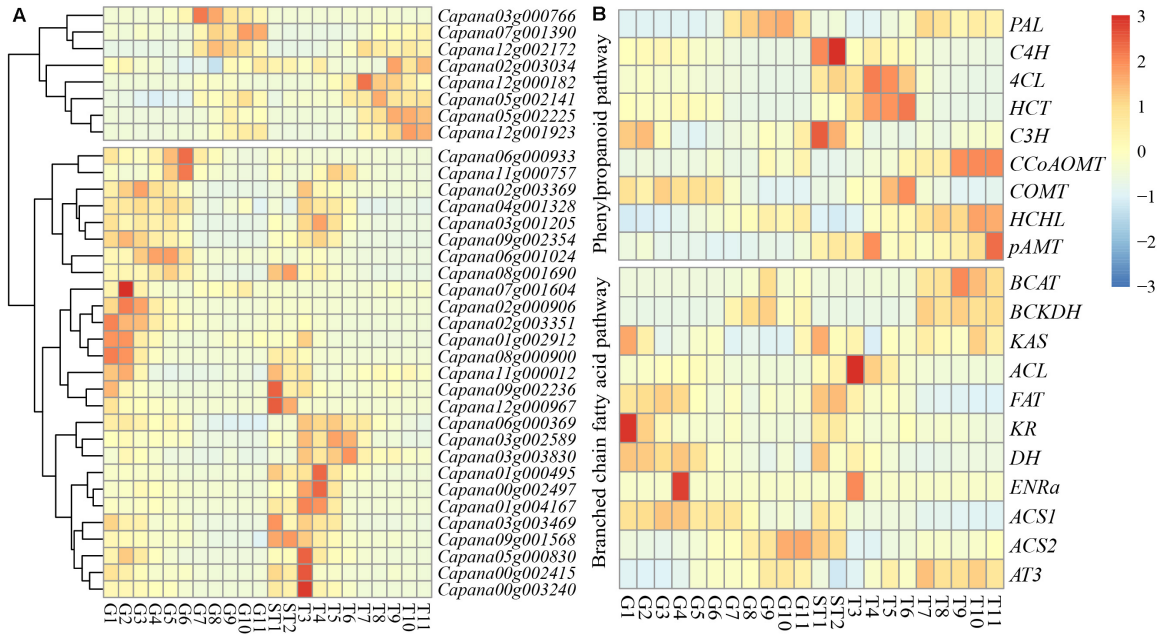


FIGURE 5 | Expression patterns of *CaR2R3-MYB* DEGs and CBGs. Expression profiles using RNA-seq Fragments Per Kilobase Million (FPKM) data in the pericarp between ten DAP and 60 DAP (G1-G11), placenta and seed between ten DAP and 15 DAP (ST1 and ST2), and the placenta between 20 DAP and 60 DAP (T3-T10). **(A)** *CaR2R3-MYB* DEGs expression level **(B)** CBGs. CBGs including PAL: phenylalanine ammonia lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumaroyl-CoA ligase; HCT: hydroxycinnamoyltransferase; C3H: coumaroylshikimate/quinate 3-hydroxylase; CCoAOMT: caffeoyl-CoA 3-O-methyltransferase; COMT: caffeic acid O-methyl transferase; HCHL: hydroxycinnamoyl-CoA hydratase/lyase; pAMT: putative aminotransferase; in phenylpropanoid pathway, and BCAT: branched-chain amino acid transferase; KAS: ketoacyl-ACP synthase; ACL: acyl carrier protein; FAT: acyl-ACP thioesterase; ENR, enoyl-ACP reductase; KR, ketoacyl-ACP reductase; DH, hydroxyacyl-ACP dehydratase; ACS: acyl-CoA synthetase; AT: acyltransferase; in branched chain fatty acid pathway. The color scale at the left of each dendrogram represents log2 expression values, with red indicating high levels and blue indicating low levels of transcript abundance.

globally for its nutritional benefits. Reference genome sequencing of the two varieties of pepper cultivars, CM334 (Mexico), and Zunla-1 (Guizhou, China), was completed in 2014 (Kim et al., 2014; Qin et al., 2014). In the pepper genome, however, gene families have been widely identified *via* incomplete annotation files. A total of 104 *CaNAC* genes (Diao et al., 2018), 35 *mTERF* genes (*CamTERF*s) (Tang et al., 2019), nine *CaCBL*, and 26 *CaCIPK* genes in the pepper genome (Ma et al., 2019). Here, 108 *CaR2R3-MYB*s were identified in both the CM334 and Zunla-1 genome, less than those identified in *Arabidopsis thaliana* (Stracke et al., 2001). The Basic Local Alignment Search Tool Protein (BLASTP) was used to align homologous genes in these two genomes. Thus, 11 *CaR2R3-MYB*s had "00g" chromosomal records in the Zunla-1 genome (also annotated in the CM334 genome), while seven *CaR2R3-MYB*s had chromosomal annotation in the Zunla-1 genome but no records in the CM334 genome (**Supplementary Table 1**). The complementarity of CM334 and "Zunla-1" can contribute to improve the complete annotation of the pepper genome. A chromosomal location circos plot illustrates that *CaR2R3-MYB*s are distributed evenly among every chromosome.

In order to study the evolution and transcriptional features of *CaR2R3-MYB*s, the number and distribution of introns and exons were analyzed. The number of CDS in the *CaR2R3-MYB* gene varies between 1 and 10, with the largest number of genes in three exons. Genes within the same subclass have similar structures and predicted motifs. These results indicate that the gene structure of *CaR2R3-MYB*s is highly conserved; a total of 16 *CaR2R3-MYB* duplicated gene pairs were identified *via* interspecies synteny analysis. The existence of duplicated genes is the main reason for *CaR2R3-MYB* gene amplification, one reason for the large number of family genes (Wang et al., 2013). A phylogenetic tree was constructed using cluster analysis such that *CaR2R3-MYB*s were divided into seven subclasses. As a result of highly conserved features, *CaR2R3-MYB*s within the same subclass tended to have similar functions.

Different *cis*-regulatory elements in the promoter sequences of genes may produce different expression patterns (Islam et al., 2019). A total of eight cellular development related *cis*-regulatory elements, 19 stress related *cis*-elements, and 13 hormone-related *cis*-regulatory elements were present. This analysis demonstrates that most *CaR2R3-MYB*s have divergent regulatory elements compared with the actin gene. The *CaR2R3-MYB* gene family has highly different *cis*-regulatory elements in the promoter region, which may lead to *CaR2R3-MYB* gene functional divergence at the transcriptional level (Haberer et al., 2004).

Capsaicin Biosynthesis Related Expression Level of *CaR2R3-MYB*s

R2R3-MYB is recognized as the dominant MYB type gene with the largest number of members in most plants, widely involved

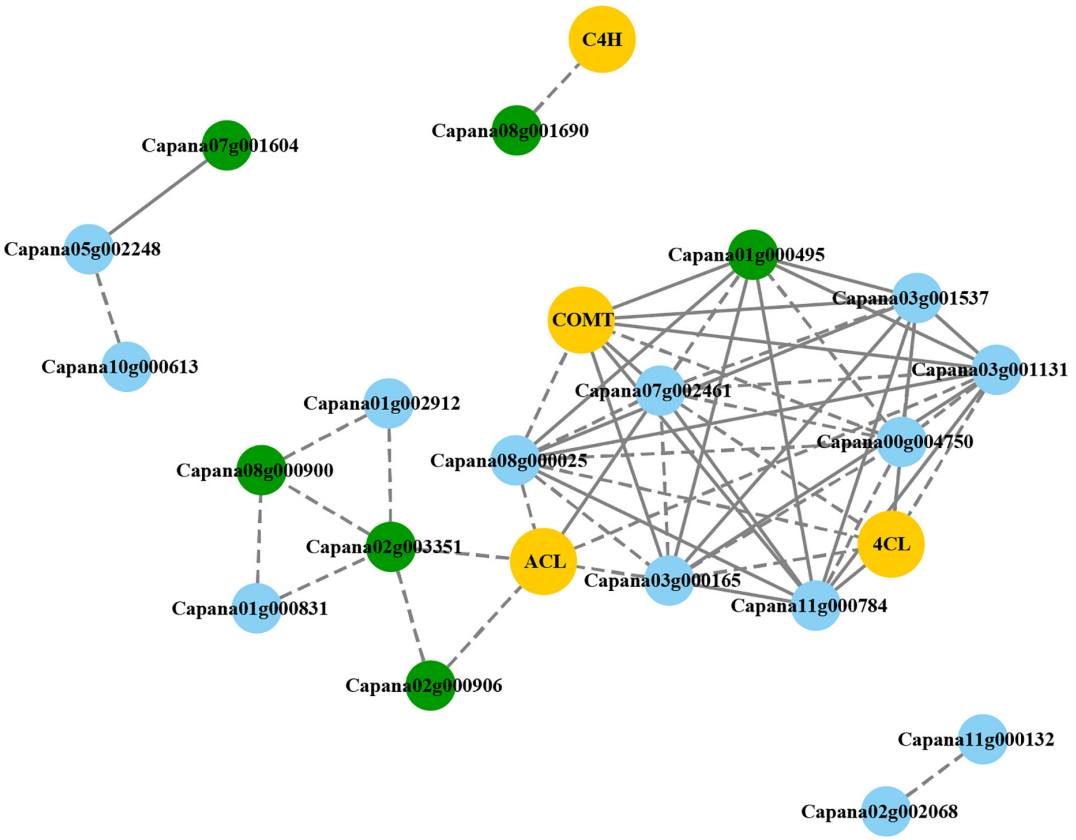


FIGURE 6 | Co-expression networks of *CaR2R3-MYB* DEGs and CBGs. Yellow boxes denote CBGs, blue boxes denote *CaR2R3-MYB* genes, and green boxes denote *CaR2R3-MYB* DEGs. The solid line indicates that the co-expression weight is higher, while the dotted line indicates that the co-expression weight is lower.

in the regulation of plant morphogenesis, growth, metabolism, developmental processes, and responses to biotic or abiotic stresses (Rosinski and Atchley, 1998). In pepper plants, *R2R3-MYB* transcription factors also play significant roles. Virus-induced gene silencing has revealed that *MYB* and *WD40* are involved in the regulation of anthocyanin biosynthesis in chili pepper fruits (Aguilar-Barragan and Ochoa-Alejo, 2014), while the *R2R3-MYB* transcription factor and homolog gene may play a major role in abiotic stress signaling pathways (Seong et al., 2008). Three *R2R3-MYB* transcription factor genes, *CaMYB1*, *CaMYB2*, and *CaMYB3*, from *C. annuum* exhibited differential expression during fruit ripening (Guo et al., 2011).

*R2R3-MYB*s also play a specific regulatory role in the pepper capsaicin synthesis pathway. The *R2R3-MYB* transcription factor *CaMYB31* is therefore also a candidate to control pungency in *C. annuum*; it is known that Jasmonate-Inducible *CaMYB108*, a typical *R2R3-MYB* gene, regulates capsaicinoid biosynthesis and stamen development in *Capsicum* (Sun et al., 2019). Capsaicin *R2R3-MYB* candidate genes were screened in this study and analyzed using pepper RNA-seq data. This enabled investigation of *CaR2R3-MYB* DEGs in pericarp, placenta and seed, and placenta. The expression levels of the 26 up-regulated DEGs in the early stages of pericarp, placenta and seed, as well as placenta development are typically increased. This illustrates that up-regulated *CaR2R3-MYB* DEGs respond during capsaicin synthesis.

The expression of *CaR2R3-MYB* DEGs is tissue-specific during plant development. Partial *CaR2R3-MYB* DEGs have the highest expression levels in the pericarp, while other genes are expressed in the placenta, remarkable tissue-specific expression differences. Partial genes such as *Capana03g001205*, *Capana04g001328*, *Capana02g003369*, *Capana11g000757*, and *Capana09g002354* have almost detectable expressions in the early stages of both pericarp and placenta. Family genes only contain the duplicate gene pairs of *Capana02g000906-Capana02g003369* and *Capana02g000906-Capana06g001024*; in *Capsicum* species, several key CBGs are also expressed at early stages in the pericarp, placenta and seed, and placenta. In both phenylpropanoid and branched chain fatty acid pathways, expression levels of CBGs in this treatment are similar to those previously determined (Kim et al., 2014). Numerous *CaR2R3-MYB* family genes are related to CBGs in expression level, indicating that *CaR2R3-MYB* genes may directly or indirectly participate in the capsaicinoid biosynthesis.

In the co-expression network of capsaicinoid biosynthetic pathway, *MYB31* was co-expressed with 174 genes; amongst these co-expressed genes, 15 have been previously reported to be involved in CAPD biosynthesis (Zhu et al., 2019). Genes in

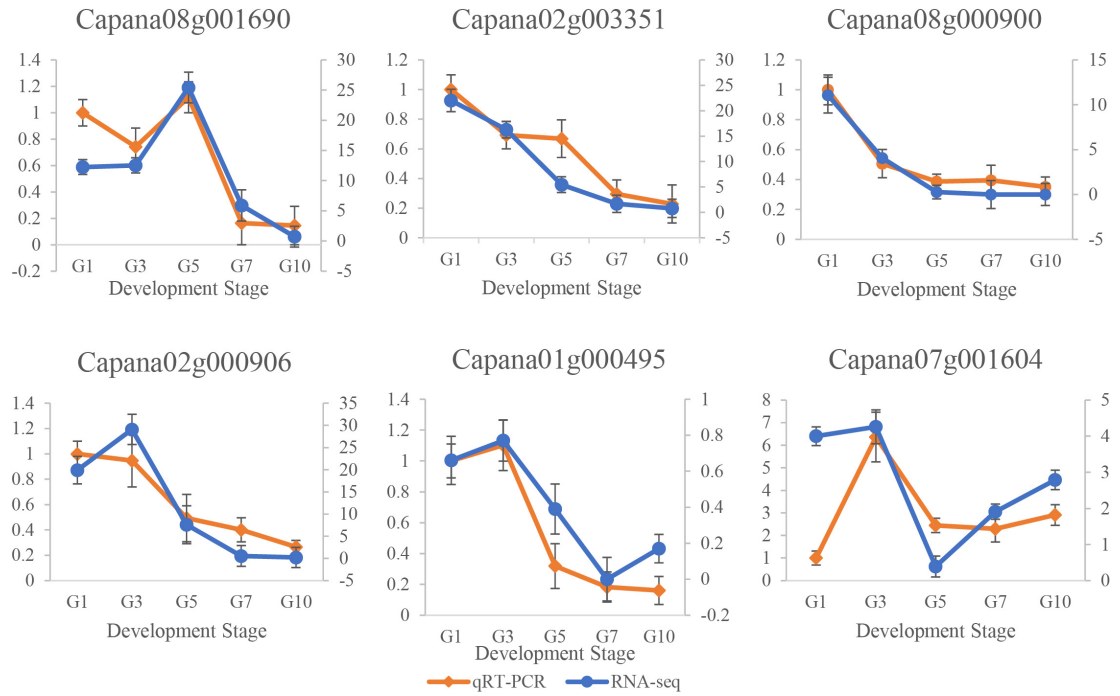


FIGURE 7 | qRT-PCR and RNA-seq data of six candidate *CaR2R3-MYB* genes. Pericarp samples from G1, G3, G5, G7, and G10 show a series of expression changes in candidate *CaR2R3-MYB* genes. The orange line represents the qRT-PCR level while the blue line represents the RNA-seq level. CBGs are shown in **Supplementary Figure S2**.

the *CaR2R3-MYB* family co-expression network were identified; 19 were selected in the pericarp, placenta and seed, and placenta development process, which were co-expressed with CBGs that might be involved in capsaicinoid biosynthetic synthesis. Divergence expression of *CaR2R3-MYB* genes shape the pungent diversification in peppers. Six *CaR2R3-MYB* DEGs are candidate capsaicinoid biosynthetic related genes in this study, including *MYB31* and other five *CaR2R3-MYB*. This research systematically studied the main characteristics of the *CaR2R3-MYB* gene family, and also provides an important reference for the study of transcription factors related to the capsaicin synthesis pathway. This study provides comprehensive information about pepper *R2R3-MYB* genes and can help determine the function of pepper *R2R3-MYB* genes. It also provides important candidate genes for capsaicinoid biosynthesis research.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://pepperhub.hzau.edu.cn/.

AUTHOR CONTRIBUTIONS

XZ, FL, XD, and JW designed the research. JW, YL, and BT conducted the experiments. JW and YL analyzed the data. JW wrote the manuscript. FL, LX, and JW revised the manuscript and improved the English. XZ and FL acquired the funding. All authors have read, reviewed, and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (U19A2028).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.598183/full#supplementary-material

Supplementary Figure 1 | The DEGs of *CaR2R3-MYB* in pepper between pericarp in different stages (adjust *P*-value < 0.01, | Log2foldchange| > 1). Orange bars are up-regulated genes and blue are down-regulated DEGs.

Supplementary Figure 2 | qRT-PCR and RNA-seq data of four CBGs.

Supplementary Figure 3 | Expression patterns of capsaicinoid-biosynthetic related *CaR2R3-MYB* genes.

Supplementary Table 1 | The basic information of *R2R3-MYB* gene family in *Capsicum*. List of predicted genes and related information include gene name (Zunla_1), CM334 homologous gene and gene locus, molecular details, and predicting subcellular localization.

Supplementary Table 2 | Genes and primers used in the quantitative real-time PCR.

Supplementary Table 3 | *Cis*-regulatory elements annotations.

REFERENCES

Aguilar-Barragan, A., and Ochoa-Alejo, N. (2014). Virus-induced silencing of *MYB* and *WD40* transcription factor genes affects the accumulation of anthocyanins in chilli pepper fruit. *Biol. Plant.* 58, 567–574. doi: 10.1007/s10535-014-0427-4

Arce-Rodriguez, M. L., and Ochoa-Alejo, N. (2015). Silencing *AT3* gene reduces the expression of *pAmt*, *BCAT*, *Kas*, and *Acl* genes involved in capsaicinoid biosynthesis in chili pepper fruits. *Biol. Plant.* 59, 477–484. doi: 10.1007/s10535-015-0525-y

Arce-Rodriguez, M. L., and Ochoa-Alejo, N. (2017). An *R2R3-MYB* transcription factor regulates capsaicinoid biosynthesis. *Plant Physiol.* 174, 1359–1370. doi: 10.1104/pp.17.00506

Arce-Rodriguez, M. L., and Ochoa-Alejo, N. (2019). Biochemistry and molecular biology of capsaicinoid biosynthesis: recent advances and perspectives. *Plant Cell Rep.* 38, 1017–1030. doi: 10.1007/s00299-019-02406-0

Aza-González, C., Núñez-Palenius, H. G., and Ochoa-Alejo, N. (2011). Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). *Plant Cell Rep.* 30, 695–706. doi: 10.1007/s00299-010-0968-8

Bolger, A. M., Marc, L., and Bjoern, U. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brown, J., Pirrung, M., and Mccue, L. A. (2017). FQC dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33, 3137–3139. doi: 10.1093/bioinformatics/btx373

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools – an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Choi, S., Suh, B., Kozukue, E., Kozukue, N., Levin, C., and Friedman, M. (2006). Analysis of the contents of pungent compounds in fresh korean red peppers and in pepper-containing foods. *J. Agr. Food Chem.* 54, 9024–9031. doi: 10.1021/jf061157z

Diao, W., Snyder, J. C., Wang, S., Liu, J., Pan, B., Guo, G., et al. (2018). Genome-wide analyses of the *NAC* transcription factor gene family in pepper (*Capsicum annuum* L.): chromosome location, phylogeny, structure, expression patterns, cis-elements in the promoter, and interaction network. *Int. J. Mol. Sci.* 19:1028. doi: 10.3390/ijms19041028

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2002). *MYB* transcription factors in *Arabidopsis*. Trends Plant Sci. 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

Elgebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, 29–37. doi: 10.1093/nar/gkr367

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. M. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/gkg563

Gates, D. J., Strickler, S. R., Mueller, L. A., Olson, B. J., and Smith, S. D. (2016). Diversification of *R2R3-MYB* transcription factors in the tomato family *Solanaceae*. *J. Mol. Evol.* 83, 26–37. doi: 10.1007/s00239-016-9750-z

Guo, L. J., Liang, L. H., and Peng, S. Q. (2011). Three *R2R3-MYB* transcription factor genes from *Capsicum annuum* showing differential expression during fruit ripening. *Afr. J. Biotechnol.* 10, 8267–8274. doi: 10.1186/1471-2164-12-399

Haberer, G., Hindemitt, T., Meyers, B. C., and Mayer, K. F. (2004). Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. *Plant Physiol.* 136, 3009–3022. doi: 10.1104/pp.104.046466

Hu, B., Jin, J., Guo, A., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Islam, S., Sajib, S. D., Jui, Z. S., Arabia, S., and Ghosh, A. (2019). Genome-wide identification of glutathione S-transferase gene family in pepper, its classification, and expression profiling under different anatomical and environmental conditions. *Sci. Rep.* 9:15. doi: 10.1038/s41598-019-45320-x

Jia, L., Clegg, M. T., and Jiang, T. (2004). Evolutionary dynamics of the DNA-binding domains in putative *R2R3-MYB* genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.* 134, 575–585. doi: 10.1104/pp.103.027201

Keyhaninejad, N., Curry, J., Romero, J., and Oconnell, M. A. (2014). Fruit specific variability in capsaicinoid accumulation and transcription of structural and regulatory genes in *Capsicum* fruit. *Plant Sci.* 21, 59–68. doi: 10.1016/j.plantsci.2013.10.013

Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* 46, 270–278. doi: 10.1038/ng.2877

Koeda, S., Sato, K., Saito, H., Nagano, A. J., Yasugi, M., Kudoh, H., et al. (2019). Mutation in the putative ketoacyl-ACP reductase *CaKR1* induces loss of pungency in *Capsicum*. *Theor. Appl. Genet.* 132, 65–80. doi: 10.1007/s00122-018-3195-2

Krzywinski, M., Schein, J. E., Birol, I., Connors, J. M., Gascoyne, R. D., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Langfelder, P., and Horvath, S. (2008). WGCNA, an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496. doi: 10.1093/nar/gkx922

Liu, F., Yu, H., Deng, Y., Zheng, J., Liu, M., Ou, L., et al. (2017). PepperHub, an informatics HUB for the chili pepper research community. *Mol. Plant* 10, 1129–1132. doi: 10.1016/j.molp.2017.03.005

Liu, S., Li, W., Wu, Y., Chen, C., and Lei, J. (2013). *De Novo* Transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. *PLoS One* 8:e48156. doi: 10.1371/journal.pone.0048156

Ma, X., Gai, W., Qiao, Y., Ali, M., Wei, A., Luo, D., et al. (2019). Identification of *CBL* and *CIPK* gene families and functional characterization of *CaCIPK1* under *Phytophthora capsici* in pepper (*Capsicum annuum* L.). *BMC Genomics* 20:18. doi: 10.1186/1471-2229-14-8

Naves, E. R., Silva, L. D., Sulpice, R., Araujo, W. L., Nunesnesi, A., Peres, L. E., et al. (2019). Capsaicinoids: pungency beyond *Capsicum*. *Trends Plant Sci.* 24, 109–120. doi: 10.1016/j.tplants.2018.11.001

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rosinski, J. A., and Atchley, W. R. (1998). Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. *J. Mol. Evol.* 46, 74–83. doi: 10.1007/PL00006285

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385. doi: 10.1073/pnas.0807484105

Seong, E. S., Guo, J., and Wang, M. H. (2008). The chilli pepper (*Capsicum annuum*) MYB transcription factor (*CaMYB*) is induced by abiotic stresses. *J. Plant Biochem. Biotechnol.* 17, 193–196. doi: 10.1007/BF03263285

Shannon, P., Markiel, A., Ozier, O., Wang, N. S., and Ramage, J. T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Stewart, C., Kang, B., Liu, K., Mazourek, M., Moore, S., Yoo, E. Y., et al. (2005). The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant J.* 42, 675–688. doi: 10.1111/j.1365-313X.2005.02410.x

Stewart, C., Mazourek, M., Stellari, G. M., Oconnell, M. A., and Jahn, M. (2007). Genetic control of pungency in *C. chinense* via the *Pun1* locus. *J. Exp. Bot.* 58, 979–991. doi: 10.1093/jxb/erl243

Stracke, R., Werber, M., and Weisshaar, B. (2001). The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 447–456. doi: 10.1016/S1369-5266(00)00199-0

Sun, B., Zhu, Z., Chen, C., Chen, G., Cao, B., Chen, C., et al. (2019). Jasmonate-inducible *R2R3-MYB* transcription factor regulates capsaicinoid biosynthesis and stamen development in *Capsicum*. *J. Agr. Food Chem.* 67, 10891–10903. doi: 10.1021/acs.jafc.9b04978

Tanaka, Y., Asano, T., Kanemitsu, Y., Goto, T., Yoshida, Y., Yasuba, K., et al. (2019). Positional differences of intronic transposons in *pAMT* affect the pungency level in chili pepper through altered splicing efficiency. *Plant J.* 100, 693–705. doi: 10.1111/tpj.14462

Tang, B., Xie, L., Yi, T., Lv, J., Yang, H., Cheng, X., et al. (2019). Genome-wide identification and characterization of the mitochondrial transcription termination factors (*mTERFs*) in *Capsicum annuum* L. *Int. J. Mol. Sci.* 21:269. doi: 10.3390/ijms21010269

Tewksbury, J. J., Reagan, K. M., Machnicki, N. J., Carlo, T. A., Haak, D. C., Penaloza, A. L., et al. (2008). Evolutionary ecology of pungency in wild chilies. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11808–11811. doi: 10.1073/pnas.0802691105

Varet, H., Brilletgueguen, L., Coppee, J., and Dillies, M. (2016). SARTools: a DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PLoS One* 11:e0157022. doi: 10.1371/journal.pone.0157022

Von Wettsteinknowles, P., Olsen, J. G., Mcguire, K. A., and Larsen, S. (2000). Molecular aspects of β-ketoacyl synthase (KAS) catalysis. *Biochem. Soc. Trans.* 28, 601–607. doi: 10.1042/BST0280601

Wang, J., Liu, Y., Chen, X. L., and Kong, Q. S. (2020). Characterization and divergence analysis of duplicated *R2R3-MYB* genes in watermelon. *J. Am. Soc. Hortic. Sci.* 145, 1–8. doi: 10.21273/JASHS04849-19

Wang, Y., Tan, X., and Paterson, A. H. (2013). Different patterns of gene structure divergence following gene duplication in *Arabidopsis*. *BMC Genomics* 14:652. doi: 10.1186/1471-2164-14-652

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Yang, L., Smyth, G. K., and Wei, S. (2013). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Zhang, Z., Li, J., Zhao, X., Wang, J., Wong, G., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4, 259–263. doi: 10.1016/S1672-0229(07)60007-2

Zhu, Z., Sun, B., Cai, W., Zhou, X., Mao, Y., Chen, C., et al. (2019). Natural variations in the *MYB* transcription factor *MYB31* determine the evolution of extremely pungent peppers. *New Phytol.* 223, 922–938. doi: 10.1111/nph.15853

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Liu, Tang, Dai, Xie, Liu and Zou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 December 2020
doi: 10.3389/fgene.2020.593426

Pyramiding of High Grain Weight With Stripe Rust and Leaf Rust Resistance in Elite Indian Wheat Cultivar Using a Combination of Marker Assisted and Phenotypic Selection

Satinder Kaur[1]†, Jaspreet Kaur[2], G. S. Mavi[2], Guriqbal Singh Dhillon[1†], Achla Sharma[2], Rohtas Singh[1], Urmila Devi[1] and Parveen Chhuneja[1†]*

[1] School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, [2] Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

OPEN ACCESS

Edited by:
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Sreepriya Pramod,
Altria, United States
Vijay Gahlaut,
University of Delhi, India

***Correspondence:**
Satinder Kaur
satinder.biotech@pau.edu

†ORCID:
Satinder Kaur
orcid.org/0000-0003-3704-3074
Guriqbal Singh Dhillon
orcid.org/0000-0001-6766-810X
Parveen Chhuneja
orcid.org/0000-0002-8599-9479

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 10 August 2020
Accepted: 30 November 2020
Published: 22 December 2020

Citation:
Kaur S, Kaur J, Mavi GS, Dhillon GS, Sharma A, Singh R, Devi U and Chhuneja P (2020) Pyramiding of High Grain Weight With Stripe Rust and Leaf Rust Resistance in Elite Indian Wheat Cultivar Using a Combination of Marker Assisted and Phenotypic Selection. Front. Genet. 11:593426. doi: 10.3389/fgene.2020.593426

Wheat (*Triticum aestivum* L.) is an important cereal crop globally as well as in India and yield improvement programs encounter a strong impediment from ever-evolving rust pathogens. Hence, durable rust resistance is always a priority trait for wheat breeders globally. Grain weight, represented as thousand grain weight (TGW), is the most important yield-contributing trait in wheat. In the present study high TGW has been transferred into two elite Indian wheat cultivars PBW343 and PBW550 from a high TGW genotype, Rye selection 111, selected from local germplasm. In the background of PBW343 and PBW550, an increase in TGW upto 27.34 and 18% was observed, respectively (with respect to recipient parents), through conventional backcross breeding with phenotypic selections in 3 years replicated RBD trials. Resistance to leaf rust and stripe rust has been incorporated in the high TGW version of PBW550 through marker assisted pyramiding of stripe rust resistance gene *Yr15* using marker *Xuhw302,* and a pair of linked leaf rust and stripe rust resistance genes *Lr57-Yr40* using marker *Ta5DS-2754099_kasp23*. Improved versions of PBW550 with increased TGW ranging from 45.0 to 46.2 g (up to a 9% increase) and stacked genes for stripe and leaf rust resistance have been developed. This study serves as proof of utilizing conventional breeding and phenotypic selection combined with modern marker assisted selection in improvement of important wheat cultivars as a symbiont of conventional and moderan techniques.

Keywords: wheat, PBW550, PBW343, pyramiding, grain weight, rust resistance

INTRODUCTION

Wheat (*Triticum aestivum* L.) is an important cereal crop in India, ranking second after rice for the area (29.31 million hectares) and production (103.6 million metrics) with the state of Punjab sharing 18% of production. With global per capita consumption of 67.4 kg/year, wheat is the most widely consumed food grain (Djanaguiraman et al., 2019). In India only, the population is

projected to cross the 1.70 billion mark by 2050 with a domestic demand of wheat exceeding 140 million tons (Nagarajan, 2005)[1]. A consistent increase in the wheat yields is a primary goal for food security of the growing population (Singh et al., 2007; Ye and Smith, 2008). Crop yield is a complex quantitative trait determined by different parameters of tiller number, grain number, grain weight, etc. Grain or kernel weight (1,000 grain weight in g, TGW), consisting of grain length, width, and area has high heritability (> 0.68) (Giura and Saulescu, 1996; Kuchel et al., 2007b) which not only translates into higher yields but also has a favorable effect over flour yield (Gegas et al., 2010). Moreover, uniform and larger sized grains are visually appealing and fetch higher market prices.

As rusts are major diseases in India, keeping pace of increasing yields with rust resistance is one of the most challenging tasks. The three species of rust viz., stripe rust caused by *Puccinia striiformis*, leaf rust caused by *Puccinia triticina*, and stem rust caused by *Puccinia graminis* are severely affecting wheat yield. An estimated loss of 200 million rupees occur every year due to rusts (Mehta, 1950). Leaf rust is prevalent in all the wheat growing zones of India and its widespread occurrence was observed during periods of 1971–1973, 1993–1994 (Joshi et al., 1975; Nayar et al., 1997). The 70% of the total area under wheat cultivation in cooler parts Northern India is under constant threat of ever evolving epidemic stripe rust pathogenic races. It occurred almost every year from 1967 to 1974, with high incidence recorded in 2001 and 2011 (Nayar et al., 1997; Prashar et al., 2007; Pannu et al., 2010; Tomar et al., 2014; Pal et al., 2015). An approximate loss of rupees 236 crore have been recorded in Punjab state during the epidemic year of 2009–2010 (Jindal et al., 2012). Stem rust on other hand is important central and peninsular India since its first epidemic reported in 1786AD (Nagarajan and Joshi, 1975).

Despite the devastating nature and continuous occurrence of rusts, wheat production is growing linearly due to the continuous addition of new rust resistance genes in the wheat gene pool. Use of resistant wheat cultivars is not only effective and economical but also environment friendly (Peng et al., 2003; Singh et al., 2020). Due to pathogen evolution, combination of resistance genes is being pursued by the crop breeders world over for increasing durability of resistance. Resistance gene pyramiding has been found to increase the life of each gene though the synergistic effect of pyramided genes (Klymiuk et al., 2018; Mundt, 2018). One of the greatest successes in the history of resistance breeding is with the pyramiding of resistance genes in controlling stem rust (Singh et al., 2015). The Ug99 race group of pathotypes of stem rust defeated rust resistance genes *Sr31* and *Sr38* but their differential pyramiding combinations of different resistant genes *Sr22*, *Sr25*, *Sr26*, *Sr33*, *Sr35*, *Sr45,* and *Sr50* were found to be effective. For leaf rust, pyramiding of different genes in combination with leaf rust resistance genes provided long lasting resistance (Kolmer, 1996; Bhawar et al., 2011; Aboukhaddour et al., 2020; Babu et al., 2020). Resistance to stripe rust has also been significantly improved by pyramiding of different resistance genes (Zheng et al., 2017; Randhawa et al., 2019).

The present study reports the introgression of high thousand grain weight (TGW) to wheat varieties PBW343 and PBW550 from a local selection named "Rye Selection 111." Pyramiding of three rust resistance genes viz *Yr15* and linked *Lr57-Yr40* into improved TGW version of PBW550 was done using "transfer first and assemble later" strategy (Ishii and Yonezawa, 2007a,b). Wheat cultivar PBW343, released in India in 1995, was the most widely grown cultivar in the country and soon was recognized as the new miracle genotype with wider adaptability and yield potential (Gupta et al., 2018). PBW550 was released in 2008 and farmers' receptivity to this short duration variety with bold grains has accelerated its spread in the first few years of its release (Gupta et al., 2018). However, both of these hexaploid elite cultivars succumbed to emerging races of YR thus it became imperative to restructure the genetic makeup of these cultivars for improved yield and rust resistance.

MATERIALS AND METHODS

Plant Material

The plant material used in the present study included two cultivated hexaploid bread wheat varieties named PBW343 (ND/VG9144//KAL/BB/3/YACO'S'/4/VEE#5 "S," notification number 1(E) dated 01.01.1996, Indian Council of Agricultural Research, Government of India) and PBW550 (WH 594/RAJ 3856//W 485 notification number 72(E) dated 10.01.2008, Indian Council of Agricultural Research, Government of India), two near isogenic lines (NILs) developed at Punjab Agricultural University (PAU) in PBW550 background viz PBW550 + *Yr15* and PBW550 + *Lr57-Yr40*, a high thousand grain weight selection from local germplasm named Rye selection 111 (RyeSel 111) and the progenies from crosses of RyeSel 111 × PBW343, RyeSel 111 × PBW550, HGW343 × PBW550, HGW550 × PBW550 + *Yr15* and HGW550 + *Yr15* × PBW550 + *Lr57-Yr40*. (The "PBW" stands for Punjab Bread Wheat followed by numeric value of varietial release number, HGW = high grain weight).

Introgression of High Grain Weight in PBW343 and PBW550

For transfer of high TGW to wheat variety PBW343, cross was made with RyeSel 111 as male parent and the phenotypic selection was done in subsequent backcross generations (**Figure 1a**). The BC_2F_5 progenies of cross PBW343-RyeSel 111 with high TGW will be referred to as HGW343 hereon. Selected HGW343 plants were used as donor to transfer high TGW to wheat variety PBW550 and BC_2F_5 progenies with high TGW (**Figure 1b**) thus obtained will be termed as HGW550 hereon. All the crosses and selections were done at PAU, Ludhiana, India.

Phenotyping and Experimental Design

In both the above crosses, traits were recorded on single plants in BC_2-F_1/F_2 generations while in BC_2-$F_3/F_4/F_5$ progenies, traits were recorded on 10 single plants of each progeny as well as

[1]https://www.thehindubusinessline.com; https://www.eastasiaforum.org

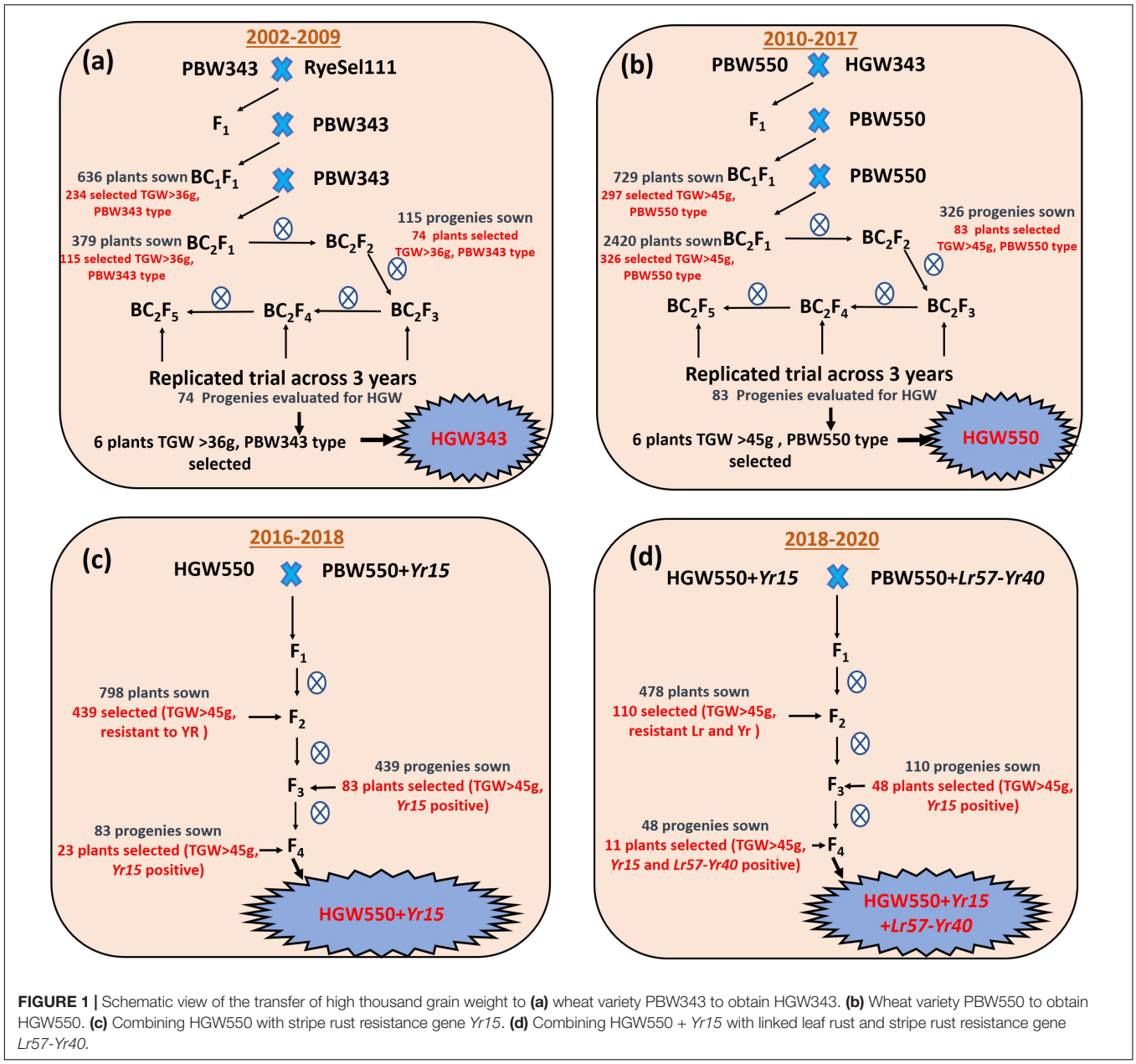


FIGURE 1 | Schematic view of the transfer of high thousand grain weight to **(a)** wheat variety PBW343 to obtain HGW343. **(b)** Wheat variety PBW550 to obtain HGW550. **(c)** Combining HGW550 with stripe rust resistance gene *Yr15*. **(d)** Combining HGW550 + *Yr15* with linked leaf rust and stripe rust resistance gene *Lr57-Yr40*.

bulk of progeny. Single plants from uniform and promising progenies were selected and carried forward. Evaluation of each of the BC_2-F_3/F_4/F_5 progenies along with parental lines were done in three replications (1.5 m paired rows with the plant to plant distance of 10 cm and row to row distance of 20 cm) in randomized block design. Wherever single plant selections were done, the weight of 300–500 grains (as available per plant) was recorded and converted into TGW, while the weight of thousand grains was recorded for progeny bulks. Selections were done for plant/progenies having TGW higher than recurrent parent. Similarly other phenotypic traits of plant height, tiller number, spikelets per spike, spike length, and grains per spike were also recorded and plants/progenies having either similar value to or higher than recurrent parent were selected (data not given).

For each progeny the mean values were calculated for recorded traits in each replication and adjusted means of replications were finalized for each year. Three years data was again used to calculate adjusted means across the years for final selections.

Combining High Grain Weight and Rust Resistance in PBW550

For pyramiding HGW with stripe rust resistance gene *Yr15*, first the selected HGW550 progenies obtained from the above cross were crossed with NIL PBW550 + *Yr15*. The HGW550 + *Yr15* thus obtained was again crossed with another NIL, PBW550 + *Lr57-Yr40*, to develop the HGW550 + *Yr15* + *Lr57-Yr40* line. Selections were done

for high TGW and rust resistance as for single plant in F_1/F_2 generations, while in $F_3/F_4/F_5$ progenies 10 single plants along with progeny bulks recorded to select single plants from best performing progenies. Since both the parental lines were from the same background, no backcrossing was done.

Marker Assisted Selection for *Yr15* and *Lr57-Yr40* Genes

Selection for rust resistance was done in a combined phenotypic–genotypic manner. For marker assisted selection, DNA from parental lines and segregating progenies was extracted using CTAB method (Saghai-Maroof et al., 1984) with the small modification of reducing incubation time to 30 min at 65°C with CTAB buffer and to 20 min at room temperature for solvent extraction using chloroform: isoamyl alcohol (24:1). *Yr15* positive plants were selected by amplifying gene specific marker *Xuhw302* (Klymiuk et al., 2018). For the selection of *Lr57-Yr40* gene, linked Kasp marker *Ta5DS-2754099_kasp23* (Bansal et al., 2020) and caps marker *Lr57-Yr40_caps16* (Toor et al., 2016) were used. The PCR reaction for gel based marker *Xuhw302* and *Lr57-Yr40_caps16* was done in 10 µl reaction volume (60 ng DNA, 5 µl of 2× EmeraldAmp® GT PCR Master Mix, 0.75 µl of 5 µM each primer) in 384 well microtiter plate in an Applied Biosciences 384 thermal cyclers. The PCR products were resolved using 2.5% agarose gel electrophoresis and visualized and photographed using gel documentation system. The scoring was done identifying the presence of gene specific amplicon as per the positive control. The Kasp marker *Ta5DS-2754099_kasp23* was amplified in 4 µl reaction (20 ng DNA, 1.944 µl of 2X KASP V3.0 master mix from LGC, Biosearch Technologies, 0.056 µl of primer mix in ratio of 12:12:30:46 Allele specific Primer I:Allele specific II primer: Common primer: Water) in a 384 well microtiter plate. The amplicons were identified by measuring allele specific flourescence in a high throughput TECAN infinite F200 PRO plate reader. Kluster Caller software was used to view the allele specific calls in an *x–y* plot to identify the positive and negative allels against the positive control.

Screening Against Stripe Rust and Leaf Rust

Screening for stripe rust and leaf rust was also done in the field along with marker selection to validate the effectiveness of gene pyramiding. F_3 and F_4 progenies of HGW550 X PBW550 + *Yr15* were screened against stripe rust at the adult plant stage in the field during season 2016–2017 and 2017–2018. Similarly, F_3 and F_4 progenies of HGW550 + *Yr15* X PBW550 + *Lr57-Yr40* were screened both against stripe rust and leaf rust at the adult plant stage in the field during season 2018–2019 and 2019–2020. For screening, artificial epiphytotic conditions for rust were created by spraying the urediniospores mixed and diluted in water, containing Tween-20, of *Pst* pathotypes (100S119, 78S84), and *Pt* pathotypes (77-1, 77-2, 77-5, 104-2) and a mixture of leaf rust and stripe rust inoculum collected from farmers' fields. For the uniform spread of disease, highly susceptible cultivar WL711 was planted as spreader rows all around the field and after every 20 rows. Rust data was recorded when WL711 showed complete susceptibility. Rust was recorded using Cobbs scale, as illustrated in McIntosh et al. (1995) where types of spores were recorded as zero (immune); TR (traces of severity); MR (moderately resistant), MS (moderately susceptible); and S (susceptible), and numeric numbers associated with these scores signified the percentage of leaf area covered by rust. The TR score is given to a plant when there are no lesions, MR when there are no visual postules of spores on the leaf but apoptotic lesions are visible, MS when there are minute visual postules of spores on the leaf with apoptotic lesions while *S* score is given to a plant when advanced postules of spores are visible to the leaf.

RESULTS

Development and Selection of HGW343

Different number of plants phenotypically similar to PBW343 with TGW > 36 g were selected from the cross of PBW343-Rye Sel 111 across different generations. In the crop season 2003–2004, 636 BC_1F_1 plants were sown and BC_2 progenies of only 234 BC_1F_1 plants (**Figure 1a**) were carried forward. In BC_2F_1, 115 plants were selected and planted as single plant-to-row BC_2F_2 progenies. During the years 2005–2006, from 943 well-established BC_2F_2 single plants, only 74 superior plants were selected.

For three consecutive seasons (2006–2009), 74 BC_2F_{3-5} progenies, along with RyeSel 111 and PBW343, were evaluated in replicated trials. Adjusted means for 3 years showed that TGW ranged from 36.66 to 46.81 g (0.30–28.10% increase), where TGW for PBW343 was 36.54 g and for RyeSel 111 was 54.17 g (**Table 1** and **Figure 2**). Six BC_2F_5 plants of progenies with significantly high TGW, namely HGW343-3 (TGW-45.36 g, 24.13% increment), HGW343-6 (TGW-45.54 g, 24.63% increment), HGW343-8 (TGW-46.08 g, 26.11% increment), HGW343-60 (TGW-46.48 g, 27.20% increment), HGW343-61 (TGW-46.71 g, 27.83% increment), and HGW343-11 (TGW-46.81 g, 28.10% increment) were

TABLE 1 | Statistical analysis of HGW343 and HGW550 progenies across F_3, F_4, and F_5 generations planted in RBD design.

	HGW343	HGW550
RP	36.54	42.23
RyeSel111	54.17	51.36
Range	36.66–48.36	39.29–49.83
Heritability	0.85	0.76
Genotype variance	11.52	4.88
Residual variance	5.89	4.5
Grand mean	41.11	44.74
LSD	3.62	2.99
CV	5.9	4.74
n replicates	3	3
n years	3	3

RP, recurrent parent; LSD, least significant difference; CV, coefficient of variation.

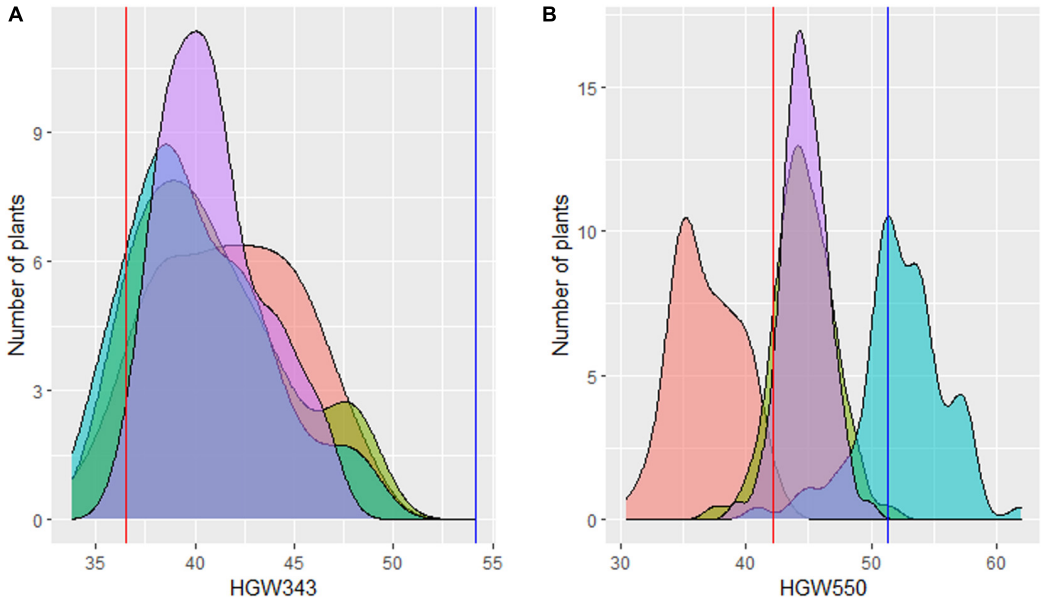


FIGURE 2 | Graphical representation of distribution of thousand grain weight of derived BC_2F_{3-5}. **(A)** HGW343 progenies and **(B)** HGW550 progenies with the BLUPs in each environment (purple). The BLUPs of Rye Selection111 and recurrent parents (PBW343 and PBW550) are given as vertical red lines and blue lines, respectively.

selected (**Figure 3**). The popularity of PBW343 in the late 1990s and early 2000s led to its widespread sowing which facilitated the selection of virulence for a super aggressive stripe rust race *78S84* (Prashar et al., 2007), breaking down resistance of its predominant gene *Yr27*, crumbling the wheat production in India. For restructuring the resistant version of PBW343, the high TGW progenies in PBW343 background were selected and disease resistance gene pyramiding carried out under a separate breeding program (not included in the present manuscript).

Development and Selection of HGW550

Four HGW343 plants named HGW343-8, HGW343-60, HGW343-61, and HGW343-11 were crossed and backcrossed with PBW550 (**Figure 1b**). Phenotypic selections for high TGW were done in BC_1F_1, BC_2F_1, and BC_2F_2. From 729 BC_1F_1 plants, backcrosses from 297 single plants were selected to obtain 2,420 BC_2F_1 plants. A total of 83 BC_2F_2 plants were selected and BC_2F_3, BC_2F_4, and BC_2F_5 progenies were evaluated across 3 years (2014–2017) in replicated trials, each having three replications (**Table 1** and **Figure 2**). Overall adjusted means for 3 years showed that TGW for the progenies ranged from 39.29 to 49.83 g where TGW for PBW550 was 42.23 g and for RyeSel 111 was 51.36. Six plants from progenies with TGW higher than 47.0 g, namely HGW550-8 (TGW-49.83 g, 17.99% increment), HGW550-6 (TGW-48.1 g, 13.90% increment), HGW550-63 (TGW-47.93 g, 13.50% increment), HGW550-3 (TGW-47.64 g, 12.81% increment), HGW550-7 (TGW-47.54 g, 12.57% increment), and HGW550-21 (TGW-47.5 g, 12.48% increment) were selected (**Figure 4**) for introducing rust resistance.

Marker Assisted Pyramiding of HGW550 and *Yr15*

Three selected HGW550 plants; HGW550-8, HGW550-6, and HGW550-63 were crossed with NIL PBW550 + *Yr*15 (**Figure 1c**). F_1s thus obtained were selfed to develop 798 F_2 plants and 439 stripe rust resistant plants with TGW of more than 42 g were selected. In F_3, plants having a TGW equal to or more than 45.0 g were selected from promising uniform stripe rust resistant progenies. The selected plants were screened with molecular marker *Xuhw302,* and 83 plants, homozygous for *Yr15* were selected (**Supplementary Figure S1A**). These 83 plants were further sown in plant to progeny rows in F_4 and 23 single F_4 plants with TGW ranging between 45.0 and 48.0 g and positive for *Yr15* were selected (**Figure 4**). These plants were also completely resistant to stripe rust in the field.

Marker Assisted Pyramiding of HGW550 + *Yr15* and *Lr57-Yr40*

F_1s generated by crossing four selected single F_4 plants HGW550 + *Yr15*-34 (TGW-46.3), HGW550 + *Yr15*-29 (TGW-47.1), HGW550 + *Yr15*-3 (TGW-47.3), HGW550 + *Yr15*-34 (TGW-47.4) with PBW550 + *Lr57-Yr40* were selfed to generate 478 F_2 plants, of which 110 individual F_2 plants were selected with TGW > 45 g (**Figure 1d**). These 110 plants were sown as plants to progeny rows and 48 single F_3 plants with the positive allele for *Yr15* and *Lr57-Yr40* markers were selected (**Supplementary Figure S1B**). These plants were also screened against stripe rust and leaf rust at the adult plant stage and found to be completely resistant to both diseases. These 48 plants were again evaluated in F_4 generation in plant to row progenies and 11 F_4 plants

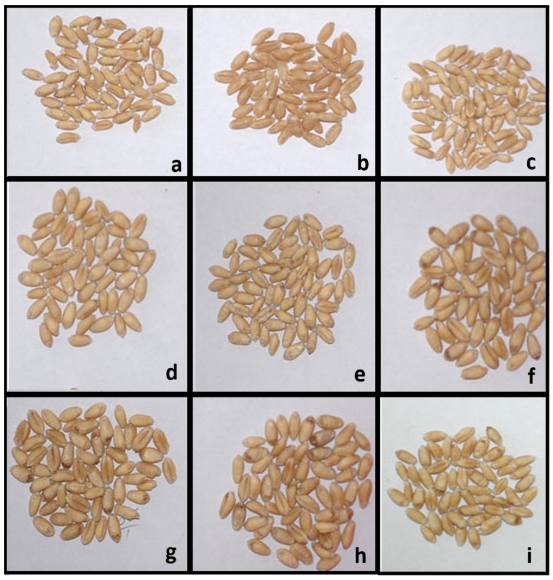


FIGURE 3 | Representative grains from the field grown BC_2F_5 progenies along with parental genotypes. **(a)** PBW343. **(b)** Rye selection 111. **(c–i)** BC_2F_5 progenies obtained from cross of PBW343/Rye selection 111.

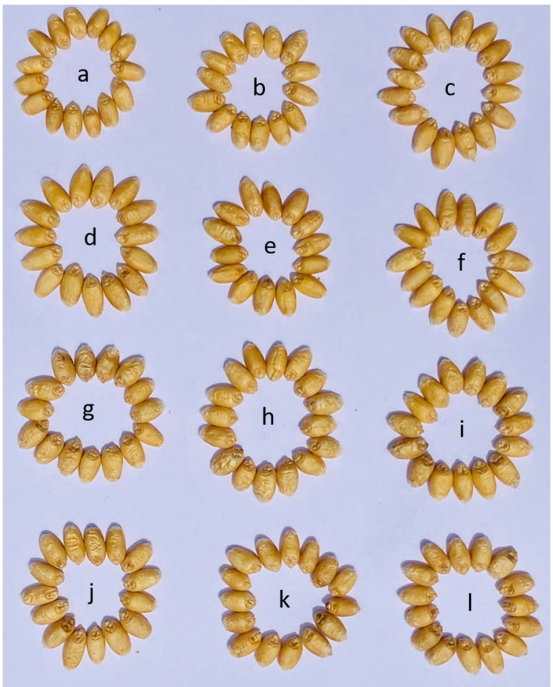


FIGURE 4 | Representative grains from the field grown pyramided lines along with parental genotypes. **(a)** PBW550. **(b)** PBW550 + *Yr15*. **(c)** PBW550 + *Yr15* + *Lr57-Yr40*. **(d–f)** BC_2F_5 progenies obtained from cross of HGW343 × PBW550. **(g–i)** F_4 progenies obtained from cross of HGW550 × *Yr15*; **(j–l)** F_4 progenies obtained from cross of HGW550 + *Yr15* X *Lr57-Yr40*.

from superior progenies with TGW between 45.0 and 46.2 g and complete resistance to stripe rust and leaf rust have been selected (**Figure 4**).

DISCUSSION

The stacking of genes governing multiple traits enhances the value of breeding material besides improving its durability. Yield is a trait of foremost priority for the commercial success of a variety, but combining improved grain yield with disease resistance, and high grain quality is a need of the changing time and environment. The introgression and pyramiding of major genes/QTL for different traits through marker-assisted selection (MAS) has been reported in wheat (Gupta et al., 2010; Kumar et al., 2010; Tyagi et al., 2014; Gautam et al., 2020).

We used RyeSel 111 as the donor to transfer the high TGW trait to elite cultivar PBW343 and PBW550 through conventional phenotypic selection. Though QTLs for high grain weight have been reported in the RyeSel 111-Chinese Spring RIL population (Kumar et al., 2006), the reported linked markers were found to be non-polymorphic between the receipient and donor combinations used in the present study. Thus a conventional route is followed through phenotypic selection for high TGW. Mapping and transfer of QTLs for yield related traits often sound more logical theoretically, but these complex traits are controlled by many QTLs of large and small effects and pyramiding all these QTLs in one background through MAS to retrieve the donor effect is not always achieved. This is more often pronounced in cases where the minor QTLs additively contribute toward the trait along with major ones. Thus phenotypic selection especially for an easily quantified trait like grain weight, gives more realistic results. Wheat yield is controlled by traits of yield per area (includes grains per spike, grain weight, and spikes per area) and yield per spike (spikelet number per spike, grain number, and grain size) (Slafer et al., 2014). Several QTLs related to grain yield related traits have been mapped on different chromosomes but none have been cloned or effectively utilized for marker assisted selection. From different factors influencing wheat yield, grain weight was found to be one of the most stably inherited (Sidwell et al., 1976; Kuchel et al., 2007a,b; Aisawi et al., 2015) suggesting selection for heavier grains lead to effective yield improvement. Raising plant yield by conventional breeding methods (Reynolds et al., 2009) has remained very successful in the past, though new emerging technologies are creating a different niche. High TGW was introgressed effectively into two popular wheat cultivars cv. PBW343 and PBW550 through the phenotypic selection, and improved progenies with an increase of TGW by 27.34 and 18% were obtained, respectively. Initially, direct crosses of RyeSel 111 were done with PBW343 then using improved HGW343 progenies as the donor, HGW introgressed into PBW 550. Similarly improvement in grain yield by improving TGW has been reported by several studies (Giunta et al., 2007; Beche et al., 2014; Qin et al., 2015; Zhang et al., 2016; Gao et al., 2017).

PBW343 has been one of the most popular wheat varieties and was grown in about 25% of the 27 million hectares under

the wheat cultivation in the country and contribute roughly 55% of the total wheat output in the country (Pavithra et al., 2017). Due to its wider adaptability to a range of environments, it has been a variety of choice for many improvement programs and was selected in this program for improving grain weight. After the PBW343 succumbed to *78S84* pathotype of stripe rust in 2004 and so did the whole breeding pipeline at PAU, it became mandatory to quickly mobilize some known stripe rust resistance genes to PBW343 and other advance germplasms to have a rust resistant variety for the farmers of the region. It was then that the development of rust resistant version of PBW343 became the main mandate of the wheat breeders, and reintroduction of its rust resistant version led to the shift of this work to a separate program.

PBW550, on the other hand, has been recognized as a good quality cultivar with bold grains and was spread to whole wheat growing regions of the country in the few years since its release in the year 2008. This variety was tested and also recommended for cultivation in Eastern and Central India for processing the commercial wheat flour under the PAU-ITC (Indian Tobacco Company) agreement. However, the evolution of the 78S84 pathogen was so targeted and led to the development of a very aggressive strain which not only rendered PBW343 susceptible to stripe rust but also slowly wrapped up PBW550 including all the newly released varieties during that time. As a result, PBW550 too became susceptible to yellow rust after 6–7 years of release despite having multi-pathogen resistance gene loci *Lr34/Yr18/Sr57*. The short-lived resistance makes it of utmost importance to introgress rust resistance in addition to yield improvement in these cultivars.

Though HGW has been introduced with backcross breeding only while the rust resistance was facilitated through MAS. Increased adoption of combining conventional breeding with the MAS approach (Gupta et al., 2010) in recent years has given multifold advantages, the major one being accelerated mobilization of the desirable gene(s) and their efficient stacking in elite backgrounds. Two different NILs carrying stripe rust resistance gene *Yr15* and linked leaf rust-stripe rust resistance gene *Lr57-Yr40* have been developed by PAU. NIL PBW550 + *Yr15* has also been released as a variety for cultivation in Punjab under timely sown irrigated conditions as *Unnat* PBW550. *Yr15* mapped near distal Nor-B1 on the short arm of chromosome 1B was first discovered in the 1980s in wild emmer wheat *T. turgidum var. dicoccoides*. Avocet + *Yr15* was initially used as a donor to transfer this gene in PBW550 background. As per the Global Rust Reference Centre[2], *Yr15* has been providing a complete and broad spectrum resistance (Klymiuk et al., 2018). Linked *Lr57-Yr40* (on chromosome 5DS) has been transferred from an introgression line INGR15046, developed in the wheat wide hybridization program of PAU by transfer of leaf rust and stripe rust resistance from *Aegilopes geniculata* to wheat cultivar WL711 (Kuraparthy et al., 2007). This linked gene provides complete resistance to leaf rust and partial resistance (20MR) to stripe rust in field.

[2]https://agro.au.dk/forskning/internationale-platforme/wheatrust/

For pyramiding, the "transfer first and assemble later" approach (Ishii and Yonezawa, 2007a,b) already given above was followed where PBW550 + *Yr15* and PBW550 + *Lr57-Yr40*, NILs were used as donors of respective genes. Since both the parental genotypes were in the PBW550 background, backcrossing was not required. Similarly, HGW550 + *Yr15* and PBW550 + *Lr57-Yr40* were crossed, and selfing generations F_2, F_3, and F_4 were evaluated for HGW and two resistances to leaf and stripe rust leading to the selection of 11 HGW550 + *Yr15* + *Lr57-Yr40* progenies with three rust resistance genes and high TGW. Gene pyramiding of disease resistance genes reported improving the durability of resistance, though the combined effect depends upon the nature of the individual genes and their synergism. Resistance breeding by marker led pyramid in the last decade is being used in wheat programs globally. Moreover, the combination of major and minor genes was also found to have more significant effects. Although *Yr15* has already succeeded in conferring stripe rust resistance for many years in different introgression lines around the world, marker assisted gene pyramiding of *Yr15,* and *Lr57-Yr40* genes provides durable resistance to stripe rust in combination with leaf rust resistance. Advanced breeding lines of PBW550 with three rust resistance genes and high TGW can lead to the development of new varieties and also serve as valuable germplasm for breeders to be used in the varietal development program to aid future breeding programs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s. The raw data for trials has been added as **Supplementary Table S1**.

AUTHOR CONTRIBUTIONS

SK conceived and designed the research, planned, and conducted the experiments, did rust screening, compiled the work, and produced the final draft of the manuscript. JK and RS conducted field experiments and did marker assisted selection. GM and AS conducted the field experiments. GD did statistical analysis and compilation of the manuscript. UD helped in field experiments. PC conceived and designed the research, and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Department of Biotechnology (DBT) under the grant number BT/PR21024/AGIII/103/925/2016.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.593426/full#supplementary-material

Supplementary Figure 1 | (A) PCR amplification profile of marker *Xuhw302* amplifying *Yr15* gene in F_3 plants derived from cross HGW550 × PBW550 + Yr15 as on agarose gel. **(B)** KlusterCaller output view of segregation of Co-dominant *Ta5DS-2754099_kasp23* markers in F_3 plants derived from cross HGW550 + Yr15 × PBW550 + *Lr57-Yr40*. FAM tailed *Lr57-Yr40* (blue color) allele on *X*-axis, HEX tailed PBW550 alleles (red color) on *Y*-axis, heterozygous individuals on mid-axis (green color). Black color represents non-template control and pink color represents unamplified or unclustered samples.

Supplementary Table 1 | Raw data of thousand grain weight in three-year replicated trials as recorded on BC_2F_{3-5} progenies of cross PBW343 X Rye Sel 111 and BC_2F_{3-5} progenies of cross PBW550 × HGW343 as Env1, Env2 and Env3. The Env 4 Represents the adjusted means of the three years.

REFERENCES

Aboukhaddour, R., Fetch, T., McCallum, B. D., Harding, M. W., Beres, B. L., and Graf, R. J. (2020). Wheat diseases on the prairies: a Canadian story. *Plant Pathol.* 69, 418–432. doi: 10.1111/ppa.13147

Aisawi, K. A. B., Reynolds, M. P., Singh, R. P., and Foulkes, M. J. (2015). The physiological basis of the genetic progress in yield potential of CIMMYT spring wheat cultivars from 1966 to 2009. *Crop Sci.* 55, 1749–1764. doi: 10.2135/cropsci2014.09.0601

Babu, P., Baranwal, D. K., Harikrishna, Pal, D., Bharti, H., Joshi, P., et al. (2020). Application of genomics tools in wheat breeding to attain durable rust resistance. *Front. Plant Sci.* 11:567147. doi: 10.3389/fpls.2020.567147

Bansal, M., Adamski, N. M., Toor, P. I., Kaur, S., Molnár, I., Holušová, K., et al. (2020). *Aegilops umbellulata* introgression carrying leaf rust and stripe rust resistance genes *Lr76* and *Yr70* located to 9.47-Mb region on 5DS telomeric end through a combination of chromosome sorting and sequencing. *Theor. Appl. Genet.* 133, 903–915. doi: 10.1007/s00122-019-03514-x

Beche, E., Benin, D., Da Silva, C. L., Munaro, L. B., and Marchese, J. A. (2014). Genetic gain in yield and changes associated with physiological traits in Brazilian wheat during the 20th century. *Eur. J. Agron.* 61, 49–59. doi: 10.1016/j.eja.2014.08.005

Bhawar, K., Vinod Sharma, J., Singh, A., Sivasamy, M., Singh, M., Prabhu, K., et al. (2011). Molecular marker assisted pyramiding of leaf rust resistance genes Lr19 and Lr28 in wheat variety HD2687. *Ind. J. Genet. Plant Breed.* 71, 304–311.

Djanaguiraman, M., Prasad, P.V. V., Kumari, J., Sehgal, S. K., Friebe, B., Djalovic, I., et al. (2019). Alien chromosome segment from *Aegilops speltoides* and *Dasypyrum villosum* increases drought tolerance in wheat via profuse and deep root system. *BMC Plant Biol.* 19, 242–257. doi: 10.1186/s12870-019-1833-8

Gao, F., Ma, D., Yin, G., Rasheed, A., Dong, Y., Xiao, Y., et al. (2017). Genetic progress in grain yield and physiological traits in Chinese wheat cultivars of southern Yellow and Huai Valley since 1950. *Crop Sci.* 57, 760–773. doi: 10.2135/cropsci2016.05.0362

Gautam, T., Dhillon, G. S., Saripalli, G., Rakhi, Singh, V. P., Prasad, P., et al. (2020). Marker-assisted pyramiding of genes/QTL for grain quality and rust resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 40:49. doi: 10.1007/s11032-020-01125-9

Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., et al. (2010). A genetic framework for grain size and shape variation in wheat. *Plant Cell* 22, 1046–1056. doi: 10.1105/tpc.110.074153

Giunta, F., Motzo, R., and Pruneddu, G. (2007). Trends since 1900 in the yield potential of Italian-bred durum wheat cultivars. *Eur. J. Agron.* 27, 12–24. doi: 10.1016/j.eja.2007.01.009

Giura, A., and Saulescu, N. N. (1996). Chromosomal location of genes controlling grain size in a large grained selection of wheat (*Triticum aestivum* L.). *Euphytica* 89, 77–80. doi: 10.1007/bf00015722

Gupta, A., Singh, C., Kumar, V., Tyagi, B. S., Tiwari, V., Chatrathand, R., et al. (2018). *Wheat Varieties Notified in India Since 1965*. Karnal: ICAR- Indian Institute of Wheat & Barley Research, 101.

Gupta, P. K., Kumar, J., Mir, R. R., and Kumar, A. (2010). Marker-assisted selection as a component of conventional plant breeding. *Plant Breed. Rev.* 33, 145–217. doi: 10.1002/9780470535486.ch4

Ishii, T., and Yonezawa, K. (2007a). Optimization of the marker-based procedures for pyramiding genes from multiple donor lines: I. Schedule of crossing between the donor lines. *Crop Sci.* 47, 537–546. doi: 10.2135/cropsci2006.06.0435

Ishii, T., and Yonezawa, K. (2007b). Optimization of the marker-based procedures for pyramiding genes from multiple donor lines: II. Strategies for selecting the objective homozygous plant. *Crop Sci.* 47, 1878–1886. doi: 10.2135/cropsci2006.11.0750

Joshi, L. M., Srivastava, K. D., and Ramanujam, K. (1975). An analysis of brown rust epidemics of 1971- 72 and 1972-73. *Indian Phytopathol.* 28:137.

Jindal, M. M., Sharma, I., and Bains, N. S. (2012). Losses due to stripe rust caused by *Puccinia striiformis* in different varieties of wheat. *J. Wheat Res.* 4, 33–36.

Klymiuk, V., Yaniv, E., Huang, L., Raats, D., Fatiukha, A., Chen, S., et al. (2018). Cloning of the wheat Yr15 resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nat. Commun.* 9:3735. doi: 10.1038/s41467-018-06138-9

Kolmer, J. A. (1996). Genetics of resistance to wheat leaf rust. *Annu. Rev. Phytopathol.* 34, 435–455. doi: 10.1146/annurev.phyto.34.1.435

Kuchel, H., Fox, R., Reinheimer, J., Mosionek, L., Willey, N., Bariana, H., et al. (2007a). The successful application of a marker-assisted wheat breeding strategy. *Mol. Breed.* 20, 295–308. doi: 10.1007/s11032-007-9092-z

Kuchel, H., Williams, K. J., Langridge, P., Eagles, H. A., and Jefferies, S. P. (2007b). Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor. Appl. Genet.* 115, 1029–1041.

Kumar, J., Mir, R. R., Kumar, N., Kumar, A., Mohan, A., Prabhu, K. V., et al. (2010). Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. *Plant Breed.* 129, 617–621. doi: 10.1111/j.1439-0523.2009.01758.x

Kumar, N., Kulwal, P. L., Gaur, A., Tyagi, A. K., Khurana, J. P., Khurana, P., et al. (2006). QTL analysis for grain weight in common wheat. *Euphytica* 151, 135–144. doi: 10.1007/s10681-006-9133-4

Kuraparthy, V., Chhuneja, P., Dhaliwal, H. S., Kaur, S., Bowden, R. L., and Gill, B. S. (2007). Characterization and mapping of cryptic alien introgression from *Aegilops geniculata* with new leaf rust and stripe rust resistance genes Lr57 and Yr40 in wheat. *Theor. Appl. Genet.* 114, 1379–1389. doi: 10.1007/s00122-007-0524-2

McIntosh, R. A., Wellings, C. R., and Park, R. F. (1995). *Wheat Rusts: An Atlas of Resistance Genes CSIRO Publications*, Vol. 1. Melbourne: CSIRO Publications, doi: 10.1007/bf03214019

Mehta, K. C. (1950). Control of rust epidemics of wheat in India- a national emergency. *Sci. Cult.* 15, 263–270.

Mundt, C. C. (2018). Pyramiding for resistance durability: theory and practice. *Phytopathology* 108, 792–802. doi: 10.1094/phyto-12-17-0426-rvw

Nagarajan, S. (2005). Can India produce enough wheat even by 2020? *Curr. Sci.* 89, 1467–1471.

Nagarajan, S., and Joshi, L. M. (1975). A historical account of wheat rust epidemics in India and their significance. *Cereal Rusts Bull.* 3, 29–33.

Nayar, S. K., Prashar, M., and Bhardwaj, S. C. (1997). Mannual of current techniques in wheat rusts. *Res. Bull.* 2, 1–32. doi: 10.1007/978-94-011-0083-0_1

Pal, D., Bhardwaj, S. C., Sharma, D., Kumari, S., Patial, M., Vinod, P., et al. (2015). Assessment of genetic diversity and validating rust resistance gene sources using molecular markers in wheat (*Triticum aestivum* L.). *SABRAOJ. Breed. Genet.* 47, 89–98.

Pannu, P. P. S., Mohan, C., Singh, G., Kaur, J., Mann, S. K., Bala, G. K., et al. (2010). Occurrence of yellow rust of wheat, its impact on yield viz-a-viz its management. *Plant Dis. Res.* 25, 144–150.

Pavithra, S., Mittal, S., Bhat, S. A., Birthal, P. S., Shah, S. A., and Hariharan, V. K. (2017). Spatial and temporal diversity in adoption of modern wheat varieties in India. *Agric. Econ. Res. Rev.* 30, 57–72. doi: 10.5958/0974-0279.2017.00005.2

Peng, J., Ronin, Y., Fahima, T., Röder, M. S., Li, Y., Nevo, E., et al. (2003). Domestication quantitative trait loci in Triticum dicoccoides, the progenitor of wheat. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2489–2494. doi: 10.1073/pnas.252763199

Prashar, M., Bhardwaj, S. C., Jain, S. K., and Datta, D. (2007). Pathotypic evolution in Puccinia striiformis in India during 1995-2004. *Aust. J. Agric. Res.* 58, 603–604.

Qin, X., Zhang, F., Liu, C., Yu, H., Cao, B., Tian, S., et al. (2015). Wheat yield improvements in China: past trends and future directions. *Field Crops Res.* 177, 117–124. doi: 10.1016/j.fcr.2015.03.013

Randhawa, M. S., Bains, N. S., Sohu, V. S., Chhuneja, P., Trethowan, R. M., Bariana, H. S., et al. (2019). Marker assisted transfer of stripe rust and stem rust resistance genes into four wheat cultivars. *Agronomy* 9:497. doi: 10.3390/agronomy9090497

Reynolds, M., Foulkes, M. J., Slafer, G. A., Berry, P., Parry, M. A. J., Snape, J. W., et al. (2009). Raising yield potential in wheat. *J. Exp. Bot.* 60, 1899–1918. doi: 10.1093/jxb/erp016

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 81, 8014–8028. doi: 10.1073/pnas.81.24.8014

Sidwell, R. J., Smith, E. L., and McNew, R. W. (1976). Inheritance and interrelationships of grain yield and selected yield-related traits in a hard red winter wheat cross. *Crop Sci.* 16, 650–654. doi: 10.2135/cropsci1976.0011183x001600050013x

Singh, H., Kaur, J., Bala, R., Srivastava, P., and Bains, N. S. (2020). Virulence and genetic diversity of Puccinia striiformis f. sp. tritici isolates in sub-mountainous area of Punjab, India. *Phytoparasitica.* 48, 383–395. doi: 10.1007/s12600-020-00809-4

Singh, R. P., Hodson, D. P., Jin, Y., Lagudah, E. S., Ayliffe, M. A., Bhavani, S., et al. (2015). Emergence and spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control. *Phytopathology* 105, 872–884. doi: 10.1094/phyto-01-15-0030-fi

Singh, R. P., Huerta-Espino, J., Sharma, R., Joshi, A. K., and Trethowan, R. (2007). High yielding spring bread wheat germplasm for global irrigated and rainfed production systems. *Euphytica* 157, 351–363. doi: 10.1007/s10681-006-9346-6

Slafer, G. A., Savin, R., and Sadras, V. O. (2014). Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Res.* 157, 71–83. doi: 10.1016/j.fcr.2013.12.004

Tomar, S. M. S., Singh, S. K., Sivasamy, M., and Vinod. (2014). Wheat rusts in India: resistance breeding and gene deployment - A review. *Indian J. Genet.* 74, 129–156. doi: 10.5958/0975-6906.2014.00150.3

Toor, P. I., Kaur, S., Bansal, M., Yadav, B., and Chhuneja, P. (2016). Mapping of stripe rust resistance gene in an *Aegilops caudata* introgression line in wheat and its genetic association with leaf rust resistance. *J. Genet.* 95, 933–938. doi: 10.1007/s12041-016-0718-y

Tyagi, S., Mir, R. R., Kaur, H., Chhuneja, P., Ramesh, B., Balyan, H. S., et al. (2014). Marker-assisted pyramiding of eight QTLs/genes for seven different traits in common wheat (*Triticum aestivum* L.). *Mol. Breed.* 34, 167–175. doi: 10.1007/s11032-014-0027-1

Ye, G. B., and Smith, K. (2008). Marker-assisted gene pyramiding for inbred line development: basic principles and practical guidelines. *Int. J. Plant Breed.* 2, 1–10. doi: 10.17221/51/2016-cjgpb

Zhang, Y., Xu, W., Wang, W., Dong, H., Qi, X., Zhao, M., et al. (2016). Progress in genetic improvement of grain yield and related physiological traits of Chinese wheat in henan province. *Field Crops Res.* 199, 117–128. doi: 10.1016/j.fcr.2016.09.022

Zheng, S., Li, Y., Lu, L., Liu, Z., Zhang, C., Ao, D., et al. (2017). Evaluating the contribution of Yr genes to stripe rust resistance breeding through marker-assisted detection in wheat. *Euphytica* 213:50.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kaur, Kaur, Mavi, Dhillon, Sharma, Singh, Devi and Chhuneja. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fgene.2020.581214

Genome-Wide Association Study of Tan Spot Resistance in a Hexaploid Wheat Collection From Kazakhstan

Alma Kokhmetova[1,2], Deepmala Sehgal[3]*, Shaukat Ali[4], Makpal Atishova[1], Madina Kumarbayeva[1,2], Irina Leonova[5] and Susanne Dreisigacker[3]*

[1] Laboratory of Breeding and Genetics, Institute of Plant Biology and Biotechnology (IPBB), Almaty, Kazakhstan, [2] Faculty of Agronomy, Kazakh National Agrarian University, Almaty, Kazakhstan, [3] Global Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [4] Department of Agronomy, Horticulture & Plant Science, South Dakota State University, Brookings, SD, United States, [5] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

Tan spot, caused by *Pyrenophora tritici-repentis*, is a serious foliar disease of wheat in Kazakhstan with reported yield losses as high as 50% during epidemic years. Here, we report the evaluation of a collection of 191 hexaploid spring and winter wheat lines for tan spot resistance and its underlying genetic architecture using genome-wide association study (GWAS). Our wheat collection comprised candidate varieties from Kazakhstan, Russia, and CIMMYT. It was genotyped using the DArTseq technology and phenotyped for resistance to tan spot at seedling and adult plant stages in Kazakhstan. DArTseq SNPs revealed high genetic diversity (average polymorphic information content = 0.33) in the panel and genome-wide linkage disequilibrium decay at 22 Mb (threshold $r^2 = 0.1$). Principal component analysis revealed a clear separation of Eurasian germplasm from CIMMYT and IWWIP lines. GWAS identified 34 marker-trait associations (MTA) for resistance to tan spot and the amount of phenotypic variation explained by these MTA ranged from 4% to 13.7%. Our results suggest the existence of novel valuable resistant alleles on chromosomes 3BS, and 5DL and 6AL for resistance to Race 1 and Race 5, respectively, in addition to known genes *tsn1* and *tsc2*. On chromosome 6AL, a genomic region spanning 3 Mb was identified conferring resistance to both Race 1 and Race 5. Epistatic interaction of associated loci was revealed on chromosomes 1B, 5B, 7B, 5A, and 6A contributing to additional variation of 3.2–11.7%. Twenty-five lines with the best allele combinations of SNPs associated with resistance to both races have been identified as candidates for future variety release and breeding. The results of the present study will be further validated in other independent genetic backgrounds to be able to use markers in breeding.

Keywords: DArTseq, genome-wide association study, *Pyrenophora tritici-repentis*, tan spot, wheat

OPEN ACCESS

Edited by:
Vijay Kumar Tiwari, University of Maryland, College Park, United States

Reviewed by:
Vikas Belamkar, University of Nebraska–Lincoln, United States
Ze Peng, University of Florida, United States

****Correspondence:***
Alma Kokhmetova
akokhmetova@gmail.com;
gen_kalma@mail.ru
Deepmala Sehgal
d.sehgal@cgiar.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *08 July 2020*
Accepted: *09 December 2020*
Published: *11 January 2021*

Citation:
Kokhmetova A, Sehgal D, Ali S, Atishova M, Kumarbayeva M, Leonova I and Dreisigacker S (2021) Genome-Wide Association Study of Tan Spot Resistance in a Hexaploid Wheat Collection From Kazakhstan. Front. Genet. 11:581214. doi: 10.3389/fgene.2020.581214

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is grown in more than 85 countries with a gross annual production of 761.5 million tons (FAO, 2020). Consumed by more than 40% world population, it is the primary source of calories for millions of people worldwide. Central Asia, including Kazakhstan, plays a significant role in regional and global food security as most of the grain produced is traded

in these regions (Morgounov et al., 2014). The total area sown to wheat in Kazakhstan represents over 85% of total cereal production. Tan spot caused by *Pyrenophora tritici-repentis* (Died.) is an economically important disease in most wheat-growing regions worldwide, including Europe, North and South America, Australia, and Asia (Duveiller et al., 2005). In Central Asia, the disease was discovered in the 1980s in Tajikistan (Khasanov, 1988), and since then, it has spread throughout Central Asia and Kazakhstan (Postnikova and Khasanov, 1998; Lamari et al., 2005; Koyshibaev, 2018). On an average, losses due to tan spot vary from 10% to 15% but may reach up to 50% during epidemic years (Rees et al., 1982; Shabeer and Bockus, 1988). Tan spot reduces total yield, grain weight, number of grains per head, total biomass, and grain quality (Shabeer and Bockus, 1988; Fernandez et al., 1994).

Currently, eight races of *P. tritici-repentis* (PTR) have been identified based on necrosis and chlorosis symptoms induced by host-selective toxins (HST) on a set of differential wheat varieties (Lamari et al., 2003). Races 2, 3, and 5 can be designated as basic races, while Races 1, 6, 7, and 8 are combinations of the three basic races except for Race 4, which is avirulent (Lamari et al., 2003). To date, three host-specific toxins, Ptr ToxA, Ptr ToxB, and Ptr ToxC, of *P. tritici-repentis* have been identified and well characterized. Ptr ToxA is produced by Races 1, 2, 7, and 8 (Lamari et al., 2003) and associated with necrotic symptoms in Ptr ToxA-sensitive cultivars. Ptr ToxB is produced by Races 5, 6, 7, and 8 and is responsible for the induction of chlorosis in Ptr Tox-B-sensitive cultivars. Both Ptr ToxA and Ptr ToxB are proteins in nature, while Ptr ToxC (produced by Race 1 and Race 3) is a non-ionic, polar non-protein (Effertz et al., 2002).

There are a number of studies conducted on the racial composition of *P. tritici-repentis* worldwide (Lamari and Bernier, 1989; Ali and Francl, 2002, 2003; Lamari et al., 2003, 2005; Maraite et al., 2006; Gamba et al., 2012; Kokhmetova et al., 2016, 2017). The greatest diversity in the pathogen population was observed in Azerbaijan, where Races 1, 2, 3, 5, 7, and 8 were identified, and in Syria, where Races 1, 3, 5, 7, and 8 were observed. Race 1 is the most widespread race in Central Asia and Kazakhstan (87%), and Races 2, 3, and 4 are prevalent infrequently (Maraite et al., 2006). Recently, Race 8 was also found in high frequency in Kazakhstan (Kokhmetova et al., 2016, 2017).

The inheritance of resistance to tan spot is both qualitative and quantitative; toxicity resistance genes and quantitative trait loci (QTL) are known (Faris et al., 1996, 2013; Gamba and Lamari, 1998; Anderson et al., 1999; Faris and Friesen, 2005; Tadesse et al., 2006a,b; Chu et al., 2008, 2010; Singh et al., 2008, 2010, 2016; Li et al., 2011; Hu et al., 2019). Qualitative genes identified through conidial inoculations have been given the designation "*Tsr*" for "tan spot resistance," and genes associated with reaction to HST-containing cultures are designated as "*Tsc*" or "*Tsn*" depending on the necrosis or chlorosis symptom exhibited by the HST (McIntosh et al., 2008). To date, eight major *Tsr* genes (*Tsrl*, *Tsr2*, *Tsr3*, *Tsr4*, *Tsr5*, *Tsr6*, *TsrHar*, and *TsrAri*) located on chromosomes 2BS, 3AS, 3BL, 3DS, and 5BL have been identified (McIntosh et al., 2013). Two *Tsc* genes (*Tsc1 and Tsc2*), conferring sensitivity to Ptr ToxC and Ptr ToxB, have been mapped on chromosomes 1AS (Effertz et al., 2002) and 2BS (Friesen and Faris, 2004; Abeysekara et al., 2010), respectively.

With the availability of millions of single-nucleotide polymorphisms (SNPs) in almost all crops, genome-wide association study (GWAS) has become a common approach in dissecting genetic architecture of traits and in identifying beneficial alleles for use in marker-assisted selection (Santure and Garant, 2018). In wheat, GWAS has been conducted for several diseases, including resistance to *Stagonospora nodorum blotch* (Tommasini et al., 2007), stem rust (Bajgain et al., 2016; Elbasyoni et al., 2017), stripe rust (Yu et al., 2011, 2012; Sehgal et al., 2016), fusarium head blight (Miedaner et al., 2011), and *Septoria tritici blotch* (Gerard et al., 2017). GWAS studies for tan spot resistance have led to significant advances in the identification of loci on chromosomes 2B, 3B, 4A, 6B, 6A, and 7B (Gurung et al., 2011; Patel et al., 2013; Kollers et al., 2014; Singh et al., 2016; Hu et al., 2019). However, hitherto, the genetic basis of tan spot resistance in wheat germplasm from Kazakhstan has not been yet investigated. Our objectives were therefore (1) to evaluate a Kazakhstan collection of winter and spring wheat cultivars/breeding lines for tan spot resistance, (2) to identify genetic loci associated with tan spot resistance via GWAS, (3) to explore gene-by-gene interactions among significant genetic loci, and (4) to identify best combinations of alleles and lines for future phenotyping trials and breeding.

MATERIALS AND METHODS

Plant Materials

A total of 191 spring and winter hexaploid wheat accessions (*Triticum aestivum* L.) were evaluated in this study. The collection included 111 cultivars and breeding lines from Kazakhstan, 17 cultivars from Russia, 1 cultivar from Brazil, 30 lines released by CIMMYT, and 31 lines released by CIMMYT-ICARDA-IWWIP (International Maize and Wheat Improvement Center-International Center for Agricultural Research in the Dry Areas–International Winter Wheat Improvement Program) (**Supplementary Table 1**). The collection comprised important wheat genotypes that have been widely used as parental lines in breeding programs across the Kazakhstan and Central Asian countries. The 111 cultivars from Kazakhstan included in the collection were chosen based on their contrasting phenotypic expression for traits of agronomic and disease resistance. Three differential lines, cultivar "Glenlea" carrying the *Tsc1* gene, line "6B662" carrying *Tsc2*, and cultivar Salamouni resistant to all known races and insensitive to toxins Ptr ToxA, Ptr ToxB, and Ptr ToxC were included as checks.

Fungal Isolates, Inoculum Production, and Inoculations

The isolates Pti2 and DW7 used in this study were previously obtained from bread wheat (*T. aestivum* L.) and durum wheat (*Triticum durum* Desf), respectively, and stored as dried mycelial plugs at −20°C (Jordahl and Francl, 1992). The isolate Pti2 was from spring wheat and the isolate DW7 was collected from a durum wheat field in North Dakota (Ali and Francl, 2003),

and they were used for inoculum production in this study. The inoculum and inoculations were carried out as described in Ali and Francl (2001). To prepare the inoculum, a single mycelial plug (0.5 cm in diameter) was placed on V8PDA (150 ml of V8 juice, 10 g of Difco PDA, 10 g of Difco agar, 3 g of calcium carbonate, and 850 ml of sterile distilled water) (Lamari and Bernier, 1989) in 9-cm petri plates. After placing the mycelia plug of both isolates individually in petri plates, the plates were wrapped with aluminum foil and incubated at 21°C for 5 days. Thereafter, the petri plates were filled with sterilized distilled water, and hyphal growth was knocked down with a flamed-sterilized glass test tube. After suppressing the hyphae, excess water was removed from the plates, and then they were incubated in an alternate cycle of 24 h light at 22°C and 24 h dark at 16°C to induce conidiophores and conidia. The conidia were dislodged with an inoculating loop wired needle. The conidial suspension was obtained by adding 30 ml of distilled sterile water in each plate, and conidia were dislodged with a looped wire needle. The conidial suspension was adjusted to 3,000 spores/ml using the hemocytometer. One drop of Tween 20 was added/100 ml of conidial suspension before inoculation. Inoculated seedlings were moved to a mist chamber at 21°C with a 16-h photoperiod for 24 h to enhance the chances of infection. The chamber was misted for 16 s at 3-min intervals to keep 100% RH. The plants then were moved to a growth chamber at 22°C.

Seedling Test

Seedlings of all 191 wheat accessions were raised in containers (3.8 cm in diameter and 20 cm long) as described in Ali and Francl (2001). For each race of *P. tritici-repentis* (Races 1 and 5), separate experiments were conducted in the growth chamber. Three seeds were planted in a plastic container (Stuewe & Sons, Inc., OR) filled with Sunshine Mix #1 (Fison Horticulture, Vancouver, BC, Canada). All containers were placed in 96-slot racks to hold them straight. Each container was considered as an experimental unit, and each single plant in a container with three seedlings served as an entry. All entries were arranged in a randomized complete design with three replications. Thus, nine seedlings of each genotype were evaluated individually at the two-leaf stage against Race 1 (isolate Pti2) and Race 5 (isolate DW7). In each experiment, replications were treated as random effect and the wheat accessions as fixed effects. All experiments were conducted in growth chambers at South Dakota State University (SDSU), Brookings, SD, in 2015. Seedlings were rated 8 days post-inoculation using the rating scale 1–5 (Lamari and Bernier, 1989) where 1–2 is resistant to moderately resistant, 3 is moderately susceptible, and 4–5 is susceptible. BLUE (best linear unbiased estimators) across replications were calculated in META-R (Vargas et al., 2013).

Phenotyping for Sensitivity to Ptr ToxA and Ptr ToxB

The 191 wheat entries and three checks were tested for their reaction to purified toxins Ptr ToxA and Ptr ToxB at a concentration of 10 μg/ml. Ptr ToxA and Ptr ToxB were kindly provided by Dr. Steven Meinhardt, North Dakota State University, Fargo, and Dr. Timothy Friesen, USDA, Fargo, ND. Four leaves (second leaf fully expanded), of each genotype were infiltrated as described by Faris et al. (1996) with pure Ptr ToxA and Ptr ToxB culture filtrates separately. After infiltration, the plants were kept in a growth chamber at 21°C during the day and 18°C at night with 16 h photoperiod in the growth chamber. Leaves evaluated 4 days post infiltration and scored as insensitive (−) or sensitive (+). This experiment was repeated twice. Wheat genotypes "Glenlea" (ToxA sensitive), "6B662" (ToxB sensitive), and "Salamouni" (ToxA and ToxB insensitive) were also included in the experiment to verify the results and toxin viability (Ali and Francl, 2003; Lamari et al., 2003).

Field Phenotyping

The experimental material was phenotyped at the Kazakh Research Institute of Agriculture and Plant Growing (KRIAPG), Almalybak (43°13′09″ N, 76°36′17″ E) in Southeast Kazakhstan, Almaty region, during the 2016 to 2018 cropping seasons. Experiments were conducted as a completely randomized design with three replicates. Individual plot size was 1 m^2. Fertilizer treatments, 60 and 30 kg/ha of N and P_2O_5, respectively, and other management practices were corresponded to those normally recommended for the region (Dospekhov, 1985).

Experiments were planted in mid-September and were harvested in mid-August the following year for 3 years. The weather conditions at Almalybak were characterized annually by over 400 mm of rainfall. The irrigated foothill zone, where KRIAPG is located, is a relatively high moistured location; the experimental wheat materials were irrigated thrice during the growing season and were kept free from weeds.

The field evaluation was carried out under natural epidemic conditions in 2016 and 2017, whereas in 2018, it was carried out under natural conditions as well as using artificial inoculation. Ten Flag-1 leaves were evaluated for each disease assessment of genotypes. The disease was assessed three times. The experiments on an artificial infectious background were made with naturally infected straw stubbles. In October, before sowing, the infected straw (1 kg/m^2) was incorporated into the soil. For evaluation of field response, disease severities were assessed on first leaves and flag leaves in GS 65–69, Zadoks scale (Zadoks et al., 1974). The percentage of Ptr-infected leaf area was determined on each leaf, and the average value for all evaluated leaves was calculated for each wheat entry in order to determine the PTR score. A rating system based on% leaf area infected developed for appraising the foliar intensity of diseases (Kremneva and Volkova, 2007) was used to categorize host reaction to *P. tritici-repentis*. This scale of disease severity was rated numerically based on% necrotic or chlorotic area as follows: 0–10%—resistant (R), 11–20%—moderately resistant (MR), 21–30%—moderately susceptible (MS), and 31–100%—susceptible (S). The standard wheat differentials included "Glenlea" (susceptible check) and "Salamouni" (resistant check) and were included in the field trials. Plant height (PH) was recorded in centimeters from the soil surface to the tip of the spike of 10–15 plants per plot in GS 90–99. Days to heading (DH) were recorded as the number of days from planting to the 50% spike emergence in GS 49–50 (Zadoks et al., 1974).

Analysis of phenotypic resistance ratings to tan spot based on the area under disease progress curve (AUDPC) was performed. AUDPC was calculated annually by summarizing the progress of disease severity. AUDPC values from double digit and AUDPC from flag leaf (F) and penultimate leaf (F-1) were separately calculated by using the following formula described by Das et al. (1992). ANOVA analyses of phenotypic data was done using linear mixed effect model with package *lme4* in R using replications as fixed effect and entries as random effect. The package *reshape* was used to transform the data for ANOVA analysis. Broad-sense-heritability was estimated using the formula: $h^2 = V_g/(V_g + V_{err}/r)$, where V_g is the genotypic variance, V_{err} is the error variance, and r = the number of replications.

DNA Extraction and Genotyping

Genomic DNA was extracted from fresh leaves collected from a single individual plant using a modified CTAB (cetyltrimethylammonium bromide) method described in the CIMMYT laboratory protocols (Dreisigacker et al., 2012, 2016) and quantified using a NanoDrop 8000 spectrophotometer V 2.1.0. Genotyping was performed using the DArTseq™ technology provided by the Genetic Analysis and Service for Agriculture (SAGA) laboratory in Mexico. Briefly, the genotypes were sequenced at 192-plexing on Illumina HiSeq2500 with 1 × 77-bp reads. A proprietary analytical pipeline developed by DArT P/L was used to generate allele calls for SNP and presence/absence variation (PAV) markers (Sansaloni et al., 2011). A 100K consensus map provided by SAGA was used to obtain genetic positions of the SNPs (Sansaloni et al. unpublished). To obtain physical positions of SNPs, sequence reads of the SNPs were blasted to the reference genome of RefSeq V1.0 in the Ensemble Plants database[1].

Diversity, Population Structure, and Linkage Disequilibrium Analysis

Polymorphic information content (PIC) (Botstein et al., 1980) was calculated to characterize the genetic diversity of the panel using PowerMarker version 3.25 (Liu and Muse, 2005). Population structure was investigated by principal component analysis (PCA) using the STATS package in R and using STRUCTURE v 2.3.4 (Pritchard et al., 2000). The STRUCTURE program was run by setting replication number to 50,000 for the burn-in and the Markov Chain Monte Carlo (MCMC) periods. *K* values were run from 1 to 7 using "admixture" and "correlated allele frequency" models. The correct estimation of *K* was provided by an *ad hoc* statistic delta *K*, calculated using the program STRUCTURE HARVESTER[2]. A weighted neighbor joining tree was constructed in DARwin version 6.0[3].

GAPIT v. 2.0 (Lipka et al., 2012) was used to obtain squared correlation coefficient (r^2), a measure of linkage disequilibrium (LD), for all pairwise comparisons among markers. Pattern of LD decay was visualized in individual subpopulations and in the whole panel by plotting pair-wise r^2 values against the physical distance. A smooth line was fit to the data using second-degree locally weighted scatterplot smoothing, LOESS (Breseghello and Sorrells, 2006) as implemented in SAS. For the LOESS estimation of LD decay, genetic distance was estimated as the point where the LOESS curve first crosses the baseline r^2 of 0.1.

Genome-Wide Association Analysis

The BLUE and average scores for disease severity at seedling stage and AUDPC values at adult plant stage in the field were used for conducting GWAS. To map Ptr ToxA and Ptr ToxB resistance loci, all insensitive genotypes were scored as 1 and all sensitive genotypes as 5, and the average of the two scores was used for GWAS. VanRaden algorithm (VanRaden, 2008) was used to calculate the kinship matrix in the GAPIT package vs. 2.0 (Lipka et al., 2012). GWAS was conducted in the TASSEL software vs. 5 (Bradbury et al., 2007). A mixed linear model (MLM) was applied in which PCA was a fixed variate (first three PCs) and kinship as random. A false-discovery rate (FDR) was used to assess the significance of the p value (<0.05). The allelic effect of significant marker-trait associations was estimated as the difference between the mean value of lines with and without favorable allele and was presented as box plots. A second GWAS analysis was conducted in which we included 8,947 unmapped markers along with a filtered set of markers. Interestingly, eight unmapped markers crossed the FDR threshold, but these showed very low R^2 values, ranging from 2.2 to 3.1%. We have not shown these results.

Epistatic Interaction Analysis

Two- and three-locus epistatic interactions among associated SNPs and among genome-wide SNPs were explored using a custom-made in-house script described in Sehgal et al. (2017). Briefly, a stepwise multiple regression was performed using a linear regression model to calculate *P* values for pairwise marker interactions. A threshold of $P < 0.0001$ was used to declare significant marker–marker interactions. The parameter R^2 was used to describe percentage variation explained by the significant interactions. The interactions showing percentage variation <10% were discarded from the output.

RESULTS

Resistance to *Pyrenophora tritici-repentis* Race 1 and Race 5 at Seedling Stage

Uniform and consistent tan spot development was observed in the seedling evaluation in growth chambers. ANOVA showed significant differences among genotypes ($P < 0.001$) for the reaction scores to both races (**Supplementary Table 2**). The checks Glenlea and 6B662 developed necrosis and chlorosis to Race 1 and Race 5, respectively, and verified the Race identity and the success of inoculation. Salamouni did not develop any symptoms and exhibited resistant reaction. Salamouni and Glenlea showed scores of 1.0 and 4.5 for Race 1, whereas for Race 5, Salamouni and 6B-662 showed scores of 1.1 and 4.3.

[1]https://plants.ensembl.org/Triticum_aestivum/Info/Index

[2]http://taylor0.biology.ucla.edu/structureHarvester/

[3]https://darwin.cirad.fr/product.php

Of the 191 genotypes, 50 genotypes (26.2%) revealed a disease score of less than 2 and were considered resistant to Race 1, whereas 130 wheat accessions (68.6%) were resistant to Race 5 (**Figure 1A**). Sixty-two accessions showed resistance against both races.

Sensitivity to Ptr ToxA and Ptr ToxB

One hundred thirty-two genotypes (69.1%) showed no symptoms of necrosis and were determined to be Ptr ToxA insensitive, while the remaining 59 genotypes (30.9%) were scored as sensitive (**Figure 1B**). Salamouni and Glenlea exhibited insensitive and sensitive reactions to Ptr ToxA, respectively. Ptr ToxA sensitivity was associated with disease susceptibility in field conditions in all years (Pearson's correlation $r = 0.22$ to 0.38, $P < 0.001$).

For Ptr ToxB, Salamouni and 6B-662 exhibited insensitive and sensitive reactions, respectively. Twelve genotypes (6.3%) exhibited chlorosis in the toxin-infiltrated leaf area and were rated to be Ptr ToxB sensitive. The remaining 93.7% (179) did not develop chlorosis symptoms and were rated as Ptr ToxB insensitive (**Figure 1B**). The correlation of Ptr Tox B sensitivity with disease susceptibility in the field was not significant ($r = 0.03$–0.11, $P = 0.3029$).

Field Resistance to *Pyrenophora tritici-repentis*

Of the three field experiments conducted in 2016, 2017, and 2018 under natural epidemic conditions, the disease pressure was minimum in 2017, and more than 95% of lines were resistant. Hence, phenotypic data from 2017 was not used for any further analysis, i.e., calculating AUDPC scores or GWAS. There were significant ($P < 0.001$) differences for tan spot response in the field in 2016 and 2018 (under both natural conditions and under artificial inoculation in 2018). Days to heading (DH) and plant height (PH) also showed significant variation ($P < 0.001$) (**Supplementary Table 3**). Field severity of tan spot varied between three experiments (2016–2018). The disease severity was generally higher under artificial inoculation conditions (2018_infect) than under natural inoculum conditions in 2016 and 2018. The number of accessions in each reaction class across three experiments is presented in **Figure 2**. On an average, 9, 27, 44, and 25% of genotypes were resistant (R), moderately resistant (MR), moderately susceptible (MS) and susceptible (S) in the field. Under artificial inoculation conditions in 2018 (2018_infect), the number of accessions in class S increased to 73%. Six genotypes, i.e., Tungysh, JAC161/TEMU51.80, TOO11/TOOOO7, SOMO/SORA/ACTS5, NANJING 82149/KAUZ, and ALTAR 84/AE. SQUARROSA exhibited the highest level of resistance. A significant number of wheat entries from CIMMYT (13%) demonstrated high or moderate level of resistance. Correlation of days to heading and plant height with AUDPC scores was not significant in all field seasons (**Supplementary Table 4**).

Comparison of the phenotypic data under field conditions with the green house experiment showed that out of 69 entries that showed R and MR response in field, 31 and 62 genotypes had disease scores less than 2 for Race 1 and Race 5, respectively.

Marker Distribution and Diversity

A total of 40,429 SNPs were obtained across the 191 genotypes after allele calling. Of these, 10,186 markers with missing data >20% and 8,947 unmapped markers were removed from the dataset. Further, markers with minor allele frequency (MAF) <0.05 and >0.95 were culled. Three lines, CATBIRD, BR35/BR14, and KR12-5001, showing more than 25% missing data were also culled, and a final filtered set of 8,154 SNPs on 187 lines was utilized for further analyses. Marker distribution on the 21 wheat chromosomes, PIC, and LD statistics are shown in **Table 1**. The least number of markers were distributed on the D genome (11.2%). The highest numbers of markers were found on chromosome 2B (870; 10.7%), followed by chromosome 5B (769; 9.4%) and 3B (762; 8.1%) (**Table 1**).

The average PIC for the 8,154 SNPs was 0.33. Comparison of the average PIC of winter versus spring wheat lines showed that diversity was lower in winter wheat (0.38) compared to spring wheat germplasm (0.41).

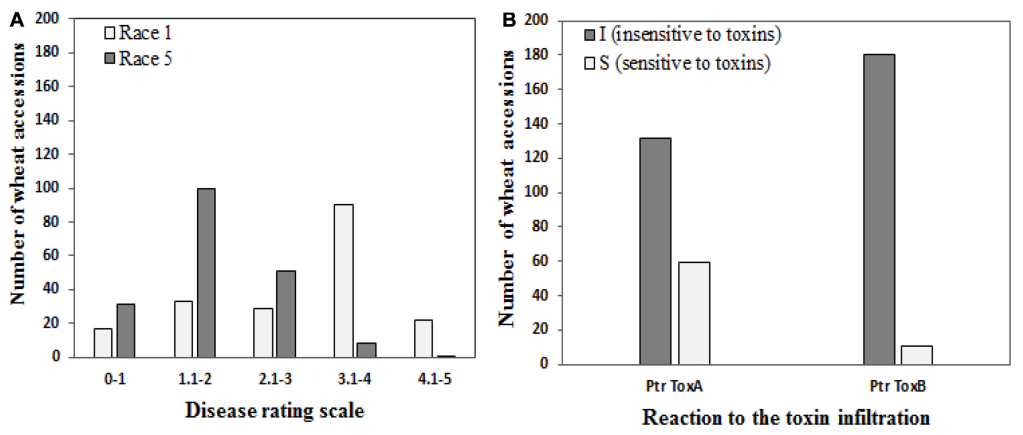


FIGURE 1 | Frequencies of 191 wheat genotypes in different disease score groups **(A)** to tan spot (*Pyrenophora tritici-repentis* Race 1 and Race 5) and **(B)** host-selective Ptr ToxA and Ptr ToxB.

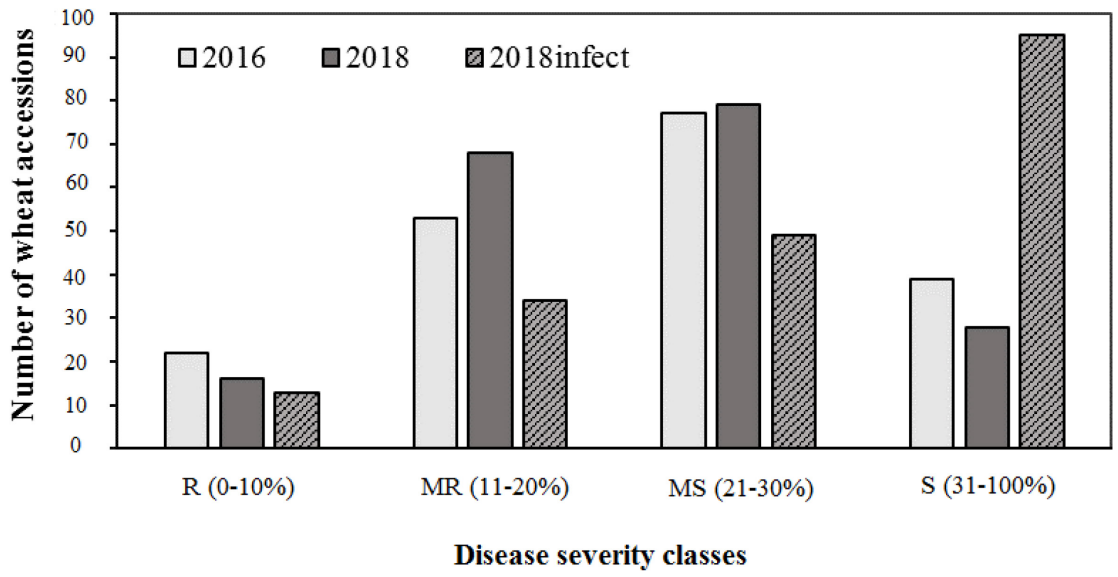


FIGURE 2 | Frequencies of 191 wheat genotypes reaction to tan spot in field.

Population Structure and Linkage Disequilibrium Decay

Principal component analysis (PCA) revealed three broad groups in the panel, and the first three principal components explained 11.8, 9.9, and 5.5% of the genetic variation, respectively (**Figure 3A**). Groups 1 and 2 containing Eurasian germplasm with winter and spring wheat lines, respectively, were separated from the group of lines developed by the CIMMYT and CIMMYT-ICARDA-IWWIP program (Group 3). The weighted neighbor-joining (NJ) dendrogram confirmed the three groups obtained in PCA (**Figure 3B**). Further ΔK vs. K plot, based on results obtained in Bayesian model-based STRUCTURE analysis, showed the highest likelihood at $K = 3$ (**Figure 3C**). At $K = 3$, the three subpopulations corresponded with both PCA and NJ-based groups (**Figure 3D**).

The average LD was highest in the D genome (0.42) and was higher compared to A (0.35) and B (0.34) genomes. Genome-wide LD decay patterns were investigated in the three subpopulations individually and in the whole panel (**Supplementary Figure 1**). The results revealed that LD decay was faster in the CIMMYT and IWWIP germplasm, i.e., Group 3 (18 Mb at cut off $r^2 = 0.1$) compared to Eurasian lines (30 and 25 Mb in groups 1 and 2, respectively). The LD decayed at 22 Mb in the whole panel (**Supplementary Figure 1D**).

TABLE 1 | Polymorphic information content (PIC) and linkage disequilibrium (LD) estimated for all chromosomes using 8,154 GBS markers.

Chromosome	Number of SNPs	Average PIC	Total number of marker pairs with $r^2 \geq 0.1$	Average LD at $r^2 \geq 0.1$
1A	464	0.34	4,521	0.31
1B	608	0.36	6,249	0.47
1D	167	0.31	2,488	0.48
2A	579	0.30	4,312	0.38
2B	870	0.36	8,016	0.34
2D	295	0.29	6,322	0.56
3A	469	0.37	3,911	0.32
3B	662	0.36	6,548	0.29
3D	129	0.31	828	0.45
4A	367	0.36	5,108	0.39
4B	202	0.35	1,747	0.38
4D	28	0.33	45	0.35
5A	442	0.33	5,498	0.40
5B	769	0.36	6,631	0.36
5D	79	0.31	621	0.36
6A	412	0.35	3,542	0.33
6B	476	0.38	5,567	0.30
6D	109	0.30	733	0.37
7A	565	0.34	5,924	0.34
7B	354	0.37	3,173	0.30
7D	108	0.28	495	0.41

Marker-Trait Associations

MTA for Resistance to Race 1 and Race 5 at Seedling Stage

Ten SNPs on chromosomes 1B (3), 3B (1), 4B (1), 5A (1), 5B (1), 6A (1), and 7B (2) were associated with resistance to Race 1 of *P. tritici-repentis* with 7.3–10.5% variation explained (**Table 2** and **Supplementary Figures 2A,B**). The phenotypic mean difference between the alleles for the significant SNPs ranged from 0.2 to 1.3. The SNP identified on chromosome 6A with clone ID 1004240 showed the strongest allele effect followed by SNPs on chromosomes 4B (ID 3958510) and 3B (ID 1147153). Favorable alleles at these loci were present in 5–9% frequencies.

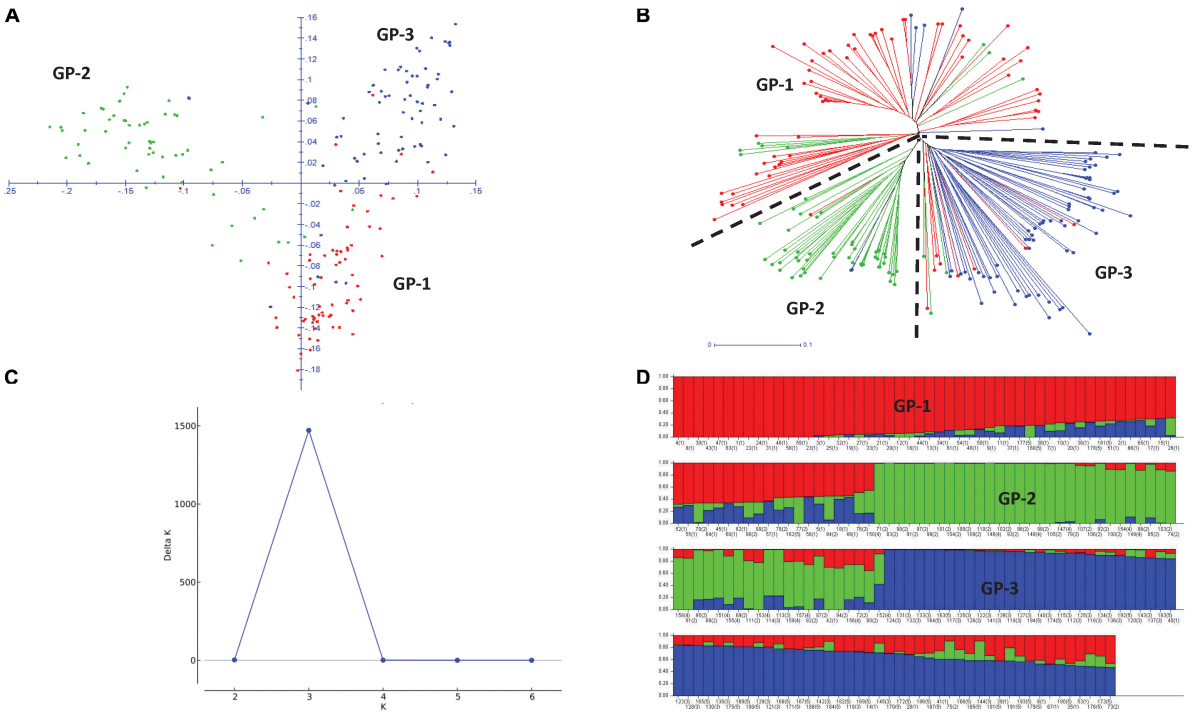


FIGURE 3 | Two-dimensional principal component analysis plot **(A)** and weighted neighboring-joining dendrogram **(B)** of panel showing three groups with both analyses. GP1, Winter type from Kazakhstan; GP2, Spring type from Kazakhstan and Russia; GP3, Spring and winter types from the CIMMYT and CIMMYT-ICARDA-IWWIP program. Delta *K* vs. *K* plot showing the best *K* at 3 **(C)** and bar plot showing subpopulation division at *K* = 3 **(D)**.

For Race 5, three genomic regions on chromosomes 5A, 5D, and 6A were identified to be associated with percentage variation from 6.8 to 8.0% (**Table 2** and **Supplementary Figures 2C,D**). Compared to associations identified for Race 1, markers associated with Race 5 showed lower phenotypic mean differences between alleles. Interestingly, the genomic region spanning 3 Mb on chromosome 6AL was identified to be associated with resistances to both races, Race 1 and Race 5. This indicates the importance of chromosome 6AL in imparting race-specific resistance to *P. tritici-repentis*. *In silico* analysis of the genomic region identified on chromosome 6A revealed two candidate gene hits, TraesCS6A02G37880 and TraesCS6A02G384600 with SNPs 1004240 and 1862737, respectively, both with oxidoreductase activity. Oxidoreductase family proteins are known as important components of pathogen-associated molecular pattern-triggered immunity *via* production of reactive oxygen species in response to pathogen attack (Heiser and Elstner, 2002). Of the two candidate genes, TraesCS6A02G384600 has been investigated for expression analysis (Expression Atlas[4]), and it is known to express in outer pericarp layer of developing grain.

Allelic effects of important associations identified for resistance to Races 1 and 5 are shown in **Figure 4**. Twenty-five tan spot resistant lines have been identified with different allelic combinations for seedling resistance to Race 1 and Race 5 (**Supplementary Table 5**).

[4]https://www.ebi.ac.uk/gxa/home

Marker-Trait Associations for Insensitivity to Toxins Ptr ToxA and Ptr ToxB

Four MTA were identified associated with insensitivity to Ptr ToxA on chromosomes 2B (1) and 5B (3) with percentage variation explained by them ranging from 7.0 to 13.7% (**Table 2**). The phenotypic mean difference between the alleles for the three significant SNPs on chromosome 5B ranged from 1.0 to 2.4, whereas for the SNP on chromosome 2B, the allelic difference in alleles was 0.6. For insensitivity to Ptr ToxB, important loci were identified on chromosomes 2A (1), 2B (2), 3A (1), 3B (1), and 4B (1) with percentage variation from 7.2 to 10.6%. The QTL identified on chromosome 2B related to SNPs 1095982 and 1065699 had the strongest allelic effect among all (**Table 2**).

Marker-Trait Associations for Area Under Disease Progress Curve Scores From Three Field Experiments

Analysis of AUDPC for two seasons (2016 and 2018) under natural epidemic conditions revealed eight associations on chromosomes 1B (1), 2B (4), 3B (1), 4A (1), and 6B (1) (**Table 3**). On chromosome 2B, four MTA were identified, which were delineated into two QTL, one between 18.5 and 28.0 cM associated with 2018 AUDPC scores and the second between 86.5 and 89.8 cM associated with 2016 AUDPC scores. The second QTL located 86.5–89.8 cM explained higher (8.7–11.9%) percentage variation compared to the other 2BS QTL (4.0–7.3%). In addition, the association identified on 1B also explained high percentage variation (11.6%). Under artificial

TABLE 2 | Markers associated with resistance to Race 1, Race 5, and insensitivities to Ptr ToxA and Ptr ToxB at seedling stage.

Trait	Marker name (clone ID)	Chr	Genetic position on consensus map (cM)	Physical position (Chinese_ Spring_v1.0)	*P*-value	R^2	Allele 1	Mean	Allele 2	Mean	Fav. allele frequency
Race 1 (Average)	1103424	1BS	86.7	62,284,931	3.29E–04	7.3	T	3.0	C	2.0	0.13
	1406319	1BS	90.8	–	3.18E–04	7.7	C	3.0	T	2.2	0.17
	1090475	1BS	92.2	57,766,497	6.18E–04	7.9	G	3.0	C	2.0	0.13
	1147153	3BS*	53.7	227,835,794	9.28E–04	8.6	C	3.1	G	2.2	0.05
	3958510	4BS	39.4	–	8.61E–04	9.2	T	3.0	C	1.9	0.09
	3946488	5BL	40.8	–	9.37E–04	8.5	T	3.6	C	2.8	0.80
Race 1 (BLUE)	1134493	5AL	153.01	698,528,378	4.61E–04	8.1	T	3.0	C	2.7	0.51
	1004240	6AL	89.8	602,914,149	5.60E–04	10.5	T	3.1	C	1.8	0.08
	3021234	7BL	49.4	–	1.17E–04	8.2	T	3.1	C	2.5	0.28
	1081730	7BL	49.4	538,125,772	4.53E–04	8.3	G	3.1	A	2.5	0.32
Race 5 (Average)	1008802	5AL	44.1	441,476,773	3.09E–04	6.8	G	2.0	C	1.7	0.23
	1093048	5DL*	151.1	629,968,953	8.65E–04	7.7	T	1.9	G	1.4	0.10
	1862737	6AL*	90.3	599,831,232	6.50E–04	8.0	G	2.4	C	1.7	0.79
Race 5 (BLUE)	1008802	5AL	44.1	441,476,773	7.60E–04	7.2	G	2.1	C	1.7	0.23
Insensitivity to Ptr ToxA (Average)	4003201	2BS	86.6	–	7.34E–05	7.0	T	2.8 (sensitive)	C	2.2 (insensitive)	0.60
	1138872	5BL	35.7	429,865,271	9.06E–06	10.7	G	2.4 (sensitive)	T	1.4 (insensitive)	0.11
	3955588	5BL	70.1	546,831,358	6.98E–06	13.5	G	3.7 (sensitive)	T	1.3 (insensitive)	0.59
	1128605	5BL	73.1	547,522,421	3.20E–05	13.7	G	3.3 (sensitive)	T	1.6 (insensitive)	0.46
Insensitivity to Ptr ToxB (average)	1038630	2AS	18.09	29,263,722	1.42E–05	9.1	C	2.3 (sensitive)	G	1.2 (insensitive)	0.86
	1095982	2BS	75.08	–	8.52E–05	10.6	A	2.6 (sensitive)	G	1.1 (insensitive)	0.87
	1065699	2BS	71.8	184,907,691	3.57E–04	9.4	C	2.4 (sensitive)	T	1.1 (insensitive)	0.77
	1093535	3AS*	41.1	64,777,406	7.95E–04	8.6	T	2.0 (sensitive)	C	1.2 (insensitive)	0.81
	2263392	3BL	156.6	824,374,808	7.51E–04	8.0	C	2.0 (sensitive)	T	1.2 (insensitive)	0.83
	1217569	4AL	29.8	–	4.14E–05	7.2	C	1.8 (sensitive)	T	1.2 (insensitive)	0.90

Asterisk represent novel quantitative trait loci (QTL) identified in the study, – BLAST identified either a different chromosome than reported based on consensus map provided by Genetic Analysis and Service for Agriculture (SAGA) or reported single nucleotide polymorphism (SNP) was not identified in the Ensemble Plants database.

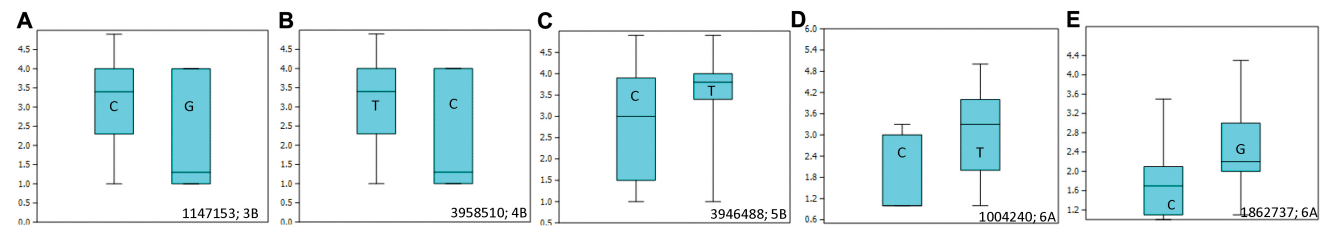


FIGURE 4 | Box plots showing allelic effects of important associations identified for resistance to Race 1 with average **(A–C)** and best linear unbiased estimator (BLUE) scores **(D)** and for resistance to Race 5 with average scores **(E)**. The number on the right side below shows the name of the associated marker and chromosome number.

inoculation conditions in 2018, two associations were identified on chromosomes 3A (1) and 5B (1), both explaining moderate percentage variation. QQ plots for all traits investigated by GWAS are shown in **Supplementary Figure 3**.

Epistatic Interactions

Two- and three-locus interactions were estimated among associated loci, and among associated and genome wide loci (**Figure 5**, **Supplementary Table 6**, and **Supplementary Figures 4–6**). For Race 1 and Race 5, associated loci showed epistatic interactions among them with R^2 explained from 10.7 to 18.7% for Race 1 and 16.0–21.1% for Race 5, respectively (**Figure 5**). The R^2 explained by marker combinations was higher than individual markers for both Race 1 and Race 5. Further, the three-marker combination resulted in the highest additive interactions explained by R^2 of 18.7 and 21.1% for Race 1 and Race 5, respectively. For other traits, no significant interactions were observed among associated loci. However, significant interactions among unassociated genome-wide markers were observed for

TABLE 3 | Markers associated with area under disease progress curve (AUDPC) scores under natural epidemic conditions in two field seasons (2016 and 2018) and by artificial inoculation in 2018 (AUDPC 2018_infect).

Trait	Marker name (clone ID)	Chr	Genetic position on consensus map (cM)	Physical position (Chinese_Spring_ v1.0)	*P*-Value	R^2	Allele 1	Mean	Allele 2	Mean	Fav. allele frequency
AUDPC 2016	1022665	2BS	86.5	760,015,871	6.94E–05	11.9	T	23.1	C	16.6	0.75
	1367534	2BS	89.8	758,922,449	8.47E–04	8.7	G	20.7	A	16.6	0.72
	2256531	4AL	103.3	–	8.74E–04	7.2	C	19.4	A	16.4	0.61
AUDPC 2018	4993225	1BS	87.3	57,643,782	2.51E–05	11.6	G	19.3	C	7.7	0.12
	3533672	2BS	18.5	–	1.27E–04	7.3	C	18.3	T	11.3	0.11
	1061674	2BS	28.0	33,729,208	2.47E–04	4.0	G	17.8	C	14.0	0.05
	2295410	3BL	145.1	–	5.64E–04	6.0	G	18.3	A	14.2	0.58
	3020636	6BS	6.0	–	6.65E–05	9.7	C	18.9	A	11.2	0.23
AUDPC 2018_infect	4988948	3AL*	130.9	731,226,020	9.19E–04	7.2	T	59.0	C	44.0	0.27
	1234099	5BL	60.1	531,195,246	3.01E–04	9.9	G	55.2	T	37.2	0.13

Asterisk represent novel QTL identified for adult plant resistance in the study, – BLAST identified either a different chromosome than reported based on consensus map provided by SAGA or reported SNP was not identified in Ensemble Plants database.

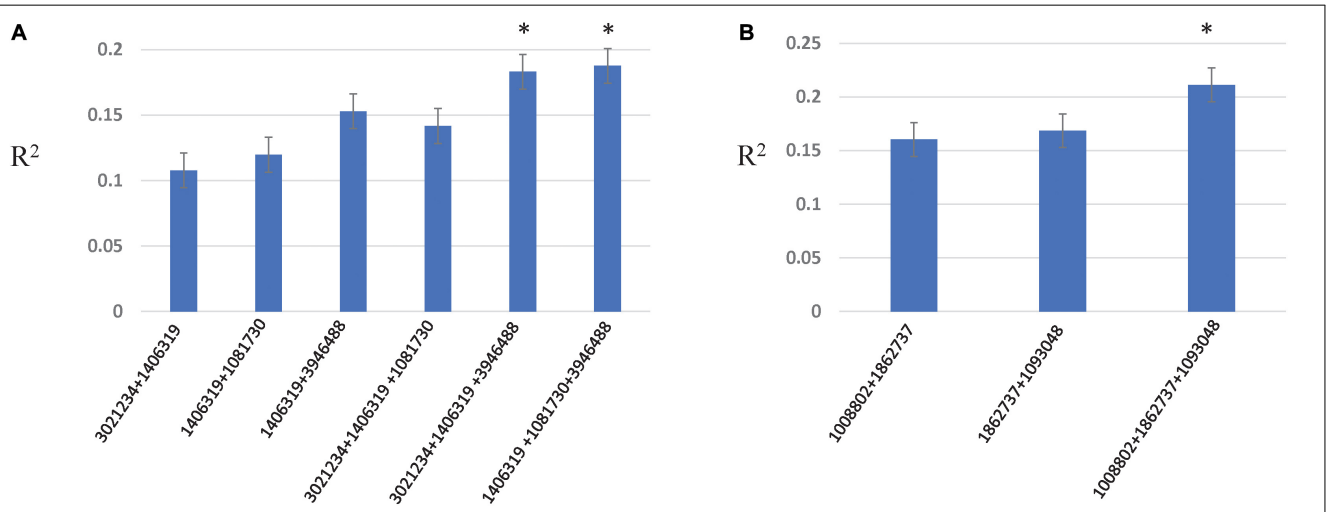


FIGURE 5 | Additive epistatic interactions among associated loci for resistance to Race 1 **(A)** and Race 5 **(B)**. The *X*-axis represents two- and three-marker combinations interacting epistatically. The *Y*-axis represents the percentage variation as R^2 explained by marker combinations. The asterisk represents the marker combinations that resulted in the highest R^2.

all traits except insensitivity to Ptr ToxB (**Supplementary Figures 4–6**).

DISCUSSION

Tan spot, caused by *P. tritici-repentis*, is a serious foliar disease affecting wheat production in Kazakhstan, especially in the northern region of the country where farmers lose anywhere from 10% to as much as 50% of their crop due to this disease (Koyshibaev, 2018). It has been predicted that outbreaks of this disease in Kazakhstan are likely to increase in severity and frequency due to an increase in mean annual temperatures and altered precipitation patterns (Salnikov et al., 2015). Characterization of the current germplasm for resistance to prominent races of the pathogen in Kazakhstan has therefore become more imperative than ever before (Kokhmetova et al., 2018, 2019). Here, we evaluated registered cultivars and lines from Kazakhstan and Russia representing promising spring and winter wheat germplasm along with lines released by the CIMMYT and CIMMYT-ICARDA-IWWIP program.

Most wheat entries evaluated in the study showed a susceptible response to Race 1 (72.6%), whereas for Race 5, 67.5% of lines were resistant. It is a common observation in most germplasm collections (Liu et al., 2015; Halder et al., 2019), and our results reinforce that finding resistance against Ptr Race 1 is very challenging compared to other races. Race 1 is the most prevalent worldwide (Ali and Francl, 2003; Abdullah et al., 2017) and is reported to contain the virulence of Races 2 and 3, making it more aggressive than other races (Lamari et al., 2003). Nevertheless, 25 lines (13%) resistant to both races (Race 1 and Race 5) have been identified in the present study. Regarding

toxin sensitivities of the genotypes, a low but positive correlation was observed between Ptr ToxA sensitivity of the genotypes and disease susceptibility in the field, which is in contrast with the observations made by Friesen et al. (2003); Faris and Friesen (2005), Noriel et al. (2011), and Halder et al. (2019). We did not obtain any significant correlations between days to heading and plant height with Ptr resistance in field, suggesting that none of these phenological traits significantly affected tan spot resistance in the germplasm investigated. Li et al. (2011) and Pandey et al. (2018) obtained similar results, i.e., no association of tan spot resistance with phenological traits. Many previous studies have detected a negative correlation between heading date or plant height with disease resistance (Srinivasachary et al., 2009; Mao et al., 2010; Kollers et al., 2013a,b), including tan spot resistance (Kollers et al., 2014). The broad sense heritability ($h2$) estimates for tan spot across years and different infectious background were high (from 0.83 to 0.91) indicating that resistance to tan spot can be improved by selection (**Supplementary Table 3**). Similar heritability estimates for tan spot disease reaction have been reported by Singh et al. (2016, 2019). Heritability ranged from 0.70 to 0.96 for DH and was 0.98 for PH in different years. Jamil et al. (2019) reported heritability estimates for these traits in a similar range.

The mean PIC of the panel based on 8,154 SNPs was 0.33, which is in the same range as observed recently in wheat cultivars from China (Gao et al., 2016) and higher than those obtained in United States wheat cultivars (Chao et al., 2009, 2010) or other germplasm sets (Somers et al., 2007; Alemu et al., 2020). The higher diversity can be explained by the diverse origin of the lines included in the panel such as those from CIMMYT. CIMMYT germplasm has shown substantial genetic diversity in previous studies (Dreisigacker et al., 2005; Reif et al., 2005; Warburton et al., 2006; Sehgal et al., 2015). Population structure through PCA, NJ, and STRUCTURE analyses of the panel revealed a clear distinction of spring and winter wheat types from Kazakhstan. Chao et al. (2010) also obtained distinct groups of spring and winter wheat from the US. A third group of spring and winter wheat was exclusive to lines released by the CIMMYT and IWWIP program, thus clearly separating CIMMYT germplasm from Eurasian lines. The panel, therefore, shows a moderate population structure along with high diversity rendering it fit for trait dissection using the GWAS approach.

The pattern of LD across the three genomes revealed the highest LD in the D genome followed by the A and B genomes. This pattern is common in wheat and reflects the recombination history of the three genomes and population bottleneck accompanying the origin of hexaploid wheat (Akhunov et al., 2010; Chao et al., 2010). Genome-wide LD decay was observed at 22 Mb in the whole panel, which is in the same range as reported recently in a CIMMYT spring wheat collection (Jamil et al., 2019) and in a composite collection made of CIMMYT and South Asian genotypes (Phuke et al., 2020). It was not possible to compare the LD decay of the panel with other germplasms because LD decay has been reported as a measure of genetic distance (in cM) in previous studies (Somers et al., 2007; Benson et al., 2012; Chen et al., 2012; Würschum et al., 2013; Edae et al., 2014; Zegeye et al., 2014; Sehgal et al., 2017; Erginbas-Orakci et al., 2018). A comparison of LD decay among three subpopulations of the present study revealed a faster decay in the subpopulation composed by CIMMYT and IWWIP lines compared to the two subpopulations of Eurasian lines, which is attributed to higher diversity of CIMMYT and IWWIP germplasm vis-à-vis Eurasian germplasm. Breeders at CIMMYT have successfully broadened the genetic diversity of the elite germplasm through incorporation of primary synthetics into the breeding programs and consistent introductions of additional materials from all over the world (Sehgal et al., 2015). Similarly, the pedigree of IWWIP lines incorporates diverse CIMMYT parents and a wide range of genetically un-related winter wheat from Turkey and Iran, Russia, Ukraine, Romania, Bulgaria, Hungary, and United States winter wheat.

Genome-wide association study identified a total of 34 MTA for tan spot resistance. Of all the MTA identified for seedling resistance to Races 1 and 5, the genomic region on chromosome 6AL showed the biggest phenotypic effect. In fact, two MTA were identified within a 3-Mb genomic region on 6AL for resistance to both Race 1 (Clone ID 1004240) and 5 (Clone ID 1862737) (**Table 2**). The favorable allele of the SNP with Clone ID 1004240 was predominantly present in CIMMYT lines and three Kazakhstani winter wheat types (Naz/GF55-5, 428 g/MK-122A-2 and Almaly/Obri), while the favorable allele of the SNP with Clone ID 1862737 was predominant in Kazakhstani winter and spring wheat lines. Singh et al. (2016) reported two QTL on 6AL ($6AL_1$ and $6AL_2$) for tan spot resistance in CIMMYT germplasm for resistance to Race 1, and the SNP with clone ID 1004240 represents $6AL_2$ QTL. The allelic effect obtained by this SNP is, however, three times larger than reported by Singh et al. (2016). The second marker on chromosome 6AL (Clone ID 1862737) represents a novel QTL for resistance to Race 5 because the only QTL reported on 6A for resistance to Race 5 (Gurung et al., 2014) is on 6AS. For resistance to Race 1, MTA on chromosomes 4BS and 3BS also showed strong effects. Halder et al. (2019) recently reported a novel QTL on 4BS for resistance to Race 1 in the United Kingdom Watkins core collection, and the associated marker identified here on chromosome 4BS corresponds to this QTL. No gene or QTL has so far been reported on chromosome 3BS for resistance to Race 1. Hence, MTA obtained on chromosome 3BS is likely to be novel.

The collection investigated was comprised of 12% spring wheat lines distributed by CIMMYT and evaluated to be resistant to both *Septoria tritici blotch* and tan spot. In CIMMYT's germplasm, a large QTL on chromosome 1BS (∼19 cM based on consensus map and ∼130 Mb based on physical position) has contributed significantly for resistance to Race 1, which was suggested due to the 1B.1R translocation harboring several resistance genes including *Lr26*, *Yr9*, *Sr31*, *Pm8*, and others (Singh et al., 2016; Juliana et al., 2018). The three MTA identified here on chromosome 1BS covered ∼5 Mb on reference genome and were in the middle of 1BS QTL. Similarly, MTA identified on other chromosomes for resistance to Race 1 and Race 5 coincided with known QTL except the one on chromosome 5DL for resistance to Race 5 (Faris et al., 2013; Singh et al., 2016;

Hu et al., 2019). The only reported QTL on chromosome 5DL is for resistance to Race 1 (Faris et al., 2012). Therefore, the associated marker identified on chromosome 5DL for resistance to Race 5 represents a new QTL.

Isolates of Race 1 and Race 5 are known to produce toxins Ptr ToxA and Ptr ToxB, respectively. Qualitative resistance in the host has been reported to be manifested by two toxin insensitivity genes on chromosomes 5BL (*tsn1*) and 2BS (*tsc2*), conferring resistance against Ptr ToxA and Ptr ToxB, respectively (Singh et al., 2010; Faris et al., 2013). We obtained three MTA on chromosome 5BL associated with insensitivity to Ptr Tox A, which were within the same genomic region as the *tsn1* gene (Faris et al., 2010). Of the three, two SNPs (Clone ID 3955588 and 1128605) had higher allelic effects on toxin insensitivity compared to the third SNP (Clone ID 1138872). An additional locus for insensitivity to Ptr ToxA was identified on chromosome 2BS, which coincided with previous QTL reported for resistance to Race 1 (Gurung et al., 2014). For insensitivity to Ptr ToxB, important associations were obtained on chromosomes 2AS, 2BS, 3AS, 3BL, and 4AL. Of all associations, two SNPs (clone IDs 1095982 and 1065699) on 2BS showed the strongest allelic effect and were in the same genomic region as the cataloged *tsc2* gene. The SNP on chromosome 2AS was the second most important genomic region with a strong allelic effect on insensitivity to Ptr ToxB. Previous studies till date have reported only minor effect QTLs on 2AS for insensitivity to Ptr Tox B and/or resistance to Race 5 (Friesen and Faris, 2004; Chu et al., 2008). On chromosome 3AS, *tsr4* has been mapped, and the associated marker obtained here on chromosome 3AS is around 70 cM from *tsr4* locus and hence unlikely to be in the location of *tsr4* gene. The known genes on chromosome 3BL are *tsn2* and *tsn5* that confer resistance to the necrosis induced by Race 3 and Race 5, respectively (Singh et al., 2008). The locus obtained on chromosome 3BL is of moderate effect and corresponds to minor QTL reported by Faris and Friesen (2005) and Chu et al. (2010) on the same chromosome. The locus identified on chromosome 4AL also represents another minor effect QTL as also previously reported for resistance to Race 5 (Friesen and Faris, 2004).

Regarding MTA identified for AUDPC scores, most associations except the one on chromosome 3AL were identified within 2–3 Mb genomic region of the markers associated with seedling resistance to Races 1 and 5 or insensitivities to toxins Ptr ToxA and Ptr ToxB in the present study, thus confirming to be major genes (*tsn 1*, *tsc2*). Three associations on chromosomes 2BS, 4AL, and 6BS overlap with minor QTL reported for tan spot resistance in previous studies (Gurung et al., 2011; Halder et al., 2019). The QTL found on chromosome 3AL is likely new for adult plant resistance, which was identified only at adult plant stage in the present study.

The role of epistasis in the genetic architecture of disease resistance has been investigated for stem rust resistance and stripe rust resistance in wheat using both biparental and GWAS designs (Kolmer et al., 2011; Yu et al., 2011). However, investigations on the contribution of epistatic interactions to the genetic architecture of tan spot resistance have been rarely investigated (Singh et al., 2019). Singh et al. (2019) revealed epistatic interactions between QTL on chromosomes 1B and 5B and between QTL on chromosomes 1A and 6A for seedling resistance to Race 1 in a biparental population, however, with small effects. We confirm significant epistatic interactions between associated loci on chromosomes 1B and 5B in the present study with even larger percentage variation explained (up to 18.7%). In addition, we obtained significantly higher interactions between associated loci on chromosomes 1B and 7B with a percent variation explained as large as 14.1% for resistance to Race 1. The associated locus on 7B (SNP 1081730) was also involved in genome-wide interactions with other loci on chromosomes 7A, 2D, and 3B, and contributed to additional 4% variation. For seedling resistance to Race 5, associated loci on chromosomes 5A, 5D, and 6A contributed to additive epistatic interactions of 21.1%. These results indicated that both additive and epistatic effects are important for tan spot resistance to both races, Race 1 and Race 5. The results also suggest that three marker combinations identified for resistance to Race 1 and Race 5 can be used efficiently in marker-assisted selection.

CONCLUSION

Our results suggest the existence of valuable resistant alleles on chromosomes 3AS, 3AL, 3BS, and 6AL for tan spot resistance in the investigated germplasm in addition to known genes *tsn1* and *tsc2*. The study therefore confirms that resistance to tan spot in the collection is attributed to both the known toxin insensitivity genes with major effects as well as broad-spectrum and race-non-specific genes, which have minor effects. Twenty-five tan spot-resistant lines have been identified with different allelic combinations for resistance to Race 1 and Race 5 for future variety release. The candidate genes identified on 6A are important targets for future validation studies for tan spot resistance.

DATA AVAILABILITY STATEMENT

The genotyping data is available at the CIMMYT Research Data and Software Repository Network: https://hdl.handle.net/11529/10548482.

AUTHOR CONTRIBUTIONS

AK conceived the manuscript and designed the research. DS and AK analyzed the data and wrote the manuscript. SA, IL, MA, and MK generated the phenotypic data. DS and SD generated the allele called GBS marker data. All authors reviewed the manuscript.

FUNDING

This work was supported by the Kazakhstan Ministry of Education and Sciences, project #AP05132540.

ACKNOWLEDGMENTS

We thank Z. Sapakhova and the staff of Laboratory Genetics and Breeding, Institute of Plant Biology and Biotechnology for their technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.581214/full#supplementary-material

Supplementary Figure 1 | Linkage disequilibrium decay plot for the three subpopulations; CIMMYT and IWWIP lines **(A)**, Spring wheat from Kazakhstan and Russia **(B)**, Winter wheat from Kazakhstan **(C)** and across all panel **(D)**. *X*-axis represents physical distance in Mb; *Y*-axis represents correlation coefficient.

Supplementary Figure 2 | Manhattan plots showing significant markers associated with BLUE (left) and average (right) scores for resistance to Race 1 **(A,B)** and Race 5 **(C,D)**, respectively.

Supplementary Figure 3 | QQ plots of all traits used in GWAS analysis with mixed linear model by using first three principal components as a fixed variate and kinship as a random variate.

Supplementary Figure 4 | Genome wide epistatic interactions for resistance to Race 1 **(A)**, Race 5 **(B)**, and sensitivity to Ptr Tox A **(C)**. For sensitivity to Ptr ToxB, no significant interactions were observed.

Supplementary Figure 5 | Genome wide epistatic interactions for AUDPC scores in field seasons in 2016 **(A)** and 2018 **(B)**.

Supplementary Figure 6 | Genome wide epistatic interactions for AUDPC 2018_infect; 1 associated SNP on 3A showed genome wide and explained 5.5%.

Supplementary Table 1 | Cultivars and breeding lines used in the present study.

Supplementary Table 2 | Analysis of variance for tan spot (TS) disease scores for resistance to Race 1 and Race 5 and distribution of tan spot score in checks and the panel.

Supplementary Table 3 | Analysis of variance (ANOVA) for tan spot severity and phenological traits under natural conditions and under artificial inoculation conditions and summary statistics.

Supplementary Table 4 | Correlations of days to heading (DH) and plant height (PH) with AUDPC scores in field seasons.

Supplementary Table 5 | Kazakhstani and Russian entries with resistances to Race 1 and Race 5 with best allele combinations.

Supplementary Table 6 | Epistatic interactions among associated loci for tan spot resistance.

REFERENCES

Abdullah, S., Sehgal, S. K., Jin, Y., Turnipseed, B., and Ali, S. (2017). Insights into tan spot and stem rust resistance and susceptibility by studying the pre-green revolution global collection of wheat. *Plant Pathol J.* 33, 125–132. doi: 10.5423/ppj.oa.07.2016.0157

Abeysekara, N. S., Friesen, T. L., Liu, Z., McClean, P. E., and Faris, J. D. (2010). Marker development and saturation mapping of the tan spot Ptr ToxB sensitivity locus Tsc2 in hexaploid wheat. *Plant Gen.* 3, 179–189.

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., et al. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* 11:702. doi: 10.1186/1471-2164-11-702

Alemu, A., Feyissa, T., Letta, T., and Abeyo, B. (2020). Genetic diversity and population structure analysis based on the high-density SNP markers in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *BMC Genet.* 21:18. doi: 10.1186/s12863-020-0825-x

Ali, S., and Francl, L. J. (2001). Recovery of *Pyrenophora tritici*-repentis from barley and reaction of 12 cultivars to five Races and two host-specific toxins. *Plant Dis.* 85, 580–584. doi: 10.1094/pdis.2001.85.6.580

Ali, S., and Francl, L. J. (2002). A new race of *Pyrenophora tritici*-repentis from Brazil. *Plant Dis.* 86:1050. doi: 10.1094/pdis.2002.86.9.1050c

Ali, S., and Francl, L. J. (2003). Population Race structure of *Pyrenophora tritici*-repentis prevalent on wheat and non-cereal grasses in the Great Plains. *Plant Dis.* 87, 418–422. doi: 10.1094/pdis.2003.87.4.418

Anderson, J. A., Effertz, R. J., Faris, J. D., Francl, L. J., Meinhardt, S. W., and Gill, B. S. (1999). Genetic analysis of sensitivity to a *Pyrenophora tritici*-repentis necrosis inducing toxin in durum and common wheat. *Phytopathology* 89, 293–297. doi: 10.1094/phyto.1999.89.4.293

Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., Jin, Y., et al. (2016). Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS One* 11:e0155760. doi: 10.1371/journal.pone.0155760

Benson, J., Brown-Guedira, G., Murphy, J. P., and Sneller, C. (2012). Population structure, linkage disequilibrium, and genetic diversity in soft winter wheat enriched for fusarium head blight resistance. *Plant Genome* 5, 71–80. doi: 10.3835/plantgenome2011.11.0027

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Breseghello, F., and Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177. doi: 10.1534/genetics.105.044586

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M. C., Baenziger, S. P., Matnyazov, R., et al. (2010). Population - and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L). *BMC Genomics* 11:727. doi: 10.1186/1471-2164-11-727

Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M. C., et al. (2009). Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L) cultivars. *Mol. Breed.* 23, 23–33. doi: 10.1007/s11032-008-9210-6

Chen, X., Min, D., Yasir, T. A., and Hu, Y. G. (2012). Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PLoS One* 7:e44510. doi: 10.1371/journal.pone.0044510

Chu, C. G., Chao, S., Friesen, T. L., Faris, J. D., Zhong, S., and Xu, S. S. (2010). Identification of novel tan spot resistance QTLs using and SSR-based linkage map of tetraploid wheat. *Mol. Breed.* 25, 327–338. doi: 10.1007/s11032-009-9335-2

Chu, C. G., Friesen, T. L., Faris, J. D., and Xu, S. S. (2008). Evaluation of seedling resistance to tan spot and Stagnosphora nodorum blotch in tetraploid wheat. *Crop Sci.* 48, 1107–1118. doi: 10.2135/cropsci2007.09.0516

Das, M. K., Rajaram, S., Mundt, C. C., and Kronstad, W. E. (1992). Inheritance of slow rusting resistance to leaf rust in wheat. *Crop Sci.* 32, 1452–1456. doi: 10.2135/cropsci1992.0011183x003200060028x

Dospekhov, B. A. (1985). *Methods of Field Experience (With the Basics of Statistical Processing of Research Results)*, 5th Edn. Kovalivka: Kolos.

Dreisigacker, S., Sehgal, D., Reyes Jaimez, A. E., Luna Garrido, B., Muñoz Zavala, S., Núñez Ríos, C., et al. (2016). *Laboratory Protocols and Applications to Wheat Breeding, Version 1, Mexico, DF.*

Dreisigacker, S., Shewayrga, H., Crossa, J., Arief, V. N., DeLacy, I. H., Singh, R. P., et al. (2012). Genetic structures of the CIMMYT international yield trial targeted to irrigated environments. *Mol. Breed.* 29, 529–541. doi: 10.1007/s11032-011-9569-7

Dreisigacker, S., Zhang, P., Warburton, M. L., Skovmand, B., Hoisington, D., and Melchinger, A. E. (2005). Genetic diversity among and within CIMMYT hexaploid wheat landrace accessions and implications for plant genetic resource management. *Crop Sci.* 45, 653–661. doi: 10.2135/cropsci2005.0653

Duveiller, E., Kandel, Y. R., Sharma, R. C., and Shrestha, S. M. (2005). Epidemiology of foliar blights (spot blotch and tan spot) of wheat in the plains bordering the Himalayas. *Phytopathology* 95, 248–256. doi: 10.1094/phyto-95-0248

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* 127, 791–807. doi: 10.1007/s00122-013-2257-8

Effertz, R. J., Meinhardt, S. W., Anderson, J. A., Jordahl, J. G., and Francl, L. J. (2002). Identification of a chlorosis-inducing toxin from *Pyrenophora tritici-*repentis and the chromosomal location of an insensitivity locus in wheat. *Phytopathology* 92, 527–533. doi: 10.1094/phyto.2002.92.5.527

Elbasyoni, I. S., El-Orabey, W. M., Baenziger, P. S., and Eskridge, K. M. (2017). Association mapping for leaf and stem rust resistance using worldwide spring wheat collection. *AJÎB* 4, 1–25. doi: 10.9734/ajob/2017/38120

Erginbas-Orakci, G., Sehgal, D., Sohail, Q., Ogbonnaya, F., Dreisigacker, S., Pariyar, S. R., et al. (2018). Identification of novel quantitative trait loci linked to crown rot resistance in spring wheat. *Int. J. Mol. Sci.* 19:2666. doi: 10.3390/ijms19092666

FAO (2020). *Crop Prospects and Food Situation - Quarterly Global Report No. 2*. Rome: FAO.

Faris, J. D., Abeysekara, N. S., McClean, P. E., Xu, S. S., and Friesen, L. T. (2012). Tan spot susceptibility governed by the Tsn1 locus and Race-nonspecific resistance quantitative trait loci in a population derived from the wheat lines Salamouni and Katepwa. *Mol. Breed.* 30, 1669–1678. doi: 10.1007/s11032-012-9750-7

Faris, J. D., Anderson, J. A., Francl, L. J., and Jordahl, J. G. (1996). Chromosomal location of a gene conditioning insensitivity in wheat to a necrosis-inducing culture filtrate from *Pyrenophora tritici-*repentis. *Phytopathology* 86, 459–463. doi: 10.1094/phyto-86-459

Faris, J. D., and Friesen, T. L. (2005). Identification of quantitative trait loci for Race-nonspecific resistance to tan spot in wheat. *Theor. Appl. Genet.* 111, 386–392. doi: 10.1007/s00122-005-2033-5

Faris, J. D., Liu, Z., and Xu, S. S. (2013). Genetics of tan spot resistance in wheat. *Theor. Appl. Genet.* 126, 2197–2217. doi: 10.1007/s00122-013-2157-y

Faris, J. D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., et al. (2010). A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proc. Nat. Acad. Sci. U.S.A.* 107, 13544–13549. doi: 10.1073/pnas.1004090107

Fernandez, M. R., Clarke, J. M., and DePauw, R. M. (1994). Response of durum wheat kernels and leaves at different growth stages to *Pyrenophora tritici-*repentis. *Plant Dis.* 78, 597–600. doi: 10.1094/pd-78-0597

Friesen, T. L., Ali, S., Kianian, S., Francl, L. J., and Rasmussen, J. B. (2003). Role of host sensitivity to Ptr ToxA in development of tan spot of wheat. *Phytopathology* 93, 397–401. doi: 10.1094/phyto.2003.93.4.397

Friesen, T. L., and Faris, J. D. (2004). Molecular mapping of resistance to *Pyrenophora tritici-*repentis Race 5 and sensitivity to Ptr ToxB in wheat. *Theor. Appl. Genet.* 109, 464–471. doi: 10.1007/s00122-004-1678-9

Gamba, F. M., and Lamari, L. (1998). Mendelian inheritance of resistance to tan spot (*Pyrenophora tritici-*repentis) in selected genotypes of durum wheat (*Triticum turgidum*). *Can. J. Plant Pathol.* 20, 408–414. doi: 10.1080/07060669809500412

Gamba, F. M., Strelkov, S. E., and Lamari, L. (2012). Virulence of *Pyrenophora tritici-*repentis in the southern cone region of South America. *Can J. Plant Pathol.* 34, 545–550. doi: 10.1080/07060661.2012.695750

Gao, L., Turner, M. K., Chao, S., Kolmer, J., and Anderson, J. (2016). Genome wide association study of seedling and adult plant leaf rust resistance in elite spring wheat breeding lines. *PLoS One* 11:e0148671. doi: 10.1371/journal.pone.0148671

Gerard, G. S., Borner, A., Lohwasser, U., and Simon, M. R. (2017). Genome-wide association mapping of genetic factors controlling *Septoria tritici* blotch resistance and their associations with plant height and heading date in wheat. *Euphytica* 213:27.

Gurung, S., Mamidi, S., Bonman, J. M., Jackson, E. W., del Rio, L. E., Acevedo, M., et al. (2011). Identification of novel genomic regions associated with resistance to *Pyrenophora tritici-*repentis Races 1 and 5 in spring wheat landraces using association analysis. *Theor. Appl. Genet.* 123, 1029–1041. doi: 10.1007/s00122-011-1645-1

Gurung, S., Mamidi, S., Bonman, J. M., Xiong, M., Brown-Guedira, G., and Adhikari, T. B. (2014). Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS One* 9:e108179. doi: 10.1371/journal.pone.0108179

Halder, J., Zhang, J., Ali, S., Jagdeep, S. S., Harsimardeep, S. G., Shyamal, K. T., et al. (2019). Mining and genomic characterization of resistance to tan spot, *Stagonospora nodorum* blotch (SNB), and Fusarium head blight in Watkins core collection of wheat landRaces. *BMC Plant Biol.* 19:480. doi: 10.1186/s12870-019-2093-3

Heiser, I., and Elstner, E. F. (2002). Biochemical mechanisms of plant defense a central role for reactive oxygen species. *Plant Protect. Sci.* 38, 76–S86.

Hu, W., He, X., Dreisigacker, S., Sansaloni, C. P., Juliana, P., and Singh, P. K. (2019). A wheat chromosome 5AL region confers seedling resistance to both tan spot and Septoria nodorum blotch in two mapping populations. *Crop J.* 7, 809–818.

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A. A., Ibrahim, A. M. H., et al. (2019). Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. *BMC Plant Biol.* 19:149.

Jordahl, J. G., and Francl, L. J. (1992). "Increase and storage of cultures of *Pyrenophora tritici-*repentis," in *Advances in Tan Spot Research*, eds L. J. Francl, J. M. Krupinsky, and M. P. McMullen (Fargo, ND: North Dakota Agric. Exp. Station).

Juliana, P., Singh, R. P., Singh, P. K., Poland, J. A., Bergstrom, G. C., Huerta-Espino, J., et al. (2018). Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. *Theor. Appl. Genet.* 131, 1405–1422. doi: 10.1007/s00122-018-3086-6

Khasanov, B. A. (1988). Yellow spotting of cereals caused by Pyrenopliora tritici-repentis. *II Mycol. Phytopathol.* 1, 78–84.

Kokhmetova, A., Atishova, M., Kumarbayeva, M., and Leonova, I. N. (2019). Phytopathological screening and molecular marker analysis of wheat germplasm from Kazakhstan and CIMMYT for resistance to tan spot. *Vavilov J. Genet. Breed.* 23, 879–886. doi: 10.18699/vj19.562

Kokhmetova, A., Kremneva, O., Volkova, G., Atishova, M., and Sapakhova, Z. (2017). Evaluation of wheat cultivars growing in Kazakhstan and Russia for resistance to tan spot. *J. Plant Pathol.* 99, 161–167.

Kokhmetova, A. M., Ali, S., Sapakhova, Z., and Atishova, M. N. (2018). Identification of genotypes-carriers of resistance to tan spot Ptr ToxA and Ptr ToxB of *Pyrenophora tritici-*repentis in common wheat collection. *Vavilov J. Genet. Breed.* 22, 978–986. doi: 10.18699/vj18.440

Kokhmetova, A. M., Kremneva, O., Keyshilov, Z. H. S., and Sultanova, N. Z. H. (2016). Race structure and virulence of isolates *Pyrenophora tritici-*repentis in the Republic of Kazakhstan and north Caucasus region of Russia. *Eur. J. Appl. Biotech.* 3, 57–66.

Kollers, S., Rodemann, B., Ling, J., Korzun, V., Ebmeyer, E., Argillier, O., et al. (2013a). Genetic architecture of resistance to *Septoria tritici* blotch (*Mycosphaerella graminicola*) in European winter wheat. *Mol. Breed.* 32, 411–423. doi: 10.1007/s11032-013-9880-6

Kollers, S., Rodemann, B., Ling, J., Korzun, V., Ebmeyer, E., Argillier, O., et al. (2013b). Whole genome association mapping of fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). *PLoS One* 8:e57500. doi: 10.1371/journal.pone.0057500

Kollers, S., Rodemann, B., Ling, J., Korzun, V., Ebmeyer, E., Argillier, O., et al. (2014). Genome-wide association mapping of tan spot resistance (*Pyrenophora tritici-*repentis) in European winter wheat. *Mol. Breed.* 34, 363–371. doi: 10.1007/s11032-014-0039-x

Kolmer, J. A., Garvin, D. F., and Jin, Y. (2011). Expression of a Thatcher wheat adult plant stem rust resistance QTL on chromosome arm 2BL is enhanced by Lr34. *Crop Sci.* 51, 526–533. doi: 10.2135/cropsci2010.06.0381

Koyshibaev, M. K. (2018). *Diseases of Wheat*. Ankara: FAO.

Kremneva, O. Y., and Volkova, G. V. (2007). The structure of *Pyrenophora tritici-repentis* population in north Caucasus by morphological and cultural characteristics and virulence. *Mikol. Fitopatol.* 41, 356–361.

Lamari, L., and Bernier, C. C. (1989). Evaluation of wheat lines and cultivars to tan spot (*Pyrenophora tritici-*repentis) based on lesion type. *Can. J. Plant Pathol.* 11, 49–56. doi: 10.1080/07060668909501146

Lamari, L., Strelkov, S. E., Yahyaoui, A., Amedov, M., Saidov, M., Djunusova, M., et al. (2005). Virulence of *Pyrenophora tritici-*repentis in the countries of the Silk Road. *Can. J. Plant Path.* 27, 383–388.

Lamari, L., Strelkov, S. E., Yahyaoui, A., Orabi, J., and Smith, R. B. (2003). The identification of two new Races of *Pyrenophora tritici-*repentis from the host center of diversity confirms a one-to-one relationship in tan

spot of wheat. *Phytopathology* 93, 391–396. doi: 10.1094/phyto.2003.93.4.391

Li, H. B., Yan, W., Liu, G. R., Wen, S. M., and Liu, C. J. (2011). Identification and validation of quantitative trait loci conferring tan spot resistance in the bread wheat variety Ernie. *Theor. Appl. Genet.* 122, 395–403. doi: 10.1007/s00122-010-1455-x

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, K., and Muse, S. V. (2005). Power Marker: integrated analysis environment for genetic marker data. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Liu, Z., El-Basyoni, I., Kariyawasam, G., Zhang, G., Fritz, A., Hansen, J., et al. (2015). Evaluation and association mapping of resistance to tan spot and Stagonospora nodorum blotch in adapted winter wheat germplasm. *Plant Dis.* 99, 1333–1341. doi: 10.1094/pdis-11-14-1131-re

Mao, S. L., Wei, Y. M., Cao, W., Lan, X. J., Yu, M., Chen, Z. M., et al. (2010). Confirmation of the relationship between plant height and Fusarium head blight resistance in wheat (*Triticum aestivum* L.) by QTL meta-analysis. *Euphytica* 174, 343–356. doi: 10.1007/s10681-010-0128-9

Maraite, H., Mercado-Vergnes, D., Renard, M. E., Zhanarbekova, A., and Duveiller, E. (2006). Relevance of pathogen diversity in management of leaf spot and leaf blight diseases on wheat in Central Asia. *Agromeridian* 2, 105–114.

McIntosh, R. A., Yamazaki, Y., Dubcovsky, J., Rogers, J., Morris, C., Appel, R., et al. (2013). *Catalogue of Gene Symbols for Wheat*. Available online at: www.shigen.nig.ac.jp/wheat/komugi/genes/ (accessed June 16, 2020).

McIntosh, R. A., Yamazaki, Y., Dubcovsky, J., Rogers, J., Morris, C., Somers, D. J., et al. (2008). *Catalogue of Gene Symbols For Wheat. MacGene 2008*. Available online at: http://wheat.pw.usda.gov/GG2/Triticum/wgc/2008/ (accessed June 16, 2020).

Miedaner, T., Wiirschum, T., Maurer, H. P., Korzun, V., Ebmeyer, E., and Reif, J. C. (2011). Association mapping for Fusarium head blight resistance in European soft winter wheat. *Mol. Breed.* 28, 647–655. doi: 10.1007/s11032-010-9516-z

Morgounov, A., Abugalieva, A., and Martynov, S. (2014). Effect of climate change and variety on long-term variation of grain yield and quality in winter wheat in kazakhstan. *Cereal Res. Commun.* 42, 163–172. doi: 10.1556/crc.2013.0047

Noriel, A. J., Sun, X., Bockus, W., and Bai, G. (2011). Resistance to tan spot and insensitivity to Ptr ToxA in wheat. *Crop Sci.* 51, 1059–1067. doi: 10.2135/cropsci2010.08.0464

Pandey, A. K., Burlakoti, R. R., Kenyon, L. C., and Nair, R. M. (2018). Perspectives and challenges for sustainable management of fungal diseases of mungbean [*Vigna radiata* (L.) R. Wilczek var. radiata]: a review. *Front. Environ. Sci.* 6:53. doi: 10.3389/fenvs.2018.00053

Patel, J. S., Mamidi, S., Bonman, J. M., and Adhikari, T. B. (2013). Identification of QTL in spring wheat associated with resistance to a novel isolate of *Pyrenophora tritici*-repentis. *Crop Sci.* 53, 842–852. doi: 10.2135/cropsci2012.01.0036

Phuke, R. M., He, X., Juliana, P., Bishnoi, S. K., Singh, G. P., Kabir, M. R., et al. (2020). Association mapping of seedling resistance to tan spot (*Pyrenophora tritici*-repentis race 1) in CIMMYT and South Asian wheat germplasm. *Front. Plant Sci.* 11:1309. doi: 10.3389/fpls.2020.01309

Postnikova, E. N., and Khasanov, B. A. (1998). "Tan spot in central Asia," in *Helminthosporium Blights of Wheat: Spot Blotch and Tan Spot. Proceedings of an International Workshop held at CIMMYT* (El Batan: CIMMYT), 107–113.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Rees, R. G., Platz, G. J., and Mayer, R. J. (1982). Yield losses in wheat from yellow spot: Comparison of estimates derived from single tillers and plots. *Austr. J. Agric. Res.* 33, 899–908. doi: 10.1071/ar9820899

Reif, J. C., Zhang, P., Dreisigacher, S., Warburton, M. L., Ginkel, M. V., Hoisington, D., et al. (2005). Wheat genetic diversity trends during domestication and breeding. *Theor. Appl. Genet.* 110, 859–864. doi: 10.1007/s00122-004-1881-8

Salnikov, V., Turulina, G., Polyakova, S., Petrova, Y., and Skakova, A. (2015). Climate change in Kazakhstan during the past 70 years. *Quaternary Int.* 358, 77–82. doi: 10.1016/j.quaint.2014.09.008

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., and Detering, F. (2011). Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc.* 5:54. doi: 10.1186/1753-6561-5-S7-P54

Santure, A. W., and Garant, D. (2018). Wild GWAS-association mapping in natural populations. *Mol. Ecol. Res.* 18, 729–738. doi: 10.1111/1755-0998.12901

Sehgal, D., Autrique, E., Singh, R., Ellis, M., Singh, S., and Dreisigacker, S. (2017). Identification of genomic regions for grain yield and yield stability and their epistatic interactions. *Sci. Rep.* 7:41578.

Sehgal, D., Dreisigacker, S., Belen, S., Küçüközdemir, Ü, Mert, Z., Özer, E., et al. (2016). Mining centuries old in situ conserved Turkish wheat landRaces for grain yield and stripe rust resistance genes. *Front. Genet.* 7:201.

Sehgal, D., Vikram, P., Sansaloni, C. P., Ortiz, C., Pierre, C. S., Payne, T., et al. (2015). Exploring and mobilizing the gene bank biodiversity for wheat improvement. *PLoS One* 10:e0132112. doi: 10.1371/journal.pone.0132112

Shabeer, A., and Bockus, W. W. (1988). Tan spot effects on yield and yield components relative to growth stage in winter wheat. *Plant Dis.* 72, 599–602. doi: 10.1094/pd-72-0599

Singh, P. K., Crossa, J., Duveiller, E., Singh, R. P., and Djurle, A. (2016). Association mapping for resistance to tan spot induced by *Pyrenophora tritici*-repentis Race 1 in CIMMYTs historical bread wheat set. *Euphytica* 207, 515–525. doi: 10.1007/s10681-015-1528-7

Singh, P. K., Mergoum, M., Gonzalez-Hernandez, J. L., Ali, S., Adhikari, T. B., Kianian, S. F., et al. (2008). Genetics and molecular mapping of resistance to necrosis inducing Race 5 of *Pyrenophora tritici*-repentis in tetraploid wheat. *Mol. Breed.* 21, 293–304. doi: 10.1007/s11032-007-9129-3

Singh, P. K., Singh, R. P., Duveiller, E., Mergoum, M., Adhikari, T. B., and Elias, E. M. (2010). Genetics of wheat - *Pyrenophora tritici*-repentis interactions. *Euphytica.* 171, 1–13. doi: 10.1007/s10681-009-0074-6

Singh, P. K., Singh, S., Deng, Z., He, X., Kehel, Z., and Singh, R. P. (2019). Characterization of QTLs for seedling resistance to tan spot and septoria nodorum blotch in the PBW343/kenya nyangumi wheat recombinant inbred lines population. *Int. J. Mol. Sci.* 20:5432. doi: 10.3390/ijms20215432

Somers, D. J., Banks, T., DePauw, R., Fox, S., Clarke, J., Pozniak, C., et al. (2007). Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome* 50, 557–567. doi: 10.1139/g07-031

Srinivasachary, S., Gosman, N., Steed, A., Hollins, T. W., Bayles, R., Jennings, P., et al. (2009). Semi-dwarfing Rht-B1 and Rht-D1 loci of wheat differ significantly in their influence on resistance to Fusarium head blight. *Theor. Appl. Genet.* 118, 695–702. doi: 10.1007/s00122-008-0930-0

Tadesse, W., Hsam, S. L. K., Wenzel, G., and Zeller, F. J. (2006a). Identification and monosomic analysis of tan spot resistance genes in synthetic wheat lines (*Triticum turgidum* L. × *Aegilops tauschii* Coss.). *Crop Sci.* 46, 1212–1217. doi: 10.2135/cropsci2005.10-0396

Tadesse, W., Hsam, S. L. K., and Zeller, F. J. (2006b). Evaluation of common wheat cultivars for tan spot resistance and chromosomal location of a resistance gene in the cultivar 'Salamouni'. *Plant Breed.* 125, 318–322. doi: 10.1111/j.1439-0523.2006.01243.x

Tommasini, L., Schnurbusch, T., Fossati, D., Mascher, F., and Keller, B. (2007). Association mapping of Stagonospora nodorum blotch resistance in modern European winter wheat varieties. *Theor. Appl. Genet.* 115, 697–708. doi: 10.1007/s00122-007-0601-6

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Vargas, M., Combs, E., Alvarado, G., Atlin, G., Mathews, K., and Crossa, J. (2013). META: a suite of SAS programs to analyze multi-environment breeding trials. *Agron. J.* 105, 11–19. doi: 10.2134/agronj2012.0016

Warburton, M. L., Crossa, J., Franco, J., Kazi, M., Trethowan, R., Rajaram, S., et al. (2006). Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica* 149, 289–301. doi: 10.1007/s10681-005-9077-0

Würschum, T., Langer, S. M., Longin, C. F., Korzun, V., Akhunov, E., Ebmeyer, E., et al. (2013). Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor. Appl. Genet.* 126, 1477–1486. doi: 10.1007/s00122-013-2065-1

Yu, L. X., Lorenz, A., Rutkoski, J., Singh, R. P., Bhavani, S., Huerta-Espino, J., et al. (2011). Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. *Theor. Appl.Genet.* 123, 1257–1268. doi: 10.1007/s00122-011-1664-y

Yu, L. X., Morgounov, A., Wanyera, R., Keser, M., Singh, S. K., and Sorrells, M. (2012). Identification of Ug99 stem rust resistance loci in winter wheat germplasm using genome-wide association analysis. *Theor. Appl.Genet.* 125, 749–758. doi: 10.1007/s00122-012-1867-x

Zadoks, J. C., Chang, T. T., and Konzak, M. M. (1974). A decimal code for the growth stages of wheat. *Weed Res.* 14, 415–421. doi: 10.1111/j.1365-3180.1974.tb01084.x

Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., and Ogbonnaya, F. C. (2014). Genome-Wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploidy wheat. *PLoS One* 9:e105593. doi: 10.1371/journal.pone.0105593

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kokhmetova, Sehgal, Ali, Atishova, Kumarbayeva, Leonova and Dreisigacker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fgene.2020.600378

Silencing of an Ubiquitin Ligase Increases Grain Width and Weight in *indica* Rice

Ankit Verma[1], *Geeta Prakash*[1,2], *Rajeev Ranjan*[1,3], *Akhilesh K. Tyagi*[1,3] *and Pinky Agarwal*[1]*

[1] National Institute of Plant Genome Research, New Delhi, India, [2] Department of Botany, Gargi College, University of Delhi, New Delhi, India, [3] Department of Plant Molecular Biology, University of Delhi, New Delhi, India

Many quantitative trait loci (QTLs) have been identified by molecular genetic studies which control grain size by regulating grain width, length, and/or thickness. *Grain width 2* (*GW2*) is one such QTL that codes for a RING-type E3 ubiquitin ligase and increases grain size by regulating grain width through ubiquitin-mediated degradation of unknown substrates. A natural variation (single-nucleotide polymorphism at the 346th position) in the functional domain-coding region of *OsGW2* in *japonica* rice genotypes has been shown to cause an increase in grain width/weight in rice. However, this variation is absent in *indica* rice genotypes. In this study, we report that reduced expression of *OsGW2* can alter grain size, even though natural sequence variation is not responsible for increased grain size in *indica* rice genotypes. *OsGW2* shows high expression in seed development stages and the protein localizes to the nucleus and cytoplasm. Downregulation of *OsGW2* by RNAi technology results in wider and heavier grains. Microscopic observation of grain morphology suggests that OsGW2 determines grain size by influencing both cell expansion and cell proliferation in spikelet hull. Using transcriptome analysis, upregulated genes related to grain size regulation have been identified among 1,426 differentially expressed genes in an *OsGW2*_RNAi transgenic line. These results reveal that *OsGW2* is a negative regulator of grain size in *indica* rice and affects both cell number and cell size in spikelet hull.

Keywords: E3 ubiquitin ligase, grain width, GW2, *Oryza sativa*, seed development

OPEN ACCESS

Edited by:
Christian Meyer, INRA UMR1318 Institut Jean Pierre Bourgin, France

Reviewed by:
Longbiao Guo, Chinese Academy of Agricultural Sciences, China
Himanshu Sharma, National Agri-Food Biotechnology Institute, India

***Correspondence:**
Pinky Agarwal
pinky.agarwal@nipgr.ac.in; pinky.agarwal@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *29 August 2020*
Accepted: *27 November 2020*
Published: *12 January 2021*

Citation:
Verma A, Prakash G, Ranjan R, Tyagi AK and Agarwal P (2021) Silencing of an Ubiquitin Ligase Increases Grain Width and Weight in indica Rice. Front. Genet. 11:600378. doi: 10.3389/fgene.2020.600378

INTRODUCTION

Globally, rice (*Oryza sativa* L.) is a staple cereal food crop since it is the primary source of caloric intake and is consumed by more than half of the world's population (Sasaki, 2008; Zuo and Li, 2014; Azizi et al., 2019). In upcoming decades, the overall demand of rice will increase, particularly in Asia and Africa (Kubo and Purevdorj, 2004). Hence, there is an urgent need to increase rice grain yield. Rice grain size/weight is the key agronomic trait for the improvement of yield. Rice grain yield is a complex trait, which is determined by four typical quantitative component traits, i.e., number of panicles per plant, number of filled grains per panicle, grain size, and grain weight (Xing and Zhang, 2010; Zuo and Li, 2014). Grain size (grain length and width)/weight is the most vital yield-contributing complex quantitative trait in rice (Fan et al., 2006; Mao et al., 2010; Ying et al., 2012). Therefore, the best approach for enhancing rice yield is to find/develop new high yielding varieties with increased grain size/weight and superior grain nutrient quality (Rosegrant and Cline, 2003). Grain weight and grain size are positively associated with each other. Grain size

is defined in terms of its length, width, and thickness. Rice grain size is mainly controlled by genes that determine cell expansion and/or proliferation in spikelet hull and contribute in endosperm development. All these yield component traits are controlled by naturally occurring quantitative trait loci, QTLs (Zuo and Li, 2014). Several QTLs and a number of genes regulating grain size have been identified and functionally characterized by different approaches such as genomics, proteomics, metabolomics, genome editing, and genome-wide association to find out molecular components and genetic regulatory mechanisms controlling grain size trait in rice (Li and Li, 2014; Li et al., 2018; Azizi et al., 2019). Hitherto, much needs to be explored regarding the molecular mechanisms as well as the underlying genes controlling rice grain size/weight to understand the mechanism of grain development, which will provide a way to improve yield and quality.

Ubiquitination has been shown to play a significant role in determining seed size in crop plants (Song et al., 2007; Li et al., 2008; Choi et al., 2018). It is an enzymatic post-translational modification in which ubiquitin protein (76 residues long) covalently attaches with a target protein (Li and Li, 2014), with the consecutive participation of three unique enzymes, E1, E2, and E3 (Moon et al., 2004). E1 is a ubiquitin-activating enzyme. E2 conjugates with activated ubiquitin. Ubiquitin ligase E3 determines substrate specificity (Smalle and Vierstra, 2004). This process regulates the stability, activity, and localization of modified target proteins. Mono-ubiquitination of a target protein affects its function and localization, whereas conjugation of multi-ubiquitin can degrade modified proteins through the ubiquitin−26S proteasome complex (Vierstra, 2009).

Grain yield and quality are controlled by several QTLs and genes. A number of grain size related QTLs (*GW2*, *GW5*, *GS3*, *GS5*, *qGL3*, *TGW6*, *GW8*, etc.) and genes have been functionally characterized in the last few decades. They regulate either cell division or cell expansion, or both, during the process of seed development, by being involved in signaling pathways facilitated by ubiquitin-mediated proteasomal degradation, transcription factors, guanine nucleotide-binding proteins (G-proteins), protein kinases and phytohormones (Zuo and Li, 2014; Wang et al., 2019; Zhao et al., 2019). *GW2* is one such QTL that regulates grain weight/yield by regulating grain width in *japonica* rice genotypes. It codes for a RING-type protein having E3 ubiquitin ligase activity that binds with its substrates and targets them for degradation through the ubiquitin−26S proteasomal complex. A natural variation in long grain WY3 cultivar results in truncation of the functional domain of GW2. This results in a wider spikelet hull and an increase in the number of cells in the outer parenchyma cell layer, without affecting cell numbers in the endosperm. Hence, it has been identified that *GW2* negatively regulates grain size and weight in *japonica* rice cultivars (Song et al., 2007). Chitinase 14 (CHT14) and phosphoglycerate kinase (PGK) are involved in carbohydrate metabolism and show strong interaction with GW2, but are not involved in ubiquitin-mediated degradation (Lee et al., 2018). GW2 also targets expansin-like 1 (EXPLA1), which is a cell wall-releasing protein, for degradation through the proteasome pathway and hence participates in determining seed size and weight (Choi et al., 2018). WIDE AND THICK GRAIN 1 (WTG1) also controls rice grain size through the ubiquitin–proteasome pathway. WTG1 is a deubiquitinating enzyme, an otubain-like protease. WTG1/OsOTUB1 regulates seed size/shape by affecting only cell expansion in the hull, suggesting a different pathway than GW2 (Jiao et al., 2010; Huang et al., 2017; Li and Li, 2019). DA2 is the homolog of GW2 in *Arabidopsis thaliana* that also controls seed and organ size. Rice GW2 homologs have been identified in maize and wheat, where they regulate kernel size (Li et al., 2010; Sestili et al., 2019).

Although the specific natural variation in *GW2*, causing its inactivation, results in higher grain width/weight in *japonica* rice genotypes, such a sequence variation is absent in *indica* genotypes (Dixit et al., 2013). The aim of the current study was to examine if, despite the absence of this variation, downregulation of *GW2* in *indica* rice would alter grain size. Downregulation in transgenic rice plants, by RNAi, revealed that the grain width/weight in *indica* rice genotypes increases due to higher cell division as well as expansion in spikelet hull which in turn is controlled by differential gene expression of many seed development-related genes.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Rice *indica* and *japonica* genotypes (PB1, IR64, Sonasal, LGR, and Nipponbare) were grown in experimental fields of NIPGR (New Delhi, India) during the months of June to October (T_{max} 35–40°C; T_{min} 25–28°C). Rice transgenic plants along with the wild-type lines (WT) were grown in a mix of soil and compost (3:2), which was supplemented with NPK, in a greenhouse with a 12-h/12-h light/dark cycle, at 28/23°C temperature, 65% relative humidity, and 250 μmol m^{-2} s^{-1} photosynthetic active radiation (PAR) light. The complete development of rice seed occurs in five different stages based on morphological changes occurring in the embryo and endosperm categorized as S1 (0–2 DAP, days after pollination), S2 (3–4 DAP), S3 (5–10 DAP), S4 (11–20 DAP), and S5 (21–29 DAP) (Agarwal et al., 2007, 2011). In accordance, rice panicles were tagged and harvested as mentioned previously (Sharma et al., 2012). Flag leaf and seed tissue from each variety, at each stage, were harvested and frozen in liquid nitrogen and stored at −80°C until further use.

RNA Isolation, cDNA Preparation, and Quantitative RT-PCR

Isolation of RNA from the leaves and seeds (S1–S5 stages) of different rice genotypes (PB1, IR64, Sonasal, LGR, and Nipponbare) as well as transgenic plants was carried out using TRIzol® reagent (Invitrogen™ Life Technologies, USA) as detailed previously (Singh et al., 2003; Mathew et al., 2016, 2020; Das et al., 2019). The quantification of RNA samples was done by NANODROP 2000c Spectrophotometer (Thermo Scientific, USA). For the amplification of the *OsGW2* gene (LOC_Os02g14720), full-length cDNAs of five rice genotypes, mentioned above, were used. For full-length cDNA synthesis, 2 μg of purified RNA and Superscript III® first-strand cDNA synthesis kit (Invitrogen™ Life Technologies, USA) was used.

Reverse transcription of purified RNA was performed with High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, USA) with random primers. Gene-specific primers were designed by using Primer Express Version 3.0 (Applied Biosystems, USA). For expression analysis of *OsGW2*, qRT-PCR was performed in three biological replicates using gene-specific primers (**Supplementary Table 1**), with Fast SYBR® Green Master Mix (Applied Biosystems, USA). Real-time PCR was done in 7500 Fast Real-Time System (Applied Biosystems, USA) using KAPA SYBR® FAST qPCR kit and 7500 software v2.0.1 for data analysis. *OsACTIN1* gene was the endogenous control for leaf and seed tissue throughout the study. Relative gene expression level was calculated by the comparative Ct ($2^{-\Delta\Delta Ct}$) method.

Sequence Analysis of *OsGW2* in Different Rice Genotypes

The *OsGW2* gene coding region was amplified from different rice genotypes (IR64, Sonasal, LGR, and PB1) using full-length cDNA for each (**Supplementary Table 1**) and cloned in pJET$^{1.2}$ vector (Thermo Scientific, USA). Cloning was confirmed by restriction digestion by *Bgl*II and confirmed clones were sequenced on ABI 3730xl DNA Analyzer (Applied Biosystems, USA) in three replicates. The obtained sequences were analyzed by Clustal Omega tool (Sievers et al., 2011) for aligning the insert sequence with template sequence extracted from the Rice Genome Annotation Project (RGAP, http://rice.plantbiology.msu.edu/). Any difference in either nucleotide or subsequent protein sequence was noted.

Preparations of RNAi Construct and Rice Plant Transformation

To generate the RNAi construct, a 510-bp unique region of *OsGW2* was delineated by Rational siRNA Design program (Reynolds et al., 2004). This unique region of *OsGW2* was amplified along with CACC sequence added to the 5′ end of the forward primer (**Supplementary Table 1**), from the *OsGW2*_pJET clone. The amplified fragment was purified using GeneJET gel extraction kit (Thermo Scientific, USA), followed by cloning into Gateway® entry vector pENTR™/D-TOPO® (Invitrogen™, USA) containing recombination sites (*attL1* and *attL2*) using pENTR™/Directional TOPO® cloning kit as per the manufacturer's protocol. Cloning was confirmed by restriction digestion with *Asc*I and *Not*I restriction enzymes and sequencing of two positive clones. Plants with reduced expression or knockdown of *OsGW2* were made using RNAi vector pANDA (Miki and Shimamoto, 2004), in which transgene expression was under the control of the maize *UBIQUITIN* promoter. The insert was transferred to pANDA having kanamycin and hygromycin selection marker genes using Gateway™ LR Clonase™ II Enzyme Mix (Invitrogen™, USA) through Gateway® cloning technology. Target sequence was incorporated in opposite orientation in two *attR* sites located at both sides of the *GUS* linker sequence, confirmed by restriction digestion by using *Kpn*I and *Sac*I enzymes. The confirmed clone was introduced into the *Agrobacterium tumefaciens* strain *EHA105*. Transgenic plants were generated by transformation of the RNAi construct into rice calli (PB1 genotype IET-10364) through the *Agrobacterium*-mediated transformation method (Toki et al., 2006; Das et al., 2019; Mathew et al., 2020).

Phenotypic Characterization of Rice Transgenic Plants

Vegetative phenotypic characters (plant height, flag leaf length and width, number of tillers per plant, number of panicles per plant, and number of grains per panicle) were measured manually and compared with the respective wild-type plant in T_0, T_1, and T_2 generations. Six biological replicates were taken to quantitate each vegetative character for each line. Grain parameters (length and width) were measured after the grains were harvested and dried, by using WinSEEDLE™ (Regent Instruments Inc., Canada). Photographs of grain length and width were taken using a Nikon D5200 48 MP DSLR camera.

Scanning Electron Microscopy

For scanning electron microscopy analysis, the mature grains of homozygous T_2 transgenic line *OsGW2*_RNAi_7AP9 and wild-type plants were harvested. To study cell size and number in the outer epidermis, a scanning electron microscope (SEM, Zeiss, Germany) was used to observe the central parts of the lemma surface of mature grains, under different magnifications (500×, 1,000×, and 2,500×). For analyzing starch granules, mature rice grains were dehusked manually. These were sectioned longitudinally with a scalpel and the central parts of the endosperm were photographed by SEM. To analyze cell length and cell area of the outer epidermal cells, ImageJ software was used (Schneider et al., 2012).

Subcellular Localization of OsGW2

The full-length coding sequence of *OsGW2* from PB1, lacking the stop codon, was amplified (**Supplementary Table 1**) and cloned downstream of the YFP coding region in the pSITE-3CA-YFP vector (Chakrabarty et al., 2007) under the control of CaMV *35S* promoter to generate the recombinant plasmid YFP-OsGW2. To study the subcellular localization of OsGW2, YFP-OsGW2 recombinant plasmid and control YFP vector were bombarded separately into onion peel cells by particle bombardment using Biolistic®-PDS-1000/He™ system (Bio-Rad, USA). After incubation for 16 h in the dark at 28°C, YFP fluorescence was observed with a confocal laser microscope (TCS SP2, Leica, Germany).

Transcriptome Analysis of *OsGW2*_RNAi Plants

To study the functional relevance of *OsGW2*, RNA-seq analysis of T_3 grains, from S4 stage, of *OsGW2*_RNAi_7AP9 transgenic plants, along with the wild-type lines, was performed in two biological replicates. RNA isolation and quantification was done as mentioned above and RNA integrity was checked on Bioanalyzer (2100 Agilent Technologies, USA). RNA having RIN ≥8.0 were used for library preparation. Total RNA (∼5 μg) for each sample was used for cDNA library preparation and sequencing. Transcriptome sequencing was performed

on Illumina Hiseq^{TM} 2000 platform with two independent biological replicates, each for WT and *OsGW2*_RNAi_7AP9. After obtaining sequencing data, different tools were used for performing preprocessing of data, alignment with reference, expression estimation, and comparison analysis, as detailed previously (Mathew et al., 2020). The preprocessing tools used were Adapter Removal v2 (version 2.2.0) and bowtie2 (version 2.2.9). Processed reads were aligned in STAR (version 2.5.3a) with the MSU rice genome database (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir). Cufflinks was used for estimating gene expression (version 2.2.1) and output was in terms of fragments per kilobase million reads (FPKM). To identify differentially expressed genes (DEGs), Cuffdiff utility provided in Cufflinks package was used. The $\log_2$ fold change cutoff was $\geq$1.5 (upregulated genes) and ≤ -1.5 (downregulated genes) with a *P*-value cutoff of ≤ 0.05.

Functional annotation of DEGs in metabolic pathways/processes was studied by using databases including the KEGG pathway analysis and MSU-RGAP 7 (Ouyang et al., 2006). Characterized differentially expressed genes were separated and analyzed using RICENCODE, and seed-related genes were extracted by comparing DEGs with the data available from previous studies in the lab and online sources, as done previously (Mathew et al., 2020). Heat maps were constructed using MeV Version 4.9.0 (Saeed et al., 2003). Gene ontology enrichment for various DEGs was performed using the AgriGO software (v1.2; Du et al., 2010) with MSU-RGAP 7 as the reference annotation set. Rice GO information for biological process, cellular component, and molecular function was used for gene ontology enrichment analysis. For pathway analysis of differentially expressed genes, the MapMan software (version 3.5.1; http://mapman.gabipd.org/web/guest) was used with a *P*-value cutoff of $\leq$0.05 and mapping files from MSU-RGAP 7.

RESULTS

OsGW2 Sequence Variation in Rice Genotypes

A natural variation (single-nucleotide polymorphism) was present in large grain-sized *japonica* rice genotypes in comparison with small grain-sized genotypes which was responsible for truncation of encoded OsGW2 protein functional domain in the large grain-sized genotypes (Song et al., 2007). No such variation was observed in *indica* genotypes (Dixit et al., 2013). In our analysis, the coding region of *OsGW2* (1,278 bp) from four *indica* rice genotypes, selected on the basis of different grain sizes as extra large, large, medium, and small (LGR, PB1, IR64, and Sonasal), was sequenced and compared with *japonica* cultivar Nipponbare (**Figure 1A**). The coding region showed 100% similarity amongst all four *indica* rice genotypes (LGR, PB1, IR64, and Sonasal). However, the *japonica* cultivar, Nipponbare, showed variation with respect to *indica* sequence. Comparison of the nucleotide sequences of the alleles of *OsGW2* from both *indica* and *japonica* cultivars revealed three nucleotide changes, namely two single-nucleotide polymorphism (SNP) variations and one 6-bp addition in the coding region of *indica* genotypes. SNP variation (G to A) at position 572 bp resulted in an amino acid change (Gly to Glu), six base pair addition at positions 714–719 resulted in the addition of two amino acids (Gln and Glu), whereas (A–G) SNP variation at 1,116 bp did not result in any amino acid variation (**Figure 1A**). Despite these sequence variations, the conserved domain structure of the OsGW2 protein remained the same, and it did not show truncation as reported for *japonica* rice genotypes with increased grain size (Song et al., 2007).

OsGW2 Expresses Mainly in Seed Stages and Its Protein Localizes to the Nucleus and Cytoplasm

To determine the expression patterns of *OsGW2*, transcript levels were examined by qRT-PCR in S1–S5 seed developmental stages of four *indica* and one *japonica* rice genotype in comparison with their respective flag leaf. These rice genotypes show variation in grain size parameters, viz. grain width, length, and weight (**Supplementary Figure 1**). Previous data showed that *OsGW2* expressed constitutively in vegetative as well as reproductive tissues (Song et al., 2007). qRT-PCR data revealed that *OsGW2* expressed at lower levels in S1–S3 stages and had higher expression in S4–S5 stages of seed development, which indicates its essential role in rice seed development (**Figure 1B**). To study the subcellular localization of OsGW2, OsGW2 yellow fluorescent protein (YFP-OsGW2) fusion construct was designed where expression was controlled by the CaMV *35S* promoter. *In vivo* protein targeting in onion epidermal cells by bombarding with the YFP-OsGW2 fusion construct showed that it localized to both the nucleus and cytoplasm. This was observed in six independent cells (**Figure 1C**).

Suppressed Expression of *OsGW2* Alters Grain Morphology

To functionally validate *OsGW2*, we generated *OsGW2* RNA interference (RNAi) lines in *indica* PB1 (IET-10364) background, and a 510-bp unique region was targeted to suppress the expression of *OsGW2*. A total of 10 *OsGW2*_RNAi transgenic lines (RNAi_2, 4, 7, 9, 13, 15, 17, 21, 23, 24) were obtained through tissue culture. Each plant of all these transgenic lines in T_0 generation was found positive by PCR amplification of hygromycin resistance gene (*Hygromycin phosphotransferase II*) resulting in an amplicon of 850 bp (**Supplementary Figure 2**). Vegetative parameters (plant height, flag leaf length, flag leaf width, number of tillers per plant) did not show any morphological difference between RNAi and wild-type PB1 plants. However, transgenic plants in T_0 generation showed significantly higher grain width (23–38%), grain length (5–15%), and grain weight (19–58%) when compared with the wild-type lines (**Figures 2A–C**; **Supplementary Figure 2**). In T_1 generation, four lines (RNAi 2A, 7A, 9B, and 24P2) were selected out of 10 lines which showed 3:1 segregation ratio (**Supplementary Table 2**). These *OsGW2* RNAi plants had significantly higher grain width, length, and weight in comparison with PB1 seeds, in T_1 generation (**Figures 2D–F**).

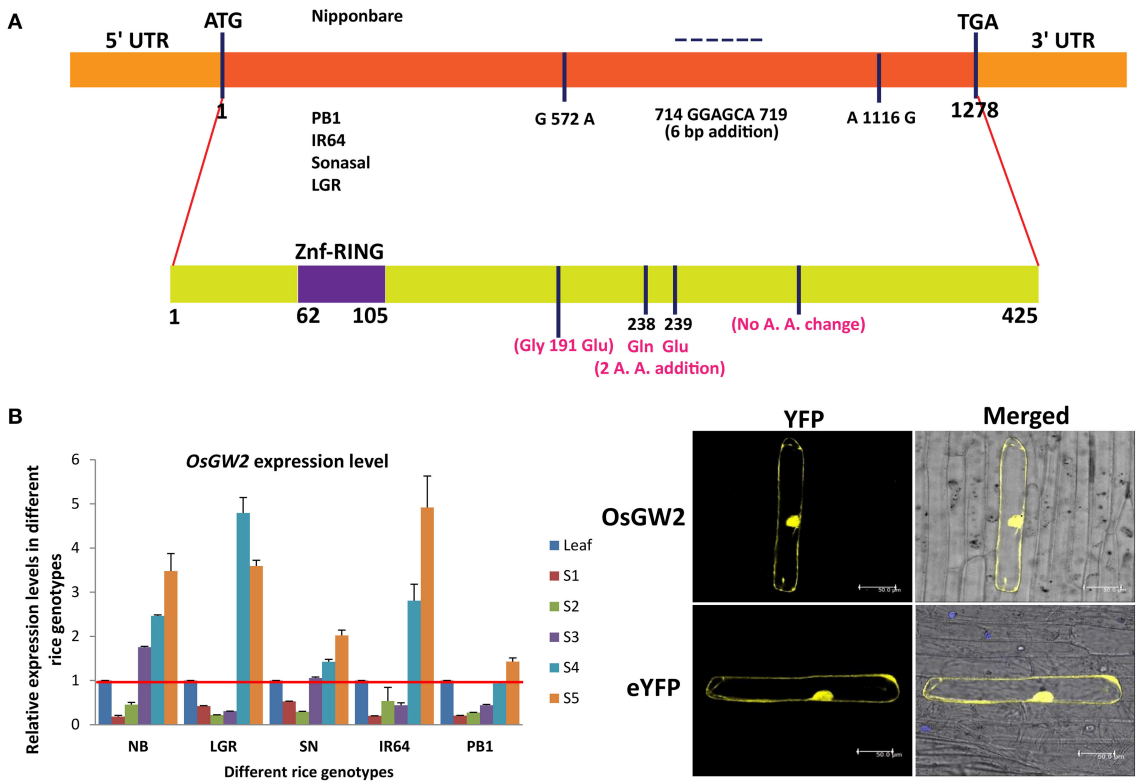


FIGURE 1 | Structure, expression, and subcellular localization of OsGW2. **(A)** *OsGW2* nucleotide (upper panel) and amino acid (lower panel, marked in pink color) sequence variation in *indica* rice genotypes (IR64, PB1, Sonasal, and LGR) including substitutions and additions in comparison with the *japonica* rice genotype (Nipponbare) and protein structure variation. A. A. represents amino acid. **(B)** Expression levels of *OsGW2* in seed tissues (S1–S5) of five different rice genotypes determined by qRT-PCR. The rice *OsACTIN1* gene was used as an internal control. Data show fold changes with respect to the flag leaf of the same genotype, and bar represents mean ± SD ($n = 3$). **(C)** Subcellular localization of YFP-tagged OsGW2 in onion peel cells (top panel). Lower panel represents empty vector control. Scale bar, 50 μm.

Other phenotypic parameters did not show significant difference in comparison with wild-type plants as before. The four lines were taken forward, and in T_2 generation, suppressed expression of *OsGW2* (RNAi 2AP9, 7AP9, 9BP3, and 24P2P5) was screened and used for further analysis. Two lines 2AP9 and 7AP9 had homozygous plants. Positive *OsGW2*_RNAi transgenic plants along with the wild-type lines were grown and analyzed (**Figure 3A**; **Supplementary Figure 3**). Expression levels were significantly reduced (53.88–84.84%) in the third leaf (as the gene expresses ubiquitously) of *OsGW2*_RNAi T_2 transgenic plants. T_3 transgenic seeds from S4 stage of the two homozygous lines also showed a significant decrease (66–88%) in the expression of *GW2* (**Figure 3B**). The grains of T_2 generation plants were significantly larger than those of the wild-type lines (**Figure 3E**), showing increased grain width (4.0–18.98%) and grain length (4.09–6.19%) (**Figures 3C,D,F,G**). GW2 is known to affect seed width in *japonica* rice (Song et al., 2007). It has also been shown to affect width and, hence, weight in wheat (Sestili et al., 2019). In our experiments, T_1 seeds showed a significant increase in length. T_2 seeds did show an increase in length but it was not significant (**Figure 2E**). Again, T_3 seeds from two homozygous lines (**Figure 3G**) showed a significant increase in length. This may be possible because the *t*-test was performed on 10 T_1 seeds from one plant, in triplicates. Further, 10 T_2 seeds were taken from three plants and 100 T_3 seeds were taken from three plants to calculate seed parameters. All T_1 seeds generated by each plant and T_3 seeds from two homozygous plants showed a significant increase in seed length, probably because of less variation in the sample taken. Furthermore, the 30-grain weight at the mature stage of *OsGW2*_RNAi T_3 seeds was 12.94–42.96% higher than that of the wild-type grains (**Figure 3H**). These results indicated that reduced expression of *OsGW2* could significantly increase grain width and weight even in *indica* rice, where a natural polymorphism was absent.

OsGW2 Regulates Rice Grain Size by Influencing Cell Division and Expansion

Spikelet hull size is delimiting to the final grain size. Since cell proliferation and/or cell expansion is responsible for alteration in grain size, the cell number of the *OsGW2*_RNAi_7AP9 line was calculated in the central part of the lemma epidermis using scanning electron microscopy images (**Figure 4A**). The cell length and cell area of the outer lemma epidermis of the wild-type and *OsGW2*_RNAi_7AP9 T_3 grains were

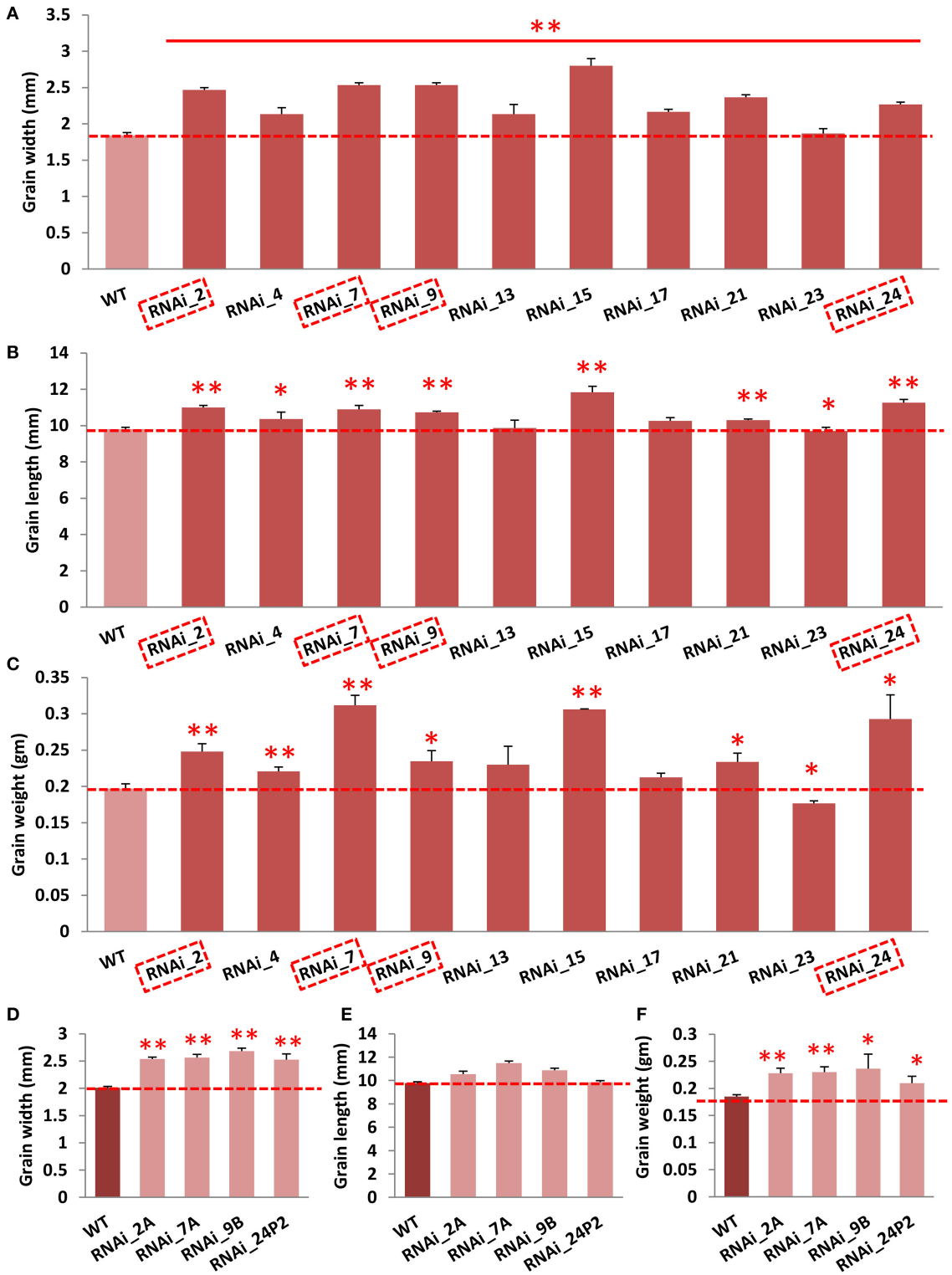


FIGURE 2 | Effect of *OsGW2* knockdown on width, length, and weight of T_1 and T_2 rice grains. **(A–C)** Grain width, length, and 10-grain weight, respectively, of *OsGW2*_RNAi T_1 seeds. **(D–F)** Grain width, length, and 10-grain weight, respectively, of *OsGW2*_RNAi T_2 seeds. WT represents wild-type PB1 grains; grain length and width were measured with WINSEEDLE^{TM}. RNAi_2, 4, 7, 9, 13, 15, 17, 21, 23, and 24 are 10 different transgenic lines in T_0 generation. Dashed boxes in **(A–C)** represent selected transgenic lines for further analyses which follow 3:1 segregation in T_1 generation. Hence, RNAi_2A, 7A, 9B, and 24P2 are four different transgenic lines in T_1 generation, selected from T_0 generation. Bar represents mean ± SD [$n = 10$ in **(A–E)**; $n = 3$ in **(C,F)**]. $^{**}P < 0.01$; $^{*}P < 0.05$, determined by Student's *t*-test.

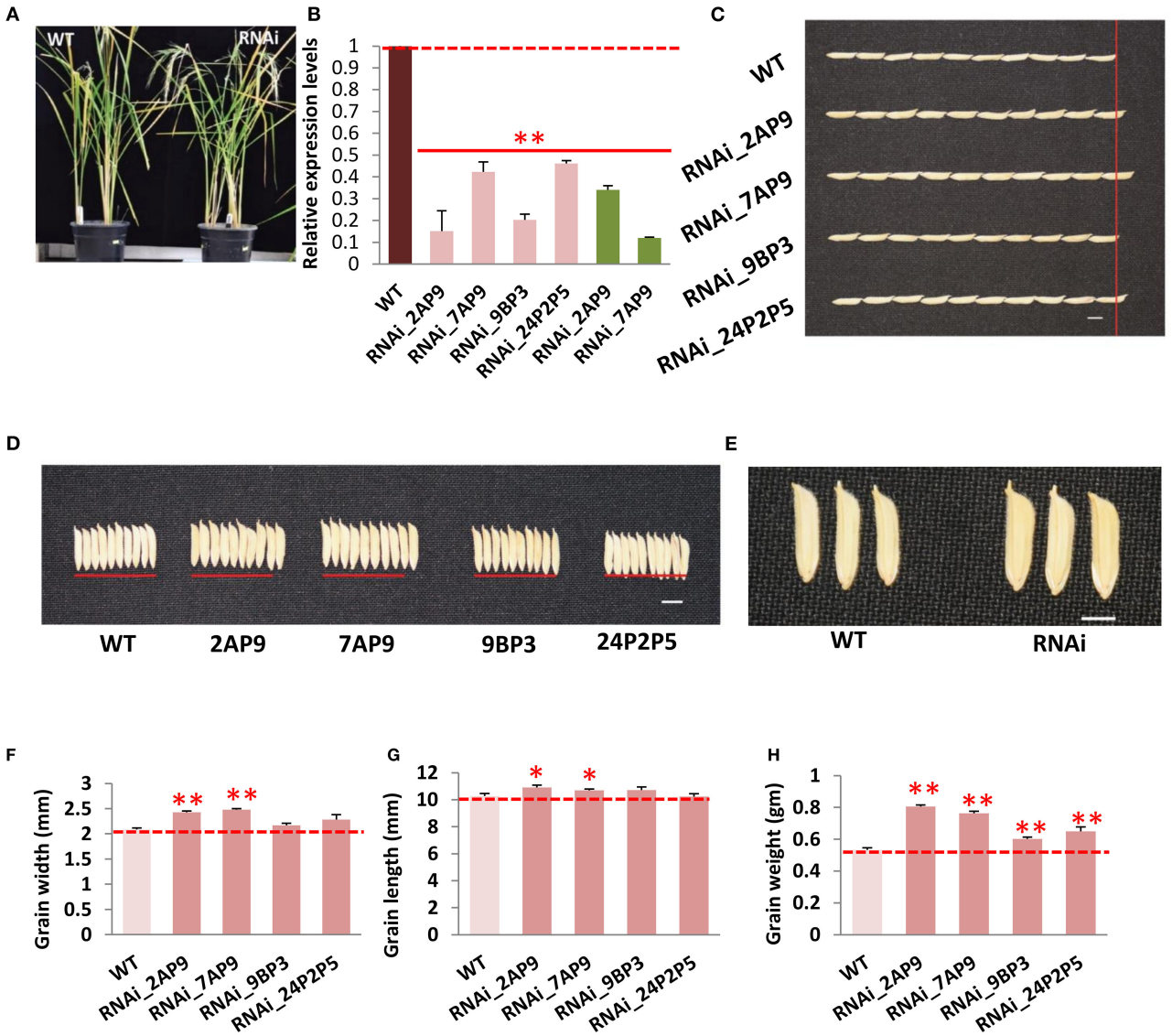


FIGURE 3 | Phenotypic characterization of *OsGW2*_RNAi T_2 generation transgenic plants. **(A)** Gross morphology of wild-type (WT) and *OsGW2*_RNAi plants at the mature stage. **(B)** Relative expression levels, by qRT-PCR, of *GW2* in *OsGW2*_RNAi plants (T_2 flag leaf, pink bars and T_3 grain, green bars) as compared with the respective wild-type tissues. Data show fold changes in expression level with respect to the wild-type lines; values represent mean ± SD ($n = 3$), $^{**}P < 0.01$, Student's *t*-test. **(C)** Grain length of *OsGW2*_RNAi T_3 seeds. **(D)** Grain width of *OsGW2* RNAi T_3 seeds. **(E)** Zoomed in image of grains of WT and *OsGW2*_RNAi_7AP9 T_3 grains. Bar = 1 cm. Comparison of **(F)** grain width, **(G)** grain length, and **(H)** 30-grain weight of *OsGW2*_RNAi T_3 seeds with wild-type PB1 seeds. Error bar represents mean ± SD [$n = 100$*3 in **(F,G)**, $n = 3$ in **(H)**]. $^{**}P \leq 0.01$, and $^{*}P \leq 0.05$, calculated by Student's *t*-test. RNAi (2AP9, 7AP9, 9BP3, 24P2P5) are four different lines in T_2 generation.

compared by SEM (**Figures 4B,C**). The images revealed that total lemma cell number in *OsGW2*_RNAi was increased by 12.94% compared with that of the wild-type lines (**Figure 4D**), resulting from a significant decrease in cell length and cell area. This suggested that decreased expression level of *OsGW2* contributes to higher grain width by affecting both cell division and cell expansion to regulate the number and size of cells of the spikelet hull during the process of development. To check whether or not the suppressed expression of *OsGW2* influences the structure of starch granules in the endosperm, we observed seed cross-sections by a scanning electron microscope. In the WT endosperm, starch granules were tightly packed. The spherical starch granules (Dong et al., 2015) were more in comparison with the polyhedron-shaped ones. However, in the *OsGW2*_RNAi_7AP9 line, starch granules were loosely packed and were larger in size, and the distribution of spherical and polyhedron granules was similar (**Figure 4E**). Thus, suppression of *OsGW2* additionally resulted in modifications in starch granule size.

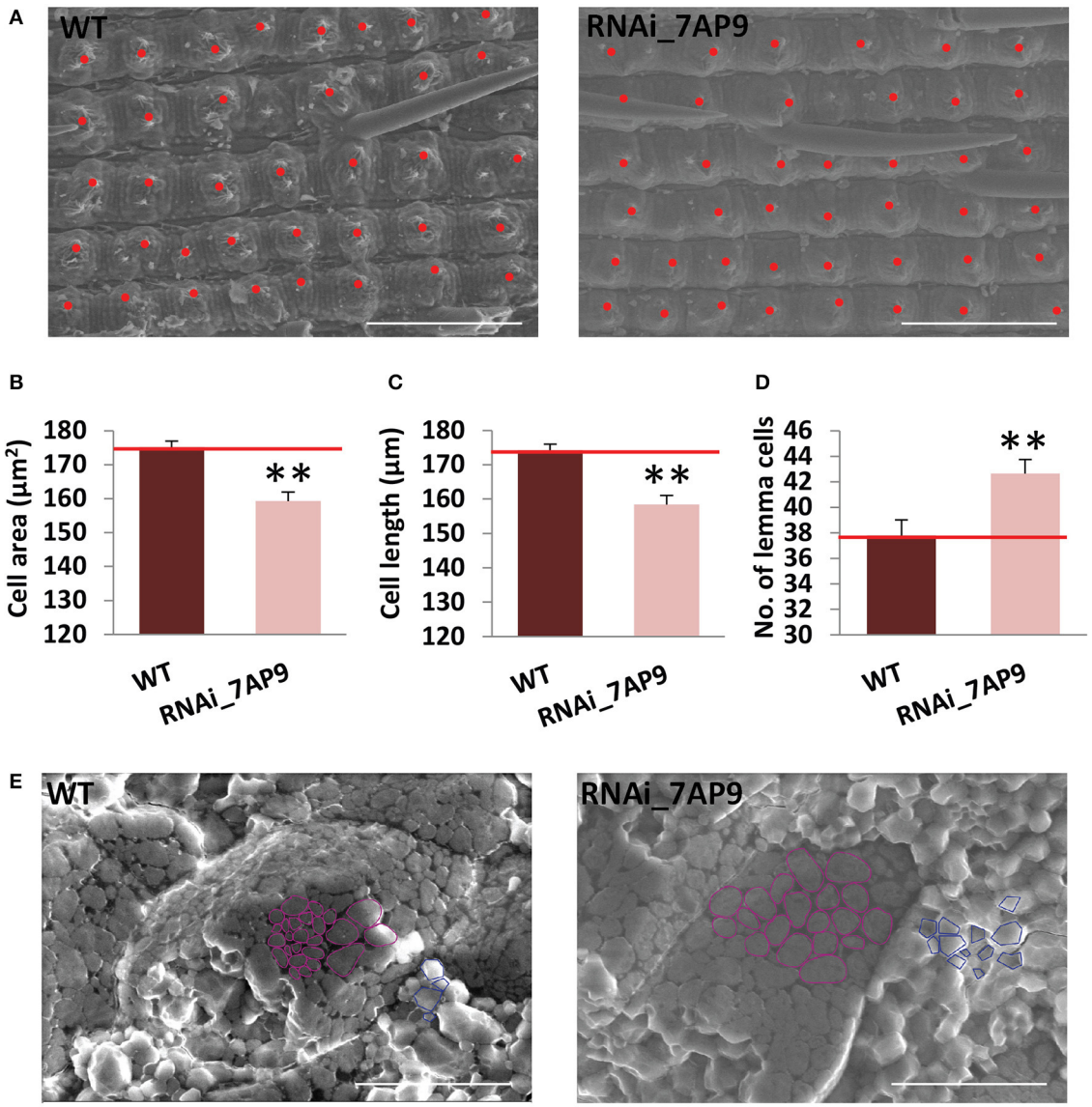


FIGURE 4 | *OsGW2* influences cell proliferation and expansion in T_3 spikelet hulls. **(A)** SEM photographs of the outer surface of lemma from wild-type and *OsGW2*_RNAi_7AP9 T_3 seeds. Each red dot represents a cell. **(B–D)** Cell area, cell length, and number of cells on the epidermis of lemma surface as seen in the area under the microscope. Bar represents mean ± SD; $n = 3$. $^{**}P \leq 0.01$, calculated by Student's *t*-test. **(E)** Scanning electron microscopy (SEM) analysis of starch granules of endosperm in the wild-type and *OsGW2*_RNAi_7AP9 T_3 seeds at the mature stage. As a representation, spherical granules are outlined in magenta, while the polyhedron ones are by a blue outline. WT represents wild-type; RNAi_7AP9 represents the knockdown line of *OsGW2*. Scale bar in **(A)** and **(E)** = 20 μm.

Transcriptome Analysis

The transcriptome of the developing seed was compared at S4 stage between WT and *OsGW2*_RNAi_7AP9 (a homozygous line) to elucidate genes and pathways which could be altering the grain size. Two biological replicates of each sample were used for sequencing and four libraries were generated in total. Each library had more than 50–78 million clean reads after quality control and 89–93% reads mapped uniquely to the rice reference sequences (**Supplementary Table 3**). Pearson correlation coefficient (R^2) was calculated between biological replicates. Correlation was higher (0.98) within the biological replicates than that between the WT and *OsGW2*_RNAi_7AP9 line. Published data shows that in rice, using RNA sequencing data from two biological replicates, ∼22–61 million clean reads have been generated and 62–92% of these reads map to the reference genome (Ding et al., 2017; Yuenyong et al., 2018; Jung et al., 2019; Zha et al., 2019). The transcripts with $\log_2$ fold change ≥1.5 (upregulated genes) and ≤(−1.5) (downregulated genes) with a *P*-value cutoff of ≤0.05, in *OsGW2*_RNAi_7AP9 transgenic line in comparison with WT, were considered as DEGs.

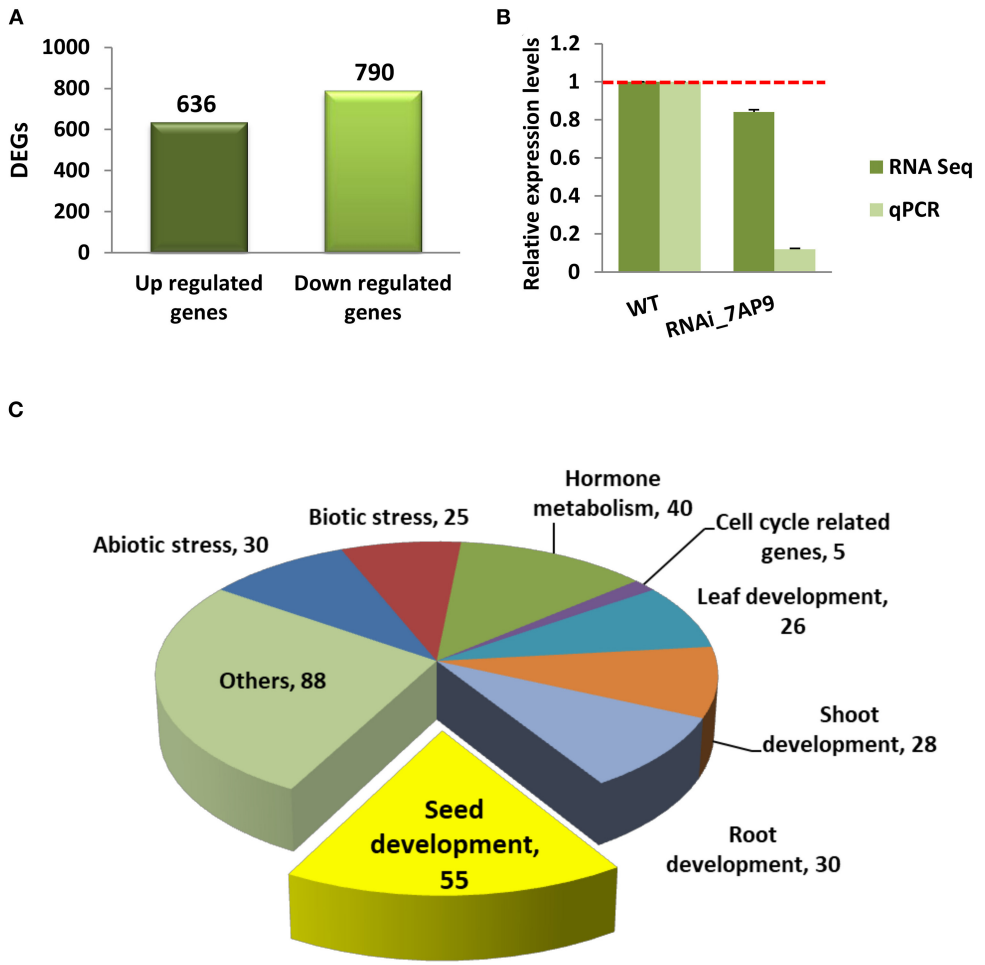


FIGURE 5 | Transcriptome analysis of *OsGW2*_RNAi T_3 generation S4 stage seeds. **(A)** The number of up- and downregulated DEGs between WT and *OsGW2*_RNAi_7AP9 seeds is represented as bars. The DEGs with $\log_2$ fold change ≥1.5 (upregulated genes) and ≤-1.5 (downregulated genes) with a *P*-value cutoff of ≤0.05 were considered as significant DEGs. **(B)** qRT-PCR validation of RNA-seq-based expression levels of *OsGW2*. qRT-PCR values are the same as in **Figure 3B**. Data are mean ± SD for two biological replicates. **(C)** Categorization of functionally characterized DEGs with their overlapping functional terms through RICENCODE. Yellow color in the pie chart represents DEGs specifically involved in seed development. Colors in the pie chart represent different plant development and metabolism-related processes.

Thus, we could identify 1,426 DEGs, of which 636 (44.60%) DEGs had higher expression in the *OsGW2*_RNAi_7AP9 line compared with the WT (upregulated genes), while 790 (55.40%) DEGs showed lower expression in the *OsGW2*_RNAi_7AP9 line compared with the WT (downregulated genes) (**Figure 5A**; **Supplementary Table 4**). The expression data obtained through RNA sequencing was validated by comparing the expression of *OsGW2* in *RNAi_7AP9* T_3 seeds generated by qRT-PCR (**Figure 3B**). The expression patterns were found consistent for *OsGW2* in both analyses (**Figure 5B**).

All identified DEGs (1,426 in total) in *OsGW2*_RNAi_7AP9 plants were analyzed by comparison with functionally characterized rice genes from the RICENCODE database. A total of 115 genes were well characterized with specific function (**Supplementary Table 5**). These genes were involved in the regulation of development of the seed, leaf, root, and shoot, apart from the genes related to the cell cycle, hormone signaling, and biotic and abiotic stress tolerance. Overlapping functions of these 115 characterized genes were represented a total of 327 times with respect to different functional terms (leaf, root, shoot, flower, and seed development, etc.). Of these, 55 genes were known to be involved in seed development in rice (**Figure 5C**, **Supplementary Table 6**).

To gain insight into the function of *OsGW2* in seeds, DEGs in *OsGW2*_RNAi_7AP9 T_3 seeds were compared with previously generated DEGs in five seed developmental stages (S1–S5) in the rice genotype IR64 (Sharma et al., 2012). The same analysis has also been successfully done for rice transgenic plants with altered expression of another seed-specific gene *ONAC025*. DEGs common to both transgenic plant and seed are the ones downstream to the transgene (Mathew et al., 2020). This analysis unveiled a correlation of DEGs in the *OsGW2*_RNAi_7AP9

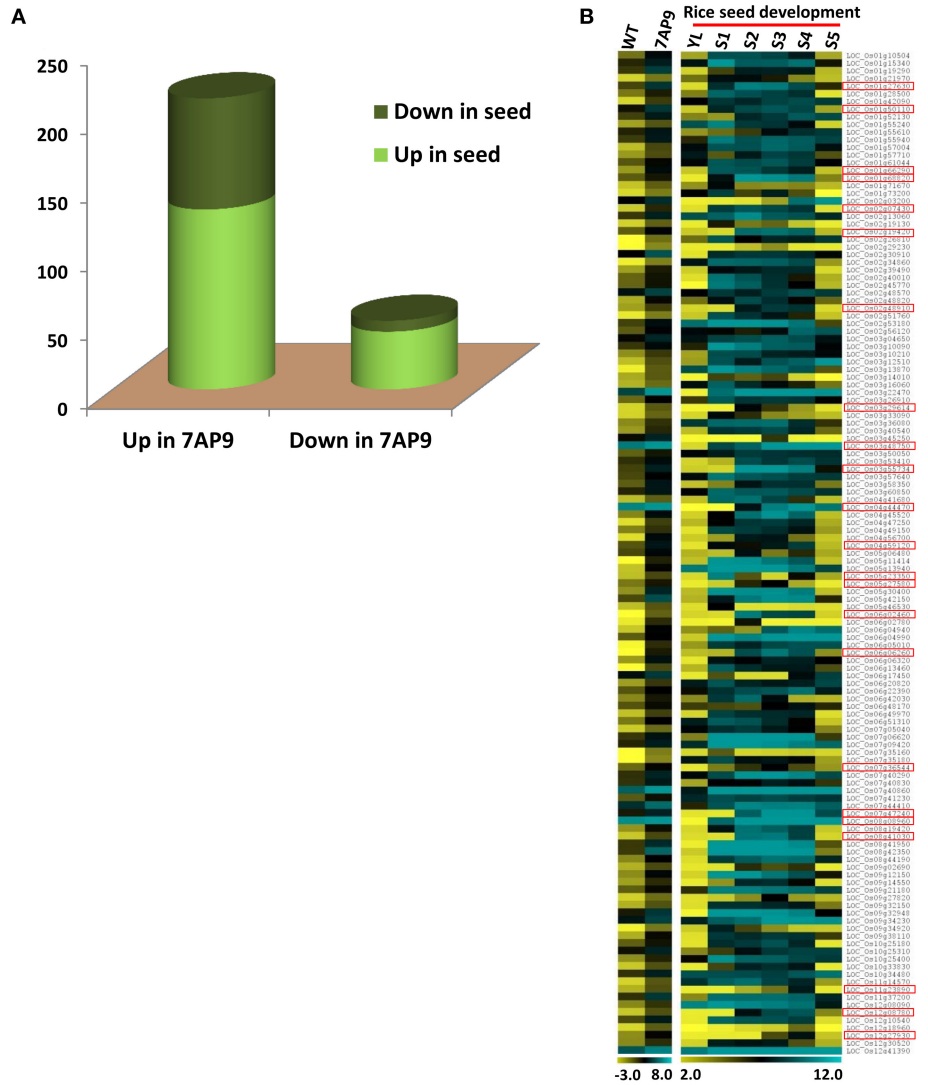


FIGURE 6 | Comparison of transcriptomes of *OsGW2*_RNAi seed and five rice seed development stages. **(A)** Graph represents DEGs in both IR64 seed development stages and in *OsGW2*_RNAi_7AP9 seeds. The first and second bars represent gene numbers up- and downregulated, respectively, in *OsGW2*_RNAi_7AP9 seeds. Gene numbers up- and downregulated in rice seed are marked in light and dark green, respectively. **(B)** Heat map shows $\log_2$ expression values of genes upregulated in both *OsGW2*_RNAi_7AP9 seeds and in five stages of rice seed development. Upregulated seed-specific genes have been demarcated with red boxes in the figure. Color bar below the heat map shows the range of $\log_2$ GCRMA or FPKM values, for microarray and RNA-seq, respectively. YL represents Y-leaf and S1-S5 represent five stages of rice seed development.

transgenic line with seed-related genes. Of the 636 genes upregulated as a result of *OsGW2* knockdown, 262 genes were also differentially expressed in the different seed developmental stages in comparison with the Y-leaf stage (**Figure 6A**). Of these, 131 genes were upregulated in both *OsGW2*_RNAi_7AP9 plants and S1–S5 seed developmental stages, of which 23 were seed-specific (**Figure 6B**). The upregulated seed-specific genes in *OsGW2*_RNAi_7AP9 plants were involved in various seed development and related processes/pathways. One such seed-specific transcription factor, *OsMADS29*, has been shown to affect many processes. Knockdown and overexpression of *OsMADS29* revealed its function in hormone homeostasis, starch biosynthesis in the endosperm, and embryo development (Yang et al., 2012; Nayar et al., 2013). Also, 81 genes showed downregulation in various seed stages compared with the Y-leaf stage (**Supplementary Figure 6A**). Similarly, among the 50 downregulated genes in *OsGW2*_RNAi_7AP9 plants, 42 showed upregulation in the seed stages and eight of them showed downregulation in the seed stages compared with the Y-leaf stage (**Figure 6A**; **Supplementary Figures 6B,C**).

Furthermore, the total DEGs were analyzed for their roles in metabolic pathways by MapMan. There was enrichment of secondary metabolites, starch and sucrose biosynthesis, and lipid metabolism, besides amino acid metabolism and cell

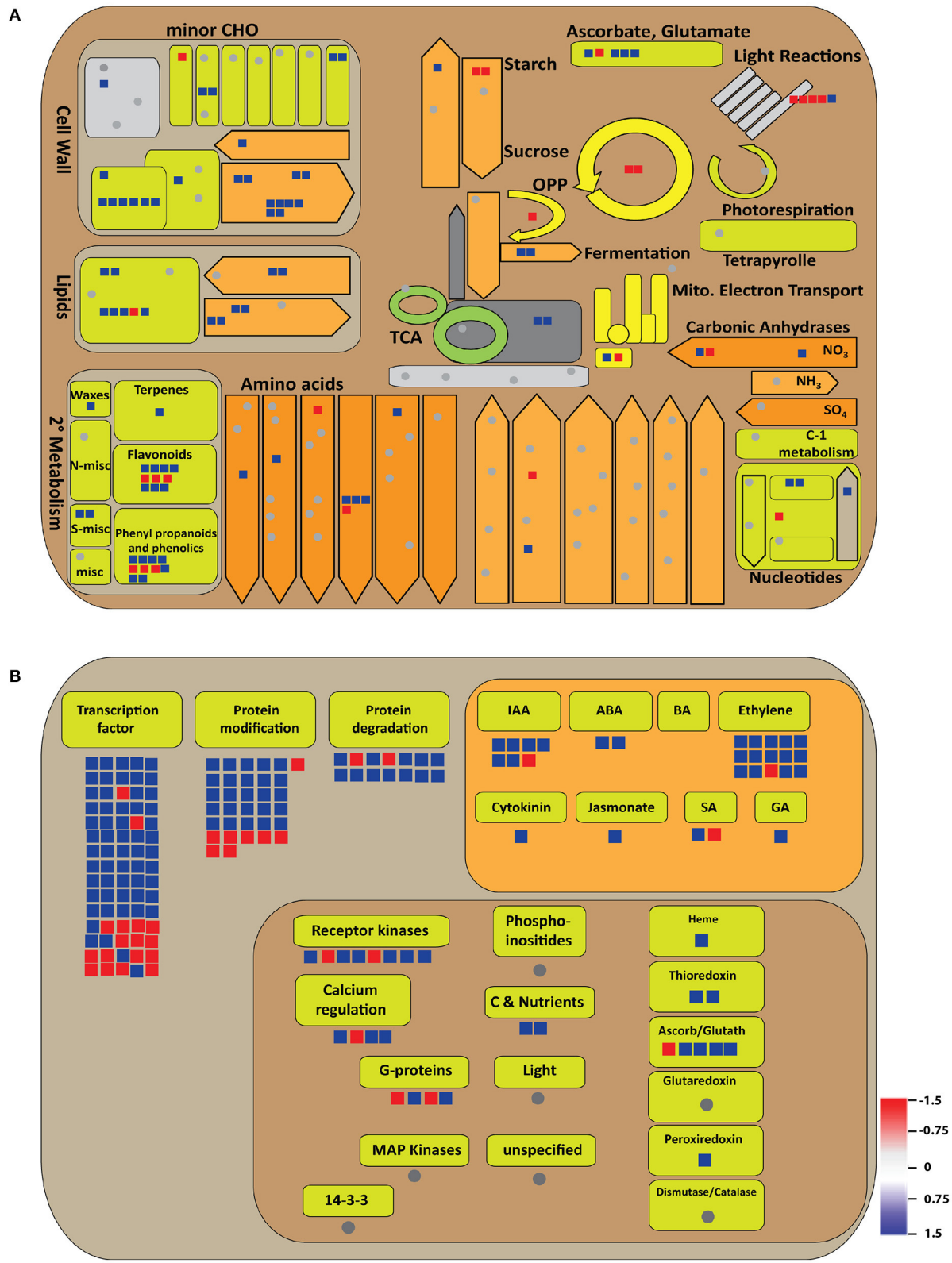


FIGURE 7 | Illustration of metabolic processes involving significant DEGs (log_2 fold change ≥ 1.5 or ≤ -1.5) as visualized by the MapMan tool. **(A)** Metabolism and **(B)** regulation overview have been shown. Blue and red colors signify up- and downregulated genes, respectively, as shown in the color bar in **(B)**.

wall biosynthetic pathways in the *OsGW2*_RNAi_7AP9 line (**Figure 7A**). A parallel level of pathway enrichment was also observed by KEGG. KEGG analysis of all DEGs revealed that photosynthesis, sucrose, and starch metabolism DEGs were mostly upregulated (**Supplementary Figure 4**). We found that genes involved in signaling pathways, including components of ABA, cytokinin, ethylene, GA, IAA, jasmonic acid, and transcription regulatory elements (including TFs), were enriched. Similarly, genes involved in protein modification and protein degradation were found to be enriched in *OsGW2*_RNAi_7AP9 (**Figure 7B**).

DISCUSSION

GRAIN WIDTH 2 codes for an E3 ubiquitin ligase. A natural variation in *GW2* causes truncation of the protein-interacting domain in a long grain *japonica* rice cultivar, WY3, resulting in increased grain width and weight (Song et al., 2007). Recent steps toward the identification of the molecular mechanism of GW2 show that it interacts with EXPLA1, a cell wall-loosening protein (Choi et al., 2018), and chitinase 14 (CHT14) and phosphoglycerate kinase (PGK) which are involved in carbohydrate metabolism (Lee et al., 2018). Oochikara (a natural *japonica* rice gw2 mutant) has loosely packed spherical starch granules (Choi et al., 2018), also observed in knockdown *indica* rice in our experiments (**Figure 4E**). Since *indica* rice lacks the SNP leading to a stop codon (Dixit et al., 2013), the role of such a *GW2* allele has hitherto not been explored in these varieties. This is the first report of functional relevance of *GW2* in *indica* rice, which shows that its downregulation promotes seed width, length and weight, even in *indica* rice, and elucidates the genes downstream to GW2 responsible for this phenotype.

In the present study, four *indica* rice genotypes, with variation in seed size, lacked the natural variation of *japonica* genotypes, though *GW2* expressed in all. In order to test if the GW2 protein has a functional role in *indica* rice, we generated a knockdown of *OsGW2* in *indica* rice. The aim was to examine if a knockdown would generate a phenotype similar to a knockout phenotype observed in *japonica* genotypes. A similar RNAi-based knockdown approach has been adopted to explore the *GW2* function in wheat which resulted in higher grain width and starch content (Sestili et al., 2019). While knockout (KO) is increasingly being used to study gene functions, the methodology has certain limitations, such as the unavailability of PAM sites, off-target effects, and most importantly a "fitness cost" (Ahmad et al., 2018). Partial knockdown of genes allows us to study the function of essential genes, while complete knockout of genes may be lethal to the plants (Boettcher and McManus, 2017; Moreira et al., 2020). Hence, RNAi-mediated knockdown of genes has its own utility. It has also been suggested to first characterize a gene by RNAi methodology and, subsequently, edit it by the CRISPR/Cas technique (Ansari et al., 2020). We did not generate KO in this attempt as the gene expresses ubiquitously, and a KO may hamper plant growth.

Grain weight, a key determinant of grain yield in rice, is positively associated with grain size parameters (grain length, width, and thickness; Shi et al., 2019). Majority of the genes (*GW5*, *qGL3*, *UBP15/LG1*, *GL4*, *SPL16*, etc.) impact grain size by affecting the rate of cell division/proliferation. Few genes (*SPL13*, *KNAT7*) were reported to control grain size by influencing cell expansion (Weng et al., 2008; Qi et al., 2012; Zhang et al., 2012; Wang et al., 2015, 2019; Si et al., 2016; Liu et al., 2017; Wu et al., 2017; Gao et al., 2019; Shi et al., 2019). In the last decade, some genes were reported that regulate grain size by controlling both cell proliferation as well as expansion (*GS2*, *GS5*; Li et al., 2011; Duan et al., 2015; Xu et al., 2015). In our study, SEM analysis of the central portion of the hull of mature grains of the knockdown line showed that an increase in grain size resulted from increased cell number. Previously, in *Arabidopsis*, it was observed that E3 ubiquitin ligase, *DA2*, and *EOD1* also regulate grain size by affecting cell proliferation (Li et al., 2008; Xia et al., 2013). GW2 controls grain size in rice in a similar way as DA2 and EOD1 in *Arabidopsis*. Downregulation of *OsGW2* led to a wider and larger grain. Collectively, *OsGW2* negatively regulates grain size by affecting cell proliferation in the spikelet hull even in *indica* rice.

In dicot plants, endosperm disappears with embryo maturation, whereas in monocots, the endosperm covers the major proportion of the seed after maturation and stores starch in amyloplasts (Olsen et al., 1999). Sugars translocate to the endosperm from vegetative tissues and are converted to amylose and amylopectin at the time of grain filling and stored in the form of starch granules in amyloplasts (Nayar et al., 2013; Sonnewald and Kossmann, 2013). The starch granules in the endosperm of the *OsGW2* knockdown lines were larger and more loosely packed. *OsGW2* knockdown affected the size, shape, and number of starch granules. Transcriptome data suggested that transcript levels of starch metabolism/synthesis-related genes (*OsIAA9*, *OsMADS6*, *OsMADS29*) are upregulated in the silenced *OsGW2*_RNAi plants (**Supplementary Table 6**).

Hormone signal transduction, cell cycle, and sucrose metabolism-related genes are well known to be involved in grain size regulation through various metabolic processes (Li and Li, 2016; Lee et al., 2017). In our transcriptome study, KEGG, MapMan, and GO analyses revealed that DEGs that were significantly enriched pertained to biological processes (cellular process, primary and secondary metabolic processes), metabolism (carbohydrate, lipid, amino acid, and nucleotide), hormone signal transduction, starch metabolism, stress response, photosynthesis, etc. (**Figure 7**; **Supplementary Figures 4**, **5**). Plant hormones are important grain size regulators. Auxins, cytokinins, and brassinosteroids play an important role in determining rice grain size. Auxin-responsive genes (*OsARF4*, *Gnp4*/*LAX2*, *BG1*, *qTGW3*, *TGW6*), a cytokinin-related gene (*OsCKX2*/*Gn1a*), and brassinosteroid-related genes (*GS5*, *qGL3*, *GS2*/*OsGRF4*) regulate grain size by being involved in signaling, biosynthesis, and transport (Azizi et al., 2019; Li and Li, 2019). In our transcriptome data, majority of auxin-responsive genes, including *OsMGH3*, *OsMADS29*, *OsRAA1*, and *OsIAA9*, showed upregulation in *OsGW2* knockdown plants (**Supplementary Table 6**). In addition to this, cytokinin- and brassinosteroid-related genes were also upregulated in the *OsGW2*_RNAi_7AP9 line. These results provide evidence for hormone-responsive genes being important regulators of grain size in rice. Several genes (*GS3*, *GS5*, *GW5*, *qGL3*, *TGW6*, *GIF1*, *GW8*, etc.) are preferentially or specifically expressed in seed

tissues and regulate grain size in rice by being part of various pathways like the ubiquitin–proteasome pathway, G-protein signaling, plant hormone signaling, mitogen-activated protein kinase (MAPK) signaling, and some transcription factors (Zuo and Li, 2014). Upregulation of starch metabolism and cell cycle-related genes (*OsMADS6*, *OsMADS29*, *OsYUC9*, and *OsKRP3*) in *OsGW2* knockdown plants also indicates the pathways through which GW2 regulates seed development in rice. The DEGs generated by transcriptome analysis will in all probability be downstream of OsGW2 substrates, which regulate grain size in rice. We can speculate that upregulated genes in *OsGW2*_RNAi plants that are also seed-specific might be involved in direct grain size regulation.

CONCLUSION

Here, we described some important facets of *OsGW2*-mediated regulation of grain width and weight in *indica* rice. Both cell division and expansion in the hull of spikelet are responsible for higher grain width as well as weight in *OsGW2* knockdown plants in *indica* rice. Additionally, *OsGW2* also affects starch granule morphology and their packing. We further conducted RNA sequencing to elucidate the active molecular mechanisms in *OsGW2* knockdown plants and identified enriched DEGs involved in hormone signal transduction and starch metabolism. Transcriptome analyses also suggested that *OsGW2* plays a crucial role in the regulation of cellular and metabolic processes relevant to grain size regulation.

DATA AVAILABILITY STATEMENT

RNA-seq was performed for one wild type line and one transgenic line (RNAi_7AP9), in biological replicates, using Illumina® HiseqTM. RNA sample preparation kit v.2 (Illumina; USA). The sequences generated were submitted to NCBI SRA database (http://www.ncbi.nlm.nih.gov/sra/PRJNA659415) with SRA accession number PRJNA659415. Biosample accession numbers for individual samples are SAMN15905438_WT_Rep1, SAMN15905439_WT_Rep2, SAMN15905440_7AP9_Rep1, SAMN15905441_7AP9_Rep2.

AUTHOR CONTRIBUTIONS

GP made the constructs, and RR and GP generated the rice transgenic plants. AV conducted all the experiments and analyzed the data. AKT and PA supervised the study and guided the experimental design and data analysis. AV and PA wrote the manuscript. All authors read/corrected the manuscript and approved it.

FUNDING

This work was supported by research grant no. BT/AB/NIPGR/SEED BIOLOGY/2012 to PA and BT/PR12394/AGIII/103/891/2014 to AKT from the Department of Biotechnology, Ministry of Science and Technology, India, and NIPGR core grant to PA; junior and senior research fellowships from the Department of Biotechnology, India, to AV; and junior and senior research fellowships from the Council of Scientific and Industrial Research (CSIR) to RR.

ACKNOWLEDGMENTS

The authors acknowledge the support provided by the central instrumentation, microscopy, and plant growth facilities of NIPGR and the DBT e-library consortium for providing online access to the articles.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.600378/full#supplementary-material

Supplementary Figure 1 | Comparison of grain parameters in five different rice genotypes. **(A)** Grain width, **(B)** grain length, and **(C)** grain weight. Error bar represents mean ± SD; $n = 100$. Grain width and length were measured with WINSEEDLE^{TM}. Nipponbare, LGR, Sonasal, IR64, and PB1 represent different rice genotypes.

Supplementary Figure 2 | Screening of *OsGW2*_RNAi transgenic plants and effect of GW2 on grain size parameters. **(A)** Screening of *OsGW2*_RNAi T0 transgenic plants by PCR amplification of hygromycin resistance gene (*Hygromycin phosphotransferase II*). L represents Gene Ruler 100 bp DNA Ladder Mix. 1–35 are different transgenic lines in T0 generation. Ten positive T0 plants marked in red (2, 4, 7, 9, 13, 15, 17, 21, 23, 24) were grown further and analyzed. Effect of *GW2* on grain size parameters; grain length **(B)** and width **(C)** of *OsGW2*_RNAi T1 seeds and; grain length **(D)** and width **(E)** of *GW2* RNAi T2 seeds. In T1 generation four lines (RNAi_2A, 7A, 9B, and 24P2) were selected out of ten lines which showed 3:1 segregation ratio. Bar = 1 cm; WT represents wild-type PB-1 grains; grain length and width measured with 30 cm scale by aligning 10 seeds of each line length wise and breadth wise. Red line in **(B–E)** represents WT measurements for comparison.

Supplementary Figure 3 | Screening of *OsGW2_RNAi* plants in T2 generation by PCR amplification of hygromycin resistance gene (*Hygromycin phosphotransferase II*). L represents Gene Ruler 1 kb DNA Ladder Mix. RNAi (2AP3, 2AP9, 7AP8, 7AP9, 9BP3, 9BP4, 24P2P5, 24P2P6) are different lines in T2 generation. N and P represent negative and positive control, respectively.

Supplementary Figure 4 | Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of DEGs. KEGG pathway analysis of **(A)** up-regulated and **(B)** down-regulated genes. The KEGG pathways with FDR value ≤0.05 were significantly enriched. Numbers represent DEGs enriched in each pathway.

Supplementary Figure 5 | GO term enrichment analysis through AgriGO of DEGs in *OsGW2*_RNAi transgenic plants. The terms enriched under **(A)** biological process, **(B)** cellular component, and **(C)** molecular function have been shown. Each box contains GO terms, GO IDs, and statistical information. GO terms with FDR value ≤0.05 show significant enrichment. Yellow, orange, and red colors represent significant GO terms and white color represents non-significant terms as shown in color legend in **(B)**. Arrows represent relationship between GO terms as shown in **(B)**.

Supplementary Figure 6 | Heat maps showing comparison of *OsGW2*_RNAi_7AP9 seeds RNAseq data with rice seed development microarray data. **(A)** DEGs down-regulated in IR64 seeds (S1–S5 stages) and up-regulated in *OsGW2*_RNAi_7AP9 transgenic seeds. **(B)** DEGs up-regulated in IR64 seeds and down-regulated in *OsGW2*_RNAi_7AP9 transgenic seeds. **(C)** DEGs down-regulated in both IR64 seeds as well as *OsGW2*_RNAi_7AP9 transgenic seeds. S1–S5 represent different seed developmental stages, YL represents Y leaf. WT and 7AP9 represent wild-type and *OsGW2*_RNAi transgenic line, respectively. The DEGs with $\log_2$ fold change ≥1.5 (up-regulated genes) and ≤−1.5 (down-regulated genes) with *p*-value cut off of ≤0.05 were considered as significant DEGs.

REFERENCES

Agarwal, P., Arora, R., Ray, S., Singh, A. K., Singh, V. P., Takatsuji, H., et al. (2007). Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. *Plant Mol. Biol.* 65, 467–485. doi: 10.1007/s11103-007-9199-y

Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33, 189–202. doi: 10.1002/bies.201000107

Ahmad, S., Wei, X., Sheng, Z., Hu, P., and Tang, S. (2018). CRISPR/Cas9 for development of disease resistance in plants: recent progress, limitations, and future prospects. *Brief. Funct. Genomics* 19, 26–39. doi: 10.1093/bfgp/elz041

Ansari, W. A., Chandanshive, S. U., Bhatt, V., Nadaf, A. B., Vats, S., Katara, J. L., et al. (2020). Genome editing in cereals: approaches, applications, and challenges. *Int. J. Mol. Sci.* 21, 1–32. doi: 10.3390/ijms21114040

Azizi, P., Osman, M., Hanafi, M. M., Sahebi, M., Rafii, M. Y., Taheri, S., et al. (2019). Molecular insights into the regulation of rice kernel elongation. *Crit. Rev. Biotechnol.* 39, 904–923. doi: 10.1080/07388551.2019.1632257

Boettcher, M., and McManus, M. T. (2017). Choosing the right tool for the job: RNAi, TALEN or CRISPR. *Physiol. Behav.* 176, 139–148. doi: 10.1016/j.molcel.2015.04.028

Chakrabarty, R., Banerjee, R., Chung, S., Farman, M., Citovsky, V., Hogenhout, S. A., et al. (2007). pSITE vectors for stable integration or transient expression of autofluorescent protein fusions in plants: probing *Nicotiana benthamiana*—virus interactions. *Mol. Plant Microbe Interact.* 20, 740–750. doi: 10.1094/MPMI-20-7-0740

Choi, B. S., Kim, Y. J., Markkandan, K., Koo, Y. J., Song, J. T., and Seo, H. S. (2018). GW2 functions as an E3 ubiquitin ligase for rice expansin-like 1. *Int. J. Mol. Sci.* 19, 1–12. doi: 10.3390/ijms19071904

Das, S., Parida, S. K., Agarwal, P., and Tyagi, A. K. (2019). Transcription factor OsNF-YB9 regulates reproductive growth and development in rice. *Planta* 250, 1849–1865. doi: 10.1007/s00425-019-03268-2

Ding, W., Tong, H., Zheng, W., Ye, J., Pan, Z., Zhang, B., et al. (2017). Isolation, characterization, and transcriptome analysis of a cytokinin receptor mutant osckt1 in rice. *Front. Plant Sci.* 8:88. doi: 10.3389/fpls.2017.00088

Dixit, N., Dokku, P., Amitha Mithra, S. V., Parida, S. K., Singh, A. K., Singh, N. K., et al. (2013). Haplotype structure in grain weight gene GW2 and its association with grain characteristics in rice. *Euphytica* 192, 55–61. doi: 10.1007/s10681-012-0852-4

Dong, X., Zhang, D., Liu, J., Liu, Q. Q., Liu, H., Tian, L., et al. (2015). Plastidial disproportionating enzyme participates in starch synthesis in rice endosperm by transferring maltooligosyl groups from amylose and amylopectin to amylopectin. *Plant Physiol.* 169, 2496–2512. doi: 10.1104/pp.15.01411

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64–W70. doi: 10.1093/nar/gkq310

Duan, P., Ni, S., Wang, J., Zhang, B., Xu, R., Wang, Y., et al. (2015). Regulation of OsGRF4 by OsmiR396 controls grain size and yield in rice. *Nat. Plants* 2:15203. doi: 10.1038/nplants.2015.203

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1

Gao, X., Zhang, J.-Q., Zhang, X., Zhou, J., Jiang, Z., Huang, P., et al. (2019). Rice qGL3/OsPPKL1 functions with the GSK3/SHAGGY-like kinase OsGSK3 to modulate brassinosteroid signaling. *Plant Cell* 31, 1077–1093. doi: 10.1105/tpc.18.00836

Huang, K., Wang, D., Duan, P., Zhang, B., Xu, R., Li, N., et al. (2017). WIDE AND THICK GRAIN 1, which encodes an otubain-like protease with deubiquitination activity, influences grain size, and shape in rice. *Plant J.* 91, 849–860. doi: 10.1111/tpj.13613

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591

Jung, I. J., Ahn, J. W., Jung, S., Hwang, J. E., Hong, M. J., Choi, H. I. I., et al. (2019). Overexpression of rice jacalin-related mannose-binding lectin (OsJAC1) enhances resistance to ionizing radiation in Arabidopsis. *BMC Plant Biol.* 19:561. doi: 10.1186/s12870-019-2056-8

Kubo, M., and Purevdorj, M. (2004). The future of rice production and consumption. *J. Food Distrib. Res.* 35, 128–142. doi: 10.22004/ag.econ.27145

Lee, K. H., Park, S. W., Kim, Y. J., Koo, Y. J., Song, J. T., and Seo, H. S. (2018). Grain width 2 (GW2) and its interacting proteins regulate seed development in rice (*Oryza sativa* L.). *Bot. Stud.* 59:23. doi: 10.1186/s40529-018-0240-z

Lee, Y., Choi, M. S., Lee, G., Jang, S., Yoon, M. R., Kim, B., et al. (2017). Sugary endosperm is modulated by starch branching enzyme IIa in rice (*Oryza sativa* L.). *Rice* 10:33. doi: 10.1186/s12284-017-0172-3

Li, N., and Li, Y. (2014). Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* 5:332. doi: 10.3389/fpls.2014.00332

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008

Li, N., and Li, Y. (2019). Molecular networks of seed size control in plants. *J. Exp. Bot.* 66, 1087–1097. doi: 10.1093/jxb/eru549

Li, N., Xu, R., Duan, P., and Li, Y. (2018). Control of grain size in rice. *Plant Reprod.* 31, 1–15. doi: 10.1007/s00497-018-0333-6

Li, Q., Li, L., Yang, X., Warburton, M. L., Bai, G., Dai, J., et al. (2010). Relationship, evolutionary fate, and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. *BMC Plant Biol.* 10:143. doi: 10.1186/1471-2229-10-143

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43, 1266–1269. doi: 10.1038/ng.977

Li, Y., Zheng, L., Corke, F., Smith, C., and Bevan, M. W. (2008). Control of final seed and organ size by the DA1 gene family in *Arabidopsis thaliana*. *Genes Dev.* 22, 1331–1336. doi: 10.1101/gad.463608

Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., et al. (2017). GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants* 3:17043. doi: 10.1038/nplants.2017.43

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

Mathew, I. E., Das, S., Mahto, A., and Agarwal, P. (2016). Three rice NAC transcription factors heteromerize and are associated with seed size. *Front. Plant Sci.* 7:1638. doi: 10.3389/fpls.2016.01638

Mathew, I. E., Priyadarshini, R., Mahto, A., Jaiswal, P., Parida, S. K., and Agarwal, P. (2020). SUPER STARCHY1/ONAC025 participates in rice grain filling. *Plant Direct* 4, 1–25. doi: 10.1002/pld3.249

Miki, D., and Shimamoto, K. (2004). Simple RNAi vectors for stable and transient suppression of gene function in rice. *Plant Cell Physiol.* 45, 490–495. doi: 10.1093/pcp/pch048

Moon, J., Parry, G., and Estelle, M. (2004). The ubiquitin-proteasome pathway and plant development. *Plant Cell* 16, 3181–3195. doi: 10.1105/tpc.104.161220

Moreira, D., Pereira, A. M., Lopes, A. L., and Coimbra, S. (2020). The best CRISPR/Cas9 versus RNA interference approaches for Arabinogalactan proteins' study. *Mol. Biol. Rep.* 47, 2315–2325. doi: 10.1007/s11033-020-05258-0

Nayar, S., Sharma, R., Tyagi, A. K., and Kapoor, S. (2013). Functional delineation of rice MADS29 reveals its role in embryo and endosperm development by affecting hormone homeostasis. *J. Exp. Bot.* 64, 4239–4253. doi: 10.1093/jxb/ert231

Olsen, O.-A., Linnestad, C., and Nichols, S. E. (1999). Developmental biology of the cereal endosperm. *Trends Plant Sci.* 4, 253–257. doi: 10.1016/S1360-1385(99)01431-4

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2006). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 35, D883–D887. doi: 10.1093/nar/gkl976

Qi, P., Lin, Y.-S., Song, X.-J., Shen, J.-B., Huang, W., Shan, J.-X., et al. (2012). The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res.* 22, 1666–1680. doi: 10.1038/cr.2012.151

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., and Khvorova, A. (2004). Rational siRNA design for RNA interference. *Nat. Biotechnol.* 22, 326–330. doi: 10.1038/nbt936

Rosegrant, M. W., and Cline, S. A. (2003). Global food security: challenges and policies. *Science* 302, 1917–1919. doi: 10.1126/science.1092958

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374–378. doi: 10.2144/03342mt01

Sasaki, T. (2008). From the editor's desk. *Rice* 1, 1–2. doi: 10.1007/s12284-008-9010-y

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Sestili, F., Pagliarello, R., Zega, A., Saletti, R., Pucci, A., Botticella, E., et al. (2019). Enhancing grain size in durum wheat using RNAi to knockdown GW2 genes. *Theor. Appl. Genet.* 132, 419–429. doi: 10.1007/s00122-018-3229-9

Sharma, R., Agarwal, P., Ray, S., Deveshwar, P., Sharma, P., Sharma, N., et al. (2012). Expression dynamics of metabolic and regulatory components across stages of panicle and seed development in *indica* rice. *Funct. Integr. Genomics* 12, 229–248. doi: 10.1007/s10142-012-0274-3

Shi, C., Ren, Y., Liu, L., Wang, F., Zhang, H., Tian, P., et al. (2019). Ubiquitin specific protease 15 has an important role in regulating grain width and size in rice. *Plant Physiol.* 180, 381–391. doi: 10.1104/pp.19.00065

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Singh, G., Kumar, S., and Singh, P. (2003). A quick method to isolate RNA from wheat and other carbohydrate-rich seeds. *Plant Mol. Biol. Report.* 21, 93–93. doi: 10.1007/BF02773401

Smalle, J., and Vierstra, R. D. (2004). The ubiquitin 26S proteasome proteolytic pathway. *Annu. Rev. Plant Biol.* 55, 555–590. doi: 10.1146/annurev.arplant.55.031903.141801

Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z., and Lin, H.-X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014

Sonnewald, U., and Kossmann, J. (2013). Starches-from current models to genetic engineering. *Plant Biotechnol. J.* 11, 223–232. doi: 10.1111/pbi.12029

Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S., et al. (2006). Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. *Plant J.* 47, 969–976. doi: 10.1111/j.1365-313X.2006.02836.x

Vierstra, R. D. (2009). The ubiquitin−26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* 10, 385–397. doi: 10.1038/nrm2688

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47, 949–954. doi: 10.1038/ng.3352

Wang, S., Yang, H., Mei, J., Liu, X., Wen, Z., Zhang, L., et al. (2019). Rice homeobox protein KNAT7 integrates the pathways regulating cell expansion and wall stiffness. *Plant Physiol.* 181, 669–682. doi: 10.1104/pp.19.00639

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* 18, 1199–1209. doi: 10.1038/cr.2008.307

Wu, W., Liu, X., Wang, M., Meyer, R. S., Luo, X., Ndjiondjop, M.-N., et al. (2017). A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. *Nat. Plants* 3:17064. doi: 10.1038/nplants.2017.64

Xia, T., Li, N., Dumenil, J., Li, J., Kamenski, A., Bevan, M. W., et al. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. *Plant Cell* 25, 3347–3359. doi: 10.1105/tpc.113.115063

Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442. doi: 10.1146/annurev-arplant-042809-112209

Xu, C., Liu, Y., Li, Y., Xu, X., Xu, C., Li, X., et al. (2015). Differential expression of GS5 regulates grain size in rice. *J. Exp. Bot.* 66, 2611–2623. doi: 10.1093/jxb/erv058

Yang, X., Wu, F., Lin, X., Du, X., Chong, K., Gramzow, L., et al. (2012). Live and let die–The Bsister MADS-Box gene OsMADS29 controls the degeneration of cells in maternal tissues during seed development of rice (*Oryza sativa*). *PLoS ONE* 7:e51435. doi: 10.1371/journal.pone.0051435

Ying, J. Z., Gao, J. P., Shan, J. X., Zhu, M. Z., Shi, M., and Lin, H. X. (2012). Dissecting the genetic basis of extremely large grain shape in rice cultivar "JZ1560." *J. Genet. Genomics* 39, 325–333. doi: 10.1016/j.jgg.2012.03.001

Yuenyong, W., Chinpongpanich, A., Comai, L., Chadchawan, S., and Buaboocha, T. (2018). Downstream components of the calmodulin signaling pathway in the rice salt stress response revealed by transcriptome profiling and target identification. *BMC Plant Biol.* 18:335. doi: 10.1186/s12870-018-1538-4

Zha, K., Xie, H., Ge, M., Wang, Z., Wang, Y., Si, W., et al. (2019). Expression of maize MADS transcription factor ZmES22 negatively modulates starch accumulation in rice endosperm. *Int. J. Mol. Sci.* 20:483. doi: 10.3390/ijms20030483

Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21534–21539. doi: 10.1073/pnas.1219776110

Zhao, M., Zhao, M., Gu, S., Sun, J., Ma, Z., Wang, L., et al. (2019). DEP1 is involved in regulating the carbon–nitrogen metabolic balance to affect grain yield and quality in rice (*Oryza sativa* L.). *PLoS ONE* 14:e0213504. doi: 10.1371/journal.pone.0213504

Zuo, J., and Li, J. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* 48, 99–118. doi: 10.1146/annurev-genet-120213-092138

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Verma, Prakash, Ranjan, Tyagi and Agarwal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 January 2021
doi: 10.3389/fgene.2020.613217

New QTLs for Spot Blotch Disease Resistance in Wheat (*Triticum aestivum* L.) Using Genome-Wide Association Mapping

Vipin Tomar[1,2], Daljit Singh[3], Guriqbal Singh Dhillon[4], Ravi Prakash Singh[5], Jesse Poland[6], Arun Kumar Joshi[1,7], Pawan Kumar Singh[5], Pradeep Kumar Bhati[1], Suneel Kumar[1], Mokhlesur Rahman[8], Budhi Sagar Tiwari[2] and Uttam Kumar[1,5]**

[1] Borlaug Institute for South Asia, Ludhiana, India, [2] Institute of Advanced Research, Gandhinagar, Gujarat, India, [3] The Climate Corporation, Bayer Crop Science, Creve Coeur, MO, United States, [4] Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, India, [5] Global Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico, [6] Department of Plant Pathology, Kansas State University, Manhattan, KS, United States, [7] International Maize and Wheat Improvement Centre, New Delhi, India, [8] Wheat Research Center, Regional Agricultural Research Station, Bangladesh Agricultural Research Institute, Jamalpur, Bangladesh

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University,
United States

Reviewed by:
Rohit Kumar,
Clemson University, United States
Shyam Solanki,
Washington State University,
United States
Jason Fiedler,
United States Department
of Agriculture (USDA), United States

***Correspondence:**
Uttam Kumar
u.kumar@cgiar.org
Vipin Tomar
viomics@gmail.com

Specialty section:
This article was submitted to
Plant Genomics,
a section of the journal
Frontiers in Genetics

Received: 01 October 2020
Accepted: 10 December 2020
Published: 14 January 2021

Citation:
Tomar V, Singh D, Dhillon GS,
Singh RP, Poland J, Joshi AK,
Singh PK, Bhati PK, Kumar S,
Rahman M, Tiwari BS and Kumar U
(2021) New QTLs for Spot Blotch
Disease Resistance in Wheat (Triticum
aestivum L.) Using Genome-Wide
Association Mapping.
Front. Genet. 11:613217.
doi: 10.3389/fgene.2020.613217

Spot blotch disease caused by *Bipolaris sorokiniana* is a major constraint for wheat production in tropics and subtropics. The introgression of spot blotch resistance alleles to the disease susceptible lines is critical to securing the wheat production in these regions. Although genome-wide association studies (GWASs) for spot blotch were attempted earlier, the present study focused on identifying new quantitative trait loci (QTLs) for spot blotch under natural disease pressure in diverse field conditions. A total of 139 advanced spring wheat lines were evaluated in three environments (three years and two locations) in India and Bangladesh. The GWAS using 14,063 polymorphic genotyping-by-sequencing (GBS) markers identified eight QTLs associated with spot blotch disease resistance belonging to eight chromosomes across the wheat genome. Here, we report the identified marker–trait associations (MTAs), along with the allele effects associated with the disease. The functional annotation of the significant markers identified NBS-LRR, MADS-box transcription factor, and 34 other plant-related protein families across multiple chromosomal regions. The results indicate four promising new QTLs on chromosomes 1A (497.2 Mb), 1D (89.84 Mb), 2B (421.92 Mb), and 6D (6.84 Mb) associated with several disease resistance protein families. These results provide insights into new genomic regions associated with spot blotch disease, and with additional validation, could be utilized in disease resistance breeding efforts in wheat development.

Keywords: GWAS, GBS, spot blotch, QTLs, BLUPs, wheat

INTRODUCTION

Wheat (*Triticum aestivum* L.) is the primary staple food for more than 35% of the world's population (FAO, 2018). The pace of wheat improvement must accelerate to meet the projected global food demand by 2050. The Green Revolution played a crucial role in India, Pakistan, Nepal, and Bangladesh, ensuring food security in this densely populated region of the world

(Joshi et al., 2007b). However, wheat production faces multiple threats via rapidly evolving pathogen variants, pests, and increased climate variability, which considerably jeopardizes crop productivity (Sharma et al., 2007b; Friesen et al., 2008; Gurung et al., 2009, 2012). Breeding wheat for climatic resilience and disease resistance combined with good agronomic value can potentially improve wheat productivity to meet future food demands (Mondal et al., 2016).

Spot blotch caused by *Bipolaris sorokiniana* is a major constraint in wheat production in tropics and subtropics (Dubin and Van Ginkel, 1991; van Ginkel and Rajaram, 1998). The pathogen has a worldwide dispersal, but it is predominantly aggressive under warm, high relative humidity, and high temperature associated conditions with imbalanced soil fertility. Major yield losses are observed in the fields with lower inputs and under late-sown conditions (Joshi et al., 2007b). *B. sorokiniana* acts as a causal agent for numerous diseases like head blight, seedling blight, foliar blight/spot blotch, common root rot, and black point of wheat, barley, other small cereal grains and grasses (Lennard, 1984). Therefore, spot blotch of wheat is considered as one of the most crucial diseases caused by this pathogen in the mega-environments characterized by temperature above 17°C and high humidity (van Ginkel and Rajaram, 1998).

It is often difficult to achieve the desired level of host resistance to several diseases through conventional breeding. In the case of wheat disease, resistance is inherited both qualitatively and quantitatively (Kumar et al., 2009, 2010; Marone et al., 2013; French et al., 2016). The genetic basis of spot blotch resistance has earlier been documented as eight major quantitative trait loci (QTLs), namely, *QSb.bhu-2A*, *QSb.bhu-2B*, *QSb.bhu-2D*, *QSb.bhu-3B*, *QSb.bhu-5B*, *QSb.bhu-6D*, *QSb.bhu-7B*, and *QSb.bhu-7D* (Kumar et al., 2009, 2010). Sharma et al. (2007a) reported three microsatellite markers (*Xgwm67*, *Xgwm570*, and *Xgwm469*) linked with spot blotch resistance. The QTL *QSb.bhu-5B*, *QSb.bhu-7D*, and *QSb.bhu-3B* have been designated as genes *Sb1*, *Sb2*, and *Sb3*, respectively, in further studies (Lillemo et al., 2013; Kumar et al., 2015; Lu et al., 2016). *Lr34* and *Lr46*, the two broadly used genes conferring leaf rust resistance, have also been reported to contribute to spot blotch resistance. While the *Lr34* gene has been reported to be the major locus on chromosome 7D explaining up to 55% phenotypic variation for spot blotch disease resistance, this locus was given the gene designation *Sb1* (Lillemo et al., 2013). Nevertheless, in the past few years, several QTLs and genetic markers for spot blotch resistance have been identified in multiple studies in wheat (Gurung et al., 2014; Zhu et al., 2014; Singh et al., 2018).

Due to evolutionary changes in pathogen populations, resistance genes may lose their effectiveness over time. Given these challenges, identification and mapping of novel resistance genes would aid breeding for disease resistance in wheat. One of the approaches is to identify spot blotch resistance QTLs through association mapping. This approach leverages historical recombination and generally has better mapping resolution compared to biparental mapping (Zhu et al., 2008). The genome-wide association study (GWAS) using genotyping by sequencing (GBS) to identify QTLs for traits of interest provides advantage through better genome coverage compared to conventional marker systems like SSRs, AFLP, and CAPs. The GBS utilizes the advantage of high-throughput genotyping assays with relatively low data costs, which are much higher in genome sequencing and re-sequencing techniques (Elshire et al., 2011). A fundamental approach in GWAS is to have enough genome coverage so that functional alleles will be in linkage disequilibrium (LD) with at least one marker (Myles et al., 2009). The association studies for disease resistance, including spot blotch, have been reported in some of earlier studies (Maccaferri et al., 2010, 2015; Yu et al., 2011; Zegeye et al., 2014; Ayana et al., 2018).

There are limited reports where the same set of genotypes is exposed to natural disease pressure to identify genomic regions underpinning wheat spot blotch resistance. Thus, the main objective of this study was to establish marker–trait associations (MTAs) for spot blotch resistance using GBS markers in spring wheat specific to the South Asian regions, *viz*., India and Bangladesh, the well-known hot spots for this disease. We also aimed to identify novel QTLs and validated known genomic loci conferring spot blotch resistance in wheat.

MATERIALS AND METHODS

Plant Material and Field Layout

The population was a collection of 139 advanced breeding lines of spring wheat (**Supplementary Table S1**) derived from the crosses where elite high-yield breeding lines were used as parents. The advanced lines were sent to South Asia as part of the CIMMYT wheat breeding program aiming to develop high-yielding varieties suitable for South Asia. They were carefully assembled to avoid the confounding effects of phenology after multiyear and multilocation trials. The lines were evaluated in two replicates at two field locations: Bangladesh Agricultural Research Institute (BARI), Jamalpur, Bangladesh (24°22′07.7″N, 88°39′42.0″E), during 2017 and Borlaug Institute for South Asia (BISA), Pusa (25°57′22.8″N, 85°40′13.1″E), in north India during 2019 and 2020 seasons in a randomized block design. For convenience, the different location–season combinations were termed as Env1 (Pusa19), Env2 (Pusa20), and Env3 (BARI17). Pusa, India, and Jamalpur, Bangladesh, are known hot spots for spot blotch disease (Sharma et al., 2007b). The trials were timely sown with full irrigation applied through gravity flood irrigation. The spreader rows of susceptible variety Sonalika were also planted for creating epiphytotic disease conditions. Besides, four auxiliary gravity flood irrigations were also given at regular intervals. All agronomic practices like fertilization and weeding were performed as recommended for each location.

Screening for Spot Blotch Disease Resistance

The material was evaluated under natural infection conditions in the field. Spot blotch response was evaluated thrice during the mid to advanced phases of disease development, i.e., between heading (Growth Stage 50 or GS50 on Zadoks scale) and grain filling stage (GS80) (Zadoks et al., 1974). The disease severity (SEV) was recorded visually on a 0–100 scale where 0 is complete resistance and 100 is completely susceptible.

Genotyping

Seeds of all lines were obtained from the CIMMYT genetic resource program, and genomic DNA was extracted from five bulked leaves using a modified CTAB procedure as described in Dreisigacker et al. (2016) in CIMMYT, Mexico. The DNA samples were sent to Kansas State University, United States, for GBS, described by Poland et al. (2012) and sequenced with Illumina HISeq2500. GBS-SNP markers were called with TASSEL *v*5.2 pipeline GBSv2 (Bradbury et al., 2007) and aligned to the reference Chinese Spring Wheat Assembly (RefSeq v1.0). The following SNP filtering criteria were applied on raw SNP calls: less than 30% missing markers, minimum 5% minor allele frequency (MAF), and less than 20% heterozygosity. The filtering step yielded 14,063 markers, and the remaining missing values were imputed using Beagle v4.1 (Browning and Browning, 2016).

Statistical Analysis

The experimental design in each environment was a randomized complete block design with two replications per environment/location. The best linear unbiased prediction (BLUP) values were obtained through META-R v6.03 (Alvarado et al., 2020), developed by CIMMYT, Mexico, using the following formula:

$$Y_{ik} = \mu + Rep_i + Gen_k + \epsilon_{ik} \text{ (within the environment)}$$

$$Y_{ijk} = \mu + Rep_i\left(Env_j\right) + Env_j \times Gen_k + Gen_k + Env_j + \epsilon_{ijk}$$

(across environments)

where Y_{ik} is the trait of interest, μ is the mean effect, Rep_i is the effect of the *i*th replicate, Gen_k is the effect of the *k*th genotype, ϵ_{ik} is the error associated with the *i*th replication, and the *k*th genotype, which is assumed to be normally and independently distributed, with mean zero and homoscedastic variance σ^2. For across environments, Y_{ijk} is the trait response, Env_j is the *j*th environment, $Rep_i\left(Env_j\right)$ is the effect of *i*th replication in the *j*th environment, and $Env_j \times Gen_k$ is the environment and genotype interaction. The resulting analysis produced the adjusted trait phenotypic values in the form of BLUP within and across environments. The BLUP model considers genotypes as random effects, minimizing the effect of screening time and other environmental effects on the spot blotch severity.

Besides, the components of the phenotypic variance of a given trait at an individual environment were also extracted to calculate broad-sense heritability using the formula as:

$$h_2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/\text{n reps}} \text{ (within the environment)}$$

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2/\text{ n env} + \sigma_e^2/(\text{n reps} \times \text{ n env})}$$

(across environments)

where σ_g^2 and σ_e^2 are the genotype and error variance components, respectively, σ_{ge}^2 is the genotype by environment interaction variance, *n env* is the number of environments, and *n reps* is the number of replicates. All effects are considered random for calculating the BLUP and broad-sense heritability. The BLUP phenotypic distributions of disease scores at each environment were plotted to check normality assumptions.

Principal Component Analysis and Linkage Disequilibrium

Principal component analysis (PCA) was performed using 14,063 SNPs and 139 genotypes in FarmCPU (Liu et al., 2016). The first two principal components were drawn to show the clustering among genotypes. The population structure (PC) and LD were estimated in TASSELv5.2 (Bradbury et al., 2007). The intra-chromosomal LD was calculated as the pairwise marker correlations (r^2) between the SNP markers plotted against the physical distance for significant MTAs. The long-distance LD and spline were fitted to the LD-decay graph using r^2 values of less than 0.99 using ggplot2 v3.30 in R v3.5.2 (R Core Team, 2019).

Genome-Wide Association Analysis

The FarmCPU (Liu et al., 2016) model of GAPIT 3.0 was used to test the MTA between the SNP markers and spot blotch disease severity (SEV) and to take advantage of the mixed linear model (MLM) and stepwise regression [fixed-effect model (FEM)]. This algorithm uses both the FEM and the random-effect model (REM) iteratively where FEM is employed to test *m* genetic markers, and associations or pseudo-quantitative trait nucleotides are included as covariates to control false positives in REM.

Subsequent GWAS analysis was performed using 14,063 SNPs scored on 139 lines with phenotypic data of disease score from the seasons 2017, 2019, and 2020. Given the exploratory nature of this study, we used a relatively less-stringent *p-value* threshold of 0.003 ($-\log_{10}P = 2.5$) to avoid removing true positive associations. To uncover the stable disease resistance QTLs, association signals that were significant across two or more environments were selected. The allelic effects were further investigated to identify significantly associated markers in lieu of phenotypic data for studying the importance of individual alleles in spot blotch disease resistance.

Gene Functional Annotations

Genome-wide association study results were further analyzed to test if the MTAs fall within the known genic regions using functional annotation from the reference genome assembly (IWGSC Ref Seq v1.0). Functional annotation of the genes either harboring significant SNPs or adjacent to the SNPs was retrieved and examined for their association with spot blotch resistance from the genome annotations provided by IWGSC. Subsequently, protein functions were literature mined from annotated information.

Physical Mapping

From the physical map prepared, we identified the most significant QTLs found on eight wheat chromosomes. A total of 29 SNPs were mapped and used from GWAS wherein a group was considered to be different if physical distance is more than 5 Mb. New QTLs of the present study and genes/markers associated with SB resistance from previous studies were plotted for physical mapping.

RESULTS

Estimation of Heritability

The mean disease severity of the population ranged from 2.65 to 39.35 in three environments (2017, 2019, and 2020), including Pusa, India, and BARI, Bangladesh (**Supplementary Table S2**). The highest mean disease severity was recorded in Pusa20 (Env2) while the lowest was at BARI17 (Env3). The analysis of variance revealed the highest heritability in Env3 (0.82) and the lowest in Env1 (0.72). Based on the combined analysis of all environments, we observed high heritability (0.76). There were significant Genotype × Environment interactions ($P < 0.0001$; **Table 1**). The populations displayed significant phenotypic variation for spot blotch resistance with a nearly continuous distribution of lines in all environments (**Figure 1**).

TABLE 1 | Analysis of variance of 139 advanced lines evaluated for spot blotch disease resistance in four environments based on BLUP of disease severity recorded at GS 77 on Zadoks scale at BISA Pusa (India) and BARI, Jamalpur (Bangladesh).

Statistics	Env1	Env2	Env3	Combined
Heritability	0.72	0.76	0.82	0.76
Genotype variance	45.12	45.04	49.80	31.21
Residual variance	35.12	28.95	22.60	28.89
Grand mean	21.88	21.78	7.72	17.13
LSD	7.04	6.56	6.00	05.42
CV	27.09	24.70	61.59	31.37
N replicates	2	2	2	2
Genotype significance	4.58E-13	1.44E-15	0	1.55E-23
Gen × Env significance	–	–	–	2.30E-11
Gen × Loc variance	–	–	–	15.43

LSD, least significant difference; CV, coefficient of variation; N replicates, number of replicates; Env, environment.

SNP Density and Principal Component Analysis

Among polymorphic SNP markers, 40.9% (5754), 50.8% (7142), and 8.3% (1167) were from the A, B, and D genomes, respectively. With a genomic coverage of 13.9 GB and 14,063 markers across the genome, the average marker density was 1.9 Mb per marker. The lowest marker density 7.03 Mb per marker was at chromosome 4D while the highest 0.54 Mb per marker was observed at chromosome 2B. The average distance between markers for A, B, and D genomes was 0.89, 0.84, and 3.92 Mb, respectively (**Figure 2**).

Population structure was determined using PCA, where genotypes were clustered into 12 groups using the Ward method in JMP v.14 (**Figure 3**). Group 1 (G–I) consisted of nine lines, including the resistant check HD2733. Group G–X consisted of a maximum of 16 lines, while the minimum number of lines is in the group G–XI (six lines). Based on pedigree information, most of the lines in a group shared allele descended from common parents. The lines without common parents or less than three sibs per family were classified as group G-XII. The largest group (G–X) consisted of lines with mixed pedigrees, including TEPOKA, TRCH, SAUAL, WBLL#1, Kachu#1, BAV92//IRENA/KAUZ, FRANCOLIN#1, AMUR, ROLF07, FRET2, BABAX, and BORL14. The parental line SAUAL was the most common parent in group G–IV. The parental lines with TRCH/SRTU//KACHU cross in their genetic backgrounds dominated group G–V.

Marker–Trait Associations of Spot Blotch

The GWAS of spot blotch resistance was performed based on the disease scores collected at adult plant stages. A *p-value* < 0.003 was used as a threshold to identify significant MTAs. The GWAS results from three environments are given in **Figure 4**. A total of 29 significant MTAs appeared in a minimum of two environments belonging to eight QTLs on chromosome 1A, 1B, 1D, 2A, 2B, 4A, 5B, and 6D (**Figure 5**). The allele effects of those QTLs ranged from −31.37% (Env1) to 30.67% (Env2) while the allele effects for individual environments ranged from −31.37 to 30.07%, −22.83 to 30.67%, and −23.30 to 23.24% in Env1, Env2, and Env3, respectively (**Table 2**). The largest allele effect is explained by the SNP S2B_422983662 located on chromosome 2B in Env2 (30.67%) (**Table 2**). The allele effects for alternative alleles from each of the associated SNP markers were plotted (**Figure 6**). We detected a significant variance in the mean values of the favorable alleles that led to an increase in resistance varying from 12.5 to 70% for spot blotch. The "Kruskal–Wallis" test was used to determine whether there are significant differences between the mean values of two alleles. Major alleles have lower mean values compared to minor alleles except for two QTLs (*Q.Sb.bisa-1D* and *Q.Sb.bisa-5B*), where minor alleles are found to be effective (**Figure 6** and **Table 2**).

Physical Mapping

A physical map was prepared using the most significant QTLs placed on eight hexaploid wheat chromosomes. The significant QTLs were clustered into eight linkage groups (**Figure 5**).

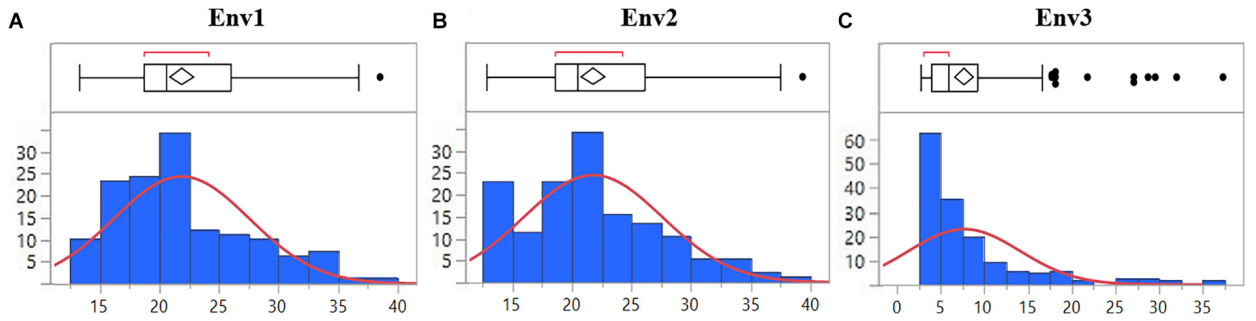


FIGURE 1 | Distribution of 139 advanced wheat breeding lines for spot blotch severity (%) based on BLUPs in **(A)** Env1, **(B)** Env2, and **(C)** Env3 environments. The lines in the boxplot represent the median of the distribution and the diamond represents the model disease score, while the black dots are outliers.

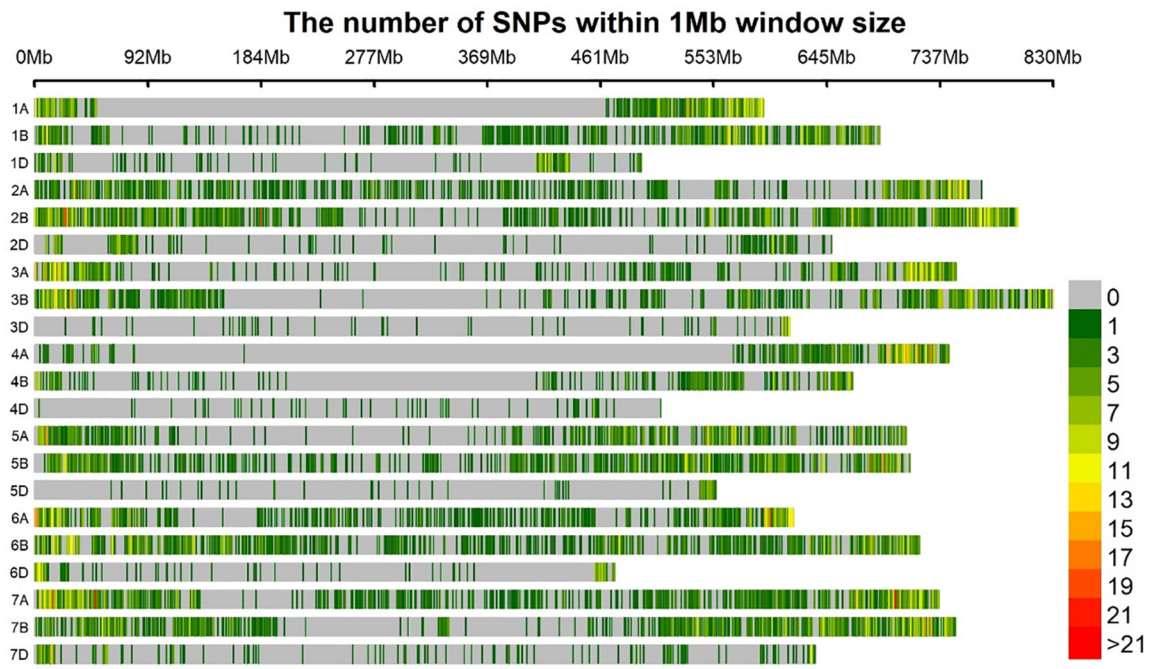


FIGURE 2 | Bar plot showing the densities of 14,063 GBS markers on the 21 wheat chromosomes. The color key with marker densities indicates the number of markers within a window size of 1 Mb.

A group was considered to be different based on critical LD ($r^2 \sim 0.2$) (**Figure 7**). We observed one linkage group on each of 1A, 1B, 1D, 2A, 2B, 4A, and 6D. The maximum number of significant MTAs in two environments (six MTAs) is observed on chromosome 2A (**Table 2**).

Putative Candidate Genes and Annotations

The significant QTLs identified from the GWAS analysis were further studied for the known candidate genes relevant to disease resistance using the recently annotated wheat reference sequence (RefSeq V1.0). We identified NBS-LRR, MADS transcription factor, and 34 other plant protein families across chromosomal regions associated with significant QTLs in the study. The SNP S1A_497201550 associated with *Q.Sb.bisa-1A* identified in Env1 and Env2 on chromosome 1A was located between the *TraesCS1A01G303600*, a gene that encodes LURP-one-like protein, and *TraesCS1A01G303700*, the gene encoding GTP cyclohydrolase 1. Similarly, the SNP S1B_636840957 belongs to *Q.Sb.bisa-1B* identified in Env1 and Env2 on chromosome 1B with GeneID *TraesCS1B01G409800* that encodes 60S ribosomal protein L35a-like protein and *TraesCS1B01G40900* encoding transmembrane protein (**Table 3**). The SNP S1D_89835681, which belongs to the *Q.Sb.bisa-1D* identified in Env1 and Env2, was located close to *TraesCS1D01G101800* associated with Pre-rRNA-processing protein TSR2 and *TraesCS1D01G101900* encoding glucan 1,3 beta-glucosidase. Similarly, the annotation of *Q.Sb.bisa-2A* with SNPs S2A_703111105, S2A_703358397, S2A_703391915, S2A_703391992, and S2A_703427639 revealed

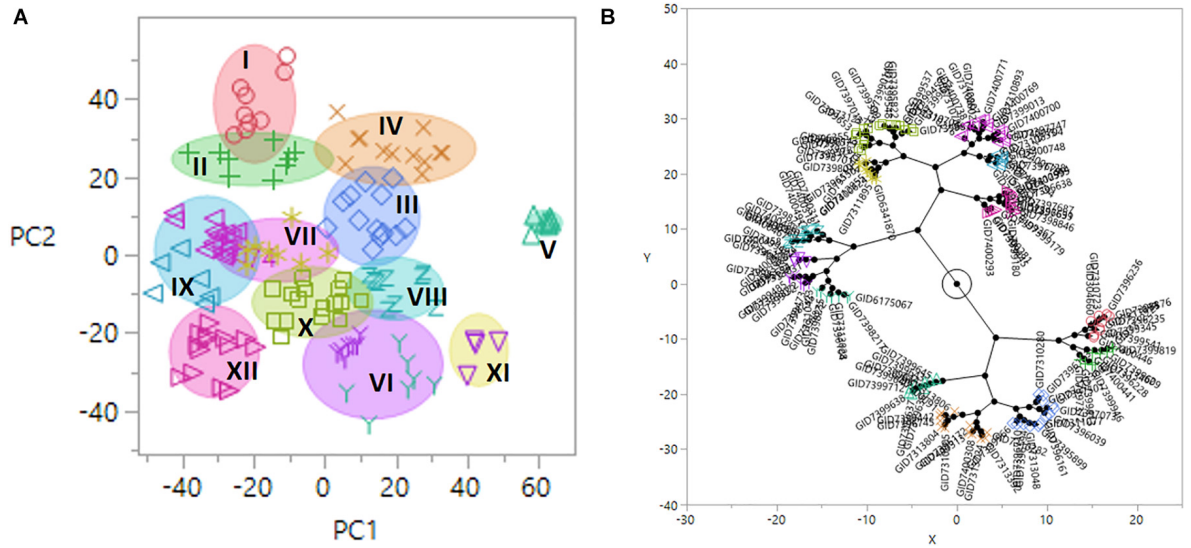


FIGURE 3 | (A) Diagram of principal component (PC) 1 and 2 vectors calculated by the principal component analysis (PCA) of 14,063 genotyping-by-sequencing (GBS) single-nucleotide polymorphism (SNP) markers in 139 advanced wheat breeding lines, plotted in 12 population groups (I–XII). The *x*-axis and *y*-axis represent projections of the PC1 and PC2, respectively. **(B)** Constellation plot using the Ward method in JMP *v.*14.

that the associated genes encode the proteins related to senescence-associated family protein.

Interestingly *Q.Sb.bisa-2B* was detected in all three environments. The Gene IDs of S2B_422983662 from *Q.Sb.bisa-2B* encode MADS-box gene (*TraesCS2B01G300500*), and *TraesCS2B01G300600* encodes a transmembrane protein. The *Q.Sb.bisa-4A* on chromosome 4A with SNP S4A_725660945 encodes *TraesCS4A01G460000* Eyes absent-like protein and (*TraesCS4A01G460100*) Cytochrome P450. The *Q.Sb.bisa-5B* carrying SNP S5B_682958475 and SNP S5B_683240735. Gene IDs (*TraesCS5B01G521000* and *TraesCS5B01G521100*) from SNP S5B_682958475 encode transmembrane protein and calmodulin-binding transcription. SNP S5B_683240735 mapped between *TraesCS5B01G521200* and *TraesCS5B01G521300* encodes receptor kinase 1 and NBS-LRR disease resistance protein family-1, respectively. The *Q.Sb.bisa-6D* of SNP S6D_6395796 detected in Env1 and Env2 lies between *TraesCS6D01G015700* and *TraesCS6D01G015800*. Those genes encode *Leucine-rich repeat receptor-like protein kinase* associated with *TaWRKY76* and *TaWRKY62*. The largest allele effect was explained by the QTL located on chromosome 2B in the gene region that encodes the MADS-box transcription factor and transmembrane protein. However, the QTLs at 1A, 1B, 1D, 2A, 2B, 4A, 5B, and 6D chromosomes are found to be involved directly in the disease resistance mechanism (**Table 3**).

DISCUSSION

Phenotypic Evaluation for Spot Blotch

The field trials were conducted at BISA research farm, Pusa, in India for two consecutive crop seasons 2019 (Env1) and 2020 (Env2). The spot blotch data recorded at BARI, Jamalpur, in Bangladesh during the 2017 crop season was included in the analysis and named Env3. Both the locations fall under the traditional, warmer wheat-growing regions belonging to Mega-Environment 5, characterized by hot and humid conditions as per CIMMYT's system for classifying wheat-growing environments in developing countries (van Ginkel and Rajaram, 1993). The average temperature during the wheat plant reproductive phase at Jamalpur and Pusa is higher than 19°C, with high relative humidity, and serves as a congenial environment for the pathogen (**Supplementary Table S3**). The spot blotch disease incidence was recorded as a percentage of the infected leaf area at three different growth stages to minimize the chances of disease escape. Since the susceptible parent displayed the highest disease severity at GS77 on the Zadoks scale (Zadoks et al., 1974), the scoring at this stage (GS77) was used in further data analysis. Although the testing sites are considered as hot spots for spot blotch (Sharma et al., 2007b), to avoid escaping of genotypes from the pathogen, the highly susceptible cultivar Sonalika was planted in alleys. The BLUPs of disease severity in three environments ranged from 2.65 to 39.35 among advanced wheat breeding lines. The right-tailed skewness of the data in Env3 highlights a likely impact of low disease pressure under natural infection without inoculation in Bangladesh. The nearly continuous distribution of disease scores in all environments indicates the quantitative nature of resistance being caused by the additive effect of various QTLs/genes. A similar trend was reported in some of the earlier findings, where more than two genes (Joshi et al., 2004; Kumar et al., 2007, 2009) and multiple alleles with minor effects (Neupane et al., 2007; Ayana et al., 2018; Singh et al., 2018) were found to be controlling the spot blotch disease resistance.

We observed significant genetic variation for disease susceptibility in the population. The genetic variances and

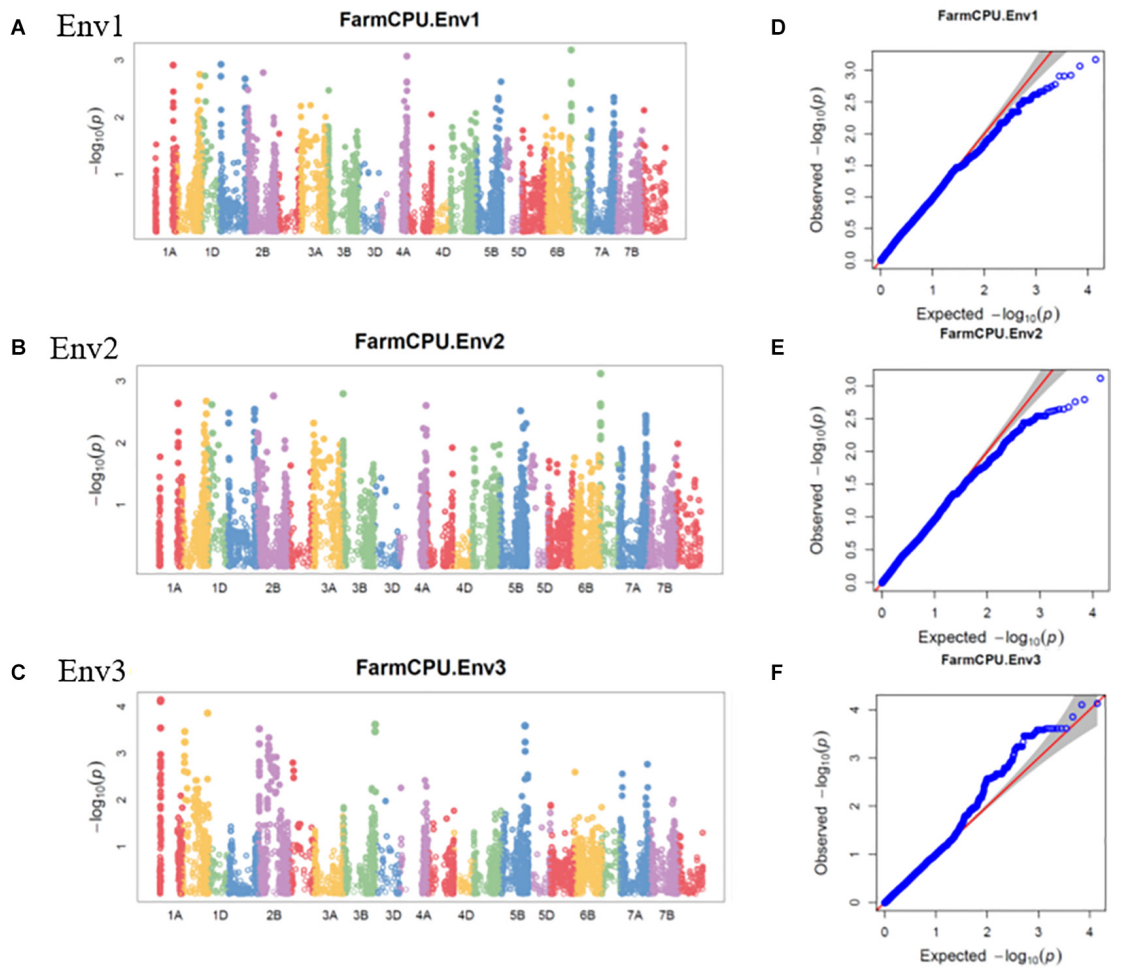


FIGURE 4 | Manhattan plots representing 21 chromosomes showing the significant markers detected by FarmCPU model using BLUP values for spot blotch in **(A)** Env1, **(B)** Env2, and **(C)** Env3 environments. Quantile–quantile (Q–Q) plots for spot blotch in Env1 **(D)**, Env2 **(E)**, and Env3 **(F)** environments showing expected null distribution of *p-values*, assuming no associations, represented as a red solid line v/s; distribution of *p-values* observed using FarmCPU model represented as the blue dots.

high heritabilities for spot blotch were comparable with earlier findings in wheat (Sharma et al., 1997; Joshi et al., 2004). Despite significant genotype × environment interactions, we observed high broad-sense heritability for disease across environments (**Table 1**), highlighting the considerable genetic variation germane for further genetic dissection. The environmental interactions might ascribe to temperature differences at the time of disease development or the difference in the pathogen isolates prevalent in the North Eastern Plains Zone (NEPZ) of India and Bangladesh as well as the varied weather conditions within the location (**Supplementary Table S3**).

Principal Component Analysis

Twelve groups formed in the phylogeny were genetically distinct based on PCA (**Figure 3**). It shows that the advanced lines used in the current study had considerable diversity to identify multiple alleles due to the higher number of resistance sources. Common parent TRCH/SRTU//KACHU are found in group G–V, while the largest group (G–X) consists of lines with mixed pedigrees dominated by TEPOKA, TRCH, SAUAL, WBLL#1, KACHU#1, BAV92//IRENA/KAUZ, FRANCOLIN#1, AMUR, ROLF07, FRET2, BABAX, and BORL14, providing the clues for the resistant genetic resources. Wheat accessions FRANCOLIN, MUCUY, TACUPETO F2001, NADI, BOKOTA, KAUZ, ROLF07, and KACHU were the major contributors to the parentage of most of the genotypes. These results confirm the notion that some elite cultivars have been frequently utilized in the pedigree of germplasm.

SNP Effects in Different Environments of Spot Blotch

Several spot blotch resistance QTLs have been reported on different chromosomes (Neupane et al., 2007; Sharma et al., 2007a; Gonzalez-Hernandez et al., 2009; Kumar et al., 2009, 2010, 2015; Adhikari et al., 2012; Lillemo et al., 2013; Zhuang et al., 2013; Gurung et al., 2014; Lu et al., 2016). However,

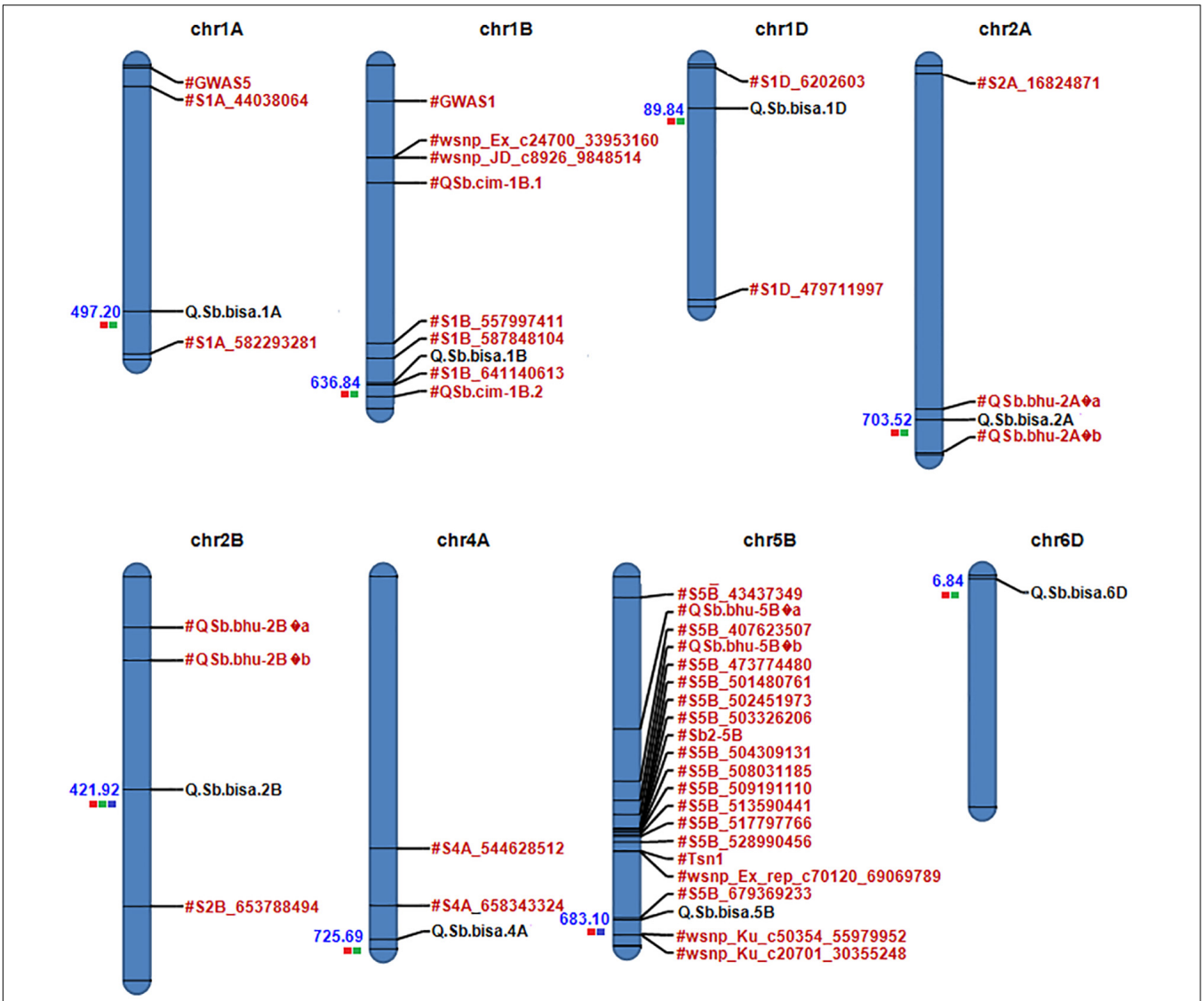


FIGURE 5 | Physical map of candidate QTLs on 1A, 1B,1D, 2A, 2D, 4A, 5B, and 6D chromosomes. Significant QTLs/associated with spot blotch are highlighted in black, while previously reported QTLs/markers are labeled in red (the red color box next to QTLs for Env1, green for Env2, and blue for Env3). The physical position is based on IWGSC 2018 (RefSeq v1.0).

only four major QTLs designated as *Sb1* on 7D (Lillemo et al., 2013), *Sb2* on 5B (Kumar et al., 2015), *Sb3* on 3B (Lu et al., 2016), and *Sb4* on 4BL (Zhang et al., 2020) are well described. We also observed consistent chromosomal regions on 2B and 5B, which were detected in two or all three environments (**Table 2**). Regardless of phenotypic effects explained by an allele, most of the wheat chromosomes have been reported for their contribution to spot blotch disease resistance (Sharma et al., 2007a; Kumar et al., 2009, 2010; Adhikari et al., 2012; Lillemo et al., 2013; Gurung et al., 2014; Zhu et al., 2014; Lu et al., 2016; Gupta et al., 2017; Ayana et al., 2018; Singh et al., 2018; Zhang et al., 2020). The broad range of environmental conditions at our field-testing sites allowed us to capture considerable genetic variation underlying spot blotch resistance. After GWAS analysis, significant SNP markers on eight chromosomes harboring QTL regions forming eight QTLs were physically mapped and used for further analysis. The previously reported genes and markers linked to spot blotch resistance were physically mapped along with the most significant SNPs detected in this study to identify if any new regions were uncovered (**Figure 5**).

To study the importance of significant SNPs in disease resistance, we annotated all SNPs using the wheat reference genome annotation (IWGSC Ref Seq v1.0). The literature was mined intensively to look for the putative functions of those genes/proteins. We found that the functions of several genes were strongly associated with disease resistance across the year and environments (**Table 3**). Two significant chromosomal

TABLE 2 | Summary of the significant SNPs associated with spot blotch resistance.

Names of QTLs	SNP	Alleles	Effective allele	Env	Position (Mb)	*p*-value	MAF	Marker effect	Effect 10X
Q.Sb.bisa-1A (Env 1 and 2)	S1A_497201550	C/T	C	Env1	497.20	0.001232	0.28	1.88	18.83
			C	Env2	497.20	0.002382	0.28	1.83	18.26
	S1A_497201682	G/A	G	Env1	497.20	0.001232	0.28	−1.88	−18.83
			G	Env2	497.20	0.002382	0.28	−1.83	−18.26
Q.Sb.bisa-1B (Env 1 and 2)	S1B_636840957	A/G	A	Env1	636.84	0.001754	0.37	1.61	16.14
			A	Env2	636.84	0.002104	0.37	1.63	16.27
Q.Sb.bisa-1D (Env 1 and 2)	S1D_89835681	T/A	A	Env1	89.84	0.001896	0.13	2.40	23.95
			A	Env2	89.84	0.002417	0.13	2.40	24.00
Q.Sb.bisa-2A (Env 1 and 2)	S2A_703111105	C/T	C	Env2	703.11	0.002881	0.16	2.28	22.83
			C	Env1	703.11	0.002946	0.16	2.22	22.23
	S2A_703358397	T/A	T	Env2	703.36	0.002881	0.16	−2.28	−22.83
			T	Env1	703.36	0.002946	0.16	−2.22	−22.23
	S2A_703391915	T/A	T	Env2	703.39	0.002881	0.16	−2.28	−22.83
			T	Env1	703.39	0.002946	0.16	−2.22	−22.23
	S2A_703391992	C/G	C	Env2	703.39	0.002881	0.16	2.28	22.83
			C	Env1	703.39	0.002946	0.16	2.22	22.23
	S2A_703427639	C/T	C	Env2	703.43	0.002881	0.16	2.28	22.83
			C	Env1	703.43	0.002946	0.16	2.22	22.23
	S2A_704446408	C/T	C	Env2	704.45	0.002881	0.16	2.28	22.83
			C	Env1	704.45	0.002946	0.16	2.22	22.23
Q.Sb.bisa-2B (Env1,2, and 3)	S2B_419320960	C/T	*	Env3	419.32	0.002276	0.46	2.32	23.25
	S2B_419456700	T/C	*	Env3	419.46	0.002138	0.45	−2.33	−23.30
	S2B_420723687	G/T	*	Env3	420.72	0.002497	0.46	2.30	23.03
	S2B_421708152	A/G	*	Env3	421.71	0.002655	0.46	2.30	22.98
	S2B_422031949	T/C	*	Env3	422.03	0.002655	0.46	−2.30	−22.98
	S2B_422190984	C/A	*	Env3	422.19	0.002655	0.46	−2.30	−22.98
	S2B_422191033	A/G	*	Env3	422.19	0.002655	0.46	2.30	22.98
	S2B_422507900	C/T	*	Env3	422.51	0.002655	0.46	2.30	22.98
	S2B_422983662	A/C	A	Env1	422.98	0.001657	0.11	3.01	30.07
			A	Env2	422.98	0.001748	0.11	3.07	30.68
	S2B_423094651	G/A	*	Env3	423.09	0.002655	0.46	−2.30	−22.98
	S2B_423836280	C/G	*	Env3	423.84	0.002655	0.46	2.30	22.98
Q.Sb.bisa-4A (Env 1 and 2)	S4A_725538462	C/T	*	Env1	725.54	0.002455	0.21	1.94	19.41
	S4A_725660945	A/G	A	Env1	725.66	0.000855	0.24	2.02	20.18
			A	Env2	725.66	0.002464	0.24	1.89	18.86
	S4A_725880148	A/G	*	Env1	725.88	0.002388	0.21	1.97	19.72
Q.Sb.bisa-5B (Env1 and 3)	S5B_682958475	G/A	A	Env3	682.96	0.002912	0.46	2.14	21.41
	S5B_683240735	G/A	G	Env1	683.24	0.002373	0.07	−3.14	−31.37
Q.Sb.bisa-6D (Env 1 and 2)	S6D_6395796	T/C	T	Env2	6.4	0.002335	0.27	−1.78	−17.77
			T	Env1	6.4	0.002360	0.27	−1.73	−17.32
	S6D_6944636	T/A	T	Env2	6.94	0.002511	0.27	−1.77	−17.66
			T	Env1	6.94	0.002589	0.27	−1.72	−17.18
	S6D_7194112	G/A	G	Env1	7.19	0.000669	0.27	−1.99	−19.94
			G	Env2	7.19	0.000763	0.27	−2.02	−20.23

*Env, environment; MAF, minor allele frequency. *Asterisk sign in the effective allele column represents that the allelic effect of the two alleles was not significantly different.*

regions/QTLs on 2B and 5B were consistent between Pusa, India, and Jamalpur, Bangladesh. This may be due to the prevalence of the most aggressive isolate of spot blotch pathogen (isolate no. ICMP 13584, Auckland, New Zealand) common in South Asia (Chaurasia et al., 2000). Although we were able to capture large allelic effects ranging from −31.4% (Env1-5B) to 30.7% (Env2-2B) for spot blotch disease severity, some genomic regions of small effect may remain undetected due to the multiple testing criteria of GWAS.

In our study, eight diverse genomic regions were found to be associated with chromosomes 1A, 1B, 1D, 2A, 2B, 4A, 5B, and 6D. The present study not only reports the region on 1A elucidating up to 18.8% of phenotypic variation but also provides insight on the markers S1A_497201550 and

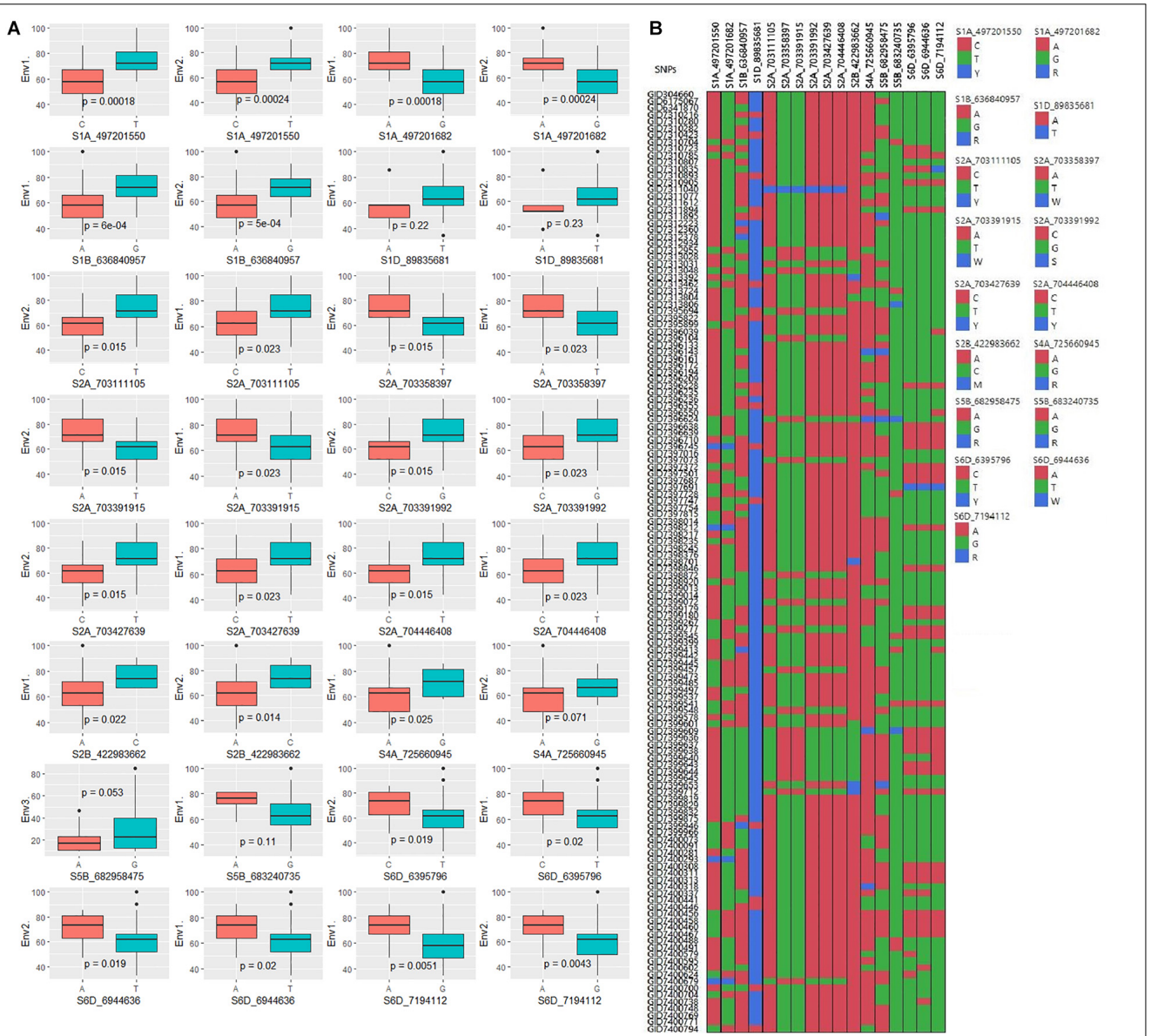


FIGURE 6 | (A) Boxplots showing the effect of phenotypic variation between the two alleles of the SNPs for disease score of bread wheat; the Kruskal–Wallis test was used to determine the significant differences between the mean values of two alleles. **(B)** Cell plot of the distribution of the different alleles of significant SNPs in the advanced wheat breeding lines.

S1A_497201682 by gene annotation. The previous studies (Adhikari et al., 2012; Jamil et al., 2018; Bainsla et al., 2020) on spot blotch reported the chromosomal region on 1A explained 2.1–10% of phenotypic variation. Both SNP markers (S1A_497201550 and S1A_497201682) were associated with LURP-one-like protein that mediates resistance by coordinated transcriptional upregulation of plant defense genes (Knoth et al., 2009). It is noteworthy that the SNP marker S1B_636840957 on chromosome 1B was associated with spot blotch severity, where the annotation study revealed that S1B_636840957 is close to transmembrane protein, which regulates fungal development and pathogenicity via the MAPK module (Gu et al., 2015). Another significant marker S1D_8983568 accounted for 24.0% of the allelic effect, consistent with the earlier studies (Bainsla et al., 2020). However, the QTL reported in our study and Bainsla et al. (2020) was more than 83 Mb apart at chromosome 1D.

The *Q.Sb.bisa-2A* at chromosome 2A between the region (703.11–704.44 Mb) was found to be linked with the senescence-associated family protein. For instance, if one allele is involved with senescence, then the alternate allele is involved with the "stay-green" trait in wheat. Stay-green has been reported to be associated with spot blotch disease resistance (Joshi et al., 2007a; Rosyara et al., 2008). The present results and other independent studies also indicated the importance of this chromosome region

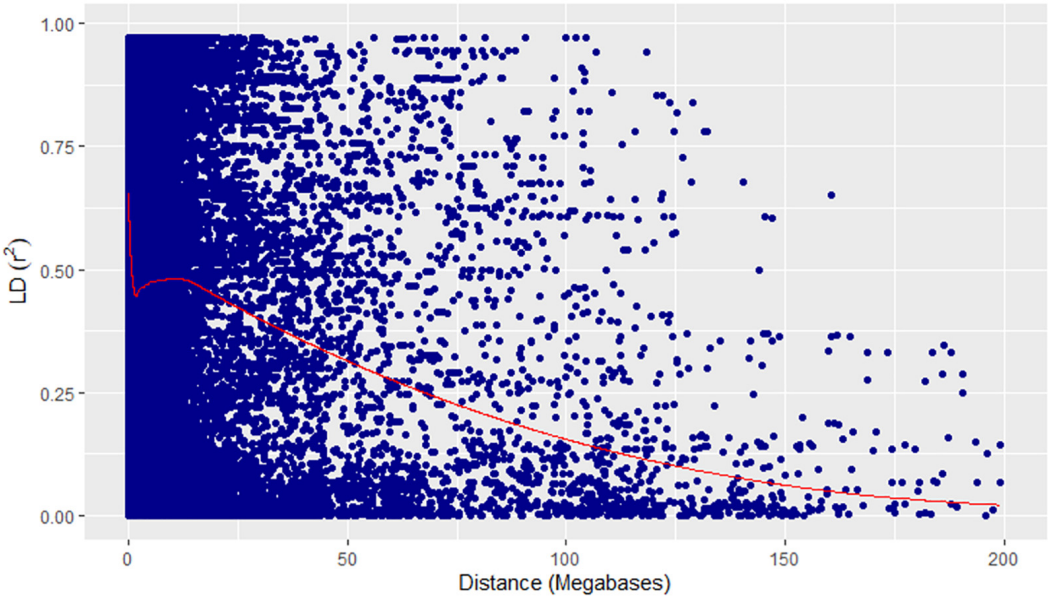


FIGURE 7 | Linkage disequilibrium decay plot (pairwise) showing squared allele-frequency correlation (r^2) vs the megabase pair distance (Mbp) between the pairs of GBS markers showing in blue dots. Megabase pair distances are shown on the *x*-axis and coefficient of determination (r^2) on the *y*-axis. The red line represents the loess curve fitted on the decay plot.

in spot blotch disease resistance (Joshi et al., 2007a; Rosyara et al., 2008; Bainsla et al., 2020).

Kumar et al. (2010) analyzed two bi-parental mapping populations and reported four QTLs for SB resistance including chromosomes 2AS explaining up to 22.7% of phenotypic variation. The other interesting *Q.Sb.bisa-2B* lie between MADS-box genes that were reported to be differentially expressed in response to stripe rust pathogen in wheat (Guo et al., 2013) and transmembrane protein, which regulates fungal development and pathogenicity via the MAPK module (Gu et al., 2015). Similar to 1D, the *Q.Sb.bisa-2B* on 2B mapped nearly 231 Mb apart from the QTL reported by Kumar et al. (2010) and Bainsla et al. (2020). The SNP S4A_725660945 of *Q.Sb.bisa-4A* is associated with wheat Cytochrome P450 family protein. This protein enhances resistance to mycotoxin, namely, deoxynivalenol (DON) and grain yield (Gunupuru et al., 2018). This QTL is in agreement with the earlier finding of Ayana et al. (2018) which was mapped in the same chromosomal region but physically a few Mb away. Since QTL mapping is based on recombination frequency, the possibility of both the QTLs/SNPs being in the same chromosomal region may not be ruled out.

The markers of *Q.Sb.bisa-5B* QTL (S5B_682958475 and S5B_683240735) were associated with calmodulin-binding transcription activator, receptor kinase 1, and NBS-LRR disease resistance protein family-1 which are involved in the defense response of wheat to *Puccinia triticina* (Wang Y. et al., 2019) and fungal pathogen *Zymoseptoria tritici*. On the other hand, the NBS-LRR disease resistance protein family is well known to contribute to fungal disease resistance (He et al., 2018). The QTL on 5B, named the *Sb2* gene, has been studied in detail (Lu et al., 2016), known to interact with the *Tsn1* gene, conferring susceptible reaction to tan spot and *Septoria nodorum* blotch (McDonald et al., 2018). The gene *ToxA* virulent to *Tsn1* exists in both *Pyrenophora tritici-repentis* and *Parastagonospora nodorum* which confer susceptible reaction to tan spot and *S. nodorum* blotch, respectively (Singh et al., 2018). Friesen et al. (2018) demonstrated the major effects of the *Tsn1* locus on chromosome 5B. However, the importance of *Tsn1* in spot blotch disease resistance under field conditions is not known. Another QTL for spot blotch resistance was mapped earlier in the same region by Jamil et al. (2018).

The *Q.Sb.bisa-6D* carrying SNP S6D_6395796 on chromosome 6D had an allelic effect of −17.8% with the stable response in two environments and annotated to be located close to a gene that synthesizes Leucine-rich repeat receptor-like protein kinase family protein, which plays an important role in disease resistance. The LRR-Like kinase gene associated with *TaWRKY76* and *TaWRKY62* plays a positive role in wheat high-temperature plant resistance to *Puccinia striiformis* f. sp. *tritici* (Wang J. et al., 2019). The 6D chromosome has earlier been identified to impart resistance against spot blotch in QTL mapping but via different genomic regions, i.e., near the centromeric region (Kumar et al., 2009, 2010) and proximal region of 6DS (Singh et al., 2018). The exact physical regions of these QTLs could not be estimated as they were either based on SSR markers or a genetic map. The consistency of SNPs on eight chromosomes (1A, 1B, 1D, 2A, 2B, 4D, 5B, and 6D) in a minimum of two or all environments indicates their potential significance in the breeding of disease-resistant varieties. Further, the resistance mechanism through protein annotation was confirmed where the same or common gene/protein family was identified independently in different environments (**Figure 8**).

TABLE 3 | SNPs with the corresponding proteins and possible function elucidated based on the gene annotation using wheat reference sequence (RefSeq V1.0) annotation database.

QTLs and Env	SNP	Gene ID	Dist. from SNP (in bp)	Gene annotation
Q.Sb.bisa-1A (Env 1 and 2)	S1A_497201550	TraesCS1A01G303600	+9400	LURP-one-like protein
		TraesCS1A01G303700	−23,504	GTP cyclohydrolase 1
	S1A_497201682	TraesCS1A01G303600	+9532	LURP-one-like protein
		TraesCS1A01G303700	−23,372	GTP cyclohydrolase 1
Q.Sb.bisa-1B (Env 1 and 2)	S1B_636840957	TraesCS1B01G409800	+2658	60Sribosomal protein L35a-like protein
		TraesCS1B01G409900	−19,389	Transmembrane protein, putative
Q.Sb.bisa-1D (Env 1 and 2)	S1D_89835681	TraesCS1D01G101800	+237,973	Pre-rRNA-processing protein TSR2
		TraesCS1D01G101900	−312,998	Glucan 1,3-beta-glucosidase
Q.Sb.bisa-2A (Env 1 and 2)	S2A_703111105	TraesCS2A01G453900	+92	Senescence-associated family protein
		TraesCS2A01G454000	−11,947	Senescence-associated family protein
	S2A_703358397	TraesCS2A01G454400	+12,352	Senescence-associated family protein
		TraesCS2A01G454500	−23,631	Senescence-associated family protein
	S2A_703391915	TraesCS2A01G454500	+9887	Senescence-associated family protein
		TraesCS2A01G454600	−10,955	Senescence-associated family protein
	S2A_703391992	TraesCS2A01G454500	+9964	Senescence-associated family protein
		TraesCS2A01G454600	−10,878	Senescence-associated family protein
	S2A_703427639	TraesCS2A01G454600	+24,769	Senescence-associated family protein
		TraesCS2A01G454700	−20,520	senescence-associated family protein
	S2A_704446408	TraesCS2A01G455500	+427,884	Cortactin-binding protein 2
		TraesCS2A01G455600	−307,046	Lectin receptor kinase
Q.Sb.bisa-2B (Env1,2, and 3)	S2B_419320960	TraesCS2B01G299000	+365,182	SNF1-related protein kinase regulatory subunit gamma 1
		TraesCS2B01G299100	−717	Cotton fiber-like protein
	S2B_419456700	TraesCS2B01G299100	+135,023	Cotton fiber-like protein
		TraesCS2B01G299200	−45,804	Amino acid transporter, putative
	S2B_420723687	TraesCS2B01G299700	+520,847	MYND type zinc finger protein
		TraesCS2B01G299800	−358,751	RNA binding protein
	S2B_421708152	TraesCS2B01G300100	+3383	UPF0250 protein mma_3250
		TraesCS2B01G300200	−623	Isocitrate lyase

(Continued)

TABLE 3 | Continued

QTLs and Env	SNP	Gene ID	Dist. from SNP (in bp)	Gene annotation
	S2B_422031949	TraesCS2B01G300200	+323,174	Isocitrate lyase
		TraesCS2B01G300300	−202,765	Auxin response factor
	S2B_422190984	TraesCS2B01G300200	+482,209	Isocitrate lyase
		TraesCS2B01G300300	−43,730	Auxin response factor
	S2B_422191033	TraesCS2B01G300200	+482,258	Isocitrate lyase
		TraesCS2B01G300300	−43,681	Auxin response factor
	S2B_422507900	TraesCS2B01G300400	+7502	DNAprimase large subunit
		TraesCS2B01G300500	−458,849	MADS box transcription factor
	S2B_422983662	TraesCS2B01G300500	+16,913	MADS box transcription factor
		TraesCS2B01G300600	−114,123	Transmembrane protein, putative
	S2B_423094651	TraesCS2B01G300500	+127,902	MADS box transcription factor
		TraesCS2B01G300600	−3134	Transmembrane protein, putative
	S2B_423836280	TraesCS2B01G301100	+868	Chaperone protein dnaJ
		TraesCS2B01G301200	−63,852	Zinc knuckle family protein, expressed
Q.Sb.bisa-4A (Env 1 and 2)	S4A_725538462	TraesCS4A01G459600	+272,847	Protein phosphatase 2C
		TraesCS4A01G459700	−33,166	Seed maturation protein
	S4A_725660945	TraesCS4A01G460000	+597	Eyes absent-like protein
		TraesCS4A01G460100	−8438	Cytochrome P450
	S4A_725880148	TraesCS4A01G460800	+19,993	Invertase inhibitor
		TraesCS4A01G460900	−13,027	Invertase inhibitor
Q.Sb.bisa-5B (Env1 and 3)	S5B_682958475	TraesCS5B01G521000	+5865	Transmembrane protein, putative
		TraesCS5B01G521100	−150,684	Calmodulin-binding transcription activator
	S5B_683240735	TraesCS5B01G521200	+100,340	Receptor kinase 1
		TraesCS5B01G521300	−99,434	NBS-LRR disease resistance protein family-1
Q.Sb.bisa-6D (Env 1 and 2)	S6D_6395796	TraesCS6D01G015700	+25,746	Leucine-rich repeat receptor-like protein kinase family protein
		TraesCS6D01G015800	−29,399	Leucine-rich repeat receptor-like protein kinase family protein
	S6D_6944636	TraesCS6D01G016400	+120,493	Transposon protein, putative, CACTA, En/Spm subclass
		TraesCS6D01G016500	−19,307	Centromere O
	S6D_7194112	TraesCS6D01G017300	+30,119	Auxin transport protein BIG
		TraesCS6D01G017400	−15,620	Polygalacturonase-1 non-catalytic beta subunit

Distance from the SNP represents distance of the start site of gene to SNP linked with QTL, where the (+) sign represents that the gene was found downstream of the SNP and the (–) sign represents that the gene was found upstream.

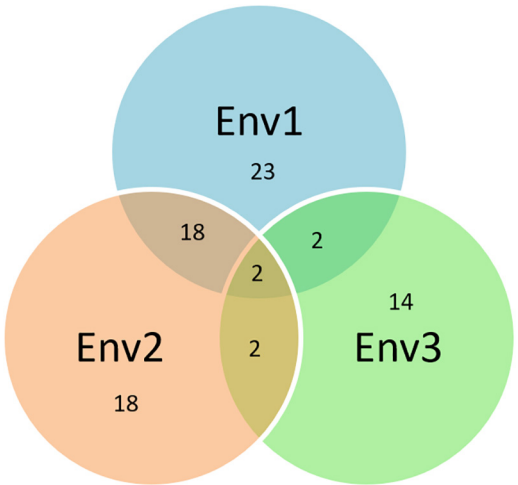


FIGURE 8 | Venn diagram based on the common protein synthesized by the same genes associated with different SNPs over years (the numbers in the circle represent the total numbers of proteins in the respective environments and the numbers at intersects represent the common proteins between the environments).

With additional validation, these genetic regions reported in this study can potentially be used in fine mapping and map-based cloning to further characterize the mechanisms of spot blotch disease resistance.

CONCLUSION

We identified genetic regions underlying spot blotch resistance in the elite spring wheat genotypes. The variable conditions at three field environments in India and Bangladesh allowed us to capture the considerable phenotypic variation for spot blotch disease with the GWAS resulting in a total of eight QTLs belonging to eight wheat chromosomes. The literature mining of the functional gene annotations of identified SNPs encoding the single protein or protein family directly or indirectly involved in disease resistance has led to the identification of putative target genes and functions to identify the disease-resistance mechanism. The new QTLs appeared on chromosomes 1A (497.2), 1D (89.84), 2B (421.92), and 6D (6.84) associated with many disease resistance family proteins. The SNP on chromosome 2A was found to be associated with a known gene that encodes "senescence-associated family protein" and is directly involved in spot blotch resistance. The results are of importance for the breeders in developing spot blotch-resistant varieties targeting the South Asian wheat-growing regions. Given the aggressive pathogen spread and food security concerns, the breeding programs in South Asia could benefit from the present study. The mapping of favorable alleles can facilitate introgression of the alleles into present-day elite cultivars to impart disease resistance. It is apparent from the results that some individual alleles cumulatively contributed as high as 70% for spot blotch disease resistance. Additional investigations are underway, which would further confirm the importance of these chromosomal regions/genes associated with spot blotch.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in **Supplementary Material**.

AUTHOR CONTRIBUTIONS

VT wrote the initial draft of the manuscript. VT, DS, and GD analyzed the data. RS provided the breeding material. JP provided genotyping data. VT and MR managed field trials and recorded phenotypic data. PB, SK, and PS provided scientific inputs during the initial manuscript drafting. VT, DS, UK, GD, and BT revised the manuscript. UK designed the experiment. UK, AJ, and BT supervised the research project. All authors read and approved the final manuscript.

FUNDING

This work was supported by the US Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics grant (Cooperative Agreement No. AID-OAA-A-13-00051).

ACKNOWLEDGMENTS

The technical support provided by Manish Kumar, Dr. Prashanth N. Suravajhala, and Dr. Raj Kumar Jat is duly acknowledged. This manuscript has been released as a pre-print at (https://www.researchsquare.com/article/rs-13392/v1), [Tomar et al. (2020) Genome-wide association study and genomic prediction of spot blotch disease in wheat (*Triticum aestivum* L.). doi: 10.21203/rs.2.22818/v1]. However, the current manuscript has considerable updates compared to the pre-print (additional year data have been used and some of the data and analysis have been removed).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.613217/full#supplementary-material

Supplementary Table 1 | List of 139 genotypes with clustering information used in the present study.

Supplementary Table 2 | Statistical analysis of spot blotch incidence in 139 wheat genotypes evaluated at two locations in three environments.

Supplementary Table 3 | Monthly average, minimum, and maximum temperatures at BISA, Pusa in India.

REFERENCES

Adhikari, T. B., Gurung, S., Hansen, J. M., Jackson, E. W., and Bonman, J. M. (2012). Association mapping of quantitative trait loci in spring wheat landraces conferring resistance to bacterial leaf streak and spot blotch. *Plant Genome.* 5, 1–16. doi: 10.3835/plantgenome2011.12.0032

Alvarado, G., Rodríguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010

Ayana, G. T., Ali, S., Sidhu, J. S., Gonzalez Hernandez, J. L., Turnipseed, B., and Sehgal, S. K. (2018). Genome-wide association study for spot blotch resistance in hard winter wheat. *Front. Plant Sci.* 9:926. doi: 10.3389/fpls.2018.00926

Bainsla, N. K., Phuke, R. M., He, X., Gupta, V., Bishnoi, S. K., Sharma, R. K., et al. (2020). Genome-wide association study for spot blotch resistance in Afghan wheat germplasm. *Plant Pathol.* 69, 1161–1171. doi: 10.1111/ppa.13191

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Browning, B. L., and Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126. doi: 10.1016/j.ajhg.2015.11.020

Chaurasia, S., Chand, R., and Joshi, A. K. (2000). Relative dominance of *Alternaria triticina* Pras. et Prab. and Bipolaris sorokiniana (Sacc.) shoemaker in different growth stages of wheat (*T. aestivum* L.). *Z. Pflanzenkr. Pflanzenschutz.* 107, 176–181.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Dreisigacker, S., Deepmala, S., Jaimez, R. A., Luna-Garrid, B., Muñoz-Zavala, S., Núñez-Ríos, C., et al. (2016). *CIMMYT Wheat Molecular Genetics: Laboratory Protocols and Applications to Wheat Breeding*. Mexico, DF: CIMMYT.

Dubin, H. J., and Van Ginkel, M. (1991). "The status of wheat diseases and disease research in warmer areas," in *Proceeding of the International Conference, Wheat for non traditional warmer areas*, ed. D. A. Saunders (Mexico: CIMMYT), 125–145.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

FAO (2018). *Statistical Pocketbook 2018. World Food and Agriculture*. Rome: FAO, doi: 978-92-5-108802-9

French, E., Kim, B. S., and Iyer-Pascuzzi, A. S. (2016). Mechanisms of quantitative disease resistance in plants. *Semin. Cell Dev. Biol.* 56, 201–208. doi: 10.1016/j.semcdb.2016.05.015

Friesen, T. L., Holmes, D. J., Bowden, R. L., and Faris, J. D. (2018). ToxA is present in the U.S. bipolaris sorokiniana population and is a significant virulence factor on wheat harboring Tsn1. *Plant Dis.* 102, 2446–2452. doi: 10.1094/PDIS-03-18-0521-RE

Friesen, T. L., Xu, S. S., and Harris, M. O. (2008). Stem rust, tan spot, stagonospora nodorum blotch, and hessian fly resistance in langdon durum-Aegilops tauschii synthetic hexaploid wheat lines. *Crop. Sci.* 48, 1062–1070. doi: 10.2135/cropsci2007.08.0463

Gonzalez-Hernandez, J. L., Singh, P. K., Mergoum, M., Adhikari, T. B., Kianian, S. F., Simsek, S., et al. (2009). A quantitative trait locus on chromosome 5B controls resistance of *Triticum turgidum* (L.) var. diccocoides to *Stagonospora* nodorum blotch. *Euphytica* 159, 167–179. doi: 10.1007/s10681-008-9825-z

Gu, Q., Chen, Y., Liu, Y., Zhang, C., and Ma, Z. (2015). The transmembrane protein FgSho1 regulates fungal development and pathogenicity via the MAPK module Ste50-Ste11-Ste7 in *Fusarium* graminearum. *New Phytol.* 206, 315–328. doi: 10.1111/nph.13158

Gunupuru, L. R., Arunachalam, C., Malla, K. B., Kahla, A., Perochon, A., Jia, J., et al. (2018). A wheat cytochrome P450 enhances both resistance to deoxynivalenol and grain yield. *PLoS One* 13:e0204992. doi: 10.1371/journal.pone.0204992

Guo, J., Shi, X. X., Zhang, J. S., Duan, Y. H., Bai, P. F., Guan, X. N., et al. (2013). A type I MADS-box gene is differentially expressed in wheat in response to infection by the stripe rust fungus. *Biol. Plant.* 57, 540–546. doi: 10.1007/s10535-012-0297-6

Gupta, P. K. K., Chand, R., Vasistha, N. K. K., Pandey, S. P. P., Kumar, U., Mishra, V. K. K., et al. (2017). Spot blotch disease of wheat: the current status of research on genetics and breeding. *Plant Pathol.* 67, 508–531. doi: 10.1111/ppa.12781

Gurung, S., Bonman, J. M., Ali, S., Patel, J., Myrfield, M., Mergoum, M., et al. (2009). New and diverse sources of multiple disease resistance in wheat. *Crop. Sci.* 49, 1655–1666. doi: 10.2135/cropsci2008.10.0633

Gurung, S., Hansen, J. M., Michael Bonman, J., Gironella, A. I. N., and Adhikari, T. B. (2012). Multiple disease resistance to four leaf spot diseases in winter wheat accessions from the USDA national small grains collection. *Crop. Sci.* 52:1640. doi: 10.2135/cropsci2011.08.0408

Gurung, S., Mamidi, S., Bonman, J. M., Xiong, M., Brown-Guedira, G., and Adhikari, T. B. (2014). Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS One* 9:e108179. doi: 10.1371/journal.pone.0108179

He, H., Zhu, S., Zhao, R., Jiang, Z., Ji, Y., Ji, J., et al. (2018). Pm21, encoding a typical CC-NBS-LRR protein, confers broad-spectrum resistance to wheat powdery mildew disease. *Mol. Plant.* 11, 879–882. doi: 10.1016/j.molp.2018.03.004

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Ibrahim, A. M. H., and Mujeeb-Kazi, A. (2018). Genome-wide association studies for spot blotch (*Cochliobolus sativus*) resistance in bread wheat using genotyping-by-sequencing. *Phytopathology* 108, 1307–1314. doi: 10.1094/PHYTO-02-18-0047-R

Joshi, A. K., Kumar, S., Chand, R., and Ortiz-Ferrara, G. (2004). Inheritance of resistance to spot blotch caused by Bipolaris sorokiniana in spring wheat. *Plant Breed.* 123, 213–219. doi: 10.1111/j.1439-0523.2004.00954.x

Joshi, A. K., Kumari, M., Reddy, V. P. S., Kumar, S., Rane, J., Chand, R., et al. (2007a). Stay green trait: variation, inheritance and its association with spot blotch resistance in spring wheat (*Triticum aestivum* L.). *Euphytica* 153, 59–71. doi: 10.1007/s10681-006-9235-z

Joshi, A. K., Mishra, B., Chatrath, R., Ortiz Ferrara, G., and Singh, R. P. (2007b). Wheat improvement in India: present status, emerging challenges and future prospects. *Euphytica* 157, 431–446. doi: 10.1007/s10681-007-9385-7

Knoth, C., Salus, M. S., Girke, T., and Eulgem, T. (2009). The synthetic elicitor 3,5-dichloroanthranilic acid induces NPR1-dependent and npr1-independent mechanisms of disease resistance in arabidopsis. *Plant Physiol.* 150, 333–347. doi: 10.1104/pp.108.133678

Kumar, S., Prasad, L. C., Kumar, U., Tyagi, K., Balasubramaniyam, A., and Joshi, A. K. (2007). "Inheritance and allelic relationship of resistance genes to spot blotch of wheat caused by bipolaris sorokiniana," in *Wheat Production in Stressed Environments Wheat Production in Stressed Environments*, eds H. T. Buck, et al. (Berlin: Springer), 113–118. doi: 10.1007/1-4020-5497-1_15

Kumar, S., Röder, M. S., Tripathi, S. B., Kumar, S., Chand, R., Joshi, A. K., et al. (2015). Mendelization and fine mapping of a bread wheat spot blotch disease resistance QTL. *Mol. Breed.* 35:218. doi: 10.1007/s11032-015-0411-5

Kumar, U., Joshi, A. K., Kumar, S., Chand, R., and Röder, M. S. (2009). Mapping of resistance to spot blotch disease caused by bipolaris sorokiniana in spring wheat. *Theor. Appl. Genet.* 118, 783–792. doi: 10.1007/s00122-008-0938-5

Kumar, U., Joshi, A. K., Kumar, S., Chand, R., and Röder, M. S. (2010). Quantitative trait loci for resistance to spot blotch caused by bipolaris sorokiniana in wheat (*T. aestivum* L.) lines "Ning 8201" and "Chirya 3.". *Mol. Breed.* 26, 477–491. doi: 10.1007/s11032-009-9388-2

Lennard, J. H. (1984). *Common Diseases of Small Grain Cereals: A Guide to Identification. By F. J. Zillinsky. Mexico: CIMMYT (1983), pp. 141*. Cambridge: Cambridge University Press, doi: 10.1017/s0014479700017695

Lillemo, M., Joshi, A. K., Prasad, R., Chand, R., and Singh, R. P. (2013). QTL for spot blotch resistance in bread wheat line Saar co-locate to the biotrophic disease resistance loci Lr34 and Lr46. *Theor. Appl. Genet.* 126, 711–719. doi: 10.1007/s00122-012-2012-6

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12:e1005767. doi: 10.1371/journal.pgen.1005767

Lu, P., Liang, Y., Li, D., Wang, Z., Li, W., Wang, G., et al. (2016). Fine genetic mapping of spot blotch resistance gene Sb3 in wheat (*Triticum aestivum*). *Theor. Appl. Genet.* 129, 577–589. doi: 10.1007/s00122-015-2649-z

Maccaferri, M., Sanguineti, M. C., Mantovani, P., Demontis, A., Massi, A., Ammar, K., et al. (2010). Association mapping of leaf rust response in durum wheat. *Mol. Breed.* 26, 189–228. doi: 10.1007/s11032-009-9353-0

Maccaferri, M., Zhang, J., Bulli, P., Abate, Z., Chao, S., Cantu, D., et al. (2015). A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. tritici) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3* 5, 449–465. doi: 10.1534/g3.114.014563

Marone, D., Russo, M. A., Laidò, G., De Vita, P., Papa, R., Blanco, A., et al. (2013). Genetic basis of qualitative and quantitative resistance to powdery mildew in wheat: from consensus regions to candidate genes. *BMC Genomics* 14:562. doi: 10.1186/1471-2164-14-562

McDonald, M. C., Ahren, D., Simpfendorfer, S., Milgate, A., and Solomon, P. S. (2018). The discovery of the virulence gene ToxA in the wheat and barley pathogen bipolaris sorokiniana. *Mol. Plant Pathol.* 19, 432–439. doi: 10.1111/mpp.12535

Mondal, S., Rutkoski, J. E., Velu, G., Singh, P. K., Crespo-Herrera, L. A., Guzman, C., et al. (2016). Harnessing diversity in wheat to enhance grain yield, climate resilience, disease and insect pest resistance and nutrition through conventional and modern breeding approaches. *Front. Plant Sci.* 7:991. doi: 10.3389/fpls.2016.00991

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., et al. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell.* 21, 2194–2202. doi: 10.1105/tpc.109.068437

Neupane, R. B., Sharma, R. C., Duveiller, E., Ortiz-Ferrara, G., Ojha, B. R., Rosyara, U. R., et al. (2007). Major gene controls of field resistance to spot blotch in wheat genotypes "Milan/Shanghai #7" and "Chirya.3.". *Plant Dis.* 91, 692–697. doi: 10.1094/PDIS-91-6-0692

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7:e32253. doi: 10.1371/journal.pone.0032253

Rosyara, U. R., Khadka, K., Subedi, S., Sharma, R. C., and Duveiller, E. (2008). Heritability of stay green traits and association with spot blotch resistance in three spring wheat populations. *J. Genet. Breed.* 62, 75–82.

Sharma, R. C., Dubin, H. J., Devkota, R. N., and Bhatta, M. R. (1997). Heritability estimates of field resistance to spot blotch in four spring wheat crosses. *Plant Breed.* 116, 64–68. doi: 10.1111/j.1439-0523.1997.tb00976.x

Sharma, R. C., Duveiller, E., and Jacquemin, J. M. (2007a). Microsatellite markers associated with spot blotch resistance in spring wheat. *J. Phytopathol.* 155, 316–319. doi: 10.1111/j.1439-0434.2007.01238.x

Sharma, R. C., Duveiller, E., and Ortiz-Ferrara, G. (2007b). Progress and challenge towards reducing wheat spot blotch threat in the Eastern Gangetic Plains of South Asia: is climate change already taking its toll? *F. Crop. Res.* 103, 109–118. doi: 10.1016/j.fcr.2007.05.004

Singh, P. K., He, X., Sansaloni, C. P., Juliana, P., Dreisigacker, S., Duveiller, E., et al. (2018). Resistance to spot blotch in two mapping populations of common wheat is controlled by multiple QTL of minor effects. *Int. J. Mol. Sci.* 19:4054. doi: 10.3390/ijms19124054

Tomar, V., Singh, R. P., Poland, J., Singh, D., Joshi, A. K., Singh, P. K., et al. (2020). *Genome-Wide Association Study and Genomic Prediction of Spot Blotch Disease in Wheat (Triticum aestivum L.) using Genotyping by Sequencing*. Durham, NC: Research Square, doi: 10.21203/rs.2.22818/v1

van Ginkel, M., and Rajaram, S. (1993). "Breeding for durable resistance to diseases in wheat an additional perspective," in *Durability of Disease Resistance*, eds T. Jacobes and J. E. Parlevliet (Dordrecht: Kluwer Academic Publishers), 259–272. doi: 10.1007/978-94-011-2004-3_28

van Ginkel, M., and Rajaram, S. (1998). "Breeding for resistance to spot blotch in wheat: a global perspective," in *Helminthosporium Blights of Wheat: Spot Blotch and Tan Spot*, eds E. Duveiller, H. J. Dubin, J. Reeves, and A. McNab (Mexico DF: CIMMYT), 162–170.

Wang, J., Wang, J., Shang, H., Chen, X., Xu, X., and Hu, X. (2019). TaXa21, a leucine-rich repeat receptor-like kinase gene associated with TaWRKY76 and TaWRKY62, plays positive roles in wheat high-temperature seedling plant resistance to puccinia striiformis f. sp. tritici. *Mol. Plant Microbe Interact.* 32, 1526–1535. doi: 10.1094/MPMI-05-19-0137-R

Wang, Y., Wei, F., Zhou, H., Liu, N., Niu, X., Yan, C., et al. (2019). TaCAMTA4, a calmodulin-Interacting protein, involved in defense response of wheat to *Puccinia* triticina. *Sci. Rep.* 9:641. doi: 10.1038/s41598-018-36385-1

Yu, L. X., Lorenz, A., Rutkoski, J., Singh, R. P., Bhavani, S., Huerta-Espino, J., et al. (2011). Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. *Theor. Appl. Genet.* 123, 1257–1268. doi: 10.1007/s00122-011-1664-y

Zadoks, J. C., Chang, T. T., and Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* 14, 415–421. doi: 10.1111/j.1365-3180.1974.tb01084.x

Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., and Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One* 9:e105593. doi: 10.1371/journal.pone.0105593

Zhang, P., Guo, G., Wu, Q., Chen, Y., Xie, J., Lu, P., et al. (2020). Identification and fine mapping of spot blotch (Bipolaris sorokiniana) resistance gene Sb4 in wheat. *Theor. Appl. Genet.* 133, 2451–2459. doi: 10.1007/s00122-020-03610-3

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20. doi: 10.3835/plantgenome2008.02.0089

Zhu, X., Qi, L., Liu, X., Cai, S., Xu, H., Huang, R., et al. (2014). The wheat ethylene response factor transcription factor PATHOGEN-INDUCED ERF1 mediates host responses to both the necrotrophic pathogen rhizoctonia cerealis and freezing stresses. *Plant Physiol.* 164, 1499–1514. doi: 10.1104/pp.113.229575

Zhuang, Y., Gala, A., and Yen, Y. (2013). Identification of functional genic components of major Fusarium head blight resistance quantitative trait loci in wheat cultivar Sumai 3. *Mol. Plant Microbe Interact.* 26, 442–450. doi: 10.1094/MPMI-10-12-0235-R

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Tomar, Singh, Dhillon, Singh, Poland, Joshi, Singh, Bhati, Kumar, Rahman, Tiwari and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fgene.2020.601500

The Barley *HvWRKY6* Transcription Factor Is Required for Resistance Against *Pyrenophora teres* f. *teres*

*Prabin Tamang[1], Jonathan K. Richards[2], Shyam Solanki[3], Gazala Ameen[3], Roshan Sharma Poudel[1], Priyanka Deka[4], Karl Effertz[3], Shaun J. Clare[3], Justin Hegstad[5], Achintya Bezbaruah[4], Xuehui Li[5], Richard D. Horsley[5], Timothy L. Friesen[1,6] and Robert S. Brueggeman[1,3]**

[1] Department of Plant Pathology, North Dakota State University, Fargo, ND, United States, [2] Department of Plant Pathology and Crop Physiology, Louisiana State University, Baton Rouge, LA, United States, [3] Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States, [4] Department of Civil and Environmental Engineering, North Dakota State University, Fargo, ND, United States, [5] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [6] Cereal Crops Research Unit, United States Department of Argiculture – Agricultural Research Service, Edward T. Schafer Agricultural Research Center, Fargo, ND, United States

OPEN ACCESS

Edited by:
Deepmala Sehgal,
International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Vandana Jaiswal,
Institute of Himalayan Bioresource Technology (CSIR), India
Vijay Gahlaut,
University of Delhi, India
Nepolean Thirunavukkarasu,
Indian Institute of Millets Research (ICAR-IIMR), India

***Correspondence:**
Robert S. Brueggeman
Bob.brueggeman@wsu.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 01 September 2020
Accepted: 14 December 2020
Published: 15 January 2021

Citation:
Tamang P, Richards JK, Solanki S, Ameen G, Sharma Poudel R, Deka P, Effertz K, Clare SJ, Hegstad J, Bezbaruah A, Li X, Horsley RD, Friesen TL and Brueggeman RS (2021) The Barley HvWRKY6 Transcription Factor Is Required for Resistance Against Pyrenophora teres f. teres. Front. Genet. 11:601500. doi: 10.3389/fgene.2020.601500

Barley is an important cereal crop worldwide because of its use in the brewing and distilling industry. However, adequate supplies of quality malting barley are threatened by global climate change due to drought in some regions and excess precipitation in others, which facilitates epidemics caused by fungal pathogens. The disease net form net blotch caused by the necrotrophic fungal pathogen *Pyrenophora teres* f. *teres* (*Ptt*) has emerged as a global threat to barley production and diverse populations of *Ptt* have shown a capacity to overcome deployed genetic resistances. The barley line CI5791 exhibits remarkably effective resistance to diverse *Ptt* isolates from around the world that maps to two major QTL on chromosomes 3H and 6H. To identify genes involved in this effective resistance, CI5791 seed were γ-irradiated and two mutants, designated CI5791-γ3 and CI5791-γ8, with compromised *Ptt* resistance were identified from an M_2 population. Phenotyping of CI5791-γ3 and -γ8 × Heartland F_2 populations showed three resistant to one susceptible segregation ratios and CI5791-γ3 × -γ8 F_1 individuals were susceptible, thus these independent mutants are in a single allelic gene. Thirty-four homozygous mutant (susceptible) CI5791-γ3 × Heartland F_2 individuals, representing 68 recombinant gametes, were genotyped via PCR genotype by sequencing. The data were used for single marker regression mapping placing the mutation on chromosome 3H within an approximate 75 cM interval encompassing the 3H CI5791 resistance QTL. Sequencing of the mutants and wild-type (WT) CI5791 genomic DNA following exome capture identified independent mutations of the *HvWRKY6* transcription factor located on chromosome 3H at ∼50.7 cM, within the genetically delimited region. Post transcriptional gene silencing of *HvWRKY6* in barley line CI5791 resulted in *Ptt* susceptibility, confirming that it functions in NFNB resistance, validating it as the gene underlying the mutant phenotypes. Allele analysis and transcript regulation of *HvWRKY6* from resistant and susceptible lines revealed sequence identity and upregulation upon

pathogen challenge in all genotypes analyzed, suggesting a conserved transcription factor is involved in the defense against the necrotrophic pathogen. We hypothesize that *HvWRKY6* functions as a conserved signaling component of defense mechanisms that restricts *Ptt* growth in barley.

Keywords: barley, disease resistance and susceptibility, *Pyrenophora teres* f. *teres*, net blotch of barley, mutants, exom capture, WRKY transcription factor

INTRODUCTION

A recent study determined that climate change is a major threat to malt barley production as yield loss is projected due to high temperatures and water deficiency in some growing regions (Xie et al., 2018). However, as climate change brings drought to some regions, others will experience excess precipitation and combined with elevated temperatures, will provide environments that are more conducive to fungal disease epidemics. Thus, without adequate management tools, disease problems can be expected to be exacerbated in these regions. The threat of greater disease epidemics due to rising temperatures in regions with excess precipitation was not accounted for in the predictions of world barley shortages (Xie et al., 2018), thus production shortfalls could be even greater than predicted. For sustainable barley production, the intelligent deployment of durable genetic resistance to important fungal pathogens is critical and is a major focus of breeding programs with the primary goal of releasing high yielding and broadly adapted varieties that produce quality grain across dynamic environments (Horsley et al., 2009).

The net blotch of barley (*Hordeum vulgare* L.), caused by *Pyrenophora teres* Drechs. is an economically important foliar disease in barley growing regions worldwide with epidemics causing 10–40% of yield loss when susceptible varieties are grown, but under environmental conditions conducive to disease epidemics losses can reach 100% (Mathre, 1997; Murray and Brennan, 2010). Because the brewing and distilling industries demand quality malting barley, which brings premium prices, producers are concerned by biotic or abiotic stresses that negatively affect yield and quality (Grewal et al., 2008). Foliar infection by *P. teres* is a major concern as it has a large impact on yield, but foliar and kernel infection can also impact quality (Liu et al., 2011). The most sustainable and environmentally friendly way to manage net blotch is deploying effective genetic resistance into varieties, yet, a better understanding of this complex pathosystem and the quantitative nature of the host pathogen genetic interactions is needed. Practicing the stewardship of effective resistance sources is important when deploying resistance so that resistance is durable and genes are conserved. One way of accomplishing this goal is through gene discovery and subsequent functional analysis which is more practical with the new array of genomic tools available to the barley research community.

Net blotch exists in two forms; net form net blotch (NFNB) caused by *Pyrenophora teres* f. *teres* (*Ptt*) and spot form net blotch (SFNB) caused by *Pyrenophora teres* f. *maculata* (*Ptm*) (Smedegård-Petersen, 1971; Steffenson and Webster, 1992). The symptoms of NFNB first appear as small dark brown necrotic lesions that expand over time forming net like longitudinal and transverse striated necrotic lesions commonly surrounded by chlorosis on susceptible host genotypes (**Figure 1A**). The symptoms of SFNB also initially appear as small dark brown necrotic lesions that expand over time producing elliptical necrotic lesions typically surrounded by chlorosis on susceptible host genotypes (**Figure 1B**). Although, these two pathogens are morphologically identical (conidia and mycelium), their genetics as well as host-pathogen interactions are considered to be relatively distinct (Liu et al., 2011), thus, are considered to be different diseases. However, contrary to this statement the recent review of barley-*Pyrenophora teres* genetic interactions showed that 17 of the 19 *Ptm* resistance/susceptibility loci known to date overlap with *Ptt* resistance/susceptibility loci (Clare et al., 2020). This may include the 3H QTL reported by Koladia et al. (2017), where they identified the *HvWRKY6* transcription factor required for NFNB resistance as there were also SFNB resistance QTL identified by biparental and association mapping within the region. Interestingly, we also generated data showing that the *HvWRKY6* gene described here is also required for *Ptm* resistance (data not presented).

The barley line CIho 5791 (hereafter referred to as CI5791), is an Ethiopian breeding line, that is highly resistant to most *Ptt* isolates collected from barley growing regions worldwide (Mode and Schaller, 1958; Steffenson and Webster, 1992; Wu et al., 2003; Richards et al., 2016; Koladia et al., 2017). CI5791 consistently exhibits a highly resistant reaction at the seedling stage (secondary leaf) mainly consisting of small pinpoint necrotic lesions (**Figure 1A**) and the resistance translates to the field at the adult plant stages. However, CI5791 resistance has been compromised by Canadian, French (Arabi et al., 1992; Akhavan et al., 2016), and Moroccan isolates that are moderately virulent on CI5791 (**Figure 1C**). Although, the dominant CI5791 resistance located on chromosome 6H is remarkably broad and effective, it is apparent that pathogen populations have the diversity to overcome this resistance.

Several studies mapped dominant and recessive NFNB resistance genes with different specificities to the centromeric region of barley chromosome 6H (Abu Qamar et al., 2008; Liu et al., 2015; Richards et al., 2016, 2017; Koladia et al., 2017), thus, this important NFNB resistance locus is considered complex, possibly harboring multiple dominant resistance genes and recessive susceptibility genes (Steffenson et al., 1996; Cakir et al., 2003; Wu et al., 2003; Friesen et al., 2006; Koladia et al., 2017). Koladia et al. (2017) also mapped two dominant resistance QTL contributed by CI5791 on chromosome 3H and 6H using a CI5791 × Tifang recombinant inbred line (RIL) population individually inoculated with nine geographically

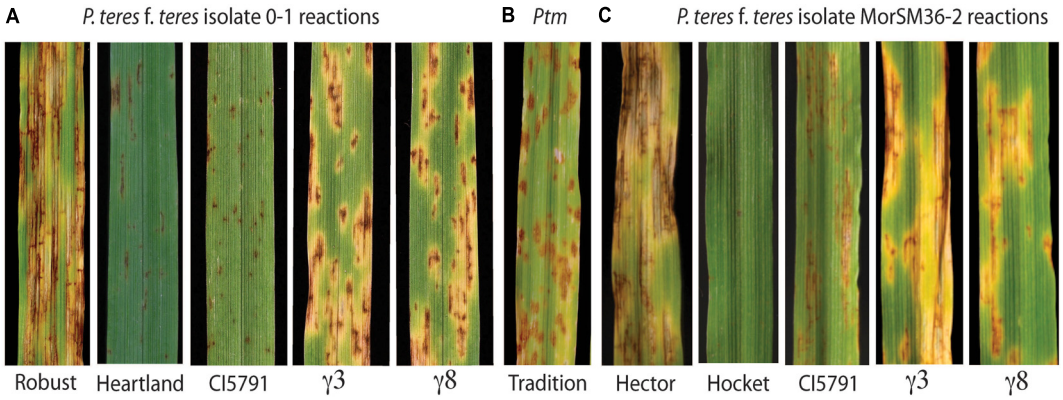


FIGURE 1 | Phenotypic reactions of wild-type CIho 5791, mutants, and resistant and susceptible checks inoculated with *P. teres*. **(A)** Phenotypic reaction of Robust, Heartland, CI5791, CI5791-γ3 (γ3), and CI5791-γ8 (γ8) to *Pyrenophora teres* f. *teres* (*Ptt*) isolate 0-1. **(B)** Phenotypic reaction of the barley cultivar Tradition to *P. teres* f. *maculata* (*Ptm*) isolate FGOB10Ptm-1. **(C)** Phenotypic reaction of Hector, Hockett, CI5791, CI5791-γ3 (γ3), and CI5791-γ8 (γ8) to *Ptt* isolate MorSM36-2. The disease was scored based on a 1–10 rating scale with 1 being highly resistant and 10 being highly susceptible.

distinct *Ptt* isolates. The major CI5791 6H QTL was shown to be dominantly resistant and was effective against all isolates used in the study. The gene underlying the CI5791 3H QTL was also dominant in nature, but was only effective against two Japanese *Ptt* isolates, JPT0101 and JPT9901. Interestingly, a similar chromosome 3H QTL was also contributed by Tifang and was shown to confer dominant resistance against *Ptt* isolates Br. Pteres (Brazil), BB06 (Denmark), 6A (California, United States), and 15A (California, United States).

Exome capture is a cost-effective yet powerful genomics tool that allows for targeted sequencing of the coding regions (exons) of specific specie genomes. This tool has been used for the identification of polymorphism within coding regions that contribute to disease in humans, other animals, and plants (Choi et al., 2009; Raca et al., 2010; Wang et al., 2010; Bamshad et al., 2011; Cosart et al., 2011; Mascher et al., 2013b, 2014, 2016; Warr et al., 2015; Russell et al., 2016). The exome capture array specific to barley (*Hordeum vulgare*), representing 61.6 mega bases of coding region from the complex ∼5.1 Gb barley genome, is available (Mascher et al., 2013b). This array was used to study the domestication and evolution of barley by resequencing and identifying variants in wild barley (*Hordeum sponteneum*) and land race (*Hordeum vulgare*) coding regions, including ancient barley germplasm (Mascher et al., 2016; Russell et al., 2016). Mascher et al. (2014) also identified the barley *HvMND* gene that governs increased tiller numbers utilizing this array. Schreiber et al. (2019) utilized the array on a highly mutagenized TILLING population of the barley variety Golden Promise to identify and evaluate mutation density. They also assembled a collection of semi-sterile mutants from the population and developed a custom exome capture array of 46 candidate genes to identify potential mutations causing the sterility phenotype (Schreiber et al., 2019). The exome capture methodology has also been applied to resistance gene enrichment sequencing (RenSeq), single-molecule real-time RenSeq (SMRT RenSeq), mutagenesis RenSeq (MutRenSeq), and association genetics (AgRenSeq) technology designed to identify nucleotide binding site-leucine rich repeat (NLR) disease resistance genes (*R* genes) in plants (Jupe et al., 2013; Steuernagel et al., 2016; Witek et al., 2016; Arora et al., 2019). A cDNA RenSeq method was also utilized to accelerate the identification of *R* genes in tomato (Andolfo et al., 2014).

The exome capture method coupled with forward genetics screens was considered an efficient genomics tool for the identification of resistance/susceptibility genes. Thus, we utilized this methodology to efficiently identify a gene required for broad and effective CI5791 NFNB resistance underlying the major QTL located on chromosomes 3H and 6H previously identified by Koladia et al. (2017). A CI5791 γ-irradiated mutant population was created and screened for individuals with compromised *Ptt* resistance. Utilizing a forward genetics screen, exome capture, and comparative sequence analysis, the *HvWRKY6* transcription factor (TF) gene underlying the CI5791 chromosome 3H *Ptt* resistance QTL was identified as being required for broad *Ptt* resistance. To the best of our knowledge this is the first gene identified that contributes to NFNB resistance in barley.

MATERIALS AND METHODS

Mutants Development

The Ethiopian barley line CI5791 is highly resistant to most NFNB isolates collected worldwide. A g-irradiation approach was used to develop a CI5791 mutant population. Briefly, ∼500 g of seed was hydrated in an airtight container with 60% glycerol for about 7–10 days. The hydrated seeds were irradiated with 35 kilorads (350 Gy) of g rays in a Gammator (M38-4, Radiation Machinery Corporation) prior to planting. Approximately, 1,400 M_1 seeds were planted in trays and allowed to self-generate the M_2 generation. Approximately 10,000 M_2 seedlings, derived from the original 1,400 M_1 individuals were screened by inoculation with the *Ptt* isolate LDNH04Ptt19 (hereafter referred to as LDN) collected from Langdon, North Dakota. Planting, inoculum preparation, inoculation, and disease evaluations were performed

as described in Friesen et al. (2006). After identifying putative mutants, these seedlings were transplanted to 15.24 cm (6 inch) pots and allowed to self-generate M_3 generation seeds. The M_3 generation seeds were planted in cone-tainers and screened with *Ptt* isolate 0-1 using three replicates with WT CI5791 and Heartland as the resistant checks and the susceptible check Robust. *Ptt* isolate 0-1 is a Canadian isolate collected from Ontario (Weiland et al., 1999; Wyatt et al., 2018) and has similar virulence as the isolate LDN which was used to identify the original CI5791-γ3 and CI5791-γ8 mutants from the M_2 generation. The planting, inoculum preparation, inoculation, and disease reading were performed as described in Friesen et al. (2006).

Mapping Populations and Phenotyping

Two F_2 mapping populations were developed by crossing CI5791-γ3 and CI5791-γ8 homozygous mutant M_3 individuals with the NFNB resistant barley line Heartland. Heartland is a six-rowed spring feed barley that was developed at the Agriculture Canada Research Station, Brandon, Manitoba and registered and released in 1984 (Therrien and Wolfe, 1985). Heartland was shown to be resistant to three major Canadian races of *Ptt* before its release. Heartland is hypothesized to contain a similar dominant resistance gene as CI5791 at the chromosome 6H locus designated as *Rpt5*. The planting, inoculum preparation, inoculation, and disease reading for the F_2 individuals from each of the CI5791-γ3 and CI5791-γ8 × Heartland populations were performed as described in Friesen et al. (2006). The disease reading was performed 7 days after inoculation (DAI) using a 1–5 rating scale (Neupane et al., 2015) for CI5791-γ3 × Heartland due to the phenotypic resemblance to *Ptm* infection and a 1-10 rating scale developed by Tekauz (1985) for CI5791-γ8 × Heartland. The CI5791-γ3 × Heartland F_2 susceptible individuals using a rating cutoff of > 2 representing the homozygous CI5791 genotype at the mutant region were used for mapping the gene. We utilized PCR genotyping by sequencing (PCR-GBS) to genotype CI5791-γ3 × Heartland F_2 homozygous susceptible lines (a total of 34 lines representing 68 recombinant gametes) on an Ion Torrent™ PGM: a PCR-GBS marker panel designed for polymorphism between CV. Tradition and barley line PI67381 consisting of 365 markers (**Supplementary Table 1**) was used to genotype all 34 susceptible F_2 lines. Primer development, DNA extraction, PCR cycle parameters, library preparation, and sequencing on the Ion Torrent™ PGM were performed as previously described in Richards et al. (2016).

The disease severity of the 34 CI5791-γ3 × Heartland F_2 homozygous susceptible lines along with the genotypic data were used for QTL mapping using MapDisto 2.0 (Heffelfinger et al., 2017) and QGene 4.0 (Joehanes and Nelson, 2008). The individual SNP calls were filtered for a minimum genotype quality of 10, and a minimum read depth of 3. The markers with more than 30% missing data and MAF $< 25\%$ were removed from further analysis. Single marker regression was used to identify the susceptible QTL in the γ3 × Heartland F_2 population. CI5791 and the two mutants were also phenotyped with the two Moroccan *Ptt* isolates SM36-2 and SM36-3 that were shown to be moderately virulent on CI5791. The spot form net blotch susceptible barley line Tradition was inoculated with *Ptm* isolate FGOB10Ptm-1 as previously described in Neupane et al. (2015) to compare a typical SFNB phenotype with the NFNB phenotypes on the CI5791-γ3 and CI5791-γ8 mutants (**Figure 1**).

Allelism Tests

Reciprocal crosses were made between CI5791-γ3 and CI5791-γ8 to determine if the two putative independent mutants were allelic. Six F_1s of CI5791-γ3 × CI5791-γ8 and ten F_1s of CI5791-γ8 × CI5791-γ3 were phenotyped as previously described using the 1–10 rating scale developed by Tekauz (1985) with *Ptt* isolate 0–1. The F_1 individuals from both crosses were also genotyped utilizing the STS markers targeting the putative mutant gene underlying the region delimited by genetic mapping. The primers

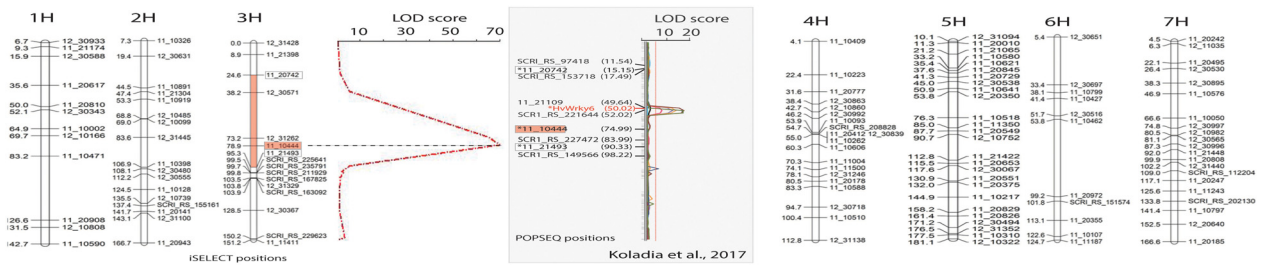


FIGURE 2 | Linkage map developed using 34 CI5791-γ3 × Heartland F_2 susceptible individuals (representing 68 recombinant gametes) developed with 123 polymorphic SNP markers showing the seven barley chromosomes. For chromosome 3, the QTL map of resistance/susceptibility to *Pyrenophora teres* f. *teres* isolate 0–1 using single marker regression analysis showing the only significant peak (red dashed line) with a LOD score of 71. The *X*-axis represents LOD values and the *Y*-axis represents the PCR-GBS SNP markers. The most significant marker was 11_10444 (red boxes) positioned at 78.9 cM with iSelect positions used to develop the map and 74.99 cM with POPSEQ positions with a LOD value of 71. The red region filled in on the chromosome 3 map flanked by the markers 11_20742 and 11_21493 (white boxes) shows the high confidence interval region containing the CI5791-γ3 mutation. The gray box shows the comparative region mapped by Koladia et al. (2017) with the flanking (white boxes) and most significant (red box) markers from our mapping shown relative to their map. The *HvWRKY6* gene is shown within the QTL detected in their CI5791 × Tifang biparental mapping population. Markers and the *HvWRKY6* gene marked with an asterisk were not mapped by Koladia et al. (2017) but were placed on the map based on their POPSEQ positions for comparison. The markers without an asterisk were mapped by Koladia et al. (2017) but were given POPSEQ positions.

used for STS marker development are described below under the STS marker development and mutation validation section.

Exome Capture Sequencing and Analysis

The exome capture, sequencing, and analysis of the CI5791-γ3 and CI5791-γ8 mutants and CI5791 WT were performed as thoroughly described in Solanki et al. (2019). The POPSEQ positions of the markers flanking the QTL identified in the segregating F_2 population (**Figure 2**), described above, were obtained and used to identify exome capture targets within the mapped region. BAM files from the analysis were imported into CLC Genomics Workbench version 8.0.3 (Qiagen) for the visualization of sequence alignments (**Figure 3**).

STS Marker Development and Mutation Validation

Based on the identified nucleotide deletion in the CI5791-γ8 mutant specific sequence tagged site (STS), markers was developed. Oligonucleotide primer pairs were developed specific to CI5791 WT and the CI5791-γ8 mutants. Primers WRKY6-F1 (5′-GCCGCTGGTTCTCGTCG TTCATGCG-3′) and WRKY6-Wt-R1 (5′-TAGTCGACGACGACGGGGCGTCCC-3′) only produced an amplicon from CI5791 WT (**Figure 3**) whereas the primer combination of WRKY6-F1 and WRKY6-Mt-R2 (5′-TAGTCGACGACGACGGGGCGT CCG-3′) only produced an amplicon from the CI5791-γ8 mutant due to designing specificity in the 3 bases at the 3′ terminus of the primer that are specific to the 1 bp deletion discovered in the CI5791-γ8 mutant from the exome capture data. The PCR was optimized so the discriminant amplicons were specific to the WT or mutant genotypes. The PCR amplification program was set as: denaturation at 95°C for 5 min, 25 cycles of 95°C for 30 s, 76°C for 1 min, and 76°C for 30 s, and a final extension of 72°C for 5 min. Wild-type CI5791, Heartland, CI5791-γ3, CI5791-γ8, homozygous susceptible F_2 individuals from two populations, and 15 randomly selected resistance F_2 lines (CI5791-γ8 × Heartland) were genotyped with the WT and mutant specific primers. The F_1 reciprocal cross between CI5791-γ3 and CI5791-γ8 were also genotyped with these primer sets. All PCR amplicons were visualized on a 1% agarose gel with GelRed® (Biotium).

HvWRKY6 Allele Sequencing and Analysis

To compare if there was the presence of any allelic variation between resistant and susceptible barley cultivars, we sequenced

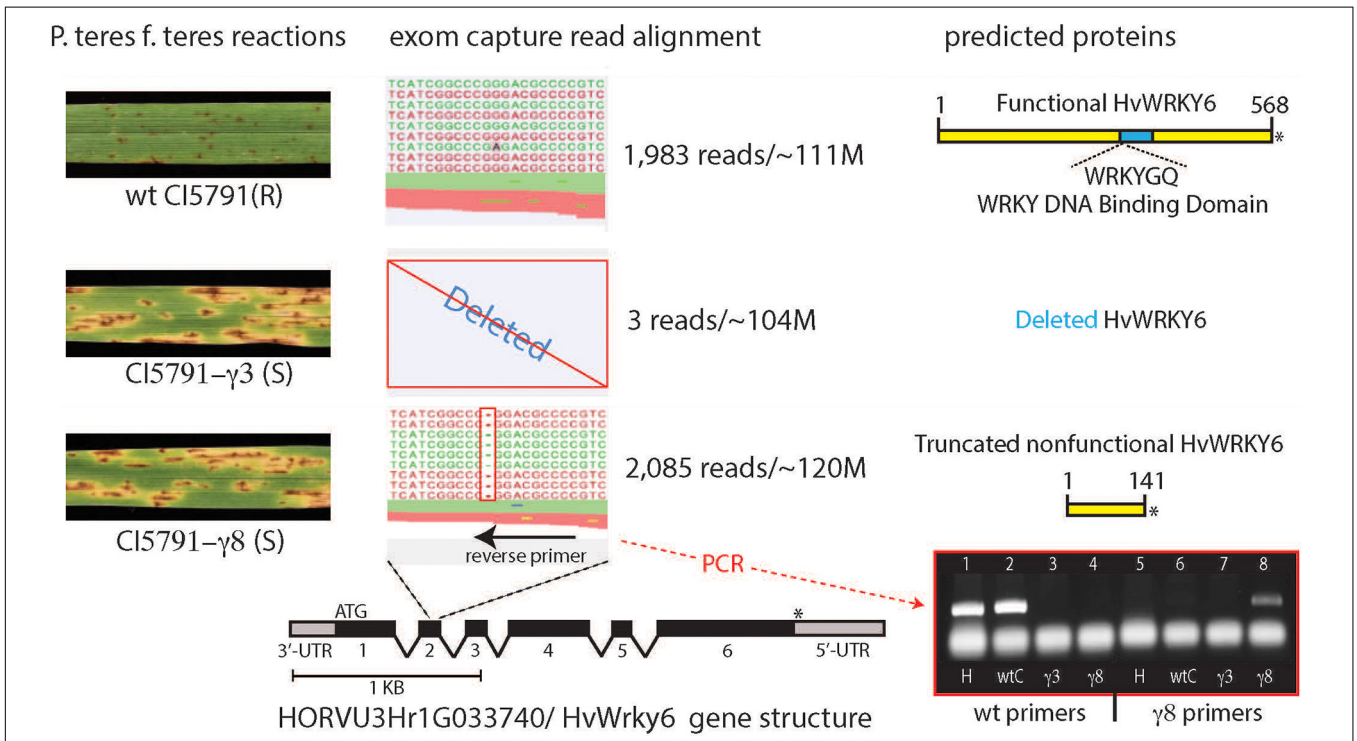


FIGURE 3 | Alignment of exome capture reads from wild-type (WT) CI5791, and the CI5791-γ3 and CI5791-γ8 mutants to the HORVU3Hr1G033740.2 (*HvWRKY6*) gene model from the cultivar Morex genome sequence. The left shows the reaction of WT CI5791, CI5791-γ3, and CI5791-γ8 to *Pyrenohpora teres* f. *teres* isolate 0-1. The center shows the pile up of reads from each genotype to the HORVU3Hr1G033740.2 (*HvWRKY6*) gene model. The alignment shows only three reads in CI5791-γ3, and ∼2,000 reads in both CI5791-γ8 and WT CI5791 confirming the complete gene deletion in CI5791-γ3. The CI5791-γ8 pileup shows a single bp deletion in the 2nd exon. Reverse PCR primers were developed from the deletion region that produced WT specific amplicons and CI5791-γ8 specific amplicons as shown in the bottom right. The WT specific primers amplified only from Heartland and WT CI5791. The mutant specific primers amplified only from CI5791-γ8 and no amplification in CI5791-γ3 as it is missing the entire gene sequence. The right side shows the predicted proteins encoded by the *HvWRKY6* gene in each genotype.

HvWRKY6 from CI5791 (resistant), cvs. Tifang (susceptible), and Morex (moderately susceptible). We designed four primer pairs at 1 kb intervals to sequence the entire gene including the promoter region (∼3,544 bp) (**Supplementary Table 2**). The gDNA extractions were performed as previously described in the "DNA Extraction, Exome Capture, and Sequencing" section above and were quantified using the Qubit™ 2.0 Fluorometer with a Qubit™ dsDNA Broad Sensitivity Kit (Thermo Fisher Scientific). PCR parameters were initial denaturation at 95°C for 5 min, 35 cycles of 94°C for 30 s, 60°C for 60 s, and 72°C for 60 s, followed by a final extension at 72°C for 5 min. PCR amplicons were visualized on a 1% agarose gel containing GelRed® (Biotium) and purified using an E.Z.N.A.® Cycle Pure centrifugation columns (Omega Bio-tek) following the manufacturer's protocol. Purified PCR products, ∼40 ng, were sent to GenScript for sequencing following their guidelines.

RNA Extraction, cDNA Synthesis, and qPCR

Quantitative PCR (qPCR) was conducted to measure if differential regulation of the *HvWRKY6* gene occurred upon interaction with *Ptt* isolates in compatible or incompatible interactions. The isolates *Ptt* 0-1, SM36-2, and SM36-3 were used to inoculate CI5791 and the barley line Tifang was inoculated with *Ptt* 0-1 only. Primers were designed across exons 1 and exon 2 (**Figure 4**): wrky6-qpcr-F2 (5′-GTTCCTGCCGTTACTGTCCTCATC-3′) and wrky6-qpcr-R2 (5′-TCGCCATCAAGAAGGAGGACCTCAC-3′), that specifically amplify ∼120 bp from cDNA and ∼270 bp from gDNA. At least three biological replications were collected from each mock (water + tween 20) and *Ptt* inoculated plants. Tissues from the first leaves were collected at time point 0 (non-inoculated control), 5, 30 min, 1, 2, 4, 6, 12, 24, 48, 72, 96, 120, 144, and 168 h post inoculation. Tissue samples were immediately flash frozen in liquid nitrogen and stored at −80°C for further processing. Total RNA was extracted from the collected tissue using an RNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions. The total RNA was quantified using the Qubit™ Fluorometer and the Qubit™ RNA BR assay kit (Thermo Fisher Scientifics) per the manufacturer's instructions. To ensure RNA integrity and that the RNA was free of gDNA contamination, 1 μl of total RNA was denatured in 4 volumes of denaturing buffer (Formaldehyde Load Dye, Ambion) at 80°C for 5 min and visualized on a 1% agarose gel with GelRed® (Biotium). RNA samples with the four-intact ribosomal RNA (rRNA) bands at the expected molecular weights of ∼ 3.4, 1.8, 1.5, and 1.1 kb corresponding to the nuclear 28S and 18S rRNAs and the 23S and 16S plastid rRNAs, respectively, without high molecular weight gDNA contamination were considered as quality RNA and used for cDNA synthesis. The GoScript™ Reverse Transcription System (Promega) was used to synthesize cDNA following the manufacturer's protocol. Briefly, ∼1 μg of total RNA was mixed with oligo$(dT)_{15}$ primer (0.5 μg) and incubated at 70°C for 5 min. The RNA sample was then mixed with 15 μl of reverse transcription reaction mix (GoScript™ Reaction Buffer (5X), $MgCl_2$ (1.5 mM), PCR Nucleotide Mix (0.5 mM each dNTP), Recombinant RNasin Ribonuclease Inhibitor (20 units), and Reverse Transcriptase, and incubated at 25°C for 5 min followed by 42°C for 60 min and inactivated at 70°C for 15 min. The 20 μl cDNA synthesis reactions were diluted with 80 μl H_2O (1:5). A 10 μl qPCR reaction was prepared by mixing 4 μl of diluted cDNA, 5 μl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), and 0.5 μl of each forward and reverse primer (10 μM). The qPCR was conducted in a CFX96 Real-time system thermocycler (Bio-Rad) with cycling parameters of 95°C for 30 s followed by 40 cycles of 95°C for 15 s and 60°C for 30 s; 65°C for 30 s; and 60 cycles of temperature increasing from 60 to 95°C with fluorescence readings acquired at 0.5°C increments per cycle. Three technical replications were used for each biological rep. The barley *HvSnor14* gene was used as the reference (Ferdous et al., 2015) to normalize *HvWRKY6* gene expression. The efficiency of qPCR for the *HvWRKY6* and S*nor14* primers were calculated by generating a standard curve by running qPCR on a 10-fold serial dilution starting from 200 pg of the PCR amplified template of *HvWRKY6* and *Snor14*. Differential expression was calculated by using the ΔΔCT method on the Bio-Rad CFX Manager 3.1 software. A *t*-test was performed to check the significance of difference at $p < 0.05$ using a standard error of mean of 1.

BSMV-VIGS

The barley stripe mosaic virus-virus induced gene silencing (BSMV-VIGS) system was exploited to functionally validate the *HvWRKY6* as required for resistance in the barley line CI5791. A unique 65 bp sequence was selected from the *HvWRKY6* gene by performing a BLASTn search against the low and high confidence gene list in the IPK barley database[1] to reduce the cross amplification and off target silencing of other WRKY TF homologs in the barley genome. Two primer pairs based on the 5′ and 3′ termini of this unique sequence were designed with *Not*I and *Pac*I adaptor sequences attached to the 5′ ends of the respective primers. These adaptors were reciprocally utilized in order to develop sense and antisense constructs. The first primer set was designed with a *Not*I adaptor on the forward primer and *Pac*I adaptor on the reverse primer and the second set with the *Pac*I adaptor on the forward primer and *Not*I adaptor on the reverse primer.

First Primer Set

WRKY6_KD_NtFP1- GGAGCGGCCGCACGCCATGCC GCTAAACGTCG
WRKY6_KD_PcRP1- GGATTAATTAAGCCGGGCATC GGAACATGGAAC

Second Primer Set

WRKY6_KD_PcFP1-GGATTAATTAAACGCCATGCC GCTAAACGTCG
WRKY6_KD_NtRP1-GGAGCGGCCGCGCCGGGCAT CGGAACATGGAAC

[1]http://webblast.ipk-gatersleben.de/barley_ibsc/

These two primer sets were used to clone the unique 65 bp *HvWRKY6* fragment into the γRNA strand of the BSMV-VIGS infectious cDNA clone PSL38.1 in both sense and antisense orientations. First, the two primer sets were used to produce the gene specific amplicon from CI5791 cDNA in 20 μl PCR reactions consisting of 2 μl of cDNA template, 0.5 μl of each forward and reverse primers (10 μM), 0.3 μl of dNTPs (500 μM), 0.2 μl of GoTaq® (1.25 units), 4 μl of GoTaq® buffer (10x), and 12.5 μl of H_2O. The PCR cycle parameters had an initial denaturation at 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s followed by a final extension of 72°C for 5 min. The amplicon was purified using an E.Z.N.A® Cycle Pure centrifugation column (Omega Bio-tek). The purified PCR products was digested in a 30 μl reaction consisting of 0.5 μl of *Not*I HF (NEB), 0.5 μl of *Pac*I (NEB), 3 μl of Cut Smart Buffer (NEB), 11 μl of H_2O, and 15 μl of PCR product. The digestion reaction was allowed to incubate at 37°C for 2 h followed by inactivation at 65°C for 20 min. The BSMV vector PSL38.1-MCS for cloning the target amplicon was also digested with 3 units of *Not*I and *Pac*I double digestion reactions using 5 μg of plasmid in a 30 μl reaction. Digested PCR product (2 μl) was mixed thoroughly in an 8 μl ligation reaction mix comprised of 1 μl of predigested vector (∼80 ng), 1 μl of ligation buffer (10X), 1 μl of T4 DNA ligase, and 5 μl of H_2O and incubated at 4°C for 24 h. Chemically competent One Shot® TOP10 *E. coli* cells (Thermo Fisher Scientific) according to the manufacturer's protocol were then transformed with the ligation mix and inoculated into 250 μl of Luria Broth (LB) liquid media and incubated at 37°C with 230 rpm shaking for 1 h. A total of 100 μl of each transformation was plated onto LB agar plates with 100 μg/ml of ampicillin and incubated overnight (∼12 h) at 35°C. Ten random colonies were picked from each transformation and inoculated into 2 ml of LB broth [5 g of NaCl, 5 g of tryptone, 2.5 g of yeast extract, and 500 ml of H_2O and ampicillin (100 μg/ml)] in 12 ml borosilicate culture tubes and incubated overnight with shaking at 230 rpm at 37°C. The cell cultures were transferred to a 2-ml microcentrifuge tube and centrifuged at 12,000 rcfs for 5 min to pellet the cells and the waste supernatant was discarded. The plasmid DNA was extracted from the pelleted cells using the PureYield™ Plasmid miniprep System (Promega) following the manufacturer's protocol.

The BSMV tripartite viral genomic RNAs (α, β, and γ genomes) with the γ fragment containing the unique 65 bp fragments of the CI5791 *HvWRKY6* allele cloned in both the sense and antisense orientations, were synthesized via *in vitro* transcription using the mMESSAGE mMACHINE™ T7 Transcription Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Twenty microliter reactions of each of the α genome, β genome, γ-*HvWRKY6* sense, and γ-*HvWRKY6* antisense genomes were combined with 370 μl of FES buffer (100 ml of GP buffer, 5 g of sodium pyrophosphate decahydrate, 5 g of bentonite, 5 g of celite, and up to 500 ml of H_2O) as the BSMV-VIGS inoculum. A total of 20 μl of each of the α genome, β genome, and γ genome were combined with 390 μl FES buffer for the BSMV-VIGS control inoculum.

Single seeds of the barley line CI5791 were planted per cone-tainer and placed in racks. Newly emerged secondary leaves still whorled at the ∼10–11 days old seedling stage were inoculated with either 5 μl of each tripartite RNAs or BSMV-VIGS control virus (both in FES buffer). Approximately 40 individual plants were inoculated for each BMSV-VIGs experimental RNAs and control RNAs. Plants were first misted heavily and then inoculated by gently rubbing the leaves with 5 μl of each BSMV-VIGs construct. After incubation in the mist chamber for 24 h at 100% humidity, inoculated plants were moved back to the growth chamber set at 21°C with a 12 h photoperiod. Once typical BSMV symptoms, mottling and striping, appeared on the expanded or expanding tertiary leaves, plants were inoculated with *Ptt* isolate 0–1 as previously described in Friesen et al. (2006). Inoculum preparation, inoculation, and disease reading were performed as described before. CI5791 and Robust were used as a resistance and a susceptible check, respectively. Disease reading was performed 7 and 12 days after *Ptt* inoculation (**Figure 4** and **Supplementary Tables 3, 4**) using the 1–10 scale developed by Tekauz (1985).

mRNA Extraction, RNAseq Library Preparation, and Sequencing

Three leaf samples of equal size (∼2 cm) from each replicate per treatment, non-inoculates and inoculated with *Ptt* isolate LDN at 3, 21, and 45 h post pathogen inoculation (hpi), were combined in a single tube and flash frozen in liquid nitrogen then stored at −80°C until further processing for total RNA extraction. The total RNA was extracted from the frozen leaf samples using the RNeasy mini kit (Qiagen) following the manufacturer's standard protocol. RNA concentrations were measured using the Qubit® Broad Range RNA kit on a Qubit™ 2.0 Fluorometer (Thermo Fisher Scientific), and RNA samples were visualized on 1% agarose gels stained with GelRed® (Biotium) to confirm the integrity of the RNA samples. RNA samples with four sharp ribosomal RNA (rRNA) bands; approximate molecular weights of 3.4, 1.8, 1.5, and 1.1 kb corresponding to nuclear 28S and 18S rRNAs and 23S and 16S plastid rRNAs, respectively, without a high molecular weight genomic DNA contamination band were considered quality RNA. One microgram of total RNA was used for RNA sequencing (RNAseq). The library construction was performed using the TruSeq RNA Library Prep Kit v2 (Illumina) following the manufacturer's standard protocol. The final library was validated and quantified on the Agilent 2100 Bioanalyzer. The cDNA libraries from four different samples were pooled into one single tube and were normalized according to the manufacturer's protocol. Each of the library pools were diluted to a concentration of 1.8 pm and sequenced on the Illumina NextSeq® 500 sequencer on a single flow cell at the USDA Cereal Genotyping Centre, Fargo, ND, United States. The NextSeq® 500/550 High Output Kit v2 (150 cycles) was used for the generation of 150 bp single end reads. The raw sequence reads were demultiplexed and converted into individual FASTQ files using bcl2fastq software v2.17.1.14 (Illumina). The FASTQ reads were quality trimmed in CLC Genomics Workbench v8.0.3 (Qiagen) using default settings.

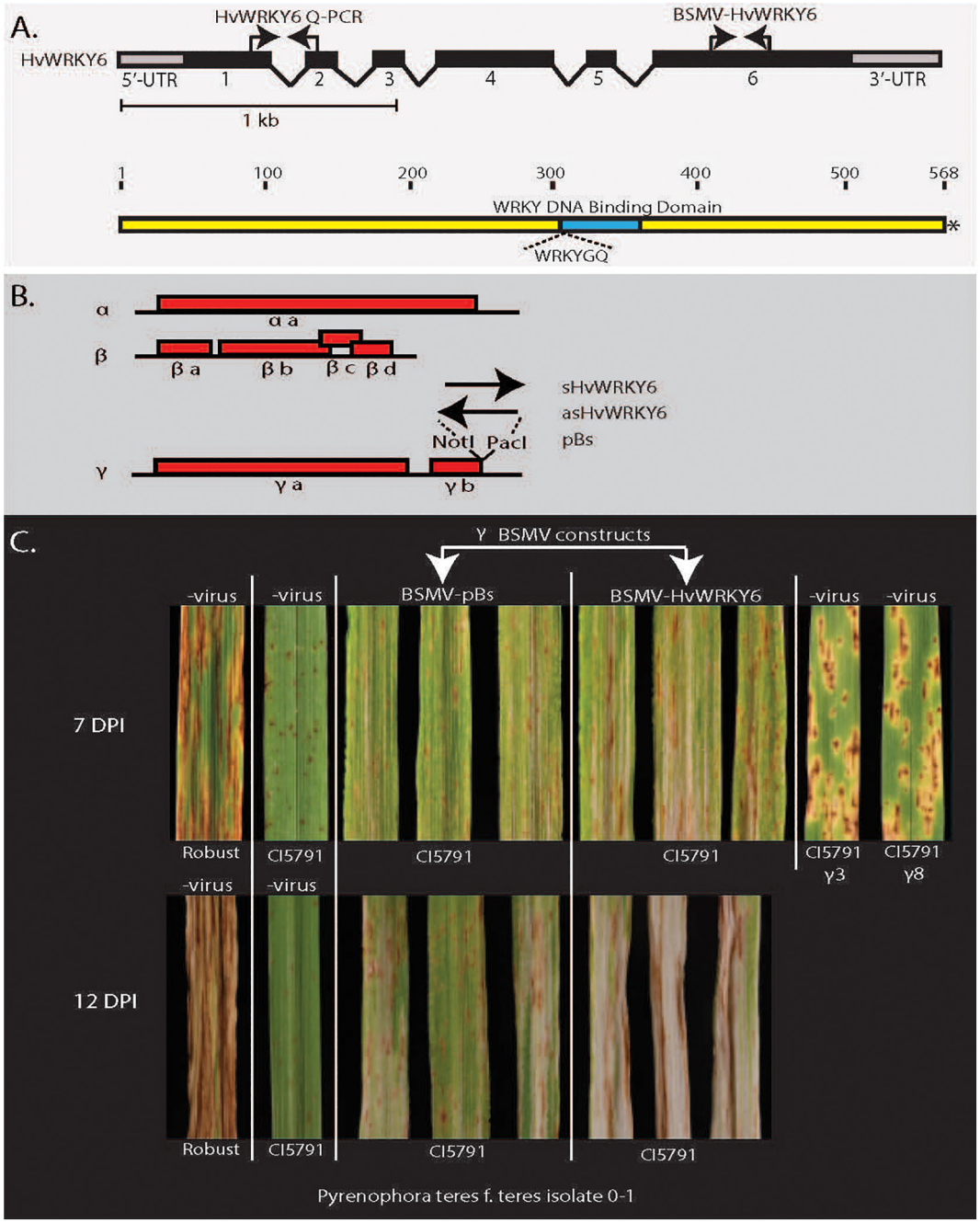


FIGURE 4 | Barley stripe mosaic virus-virus induces gene-silencing (BSMV-VIGS) validation of *HvWRKY6* function in *Pyrenophora teres* f. *teres* resistance in barley line CI5791. **(A)** The *HvWRKY6* transcription factor gene and protein structure showing location of primers used for qPCR and BSMV-VIGS constructs. Black bars at the top represent the intron/exon structure of the *HvWRKY6* gene with black bars indicating exons and the gray terminal bars indicating the 5′ and 3′ untranslated regions (UTRs). **(B)** The red bars below show the barley stripe mosaic virus genomic RNA with red boxes indicating viral genes encoded from the positive stranded RNA virus genome. The black arrows indicate where the sense and antisense *HvWRKY6* fragments were inserted into the γ genome at the *Not*I and *Pac*I restriction sites of the infectious BSMV *E. coli* plasmid pSL38.1 to develop *HvWRKY6* post transcriptional gene silencing constructs. **(C)** The BSMV-VIGS experiments showing that the specific silencing of the candidate *HvWRKY6* gene results in susceptible reactions when inoculated with *P. teres* f. *teres* isolate 0–1 at 7 and 12 days post inoculation. The BSMV-VIGS pBs vector control does not show the shift from resistance toward susceptibility.

Expression Analysis

The analysis pipeline for mapping reads to the reference genome, quality check, and for expression analyses was performed by first mapping the high quality trimmed sequencing reads to the barley RefSeq v1.0 (see text footnote 1) in CLC Genomics Workbench v8.0.3 (Qiagen). Gene specific and transcript specific reads were

obtained from reference genes as well as from the gene track and mRNA tract information. This enabled reads to align to both intronic and intergenic regions. Reads less than 90% identical for 90% of the read length and that mapped to more than 10 positions were discarded. The total reads mapped for each gene model were normalized to obtain reads per kilobase of exon model per million mapped reads (RPKM) values for each sample. In all the comparisons, the false discovery rate (FDR)-corrected *p*-values were calculated by the exact test using the EdgeR bioconductor package in CLC Genomics Workbench. Analyses were based on a threshold of 0.05 for FDR-corrected *p*-value and a fold change of 3. All treatments were compared with 0 h control (no *Ptt* inoculation).

RESULTS

Mutant Identification and Validation

To identify mutants compromised for *Ptt* resistance seeds of the highly resistant barley line CI5791 was γ-irradiated and ∼10,000 M_2 seedlings derived from ∼1,400 M_1 individuals were used in a forward genetics screen for their reaction to *Ptt* isolate LDN. Two mutant individuals designated CI5791-γ3 and CI5791-γ8 were identified and advanced to the M_3 generation. The two mutant lines were confirmed by phenotyping M_3 individuals in replicated trials and shown to express similar phenotypes resembling susceptible SFNB reactions when inoculated with *Ptt* isolate 0–1 the causal agent of NFNB (**Figures 1A,B**). The CI5791 WT and the resistant variety Heartland, which was used to develop the mutant mapping populations described below, exhibited highly resistant reactions (pin point necrotic lesions) to *Ptt* isolate 0–1 with an average infection type (IT) of 1.5 based on the 1–10 NFNB rating scale developed by Tekauz (1985) (**Figure 1A** and **Supplementary Tables 1, 5**). Phenotyping CI5791-γ3 and CI5791-γ8 showed average ITs of 6.5 and 6.0, respectively with *Ptt* isolate 0–1 using the NFNB 1–10 scale (Tekauz, 1985; **Figure 1A** and **Supplementary Table 5**). The susceptible barley variety Robust exhibited typical NFNB susceptible infection types (**Figure 1A**). Inoculation of CI5791 WT with the Moroccan *Ptt* isolates SM36-2 and SM36-3 exhibited moderately susceptible ITs averaging 4.5 and 3.5, but the mutants were susceptible showing average disease scores of 6.5 and 6 for CI5791-γ3 and 7.0 and 6.5 for CI5791-γ8, respectively (**Figure 1C** and **Supplementary Table 6**). The variety Hockett, exhibited highly resistant reactions to *Ptt* isolates SM36-2 and SM36-3, with average ITs of 1 (**Figure 1C** and **Supplementary Table 6**). The susceptible barley variety Hector exhibited highly susceptible ITs with an average disease score of 9.5 and 8.0 to isolates SM36-2 and SM36-3, respectively (**Figure 1C** and **Supplementary Table 6**).

CI5791-γ3 and CI5791-γ8 were crossed with cultivar Heartland and F_1 individuals were allowed to self-pollinate to produce the F_2 populations. The CI5791-γ8 × Heartland F_2 populations contained 116 individuals. The F_2 population was inoculated with *Ptt* isolate 0–1 and an IT of 4 was used as the cutoff for resistance/susceptibility with the CI5791-γ8 × Heartland F_2s on the 1–10 NFNB scale (Tekauz, 1985). The F_2 population showed a segregation ratio not significantly different from a 3 resistant: 1 susceptible ratio as would be expected for a single recessive mutant gene (**Supplementary Tables 1, 5, 7**). Since the F_2 phenotyping data determined that a single recessive mutation was responsible for the susceptible phenotype in the CI5791-γ8 mutant, reciprocal crosses between CI5791-γ3 and -γ8 were made to determine if the mutations were allelic. Six CI5791-γ3 × -γ8 and ten CI5791-γ8 × -γ3 F_1 individuals were challenged with *Ptt* isolate 0–1 and all the F_1 individuals showed susceptible reactions that were similar to each of the mutant parental lines with average scores of 6.45 and 6.25, respectively (**Supplementary Table 3**).

Mutant Mapping

Thirty-four F_2 individuals from a CI5791-γ3 × Heartland population showing the characteristic SFNB-like susceptible lesions indicating that they were homozygous for the mutated gene, were genotyped using PCR-GBS to identify 123 polymorphic SNP markers spread across the seven barley chromosomes (**Figure 2**). The QTL mapping utilizing the genotyping of the 34 CI5791-γ3 × Heartland F_2 homozygous susceptible lines, representing 68 recombinant gametes, localized the mutation to chromosome 3H within a high confidence ∼75 cM interval flanked by the SNP markers 11_20742 (POPSEQ position; chr = 3H cM = 15.15) and 11_21493 (POPSEQ position; chr = 3H cM = 90.33) (**Figure 2**). The most significant marker 11_10444 (POPSEQ position; chr = 3H cM = 74.99) had a LOD score of 71 (**Figure 2**).

Exome Capture Sequencing and Analysis

Sequencing of CI5791 WT, CI5791-γ3, and CI5791-γ8 gDNA enriched via exome capture on an Illumina NextSeq® flow cell resulted in a total of 111,251,482, 103,796,564, and 120,530,567 reads, respectively. Thus, the parallel sequencing of the three exome-captured genotypes represented a balanced sequencing library. Utilizing the gene models underlying the ∼75 cM interval containing the mutant gene, deletion variant analysis identified a 1 bp cytosine deletion in the predicted coding region of the HORVU3Hr1G033740.2 gene model in CI5791-γ8 (**Figure 3**). Of the ∼120 million sequence reads from the CI5791-γ8 exome capture, 2,085 reads were aligned to the HORVU3Hr1G033740.2 gene model. The single base deletion in gene model HORVU3Hr1G033740.2 was the only deletion identified within the chromosome 3H region containing the mutant gene (**Figure 2**). The single base deletion in HORVU3Hr1G033740.2 is predicted to be within the second predicted exon, of the barley ortholog of the Arabidopsis WRKY transcription factor 6 gene, designated *HvWRKY6*. Coverage analysis showed that *HvWRKY6* is completely deleted from the CI5791-γ3 mutant as only three sequence reads from the CI5791-γ3 mutant were mapped to the HORVU3Hr1G033740.2 reference sequence gene model (**Figure 3**). Considering the balance of the multiplexed sequencing library which yielded 111,251,482, 103,796,564, and 120,530,567 reads for CI5791 WT, CI5791-γ3, and CI5791-γ8, respectively, and the numbers of reads that mapped to *HvWRKY6* for CI5791 WT (1,983) and CI5791-γ8 (2,085), the three reads that mapped to CI5791-γ3 is well below the threshold of contamination (**Figure 3**).

Characterization of the Candidate Gene

The variant analysis of the exome capture data pinpointed a single cytosine base deletion at nucleotide position 545 in relation to the adenosine of the start methionine as base 1 of the barley *HvWRKY6* gene model (**Figure 3**). The mutation in the CI5791-γ8 mutant occurs in the second exon, which resulted in a frame shift and the predicted translation of a non-functional 141 amino acid (aa) truncated protein (**Figure 3**). Read depth analysis showed that *HvWRKY6* was completely deleted from the CI5791-γ3 mutant as only three reads out of ∼104 million mapped to the HORVU3Hr1G033740.2 gene model compared to 1,983/∼111 million for CI5791 WT and 2,085/∼120 million for CI5791-γ8 (**Figure 3**). The *HvWRKY6* gene spans 8,026 bp of gDNA localized to barley chromosome 3H at ∼50.03 cM (**Figure 2**) based on POPSEQ positions (Mascher et al., 2013a). *HvWRKY6* is predicted to transcribe a 1,710 nucleotides mRNA consisting of six exons (**Figure 3**) predicted to encode a 569 aa functional protein (∼59.67 kDa) containing WRKY transcription factor domains including the highly conserved WRKYGQK DNA binding motif (**Figures 3**, **4**). Homology searches utilizing NCBI BLASTp identified the candidate HORVU3Hr1G033740.2 predicted protein as an ortholog of the *Arabidopsis* WRKY transcription factor 6, thus, was designated *HvWRKY6*. The predicted HvWRKY6 protein has 50% aa identity and 59% aa similarity with the Arabidopsis WRKY6 protein (query cover 89% and *e*-value $4e^{-130}$) (**Figure 5**). A reciprocal result was obtained when the *AtWRKY6* protein was used as the query in a BLASTp search using the IPK barley blast server, identifying only one matching WRKY protein in the barley genome corresponding to HORVU3Hr1G033740.2. Thus, *HvWRKY6* represents the only known *AtWRKY6* ortholog in the barley genome. InterProScan SMART domain identified a conserved WRKY domain at 300-360 aa in *HvWRKY6* with high confidence prediction[2] (Letunic et al., 2015). Analysis of the full length *HvWRKY6* gene sequence from CI5791, Morex, and Tifang were identical (**Supplementary Figure 1**) suggesting that the gene is conserved for its primary aa sequence across both resistant and susceptible barley genotypes. BLAST analysis of the barley pan-genome (Jayakodi et al., 2020) showed 100% sequence identity for the lines Akashinriki, Barke, Golden Promise, Hockett, HOR 3081, HOR 3365, HOR 7552, HOR 8148, HOR 9043, HOR 10350, HOR 13821, HOR 13942, HOR 22559, Igri, OUN333, RGT Planet, ZDM02064 (Jayakodi et al., 2020; Schreiber et al., 2020), Bowman (IBGSC, 2012), Haruna Nijo (Sato et al., 2016), and Lasa Goumang (Zeng et al., 2020). Whereas, ZDM01467 (Jayakodi et al., 2020) contained two concatenated hits of highly identical sequences suggesting a mis-assembly of the region. Zangqing320, a Tibeten hulless variety (Dai et al., 2018) and B1K-04-12, a wild barley (*H. vulgare* subsp. *spontaneum,* Jayakodi et al., 2020) both show 99.9% sequence identity. Two additional wild *hordeum* lines *H. bulbosum* and *H. pubiflorum* (Mascher et al., 2013b) showed, 96.8 and 95.4% sequence identity, respectively, further exemplifying the high conservation of HvWRKY6 in *hordeum*. In addition, *WRKY6* appears to be more highly conserved amongst monocots (wheat and *Brachypodium distachyyon*) than dicots (*Arabidopsis* and soybean, **Figure 5**).

[2]http://smart.embl.de/

Validation of HvWRKY6 Function in CI5791 NFNB Resistance

The primer pair, WRKY6-F1 and WRKY6-Mt-R2, produced a CI5791-γ8 mutant specific *HvWRKY6* amplicon utilizing 3′-terminus specificity to the single nucleotide deletion based on the WRKY6-Mt-R2 primer sequence. The CI5791-γ8 mutant specific *HvWRKY6* primers did not produce an amplicon from the gDNA of WT CI5791, Heartland, or CI5791-γ3 (complete gene deletion) (**Figure 3**). The wild-type specific primer pair (WRKY6-F1 + WRKY6-Wt-R1) produced *HvWRKY6* specific amplicons from WT CI5791 and Heartland and did not produce amplicons from CI5791-γ3 or CI5791-γ8 mutants (**Figure 3**).

All homozygous susceptible F_2 individuals from both the CI5791-γ3 × Heartland and CI5791-γ8 × Heartland populations showed a mutant *HvWRKY6* genotype with CI5791-γ8 mutant specific primers (WRKY6-F1 + WRKY6-Mt-R2; **Figure 3**) or WT specific primers. This was determined by no observed amplification with either primer pair on the CI5791-γ3 × Heartland homozygous susceptible F_2 individuals, which is consistent with the entire gene deletion detected with the exome capture experiment (**Figure 3**). With the CI5791-γ8 × Heartland homozygous susceptible F_2 individuals there were amplicons produced with the mutant specific primer pair (WRKY6-F1 + WRKY6-Mt-R2) but no observed amplification with the WT specific primer pair, which is consistent with the 1 bp deletion detected with the exome capture experiment showing that they were homozygous mutant individuals (**Figure 3**). Fifteen randomly selected resistant F_2 individuals from the CI5791-γ8 × Heartland showed a 1 homozygous: 2 heterozygous genotype segregation (data not shown). This genotyping perfectly linked the genetic mutation with the mutant phenotype in this small F_2 population representing 68 recombinant gametes. Also, the genotypes of all the reciprocal F_1 (CI5791-γ3 × -γ8 or CI5791-γ8 × -γ3) individuals had a CI5791-γ8 mutant like genotype, lacking a WT allele, providing further evidence that the two mutants CI5791-γ3 and CI5791-γ8 are allelic.

BSMV-VIGS

To validate the function of *HvWRKY6* in NFNB resistance the barley stripe mosaic virus (BSMV) tripartite genome was utilized for post-transcriptional gene silencing constructs (**Figure 4**). The disease reactions of the BSMV-*WRKY6* inoculated plants targeted for the post-transcriptional gene silencing of *HvWRKY6* were significantly more susceptible than the BSMV-*pBS* virus inoculated controls at both 7 and 12 days post inoculation when inoculated with *Ptt* isolate 0–1 (**Figure 4** and **Supplementary Tables 4, 8, 9**). The BSMV-*pBs* virus control inoculations did not show the shift from resistance toward susceptibility.

qPCR and RNAseq

The qPCR experiment conducted on WT CI5791 inoculated with the *Ptt* isolate 0–1 showed that *HvWRKY6* is upregulated at 4 h post inoculation (hpi) at least 5-fold until 6 hpi, then it gradually

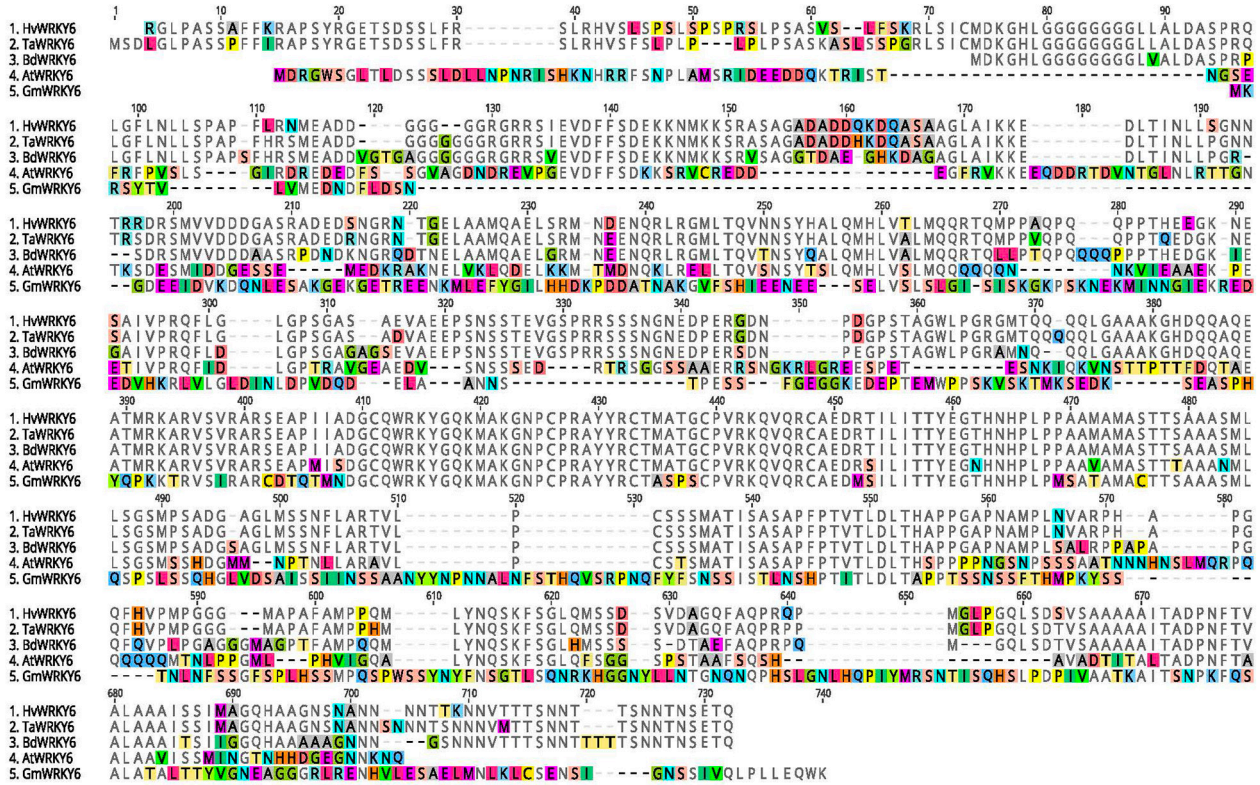


FIGURE 5 | Protein sequence alignment of barley (*Hv*), wheat (*Ta*), *Brachypodium distachyon* (*Bd), Arabidopsis thaliana* (*At*), and soybean (*Gm*). Amino acids are highlighted when in disagreement with the consensus sequence. All five sequences and hyphens represent indels. Protein alignments were generated with Clustal Omega 1.2.2 (Sievers et al., 2011) plugin of Geneious Prime 2020.2.2 (https://www.geneious.com).

decreases and maintains a level of ∼1-fold upregulation until 168 hpi (**Figure 6**). With the Moroccan *Ptt* isolate SM36-3, which is moderately virulent on CI5791, *HvWRKY6* was upregulated 1 hpi by 1.6-fold and increased to 12.6-fold upregulation at 4 hpi, 4-fold upregulation at 6 hpi, 16-fold at 12 hpi, and maintained at least 5-fold upregulation after 96 through 168 hpi (**Figure 6**). When susceptible cultivar Tifang was challenged with *Ptt* isolate 0-1, *HvWRKY6* was upregulated 2.8-fold at 30 min post inoculation to 22-fold at 6 hpi, 4.8-fold at 12 hpi, 9-fold at 24 hpi, and maintained at least 8-fold upregulation after 96 through 168 hpi (**Figure 6**). The susceptible line Tifang was utilized in this study due to the fact that the 3H locus encompassing HvWRKY6 (**Figure 2**) was identified as a resistance QTL in a CI5791 x Tifang cross (Koladia et al., 2017). The qPCR analyses confirmed that the expression of *HvWRKY6* in CI5791 was significantly higher with the moderately virulent *Ptt* isolate SM36-3 than the avirulent isolate 0–1 from 1 to 4 h, and 96 to 168 h. Similarly, the expression of *HvWRKY6* in the susceptible line Tifang challenged with the virulent *Ptt* isolate 0–1 was much higher and significantly different than SM36-3 and 0–1 on CI5791 at the times between 30 min through 2 and 96 h through 168 h.

RNA sequencing was utilized for comparative global transcriptomics between non-inoculated and LDN-inoculated CI5791 at 3, 21, and 45 h post-inoculation. An upregulation of *HvWRKY6* (HORVU3Hr1G033740) in the inoculated samples was observed at all of the aforementioned time points (**Supplementary Table 10**) with *HvWRKY6* upregulated 7. 9-, 4. 6-, and 8.3-fold at 3, 21, and 45 h post-inoculation, respectively, in the CI5791 resistance response. As expected the gene enrichment analysis of differentially expressed genes at all the time points showed a significant enrichment of several classes of genes that are involved in the response to fungal pathogens (**Supplementary Table 11**).

DISCUSSION

Allelic deletion mutations that compromise NFNB resistance in barley line CI5791 were genetically mapped to an approximate 75 cM region of barley chromosome 3H. An exome capture mapping by sequencing approach identified independent mutations of the barley ortholog of the Arabidopsis *WRKY6* TF, designated *HvWRKY6,* in the delimited region. A 1 bp and whole *HvWRKY6* gene deletion in the two mutant lines CI5791-γ3 and CI5791-γ8 are consistent with the reports that γ-irradiation induce 1–10 kb deletions (Morita et al., 2009). The *HvWRKY6* gene was validated as playing a role in *Ptt* resistance via virus-induced gene silencing (VIGS) utilizing the barley stripe mosaic virus. Thus, we hypothesize that the *HvWRKY6*

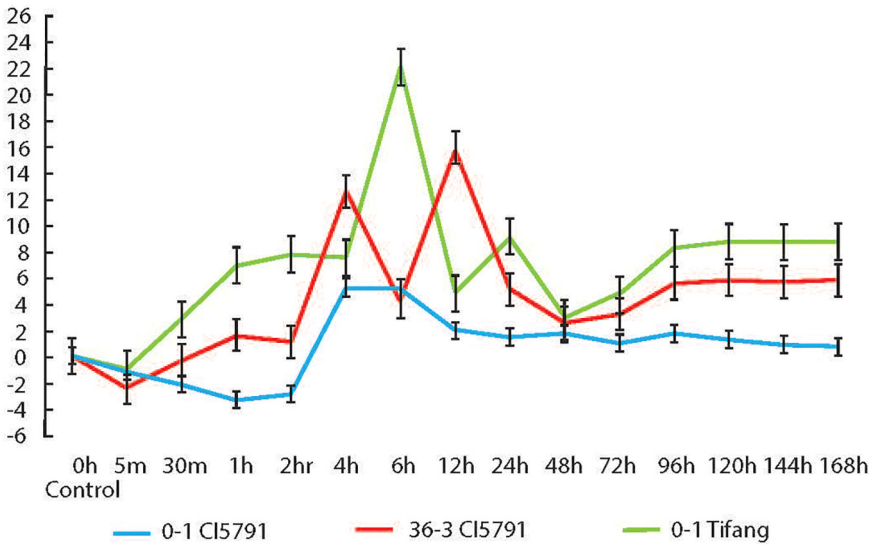


FIGURE 6 | Relative expression of *HvWRKY6* in resistant CI5791 inoculated with *Pyrenophora teres* f. *teres* (*Ptt*) isolate 0–1 (blue) and moderately virulent isolate SM36-3 (red); and the susceptible cv. Tifang inoculated with *Ptt* isolate 0–1 (green). The *Y*-axis represents the fold change relative to the non-inoculated control (time point 0) and the *X*-axis represents the time point at which the leaf samples were collected.

TF plays an important role in defense signaling that results in NFNB resistance.

The two independent mutants CI5791-γ3 and CI5791-γ8 exhibited susceptible symptoms to *Ptt* isolates that were not typical NFNB symptoms, but rather resemble SFNB lesions when inoculated with the *Ptt* isolates LDN and 0–1. The symptoms exhibited by the mutants were dark brown elliptical necrotic lesions surrounded by an extensive expanding yellow chlorotic margin (**Figure 1A**). The chlorosis expands and eventually coalesces with other lesions suggesting underlying pathogen growth, yet the necrotic regions remain relatively confined and elliptical, resembling an SFNB type of susceptible reaction (**Figures 1A,B**). Therefore, we initially phenotyped the CI5791-γ3 x Heartland F_2 population using a 1-5 SFNB rating scale as described in Neupane et al. (2015). However, the CI5791-γ8 × Heartland F_2 population, BSMV-VIGS experiment, and CI5791-γ3/CI5791-γ8 reciprocal allelism crosses were phenotyped at the F_1 stage using the NFNB 1–10 rating scale as described by Tekauz (1985) as this mutant compromises NFNB disease resistance yet results in SFNB-like symptoms. Interestingly, the mutant's symptoms when inoculated with the two Moroccan *Ptt* isolates, SM36-2 and SM36-3, resembled typical net type lesions with enlarged chlorosis that coalesced and has visible longitudinal and vertical striation but not as prominent as seen in a typical NFNB susceptible reaction. Thus, we speculate that the *HvWRKY6* transcription factor underlying the 3H QTL may function in regulating genes involved in restricting the growth of the pathogen. In a typical CI5791 resistance reaction the pathogen apparently penetrates the host as indicated by the formation of the pin point lesions, yet the pathogen growth is arrested early in the infection process and the lesion growth is effectively stopped at this early stage in the infection process. The CI5791 major dominant resistance gene responsible for this early resistance mechanism maps to the centromeric region of chromosome 6H (Koladia et al., 2017), which was the initial gene targeted in the mutant screening and was suspected to represent an immunity receptor. However, the first two mutants identified were allelic and mapped to the 3H QTL identified in the CI5791 x Tifang biparental mapping by Koladia et al. (2017) (**Figure 2**; Koladia et al., 2017) and disrupted the *HvWRKY6* transcription factor that plays a role in arresting pathogen colonization and spread after penetration. Additional studies have also reported large QLT intervals over chromosome 3H that spanned from 36.26 to 76.56 cM (Dinglasan et al., 2019). However, a smaller interval (51.27–51.77 cM) was consistently reported and designated as *qPttCLS,* that maps physically to 398.2–435.5 Mbp on the Morex reference genome (Mascher et al., 2017) and therefore is not suspected to encompass *HvWRKY6*. Several QTLs have been mapped to the 3H region, one at 28.7–36.6 cM (Steffenson et al., 1996), *QTLUHs-3H* at 29–31 cM (König et al., 2014), QTL_{UH}*-3H* at 45–51 cM (König et al., 2013), and *Rpt-3H-4* at 57.0–66.6 cM (Yun et al., 2005) that may all be in close proximity, although these cannot be directly compared due to differing populations and absence of common marker sets. The locus, *NBP_QRptt3-2,* identified using GWAS spans 49.65-52.03 cM using POPSEQ positions and physically maps to 160.7–491.8 Mbp (Wonneberger et al., 2017) and therefore includes *HvWRKY6*. Additional studies have also reported significant markers in close physical and genetic proximity to *HvWRKY6* (Burlakoti et al., 2017; Islamovic et al., 2017; Koladia et al., 2017; Daba et al., 2019; Novakazi et al., 2019; Rozanova et al., 2019; Vatter et al., 2017).

Necrotrophic pathogens often produce several host specific necrotrophic effectors (NE) including low molecular weight metabolites and small secreted proteins that interact with dominant host susceptibility genes (Wolpert et al., 2002;

Liu et al., 2011, 2012, 2015; Stergiopoulos et al., 2013; Shjerve et al., 2014). These interactions often follow the inverse-gene-for-gene model (Friesen et al., 2007) triggering programmed cell death (PCD) to facilitate necrotrophic fungal growth resulting in compatible interactions or a susceptible reaction called necrotrophic-effector triggered susceptibility (NETS) (Faris et al., 2010; Friesen and Faris, 2010; Liu et al., 2015; Faris and Friesen, 2020). *Ptt* is a necrotrophic pathogen that has been shown to produce a proteinaceous effector designated as PttNE1 that targets dominant susceptibility gene/s on chromosome 6H in barley (Liu et al., 2015) in an inverse gene-for-gene manner resulting in NETS. However, the CI5791 dominant resistance mechanism appears to follow the gene-for-gene model and possibly represents an R-gene-Avr gene interaction that results in an early dominant resistance response. The *HvWRKY6* gene appears to be a highly conserved TF that is most likely required for regulating early defense response genes post recognition that function to arrest pathogen spread after penetration. Thus, we hypothesize that it may be activated early in the response to the pathogen providing early resistance, which translates into preventing further proliferation of the fungus after penetration and thereby limiting the growth of the lesions. *Ptt* isolates SM36-2 and SM36-3 may express other effectors involved in virulence or have a variable Avr gene that evades early recognition and therefore activation of the early resistance signaling pathway, resulting in a moderately susceptible reaction in CI5791. Yet, this more prolific early pathogen growth and the chlorosis results in a higher level of susceptibility in the *HvWRKY6* mutants than in the WT, suggesting that the *HvWRKY6* transcription factor plays a role in regulating the early defense response genes that function to sequester pathogen spread after penetration thus inhibiting pathogen establishment in the host.

The qPCR analysis was performed because there appeared to be no polymorphism in the primary aa sequence of the *HvWRKY6* protein from a small number of resistant and susceptible genotypes analyzed, and BLAST analysis of 25 barley genotypes sequenced as part of a barley pan-genome project also showed 100% aa sequence identity for all but two of the *Hordeum vulgare* accessions sequenced suggesting that functional polymorphism between incompatible (resistant) and compatible (susceptible) interactions may be due to differential transcription. The qPCR analysis of HvWRKY6 was performed during infection processing. The qPCR data showed that the differential expression of *HvWRKY6* occurs in the barley line CI5791 in response to pathogen challenge as early as 1 hpi with the moderately virulent *Ptt* isolates SM36-3 and 4 hpi with the avirulent *Ptt* isolate 0-1. Similarly, in the susceptible line Tifang, the differential expression occurred as early as 30 min after inoculation and reached a maximum of 22-fold upregulation at 6 hpi. The line Tifang was utilized as the susceptible line in these analyses due to its use in the CI5791 x Tifang biparental population used to map the 3H QTL. Overall, the expression level in the susceptible cultivar Tifang with *Ptt* isolate 0-1 was significantly higher than CI5791 with 0-1 and SM36-3. The BSMV-VIGS experiment showed that the specific silencing of the candidate *HvWRKY6* gene results in a susceptible phenotype when inoculated with *Ptt* isolate 0-1. The time course RNA sequencing between non-inoculated and *Ptt* isolate LDN inoculated CI5791 seedlings showed upregulation of *HvWRKY6* in all the inoculated samples at all the time points (**Supplementary Table 10**). The gene enrichment analysis of differentially expressed genes at all the time points showed a significant enrichment and upregulation of several classes of genes that are involved in response to fungal pathogens including chitin responsive genes, salicylic acid (SA) and jasmonic acid (JA) response genes, and interestingly the upregulation of positive regulators of leaf senescence genes, which could play a role in suppressing the inverse gene-for-gene induction of PCD by this necrotrophic pathogen to facilitate disease (**Supplementary Table 11**).

The WRKY TFs are one of the largest groups of plant transcription regulators with protein domain architecture consisting of a highly conserved amino acid sequence (WRKYGQK) at the N-terminus and a zinc-finger-motif (C-C-H-H/C) at the C-terminus (Eulgem et al., 2000; Eulgem, 2006; Bakshi and Oelmüller, 2014). WRKY proteins bind specific W-box elements with the consensus sequence (TTGAC/T) at the promotor regions of targeted genes, resulting in transcriptional activation, and for some genes, repression (Eulgem et al., 2000; Yu et al., 2001; Rushton et al., 2010; Agarwal et al., 2011). WRKY TFs are important in diverse plant physiological activities such as pathogen defense responses and abiotic stress responses such as wounding, nutrient deficiency, salt stress (Chen et al., 2009; Kasajima et al., 2010; Cai et al., 2017; Hichri et al., 2017; Li et al., 2017), and developmental processes including senescence, and root growth (Robatzek and Somssich, 2001, 2002; Skibbe et al., 2008).

The *AtWRKY6* gene regulates both plant defense responses against *Pseudomonas syringae* pv. *tomato* as well as senescence in *Arabidopsis* (Robatzek and Somssich, 2002). *WRKY6* and *WRKY3* were also shown to regulate defense response in *Nicotiania attenuate* against the larvae of the insect herbivore *Manduca sexta* (Skibbe et al., 2008). In wheat, the *TaWRKY70* TF underlying the fusarium head blight QTL, *QTL-2DL,* governs resistance against *Fusarium graminearum* by regulating the downstream genes, *TaACT*, *TaDGK*, and *TaGLI,* which are involved in resistance responses (Kage et al., 2017). Other WRKY gene families have also been reported to play vital roles in defense responses in rice against *Magnaporthe grisea* (rice blast) and *Xanthomonas oryzea* (bacterial leaf blight) (Liu et al., 2007; Shimono et al., 2007; Wang et al., 2007). Li et al. (2004) reported enhanced resistance to the biotrophic fungal pathogen *Erysiphe cichoracearum,* whereas an increase in susceptibility to the bacterial necrotroph *Erwinia carotovora* subsp. *carotovora* occurred upon the upregulation of *WRKY70* in *Arabidopsis*.

The WRKY TFs also negatively regulate plant defense responses (Robatzek and Somssich, 2002; Eulgem and Somssich, 2007) as the overexpression of *WRKY38* and/or *WRKY62* were found to compromise immunity to the bacterial pathogen *P. syringae* and appeared to be negative regulators of plant basal defense responses (Mao et al., 2007; Kim et al., 2008). Grunewald et al. (2008) identified *WRKY23* as the negative regulator of plant defense responses against the cyst nematode *Heterodera schachtii.* The *WRKY11* and *WRKY17* TFs were also shown to function

as negative regulators of basal defense responses in *Arabidopsis* (Journot-Catalino et al., 2006).

The data generated in this study show that the *HvWRKY6* gene functions in NFNB resistance and likely plays a role in the activation of defense genes that are required to restrict lesion growth once the pathogen attempts to penetrate or after it has entered the host. This question will be answered once microscopy analysis is performed on the mutant vs. wild-type CI5791 across the infection process. The *HvWRKY6* gene is expressed at higher levels during later time points in compatible interactions showing that the upregulation of the TF does not correlate with resistance. However, earlier upregulation of *HvWRKY6* in the incompatible interaction may be the key to resistance, suggesting that the TF may mediate a defense response that is only effective when induced early in the host-pathogen interaction. Also, the later induced expression of *HvWRKY6* is not deterministic of resistance as this later upregulation also occurs in the compatible interactions which actually had higher induced levels of *HvWRKY6* expression across most of the time points tested (**Figure 6**). Interestingly, some transcription factors do act as a negative regulator of the plant basal defense response when highly expressed, which has been shown in previous studies (Li et al., 2004; Journot-Catalino et al., 2006; Mao et al., 2007; Kim et al., 2008; Xing et al., 2008). However, the loss of *Ptt* resistance in the mutants suggested a positive role of *HvWRKY6* as the disruption of the gene produces a predicted non-functional protein in one mutant and the complete deletion of the gene in the other showing a predominantly positive role of *HvWRKY6* in the NFNB resistance responses.

Interestingly, the *HvWRKY6* gene falls directly under the CI5791 dominant resistance QTL identified by Koladia et al. (2017). However, allele analysis of Morex, CI5791, and Tifang alleles did not reveal any primary functional polymorphism (**Supplementary Figure 1**) and the expression analysis data did not convincingly determine any expression polymorphism that would explain the differential resistances and susceptibilities. Thus, it is likely that the *HvWRKY6* TF is involved in a basal resistance and may not represent the gene that underlies the 3H dominant resistance QTL reported by Koladia et al. (2017) as there was little information explaining functional polymorphism in a CI5791 × Tifang population. However, the possibility cannot be ruled out that slight expression differences from a finely tuned level between genotypes could result in the mapping of the resistance/susceptibility QTL on chromosome 3H. Polymorphisms could exist within additional regulatory regions that result in altered expression levels between the resistant and susceptible host genotypes and represent the functional polymorphisms that were segregated in the CI5791 × Tifang population. The fact that all currently analyzed barley lines are predicted to be translationally identical but have diverse phenotypic responses suggest that an upstream signaling component that upregulates *HvWRKY6* to a threshold for this resistance that if exceeded or not induced at all results in susceptibility. However, further genetic and functional analyses will be required to answer these questions.

Several studies have provided evidence that WRKY TFs are an integral part of the plant immune system including roles in PAMP-triggered immunity, effector-triggered immunity, and systemic acquired resistance (Li et al., 2004; Eulgem and Somssich, 2007; Rushton et al., 2010). Certain WRKY DNA-binding factors serve as components of signal transduction pathways in plant cells in response to pathogens and regulate the expression of certain plant defense genes (Riechmann et al., 2000). *AtWRKY6* regulates both plant defense responses against *P. syringae* pv. *tomato* as well as senescence in *Arabidopsis,* which regulates the *SIRK* gene (Senescence-Induced Receptor like serine/threonine protein Kinase) that encodes a receptor-like kinase that is exclusively localized to the plant cell nucleus (Robatzek and Somssich, 2002). The qPCR and RNAseq data reported here determined that *HvWRKY6* was upregulated in response to pathogen challenge and similar to *Arabidopsis* may regulate senescence-induced genes as this class of genes was shown to also be upregulated in response to the pathogen in the highly resistant barley line CI5791. Since the *TaWRKY70* TF was identified as a strong candidate gene that regulates *TaACT*, *TaDGK*, and *TaGLI* in wheat resistance to *Fusarium graminearum* (Kage et al., 2017), we hypothesize that *HvWRKY6* may regulate other defense related genes that are required to restrict pathogen/lesion growth in line CI5791.

Technological advances in genomic tools and methodology allows for the more accurate characterization of complex traits including quantitative disease resistances. Resistance or lack of susceptibility to the necrotrophic specialist pathogens like *P. teres* are typically quantitative in nature and the genes underlying the resistance loci have been difficult to localize and isolate. Utilizing forward genetics, mapping by sequencing, exome capture, and next generation sequencing data, we identified the *HvWRKY6* gene located under a chromosome 3H resistance QTL that is required for the broad and remarkable NFNB resistance in barley line CI5791. To the best of our knowledge this is the first resistance gene or component required for NFNB resistance to be identified. We hypothesize that the *HvWRKY6* transcription factor positively functions to regulate defense response genes, which are required for resistance to *Ptt* in the barley line CI5791 by limiting the growth of the pathogen in the host after initial entry. The work here exemplifies the powerful new molecular tools that will help generate knowledge and resources to genetically improve crops, through marker assisted and genomic selection strategies in a more intelligent manner. Thus, expediting agricultural productivity in a sustainable manner.

DATA AVAILABILITY STATEMENT

The exome capture data is publicly available and has been deposited on NCBI under the BioProject ID PRJNA686994. The RNAseq data has been deposited under the BioProject ID PRJNA683717. Further inquires can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

RB, TF, RH, and AB designed the experiments. XL and JH generated the mutant population. PT identified the mutants.

PT, JR, SS, GA, PD, SC, and KE performed the wet lab experiments and analysis. JR and RS carried out the bioinformatics analysis. RB, PT, SC, KE, and AB wrote the manuscript. TF, JR, SS, SC, and GA proofread the manuscript. RB and AB obtained funding for research. All authors read and approved the final manuscript.

FUNDING

The research presented in this manuscript was supported by the USDA National Institute of Food and Agriculture Hatch project 1014919, Crop Improvement and Sustainable Production Systems (WSU reference 00011), the United States Department of Agriculture National Institute of Food and Agriculture award # 2018-67014-28491, and the National Science Foundation grant award # 1759030. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Agriculture National Institute of Food and Agriculture or the National Science Foundation.

ACKNOWLEDGMENTS

We would like to thank Danielle Holmes for technical assistance at the USDA-ARS, Cereal Crops Research Unit, Fargo, ND and Thomas Gross for technical assistance at North Dakota State University, Department of Plant Pathology, Fargo, ND.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.601500/full#supplementary-material

REFERENCES

Abu Qamar, M., Liu, Z. H., Faris, J. D., Chao, S., Edwards, M. C., Lai, Z., et al. (2008). A region of barley chromosome 6H harbors multiple major genes associated with net type net blotch resistance. *Theor. Appl. Genet.* 117, 1261–1270. doi: 10.1007/s00122-008-0860-x

Agarwal, P., Reddy, M. P., and Chikara, J. (2011). WRKY: its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. *Mol. Biol. Rep.* 38, 3883–3896. doi: 10.1007/s11033-010-0504-5

Akhavan, A., Turkington, T. K., Askarian, H., Tekauz, A., Xi, K., Tucker, J. R., et al. (2016). Virulence of *Pyrenophora teres* populations in western Canada. *Can. J. Plant Pathol.* 38, 183–196. doi: 10.1080/07060661.2016.1159617

Andolfo, G., Jupe, F., Witek, K., Etherington, G. J., Ercolano, M. R., and Jones, J. D. G. (2014). Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Plant Biol.* 14:120. doi: 10.1186/1471-2229-14-120

Arabi, M. I., Barrault, G., Sarrafi, A., and Albertini, L. (1992). Variation in the resistance of barley cultivars and in the pathogenicity of *Drechslera teres* f. sp. maculata and *D. teres* f. sp. teres isolates from France. *Plant Pathol.* 41, 180–186. doi: 10.1111/j.1365-3059.1992.tb02336.x

Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., et al. (2019). Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* 37, 139–143. doi: 10.1038/s41587-018-0007-9

Bakshi, M., and Oelmüller, R. (2014). WRKY transcription factors: jack of many trades in plants. *Plant Signal. Behav.* 9:e27700. doi: 10.4161/psb.27700

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., et al. (2011). Exome sequencing as a tool for mendelian disease gene discovery. *Nat. Rev. Genet.* 12, 745–755. doi: 10.1038/nrg3031

Burlakoti, R. R., Gyawali, S., Chao, S., Smith, K. P., Horsley, R. D., Cooper, B., et al. (2017). Genome-wide association study of spot form of net blotch resistance in the Upper Midwest barley breeding programs. *Phytopathology* 107, 100–108. doi: 10.1094/PHYTO-03-16-0136-R

Cai, R., Dai, W., Zhang, C., Wang, Y., Wu, M., Zhao, Y., et al. (2017). The maize WRKY transcription factor ZmWRKY17 negatively regulates salt stress tolerance in transgenic *Arabidopsis* plants. *Planta* 246, 1215–1231. doi: 10.1007/s00425-017-2766-9

Cakir, M., Gupta, S., Platz, G. J., Ablett, G. A., Loughman, R., Emebiri, L. C., et al. (2003). Mapping and validation of the genes for resistance to *Pyrenophora teres* f. *teres* in barley (*Hordeum vulgare* L.). *Aust. J. Agric. Res.* 54, 1369–1377. doi: 10.1071/AR02229

Chen, Y.-F., Li, L.-Q., Xu, Q., Kong, Y.-H., Wang, H., and Wu, W.-H. (2009). The WRKY6 transcription factor modulates PHOSPHATE1 expression in response to low Pi stress in *Arabidopsis*. *Plant Cell* 21, 3554–3566. doi: 10.1105/tpc.108.064980

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19096–19101. doi: 10.1073/pnas.0910672106

Clare, S. J., Wyatt, N. A., Brueggeman, R. S., and Friesen, T. L. (2020). Research advances in the *Pyrenophora teres*–barley interaction. *Mol. Plant Pathol.* 21, 272–288. doi: 10.1111/mpp.12896

Cosart, T., Beja-Pereira, A., Chen, S., Ng, S. B., Shendure, J., and Luikart, G. (2011). Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics* 12:347. doi: 10.1186/1471-2164-12-347

Daba, S. D., Horsley, R., Brueggeman, R., Chao, S., and Mohammadi, M. (2019). Genome-wide association studies and candidate gene identification for leaf scald and net blotch in barley (*Hordeum vulgare* L.). *Plant Dis.* 103, 880–889. doi: 10.1094/PDIS-07-18-1190-RE

Dai, F., Wang, X., Zhang, X.-Q., Chen, Z., Nevo, E., Jin, G., et al. (2018). Assembly and analysis of a qingke reference genome demonstrate its close genetic relation to modern cultivated barley. *Plant Biotechnol. J.* 16, 760–770. doi: 10.1111/pbi.12826

Dinglasan, E., Hickey, L., Ziems, L., Fowler, R., Anisimova, A., Baranova, O., et al. (2019). Genetic characterization of resistance to *Pyrenophora teres* f. *teres* in the international barley differential canadian lake shore. *Front. Plant Sci.* 10:326. doi: 10.3389/fpls.2019.00326

Eulgem, T. (2006). Dissecting the WRKY Web of plant defense regulators. *PLoS Pathog.* 2:e126. doi: 10.1371/journal.ppat.0020126

Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi: 10.1016/S1360-1385(00)01600-9

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Faris, J. D., and Friesen, T. L. (2020). Plant genes hijacked by necrotrophic fungal pathogens. *Curr. Opin. Plant Biol.* 56, 74–80. doi: 10.1016/j.pbi.2020.04.003

Faris, J. D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., et al. (2010). A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13544–13549. doi: 10.1073/pnas.1004090107

Ferdous, J., Li, Y., Reid, N., Langridge, P., Shi, B.-J., and Tricker, P. J. (2015). Identification of reference genes for quantitative expression analysis of microRNAs and mRNAs in barley under various stress conditions. *PLoS One* 10:e0118503. doi: 10.1371/journal.pone.0118503

Friesen, T. L., and Faris, J. D. (2010). Characterization of the wheat-Stagonospora nodorum disease system: what is the molecular basis of this quantitative

necrotrophic disease interaction?†. *Can. J. Plant Pathol.* 32, 20–28. doi: 10.1080/07060661003620896

Friesen, T. L., Faris, J. D., Lai, Z., and Steffenson, B. J. (2006). Identification and chromosomal location of major genes for resistance to *Pyrenophora teres* in a doubled-haploid barley population. *Genome* 49, 855–859. doi: 10.1139/g06-024

Friesen, T. L., Meinhardt, S. W., and Faris, J. D. (2007). The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. *Plant J.* 51, 681–692. doi: 10.1111/j.1365-313X.2007.03166.x

Grewal, T. S., Rossnagel, B. G., Pozniak, C. J., and Scoles, G. J. (2008). Mapping quantitative trait loci associated with barley net blotch resistance. *Theor. Appl. Genet.* 116, 529–539. doi: 10.1007/s00122-007-0688-9

Grunewald, W., Karimi, M., Wieczorek, K., Van de Cappelle, E., Wischnitzki, E., Grundler, F., et al. (2008). A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes. *Plant Physiol.* 148, 358–368. doi: 10.1104/pp.108.119131

Heffelfinger, C., Fragoso, C. A., and Lorieux, M. (2017). Constructing linkage maps in the genomics era with MapDisto 2.0. *Bioinformatics* 33, 2224–2225. doi: 10.1093/bioinformatics/btx177

Hichri, I., Muhovski, Y., Žižková, E., Dobrev, P. I., Gharbi, E., Franco-Zorrilla, J. M., et al. (2017). The Solanum lycopersicum WRKY3 Transcription Factor slwrky3 is involved in salt stress tolerance in tomato. *Front. Plant Sci.* 8:1343. doi: 10.3389/fpls.2017.01343

Horsley, R., Franckowiak, J., and Schwarz, P. (2009). "Barley," in *Cereals. Handbook of Plant Breeding*, Vol. 3, ed. M. Carena, (New York, NY: Springer), doi: 10.1007/978-0-387-72297-9_7

IBGSC (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491:711. doi: 10.1038/nature11543

Islamovic, E., Bregitzer, P., and Friesen, T. L. (2017). Barley 4H QTL confers NFNB resistance to a global set of *P. teres* f. *teres* isolates. *Mol. Breed.* 37:29. doi: 10.1007/s11032-017-0621-0

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., et al. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 588, 284–289. doi: 10.1038/s41586-020-2947-8

Joehanes, R., and Nelson, J. C. (2008). QGene 4.0, an extensible Java QTL-analysis platform. *Bioinformatics* 24, 2788–2789. doi: 10.1093/bioinformatics/btn523

Journot-Catalino, N., Somssich, I. E., Roby, D., and Kroj, T. (2006). The Transcription Factors WRKY11 and WRKY17 act as negative regulators of basal resistance in *Arabidopsis thaliana*. *Plant Cell* 18, 3289–3302. doi: 10.1105/tpc.106.044149

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 76, 530–544. doi: 10.1111/tpj.12307

Kage, U., Yogendra, K. N., and Kushalappa, A. C. (2017). TaWRKY70 transcription factor in wheat QTL-2DL regulates downstream metabolite biosynthetic genes to resist Fusarium graminearum infection spread within spike. *Sci. Rep.* 7:42596. doi: 10.1038/srep42596

Kasajima, I., Ide, Y., Yokota Hirai, M., and Fujiwara, T. (2010). WRKY6 is involved in the response to boron deficiency in *Arabidopsis thaliana*. *Physiol. Plant.* 139, 80–92. doi: 10.1111/j.1399-3054.2010.01349.x

Kim, K.-C., Lai, Z., Fan, B., and Chen, Z. (2008). *Arabidopsis* WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. *Plant Cell* 20, 2357–2371. doi: 10.1105/tpc.107.055566

Koladia, V. M., Faris, J. D., Richards, J. K., Brueggeman, R. S., Chao, S., and Friesen, T. L. (2017). Genetic analysis of net form net blotch resistance in barley lines CIho 5791 and Tifang against a global collection of P. teres f. teres isolates. *Theor. Appl. Genet.* 130, 163–173. doi: 10.1007/s00122-016-2801-4

König, J., Perovic, D., Kopahnke, D., and Ordon, F. (2013). Development of an efficient method for assessing resistance to the net type of net blotch (*Pyrenophora teres* f. *teres*) in winter barley and mapping of quantitative trait loci for resistance. *Mol. Breed.* 32, 641–650. doi: 10.1007/s11032-013-9897-x

König, J., Perovic, D., Kopahnke, D., and Ordon, F. (2014). Mapping seedling resistance to net form of net blotch (*Pyrenophora teres* f. *teres*) in barley using detached leaf assay. *Plant Breed* 133, 356–365.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949

Li, J., Brader, G., and Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell* 16, 319–331. doi: 10.1105/tpc.016980

Li, L.-Q., Huang, L.-P., Pan, G., Liu, L., Wang, X.-Y., and Lu, L.-M. (2017). Identifying the genes regulated by AtWRKY6 using comparative transcript and proteomic analysis under phosphorus deficiency. *Int. J. Mol. Sci.* 18:1046. doi: 10.3390/ijms18051046

Liu, X., Bai, X., Wang, X., and Chu, C. (2007). OsWRKY71, a rice transcription factor, is involved in rice defense response. *J. Plant Physiol.* 164, 969–979. doi: 10.1016/j.jplph.2006.07.006

Liu, Z., Ellwood, S. R., Oliver, R. P., and Friesen, T. L. (2011). *Pyrenophora teres*: profile of an increasingly damaging barley pathogen. *Mol. Plant Pathol.* 12, 1–19. doi: 10.1111/j.1364-3703.2010.00649.x

Liu, Z., Holmes, D. J., Faris, J. D., Chao, S., Brueggeman, R. S., Edwards, M. C., et al. (2015). Necrotrophic effector-triggered susceptibility (NETS) underlies the barley–*Pyrenophora teres* f. *teres* interaction specific to chromosome 6H. *Mol. Plant Pathol.* 16, 188–200. doi: 10.1111/mpp.12172

Liu, Z. H., Zhong, S., Stasko, A. K., Edwards, M. C., and Friesen, T. L. (2012). Virulence profile and genetic structure of a North Dakota population of *Pyrenophora teres* f. teres, the causal agent of net form net blotch of Barley. *Phytopathology* 102, 539–546. doi: 10.1094/PHYTO-09-11-0243

Mao, P., Duan, M., Wei, C., and Li, Y. (2007). WRKY62 transcription factor acts downstream of cytosolic NPR1 and negatively regulates jasmonate-responsive gene expression. *Plant Cell Physiol.* 48, 833–842. doi: 10.1093/pcp/pcm058

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043

Mascher, M., Jost, M., Kuon, J.-E., Himmelbach, A., Aßfalg, A., Beier, S., et al. (2014). Mapping-by-sequencing accelerates forward genetics in barley. *Genome Biol.* 15:R78. doi: 10.1186/gb-2014-15-6-r78

Mascher, M., Muehlbauer, G. J., Rokhsar, D. S., Chapman, J., Schmutz, J., Barry, K., et al. (2013a). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* 76, 718–727. doi: 10.1111/tpj.12319

Mascher, M., Richmond, T. A., Gerhardt, D. J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013b). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant J.* 76, 494–505. doi: 10.1111/tpj.12294

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., et al. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* 48:1089. doi: 10.1038/ng.3611

Mathre, D. E. (1997). *Compendium of Barley Diseases*, 2nd Edn. St. Paul, Minnesota: APS Press.

Mode, C. J., and Schaller, C. W. (1958). Two additional factors for host resistance to net blotch in barley. *Agron. J.* 50, 15–18.

Morita, R., Kusaba, M., Iida, S., Yamaguchi, H., Nishio, T., and Nishimura, M. (2009). Molecular characterization of mutations induced by gamma irradiation in rice. *Genes Genet. Syst.* 84, 361–370. doi: 10.1266/ggs.84.361

Murray, G. M., and Brennan, J. P. (2010). Estimating disease losses to the Australian barley industry. *Australas. Plant Pathol.* 39, 85–96. doi: 10.1071/AP09064

Neupane, A., Tamang, P., Brueggeman, R. S., and Friesen, T. L. (2015). Evaluation of a barley core collection for spot form net blotch reaction reveals distinct genotype-specific pathogen virulence and host susceptibility. *Phytopathology* 105, 509–517. doi: 10.1094/PHYTO-04-14-0107-R

Novakazi, F., Afanasenko, O., Anisimova, A., Platz, G. J. J., Snowdon, R., Kovaleva, O., et al. (2019). Genetic analysis of a worldwide barley collection for resistance to net form of net blotch disease (*Pyrenophora teres* f. *teres*). *Theor. Appl. Genet.* 132, 2633–2650. doi: 10.1007/s00122-019-03378-1

Raca, G., Jackson, C., Warman, B., Bair, T., and Schimmenti, L. A. (2010). Next generation sequencing in research and diagnostics of ocular birth defects. *Mol. Genet. Metab.* 100, 184–192. doi: 10.1016/j.ymgme.2010.03.004

Richards, J., Chao, S., Friesen, T., and Brueggeman, R. (2016). Fine mapping of the barley chromosome 6H net form net blotch susceptibility locus. *G3 Genes| Genomes| Genetics* 6, 1809–1818.

Richards, J. K. K., Friesen, T. L. L., and Brueggeman, R. S. S. (2017). Association mapping utilizing diverse barley lines reveals net form net blotch seedling resistance/susceptibility loci. *Theor. Appl. Genet.* 130, 915–927. doi: 10.1007/s00122-017-2860-1

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., et al. (2000). *Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes*. *Science* 290, 2105L–2110. doi: 10.1126/science.290.5499.2105

Robatzek, S., and Somssich, I. E. (2001). A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. *Plant J.* 28, 123–133. doi: 10.1046/j.1365-313X.2001.01131.x

Robatzek, S., and Somssich, I. E. (2002). Targets of AtWRKY6 regulation during plant senescence and pathogen defense. *Genes Dev.* 16, 1139–1149. doi: 10.1101/gad.222702

Rozanova, I. V., Lashina, N. M., Mustafin, Z. S., Gorobets, S. A., Efimov, V. M., Afanasenko, O. S., et al. (2019). SNPs associated with barley resistance to isolates of *Pyrenophora teres* f. *teres*. *BMC Genomics* 20:292. doi: 10.1186/s12864-019-5623-3

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006

Russell, J., Mascher, M., Dawson, I. K., Kyriakidis, S., Calixto, C., Freund, F., et al. (2016). Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat. Genet.* 48:1024. doi: 10.1038/ng.3612

Sato, K., Tanaka, T., Shigenobu, S., Motoi, Y., Wu, J., and Itoh, T. (2016). Improvement of barley genome annotations by deciphering the Haruna Nijo genome. *DNA Res.* 23, 21–28. doi: 10.1093/dnares/dsv033

Schreiber, M., Barakate, A., Uzrek, N., Macaulay, M., Sourdille, A., Morris, J., et al. (2019). A highly mutagenised barley (cv. Golden Promise) TILLING population coupled with strategies for screening-by-sequencing. *Plant Methods* 15:99. doi: 10.1186/s13007-019-0486-9

Schreiber, M., Mascher, M., Wright, J., Padmarasu, S., Himmelbach, A., Heavens, D., et al. (2020). A genome assembly of the barley 'Transformation Reference' cultivar golden promise. *G3 (Bethesda)* 10, 1823–1827. doi: 10.1101/2020.02.12.945550

Shimono, M., Sugano, S., Nakayama, A., Jiang, C.-J., Ono, K., Toki, S., et al. (2007). Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell* 19, 2064–2076. doi: 10.1105/tpc.106.046250

Shjerve, R. A., Faris, J. D., Brueggeman, R. S., Yan, C., Zhu, Y., Koladia, V., et al. (2014). Evaluation of a *Pyrenophora teres* f. *teres* mapping population reveals multiple independent interactions with a region of barley chromosome 6H. *Fungal Genet. Biol.* 70, 104–112. doi: 10.1016/j.fgb.2014.07.012

Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Skibbe, M., Qu, N., Galis, I., and Baldwin, I. T. (2008). Induced plant defenses in the natural environment: Nicotiana attenuata WRKY3 and WRKY6 coordinate responses to herbivory. *Plant Cell* 20, 1984–2000. doi: 10.1105/tpc.108.058594

Smedegård-Petersen, V. (1971). *Pyrenophora teres f. maculata* f. nov. and *Pyrenophora teres f. teres* on barley in Denmark. *Kgl. Vet. Landbohojsk. Arsskr* 124–144.

Solanki, S., Richards, J., Ameen, G., Wang, X., Khan, A., Ali, H., et al. (2019). Characterization of genes required for both *Rpg1* and *rpg4*-mediated wheat stem rust resistance in barley. *BMC Genomics* 20:495. doi: 10.1186/s12864-019-5858-z.

Steffenson, B. J., Hayes, P. M., and Kleinhofs, A. (1996). Genetics of seedling and adult plant resistance to net blotch (*Pyrenophora teres* f. *teres*) and spot blotch (*Cochliobolus sativus*) in barley. *Theor. Appl. Genet.* 92, 552–558. doi: 10.1007/BF00224557

Steffenson, B. J., and Webster, R. K. (1992). Pathotype diversity of *Pyrenophora teres* f. *teres* on barley. *Phytopathology* 82, 170–177.

Stergiopoulos, I., Collemare, J., Mehrabi, R., and De Wit, P. J. G. M. (2013). Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS Microbiol. Rev.* 37, 67–93. doi: 10.1111/j.1574-6976.2012.00349.x

Steuernagel, B., Periyannan, S. K., Hernández-Pinzón, I., Witek, K., Rouse, M. N., Yu, G., et al. (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 34, 652–655. doi: 10.1038/nbt.3543

Tekauz, A. (1985). A numerical scale to classify reactions of barley to *Pyrenophora teres*. *Can. J. Plant Pathol.* 7, 181–183. doi: 10.1080/07060668509501499

Therrien, M. C., and Wolfe, R. I. (1985). Heartland Barley. *Can. J. Plant Sci.* 65, 445–446. doi: 10.4141/cjps85-060

Vatter, T., Maurer, A., Kopahnke, D., Perovic, D., Ordon, F., and Pillen, K. (2017). A nested association mapping population identifies multiple small effect QTL conferring resistance against net blotch (*Pyrenophora teres* f. *teres*) in wild barley. *PLoS One* 12:e0186803. doi: 10.1371/journal.pone.0186803

Wang, H., Chattopadhyay, A., Li, Z., Daines, B., Li, Y., Gao, C., et al. (2010). Rapid identification of heterozygous mutations in *Drosophila* melanogaster using genomic capture sequencing. *Genome Res.* 20, 981–988. doi: 10.1101/gr.102921.109

Wang, H., Hao, J., Chen, X., Hao, Z., Wang, X., Lou, Y., et al. (2007). Overexpression of rice WRKY89 enhances ultraviolet B tolerance and disease resistance in rice plants. *Plant Mol. Biol.* 65, 799–815. doi: 10.1007/s11103-007-9244-x

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. *G3 Genes Genomes Genetics* 5, 1543–1550. doi: 10.1534/g3.115.018564

Weiland, J. J., Steffenson, B. J., Cartwright, R. D., and Webster, R. K. (1999). Identification of molecular genetic markers in *Pyrenophora teres* f. teres associated with low virulence on 'Harbin' barley. *Phytopathology* 89, 176–181. doi: 10.1094/PHYTO.1999.89.2.176

Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D., and Jones, J. D. G. (2016). Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* 34, 656–660. doi: 10.1038/nbt.3540

Wolpert, T. J., Dunkle, L. D., and Ciuffetti, L. M. (2002). Host-selective toxins and avirulence determinants: what's in a name? *Annu. Rev. Phytopathol.* 40, 251–285. doi: 10.1146/annurev.phyto.40.011402.114210

Wonneberger, R., Ficke, A., and Lillemo, M. (2017). Mapping of quantitative trait loci associated with resistance to net form net blotch (*Pyrenophora teres* f. *teres*) in a doubled haploid Norwegian barley population. *PLoS One* 12:e0175773. doi: 10.1371/journal.pone.0175773

Wu, H.-L., Steffenson, B. J., Zhong, S., Li, Y., and Oleson, A. E. (2003). Genetic variation for virulence and RFLP markers in *Pyrenophora teres*. *Can. J. Plant Pathol.* 25, 82–90. doi: 10.1080/07060660309507052

Wyatt, N. A., Richards, J. K., Brueggeman, R. S., and Friesen, T. L. (2018). Reference assembly and annotation of the *Pyrenophora teres* f. *teres* isolate 0-1. *G3 Genes| Genomes| Genetics* 8, 1–8.

Xie, W., Xiong, W., Pan, J., Ali, T., Cui, Q., Guan, D., et al. (2018). Decreases in global beer supply due to extreme drought and heat. *Nat. Plants* 4, 964–973. doi: 10.1038/s41477-018-0263-1

Xing, D.-H., Lai, Z.-B., Zheng, Z.-Y., Vinod, K. M., Fan, B.-F., and Chen, Z.-X. (2008). Stress- and pathogen-induced *Arabidopsis* WRKY48 is a transcriptional activator that represses plant basal defense. *Mol. Plant* 1, 459–470. doi: 10.1093/mp/ssn020

Yu, D., Chen, C., and Chen, Z. (2001). Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *Plant Cell* 13, 1527–1540. doi: 10.1105/tpc.010115

Yun, S. J., Gyenis, L., Hayes, P. M., Matus, I., Smith, K. P., Steffenson, B. J., et al. (2005). Quantitative trait loci for multiple disease resistance in wild barley. *Crop Sci.* 45, 2563–2572. doi: 10.2135/cropsci2005.0236

Zeng, X., Xu, T., Ling, Z., Wang, Y., Li, X., Xu, S., et al. (2020). An improved high-quality genome assembly and annotation of Tibetan hulless barley. *Sci. Data* 7:139. doi: 10.1038/s41597-020-0480-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Tamang, Richards, Solanki, Ameen, Sharma Poudel, Deka, Effertz, Clare, Hegstad, Bezbaruah, Li, Horsley, Friesen and Brueggeman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 21 January 2021
doi: 10.3389/fgene.2020.623736

Technological Innovations for Improving Cassava Production in Sub-Saharan Africa

*Edwige Gaby Nkouaya Mbanjo[1], Ismail Yusuf Rabbi[1], Morag Elizabeth Ferguson[2], Siraj Ismail Kayondo[1], Ng Hwa Eng[3], Leena Tripathi[2], Peter Kulakow[1] and Chiedozie Egesi[1,4,5]**

[1] International Institute of Tropical Agriculture, Ibadan, Nigeria, [2] International Institute of Tropical Agriculture, Nairobi, Kenya, [3] CGIAR Excellence in Breeding Platform, El Batan, Mexico, [4] National Root Crops Research Institute, Umudike, Nigeria, [5] Department of Global Development, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, United States

Cassava is crucial for food security of millions of people in sub-Saharan Africa. The crop has great potential to contribute to African development and is increasing its income-earning potential for small-scale farmers and related value chains on the continent. Therefore, it is critical to increase cassava production, as well as its quality attributes. Technological innovations offer great potential to drive this envisioned change. This paper highlights genomic tools and resources available in cassava. The paper also provides a glimpse of how these resources have been used to screen and understand the pattern of cassava genetic diversity on the continent. Here, we reviewed the approaches currently used for phenotyping cassava traits, highlighting the methodologies used to link genotypic and phenotypic information, dissect the genetics architecture of key cassava traits, and identify quantitative trait loci/markers significantly associated with those traits. Additionally, we examined how knowledge acquired is utilized to contribute to crop improvement. We explored major approaches applied in the field of molecular breeding for cassava, their promises, and limitations. We also examined the role of national agricultural research systems as key partners for sustainable cassava production.

Keywords: cassava, innovations, QTL mapping, association mapping, marker-assisted selection, genomic selection, genome-editing

OPEN ACCESS

Edited by:
Deepmala Sehgal, International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Zhi Zou, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, China
Jose Sebastian, Indian Institute of Science Education and Research Berhampur (IISER), India

****Correspondence:***
Chiedozie Egesi
C.Egesi@cgiar.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *30 October 2020*
Accepted: *23 December 2020*
Published: *21 January 2021*

Citation:
Mbanjo EGN, Rabbi IY, Ferguson ME, Kayondo SI, Eng NH, Tripathi L, Kulakow P and Egesi C (2021) Technological Innovations for Improving Cassava Production in Sub-Saharan Africa. Front. Genet. 11:623736. doi: 10.3389/fgene.2020.623736

INTRODUCTION

The agricultural sector is key to economic growth in Africa. The recent report on the Global Hunger Index indicates that over half of the world's food-insecure people live in Africa (FSIN, 2019). Sustainable agricultural production is imperative to curb food insecurity, reduce poverty, and impact the livelihood of smallholder farmers (Ojijo et al., 2016; Donkor et al., 2017). Cassava is among the six commodities defined by the African Heads of States as strategic crops for the continent, given its significant contribution to the livelihoods of African farmers and its potential for transforming African economies (Feleke et al., 2016).

Cassava (*Manihot esculenta* Crantz) is a root crop grown throughout the tropics by more than 800 million people (Nassar and Ortiz, 2010). It can grow with minimal inputs under marginal soil conditions and in regions prone to drought. Though mainly cultivated for its starchy roots,

nutrient-dense cassava leaves are also consumed as vegetables in many regions of Africa (Spencer and Ezedinma, 2017). Due to its long harvest window, cassava roots are used as a food reserve during periods of food shortage or during the lean season before harvest of other crops. Although its cultivation has traditionally been associated with subsistence farming, the crop is gradually becoming an industrial crop, which is processed into different products, including bread, pasta, and couscous-like products (Bechoff et al., 2018; Mtunguja et al., 2019). Apart from the food industry, cassava starch is used for textiles, the paper industry, in the manufacture of plywood and veneer adhesives, glucose and dextrin syrups (Tonukari et al., 2015; Spencer and Ezedinma, 2017; Waisundara, 2018).

Sub-Saharan Africa accounts for 61.1% of the world's cassava production (FAOSTAT, 2020). Although some increase have been achieved recently, mainly due to expansion in crop area, cassava average productivity (9.1 t/ha) on the continent is well below the average cassava fresh root yield (21.5 t/ha) recorded in Asia (Spencer and Ezedinma, 2017; FAOSTAT, 2020). In order to satisfy the forecasted increase in demand for cassava food and non-food products, and to harness the enormous potential offered by the crop, cassava production in sub-Saharan Africa must be increased (Khandare and Choomsook, 2019; Otekunrin and Sawicka, 2019; FAOSTAT, 2020). Some of the constraints inherent to its production include pests (Kalyebi et al., 2018; Koros et al., 2018) and diseases caused by bacteria (Fanou et al., 2018) and viruses (Patil et al., 2015; Alicai et al., 2019) leading to significant yield losses. Another drawback is the cyanogenic glucosides, which upon hydrolysis, produces toxic hydrogen cyanide (Akinpelu et al., 2011). Cassava is an energy-dense food, mainly composed of starch with low levels of protein and other nutrients important for a balanced diet. Populations that consume cassava as a staple are at high risk of protein, vitamin A, zinc, and/or iron deficiency (Gegios et al., 2010; Stephenson et al., 2010). These challenges guide the breeding objectives: (a) high yield in terms of dry matter per unit land area; (b) resistance to diseases, such as cassava mosaic disease (CMD), cassava brown streak disease (CBSD), and cassava bacteria blight (CBB), and pests such as cassava green mites (CGM) and whiteflies; (c) improved starch quality and quantity; (d) low hydrogen cyanide potential; (e) improved nutritional value or biofortification; (f) adaption to a wide range of environments; (g) improved plant type for mechanization; and (h) end user characteristics, including processing, cooking, and organoleptic properties (Teeken et al., 2020).

Conventional breeding has been efficient in providing a continuous supply of improved cultivars that have resulted in a dramatic increase in yield of most major crops (Prohens, 2011). Conventional cassava breeding is based on phenotype-based recurrent selection, which relies on the production of full-sib and/or half-sib progenies followed by successive clonal selection stages, including single row trials, preliminary, advanced, and uniform yield trials (Ceballos et al., 2016). Many cassava varieties have been developed and released through conventional breeding (Malik et al., 2020). Breeding cassava is a challenging task due to the heterozygous genetic make-up of the crop. The development of improved varieties is time consuming due to its long breeding cycle (12 months). New tools and technologies have the potential to improve the efficiency of conventional breeding, especially when several traits are being selected at the same time. Modernization of breeding programs, through the application of innovative tools, is vital for more efficient agriculture, especially in the context of climate change, shrinking resources, land scarcity, and increased food demand. Biotechnology and new genomic approaches have the potential to enhance genetic gain, speed up the development of better cultivars, and impact the livelihoods of smallholder farmers.

We highlighted the genomic resources available in cassava and their potential applications. We also reviewed the state of knowledge of the phenotyping technologies currently available and their shortcomings. Additionally, we explored the potential and implications for integration of technological innovations in cassava genetics and breeding. With farmers being the ultimate beneficiaries, we examined how to ensure regular, sustainable high level of cassava production in sub-Saharan Africa; thereby, contributing to food security challenges and improved livelihoods through income generation.

GENOMIC RESOURCES AND THEIR USE

Genomic resources for cassava have increased substantially in recent years. Thousands of simple sequence repeats (SSR) markers have been developed from expressed sequence tags and enriched genomic DNA libraries (reviewed in Ferguson et al., 2011). The cassava chromosome-scale reference genome and the re-sequencing of diverse accessions, has identified a large number of sequence polymorphisms. Single nucleotide polymorphisms (SNPs), methylation polymorphism, Insertions-Deletions (Indels), and structural variants have been identified (Sakurai et al., 2013; Wang et al., 2014; Xia et al., 2014; Bredeson et al., 2016). A cassava haplotype map harboring 25.9 million SNPs and 19 million indels has been developed (Ramu et al., 2017). High-throughput genotyping platforms (e.g., GoldenGate assay) have allowed researchers to simultaneously interrogate tens of thousands of SNPs at a reduced cost (Ferguson et al., 2012). The development of low cost genotyping technologies based on multiplex sequencing platforms, such as genotyping by sequencing, has enabled rapid and accurate high-density fingerprinting using SNP markers (Elshire et al., 2011; ICGMC, 2015; Rabbi et al., 2015; Kamanda et al., 2020)5). These resources provide valuable tools that have contributed to genetic research and molecular breeding of cassava.

Genetic Diversity

Genetic diversity is of paramount importance in crop improvement through breeding. The extent and nature of genetic variation existing within cassava landraces and cultivars from selected African countries have been assessed using molecular markers (Kawuki et al., 2009; Kamanda et al., 2020). These studies have revealed genetic differentiation between African and South American germplasm, as well as differentiation between cassava landraces within Africa. Those from East, South, and Central Africa are somewhat differentiated

from those from West Africa (Kawuki et al., 2009; Ferguson et al., 2019; Adjebeng-Danquah et al., 2020). The narrow genetic base of African cassava breeding lines is attributable to intense selection pressure for CMD resistance with recurrent selection using few parents and the clonal nature of the crop (Turyagyenda et al., 2012; Wolfe et al., 2016; Ferguson et al., 2019). Slightly higher genetic diversity has been reported in landraces in comparison to elite accessions (Ferguson et al., 2019), which is typical of most crops. It is likely that the genepool of pre-breeding germplasm will become more diverse as breeders begin to incorporate variation that responds better to consumer preferences.

Genetic Redundancy and Breeding Impact

Fingerprinting of cassava accessions using molecular markers has many applications. Genetic redundancy represents a challenge to efficiently manage and optimize the conservation of genetic resources in genebanks and breeders' collections. SNP markers have been used to confirm that particular accessions are not identical, and others are possible duplicates (Ferguson et al., 2012). They have also been used to assess the adoption of improved varieties (Turyagyenda et al., 2012; Rabbi et al., 2015; Wossen et al., 2017). Molecular markers have an important role to play as farmers frequently give different names to the same cultivar or landrace. Thus, they become difficult to identify, particularly as cassava varieties are not easy to distinguish morphologically. This enables the correct assessment of adoption rates, which in turn, influences breeding priorities and agricultural policies (Kretzschmar et al., 2018). Molecular markers have also been used to assess the integrity of putative mapping populations, select true-cross progeny, validate crosses, and guide parental selection (Rabbi et al., 2012; ICGMC, 2015; Masumba et al., 2017).

PHENOTYPING OF KEY TRAITS

Phenotyping is important to support crop breeding. The acquisition of phenotype data remains a bottleneck hindering cassava genetic studies and full-deployment of genomics-assisted breeding. Several approaches have been used to phenotype breeding lines and germplasm collections for nutrition (carotenoids, cyanogenic potential), yield and yield components (dry matter content), quality (starch physiochemical and functional properties, texture, and pasting properties), biotic stresses (disease resistance), and root system architecture. In the following section, we detail the current phenotyping strategies used in cassava.

Carotenoids

Breeding cassava roots with enhanced levels of provitamin A carotenoids is a high priority in some breeding programs. The color of the cassava storage root parenchyma has been correlated with the total carotenoids content; thus, used to evaluate carotenoids content (Iglesias et al., 1997; Afonso et al., 2017). A challenge has been to efficiently distinguish the subtle differences within each color group, as this is difficult by eye. To overcome this limitation, quantification of total carotenoids content by ultraviolet-visible spectrophotometry, as well as identification and quantification of β-carotene and its isomers by high-performance liquid chromatography (HPLC), have been employed (Carvalho et al., 2012; Belalcazar et al., 2016). These approaches are accurate, but have many drawbacks, including cost, time-needed for analysis, labor-intensive methods, and requirements for laboratory infrastructure and trained technical staff, which is not always available to breeding programs in Africa (Udoh et al., 2017). Alternative portable devices, such as ICheck^{TM} carotene, have been proven useful for rapid field evaluation and could be valuable in remote areas with no laboratory facilities or electricity (Esuma et al., 2016; Jaramillo et al., 2018). Color instruments designed to quantify the *Commission International de l'Eclairage* (CIELAB) color parameters have also been successfully used to evaluate carotenoids content in cassava root samples in an efficient way (Afonso et al., 2017). Near-infrared spectroscopy (NIRS) is another promising approach that has been explored for carotenoids quantification with demonstrated high prediction accuracy (Sánchez et al., 2014; Ikeogu et al., 2019).

Cyanogenic Potential

Cassava contains naturally occurring, but potentially toxic compounds called cyanogenic glycosides, which release hydrogen cyanide (HCN) on hydrolysis. This can be highly toxic to humans and animals if not removed through processing. It is important that the levels of cyanogenic glycosides are measured. Several approaches have been used to quantify cyanide potential, including the titration method (Moriasi et al., 2017; Iliya and Madumelu, 2019), the alkaline picrate method (Fukushima et al., 2016; Moriasi et al., 2017), and the metal-based chemosensors (Tivana et al., 2014). These approaches involved multi-step reactions and necessitate trained personnel. The picrate method, although easy to use, is very slow (minimum 12 h) and the chemicals used are hazardous. Recently, it was shown that NIRS could efficiently be used to distinguish roots with high or low cyanogenic potential (Sánchez et al., 2014).

Dry Matter Content

Storage root dry matter content (DMC) reflects the proportion of useable fresh root yield. DMC is commonly measured using either specific gravity through suspension of a root sample in water and air, or the oven-drying method, which has been the most widely used method. This is where a representative root sample is weighted wet and then oven dried to constant weight (Fukuda et al., 2010; Teye et al., 2011). The oven-drying method is tedious when working with a large number of samples. Likewise, it is difficult to implement this approach where the source of electricity is unreliable (Teye et al., 2011). Both oven-drying and specific gravity could be substituted by NIRS, which has

been shown to predict DMC with a high degree of accuracy (Belalcazar et al., 2016).

Starch Physiochemical and Functional Properties

Physiochemical properties among cassava cultivars determine root quality attributes important for processing and consumption. The main constituent of cassava storage roots is starch, which is composed of amylose and amylopectin. Both of these play a crucial role in retrogradation, gelling, pasting, crystallinity, gelatinization temperature, viscosity, texture, cooking, eating, and processing quality of cassava (Ayetigbo et al., 2018). Amylose content (AC) is the most important factor influencing cooking and textural quality. The AC in cassava has been estimated using iodine colorimetry (Sandoval-Aldana and Fernandez, 2013; Boonpo and Kungwankunakorn, 2017) or Megazyme amylose/amylopectin assay kit (Chisenga et al., 2019). Iodine colorimetry is prone to inter-laboratory variability due to the complexity of the procedure and relies on the development of a suitable curve of known amylose to-amylopectin ratios. The swelling power of starch determines its specific functional properties when utilized in food products (Noranizan et al., 2010). Swelling power and solubility patterns of cassava flour have been determined using the Leach (1959) and Kainuma et al. (1967) method, respectively (Chisenga et al., 2019; Ma'Aruf and Abdul, 2020). Starch gelatinization properties (onset, peak and conclusion gelatinization temperature, and enthalpy) and retrogradation have been determined using differential scanning calorimetry (DSC). DSC can be run at a rate of four samples per hour (Thirathumthavorn and Trisuth, 2008; Tappiban et al., 2020). Crystallinity measurement requires the use of an x-ray diffractometer, a piece of complex and expensive equipment (Chatpapamon et al., 2019). In sweet potato, it was shown that NIRS could predict most physiochemical and thermal properties of starch with acceptable precision (Lu et al., 2006a,b). NIRS was sufficiently accurate for the determination of total starch and amylose in barley in the study of Ping et al. (2013). Meanwhile, Cozzolino et al. (2013) demonstrated that swelling properties and water solubility could be determined in whole grain barley using NIRS spectroscopy. NIRS technology could potentially be used to predict functional and physiochemical properties of cassava or cassava-based products.

Texture and Pasting Properties

Texture is a critical factor for consumer acceptance of cassava. Sensory analysis has been used for characterizing cassava and cassava-based product texture properties. The sensory descriptors assessed included texture, appearance, odor, taste, masticability (Akely et al., 2016; Adinsi et al., 2019). The cost associated with training and maintaining a descriptive panel combined with the low throughput of sensory evaluations has prompted the development of less costly and less time-consuming approaches. Instrumental methods using a texture analyzer that mimics mastication have been used to evaluate and/or predict the texture of raw, cooked, and processed cassava products. The parameters measured include hardness, springiness, adhesiveness, gel strength, mouldability, elasticity, smoothness, appearance, thickness, and general acceptability (Rodríguez-Sandoval et al., 2008; Maieves et al., 2012; Rosales-soto et al., 2016; An et al., 2019; Ma'Aruf and Abdul, 2020). Pasting properties of cassava products are key determinants of quality. Rapid Visco Analyzer (RVA) has been used to evaluate pasting properties of cassava accessions, and the parameters estimated include peak viscosity, setback viscosity, final viscosity, pasting temperature, and time to reach peak viscosity. Using RVA, less than five samples can be processed in 1 h (Rosales-soto et al., 2016; Chatpapamon et al., 2019). Although not yet reported for cassava, NIRS has been shown to adequately predict texture and pasting properties of rice (Meullenet et al., 2002; Chueamchaitrakun et al., 2011) and sweet potato (Lu et al., 2006a,b). Therefore, the ability of NIRS to predict cassava pasting properties should be explored.

Biotic Constraints

Different purposes required different discrimination methods for quantification of disease incidence and/or severity. For example, for breeding purposes, a visual scale of 1–5 from disease-free to highly diseased may be sufficient, where breeders will only retain those cultivars with a score lower than two. However, this is subjective and dependent upon personal perceptions. On the other hand, QTL mapping or other applications may require greater resolution or more accuracy (Garcia-Oliveira et al., 2020). Accurate evaluation and novel approaches for cassava disease detection are needed to efficiently assess disease severity (Chiang et al., 2016). Root necrosis indexes for CBSD evaluation, which account for the root sample size or their economic value, have been proposed to replace traditional based evaluation (Kawuki et al., 2019). Image analysis has similarly been used (Garcia-Oliveira et al., 2020; Nakatumba-Nabende et al., 2020). Analysis of field images combined with various algorithms, including K mean clustering algorithms (Anderson et al., 2015), artificial neutral network (Abdullakasim et al., 2011), and more recently, machine learning techniques and convolutional neutral network (Owomugisha and Mwebaze, 2016; Sambasivam and Opiyo, 2020) have provided a more accurate and objective assessment of disease severity and incidence. The smart phone-based diagnostic system (NURU-AI) is being developed to support remote diagnosis by smallholder farmers in Africa for real-time prediction of the state of cassava health (Owomugisha and Mwebaze, 2016; Ramcharan et al., 2017, 2019).

Root System Architecture

Uniformity in size and shape of cassava roots is an important breeding objective. Routine assessment of storage root size and shape have relied on visual scores that are both subjective, time consuming, labor-intensive and destructive. Implementation of non-invasive approaches include image analysis of roots to provide unbiased quantitative data on important root traits. High throughput three-dimensional imaging and non-destructive methods such as ground penetrating radar (GPR) have recently been developed and used for quantifying cassava storage roots and their parameters for estimation of size and shape under field condition (Delgado et al., 2017, 2019). Kengkanna et al. (2019)

developed cassava shovelomics to evaluate root architecture traits. More recently, Yonis et al. (2020), demonstrated the feasibility of root phenotyping using image capture and analysis. Routine implementation of these new phenotyping solutions offers new opportunities for cassava breeders to efficiently and precisely select and release cultivars with root architecture that are favorable for harvesting, processing, as well as selection for early bulking characteristics. The development of robust phenotyping technologies for large-scale phenotyping with increased precision at reduced cost will enable efficient screening of larger populations. Precise phenotyping is invaluable for carrying out downstream analysis, including characterization of genetic factors that contribute to phenotypic variation.

CONNECTING GENOTYPIC AND PHENOTYPIC INFORMATION AND DISCOVERY OF USEFUL GENES AND/OR QUANTITATIVE TRAIT LOCI

Genotype variations have been linked to corresponding differences in phenotypes and the genetic basis of the phenotypic variability evaluated to identify genes and/or quantitative trait loci (QTLs) associated with the traits of interest. Two approaches have been used for genotype-phenotype association: (1) classical QTL mapping using experimental populations derived from bi-parental crosses with contrasting phenotypes (Collard et al., 2005); and (2) genome-wide association studies (GWAS) mapping that use germplasm collections and incorporate historical events that have occurred during domestication of the crop. Classical QTL analysis in cassava has been made possible with the development of numerous genetic linkage maps (**Table 1**) and use of statistical approaches that have resulted in several QTLs identified for economically important traits (**Table 2**). With the decreasing cost of DNA sequencing, new genotyping technologies and their improved accuracy, GWAS is being increasingly used for genetic analysis of traits. GWAS has enabled the exploration of allelic diversity that exists in natural populations, the discovery of beneficial alleles in several crops. Marker-trait associations in cassava have focused on a few important traits of interest.

Cassava Mosaic Disease

The dominant gene *CMD2* underlying CMD resistance was discovered in the farmer-preferred Nigerian landrace TME 3 (Akano et al., 2002). A new linked QTL underlying CMD resistance named *CMD3* that explained 11% of the phenotypic variance (PVE) was later identified 36 cM away of *CMD2* gene by Okogbenin et al. (2012). The qualitative nature of CMD resistance was confirmed in subsequent studies, and a single locus with a large effect in the vicinity of the previously mapped *CMD2* locus was uncovered (Rabbi et al., 2014; Echefu et al., 2016; **Table 2**). The first GWAS mapping study in cassava conducted by Wolfe et al. (2016) identified 198 significant SNPs associated with CMD severity on 14 chromosomes. The significant SNPs were mostly concentrated in a single region on chromosome 8 (based on *Manihot esculenta* v5 genome assembly corresponding to chromosome 12 on v5.1 and v6.1) that account for 30 to 60% of variation in genetic resistance. Additional regions with small effects, including one on chromosome 9 that co-located with the *CMD1* resistance were also reported (Wolfe et al., 2016; **Table 3**). The study of Wolfe et al. (2016) substantiated bi-parental mapping studies (Akano et al., 2002; Okogbenin et al., 2012) reporting the single major gene, *CMD2,* determining CMD resistance and a second QTL, *CMD3*, closely linked to *CMD2*. A key outcome of the Wolfe et al. (2016) study was the lack of other major-effect loci. Likewise, significant interactions between the significant SNPs were disclosed. Nzuki et al. (2017) found two QTLs associated with CMD on chromosomes 12 and 14. The percentage of variation explained (PVE) by these QTLs were 13.01 and 13.36%, respectively. The QTL on chromosome 12 was confirmed as the QTL linked to the *CMD2* locus. Masumba et al. (2017) detected two highly significant CMD resistance QTLs on chromosome 12 defining the *CMD2* locus, as well as additional putative QTLs on other chromosomes. Similar observations were made by Garcia-Oliveira et al. (2020) who found two closely linked loci on chromosome 12 and additional QTLs on chromosomes 9 and 10. Further dissection of the major QTL on chromosome 12 (*Manihot esculenta* v5.1 and 6.1 genome assembly) revealed the presence of two possible epistatic loci and/or multiple resistance alleles, which may account for the difference between moderate and strong disease resistance in the germplasm (Masumba et al., 2017; Nzuki et al., 2017; Garcia-Oliveira et al., 2020; **Table 2**). Gaps in the pseudochromosome 12 region containing the *CMD2* loci caused by highly repetitive DNA might explain the two separate loci (Kuon et al., 2019). Somo et al. (2020) identified QTLs associated with CMD severity across all chromosomes, except on chromosomes 4, 5, 6, 8, 11, 13, and 18. They validated the presence of the previously reported *CMD2* locus and detected another major QTL on chromosome 16. Nzuki et al. (2017) and Rabbi et al. (2020) confirmed the role of *CMD2* loci as the major gene for CMD resistance and reported two additional loci on chromosome 14 (**Table 3**).

Cassava Brown Streak Disease

Quantitative trait loci associated with CBSD root necrosis were identified on chromosomes 5, 11, 12, and 15 by Nzuki et al. (2017). The detected QTLs explained up to 10.18% of the PVE. Masumba et al. (2017) reported two consistent QTLs linked to resistance to CBSD-induced root necrosis on chromosomes 2 and 11, as well as a putative QTL on chromosome 18. Further additional putative QTLs were detected on other chromosomes (3, 4, 5, 6, 7, 10, 12, 15, and 16). In addition to QTLs found on chromosome 8 and 18, new putative QTL for CBSD root necrosis was identified on chromosome 14 (Garcia-Oliveira et al., 2020). Seven QTLs associated with CBSD foliar symptoms were found on chromosomes 4, 6, 15, 17, and 18. The most significant QTL explained 8.45% of the PVE (Nzuki et al., 2017). The study of Masumba et al. (2017) revealed several QTLs on all chromosomes linked with CBSD foliar symptoms, the most interesting of which was found on chromosome 2 and explained 4.6% of the PVE. Garcia-Oliveira et al. (2020) identified one major QTL on chromosome 18 that explained 12.87% of the phenotypic

TABLE 1 | Summary of the published genetic linkage maps of cassava.

Population	Mapped markers	Linkage groups/ chromosomes	Map length (cM)	References
TMS30572 (♀) × CM2177-2 (♂) F_1 (*n* = 90)	32 RFLPs, 30 RAPDs, 3SSRs, 3 Isoenzymes	20	932.6 (female)	Fregene et al., 1997
	107 RFLPs, 50 RAPDs, 1 SSR	24	1220 (male)	
TMS30572 (♀) × CM2177-2 (♂) F_2 (*n* = 268)	100 SRRs	22	1236.7	Okogbenin et al., 2006
Huay Bong × Hanatee F_1 (*n* = 100)	510 SSRs	23	1420.3	Sraphet et al., 2011
Hanate (♀) × Huay Bong (♂) F_1 (*n* = 100)	303 SSRs	27	1328	Whankaew et al., 2011
Namikonga (♀) × Albert (♂) F_1 (*n* = 130)	434 SNPs, 134 SSRs	19	1837 (one step)	Rabbi et al., 2012
	348 SNPs, 128 SSRs	18	1541 (two steps map)	
IITA-TMS-4(2)1425 (♂) x IITA-TMS-011412 (♂) F_1 (*n* = 180)	6756 SNPs	19	3771	Rabbi et al., 2014
KU50 × SC 124 F_1 (*n* = 85)	2331 SNPs, 537 Indels, 164 methylation markers	20	1970.40	Xia et al., 2014
Huay Bong (♀) × Hanatee (♂) F_1 (*n* = 100)	2110 SNPs	19	1785.62	Pootakham et al., 2014
TMS30572 (♀) × CM2177-2 (♂) F_1 (*n* = 132)	2141 SNPs	18	2571	Soto et al., 2015
9 bi-parental F_1 crosses 1 self-pollinated cross (S1) *n* = 1740	22,403 SNPs	18	2412	ICGMC, 2015
Kibora (♀) × AR37-80 (♂) F_1 (*n* = 106)	1974 SNPs	21	1698	Nzuki et al., 2017
Namikonga (♀) × Albert (♂) F_1 (*n* = 240)	943 SNPs	18	1776.2	Masumba et al., 2017
AR40-6 (♀) × Albert (♂) F_1 (*n* = 130)	2125 SNPs	18	1730	Garcia-Oliveira et al., 2020

SNP, single nucleotide polymorphism; RAPD, random amplified polymorphism DNA; RFLP, restriction fragment length polymorphisms; SSR, simple sequence repeat.

variation, but at a slightly different position. These same authors also reported a minor QTL on chromosome 11 (**Table 2**). The GWAS approach was used by Kayondo et al. (2018) to unravel the genetic architecture of CBSD. The polygenic control mechanism of resistance for CBSD and its instability across the environment was highlighted. Eighty-three (83) loci associated with foliar symptoms at 3 months after planting (MAP) were identified on chromosome 11. The top SNPs explained 6% of the phenotypic variance. Significant SNPs were identified on chromosomes 11, 4, and 12 for foliar severity score at 6 MAP. Recently, Somo et al. (2020) using a diverse panel of breeding lines, identified QTLs conferring resistance to CBSD on chromosomes 9 and 11 that accounted for 9 and 5% of PVE, respectively. Nine markers representing four loci on chromosomes 2, 3, 8, and 10 associated with resistance to CBSD for root necrosis were also reported by these authors (**Table 2**). Putative regions on chromosomes 11 and 15 were shown to be associated with both CBSD foliar and root necrosis symptoms (Nzuki et al., 2017), suggesting that CBSD root necrosis and CBSD foliar symptoms are most likely to be influenced to some extent by the same gene(s) or by closely linked genes at this locus. However, most of the QTLs reported have been associated with either root necrosis or foliar symptoms, supporting the notion that resistance to foliar and root symptoms of CBSD are largely under different genetic control (Masumba et al., 2017; Garcia-Oliveira et al., 2020). The detected QTLs were not consistent across studies. Different mapping populations were used in the case of experimental populations. There might be some variations in CBSD response. Furthermore, QTLs tend to be population specific in bi-parental populations. Likewise, different environmental conditions, population size (106–1986 samples), and the subjectivity of the scoring system used for data collection could have affected QTL detection and localization. The use of different versions of the cassava genome assembly makes it challenging to compare some of these results. Therefore, these QTLs should be mapped onto the most recent version of the cassava reference genome.

Cassava Green Mite and Related Traits

Nzuki et al. (2017) detected on chromosomes 5 and 10 QTLs associated with cassava green mite (CGM) with maximum PVE of 10.56 and 10.08%, respectively. Recently, 95 SNP markers significantly associated with CGM resistance were reported by

TABLE 2 | Summary of the QTL studies of cassava using controlled populations.

Traits	Mapping Population	Markers Used	Key Findings	References
CBB	TMS 30572 (♀) × CM 2177-2. (♂) F_1 (*n* = 150)	68 RFLPs (female map) 15 RFLPs (male map)	Eight (8) QTLs identified on five LGs in the female map explaining 9 - 20%. In the male map, four QTLs explaining 10.7–27.1% of the PVE.	Jorge et al., 2000
CBB	TMS 30572 (♀) × CM 2177-2 (♂) F_1 (*n* = 150)	Female map (95 RFLPs, 36 RAPD, 3 isoenzymes, 3 SSRs, 3 EST, 2 unknown genes) Male map (89 RFLPs, 41 RAPDs, 4 unknown genes, 1 EST)	Eight (8) QTLs detected. Alleles coming from both parents contribute to the resistance in the progenies.	Jorge et al., 2001
CMD	TME 3 × TMS 30555 F_1 (*n* = 80)	186 SSRs	First report of the *CMD2* locus associated with CMD resistance and identification of SSRs linked to the CMD resistance genes *CMD2.*	Akano et al., 2002
DMC	MCOL 1684 × THAI 1 F_2 (*n* = 199)	98 SSRs	Six (6) QTLs on four different linkage groups explaining 14 to 40% of the PVE.	Kizito et al., 2007
Cyanogenic glycoside (CN)			Two (2) QTLs on two different LGs explaining 7% and 20% of the PVE, respectively.	
Anthracnose	TME 117 (♀) × TMS 92/0326 (♂) F_1 (*n* = 60)	53 RAPDs	Using bulk segregant analysis, two markers (OPAF2 and OPF06) flanking the *CAD1* genes were identified that explained 33.7 and 27.4% of the phenotypic variation.	Akinbo et al., 2007
Fresh foliage	TMS 30572 (♀) × CM 2177-2 (♂) F_2 (*n* = 268)	122 SSRs	Three (3) QTLs for fresh foliar detected on three linkage groups. The PVE ranged from 5.5 to 31.1%.	Okogbenin et al., 2008
Harvest index			Three (3) QTLs associated with harvest index on three LGs explaining 2.10–6.67% of the phenotypic variation. QTL with major effect accounted for 36.9%.	
Dry yield root			Three (3) QTLs for dry root yield on three LGs (1, 3, 13) accounting for 6.0–9.2% of the PVE.	
Cyanogenic potential	Hanatee (♀) × Huay Bong (♂) F_1 (*n* = 100)	303 SSRs	Five (5) QTLs mapped on LG2, 5, 10, 11 explained 16.1 to 26% of the PVE. The most significant association was on LG10.	Whankaew et al., 2011
CGM	Four backcross populations (CW 65, 66, 67, 68)	500SSRs	Using bulk segregant analysis, three markers (SSRY11, SSRY 346, and NS1099) associated with CGM resistance were detected.	Choperena et al., 2012
Starch pasting viscosity	Hanatee (♀) × Huay Bong (♂) F_1 (*n* = 100)	510 SSRs	Fifteen (15) QTLs affecting five starch pasting viscosities identified when average values of the three environments were used. The PVE ranged from 10.0 to 48.4%. Data from separate environment revealed 48 QTLs for seven starch pasting properties, explaining 6.6 to 43.7% of the phenotypic variation.	Thanyasiriwat et al., 2013
Starch pasting properties	Huag Bong (♀) × Hanate (♂) F_1 (*n* = 100)	2110 SNPs	Single co-localize QTL for pasting temperature and pasting time detected on LG7. The PVE explained was 44.7 and 24.3% for pasting temperature and pasting time, respectively. Additional QTL for starch pasting time found on LG10 (PVE = 22.5%).	Pootakham et al., 2014
CMD	IITA = TMS-4(2) 1425 (♂) x IITA-TMS-011412 (♂) F_1 (*n* = 108)	6756 SNPs	Single SNP underlying CMD resistance in the vicinity of the previously mapped *CMD2* locus that explain 74% of the variance	Rabbi et al., 2014
CMD	TMS961089A × TMEB117	4394 SNPs	Single SNP underlying CMD resistance in the vicinity of the previously mapped *CMD2* locus that explain 74% of the variance	Echefu et al., 2016
Fresh starch content	Huag Bong (♀) × Hanate (♂) F_1 (*n* = 100)	510 SSRs	Nine (9) QTLs on seven LGs, including 6, 7, 9, 11, 13, and 11 that explained 11.3 to 27.3% of the phenotypic variation. Six of the QTLs were location specific while two were found across environments.	Sraphet et al., 2017
CBSD root necrosis	Namikonga (♀) × Albert (♂) F_1 (*n* = 240)	943 SNPs	Two QTLs consistent across seasons/sites on Chr11 and 2 and a putative QTL on Chr18. Further additional QTLs on Chr3, 4, 5, 6, 7, 10, 12, 15, and 16	Masumba et al., 2017

(Continued)

TABLE 2 | Continued

Traits	Mapping Population	Markers Used	Key Findings	References
CBSD foliar symptoms			Several QTLs revealed on all chromosomes. The most interesting QTL was on Chr2 and has a PVE of 4.6%.	
CMD			Two (2) QTLs identified on Chr12 that explain 16.43 and 13.5% of the phenotypic variation. Additional QTLs detected on all other chromosomes except Chr11 and 14.	
CBB	TMS30572 × CM2177-2 F_1 (n = 117)	2571 SNPs	Five-strain specific QTLs for resistance to *Xam* explaining between 15.8 and 22.1% of the PVE	Sedano J. C. S. et al., 2017
CBSD root necrosis	Kibora (♀) × AR37-80 (♂) F_1 (n = 106)	1974 SNPs	Three (3) QTL associated with CBSD root necrosis on Chr5, 11, and 12 explaining 2.62–10.18% of the phenotypic variation	Nzuki et al., 2017
CBSD foliar symptoms			A total of seven QTLs detected on Chr4, 6, 17 and 18 explaining 2.64–8.45% of the PVE	
CMD			Two (2) QTLs detected on Chr12 and Chr14 with a maximum PVE of 13.01 and 13.36%, respectively.	
Leaf pubescence	TMEB778 (♀) × TMEB419 (♂) F_1 (n = 109)	42204 SNPs	Seventy-one (71) SNPs significantly associated with leaf pubescence on Chr12. The top significant SNP explained 26% of the phenotypic variation.	Ezenwaka et al., 2020
Stay green			A total of 126 SNPs associated with stay green. The PVE ranged from 20 to 30%.	
CMD	AR40-6 (♀) × Albert (♂) F_1 (n = 130)	2125 SNPs	Provide evidence for two QTL or multi-allelic variants influencing CMD resistance within the locality of *CMD2* locus. Additional putative QTLs on Chr 9 and 10	Garcia-Oliveira et al., 2020
CGM			Five (5) minor effect QTLs specific to the planting stage. Four QTLs detected at 3 MAP on Chr 5, 9, 13, and 18. PVE ranging from 1.35 to 11.06%, while one QTL detected on Chr16 at 6 MAP accounting for 7.27% of the PVE.	
CBSD foliar symptoms			One minor QTL on Chr11 accounting for 2.5 % phenotypic variation. One major QTL identifying on Chr18 explaining 12.87% of the phenotypic variation	
CBSD root necrosis			New putative QTL for CBSD root necrosis identified on Chr14	

SNP, single nucleotide polymorphism; RAPD, random amplified polymorphism DNA; RFLP, restriction fragment length polymorphisms; SSR, simple sequence repeat; EST, expressed sequence tag; CBSD, cassava brown streak disease; PVE, phenotypic variance; LG, linkage group; Chr, Chromosome.

Ezenwaka et al. (2020). The significant markers concentrated in a single region of the left arm side on chromosome 12. The variance explained by the significant markers ranged from 18 to 31%. Garcia-Oliveira et al. (2020) detected five QTL for CGM resistance at 3 and 6 MAP, all with minor effect on chromosomes 5, 9, 13, and 18. While the QTLs on chromosome 9 was found in all four environment tested, the QTL on chromosome 8 co-localized with previously reported marker for CGM resistance (**Table 2**). The first GWAS to identify SNPs linked to CGM and CGM-related traits was performed by Ezenwaka et al. (2018) who found 35 significant SNP markers, including 12, 17, 5, and 1 associated with CGM, leaf pubescence, leaf retention, and stay green, respectively. All the significant markers were found on chromosome 8, except the SNP associated with stay green, which was identified on chromosome 13. Some of the significant SNP markers on chromosome 8 reported by these authors were also detected in the recent study of Rabbi et al. (2020) who identified other putative loci on chromosomes 1, 12, and 9. Association analysis of CGM-related traits, apical pubescence identified significant loci on five chromosomes, two of which co-located in the same regions underlying resistance to CGM on chromosomes 8 and 12 (Rabbi et al., 2020; **Table 3**).

Cassava Bacterial Blight

Two QTLs associated with cassava bacterial blight (CBB) caused by *Xanthomonas axonopodis* pv. manihotis (*Xam*) were reported on linkage group (LG) 4 and LG8 that explained 12.6 and 10.9% of the field resistance to CBB, respectively (Sedano J. S. et al., 2017). Another study conducted by Sedano J. C. S. et al. (2017) found five strain-specific QTLs conferring resistance to *Xam* that explained 15.8 and 22.1% of the phenotypic variance. Three of the associated QTLs were found to be effective against *Xam*318 strain and explained 17.3 to 18.8% of the phenotypic variance, while two were detected for *Xam*681 and accounted for 15.8–22.1% of the phenotypic variance (**Table 2**).

Cassava Root Rot

The complex nature of cassava root rot disease (CRR) was highlighted by Brito et al. (2017), who identified 38 significant SNPs associated with CRR. Of these, 8 and 22 were related to the severity of dry root rot in the pulp and peel, respectively, while

TABLE 3 | Summary of the GWAS studies of cassava.

Trait	Population/size	Markers used	Key findings	References
CMD	African breeding lines *n* = 6128	42 113 SNPs	A total of 198 SNPs mostly concentrated in a single region on Chr8, explaining 0.55 to 22% of PVE.	Wolfe et al., 2016
Dry matter content	Genetic Gain Collection *n* = 672	72 729 SNPs	Major locus at 24.1 Mbp of Chr1	Rabbi et al., 2017
Provitamin A			Major association regions at 24.1 and 30.5 Mbp on Chr1	
Root rot severity *Fusraium* (peel)	Cassava germplasm bank (EMBRAPA) *n* = 263	14 094 SNPs	Twenty-two (22) SNPs on Chr2, 5, 6, 8, 9, 10, 12, 17, 18, and 'A fragments' most significantly found on chromosome 10. The PVE ranged from 4.66 to 9.54%.	Brito et al., 2017
Root rot severity Fusarium (Pulp)			Eight (8) significant SNPs on Chr2,11.14,15,16. The most significant SNP was on Chr 11. The PVE range from 4.86 to 6.13%.	
Root rot severity *Phytophthora* (peel)			One (1) SNP on Chr13 explaining 7.99% of the phenotypic variation	
Root rot severity *Phytophthora (*peel)			One (1) SNP on Chr8 explaining 6.80% of the phenotypic variation.	
Root rot severity *Botryosphaeriaceae* (peel)			Three (3) SNPs on Chr5, 13, and genomic fragment "B" explain 8.90–10.09% of the phenotypic variation.	
Root rot severity *Botryosphaeriaceae* (pulp)			Three (3) SNPs on Chr9 and 12, explaining 5.47–7.04% of the phenotypic variation.	
Cassava green mite	Advanced breeding lines *n* = 845	61 307 SNPs	Twelve (12) significant loci on Chr8. The top significant explained 7% of the PVE.	Ezenwaka et al., 2018
Leaf pubescence			Seventeen (17) loci on Chr8. PVE ranged from 4 to 7%.	
Leaf retention			Five (5) loci on Chr8 each explaining 4% of the phenotypic variation.	
Stay green			One (1) significant SNP on Chr13 explaining 4% of the phenotypic variation.	
Carotenoid content	SC9 × white-root SC F_1 (*n* = 98)	104 059 SNPs	Eighty-four (84) SNPs and 694 genes on all chromosomes except Chr5.	Luo et al., 2018
Storage root quality	Diversity Panel *n* = 158	19 850 SNPs	Four (4) association signals, including two SNPs on Chr5, and 15 for starch content and two SNPs on Chr14 and 16 for dry mass content.	Zhang et al., 2018
Yield components			Seven (7) association signals. Number of storage root (3, Chr4, Chr 5, Chr 9); storage root weight (1, Chr 2), dry mass weight (3, Chr 2).	
Morphological characteristics			Eleven (11) association signals. Stem diameter (6; Chr3, Chr6, Chr7, Chr14). First branch height (5; Chr2, Chr3, Chr9, Chr10).	
Leaf characteristics			Fourteen (14) association signals. Lobular length (4; Chr4, Chr5, Chr6, Chr18). Lobular width (3, Chr3, Chr5, Chr10). Petiole length (1, Chr1), Leaf aspect ratio (6, Chr1, Chr3, Chr6, Chr10, Chr14).	
CBSD foliar severity (3 months)	Diversity panel *n* = 1986	63 016 SNPs	Eighty-three (83) significant SNPs located on Chr11 the top SNPs explained 6% of the observed significant variation.	Kayondo et al., 2018
CBSD foliar severity (6 months)			The significant SNPs were located on Chr11, Chr4, and Chr12.	
All-*trans* β-carotene	GS training population *n* = 594	87 380 SNPs	Seven (7) significant SNPs with SNP effect ranging from 0.328 to 0.592.	Ikeogu et al., 2019
Lutein			Twelve (12) significant SNPs and 11 unique loci with SNP effects ranging from 0.025 to 0.034	
Total carotenoids content			Twenty (20) significant SNPs with SNP effect ranging from 0.435 to 0.749.	
Violaxanthin			One (1) significant SNP with an effect of 0.01	
13-*cis* beta-carotene			Twenty-three (23) significant SNPs with SNP effect ranging from 0.039 to 0.058.	
15-*cis* beta-carotene			Twenty-one (21) significant SNPs with SNP effect ranging from 0.008 to 0.011.	

(Continued)

TABLE 3 | Continued

Trait	Population/size	Markers used	Key findings	References
9-*cis* beta-carotene			Eighteen (18) significant SNPs with SNP effects ranging from 0.032 to 0.045.	
CMD	Diverse breeding lines panels $n = 432$, $n = 408$, $n = 402$	84 434 SNPs	Genetic association detected on Chr1, 2, 3, 7,9,10, 12, 15, 15, 16, and 17. The presence of *CMD2* locus was validated. Another major CMD QTL was detected on Chr16.	Somo et al., 2020
CBSD root severity			Nine (9) SNPs on Chr2, 3, 8, 10 were significant.	
CBSDS| CMDS			One QTL on Chr4 and 2 on Chr12 significantly associated with both CBSDS/CMDS resistance.	
Waxy starch	Diversity panel $n = 382$	20,956 SNPs	10 association signals on Chr2	do Carmo et al., 2020
Cassava mosaic disease	Elite breeding clones $n = 5130$	101,521 SNPs	*CMD2* locus and other loci on Chr14 were captured. SNP effects ranged from −0.23 to 0.82	Rabbi et al., 2020
Cassava green mite			Four (4) regions associated with CMG on Chr1, 8, 12, 19. SNP effect size ranged from −0.18 to 0.10	
Apical pubescence			Significant association on Chr8, 9,11, 12, 16 with effect size from −0.19 to 0.08.	
Carotenoid content			Eight (8) significant associations on Chr1, 5, 8, 15, and 16 with SNP effect between −0.007 and 0.49. Major locus on Chr1 Around 24.1, 24.6 and 30.5 Mb.	
Dry matter content			Five (5) significant SNPs on Chr1, 6, 12, 15, 16. SNP effect between −1.32 and 0.85. The most significant SNP occurred on Chr1 around 24.64 Mb.	
Harvest index			Two (2) genomic regions associated with this trait on Chr2 and Chr12, with SNP effect of −0.04 and −0.03, respectively.	
Morphological traits			Eight (8) association signals detected. Outer cortex color 2 SNPs on Chr1 (3.05 Mbp) and Chr2 (6.56 Mbp). A single genomic region associated with periderm color on Chr3 (4.54 Mbp). 2 SNPs association signal for plant type on Chr1 (2.19 Mbp and 25.30 Mbp). Three Loci associated with stem color and Chr2 and Chrx 8 which contain the most significant at around 13.6 Mbp.	
Leaf morphology			Seven (7) association signals. Leaf petiole color single genomic region on Chr1 (23.45 Mbp); the same region was associated with mature leaf greenness. 3 association signals for apical leaf color (Chr2, 3, 8). For leaf shape two major loci on Chr 15 around 10.27 and 20.57 Mbp.	
Cyanogenic glycoside	Test population ($n = 1246$) Validation population ($n = 636$)	27,054 SNPS	Two significant peaks on chromosome 16 and 14, explaining 36 and 8% of the variance, respectively.	Ogbonna et al., 2020

CBSD, Cassava Brown Streak Disease severity; CMD, Cassava Mosaic Disease.

the other eight were associated with soft root rot and black root rot (**Table 3**).

Carotenoids and Storage Root Color

Three QTLs that control the content of carotenoids and four QTLs linked to the color of cassava roots pulp were identified by Morillo et al. (2013) (**Table 2**). Root yellowness resulting from carotenoid accumulation elucidated through GWAS has revealed major association regions that govern this trait around 24.1 and 30.5 Mbp of chromosome 1 (Rabbi et al., 2017). More recently, using a larger diversity panel, Rabbi et al. (2020) confirmed these previous findings. They found five new genomic regions associated with carotenoid content on chromosomes 5, 8, 15, and 16. Luo et al. (2018) identified 84 SNPs distributed in all chromosomes, except chromosome 5, associated with carotenoid traits. Ikeogu et al. (2019) identified 42 unique SNPs significantly associated with variation in total carotenoid content on chromosomes 1, 2, 4, 13, 14 and 15. Additional regions for variation in the total carotenoid content, as well as the individual carotenoids, were uncovered. Some regions associated with more than a single carotenoid were identified, suggesting the possibility of pleiotropic effects (**Table 3**). The level of phenotypic variability, the SNP frequencies and distributions, and the difference in sample size (98–5130 samples) between the panels used for GWAS might explain the differences in QTLs detected. The wider coverage of diversity could increase the detection of true novel

associations, while small sample size could lead to some spurious associations.

Cyanogenic Glucosides

Kizito et al. (2007) reported two QTLs (SSRY105, SSRY42) on two different linkage groups controlling cyanogenic glucosides. The two QTLs explained 7% and 20% of phenotypic variation, respectively. Also, both QTLs showed additive effects. Five QTLs associated with cyanogenic potential were identified across four linkage groups, including LG2, 5, 10, and 11 (Whankaew et al., 2011). The percentage of phenotypic variance explained from all detected QTLs ranged from 15.9 to 26.0% (**Table 2**). One (SSRY42) of the two QTLs reported by Kizito et al. (2007) was found in this latest study. More recently, Ogbonna et al. (2020) reported two genomic regions, in chromosomes 16 and 14, associated with hydrogen cyanide potential in cassava. The most significant marker was found on chromosome 16. Chromosomes 16 and 14 tagged SNPs explained 36 and 8% of phenotypic variance, respectively (**Table 3**). The significant region on chromosome 14 coincides with previously reported cyanide associated QTL reported by Kizito et al. (2007).

Dry Matter Content

Six QTLs detected on four different LGs were reported to control DMC using a bi-parental mapping population. Individual QTL explained 14 to 40% of the variance. It was shown that additive, dominance, and overdominance effects play a role in the expression of this trait (Kizito et al., 2007; **Table 2**). Using genome-wide association mapping, DMC was found to be associated with a major locus occurring on a 24.1 Mbp region of chromosome 1 (Rabbi et al., 2017). This locus was confirmed by the recent study of Rabbi et al. (2020), who reported another major locus on chromosome 6 as well as three additional loci on chromosomes 12, 15, and 16 (**Table 3**).

Starch and Starch Quality

The study of Thanyasiriwat et al. (2013) revealed the complex inheritance of starch pasting properties. Using average values of three environments, these authors reported 15 QTLs on LG1, 4, 6, 7, 8, 10 12, 13, 14, 16, and 18 affecting five starch pasting viscosities (peak viscosity, hot paste viscosity, cool paste viscosity, set back, and pasting temperature). The detected QTLs explained 10.0 to 48.4% of the phenotypic variance. Based on analysis of each environment, 48 QTLs significantly associated with seven starch pasting viscosities (peak viscosity, hot paste viscosity, break down, cool paste viscosity, setback, pasting time, and pasting temperature) were detected on all LGs except LG2, 4, and 7. The PVE ranged from 6.6 to 43.7. Thanyasiriwat et al. (2013) also reported two major QTLs on LG1 and LG6 for pasting temperature, which accounted for more than 70% of the phenotypic variation. Pootakham et al. (2014) identified a single co-localized QTL controlling pasting temperature and pasting time on LG7. The major QTL explaining 44.7 and 24.3% of the phenotypic variance. These authors reported additional QTL associated with starch pasting time on LG 10 that accounted for 22.5% of the phenotypic variance. A total of nine QTLs controlling fresh starch content was identified on seven linkage groups (LG4, 6, 7, 9, 11, 13, and 16). Among these, six QTLs were location-specific, and three QTLs were detected across three environments. The percentage of phenotypic variance explained by the QTLs ranged from 11.3 to 27.3% of the phenotypic variation (Sraphet et al., 2017; **Table 2**). Using GWAS, 10 SNP associated with waxy phenotypes were identified on chromosome 2 that co-located in genic regions that included five known genes and five genes of unknown function (do Carmo et al., 2020; **Table 3**).

Agronomy and Physiology Traits

Eight QTLs associated with fresh root yield were identified on seven linkage groups (LG1, 2, 6, 9,12,13, and 16) from a bi-parental mapping population, of which two QTLs on linkage group 16 were found across two environments. These QTLs explained 12.9 to 40% of phenotypic variation (Sraphet et al., 2017; **Table 2**). Zhang et al. (2018) reported 36 loci related to 11 agronomic traits, including leaf characteristics, morphological characteristics, yield components, and root quality that were identified by GWAS analyses. They found seven SNPs associated with yield components that explained about 14.95% of the phenotypic variance on average. Morphological characteristics exhibited 11 association signals and explained 12.23 to 20.86% of the phenotypic variance. A total of 14 SNPs were identified from leaf characteristics, which explained 11.9 to 22.6% of the phenotypic variance. Genomic regions associated with harvest index were uncovered on chromosomes 2, 3, 4, 6, 8, 9, 12, 14, and 15. Two association signals for outer cortex color were found on chromosomes 1 (3.05 Mbp) and 2 (6.56 Mb). A single genomic region on chromosome 3 (4.54 Mbp) has been linked to periderm color. Two association signals were found for plant types on chromosome 1 (2.19 and 25.30 Mbp). Five loci associated with stem color variation were reported on chromosomes 2 and 8. The most significant loci were around 13.6 Mbp on chromosome 8. A single genomic region at around 23.45 Mbp on chromosome 1 was found concomitantly associated with leaf petiole color and mature leaf greenness. Variation in leaf color was found to be associated with three loci occurring on chromosomes 2, 3, and 8. At around 10.27 and 20.57 Mbp on chromosome 15, two loci associated were identified (Rabbi et al., 2020; **Table 3**). Seventy-one (71) markers significantly associated with leaf pubescence were identified by Ezenwaka et al. (2020) on chromosome 12. The variance explained by these significant markers ranged from 18 to 26%. The same study of Ezenwaka et al. (2020) reported 126 SNP markers associated with stay green on chromosome 12, and the variance explained by the detected QTLs ranged from 20 to 30% (**Table 2**).

Numerous favorable alleles, functional loci or regions linked to traits of interest have been identified through marker-trait associations and their phenotypic contribution identified, giving a glimpse to the genetics underlying phenotype variation. However, the designation of linkage groups/chromosomes are sometimes different, making it challenging to align QTL detected in different studies (Garcia-Oliveira et al., 2020). Synchronization of cassava QTL information and the development of a cassava QTL repository is needed. A cassava QTL database that contains QTL information systematically aligned to the cassava reference

genome, as well as information about germplasm and genetic material used in the QTL studies, will be a useful resource for both cassava geneticists and breeders (Ni et al., 2009; Yonemaru et al., 2010; Said et al., 2015). A benefit envisioned from the identification of molecular markers tightly linked to traits of interest is their deployment in breeding as an indirect selection method to accelerate the rate of genetic gain.

QTL UTILIZATION IN CASSAVA BREEDING

Numerous QTL controlling a wide variety of traits have been identified to be utilized in marker-assisted selection (MAS) known as forward selection (**Tables 2**, **3**). MAS is a technique of indirect selection of traits. Its implementation is still challenging in many breeding programs (Chukwu et al., 2019; Cobb et al., 2019). Although panels of molecular markers closely linked to key cassava traits have been identified, successful applied experiences of MAS in cassava breeding are limited (**Table 4**). Markers tightly associated with *CMD2* have been effectively used to identify genotypes bearing the *CMD2* gene in West African germplasm, introgression of *CMD2* genes into various cassava germplasm, and selection of parental lines for planned crosses. This has enabled breeders to focus on fewer genotypes at an early stage of the breeding program, saving the breeder time and labor cost (Blair et al., 2007; Okogbenin et al., 2012; do Carmo et al., 2015). MAS is appropriate for moderate to highly inherited traits that are controlled by a few major genes and not sufficiently predictive for quantitatively inherited traits controlled by many genes at different loci, each contribution to a small effect of the phenotypic expression (Singh et al., 2019). The time and investment required to develop reliable markers have also been a limiting factor for MAS implementation in cassava breeding programs in SSA. Likewise, the use of convenient phenotyping proxies do not always translate into meaningful selection targets for the breeding programs (Cobb et al., 2019). Future endeavors should prioritize cassava traits for marker development using a stage-gate system (SGS) to manage trait development. This will require the development of well-defined cassava product profiles, a set of targeted attributes the new variety should meet (Ragot et al., 2018). The establishment of a SGS will ensure accountability, transparency, and data-driven advancement decisions. This strategy is currently being implemented with the support of Excellence in Breeding (EiB)[1]. MAS should be considered only when proven more efficient than phenotypic selection to justify the expenses of development and deployment of the locus in a breeding program. Likewise, marker development for the traits retained should be feasible in a short to medium term to avoid waste of resources (Collard and Mackill, 2008). The International Institute of Tropical Agriculture (IITA) and partners have designed and converted molecular markers for several major gene traits, including provitamin A, CGM, DMC, to low-plex-high-throughput marker assay[2]. The reliability and utility of these markers are currently being evaluated in different genetic backgrounds and across different environments. A KASP assay has recently been employed to develop and validated diagnostic markers for HCN content (Ogbonna et al., 2020). The use of shared genotyping services could substantially reduce the genotyping costs, enabling the screening of a larger number of accessions at the seedling stage. Development of novel breeding strategies or models that not only capture additive effects, but also account for dominance and epistatic interactions, could contribute to the introgression of the QTLs into desired varieties. Likewise, the model should account for QTL-environment interactions for an effective MAS scheme (Singh et al., 2019). Recent genomic innovations have prompt for the search of new tools for population improvement.

GENOMIC SELECTION AND IT's POTENTIAL IN CASSAVA BREEDING

Genomic selection (GS), also referred to as genomic prediction, can complement MAS; with MAS being used for highly heritable traits controlled by one or a few markers, and GS being used for more quantitative traits. GS relies on predicting the genomic estimated breeding values (GEBV) of an individual using a trait specific model built by simultaneously fitting information provided by thousands of molecular markers spread throughout the genome (Meuwissen et al., 2001). Fitting all markers simultaneously allows a substantial fraction of trait heritability missed by QTLs or association mapping to be captured. These are likely to be small effect alleles (Resende et al., 2012). The approach was first championed in dairy cattle (Jonas and de Koning, 2013). There are great expectations for the use of GS in cassava breeding. Since the first GS studies in cassava conducted by de Oliveira et al. (2012) and Ly et al. (2013), highlighting the potential of GS, several studies have been published (Wolfe et al., 2017; Ozimati et al., 2018; de Andrade et al., 2019; Yonis et al., 2020). GS has been utilized in cassava breeding to increase resistance against CMD and CBSD in cassava populations (Wolfe et al., 2016; Kayondo et al., 2018; Ozimati et al., 2018). Ozimati et al. (2018) highlighted the potential of GS as a pre-emptive breeding strategy. Torres et al. (2019) predicted good progress in selecting clones for traits such as fresh root yield (FRY), dry matter content (DMC), fresh shoot yield (FSY), harvest index (HI), dry yield (DY) using GS. Yonis et al. (2020) assessed genomic prediction for root size and shape based on root traits extracted from digital images. Greater predictive ability has been reported for DMC, CMD, and, to a lesser extent, HI compared to other traits (Wolfe et al., 2016; de Andrade et al., 2019). This success has been attributed to their high to moderate heritability, large-effect QTLs and low genotype × environment interaction (Torres et al., 2019; Yonis et al., 2020). GS is especially attractive for complex traits controlled by many QTLs with low heritability that are difficult or expensive to assess or are measured late in

[1]https://excellenceinbreeding.org/blog/applying-stage-gates-better-manage-public-breeding-programs

[2]https://excellenceinbreeding.org/module3/kasp

TABLE 4 | Key cassava traits, target regions, potential for GS and/or MAS and the current stage of trait development.

Traits	Region[a]	Priority[b]	Method of evaluation	Potential for MAS[c]	Stage 1[d]	Stage 2[e]	Stage 3[f]	Stage 4[g]
Cassava mosaic disease	All	High	Phenotyping/MAS/GS	High				X
Cassava brown streak	East Africa	High	Phenotyping/GS	Medium			X	
Cassava green mite	All	Medium	Phenotyping/MAS/GS	Medium				X
White flies	All	High	Phenotyping	Undetermined		X		
Pro-vitamin A	All	High	Phenotyping/MAS	High				
Low cyanogenic potential	East Africa	Medium	Phenotyping	High				
Root mealiness	East Africa and Central Africa	Medium	Phenotyping	High				
Dry matter content	All	High	Phenotyping/GS	Medium				X
Starch content	All	High	Phenotyping (low throughput)	Medium				X
Garri yield	West Africa	High	Phenotyping (low throughput)	Medium	X			
Fufu yield	West Africa	High	Phenotyping (low throughput)	Medium	X			
Plant type	All	High	Phenotyping/GS	Medium	X			
Harvest index	All	High	Phenotyping/GS	Low			X	
Root weight	All	High	Phenotyping/GS	Low			X	
Root number	All	High	Phenotyping/GS	Low			X	
Root size	All	High	Phenotyping/GS	Low			X	

[a]Target regions; [b]trait priority; [c]potential for marker-assisted selection; [d]Product scoping; [e]Phenotyping and donor discovery; [f]locus identification; [g]Locus deployment; GS, Genomic selection; MAS, Marker-assisted selection.

the breeding cycle (de Oliveira et al., 2012). It could drastically reduce the breeding cycle by choosing new parents based on GEBV rather than actual phenotypes and limiting the size and number of field experiments. However, refined strategies should be adopted for complex traits. Non-additive genetic variation prevails for low heritability cassava traits and should be accounted for. Therefore, models that capture non-additive effects should be applied (Wolfe et al., 2016). GS, like any other approaches, is facing various challenges. The GS models predict poorly across populations, and consequently, the strategy requires continuous re-calibration with every breeding cycle (Wolfe et al., 2017). Allele frequency changes, introgression of new alleles, SNP-QTL linkage disequilibrium association, lack of relatedness between germplasm and population structure, as well as genotype-by-environment interaction effects, compromise the efficiency of GS. To overcome some of these limitations, dual-purpose population development and variety development pipelines have been applied to ensure training data are closely related to the new clones and for continuously updating the training model (Santantonio et al., 2020). GS could be more robust by integrating biological knowledge. The inclusion of QTL markers associated with the trait of interest could increase the robustness of genetic evaluation (Ozimati et al., 2018; Lan et al., 2020). Other variables affecting the precision of the prediction model include the size of the training population, the number of markers used in the model, the trait genetic architecture, and heritability (de Oliveira et al., 2012; Ly et al., 2013; Wolfe et al., 2017; Somo et al., 2020). Poor predictions have been reported across breeding programs limiting the prospect of sharing data from different locations, breeding programs, and countries (Wolfe et al., 2017; Somo et al., 2020). An international project funded by UK Foreign, Commonwealth and Development Office (FCDO) and the Bill and Melinda Gates Foundation (Gates Foundation) named the Next Generation Cassava Breeding Project[3] is currently using this approach in four African research institutes, including the National Crop Resources Research Institute (NaCRRI) in Uganda, the National Root Crops Research Institute (NRCRI) in Nigeria, the Tanzania Agriculture Research Institute (TARI) in Tanzania, and IITA in Nigeria. A multi-trait genomic selection strategy is being applied using program-based selection indices to efficiently improve quantitative traits simultaneously. The selection is based on yield, DMC, virus resistance to CMD and CBSD, and good plant architecture. Outstanding clones are currently advanced toward the variety release pipeline. The adoption of GS has guided the mating design, enabling rapid pyramiding of favorable allele combinations and development of progeny with the improved allelic combinations (de Oliveira et al., 2012). Moreover, GS is being used to increase genetic gain by decreasing the breeding cycle time, increasing selection accuracy, and increasing selection intensity in early generations. In order to identify a more effective breeding strategy and further improve breeding efficiency, computational simulations are currently being performed with the support of EIB[4] to: (1) compare different selection methods; (2) define optimal crossing strategies; and (3) determine appropriate recycling times. The data intensive nature of GS and widespread use of this approach across more breeding programs requires development of shared computational infrastructures and analysis pipelines.

DEVELOPMENT OF COMPUTATIONAL TOOLS FOR A GREATER IMPACT

Breeding modernization will necessitate the development of robust statistical tools and innovative analytical pipelines and platforms to accommodate the enormous quantity of data, which is being generated. Due to its clonal propagation method, there is a narrow timeline between harvesting and the next cassava planting season. This necessitates quick decision making and effective data management. Information systems are required to track samples, store genomic and phenotypic information, merge data, and conduct analysis to guide decision making (Santantonio et al., 2020). It is from this perspective that an open access repository, such as Cassavabase[5], has been developed to centralize information access to cassava research data and support cassava breeding programs (Fernandez-Pozo et al., 2015). Cassavabase contains phenotypic, pedigree, and genomics data generated over 30 years by cassava programs in Africa, Asia, and South America. Cassavabase also encompasses computational tools to facilitate analyses.

GENOME EDITING ADVANCES AND PROSPECTS

Genome-editing is becoming a popular molecular tool for functional genomics, as well as crop improvement. Clustered regularly interspaced palindromic repeats (CRISPR/CRISPR-associated protein 9 (Cas9)-mediated genome-editing has rapidly become the most popular genome engineering approach due to its simplicity, efficiency, specificity, multiplexing and ease to adapt. Most of the CRISPR/Cas9-based genome-editing is reported for seed crops; however, recently, it is also established for clonally propagated crops such as potato, banana, and cassava (Butler et al., 2016; Odipio et al., 2017; Kaur et al., 2018; Naim et al., 2018; Tripathi et al., 2019; Ntui et al., 2020). The accessibility of the genetic transformation system and reference genomes for cassava made it possible to realize the potential of CRISPR-based genome-editing for basic and applied research to improve economically important cassava traits. The CRISPR/Cas9-based genome-editing system was demonstrated by knocking out the *Phytoene desaturase* (*MePDS*) gene in two cultivars of cassava, TMS60444 and TME204 (Odipio et al., 2017). Genome-editing was further applied for developing a cassava variety with resistance against two species of Ipomovirus, *cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), causing

[3]NextGen cassava, https://www.nextgencassava.org

[4]https://excellenceinbreeding.org/module2

[5]https://www.cassavabase.org/

CBSD (Gomez et al., 2019). The targeted mutations in the translation initiation factor 4E (eIF4E) isoforms nCBP-1 and nCBP-2 in the edited cassava variety showed a reduction in disease severity and virus accumulation in the storage tuberous roots upon glasshouse challenge of edited cassava lines with CBSV. Even though the mutations in the nCBP-1 and nCBP-2 genes conferred enhanced resistance to CBSD, complete resistance was not obtained. This suggests that total resistance to CBSD can be developed by stacking the genome-editing approach of disrupting eIF4E isoforms with other resistance strategies such as RNAi (Gomez et al., 2019). Later, researchers have attempted to apply this technology to develop resistance against a geminivirus, *African cassava mosaic virus* (ACMV) (Mehta et al., 2019). However, the edited cassava plants did not show significant resistance against ACMV in greenhouse inoculation experiments. It might be due to the evolution of editing-resistant geminiviruses in genome-edited cassava. CRISPR/Cas9 based genome-editing can be coupled to genetic improvements in cassava for traits such as starch improvement and early flowering. Cassava roots normally produce large quantities of starch, having high amylose levels, a crystallizable component that is more soluble in water. However, the starch with low amylose levels, known as "waxy starch," is preferred for food processing and other industrial uses. Bull et al. (2018) reported the application of CRISPR/Cas9 for manipulating starch biosynthesis and improving the starch quality in the storage. They generated edited cassava plants with mutations in two genes: *protein targeting to starch* (PTST1) or *granule bound starch synthase* (GBSS) involved in amylose biosynthesis, leading to reduction or elimination of amylose content. This, in turn, can improve the quality of starch cassava roots for commercial use. The authors also demonstrated accelerated breeding by transferring the *Arabidopsis* FLOWERING LOCUS T gene in the genome-edited events of cassava for early flowering. Genome-editing can multiplex the traits, and researchers can develop cassava varieties with the waxy starch and early flowering. Despite still being in its infancy, genome-editing offers promising prospects for cassava improvement and could shorten the breeding process. Novel plant varieties could be directly used for crop production or as pre-breeding materials (Xu et al., 2019). Technological developments like this should be followed by their adoption to increase productivity.

CURRENT STATUS OF TECHNOLOGY APPLICATIONS IN SUB-SAHARAN AFRICA

Effective implementation of technological development is key to enhancing the productivity and profitability of cassava on the continent. Genetic studies conducted by national institutes or academia in sub-Saharan Africa have been mainly focused on germplasm characterization, genetic diversity assessment, varietal identification, linkage mapping, and classical QTL mapping (Fregene et al., 1997; Kawuki et al., 2009; Rabbi et al., 2015; Adjebeng-Danquah et al., 2020). Recently, collaborative projects involving international institutions, NARS, along with substantial funding from the donor communities, have contributed to the increased number of genome-wide association studies. Several gene(s)/QTLs underlying key cassava traits have been identified and trait-linked markers, a pre-requisite for MAS have been developed (Ogbonna et al., 2020; Rabbi et al., 2020). The effective implementation of MAS hampered by economic obstacles is currently being addressed with the support of the EIB platform[2], who seeks to mainstream the use of genotyping data. The platform through subsidies from the Gates Foundation offers small breeding programs access to high quality genotyping services, including low-density (Kompetitive allele specific PCR (KASP)) and mid-density (DArTAg) genotyping to foster the progressive integration of MAS into NARS crop breeding programs. The EIB low-density genotyping service is being used by IITA and NARS partners in Uganda, Tanzania, and Nigeria for quality control, identification of cassava accessions with the desired alleles, and validation of trait-markers. Genomic selection-based pilot projects are ongoing in Nigeria, Uganda, and Tanzania and the appealing perspectives offered by this approach, including shortening of the breeding cycle and speed-up of variety development have been highlighted (Wolfe et al., 2017; Kayondo et al., 2018; Somo et al., 2020). The mid-density genotyping, DArTAg, a target genotyping provided by Diversity Arrays is being used as an alternative to GBS for genomic selection applications. Although traditional methodologies are still widely used for trait phenotyping, integration of high throughput phenotyping is on course. NIRS spectroscopy is being explored by few NARS programs to predict some key cassava traits (i.e., carotenoids and DMC) and promising results have been reported (Ikeogu et al., 2017, 2019). NIRS evaluation and optimization for other traits, including gari, fufu, starch, and root mealiness is ongoing. In collaboration with IDS GeoRadar[6], IITA is testing a prototype commercial GPR for routine cassava root phenotyping. The unmanned aerial vehicle (UAV) is currently used at IITA by the Cassava Source-Sink Project to quantify aboveground plant growth (Sonnewald et al., 2020). Image recognition has been evaluated for high accuracy disease detection and mobile, as well as web-based tools developed for cassava disease scoring and monitoring[7]. An application programming interface has been implemented within Cassavabase in order to process cassava root images. Breeders can upload cassava root necrosis sectional image captured during harvest and the result will be returned back to Cassavabase. IITA and NARS partners have digitized all processes; data are uploaded, processed, and accessed within Cassavabase to minimize human errors. CRISPR/Cas9-based genome-editing is mainly driven by IITA, in partnership with a handful of national research organizations. The difficulty in acquiring laboratory supplies, as well as the need for constant and sufficient level of funding, constitute a bottleneck for the implementation of gene-editing

[6]https://idsgeoradar.com/
[7]http://www.mcrops.org/

technology by NARS. Likewise, the legal framework and appropriate regulatory structures needed to guide the use of this technology are still lacking in several African countries, hampering the movement of plants from laboratory to the field (Tripathi et al., 2020). Furthermore, biosafety considerations remain a public concern. Winning consumer's acceptance and trust is key to the effective implementation of this new breeding technique and to harness its full potential. Few cassava breeding programs have benefited and/or explored these recent technological developments. For greater impact, the knowledge and technologies outlined here should be disseminated and transferred to other cassava breeding programs on the continent.

TECHNOLOGY TRANSFER AND CAPACITY DEVELOPMENT FOR MORE SUSTAINABLE CASSAVA PRODUCTION

Sustained research and innovation capacity is imperative for agricultural transformation in Africa (Ojijo et al., 2016). Most African national agricultural research systems (NARS) do not have well-funded cassava breeding programs with sufficient technical critical mass. Many programs routinely evaluate and multiply clones imported from larger breeding programs (i.e., from IITA or CIAT). Although most NARS would maintain certain capacity to develop and release varieties adapted to local agro-ecologies and local preferences, their breeding capacities need to be strengthened and modernized to be effective. Many of the technological innovations are new to NARS and access to equipment, reagents, and skilled personnel is challenging (Tester and Langridge, 2010). It is with this in mind that an initiative such as the Cassava Community of Practice and Partnership (CoPP) has been established through the NextGen Cassava Project to serve as a platform to disseminate and facilitate the transfer of proven tools, methods, technologies, and products. The CoPP provides the initial technical backstopping, fosters collaboration, facilitates connectivity, and creates opportunities for peer-to-peer learning between members. This will enable the establishment of best practices and procedures and implementation of recent technological innovations by NARS, increase breeding efficiencies, and create a broader and deeper impact. Successful integration and application of innovative technologies necessitate technical expertise. Scaling up the research capabilities of NARS, through training, will hone their skills and knowledge (Bull et al., 2011). The CoPP concept is not new. A successful example is the sweet potato breeding Community of Practice, which has strengthened national capacities leading to the development and released of several user-preferred varieties with impact on the quality of life of small farmers (SASHA, 2019). The cassava CoPP will need to evolve into a cohesive network with clear accountabilities and expectations. NARS and CGIAR will need to work together in a coordinated breeding network for economies of scale. Government involvement will be needed to sustain the gain achieved by the breeders.

CONCLUSION AND PERSPECTIVES

It takes several years to develop improved cassava varieties. Genomic resources have facilitated the evaluation of local and regional genetic diversity and profiling of breeding materials. New phenotyping approaches are substituting traditional trait evaluation and have been used for marker-trait associations, enabling identification of numerous QTLs for key traits. Although notable signs of progress have been achieved, especially for less complex traits, genetic architecture is still not fully understood. For quantitatively inherited traits, phenotypic plasticity and the difficult phenotypic evaluation complicates matters further. There is a need to scale up multi-environment evaluation. The development of advanced high throughput and accurate cassava phenotyping approaches is imperative. Translating those results obtained into practical breeding methodologies and coherent biological knowledge is needed. QTL results should be exploited, and validated trait-markers developed. Priorities will have to be set to ensure return on investment. Therefore, establishment of a formal advancement system with well-defined metrics is needed. Genetic gain should be routinely monitored. Strategies need to be constantly revised as priorities evolve, and new challenges emerge. MAS and GS are complementary breeding approaches that should be used in tandem. More attention should be given to quality traits, which have received less attention; whereas quality traits influence varietal adoption and product utilization. Key quality traits need to be defined and translated into biochemical parameters. Advanced technologies, such as genome-editing, could play a prominent role in cassava improvement. However, further investigation will be required to ensure maximum benefits. It will be important that the new technologies and tools developed are used by NARS. Therefore, regional networks, shared expertise, and service will be of prime importance. Last, but not the least, the support and involvement of national and regional governments is crucial for sustainable cassava production on the continent.

AUTHOR CONTRIBUTIONS

EN conceived this review and drafted the manuscript. CE conceived the idea and edited the manuscript. LT drafted a section of the manuscript. MF, IR, EH, KSI, PK, and CE provided critical edits. All authors contributed to the article and approved the submitted version.

FUNDING

The authors thank the UK's Foreign, Commonwealth and Development Office (FCDO) and the Bill and Melinda Gates Foundation (Grant INV-007637, http://www.gatesfoundation.org) for their financial support.

REFERENCES

Abdullakasim, W., Powbunthorn, K., Unartngam, J., and Takigawa, T. (2011). An images analysis technique for recognition of brown leaf spot disease in cassava. *Tarım Makinaları Bilim. Derg.* 7, 165–169.

Adinsi, L., Akissoé, N., Escobar, A., Prin, L., Kougblenou, N., Dufour, D., et al. (2019). Sensory and physicochemical profiling of traditional and enriched gari in Benin. *Food Sci. Nutr.* 00, 1–11. doi: 10.1002/fsn3.1201

Adjebeng-Danquah, J., Manu-Aduening, J., Asante, I. K., Agyare, R. Y., Gracen, V., and Offei, S. K. (2020). Genetic diversity and population structure analysis of Ghanaian and exotic cassava accessions using simple sequence repeat (SSR) markers. *Heliyon* 6:e03154. doi: 10.1016/j.heliyon.2019.e03154

Afonso, T., Moresco, R., Uarrota, V. G., Navarro, B. B., da Nunes, E. C. N., Maraschin, M., et al. (2017). UV-Vis and CIELAB based chemometric characterization of *Manihot esculenta* carotenoid contents*. *J. Integr. Bioinform.* 14:20170056. doi: 10.1515/jib-2017-0056

Akano, A. O., Dixon, A. G. O., Mba, C., Barrera, E., and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105, 521–525. doi: 10.1007/s00122-002-0891-7

Akely, P. M. T., Djina, Y., Konan, B. R., Irie, K., Kouame, L. P., and Amani, N. G. (2016). Study of Varietal Influence Post Conservation on Biochemical and Sensory Qualities of Attiéké and Boiled Cassava (*Manihot esculenta* Crantz). *Agric. Sci.* 7, 127–136. doi: 10.4236/as.2016.73012

Akinbo, O., Gedil, M., Ekpo, E. J. A., Oladele, J., and Dixon, A. G. O. (2007). Detection of RAPD markers-linked to resistance to cassava anthracnose disease. *African J. Biotechnol.* 6, 677–682. doi: 10.5897/AJB2007.000-2068

Akinpelu, A. O., Amamgbo, L. E. F., Olojede, A. O., and Oyekale, A. S. (2011). Health implications of cassava production and consumption. *J. Agric. Soc. Res.* 11, 118–125.

Alicai, T., Szyniszewska, A. M., Omongo, C. A., Abidrabo, P., Okao-Okuja, G., Baguma, Y., et al. (2019). Expansion of the cassava brown streak pandemic in Uganda revealed by annual field survey data for 2004 to 2017. *Sci. Data* 6:327. doi: 10.1038/s41597-019-0334-9

An, O., Uj, U., and Re, K. (2019). Physicochemical properties of cassava processing residue flour and sensory evaluation of fufu prepared from it. *J. Nutraceuticals Food Sci.* 4, 1–6.

Anderson, B., Eghan, M. J., Asare-Bediako, E., and Buah-Bassuah, P. K. (2015). Optical imaging method for determining symtoms severity of cassava mosaic disease. *Appl. Phys. Res.* 7:34.

Ayetigbo, O., Latif, S., Abass, A., and Müller, J. (2018). Comparing characteristics of root, flour and starch of biofortified yellow-flesh and white-flesh cassava variants, and sustainability considerations: A review. *Sustain* 10:3089. doi: 10.3390/su10093089

Bechoff, A., Tomlins, K., Fliedel, G., Becerra Lopez-lavalle, L. A., Westby, A., Hershey, C., et al. (2018). Cassava traits and end-user preference: Relating traits to consumer liking, sensory perception, and genetics. *Crit. Rev. Food Sci. Nutr.* 58, 547–567. doi: 10.1080/10408398.2016.1202888

Belalcazar, J., Dufour, D., Andersson, M. S., Pizarro, M., Luna, J., Londoño, L., et al. (2016). High-throughput phenotyping and improvements in breeding cassava for increased carotenoids in the roots. *Crop Sci.* 56, 2916–2925. doi: 10.2135/cropsci2015.11.0701

Blair, M. W., Fregene, M. A., Beebe, S. E., and Ceballos, H. (2007). "Marker-assisted selection in common beans and cassava," in *Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish*, eds E. Guimarães, J. Ruane, B. Scherf, A. Sonnino, and J. Dargie (Rome: FAO), 81–115.

Boonpo, S., and Kungwankunakorn, S. (2017). Study on amylose iodine complex from cassava starch by colorimetric method. *J. Adv. Agric. Technol.* 4, 345–349. doi: 10.18178/joaat.4.4.345-349

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–571. doi: 10.1038/nbt.3535

Brito, A. C., Oliveira, S. A. S., and Oliveira, E. J. (2017). Genome-wide association study for resistance to cassava root rot. *J. Agric. Sci.* 155, 1424–1441. doi: 10.1017/S0021859617000612

Bull, S. E., Ndunguru, J., Gruissem, W., Beeching, J. R., and Vanderschuren, H. (2011). Cassava: Constraints to production and the transfer of biotechnology to African laboratories. *Plant Cell Rep.* 30, 779–787. doi: 10.1007/s00299-010-0986-6

Bull, S. E., Seung, D., Chanez, C., Mehta, D., Kuon, J. E., Truernit, E., et al. (2018). Accelerated ex situ breeding of GBSS- and PTST1-edited cassava for modified starch. *Sci. Adv.* 4:eaat6086. doi: 10.1126/sciadv.aat6086

Butler, N. M., Baltes, N. J., Voytas, D. F., and Douches, D. S. (2016). Geminivirus-mediated genome editing in potato (*Solanum tuberosum* l.) *using sequence-specific nucleases*. *Front. Plant Sci.* 7:1045. doi: 10.3389/fpls.2016.01045

Carvalho, L. M. J., Oliveira, A. R. G., Godoy, R. L. O., Pacheco, S., Nutti, M. R., De Carvalho, J. L. V., et al. (2012). Retention of total carotenoid and β-carotene in yellow sweet cassava (*Manihot esculenta* Crantz) after domestic cooking. *Food Nutr. Res.* 56:15788. doi: 10.3402/fnr.v56i0.15788

Ceballos, H., Pérez, J. C., Barandica, O. J., Lenis, J. I., Morante, N., Calle, F., et al. (2016). Cassava breeding I: The value of breeding value. *Front. Plant Sci.* 7:1227. doi: 10.3389/fpls.2016.01227

Chatpapamon, C., Wandee, Y., Uttapap, D., Puttanlek, C., and Rungsardthong, V. (2019). Pasting properties of cassava starch modified by heat-moisture treatment under acidic and alkaline pH environments. *Carbohydr. Polym.* 215, 338–347. doi: 10.1016/j.carbpol.2019.03.089

Chiang, K. S., Bock, C. H., El Jarroudi, M., Delfosse, P., Lee, I. H., and Liu, H. I. (2016). Effects of rater bias and assessment method on disease severity estimation with regard to hypothesis testing. *Plant Pathol.* 65, 523–535. doi: 10.1111/ppa.12435

Chisenga, S. M., Workneh, T. S., Bultosa, G., and Laing, M. (2019). Characterization of physicochemical properties of starches from improved cassava varieties grown in Zambia. *AIMS Agric. Food* 4, 939–966. doi: 10.3934/agrfood.2019.4.939

Choperena, E. D. P. M., Ospina, C., Fregene, M., Montoya-Lerma, J., and Bellotti, A. C. (2012). Identificación de microsatélites en yuca asociados con la resistencia al ácaro Mononychellus tanajoa (acari: Tetranychidae). *Rev. Colomb. Entomol.* 38, 70–75.

Chueamchaitrakun, P., Chompreeda, P., Haruthaithanasan, V., Suwonsichon, T., and Kasemsamran, S. (2011). Prediction of pasting and thermal properties of mixed Hom-Mali and glutinous rice flours using near infrared spectroscopy. *Kasetsart J. Nat. Sci.* 45, 481–489.

Chukwu, S. C., Rafii, M. Y., Ramlee, S. I., Ismail, S. I., Oladosu, Y., Okporie, E., et al. (2019). Marker-assisted selection and gene pyramiding for resistance to bacterial leaf blight disease of rice (*Oryza sativa* L.). *Biotechnol. Biotechnol. Equip.* 33, 440–455. doi: 10.1080/13102818.2019.1584054

Cobb, J. N., Biswas, P. S., and Platten, J. D. (2019). Back to the future: revisiting MAS as a tool for modern plant breeding. *Theor. Appl. Genet.* 132, 647–667. doi: 10.1007/s00122-018-32663264

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142, 169–196. doi: 10.1007/s10681-005-1681-5

Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Cozzolino, D., Roumeliotis, S., and Eglinton, J. (2013). Relationships between swelling power, water solubility and near-infrared spectra in whole grain barley: A feasibility study. *Food Bioprocess Technol.* 6, 2732–2738. doi: 10.1007/s11947-012-0948-9

de Andrade, L. R. B., Bandeirae Sousa, M., Oliveira, E. J., de Resende, M. D. V., and Azevedo, C. F. (2019). Cassava yield traits predicted by genomic selection methods. *PLoS One* 14:e0224920. doi: 10.1371/journal.pone.0224920

de Oliveira, E. J., de Resende, M. D. V., da Silva, Santos, V., Ferreira, C. F., Oliveira, G. A. F., et al. (2012). Genome-wide selection in cassava. *Euphytica* 187, 263–276. doi: 10.1007/s10681-012-0722-0

Delgado, A., Hays, D. B., Bruton, R. K., Ceballos, H., Novo, A., Boi, E., et al. (2017). Ground penetrating radar: A case study for estimating root bulking rate in cassava (Manihot esculenta Crantz). *Plant Methods* 13:65. doi: 10.1186/s13007-017-0216-0

Delgado, A., Novo, A., and Hays, D. B. (2019). Data acquisition methodologies utilizing ground penetrating radar for cassava (*Manihot esculenta* Crantz) root architecture. *Geosci* 9:171. doi: 10.3390/geosciences9040171

do Carmo, C. D., da Silva, M. S., Oliveira, G. A. F., and de Oliveira, E. J. (2015). Molecular-assisted selection for resistance to cassava mosaic disease in *Manihot esculenta* Crantz. *Sci. Agric.* 72, 520–527. doi: 10.1590/0103-9016-2014-0348

do Carmo, C. D., e Sousa, M. B., Brito, A. C., and de Oliveira, E. J. (2020). Genome-wide association studies for waxy starch in cassava. *Euphytica* 216:82. doi: 10.1007/s10681-020-02615-9

Donkor, E., Onakuse, S., Bogue, J., de Los Rios, and Carmenado, I. (2017). The impact of the presidential cassava initiative on cassava productivity in Nigeria: Implication for sustainable food supply and food security. *Cogent Food Agric.* 3:1368857. doi: 10.1080/23311932.2017.1368857

Echefu, S. U., Rabbi, I. Y., Nwachukwu, E. C., and Adamu, C. L. Y. (2016). Mapping of Quantitative Trait Loci Associated With Cassava Mosaic Disease (Cmd) Resistance. *Int. J. Sci. Environ.* 5, 2637–2645

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Esuma, W., Kawuki, R. S., Herselman, L., and Labuschagne, M. T. (2016). Diallel analysis of provitamin A carotenoid and dry matter content in cassava (Manihot esculenta Crantz). *Breed. Sci.* 66, 627–635. doi: 10.1270/jsbbs.15159

Ezenwaka, L., Carpio, D. P., Del, Jannink, J. L., Rabbi, I., Danquah, E., et al. (2018). Genome-wide association study of resistance to cassava green mite pest and related traits in cassava. *Crop Sci.* 58, 1–12. doi: 10.2135/cropsci2018.01.0024

Ezenwaka, L., Rabbi, I., Onyeka, J., Kulakow, P., and Egesi, C. (2020). Identification of additional /novel QTL associated with resistance to cassava green mite in a biparental mapping population. *PLoS One* 15:e0231008. doi: 10.1371/journal.pone.0231008

Fanou, A. A., Zinsou, V. A., and Wydra, K. (2018). *Cassava bacterial blight: a devastating disease of cassava*. London: Intechopen, 14–36.

FAOSTAT. (2020). http://www.fao.org/faostat/en/#data/QC (Accessed January 24, 2020).

Feleke, S., Manyong, V., Abdoulaye, T., and Alene, A. D. (2016). Assessing the impacts of cassava technology on poverty reduction in Africa. *Stud. Agric. Econ.* 118, 101–111. doi: 10.7896/j.1612

Ferguson, M. E., Hearne, S. J., Close, T. J., Wanamaker, S., Moskal, W. A., Town, C. D., et al. (2012). Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor. Appl. Genet.* 124, 685–695. doi: 10.1007/s00122-011-1739-9

Ferguson, M. E., Shah, T., Kulakow, P., and Ceballos, H. (2019). A global overview of cassava genetic diversity. *PLoS One* 14:e0224763. doi: 10.1371/journal.pone.0224763

Ferguson, M., Rabbi, I., Kim, D., Gedil, M., Lopez-Lavalle, L. A. B., and Okogbenin, E. (2011). Molecular Markers and their Application to Cassava Breeding: Past, Present and Future. *Trop. Plant Biol.* 5, 95–109.

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., et al. (2015). The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. *Nucleic Acids Res.* 43, D1036–D1041. doi: 10.1093/nar/gku1195

Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., et al. (1997). A molecular genetic map of cassava (Manihot esculenta crantz). *Theor. Appl. Genet.* 95, 431–441.

FSIN (2019). *Global report on food crises*. Available online at: https://www.wfp.org/publications/2020-global-report-food-crises (Accessed on October 23, 2020)

Fukuda, W. M. G., Guevara, C. L., Kawuki, R., and Ferguson, M. E. (2010). *Selected morphological and agronomic descriptors for the characterization of cassava*. Ibadan:, NigeriaIITA, 19.

Fukushima, A. R., Nicoletti, M. A., Rodrigues, A. J., Pressutti, C., Almeida, J., Brandão, T., et al. (2016). Cassava flour: Quantification of cyanide content. *Food Nutr. Sci.* 7, 592–599. doi: 10.4236/fns.2016.77060

Garcia-Oliveira, A. L., Kimata, B., Kasele, S., Kapinga, F., Masumba, E., Mkamilo, G., et al. (2020). Genetic analysis and QTL mapping for multiple biotic stress resistance in cassava. *PLoS One* 15:e0236674. doi: 10.1371/journal.pone.0236674

Gegios, A., Amthor, R., Maziya-Dixon, B., Egesi, C., Mallowa, S., Nungo, R., et al. (2010). Children consuming cassava as a staple food are at risk for inadequate zinc, iron, and vitamin A intake. *Plant Foods Hum. Nutr.* 65, 64–70. doi: 10.1007/s11130-010-0157155

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Hayden, L., Renninger, K., et al. (2019). Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* 17, 421–434. doi: 10.1111/pbi.12987

ICGMC. (2015). High-resolution linkage map and chromosome-scale genome assembly for cassava (Manihot esculenta crantz) from 10 populations. *G3 Genes Genomes Genet* 5, 133–144. doi: 10.1534/g3.114.015008

Iglesias, C., Mayer, J., Chavez, L., and Calle, F. (1997). Genetic potential and stability of carotene content in cassava roots. *Euphytica* 94, 367–373. doi: 10.1023/A:1002962108315

Ikeogu, U. N., Akdemir, D., Wolfe, M. D., Okeke, U. G., Chinedozi, A., Jannink, J. L., et al. (2019). Genetic correlation, genome-wide association and genomic prediction of portable nirs predicted carotenoids in cassava roots. *Front. Plant Sci.* 10:1570. doi: 10.3389/fpls.2019.01570

Ikeogu, U. N., Davrieux, F., Dufour, D., Ceballos, H., Egesi, C. N., and Jannink, J. L. (2017). Rapid analyses of dry matter content and carotenoids in fresh cassava roots using a portable visible and near infrared spectrometer (Vis/NIRS). *PLoS One* 12:e0188918.

Iliya, J. M., and Madumelu, M. (2019). Determination of cyanide conetent in cassava tubers (Manihot esculenta) and apple seed (Pyrus malus). *Int. J. Curr. Res. Chem. Pharm. Sci.* 6, 14–20. doi: 10.22192/ijcr

Jaramillo, A. M., Londoño, L. F., Orozco, J. C., Patiño, G., Belalcazar, J., Davrieux, F., et al. (2018). A comparison study of five different methods to measure carotenoids in biofortified yellow cassava (*Manihot esculenta*). *PLoS One.* 13:e0209702. doi: 10.1371/journal.pone.\underbar{0209702}A ML. FJ. CGJF

Jonas, E., and de Koning, D. J. (2013). Does genomic selection have a future in plant breeding? *Trends Biotechnol.* 31, 497–504. doi: 10.1016/j.tibtech.2013.06.003

Jorge, V., Fregene, M. A., Duque, M. C., Bonierbale, M. W., Tohme, J., and Verdier, V. (2000). Genetic mapping of resistance to bacterial blight disease in cassava (Manihot esculenta Crantz). *Theor. Appl. Genet.* 101, 865–872. doi: 10.1007/s001220051554

Jorge, V., Fregene, M., Vélez, C. M., Duque, M. C., Tohme, J., and Verdier, V. (2001). QTL analysis of field resistance to Xanthomonas axonopodis pv. *manihotis in cassava*. *Theor. Appl. Genet.* 102, 564–571. doi: 10.1007/s001220051683

Kainuma, K., Odat, T., and Cuzuki, S. (1967). Study of starch phosphates monoesters. *J. Tech. Soc. Starch* 14, 24–28

Kalyebi, A., MacFadyen, S., Parry, H., Tay, W. T., De Barro, P., and Colvin, J. (2018). African cassava whitefly, Bemisia tabaci, cassava colonization preferences and control implications. *PLoS One* 13:e0204862. doi: 10.1371/journal.pone.0204862

Kamanda, I., Blay, E. T., Asante, I. K., Danquah, A., Ifie, B. E., Parkes, E., et al. (2020). Genetic diversity of provitamin-A cassava (*Manihot esculenta* Crantz) in Sierra Leone. *Genet. Resour. Crop Evol.* 67, 1193–1208. doi: 10.1007/s10722-020-00905-8

Kaur, N., Alok, A. S., Kaur, N., Pandey, P., Awasthi, P., et al. (2018). CRISPR/Cas9-mediated efficient editing in phytoene desaturase (PDS) demonstrates precise manipulation in banana cv. *Rasthali genome. Funct. Integr. Genom.* 18, 89–99. doi: 10.1007/s10142-017-0577-5

Kawuki, R. S., Esuma, W., Ozimati, A., Kayondo, I. S., Nandudu, L., and Wolfe, M. (2019). Alternative approaches for assessing cassava brown streak root necrosis to guide resistance breeding and selection. *Front. Plant Sci.* 10:1461. doi: 10.3389/fpls.2019.01461

Kawuki, R. S., Ferguson, M., Labuschagne, M., Herselman, L., and Kim, D. J. (2009). Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Mol. Breed.* 23, 669–684. doi: 10.1007/s11032-009-9264-0

Kayondo, S. I., Pino Del, Carpio, D., Lozano, R., Ozimati, A., Wolfe, M., et al. (2018). Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. *Sci. Rep.* 8:1549. doi: 10.1038/s41598-018-19696-1

Kengkanna, J., Jakaew, P., Amawan, S., Busener, N., Bucksch, A., and Saengwilai, P. (2019). Phenotypic variation of cassava root traits and their responses to drought. *Appl. Plant Sci.* 7:e1238. doi: 10.1002/aps3.1238

Khandare, V., and Choomsook, P. (2019). Cassava export of Thailand: Growth performance and composition. *Int. J. Res. Anal. Rev.* 6, 847–857.

Kizito, E. B., Rönnberg-Wästljung, A.-C., Egwang, T., Gullberg, U., Fregene, M., and Westerbergh, A. (2007). Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. *Hereditas* 144, 129–136. doi: 10.1111/j.2007.0018-0661.01975.x

Koros, J. C., Runo, S. M., Yusuf, M., and Orek, C. O. (2018). Screening selected cassava cultivars for resistance against cassava viruses and cassava green mites under advanced yield trials in Kenya. *IOSR J. Biotechnol. Biochem.* 4, 37–52. doi: 10.9790/264X-0405023752

Kretzschmar, T., Mbanjo, E. G. N., Magalit, G. A., Dwiyanti, M. S., Habib, M. A., Diaz, M. G., et al. (2018). DNA fingerprinting at farm level maps rice biodiversity across Bangladesh and reveals regional varietal preferences. *Sci. Rep.* 8:14920. doi: 10.1038/s41598-018-33080-z

Kuon, J. E., Qi, W., Schläpfer, P., Hirsch-Hoffmann, M., Von Bieberstein, P. R., Patrignani, A., et al. (2019). Haplotype-resolved genomes of geminivirus-resistant and geminivirus-susceptible African cassava cultivars. *BMC Biol.* 17:75. doi: 10.1186/s12915-019-0697-6

Lan, S., Zheng, C., Hauck, K., McCausland, M., Duguid, S. D., Booker, H. M., et al. (2020). Genomic prediction accuracy of seven breeding selection traits improved by QTL identification in flax. *Int. J. Mol. Sci.* 21:1577. doi: 10.3390/ijms21051577

Leach, H. W. (1959). Structure of the starch granule I swelling and solubility patterns of various starches. *Cereal Chem.* 36, 534–544.

Lu, G., Huang, H., and Zhang, D. (2006a). Application of near-infrared spectroscopy to predict sweetpotato starch thermal properties and noodle quality. *J. Zhejiang Univ. Sci. B.* 7, 475–481. doi: 10.1631/jzus.2006.B0475

Lu, G., Huang, H., and Zhang, D. (2006b). Prediction of sweetpotato starch physiochemical quality and pasting properties using near-infrared reflectance spectroscopy. *Food Chem.* 94, 632–639. doi: 10.1016/j.foodchem.2005.02.006

Luo, X., Tomlins, K. I., Carvalho, L. J. C. B., Li, K., and Chen, S. (2018). The analysis of candidate genes and loci involved with carotenoid metabolism in cassava (*Manihot esculenta* Crantz) using SLAF-seq. *Acta Physiol. Plant.* 40:66. doi: 10.1007/s11738-018-2634-7

Ly, D., Hamblin, M., Rabbi, I., Melaku, G., Bakare, M., Gauch, H. G., et al. (2013). Relatedness and genotype × environment interaction affect prediction accuracies in genomic selection: A study in cassava. *Crop Sci.* 53, 1312–1325. doi: 10.2135/cropsci2012.11.0653

Ma'Aruf, A. G., and Abdul, H. R. (2020). Efficient processing of cassava starch: Physicochemical characterization at different processing parameters. *Food Res.* 4, 143–151. doi: 10.26656/fr.2017.4(1).235

Maieves, H. A., Oliveira, D. C., De, Bernardo, C., Müller, C. M. D. O., and Amante, E. R. (2012). Microscopy and texture of raw and cooked cassava (*Manihot esculenta* Crantz) roots. *J. Texture Stud.* 43, 164–173. doi: 10.1111/j.1745-4603.2011.00327.x

Malik, A. I, Kongsil, P., Nguyễn, V. A., Ou, W., and Srean, P. (2020). Cassava breeding and agronomy in Asia: 50 years of history and future directions. *Breed. Sci.* 70, 145–166. doi: 10.1270/jsbbs.18180

Masumba, E. A., Kapinga, F., Mkamilo, G., Salum, K., Kulembeka, H., Rounsley, S., et al. (2017). QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. *Theor. Appl. Genet.* 130, 2069–2090. doi: 10.1007/s00122-017-2943-z

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S. E. A., Hirsch-Hoffmann, M., Gruissem, W., et al. (2019). Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biol.* 20:80. doi: 10.1186/s13059-019-1678-3

Meullenet, J. F., Mauromoustakos, A., Horner, T. B., and Marks, B. P. (2002). Prediction of texture of cooked white rice by near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chem.* 79, 52–57. doi: 10.1094/CCHEM.2002.79.1.52

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Moriasi, G. A., Olela, B. O., Waiganjo, B. W., Wakori, E. W. T., and Onyancha, J. M. (2017). Evaluation of cyanide levels in two cassava varieties (Mariwa and Nyakatanegi) grown in Bar-agulu, Siaya County, Kenya. *J. Food Nutr. Res.* 5, 817–823. doi: 10.12691/jfnr-5-1114

Morillo, A. C. C., Morillo, Yacenia, C., and Ceballos, H. L. (2013). Identification of QTLs for carotene content in the genome of cassava (*Manihot esculenta* Crantz) and S1 population validation. *Acta Agron.* 62, 196–206.

Mtunguja, M. K., Beckles, D. M., Laswai, H. S., Ndunguru, J. C., and Sinha, N. J. (2019). Opportunities to commercialize cassava production for poverty alleviation and improved food security in Tanzania. *African J. Food, Agric. Nutr. Dev.* 19, 13928–13946. doi: 10.18697/AJFAND.84.BLFB1037

Naim, F., Dugdale, B., Kleidon, J., Brinin, A., Shand, K., Waterhouse, P., et al. (2018). Gene editing the phytoene desaturase alleles of Cavendish banana using CRISPR/Cas9. *Transgenic Res.* 27, 451–460. doi: 10.1007/s11248-018-0083-0

Nakatumba-Nabende, J., Akera, B., Tusubira, J. F., Nsumba, S., and Mwebaze, E. (2020). A dataset of necrotized cassava root cross-section images. *Data Br.* 32:106170. doi: 10.1016/j.dib.2020.106170

Nassar, N., and Ortiz, R. (2010). Breeding Cassava to Feed the Poor. *Sci. Am.* 302, 78–85. doi: 10.1002/9781118060995.ch3

Ni, J., Pujar, A., Youens-Clark, K., Yap, I., Jaiswal, P., Tecle, I., et al. (2009). Gramene QTL database: Development, content and applications. *Database* 2009:ba005. doi: 10.1093/database/bap005

Noranizan, M. A., Dzulkifly, M. H., and Russly, A. R. (2010). Effect of heat treatment on the physico-chemical properties of starch from different botanical sources. *Int. Food Res. J.* 17, 127–135.

Ntui, V. O., Tripathi, J. N., and Tripathi, L. (2020). Robust CRISPR/Cas9 mediated genome editing tool for banana and plantain (*Musa* spp.). *Curr. Plant Biol.* 2020:100128. doi: 10.1016/j.cpb.2019.100128

Nzuki, I., Katari, M. S., Bredeson, J. V., Masumba, E., Kapinga, F., Salum, K., et al. (2017). QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from Manihot Glaziovii. *Front. Plant Sci.* 8:1168. doi: 10.3389/fpls.2017.01168

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D. A., Bart, R., and Taylor, N. J. (2017). Efficient CRISPR/cas9 genome editing of phytoene desaturase in cassava. *Front. Plant Sci.* 8:1780. doi: 10.3389/fpls.2017.01780

Ogbonna, A. C., de Andrade, L. R., Rabbi, I. Y., Mueller, L. A., de Oliveira, E. J., and Bauchet, G. J. (2020).). Genetic architecture and gene mapping of cyanide in cassava (*Manihot esculenta* Crantz). *bioRxiv preprint* doi: \url{https://org/10.1101/2020.06.19.159160}

Ojijo, N., Franzel, S., Simtowe, F., Madakadze, R., Nkwake, A., and Moleko, L. (2016). "The role for agricultural Research Systems, Advisory services and capacity developement and knowledge transfer," in *Africa agriculture status report 2016: progress towards agricultural transformation* (AGRA). (Uttar Pradesh: AGRA)200–230.

Okogbenin, E., Egesi, C. N., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., et al. (2012). Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* 52, 2576–2586. doi: 10.2135/cropsci2011.11.0586

Okogbenin, E., Marin, J., and Fregene, M. (2006). An SSR-based molecular genetic map of cassava. *Euphytica* 147, 433–440

Okogbenin, E., Marin, J., and Fregene, M. (2008). QTL analysis for early yield in a pseudo F2 population of cassava. *Afr. J. Biotechnol.* 7, 131–138. doi: 10.5897/AJB2008.000-5017

Otekunrin, O. A., and Sawicka, B. (2019). Cassava, a 21st century staple crop: How can Nigeria harness its enormous trade potentials? *Acta Sci. Agric.* 3, 194–202. doi: 10.31080/asag.2019.03.0586

Owomugisha, G., and Mwebaze, E. (2016). Machine learning for plant disease incidence and severity measurements from leaf images. in *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications. ICMLA* 2016, 158–163. doi: 10.1109/ICMLA.2016.126

Ozimati, A., Kawuki, R., Esuma, W., Kayondo, I. S., Wolfe, M., Lozano, R., et al. (2018). Training population optimization for prediction of cassava brown streak disease resistance in West African Clones. *G3 Genes Genomes Genet.* 8, 3903–3913. doi: 10.1534/g3.118.200710

Patil, B. L., Legg, J. P., Kanju, E., and Fauquet, C. M. (2015). Cassava brown streak disease: a threat to food security in Africa. *J. Gen. Virol.* 96, 956–968. doi: 10.1099/jgv.0.000014

Ping, H., Wang, J., and Ren, G. (2013). Prediction of the total starch and amylose content in barley using near-infrared reflectance spectroscopy. *Intell. Autom. Soft. Comput.* 19, 231–237. doi: 10.1080/10798587.2013.823719

Pootakham, W., Shearman, J. R., Ruang-Areerate, P., Sonthirod, C., Sangsrakru, D., Jomchai, N., et al. (2014). Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). *PLoS One* 9:e116028. doi: 10.1371/journal.pone.0116028

Prohens, J. (2011). Plant breeding: a success story to be continued thanks to the advances in genomics. *Front. Plant Sci.* 2:51. doi: 10.3389/fpls.2011.00051

Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Gedil, M. A., Ikpan, A. S., Jannink, J. L., et al. (2014). High-resolution mapping of resistance to cassava mosaic

geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* 186, 87–96. doi: 10.1016/j.virusres.2013.12.028

Rabbi, I. Y., Kayondo, S. I., Bauchet, G., Yusuf, M., Aghogho, C. I., Ogunpaimo, K., et al. (2020). Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. *Plant Mol. Biol.* 30:20. doi: 10.1007/s11103-020-01038-3

Rabbi, I. Y., Kulakow, P. A., Manu-Aduening, J. A., Dankyi, A. A., Asibuo, J. Y., Parkes, E. Y., et al. (2015). Tracking crop varieties using genotyping-by-sequencing markers: A case study using cassava (*Manihot esculenta* Crantz). *BMC Genet.* 16:115. doi: 10.1186/s12863-015-0273-1

Rabbi, I. Y., Kulembeka, H. P., Masumba, E., Marri, P. R., and Ferguson, M. (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 125, 329–342. doi: 10.1007/s00122-012-18361834

Rabbi, I. Y., Udoh, L. I., Wolfe, M., Parkes, E. Y., Gedil, M. A., Dixon, A., et al. (2017). Genome-wide association mapping of correlated traits in cassava: Dry matter and total carotenoid content. *Plant Genome* 3:10. doi: 10.3835/plantgenome2016.09.0094

Ragot, M., Bonierbale, M., Weltzein, E., and Genetics, N. F. (2018). *From market demand to breeding decisions: A framework*. Lima: CGIAR Gender and Breedinfg Initiative.

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Front. Plant Sci.* 8:1852. doi: 10.3389/fpls.2017.01852

Ramcharan, A., McCloskey, P., Baranowski, K., Mbilinyi, N., Mrisho, L., Ndalahwa, M., et al. (2019). A mobile-based deep learning model for cassava disease diagnosis. *Front. Plant Sci.* 10:272. doi: 10.3389/fpls.2019.00272

Ramu, P., Esuma, W., Kawuki, R., Rabbi, I. Y., Egesi, C., Bredeson, J. V., et al. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* 49, 959–963. doi: 10.1038/ng.3845

Resende, M. D., Resende, M. F. Jr., Sansaloni, C. P., Petroli, C. D., Missiaggia, A. A., Aguiar, A. M., et al. (2012). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* 194, 116–128.

Rodríguez-Sandoval, E., Fernández-Quintero, A., Sandoval-Aldana, A., and Cuvelier, G. (2008). Effect of processing conditions on the texture of reconstituted cassava dough. *Brazilian J. Chem. Eng.* 25, 713–722. doi: 10.1590/S0104-66322008000400008

Rosales-soto, M. U., Ross, C. F., Younce, F., John, K., Mattinson, D. S., Huber, K., et al. (2016). Advances in food technology and nutritional sciences physico-chemical and sensory evaluation cassava (*Manihot Esculenta* Crantz) Flour. *Adv. Food Technol. Nutr. Sci. Open J.* 2, 9–18. doi: 10.17140/AFTnSOJ-2-126

Said, J. I., Knapka, J. A., Song, M., and Zhang, J. (2015). Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* × *G. barbadense populations*. *Mol. Genet. Genom.* 290, 1615–1625. doi: 10.1007/s00438-015-1021-y

Sakurai, T., Mochida, K., Yoshida, T., Akiyama, K., Ishitani, M., Seki, M., et al. (2013). Genome-wide discovery and information resource development of DNA polymorphisms in cassava. *PLoS One* 8:e74056. doi: 10.1371/journal.pone.0074056

Sambasivam, G., and Opiyo, G. D. (2020). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egypt. Inform. J.* 3:119. doi: 10.1016/j.eij.2020.02.007

Sánchez, T., Ceballos, H., Dufour, D., Ortiz, D., Morante, N., Calle, F., et al. (2014). Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. *Food Chem.* 151, 444–451. doi: 10.1016/j.foodchem.2013.11.081

Sandoval-Aldana, A., and Fernandez, Q. A. (2013). Physicochemical characterization of two cassava (*Manihot esculenta* Crantz) starches and flours. *Sci. Agroaliment.* 1, 19–25.

Santantonio, N., Atanda, S. A., Beyene, Y., Varshney, R. K., Olsen, M., Jones, E., et al. (2020). Strategies for effective use of genomic information in crop breeding programs serving Africa and South Asia. *Front. Plant Sci.* 11:353. doi: 10.3389/fpls.2020.00353

SASHA (2019). *The sweetpotato breeding community of practice: Breeding in Africa for Africa (2009-2019)*. Available online at: https://www.sweetpotatoknowledge.org/wp-content/uploads/2019/12/01_SASHA_BREEDING_The-Sweetpotato-Breeding-Community-Practice-Breeding-Africa-2009-2019-4p_web.pdf (accessed on October 23, 2020).

Sedano, J. C. S., Moreno, R. E. M., Mathew, B., Léon, J., Cano, F. A. G., Ballvora, A., et al. (2017). Major novel QTL for resistance to Cassava bacterial blight identified through a multi-environmental analysis. *Front. Plant Sci.* 8:1169. doi: 10.3389/fpls.2017.01169

Sedano, J. S., Eduardo, R., Moreno, M., Calle, F., Ernesto, C., and Carrascal, L. (2017). QTL identification for cassava bacterial blight under natural infection conditions. *Acta Biol. Colomb.* 22, 19–26

Singh, A. K., Krishnan, S. G., Vinod, K. K., Ellur, R. K., Bollinedi, H., Bhowmick, P. K., et al. (2019). Precision breeding with genomic tools: A decade long journey of molecular breeding in rice. *Indian J. Genet.* 79, 181–191. doi: 10.31742/ijgpb.79s.1.7

Somo, M., Kulembeka, H., Mtunda, K., Mrema, E., Salum, K., Wolfe, M. D., et al. (2020). Genomic prediction and quantitative trait locus discovery in a cassava training population constructed from multiple breeding stages. *Crop Sci.* 60, 896–913. doi: 10.1002/csc2.20003

Sonnewald, U., Fernie, A. R., Gruissem, W., Schläpfer, P., Anjanappa, R. B., Chang, S. H., et al. (2020). The Cassava Source-Sink project: opportunities and challenges for crop improvement by metabolic engineering. *Plant J.* 103, 1655–1665. doi: 10.1111/tpj.14865

Soto, J. C., Ortiz, J. F., Perlaza-Jiménez, L., Vásquez, A. X., Lopez-Lavalle, L. A. B., Mathew, B., et al. (2015). A genetic map of cassava (Manihot esculenta Crantz) with integrated physical mapping of immunity-related genes. *BMC Genom.* 16:190. doi: 10.1186/s12864-015-13971394

Spencer, D. S. C., and Ezedinma, C. (2017). Cassava cultivation in sub-Saharan Africa. 123–148.

Sraphet, S., Boonchanawiwat, A., Thanyasiriwat, T., Boonseng, O., Tabata, S., Sasamoto, S., et al. (2011). SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 122, 1161–1170. doi: 10.1007/s00122-010-1520-5

Sraphet, S., Boonchanawiwat, A., Thanyasiriwat, T., Thaikert, R., Whankaew, S., Smith, D. R., et al. (2017). Quantitative trait loci underlying root yield and starch content in an F1 derived cassava population (*Manihot esculenta* Crantz). *J. Agric. Sci.* 155, 569–581. doi: 10.1017/S0021859616000678

Stephenson, K., Amthor, R., Mallowa, S., Nungo, R., Maziya-Dixon, B., Gichuki, S., et al. (2010). Consuming cassava as a staple food places children 2-5 years old at risk for inadequate protein intake, an observational study in Kenya and Nigeria. *Nutr. J.* 9:9. doi: 10.1186/1475-2891-9-9

Tappiban, P., Ying, Y., Pang, Y., Sraphet, S., Srisawad, N., Smith, D. R., et al. (2020). Gelatinization, pasting and retrogradation properties and molecular fine structure of starches from seven cassava cultivars. *Int. J. Biol. Macromol.* 150, 831–838. doi: 10.1016/j.ijbiomac.2020.02.119

Teeken, B., Agbona, A., Abolore, B., Olaosebikan, O., Alamu, E., Adesokan, M., et al. (2020). *Understanding cassava varietal preferences through pairwise ranking of gari—eba and fufu prepared by local farmer—processors*. New York: Wiley.

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700

Teye, E., Asare, A. P., Amoah, R. S., and Tetteh, J. P. (2011). Determination of the dry matter content of cassava (*Manihot esculenta* Crantz) tubers using specific gravity method. *APRN J. Agric. Biol. Sci.* 6, 23–28.

Thanyasiriwat, T., Sraphet, S., Whankaew, S., Boonseng, O., Bao, J., Lightfoot, D. A., et al. (2013). Quantitative trait loci and candidate genes associated with starch pasting viscosity characteristics in cassava (*Manihot esculenta* Crantz). *Plant Biol.* 16, 197–207. doi: 10.1111/plb.12022

Thirathumthavorn, D., and Trisuth, T. (2008). Gelatinization and retrogradation properties of native and hydroxypropylated crosslinked tapioca starches with added sucrose and sodium chloride. *Int. J. Food Prop.* 11, 858–864. doi: 10.1080/10942910701659567

Tivana, L. D., Da Cruz, Francisco, J., Zelder, F., Bergenståhl, B., and Dejmek, P. (2014). Straightforward rapid spectrophotometric quantification of total cyanogenic glycosides in fresh and processed cassava products. *Food Chem.* 158, 20–27. doi: 10.1016/j.foodchem.2014.02.066

Tonukari, N. J., Tonukari, N. J., Ezedom, T., Enuma, C. C., Sakpa, S. O., Avwioroko, O. J., et al. (2015). White gold: Cassava as an industrial base. *Am. J. Plant Sci.* 6, 972–979. doi: 10.4236/ajps.2015.67103

Torres, L. G., De Resende, M. D. V., Azevedo, C. F., Fonseca, E., Silva, F., and De Oliveira, E. J. (2019). Genomic selection for productive traits in biparental cassava breeding populations. *PLoS One* 14:e0220245. doi: 10.1371/journal.pone.0220245

Tripathi, J. N., Ntui, V. O., Ron, M., Muiruri, S. K., Britt, A., and Tripathi, L. (2019). CRISPR/Cas9 editing of endogenous banana streak virus in the B genome of *Musa* spp. *overcomes a major challenge in banana breeding. Commun. Biol.* 2:46. doi: 10.1038/s42003-019-0288-7

Tripathi, L., Ntui, V. O., and Tripathi, J. N. (2020). CRISPR/Cas9-based genome editing of banana for disease resistance. *Curr. Opin. Plant Biol.* 56, 118–126.

Turyagyenda, L. F., Kizito, E. B., Ferguson, M. E., Baguma, Y., Harvey, J. W., Gibson, P., et al. (2012). Genetic diversity among farmer-preferred cassava landraces in Uganda. *Afr. Crop Sci. J.* 20, 15–30.

Udoh, L. I., Gedil, M., Parkes, E. Y., Kulakow, P., Adesoye, A., Nwuba, C., et al. (2017). Candidate gene sequencing and validation of SNP markers linked to carotenoid content in cassava (*Manihot esculenta* Crantz). *Mol. Breed* 37:123. doi: 10.1007/s11032-017-0718-5

Waisundara, V. Y. (2018). Introductory Chapter: Cassava as a Staple Food. doi: 10.5772/intechopen.70324

Wang, W., Feng, B., Xiao, J., Xia, Z., Zhou, X., Li, P., et al. (2014). Cassava genome from a wild ancestor to cultivated varieties. *Nat. Commun.* 5:5110. doi: 10.1038/ncomms6110

Whankaew, S., Poopear, S., Kanjanawattanawong, S., Tangphatsornruang, S., Boonseng, O., Lightfoot, D. A., et al. (2011). A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an outbred population. *BMC Genomics* 12:266. doi: 10.1186/1471-2164-12-266

Wolfe, M. D., Del Carpio, D. P., Alabi, O., Ezenwaka, L. C., Ikeogu, U. N., Kayondo, I. S., et al. (2017). Prospects for genomic selection in cassava breeding. *Plant Genome* 10:3. doi: 10.3835/plantgenome2017.03.0015

Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016). Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9:2. doi: 10.3835/plantgenome2015.11.0118

Wossen, T., Girma, G., Abdoulaye, T., Rabbi, I., Olanrewaju, A., Alene, A., et al. (2017). *The Cassava Monitoring Survey in Nigeria.IITA*. Ibadan: Nigeria, 66.

Xia, Z., Zou, M., Zhang, S., Feng, B., and Wang, W. (2014). AFSM sequencing approach: A simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping. *Sci. Rep.* 4:7300. doi: 10.1038/srep07300

Xu, J., Hua, K., and Lang, Z. (2019). Genome editing for horticultural crop improvement. *Hortic. Res.* 6:113. doi: 10.1038/s41438-019-0196-5

Yonemaru, J. i., Yamamoto, T., Fukuoka, S., Uga, Y., Hori, K., and Yano, M. (2010). Q-TARO: QTL annotation rice online database. Rice 3, 194Ű203. doi: 10.1007/s12284-010-9041-z

Yonis, B. O., Pino, del Carpio, D., Wolfe, M., Jannink, J. L., Kulakow, P., et al. (2020). Improving root characterisation for genomic prediction in cassava. *Sci. Rep.* 10:80003. doi: 10.1038/s41598-020-64963-9

Zhang, S., Chen, X., Lu, C., Ye, J., Zou, M., Lu, K., et al. (2018). Genome-wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* crantz). *Front. Plant Sci.* 9:503. doi: 10.3389/fpls.2018.00503

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mbanjo, Rabbi, Ferguson, Kayondo, Eng, Tripathi, Kulakow and Egesi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 January 2021
doi: 10.3389/fgene.2020.621871

QTL Mapping and Identification of Candidate Genes for Heat Tolerance at the Flowering Stage in Rice

Lei Chen[1,2†], Qiang Wang[2†], Maoyan Tang[2], Xiaoli Zhang[2], Yinghua Pan[2], Xinghai Yang[2], Guoqing Gao[2], Ronghua Lv[2], Wei Tao[2], Ligeng Jiang[1] and Tianfeng Liang[2*]*

[1] Key Laboratory of Crop Cultivation and Farming System, College of Agriculture, Guangxi University, Nanning, China, [2] Guangxi Academy of Agricultural Sciences/Guangxi Key Laboratory of Rice Genetics and Breeding, Rice Research Institute, Nanning, China

OPEN ACCESS

Edited by:
Arron Carter,
Washington State University,
United States

Reviewed by:
Arif Ashraf,
University of Massachusetts Amherst,
United States
Karansher Singh Sandhu,
Washington State University,
United States
Lance Merrick,
Washington State University,
United States

***Correspondence:**
Tianfeng Liang
tfliang163@163.com
Ligeng Jiang
jiang@gxu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 27 October 2020
Accepted: 29 December 2020
Published: 22 January 2021

Citation:
Chen L, Wang Q, Tang M, Zhang X, Pan Y, Yang X, Gao G, Lv R, Tao W, Jiang L and Liang T (2021) QTL Mapping and Identification of Candidate Genes for Heat Tolerance at the Flowering Stage in Rice. Front. Genet. 11:621871. doi: 10.3389/fgene.2020.621871

High-temperature stress can cause serious abiotic damage that limits the yield and quality of rice. Heat tolerance (HT) during the flowering stage of rice is a key trait that can guarantee a high and stable yield under heat stress. HT is a complex trait that is regulated by multiple quantitative trait loci (QTLs); however, few underlying genes have been fine mapped and cloned. In this study, the $F_{2:3}$ population derived from a cross between Huanghuazhan (HHZ), a heat-tolerant cultivar, and 9311, a heat-sensitive variety, was used to map HT QTLs during the flowering stage in rice. A new major QTL, *qHTT8*, controlling HT was identified on chromosome 8 using the bulked-segregant analysis (BSA)-seq method. The QTL *qHTT8* was mapped into the 3,555,000–4,520,000 bp, which had a size of 0.965 Mb. The candidate region of *qHTT8* on chromosome 8 contained 65 predicted genes, and 10 putative predicted genes were found to be associated with abiotic stress tolerance. Furthermore, qRT-PCR was performed to analyze the differential expression of these 10 genes between HHZ and 9311 under high temperature conditions. *LOC_Os08g07010* and *LOC_Os08g07440* were highly induced in HHZ compared with 9311 under heat stress. Orthologous genes of *LOC_Os08g07010* and *LOC_Os08g07440* in plants played a role in abiotic stress, suggesting that they may be the candidate genes of *qHTT8*. Generally, the results of this study will prove useful for future efforts to clone *qHTT8* and breed heat-tolerant varieties of rice using marker-assisted selection.

Keywords: rice, flowering stage, heat tolerance, QTL, BSA-seq, candidate gene

INTRODUCTION

Rice (*Oryza sativa* L.) is a major staple food crop for nearly half of the world's population (Pan et al., 2020). As global temperatures have increased in recent years, extreme, high temperatures have led to serious losses in yield, decreases in grain quality and reductions in harvest index, especially during the flowering stage, which has a net negative impact on the normal seed setting of rice (Jagadish et al., 2012). Average global temperatures are expected to increase by 2–3°C over the next 30–50 years (Hatfield and Prueger, 2015). However, rice yields are expected to decrease by 10% for every increase in daily maximum and minimum temperature of 1°C (Welch et al., 2010). In addition, the average daily temperature is expected to exceed 35°C for several consecutive days, which will lead to spikelet sterility and abnormal pollination, seriously reducing

the seed-setting rate (IPCC, 2007). Current strategies to deal with high-temperature stress via alterations to technical and management systems are insufficient for sustaining yields (Driedonks et al., 2016). There is thus an urgent need to breed heat-tolerant rice varieties.

Many researchers claimed that the most sensitive growth stage of rice to heat stress was the flowering time (Baliuag et al., 2015; Nubankoh et al., 2020). And the study of HT at the flowering stage has become a major focus of rice breeding. Understanding the genetic mechanisms of HT and developing heat-tolerant varieties are essential for the ability of rice to cope with future global warming (Ye et al., 2015a,b). Germplasm resources are the material basis for the breeding of new rice varieties. The most effective method is to select different types of heat-tolerant materials to identify different rice germplasm resources, characterize HT and build a robust population, which can provide a foundation for the breeding of stress-tolerant varieties, a reference for the identification of heat-tolerant genes and a means for the exploration of heat-tolerant mechanisms. Effective measures for dealing with high-temperature stress in rice include the identification of heat-tolerant genes, the acquisition of intermediate materials and the cultivation of heat-tolerant varieties (Kilasi et al., 2018).

Much research over the past decades has focused on the mining of heat-tolerant genes in rice, primarily through the construction of different genetic populations (Cao et al., 2020). Yield or quality traits related to HT have been used as the evaluation indexes for rice heat-tolerant QTL analysis (Cao et al., 2002; Zhu et al., 2006; Asako et al., 2007; Xiao et al., 2011; Ye et al., 2012, 2015a,b; Nubankoh et al., 2020). These previous works have resulted in the detection of heat-tolerant QTLs in different regions of multiple chromosomes. Recent work has focused on the physical and chemical properties and agronomic characters of rice during each sensitive period, and has resulted in several breakthroughs in the study of heat-tolerance mechanisms. Specifically, numerous achievements have been made in research on rice HT molecular genetics, including the mapping of several rice HT QTLs. However, few QTLs related to rice HT have been cloned (Cao et al., 2020).

Genetic analysis has revealed that HT at the flowering stage in rice is a complex quantitative trait controlled by multiple genes. The resistance of rice to high temperatures shows variety specificity, which indicates that genetic factors contribute the most to explaining variation in HT among rice varieties. With the development and wide application of molecular biology and genomic tools in recent years, there has been an increasing number of QTL-mapping studies of rice HT using molecular markers. QTLs/genes for rice HT have been mapped across all 12 chromosomes using different types of molecular markers, such as RFLPs, SSRs, and SNPs, which has facilitated the identification of chromosomal regions associated with tolerance of high temperatures (**Supplementary Table 1**). In addition, different parents and types of mapping populations (e.g., F_2, $F_{2:3}$ lines, BC, NILs, RILs, CSSLs, and DH) have been used to analyze QTLs/genes with different yields (e.g., seed setting rate, spikelet fertility, pollen fertility, grain weight, flowering time and heading days) and quality traits (e.g., white-back kernels, basal-white grain and gel consistency) related to rice heat-stress tolerance at different stages, such as the seeding and reproductive stages (Cao et al., 2002, 2020; Zhu et al., 2006, 2017; Asako et al., 2007; Tabata et al., 2007; Chen et al., 2008; Jagadish et al., 2008, 2010; Xiao et al., 2011; Cheng et al., 2012; Ye et al., 2012, 2015a,b; Murata et al., 2014; Tazib et al., 2015; Wada et al., 2015; Zhao et al., 2016; Shanmugavadivel et al., 2017; Nubankoh et al., 2020).

The presence of similar heat-tolerant QTLs in rice indicates that the heat-tolerant metabolic pathways might be conserved among different rice varieties and that some QTLs with greater effects could be stably expressed. However, some heat-tolerant QTLs have not been consistently detected, which may be related to the different genetic backgrounds of varieties, or differences in environmental conditions among tests. Rice HT is characterized by quantitative trait inheritance, and its molecular mechanism is relatively complex. An important line of research on the molecular mechanism of rice HT is the determination of genes involved in the regulation of the response of rice to heat stress. Although QTLs for rice HT at the flowering stage have been mapped on all 12 chromosomes using various rice populations, the additive effect of each QTL is relatively low. As a result, introducing one or a few QTLs into a variety may not sufficiently increase its HT (Ye et al., 2015a,b). Therefore, the fine mapping, validation and characterization of more major QTLs and the design of functional SNP chips with QTL-linked markers are necessary for accelerating the selection and incorporation of multiple QTLs and, in turn, improving the efficiency of rice heat-tolerant breeding.

Here, the heat-tolerant variety HHZ and the heat-sensitive variety 9311 studied by our research group in a previous study were hybridized (F_1) and then continuously self-crossed to develop source materials (F_2 and $F_{2:3}$) for HT identification and QTL mapping (Cao et al., 2008; Wang et al., 2020). A total of 365 $F_{2:3}$ populations were selected for HT evaluation at the flowering stage. The QTLs for spikelet fertility under high-temperature stress were rapidly identified using the BSA-seq method combined with whole-genome resequencing (WGS) technology (Takagi et al., 2013; Zou et al., 2016). Finally, one major QTL, *qHTT8*, controlling HT at the flowering stage was identified on chromosome 8. Furthermore, we performed qRT-PCR to study the expression of ten putative genes under heat stress. A phylogenic analysis suggested that *LOC_Os08g07010* and *LOC_Os08g07440* were the two candidate genes controlling HT at the flowering stage in rice. Generally, the results of this study will aid future efforts to improve the HT in rice.

MATERIALS AND METHODS

Rice Materials

HHZ and 9311 are both conventional *indica* rice varieties that were kindly provided by the Guangdong Academy of Agricultural Sciences and Huazhong Agricultural University, respectively. In past decades, the $F_{2:3}$ population was widely used for rapid QTL mapping around different crops because of its short construction time and obvious segregation of the allelic characteristics of the parent strains parental lines (Austin and Lee, 1995; Fahliani et al., 2010; Park et al., 2013). A set of 365 $F_{2:3}$ lines derived from a

cross between HHZ, a heat-tolerant cultivar (Cao et al., 2008; Wang et al., 2020), and 9311, a heat-susceptible cultivar from our previous work, was used to evaluate rice HT at the flowering stage in this study. In 2018, $F_{2:3}$ lines and their parents were planted in the net-house of Guangxi Academy of Agricultural Sciences, Nanning, Guangxi, China.

Evaluation of HT in $F_{2:3}$ At the Flowering Stage

An $F_{2:3}$ population of 365 individuals was planted in plastic pots under natural conditions until heading. At the start of heading, 3–5 uniform panicles with the opened florets carefully removed, were marked with PVC tags (Jagadish et al., 2007). The plants were then moved into a phytotron. During this period, the spikelets in the panicle were exposed to 38/24°C day/night temperatures with 6 h (from 09:30 am to 3:30 pm) of 38°C during the day (Liang et al., 2016). After 3 days of exposure to high temperature, the plants were moved back to the net house and were grown to maturity. After harvest, the labeled panicles were tested for seed set, and the numbers of fully filled grains (N_{FG}), partially filled grains (N_{PG}) and empty grains (N_{EG}) were counted (Liang et al., 2016). Each spikelet was pressed between the thumb and forefinger to determine whether the grain was filled or not. Both partially and fully filled spikelets were categorized as filled spikelets (Mohammed and Tarpley, 2009; Rang et al., 2011). Next, the absolute spikelet fertility percentage was calculated using the formula below, and HT was assessed based on absolute fertility at high temperature (Jagadish et al., 2010; Cao et al., 2020).

Spikelet fertility (%) = $(N_{FG} + N_{PG})/ (N_{FG} + N_{PG} + N_{EG}) \times 100$.

Whole Genome Re-Sequencing and BSA-seq Analysis

According to the phenotypic characterization of HT of derived $F_{2:3}$ lines at anthesis under heat stress, pools of tolerant and sensitive bulk samples ($n = 50$, each group) were constructed from 365 F_2 individuals, and the same amount of DNA of fresh leaves was extracted from each plant and was evenly mixed.

After the sample DNA was quantified by a Nanodrop 2000, the DNA in each mixing pool was equally mixed. The parents and the mixing pool DNA sequences were segmented into random fragments using ultrasound. The segmented DNA was successively repaired at the end; A was added at the 3′ end, and the sequencing connector was connected. Next, the magnetic beads were used to absorb and enrich the fragments with lengths of approximately 400 bp, which were amplified by PCR to form a sequencing library. After inspection of the constructed library, the qualified library was sequenced using the Illumina NovaSeq 6000 platform. The sequencing approach used was Illumina PE150, and the total sequencing read length was 300 bp. After the low-quality reads (raw data) were filtered out, the remaining reads (clean data) were aligned to the *Oryza sativa L. ssp. japonica cv.* Nipponbare reference genome using BWA software. The location of the sequence (i.e., the BAM file) was then obtained. The best practices pipeline in GATK software was used to correct BAM files and detect SNPs and small InDels.

SnpEff software and gene prediction information of the reference genome were used to annotate the variation function, and the function annotation information for SNPs and InDels was obtained. Based on the characteristics of the data for parents and mutation pools, the SNP-index (the SNP frequency) value was calculated for the BSA association analysis to locate the target loci. For the SNP and InDel loci among the samples obtained by filtering and screening, the SNP-index values of each locus in the heat-tolerant mixed pool (T-pool)/ the sensitive mixed pool (S-pool) were calculated. The average SNP-index values of all SNPs in the window were then counted as the SNP-index of that window. The window was sliding, with a 500 kb window size and a 5-kb increment. The SNP-index of the T-pool and S-pool was defined as the ratio between the HHZ SNP and the total number of reads corresponding to the SNP. The Δ (SNP-index) was calculated according to the formula Δ(SNP-index) = [(SNP-index of T-pool) – (SNP-index of S-pool)]. Because of the linkage between the heat-tolerant loci and surrounding markers, the SNP-index in the T-pool was closer to 1, whereas the SNP-index in the S-pool was closer to 0. Because of weak linkage or a lack of linkage, the loci were randomly distributed, and the SNP-index of the other normal loci was 0.5. A thousand replacement tests were performed, and the region with the most differences in SNP-index values between the two pools (Δ) (using the 99.9% confidence level as the threshold for screening) was the candidate region of the target trait correlation. All candidate genes were analyzed by GO enrichment analysis (Gene Ontology, http://www.geneontology.org/) based on a Fisher's exact test and a Yekutieli multitest adjustment using a 5% false-positive detection threshold (Yang et al., 2017).

qRT-PCR Analysis of Candidate Genes in the Mapped Region

HHZ and 9311 were cultivated until the flowering stage, and 3 plants of each were subjected to 38°C heat stress from 9:30 am to 3:30 pm during the day for 3 days in a phytotron. Panicle tissues were sampled at different times from the initiation of heat stress, namely at 8:30 am, 10:30 am, 12:30 pm, 3:30 pm and 4:30 pm every day. Each sampling period was arranged in order from No. 1 to 15, among which, No. 5, 10, and 14 were not measured out of the convenience of analysis. Total RNA was extracted from panicles using the TRIzol® reagent (Invitrogen, Carlsbad, USA), and 2 μg of DNaseI-treated RNA was used as the template for cDNA synthesis using the PrimeScript™ RT reagent Kit with gDNA Eraser. A Bio-Rad CFX96 Real-Time system (Bio-Rad Laboratories, Inc., USA) was used to perform quantitative real-time PCR in 10-μL mixtures: 5 μL of 2 × Green qPCR MasterMix, 1 μL of cDNA, 0. 5 μL of each primer (10 μM), and 3.5 μL of ddH_2O. Amplification steps were 95°C for 30 s, 40 cycles of (95°C for 5 s, 60°C for 30 s), and 65°C for 5 s, 95°C for 15 s, 60°C for 30 s, 95°C for 15 s. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Ubiquitin (*UBQ*) was used as the internal control and

at least three replicates were performed for each experiment. Primers used for qRT-PCR are listed in **Supplementary Table 4**.

Phylogenetic Analysis of *qHTT8* Genes

The candidate genes for homologous genes in the other plants were obtained from the Rice Genome Annotation Project database (RGAP, http://rice.plantbiology.msu.edu/) database and the Phytozome (https://phytozome.jgi.doe.gov) database. Protein sequences of the candidate genes and their homologs from other species were retrieved from the Phytozome database. The phylogenetic tree was constructed using the Maximum Likelihood (ML) method with 1,000-replicate bootstrapping in MEGAX software (Kumar et al., 2018). An alignment of rice and its homologs from other species was performed using ClustalX v2.1. And the conserved domain of candidate gene was screened using the CD-search in the Nation Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov)database.

RESULTS

Phenotypic Characterization of the HT of $F_{2:3}$

Spikelet fertility has been previously used to screen and select for HT during the reproductive stage. To analyze the genetic basis of HT during the flowering stage in rice, we constructed an $F_{2:3}$ population derived from crosses between HHZ and 9311. And the frequency distribution of spikelet fertility in the $F_{2:3}$ population is shown in **Figure 1A**. Spikelet fertility ranged from 1.04 to 85.24%, with an average value of 50.00%, standard deviation of 0.1847 and coefficient of variation of 36.95%, indicating that the HT phenotype in the $F_{2:3}$ population was normally distributed (**Figure 1A**). Meanwhile, the spikelet fertilities of their parents HHZ and 9311 were 54.5 and 14.3%, respectively, under high-temperature stress (38°C) for 3 consecutive days (**Figure 1B**). The $F_{2:3}$ population showed a large degree of segregation and some super-parent lines because of their different genetic background, suggesting that the HT of rice at anthesis was a quantitative trait controlled by multiple QTLs/genes. To construct the heat-sensitive (S) and heat-tolerant (T) pools, 50 heat-sensitive and 50 heat-tolerant $F_{2:3}$ plants were selected. The percentage of spikelet fertility of the 50 $F_{2:3}$ plants in the HS-bulk ranged from 1.04 to 26.73% and that of the 50 $F_{2:3}$ plants of the HT-bulk ranged from 69.90 to 85.24%.

BSA-seq Analysis

Genomic DNA samples of the two parents (HHZ and 9311) and the two pools (T-pool and S-pool) were sequenced by an Illumina HiSeq^{TM} sequencer, and 85.7 Gb of clean data were generated after being filtered; the Q30 of all samples was >90% (**Table 1**), indicating the high quality of the sequencing data.

The parents and mixed-pool sequencing data were compared using Nipponbare as the reference genome, and mutations were detected (**Table 2**). The effective reads of HHZ accounted for 98.01% of the entire genome, with an average sequencing depth of 35.93×, whereas the effective reads of 9311 covered 98.14% of the entire genome with an average read depth of 35.30×.

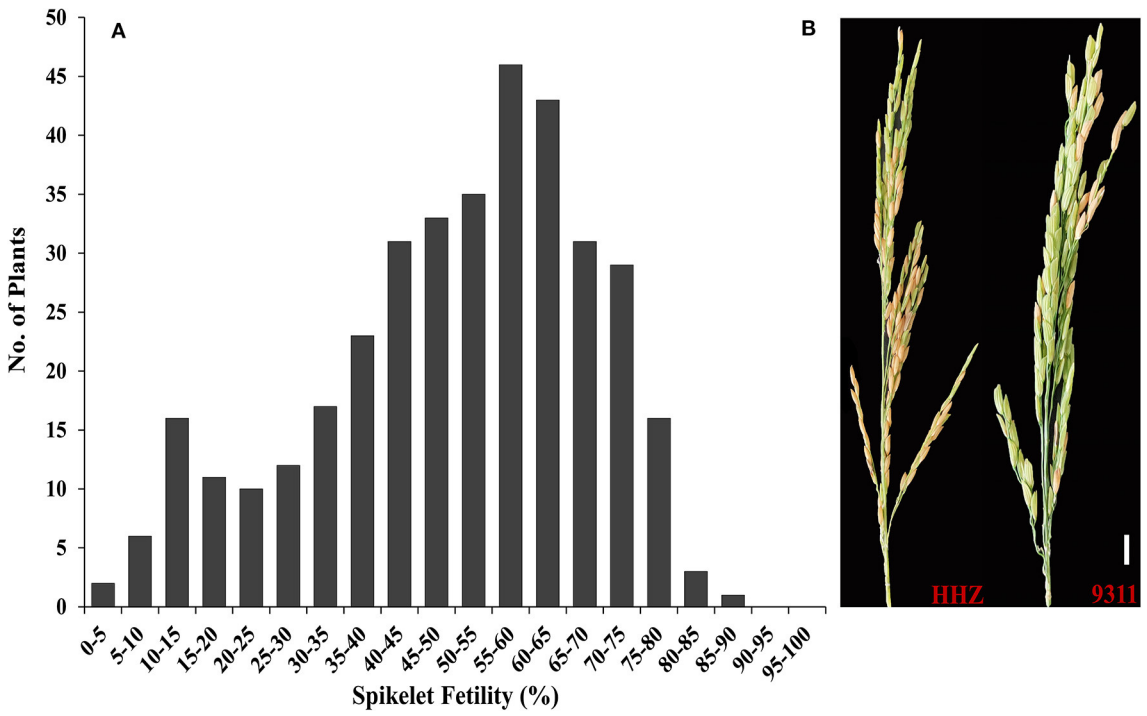


FIGURE 1 | HT evaluation of the two parents and $F_{2:3}$ population. **(A)** Frequency distribution of spikelet fertility in the $F_{2:3}$ population containing 365 plants; **(B)** HHZ and 9311 under 38°C for 3 days.

Both varieties showed good coverage and sequencing depth. BSA association analysis is a gene-mapping method based on mixed pool sequencing, which primarily analyzes regions with significant differences in the frequency of mixed pool genotypes to determine the QTL positions related to target traits. In this study, BSA was used to analyze the associated SNPs. Before association analysis, the SNPs were filtered to obtain 627,717 high-quality SNP loci. After filtering, 33,205 effective SNP loci with differences between the two pools were identified.

QTL Mapping

Based on resequencing and association analysis, the SNP-index of the T-pool and S-pool was compared (**Figures 2A,B**). At a 99.9% confidence level, the window above the threshold was considered the candidate interval. There was an imbalanced SNP between 3,555,000–4,520,000 bp on chromosome 8 (**Figure 2C** in the red-dotted box) (**Figure 2D**). In this region, the SNP-index value of the T-pool (heat-tolerant type) was greater than or equal to 0.7, while that of the S-pool (heat-sensitive type) was ≤0.3, indicating that the single plant in the heat-tolerant pool had the same fragment as HHZ in this region and that the single plant in the heat-sensitive pool had the same fragment as 9311 in this region. With a 99.9% confidence level as the screening threshold, the value of Δ(SNP-index) in this region was greater than the screening threshold. Therefore, the region of 3,555,000–4,520,000 bp on chromosome 8, which was named *qHTT8*, may be the putative locus controlling the HT of rice at the flowering stage. Analysis of the gene sequence of the *qHTT8* interval revealed a total of 6,821 SNPs and 1,155 InDels; 258 SNPs and 29 InDels caused amino acid changes.

Gene Ontology (GO) Enrichment Analysis

Gene ontology (GO) analysis was used to classify all of the genes expressed into different functional categories, including biological processes, cellular components and molecular functions. Genes located in the genomic regions for the identified QTL were extracted (**Supplementary Table 2**). The *qHTT8* QTL harbored 53 genes that were annotated in the GO database. A total of 119 GO terms were grouped into the three categories, within which genes corresponding to biological process (46) and molecular function (39) were the most abundant. The proteolysis involved in the cellular protein catabolic process (GO:0051603) was the most significant in the biological process category, indicating that the rice leaves under heat treatment had wide metabolic activities. In the molecular function category and cellular component category, the serine-type carboxypeptidase activity (GO:0004185) and integral component of membrane (GO: 0016021) were the most significantly represented groups, respectively (**Figure 3**).

TABLE 1 | The quality of sequencing data.

Sample	Clean Reads (bp)	Clean Base (bp)	GC content (%)	Q30 (%)
HHZ	55,685,598	16,765,143,297	43.77	94.13
9311	53,801,932	16,198,479,368	44.14	92.54
T-pool	86,378,436	26,006,038,749	44.14	93.85
S-pool	88,929,269	26,772,575,482	44.67	94.09

Candidate Genes Analysis

To further analyze the candidate genes in the chromosome region containing *qHTT8*, we predicted 65 putative genes in the Nipponbare genome using the RGAP database (**Supplementary Table 3**). Ten genes that were located within the mapped region have been reported to be involved in abiotic stress tolerance, such as high night temperature, drought, cold, salinity and saline-alkaline (**Table 3** and **Figure 2E**) (Ma et al., 2009; Smita et al., 2010; Nguyen et al., 2014; Lee et al., 2015; Raineri et al., 2015; Saha et al., 2015; Huang et al., 2017; Patil et al., 2017; Li N. et al., 2018; Hoang et al., 2019; Yang et al., 2019).

The relative expression levels of these 10 genes in the panicle between HHZ and 9311 were analyzed under different durations of heat stresses using qRT-PCR. Based on the cDNA sequences, 10 gene primers pairs for qRT-PCR analysis were designed. The qRT-PCR results showed that among these 10 genes, the expression levels of the genes *LOC_Os08g07010* and *LOC_Os08g07440* were both significantly higher in HHZ than in 9311 at all sampling points (**Figure 4**). In contrast, there was no clear differential expression between HHZ and 9311 in other genes (**Supplementary Figure 1**). Thus, *LOC_Os08g07010* and *LOC_Os08g07440* were determined to be the candidate genes responsible for the HT of rice.

Phylogenic Analysis of *qHTT8*

To better understand the similarities and differences in the two candidate genes in *qHTT8* between rice and other species, the phylogenetic tree was generated using different protein sequences from rice and other plants (**Figure 5**). *LOC_Os08g07010* (*OsABCG18*) encodes the ABC-2 type

TABLE 2 | Statistical analysis of sequencing depth and coverage.

Sample	Mapped Ratio (%)	Properly Mapped (%)	Duplication Ratio (%)	Average Insert Size	Average Depth (×)	Real Depth (×)	Genome Coverage (1×) (%)	Genome Coverage (5×) (%)
HHZ	98.01	90.51	18.44	378.7	35.93	38.52	93.26	89.5
9311	98.14	90.47	17.12	383.5	35.3	37.84	93.3	89.51
T-pool	97.8	90.23	17.87	384.5	56.01	58.5	95.75	92.63
S-pool	97.89	90.27	18.21	377	57.47	59.94	95.88	92.58

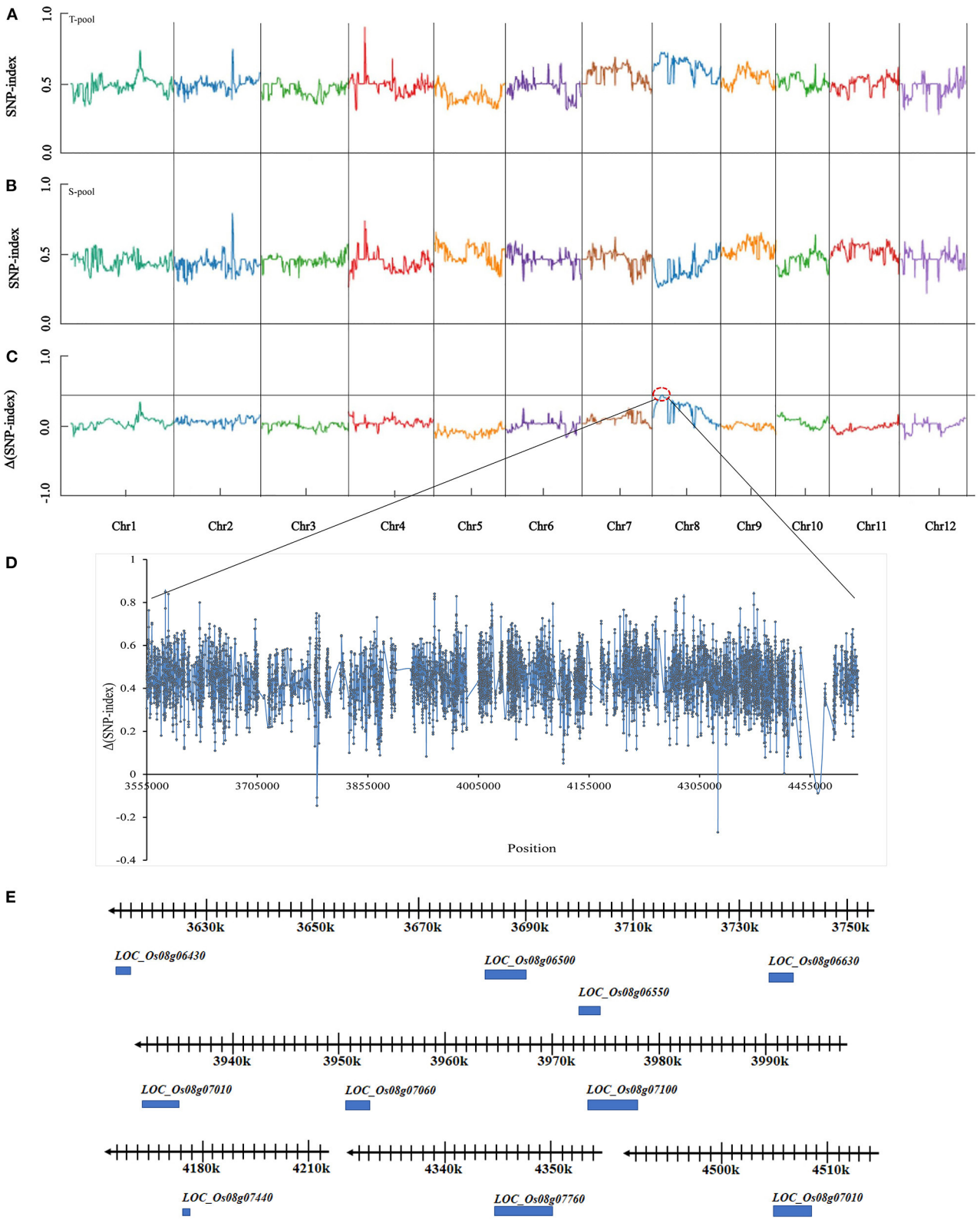


FIGURE 2 | SNP-index graphs of T-pool **(A)**, S-pool **(B)**, and Δ(SNP-index) graph **(C)** from BSA-seq analysis. X-axis represents the position of the 12 chromosomes in rice; Y-axis represents the SNP-index. The black line shows the association threshold at a 99.9% confidence level **(C)**. HT major QTL is located to chromosome 8 (red dotted box). **(D)** Value of Δ(SNP-index) in the candidate region. **(E)** Physical location on chromosome 8 of the ten putative genes.

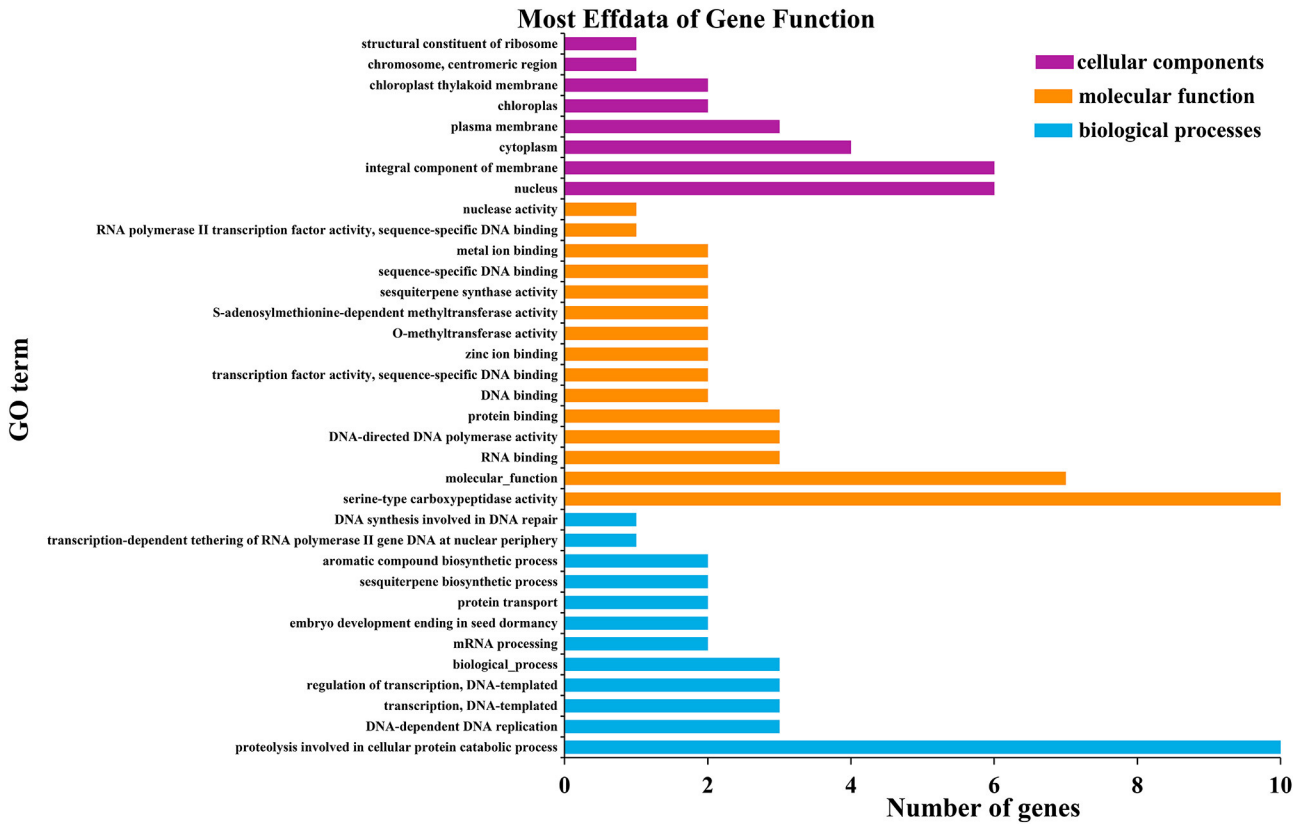


FIGURE 3 | Significantly enriched GO terms of the genes around *qHTT8*.

TABLE 3 | Putative genes associated with abiotic stress tolerance in the *qHTT8* region.

No.	Gene	Type and putative protein function	Physical location (bp)
1	*LOC_Os08g06430*	Mitochondrial NADH-ubiquinone oxidoreductase, putative, expressed	3,613,350–3,615,878
2	*LOC_Os08g06500*	PPR repeat domain containing protein, putative, expressed	3,680,925–3,688,728
3	*LOC_Os08g06550*	Acyl CoA binding protein, putative, expressed	3,698,312–3,700,553
4	*LOC_Os08g06630*	RNA polymerase sigma factor, putative, expressed	3,732,441–3,736,521
5	*LOC_Os08g07010*	ABC-2 type transporter domain containing protein, expressed	3,928,462–3,933,577
6	*LOC_Os08g07060*	CRR6, putative, expressed	3,949,017–3,951,136
7	*LOC_Os08g07100*	Terpene synthase, putative, expressed	3,972,216–3,977,334
8	*LOC_Os08g07440*	AP2 domain containing protein, expressed	4,178,549–4,175,872
9	*LOC_Os08g07760*	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed	4,344,171–4,350,502
10	*LOC_Os08g07970*	Transcription factor, putative, expressed	4,508,263–4,505,738

transporter protein. *AT1G31770* (*AtABCG14*) in *Arabidopsis thaliana*, one of the *LOC_Os08g07010* orthologous genes, was the first discovered protein related to the long-distance translocation of cytokinin which is one of the most critical signaling molecules in stress responses (Zhang et al., 2014). *OsABCG18* in rice and *AtABCG14* have similar biochemical functions in cytokinin long-distance transport from the root to the shoot (Zhao et al., 2019). Cytokinin plays an important role in plant growth, development and abiotic stress. *LOC_Os08g07440* encodes AP2 domain containing protein, which is a member

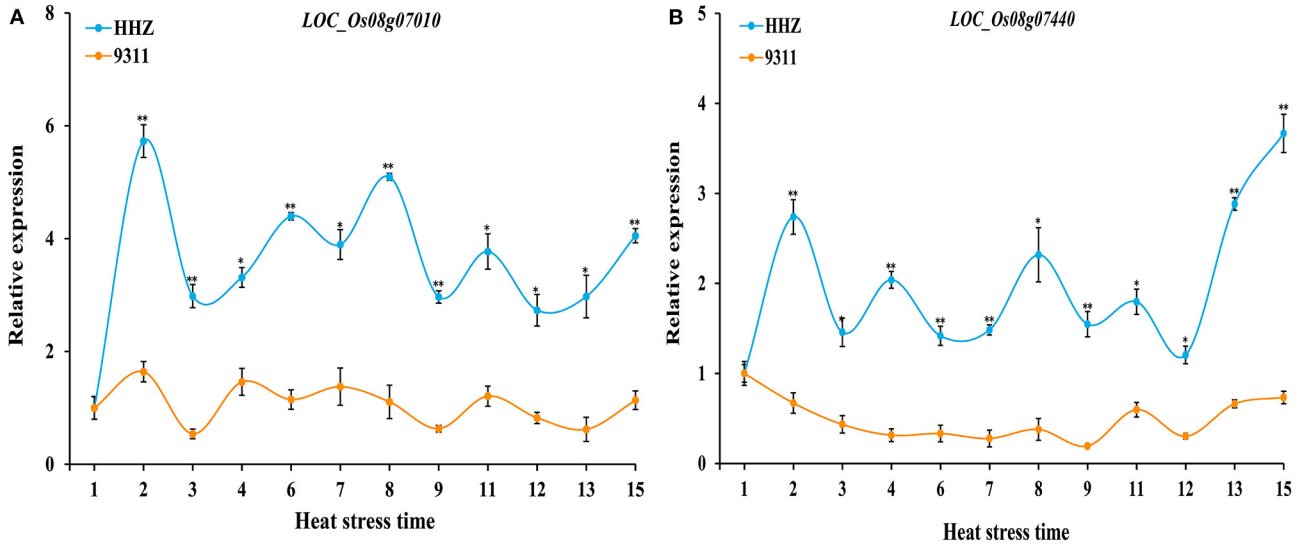


FIGURE 4 | Expression analysis of candidate genes (**A**, *LOC_Os08g07010* and **B**, *LOC_Os08g07440*) in HHZ and 9311 under heat stress for different time periods. The relative expression values were normalized to the rice *UBQ* gene. Error bars indicate standard deviation, and asterisks indicate significant differences using the Student's *t*-test (*$p < 0.05$; **$p < 0.01$). X-axis represents the different sampling time period for plants subjected to a 38°C heat stress in a phytotron; Y-axis represents the relative expression level of genes.

of the AP2-EREBP family (APETALA2/Ethylene Responsive Element Binding Protein). AP2/EREBP genes also played a major and diversified role in plants to respond to various types of biotic and environmental stress (Riechmann and Meyerowitz, 1998; Dietz et al., 2010). *GRMZM2G022359* in maize, one of the *LOC_Os08g07440* orthologs, is involved in diverse abiotic stress responses and the regulation of processes (Huang et al., 2015). The phylogenetic analysis indicated that both *LOC_Os08g07010* and *LOC_Os08g07440* had high homology with most other homologous members, revealing that these two candidate genes were highly conserved. Furthermore, one and two conserved domains were found on *LOC_Os08g07010* and *LOC_Os08g07440,* respectively (**Figures 6A,B**).

DISCUSSION

With the threat of climate change, especially increased temperatures, droughts and desertification are expected to render several regions inhospitable to agriculture; consequently, the development of heat-tolerant rice cultivars is critically important for future rice production (Costa and Farrant, 2019). Over the past decades, researchers have identified dozens of QTLs controlling heat stress tolerance (**Supplementary Table 1**). However, as HT is a quantitative trait, its underlying genetic mechanism is relatively complex. To date, the use of these candidate genes to breed high-yielding, heat-tolerant rice varieties has been rare. There is thus an urgent need to conduct more research to identify candidate genes associated with the inheritance of rice HT.

In this study, we evaluated the HT phenotype of $F_{2:3}$ families developed from HHZ crossed with 9311. The HT of $F_{2:3}$ populations at the flowering stage was a quantitative genetic trait controlled by multiple QTLs/genes. BSA-seq combined with the conventional gene mapping method can significantly accelerate the fine mapping of genes (Zhang et al., 2020). The extreme expression of HT was selected via phenotypic identification to locate heat-tolerant QTLs at the anthesis of rice using BSA combined with WGS. A heat-tolerant QTL was located between 3,555,000 and 4,520,000 bp on chromosome 8.

Compared with previous studies, some QTLs/genes for HT have been identified on chromosome 8 in recent years. For example, Tabata detected a QTL for the occurrence of white-back kernels associated with high temperatures during the ripening period of rice at 0.15 Mb (Tabata et al., 2007). The QTL *qhr8-1* of HT at the flowering stage around 17.43–21.65 Mb was mapped by Cao, and *qtl_8.2* for absolute spikelet fertility near 20.53 Mb was detected by Jagadish, which overlapped the QTL *qHTGC8* for the thermo-tolerance of gel consistency in the 19.31–20.66 Mb region located by Cao et al. (2002), Zhu et al. (2006), Jagadish et al. (2010). The QTL *qht8* located by Chen in the interval of 5.59–39.4 Mb contained the *qhr8-1* located by Cao et al., the *qtl_8.3* (27.60 Mb) located by Jagadish, the heat-tolerant QTL ranging from 2,355,534 to 37,615,523 bp mapped by Zheng and the QTL of spikelet and pollen fertility (24.72 Mb) and the early morning flowering QTL of heat escape (22.34 Mb) mapped by Cao et al. (2002), Chen et al. (2008), Jagadish et al. (2010), Baliuag et al. (2015), Zheng et al. (2017). These QTLs above are different from *qHTT8* in physical location, therefore, they belong to the different loci. And the QTL *HD8* of days-to-heading (heat escape) in the range of 3.02–4.38 Mb detected by Thanh were located close to the QTL *qHTT8* (Thanh et al., 2010). However, due to the use of different evaluation index, the two loci are also different, thus, *qHTT8* represents a new QTL related to heat tolerance at the

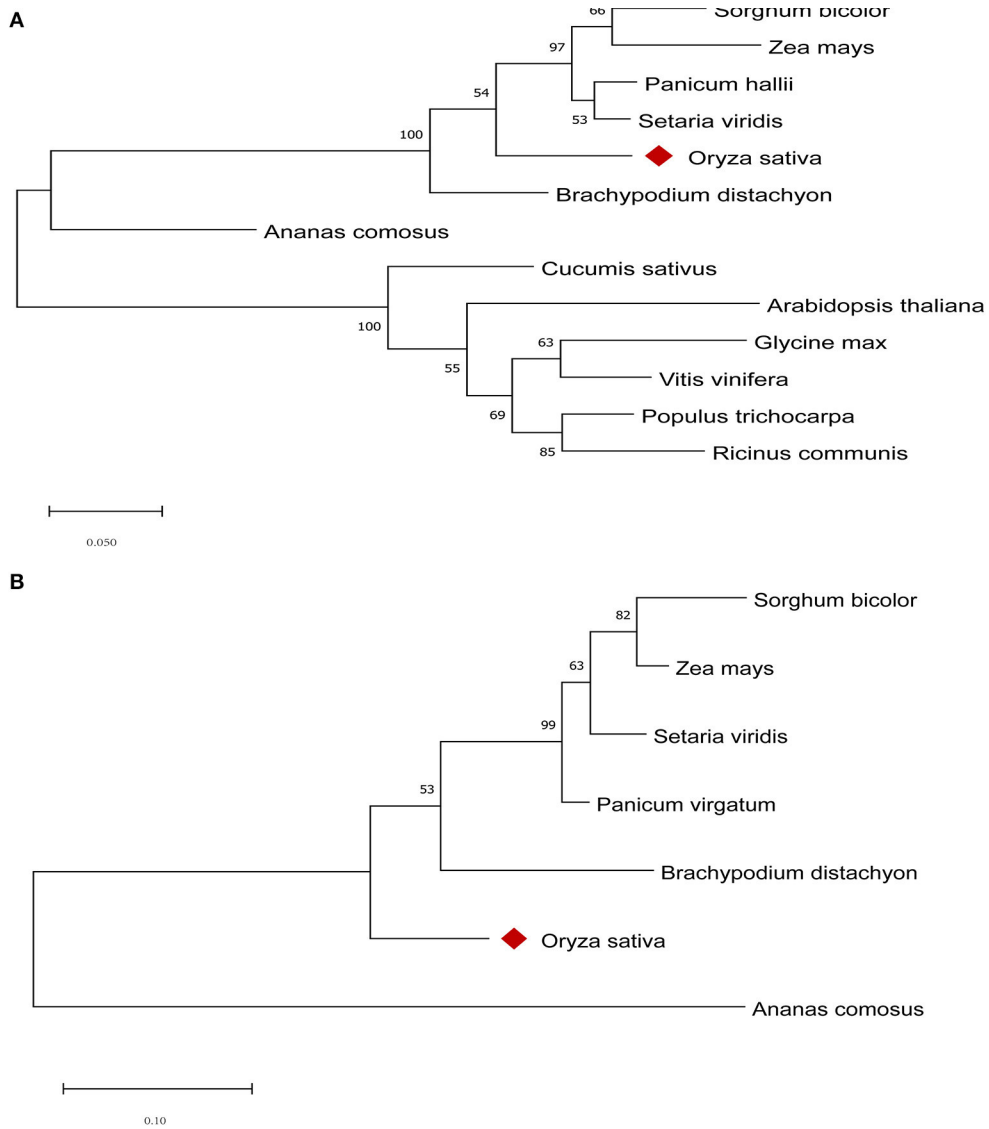


FIGURE 5 | Phylogenetic tree of *LOC_Os08g07010* **(A)** and *LOC_Os08g07440* **(B)** homologs from different species. Phylogenetic analysis was carried out using MEGAX based on the ML method with 1,000 bootstrap replications.

flowering stage in rice. Besides these, we did not find any QTL loci close to our interval on chromosome 8.

According to the RGAP database, 65 predicted genes were located in the target region containing *qHTT8*. Transcripts annotated as "hypothetical protein," "expressed protein," or "retrotransposon protein" were not included. The genes that were annotated to abiotic stress in rice (listed in **Supplementary Table 3**) were identified based on former studies. This analysis identified 10 annotated genes that were potential candidate genes for heat stress tolerance during the flowering stage in rice. Gene expression has been suggested to play the same role under different stress conditions. For example, *MYB*, a transcription factor (TF), was up-regulated when plants were exposed to a combination of drought and heat stress (Rizhsky et al., 2004). *OsbHLH148*, a basic helix-loop-helix TF, was responsive to heat, salt, dehydration and cold stress (Seo et al., 2011). *OsHCI1*, which is a rice gene encoding the RING finger protein, was specifically induced by heat and cold stress treatments but not by salinity or dehydration; its overexpression during heat and cold stress enhanced the acquired thermotolerance (Lim et al., 2013). Water deficits and high-temperature stress often occur simultaneously in the field (Bailey-Serres et al., 2019; El-Esawi and Alayafi, 2019; Shanmugavadivel et al., 2019). Furthermore, the expression of a trait might result from the contribution of many genes with similar or complementary functions (Ye et al., 2015a,b). These candidate genes, which play a role in other abiotic stress conditions, may thus have similar effects on the HT of rice at the flowering stage.

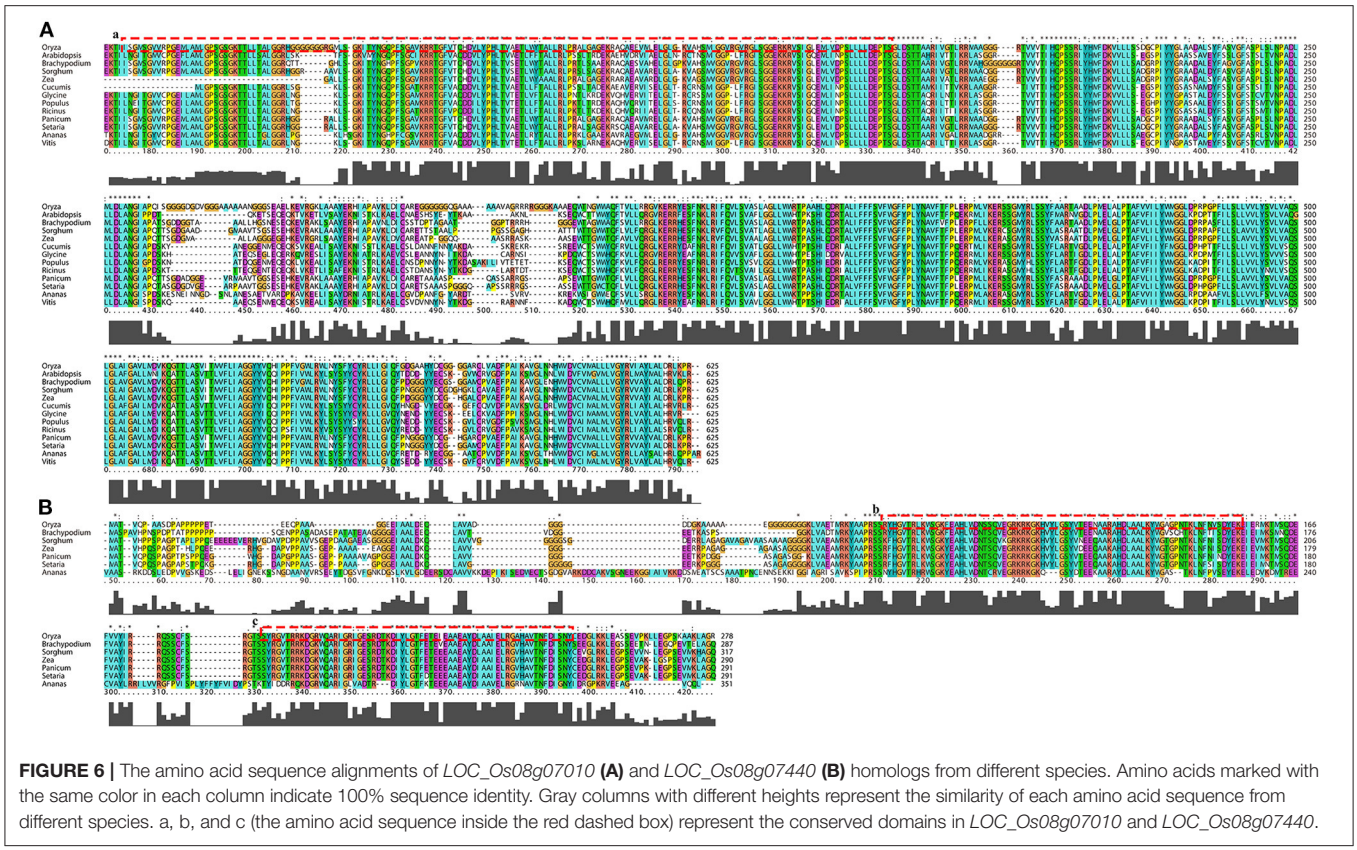


FIGURE 6 | The amino acid sequence alignments of *LOC_Os08g07010* **(A)** and *LOC_Os08g07440* **(B)** homologs from different species. Amino acids marked with the same color in each column indicate 100% sequence identity. Gray columns with different heights represent the similarity of each amino acid sequence from different species. a, b, and c (the amino acid sequence inside the red dashed box) represent the conserved domains in *LOC_Os08g07010* and *LOC_Os08g07440*.

To confirm the candidate genes in *qHTT8*, we designed qRT-PCR primers to detect the expression level of the ten genes encoding proteins in HHZ and 9311 subjected to different durations of heat stress in a phytotron. Compared with other genes, the expression levels of *LOC_Os08g07010* and *LOC_Os08g07440* were significantly higher in HHZ than in 9311, indicating that they were highly induced by heat stress. Furthermore, phylogenetic analysis revealed that *LOC_Os08g07010* (*OsABCG18*) and *LOC_Os08g07440* were both highly homologous to genes from other species. *OsABCG18* and its ortholog *AtABCG14* were identified to play the same essential roles in transporting cytokinins from the root to the shoot. ABCG (ATP-binding cassette G) is one of the transporter families known to be involved in cytokinin transport in *Arabidopsis* and rice (Zhao et al., 2019). The overexpression of *AtABCG14* in *Arabidopsis thaliana* has been reported to improve drought resistance by regulating the inhibition of stomatal opening (Li, 2019). In our work, *LOC_Os08g07010* was found to be highly stress-inducible in a heat-tolerant variety, demonstrating that *LOC_Os08g07010* also plays an important role in the regulation of HT in rice, which may be related to the involvement of cytokinin in the heat stress response. In addition, the overexpression of *OsABCG18* (*LOC_Os08g07010*) was found to improve grain yield by increasing cytokinins in the shoot, further suggesting that this gene plays an important role in crop growth and development (Zhao et al., 2019). TFs play a central role in the response to abiotic stress. The AP2/EREBP superfamily, including four subfamilies (AP2, ERF, DREB and RAV), is one of the largest plant TFs (Riechmann and Meyerowitz, 1998; Muhammad et al., 2012). Previously, dozens of AP2/EREBP genes related to abiotic stress have been identified in plants, such as *OsDREB1A*, *HYR*, *OsEREBP1*, *OsERF48,* and *StDREB2* (Dubouzet et al., 2003; Ambavaram et al., 2014; Jisha et al., 2015; Jung et al., 2017; Mohamed and Aisha, 2019). Overexpression of these stress-related genes could improve abiotic stress tolerance, including tolerance of cold, heat, drought and salt stress. *LOC_Os08g07440*, a gene encoding the AP2 domain-containing protein, was highly induced by heat stress in HHZ than in 9311, suggesting that *LOC_Os08g07440* also plays a role in HT in rice.

In sum, *LOC_Os08g07010* and *LOC_Os08g07440* are two key genes controlling HT at the flowering stage in rice. However, how *qHTT8* affects the physiological changes of rice under heat stress and its molecular functions remain unclear. Additional study is needed to elucidate the role of *qHTT8* in the molecular mechanisms underlying the heat stress response during the flowering stage. Generally, the findings of our study have important practical implications for future efforts to improve the HT of rice.

CONCLUSIONS

In conclusion, 6 h (9:30 am−3:30 pm) of exposure to high temperature (38 °C) for 3 continuous days is sufficient for identifying HT at the flowering stage in rice (*Oryza sativa* L.) (Liang et al., 2016). Ensuring that opened spikelets were removed before the plants were moved into the phytotron ensured the accuracy of the data. A new QTL for HT at the flowering stage in rice, *qHTT8*, was rapidly identified through BSA-seq within the region of 3,555,000–4,520,000 bp on chromosome 8. Ten putative genes controlling rice abiotic stress tolerance were also identified in this target region. Based on the qRT-PCR and sequence analysis, *LOC_Os08g07010* and *LOC_Os08g07440* were identified as the candidate genes controlling HT at the flowering stage in rice. This study provides a fast and effective strategy to identify heat-tolerant QTLs/genes at the flowering stage in rice. In future studies, functional analysis of parental lines will be used to validate the candidate genes by sequence analysis and genetic transformation.

DATA AVAILABILITY STATEMENT

The genetic sequencing data in our article have been deposited in the Nation Center for Biotechnology Information (NCBI) (SRA) database. The BioProject accession number is PRJNA674973, including four sub-accession numbers: SRR12998566, SRR12998567, SRR12998568 and SRR12998569 (https://www.ncbi.nlm.nih.gov/sra/PRJNA674973).

AUTHOR CONTRIBUTIONS

LJ and TL conceived and supervised the research. LC and QW designed the experiments. LC, QW, MT, XZ, YP, XY, GG, RL, and WT performed the experiments and analyzed the results. LC and QW wrote the paper. All authors have read and approved the final manuscript.

FUNDING

This work was funded by the National Natural Science Foundation of China (31560363), the Natural Science Foundation of Guangxi Province of China (2019GXNSFBA245006 and 2017GXNSFAA198188), the Key Research and Development Project of Guangxi (GuikeAB18126050 and 2020AC17001), the Superior Team Project of Guangxi Academy of Agricultural Sciences (2015YT17), the Base Business Project of Guangxi Academy of Agricultural Sciences (GuiNongKe 2020YM72 and 2019M14), and a small highland for research talents of high quality rice breeding in Guangxi (QN-08).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.621871/full#supplementary-material

REFERENCES

Ambavaram, M. M. R., Basu, S., Krishnan, A., Ramegowda, V., Batlang, U., Rahman, L., et al. (2014). Coordinated regulation of photosynthesis in rice increases yield and tolerance to environmental stress. *Nat. Commun.* 5:5302. doi: 10.1038/ncomms6302

Asako, K., Bao, G. L., Ye, S. H., and Katsura, T. (2007). Detection of quantitative trait loci for white-back and basal-white kernels under high temperature stress in Japonica rice varieties. *Breed. Sci.* 57, 107–116. doi: 10.1270/jsbbs.57.107

Austin, D. F., and Lee, M. (1995). Comparative mapping in $F_{2:3}$ and $F_{6:7}$ generations of quantitative trait loci for grain yield and yield components in maize. *Theoret. Appl. Genet.* 92, 817–826. doi: 10.1007/BF00221893

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., and Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature* 575, 109–118. doi: 10.1038/s41586-019-1679-0

Baliuag, N. N. A., Redoña, E. D., Hernandez, J. E., Cruz, P. C. S., and Ye, C. R. (2015). Genetic analysis for heat tolerance and early morning flowering traits at flowering stage in rice (*Oryza sativa* L). *Philipp. J. Crop Sci.* 40, 62–72.

Cao, L. Y., Zhu, J., Zhao, S. T., He, L. B., and Yan, Q. C. (2002). Mapping QTLs for heat tolerance in a DH population from indica-japonica cross of rice (*Oryza sativa* L). *J. Agric. Biotechnol.* 10, 210–214.

Cao, Y. Y., Duan, H., Yang, L. N., Wang, Z. Q., Zhou, S. C., and Yang, J. C. (2008). Effect of heat-stress during meiosis on grain yield of rice cultivars differing in heat-tolerance and its physiological mechanism. *Acta Agron Sin.* 34, 2134–2142. doi: 10.1016/S1875-2780(09)60022-5

Cao, Z. B., Li, Y., Tang, H. W., Zeng, B. H., Tang, X. Y., Long, Q. Z., et al. (2020). Fine mapping of the *qHTB1-1* QTL, which confers heat tolerance at the booting stage, using an *Oryza rufipogon* Griff. introgression line. *Theor. Appl. Genet.* 133, 1161–1175. doi: 10.1007/s00122-020-03539-7

Chen, Q. Q., Yu, S. B., Li, C. H., and Mu, T. M. (2008). Identification of QTLs for heat tolerance at flowering stage in rice, *Sci. Agric. Sin.* 41, 315–321.

Cheng, L. R., Wang, J. M., Uzokwe, V., Meng, L. J., Wang, Y., Sun, Y., et al. (2012). Genetic analysis of cold tolerance at seedling stage and heat tolerance at anthesis in rice. *J. Integr. Agric.* 11, 359–367. doi: 10.1016/S2095-3119(12)60020-3

Costa, M.-C. D., and Farrant, J. M. (2019). Plant resistance to abiotic stresses. *Plants* 8:553. doi: 10.3390/plants8120553

Dietz, K. J., Vogel, M. O., and Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signaling. *Protoplasma* 245, 3–14. doi: 10.1007/s00709-010-0142-8

Driedonks, N., Rieu, I., and Vriezen, W. H. (2016). Breeding for plant heat tolerance at vegetative and reproductive stages. *Plant Reprod.* 29, 67–79. doi: 10.1007/s00497-016-0275-9

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33, 751–763. doi: 10.1046/j.1365-313X.2003.01661.x

El-Esawi, M. A., and Alayafi, A. A. (2019). Overexpression of rice *Rab7* gene improves drought and heat tolerance and increases grain yield in rice (*Oryza sativa* L.). *Genes* 10:56. doi: 10.3390/genes10010056

Fahliani, R. A., Khodambashi, M., Houshmand, S., and Arzani, A. (2010). Estimation of the heritability of agro-morphological traits in rice (*Oryza sativa* L.) using $F_{2:3}$ families. *Afr. J. Agric. Res.* 5, 1297–1303. doi: 10.5897/AJAR10.010

Hatfield, J. L., and Prueger, J. H. (2015). Temperature extremes: effect on plant growth and development. *Weather Clim Extremes* 10, 4–10. doi: 10.1016/j.wace.2015.08.001

Hoang, G. T., Dinh, L. V., Nguyen, T. T., Ta, N. K., Gathignol, F., Mai, C. D., et al. (2019). Genome-wide association study of a panel of Vietnamese rice landraces reveals new QTLs for tolerance to water deficit during the vegetative phase. *Rice* 12, 4–23. doi: 10.1186/s12284-018-0258-6

Huang, J., Gao, Y., Jia, H., Liu, L., Zhang, D., and Zhang, Z. (2015). Comparative transcriptomics uncovers alternative splicing changes and signatures of selection from maize improvement. *BMC Genomics* 16:363. doi: 10.1186/s12864-015-1582-5

Huang, X. P., Zhang, H. Y., Lei, G., Wang, Z. M., Zhang, Z., He, C., et al. (2017). Analysis on comparative proteomics of rice grain between heat-tolerant and heat sensitive lines under high night temperature stress at filling stage. *Chin. J. Rice Sci.* 31, 13–22.

IPCC (2007). "Summary for policymakers," in *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, et al. (Cambridge: Cambridge University Press), 1–18.

Jagadish, S. V. K., Cairns, J., Lafitte, R., Wheeler, T. R., Price, A. H., and Craufurd, P. Q. (2010). Genetic analysis of heat tolerance at anthesis in rice. *Crop Sci.* 50, 1633–1641. doi: 10.2135/cropsci2009.09.0516

Jagadish, S. V. K., Craufurd, P. Q., and Wheeler, T. R. (2007). High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). *J. Exp. Bot.* 58, 1627–1635. doi: 10.1093/jxb/erm003

Jagadish, S. V. K., Craufurd, P. Q., and Wheeler, T. R. (2008). Phenotyping parents of mapping populations of rice for heat tolerance during anthesis. *Crop Sci.* 48, 1140–1146. doi: 10.2135/cropsci2007.10.0559

Jagadish, S. V. K., Septiningsih, E. M., Kohli, A., Thomson, M. J., Ye, C. R., Redoña, E. D., et al. (2012). Genetic advances in adapting rice to a rapidly changing climate. *J Agron Crop Sci.* 198, 360–373. doi: 10.1111/j.1439-037X.2012.00525.x

Jisha,V., Dampanaboina, L., Vadassery, J., Mithö, A., Kappara, S., and Ramanan, R. (2015). Overexpression of an AP2/ERF type transcription factor *OsEREBP1* confers biotic and abiotic stress tolerance in rice. *PLoS ONE* 10:e0127831. doi: 10.1371/journal.pone.0127831

Jung, H., Chung, P. J., Park, S.-H., Redillas, M. C. F. R., Kim, Y. S., Suh, J.-W., et al. (2017). Overexpression of *OsERF48* causes regulation of *OsCML16*, a Calmodulin-like protein gene that enhances root growth and drought tolerance. *Plant Biotechnol. J.* 15, 1295–1308. doi: 10.1111/pbi.12716

Kilasi, N. L., Singh, J., Vallejos, C. E., Ye, C. R., Jagadish, K. S. V., Kusolwa, P., et al. (2018). Heat stress tolerance in rice (*Oryza sativa* L.): identification of quantitative trait loci and candidate genes for seedling growth under heat stress. *Front. Plant Sci.* 9, 1–11 doi: 10.3389/fpls.2018.01578

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lee, J., Lee, W., and Kwon, S. W. (2015). A quantitative shotgun proteomics analysis of germinated rice embryos and coleoptiles under low-temperature conditions. *Proteome Sci.* 13, 27–37. doi: 10.1186/s12953-015-0082-5

Li, B. F. (2019). *The molecular mechanism of AtABCG14 regulating drought resistance in Arabidopsis* [Master dissertation]. Zhejiang Normal University, Jinhua, China.

Li, M. M., Li, X., Yu, L. Q., Wu, J. W., Li, H., Liu, J., et al. (2018). Identification of QTLs associated with heat tolerance at the heading and flowering stage in rice (*Oryza sativa* L.). *Euphytica* 214, 70–80. doi: 10.1007/s10681-018-2136-0

Li, N., Liu, H. L., Sun, J., Zheng, H. L., Wang, J. G., Yang, L. M., et al. (2018). Transcriptome analysis of two contrasting rice cultivars during alkaline stress. *Sci. Rep.* 8, 9586–9601. doi: 10.1038/s41598-018-27940-x

Liang, T. F., Wang, Q., Chen, L., Zhang, X. L., Lv, R. H., Tao, W., et al. (2016). Identification method for heat resistance of rice based on accumulated hourly temperature. *J Southern Agric.* 47, 537–541.

Lim, S. D., Cho, H. Y., Park, Y. C., Ham, D. J., Lee, J. K., and Jang, C. S. (2013). The rice RING finger E3 ligase, *OsHCI1*, drives nuclear export of multiple substrate proteins and its heterogeneous overexpression enhances acquired thermotolerance. *J. Exp. Bot.* 64, 2899–2914. doi: 10.1093/jxb/ert143

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Ma, T. C., Chen, R. J., Yu, R. R., Zeng, H. L., and Zhang, D. P. (2009). Differential global genomic changes in rice root in response to low-, middle-, and high-osmotic stresses. *Acta Physiol. Plant* 31, 773–785. doi: 10.1007/s11738-009-0291-6

Mohamed, A. E., and Aisha, A. A. (2019). Overexpression of *StDREB2* transcription factor enhances drought stress tolerance in cotton (*Gossypium barbadense* L.). *Genes* 10:142. doi: 10.3390/genes10020142

Mohammed, A. R., and Tarpley, L. (2009). High nighttime temperatures affect rice productivity through altered pollen germination and spikelet fertility. *Agric. For. Meteorol.* 149, 999–1008. doi: 10.1016/j.agrformet.2008.12.003

Muhammad, R., He, G., Yang, G., Javeed, H., and Yan, X. (2012). AP2/ERF Transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. *Evol. Appl.* 8, 321–355. doi: 10.4137/EBO.S9369

Murata, K., Iyama, Y., Yamaguchi, T., Ozaki, H., Kidani, Y., and Ebitani, T. (2014). Identification of a novel gene (*Apq1*) from the indica rice cultivar 'Habataki' that improves the quality of grains produced under high temperature stress. *Breed. Sci.* 64, 273–281. doi: 10.1270/jsbbs.64.273

Nguyen, V. N. T., Moon, S., and Jung, K.-H. (2014). Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses. *J. Plant Physiol.* 171, 1276–1288. doi: 10.1016/j.jplph.2014.05.006

Nubankoh, P., Wanchana, S., Saensuk, C., Ruanjaichon, V., Cheabu, S., Vanavichit, A., et al. (2020). QTL-seq reveals genomic regions associated with spikelet fertility in response to a high temperature in rice (*Oryza sativa* L.). *Plant Cell Rep.* 39, 149–162. doi: 10.1007/s00299-019-02477-z

Pan, Y. H., Liang, H. F., Gao, L. J., Dai, G. X., Chen, W. W., Yang, X. H., et al. (2020). Transcriptomic profiling of germinating seeds under cold stress and characterization of the cold-tolerant gene *LTG5* in rice. *BMC Plant Biol.* 20:371. doi: 10.1186/s12870-020-02569-z

Park, K. J., Sa, K. J., Koh, H. J., and Lee, J. K. (2013). QTL analysis for eating quality-related traits in an $F_{2:3}$ population derived from waxy corn × sweet corn cross. *Breed. Sci.* 63, 325–332. doi: 10.1270/jsbbs.63.325

Patil, S., Srividhya, A., Veeraghattapu, R., Deborah, D. A. K., Kadambari, G. M., Nagireddy, R., et al. (2017). Molecular dissection of a genomic region governing root traits associated with drought tolerance employing a combinatorial approach of QTL mapping and RNA-seq in rice. *Plant Mol. Biol. Rep.* 35, 457–468. doi: 10.1007/s11105-017-1037-z

Raineri, J., Wang, S. H., Peleg, Z., Blumwald, E., and Chan, R. L. (2015). The rice transcription factor *OsWRKY47* is a positive regulator of the response to water deficit stress. *Plant Mol. Biol.* 88, 401–413. doi: 10.1007/s11103-015-0329-7

Rang, Z. W., Jagadish, S. V. K., Zhou, Q. M., Craufurd, P. Q., and Heuer, S. (2011). Effect of high temperature and water stress on pollen germination and spikelet fertility in rice. *Environ. Exp. Bot.* 70, 58–65. doi: 10.1016/j.envexpbot.2010.08.009

Riechmann, J. L., and Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. *Biol. Chem.* 379, 633–646.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., and Mittler, R. (2004). When defense pathways collide, The response of Arabidopsis to a combination of drought and heat stress. *Plant Physiol.* 134, 1683–1696. doi: 10.1104/pp.103.033431

Saha, J., Sengupta, A., Gupta, K., and Gupta, B. (2015). Molecular phylogenetic study and expression analysis of ATP-binding cassette transporter gene family in *Oryza sativa* in response to salt stress. *Comput. Biol. Chem.* 54, 18–32. doi: 10.1016/j.compbiolchem.2014.11.005

Seo, J. S., Joo, J., Kim, M. J., Kim, Y. K., Nahm, B. H., Song, S. I., et al. (2011). *OslHLH148*, a basic helix loop helix protein, interacts with OsJAZ proteins in jasmonate signaling pathway leading to drought tolerance in rice. *Plant J.* 65, 907–921. doi: 10.1111/j.1365-313X.2010.04477.x

Shanmugavadivel, P. S., Amitha Mithra, S. V., Chandra, P., Ramkumar, M. K., Ratan, T., Trilochan, M., et al. (2017). High resolution mapping of QTLs for heat tolerance in rice using a 5K SNP array. *Rice* 10, 1–11. doi: 10.1186/s12284-017-0167-0

Shanmugavadivel, P. S., Chandra, P., and Amitha, M. S. V. (2019). *Advances in Rice Research for Abiotic Stress Tolerance, Chapter 42-Molecular Approaches for Dissecting and Improving Drought and Heat Tolerance in Rice*. Cambridge, UK: Woodhead Publishing, 839–867.

Smita, S., Katiyar, A., Lenka, S. K., Dalal, M., Kumar, A., Mahtha, S. K., et al. (2010). Gene network modules associated with abiotic stress response in tolerant rice genotypes identified by transcriptome meta-analysis. *Funct. Integr. Genomics* 20, 29–49. doi: 10.1007/s10142-019-00697-w

Tabata, M., Hirabayashi, H., Takeuchi, Y., and Ando, I. (2007). Mapping of quantitative trait loci for the occurrence of White-Back kernels associated with high temperatures during the ripening period of rice (*Oryza sativa* L.). *Breed. Sci.* 57, 47–52. doi: 10.1270/jsbbs.57.47

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105

Tazib, T., Kobayashi, Y., Koyama, H., and Matsui, T. (2015). QTL analyses for anther length and dehiscence at flowering as traits for the tolerance of extreme temperatures in rice (*Oryza sativa* L.). *Euphytica* 203, 629–642. doi: 10.1007/s10681-014-1291-1

Thanh, P. T., Phan, P. D. T., Ishikawa, R., and Ishii, T. (2010). QTL analysis for flowering time using backcross population between *Oryza sativa* Nipponbare and *O. rufipogon*. *Genes Genet Syst.* 85, 273–279. doi: 10.1266/ggs.85.273

Wada, T., Miyahara, K., Sonoda, J. Y., Tsukaguchi, T., Miyazaki, M., Tsubone, M., et al. (2015). Detection of QTLs for white-back and basal-white grains caused by high temperature during ripening period in japonica rice. *Breed. Sci.* 65, 216–225. doi: 10.1270/jsbbs.65.216

Wang, Y. L., Zhang, Y. K., Shi, Q. H., Chen, H. Z., Xiang, J., Hu, G. H., et al. (2020). Decrement of sugar consumption in rice young panicle under high temperature aggravates spikelet number reduction. *Rice Sci.* 27, 44–55. doi: 10.1016/j.rsci.2019.12.005

Welch, J. R., Vincent, J. R., Auffhammer, M., Moya, P. F., Dobermann, A., and Dawe, D. (2010). Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14562–14567. doi: 10.1073/pnas.1001222107

Xiao, Y. H., Pan, Y., Luo, L. H., Zhang, G. L., Deng, H. B., Dai, L. Y., et al. (2011). Quantitative trait loci associated with seed set under high temperature stress at the flowering stage in rice. *Euphytica* 178, 331–338. doi: 10.1007/s10681-010-0300-2

Yang, L. M., Liu, H. L., Zhao, H. W., Wang, J. G., Sun, J., Zheng, H. L., et al. (2019). Mapping quantitative trait loci and meta–analysis for cold tolerance in rice at booting stage. *Euphytica* 215, 88–102. doi: 10.1007/s10681-019-2410-9

Yang, X. H., Xia, X. Z., Zhang, Z. Q., Nong, B. X., Zeng, Y., Xiong, F. Q., et al. (2017). QTL mapping by whole genome re-sequencing and analysis of candidate genes for nitrogen use efficiency in rice. *Front. Plant Sci.* 8:1634. doi: 10.3389/fpls.2017.01634

Ye, C. R., Argayoso, M. A., Redoña, E. D., Sierra, S. L., Laza, M. A., Dilla, C. J., et al. (2012). Mapping QTL for heat tolerance at flowering stage in rice using SNP markers. *Plant Breed.* 131, 33–41. doi: 10.1111/j.1439-0523.2011.01924.x

Ye, C. R., Tenorio, F. A., Argayoso, M. A., Laza, M. A., Koh, H. J., Redoña, E. D., et al. (2015a). Identifying and confirming quantitative trait loci associated with heat tolerance at flowering stage in different rice populations. *BMC Genet.* 16, 41–50. doi: 10.1186/s12863-015-0199-7

Ye, C. R., Tenorio, F. A., Redoña, E. D., Morales-Cortezano, P. S., Cabrega, G. A., Jagadish, K. S. V., et al. (2015b). Fine-mapping and validating *QHTSF41* to increase spikelet fertility under heat stress at flowering in rice. *Theor. Appl. Genet.* 128, 1507–1517. doi: 10.1007/s00122-015-2526-9

Zhang, K., Novak, O., Wei, Z., Gou, M., Zhang, X., Yu, Y., et al. (2014). *Arabidopsis ABCG14* protein controls the acropetal translocation of root-synthesized cytokinins. *Nat. Commun.* 5:3274. doi: 10.1038/ncomms4274

Zhang, Y. X., Qin, G., Ma, Q. Q., Wei, M. Y., Yang, X. H., Ma, Z. F., et al. (2020). Identification of major locus *Bph35* resistance to brown planthopper in rice. *Rice Sci.* 27, 237–245. doi: 10.1016/j.rsci.2020.04.006

Zhao, J., Yu, N., Ju, M., Fan, B., Zhang, Y., Zhu, E., et al. (2019). ABC transporter *OsABCG18* controls the shootward transport of cytokinins and grain yield in rice. *J. Exp. Bot.* 70, 6277–6291. doi: 10.1093/jxb/erz382

Zhao, L., Lei, J. G., Huang, Y. J., Zhu, S., Chen, H. P., Huang, R. L., et al. (2016). Mapping quantitative trait loci for heat tolerance at anthesis in rice using chromosomal segment substitution lines. *Breed. Sci.* 66, 358–366. doi: 10.1270/jsbbs.15084

Zheng, J., Li, X. X., Su, H. Y., Zhou, J. J., Qin, X. J., Nan, W. B., et al. (2017). Construction of a genetic linkage map and QTL location for heat tolerance in Japonica rice resources Rejing35. *J. Nucl. Agric. Sci.* 31, 844–851.

Zhu, C. L., Jiang, L., Zhang, W. W., Wang, C. M., Zhai, H. Q., and Wan, J. M. (2006). Identifying QTLs for thermo-tolerance of amylose content and gel consistency in rice. *Chin. J. Rice Sci.* 20, 248–252.

Zhu, S., Huang, R. L., Wai, H. P., Xiong, H. L., Shen, X. H., He, H. H., et al. (2017). Mapping quantitative trait loci for heat tolerance at the booting stage using chromosomal segment substitution lines in rice. *Physiol. Mol. Biol. Plants* 23, 817–825. doi: 10.1007/s12298-017-0465-4

Zou, C., Wang, P. X., and Xu, Y. B. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J.* 14, 1941–1955. doi: 10.1111/pbi.12559

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chen, Wang, Tang, Zhang, Pan, Yang, Gao, Lv, Tao, Jiang and Liang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 January 2021
doi: 10.3389/fgene.2020.609603

Allelic Diversity, Structural Analysis, and Genome-Wide Association Study (GWAS) for Yield and Related Traits Using Unexplored Common Bean (*Phaseolus vulgaris* L.) Germplasm From Western Himalayas

Reyazul Rouf Mir[1], Neeraj Choudhary[2], Vanya Bawa[2], Sofora Jan[1], Bikram Singh[2], Mohd Ashraf Bhat[1], Rajneesh Paliwal[3], Ajay Kumar[4], Annapurna Chitikineni[5], Mahendar Thudi[5] and Rajeev Kumar Varshney[5]*

[1] Division of Genetics and Plant Breeding, Faculty of Agriculture, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir (SKUAST-K), Sopore, India, [2] Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu, India, [3] The International Institute of Tropical Agriculture, Ibadan, Nigeria, [4] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [5] Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

The north-western Indian Himalayas possesses vast diversity in common bean germplasm due to several years of natural adaptation and farmer's selection. Systematic efforts have been made for the first time for the characterization and use of this huge diversity for the identification of genes/quantitative trait loci (QTLs) for yield and yield-contributing traits in common bean in India. A core set of 96 diverse common bean genotypes was characterized using 91 genome-wide genomic and genic simple sequence repeat (SSR) markers. The study of genetic diversity led to the identification of 691 alleles ranging from 2 to 21 with an average of 7.59 alleles/locus. The gene diversity (expected heterozygosity, *He*) varied from 0.31 to 0.93 with an average of 0.73. As expected, the genic SSR markers detected less allelic diversity than the random genomic SSR markers. The traditional clustering and Bayesian clustering (structural analysis) analyses led to a clear cut separation of a core set of 96 genotypes into two distinct groups based on their gene pools (Mesoamerican and Andean genotypes). Genome-wide association mapping for pods/plant, seeds/pod, seed weight, and yield/plant led to the identification of 39 significant marker–trait associations (MTAs) including 15 major, 15 stable, and 13 both major and stable MTAs. Out of 39 MTAs detected, 29 were new MTAs reported for the first time, whereas the remaining 10 MTAs were already identified in earlier studies and therefore declared as validation of earlier results. A set of seven markers was such, which were found to be associated with multiple (two to four) different traits. The important MTAs will be used for common bean molecular breeding programs worldwide for enhancing common bean yield.

Keywords: common bean, north-western Himalayas, allelic diversity, structural analysis, GWAS, QTLs/genes for yield traits

OPEN ACCESS

Edited by:
Santosh B. Satbhai,
Indian Institute of Science Education and Research Mohali, India

Reviewed by:
Saurabh Badoni,
International Rice Research Institute (IRRI), Philippines
Ze Peng,
University of Florida, United States

****Correspondence:***
Reyazul Rouf Mir
imrouf2006@gmail.com;
rrmir@skuastkashmir.ac.in

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *23 September 2020*
Accepted: *18 December 2020*
Published: *28 January 2021*

Citation:
Mir RR, Choudhary N, Bawa V, Jan S, Singh B, Bhat MA, Paliwal R, Kumar A, Chitikineni A, Thudi M and Varshney RK (2021) Allelic Diversity, Structural Analysis, and Genome-Wide Association Study (GWAS) for Yield and Related Traits Using Unexplored Common Bean (Phaseolus vulgaris L.) Germplasm From Western Himalayas. Front. Genet. 11:609603. doi: 10.3389/fgene.2020.609603

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is one of the most important diploid grain legume crops ($2n = 2x = 22$) with a small genome size of 587 Mbp (Broughton et al., 2003). It is the major source of calories and proteins for the people in developing countries of the world (FAO[1]). Common bean is one of the most ancient crops of the Americas (Broughton et al., 2003; McClean et al., 2004) and possesses two important already diverged gene pool species: the Mesoamerican and Andean gene pool species. The Mesoamerican gene pool species is distributed from northern Mexico to Colombia, whereas the Andean gene pool species is distributed from southern Peru to north-western Argentina (Kwak and Gepts, 2009). The presence of two gene pools in common bean raises the following questions during common bean germplasm evaluation and characterization: (i) the relationship between the germplasm from two gene pools, (ii) the diversity/variation present within and between these gene pools, (iii) the quantitative differences in genetic diversity, and (iv) the levels of linkage disequilibrium (Kwak and Gepts, 2009). The characterization of genetic diversity is one of the most important subject areas of crop research. The characterized crop germplasm forms the basis of crop improvement programs and the development of genetic resources, such as mapping populations and core collections, for the genetic dissection of important traits through quantitative trait locus (QTL) mapping and genome-wide association mapping approaches (McClean et al., 2012).

A huge unexplored diversity has been observed in common bean germplasm in Jammu and Kashmir: a north-western Himalayan region in India and this region is famous for producing high-quality beans. The common bean germplasm from the area have different market classes, plant types, seed quality traits, and agro-ecological adaptation (Choudhary et al., 2018a,b). Keeping in view the diversity of common bean in this region, it will not be un-wise to call this area as "secondary center of diversity" for common bean. The huge diversity that is available in the common bean germplasm from western Himalayas of India is perhaps due to the differential adaptive evolutionary process that is happening continuously over the last several hundred years since their introduction in western Himalayas by travelers from Portugal, England, Holland, France, China, and Pakistan (Rana et al., 2015; Choudhary et al., 2018b). The extent of genetic diversity and the origin of common bean in the Jammu and Kashmir region were recently characterized using *Phaseolin locus* (*Phs*) assays and sequencing of internal transcribed spacer (ITS) region (Choudhary et al., 2018b). Out of a set of 428 common bean lines, a diverse subset of 96 lines was selected based on cluster analysis using few qualitative traits and site of collection. The core set of 96 lines comprised 54 local landraces from 11 hotspots of the Himalayan region of Jammu and Kashmir and 42 exotic lines from 11 different countries. The phaseolin patterns of these 96 lines revealed the presence of lines with "S"-type phaseolin and "T"-type phaseolin patterns. The common bean germplasm from the Kashmir region possess both S- and T-type phaseolins, whereas the germplasm from the Jammu region possess only S-type phaseolins. Few earlier studies have also attempted to characterize this huge diversity of common bean in north-western Himalayas using morphological traits (Sofi et al., 2014; Saba et al., 2016) and less reliable random amplification of polymorphic DNA (RAPD) markers or only limited simple sequence repeat (SSR) markers (Zargar et al., 2016).

Different genomics tools and molecular techniques now offer much better understanding to assess the ability of crop genetic diversity. SSR markers are considered suitable for assessing genetic variation and allele mining because they are highly informative (Powell et al., 1996; Gupta et al., 2002; Mir and Varshney, 2013; Mir et al., 2013). Their advantages for diversity studies also include uniform genome coverage, high levels of polymorphism, co-dominance, and an easy-to-implement, specific polymerase chain reaction (PCR)-based assay (Pejic et al., 1998; Gupta et al., 2002; Mir and Varshney, 2013; Mir et al., 2013). While going through the literature in common bean, it was noticed that molecular markers have played an important role in the characterization and assessment of genetic diversity of landraces and farmers varieties (Blair et al., 2006; Asfaw et al., 2009; Angioi et al., 2010; Sharma et al., 2013; Okii et al., 2014; Buah et al., 2017). The restriction fragment length polymorphism (RFLP) markers were used as the first molecular marker system for the study of genetic diversity (Becerra-Velasquez and Gepts, 1994). The amplified fragment length polymorphism (AFLP) markers were used to study wild beans germplasm (Tohme et al., 1996), diversity and origin studies of Andean local landraces (Beebe et al., 2001), and DNA fingerprinting studies to characterize yellow beans from both gene pools (Pallottini et al., 2004). The RAPD markers were mainly used to study genetic diversity and population structure among common bean germplasm and landraces (Beebe et al., 2000; Razvi et al., 2013; Bukhari et al., 2015). SSR markers have been widely used in genetic diversity and population structure studies (Blair et al., 2006, 2009; Benchimol et al., 2007; Khaidizar et al., 2012; Okii et al., 2014; Fisseha et al., 2016). However, it is important to mention here that common bean population studies with SSR markers have been performed using only a small number of landraces or breeding lines or they have focused on certain geographic regions only (Metais et al., 2002; Blair et al., 2006; Díaz and Blair, 2006). However, a systematic effort to characterize this huge diversity of Himalayan beans using molecular markers is still not available, although extremely useful for bean improvement.

Therefore, the present study was conducted to better understand the genetic diversity and population structure available in Himalayan bean germplasm using SSR markers. Efforts were also made to compare the genetic diversity revealed by genic and random SSR markers. Genome-wide association studies (GWASs) were conducted to identify molecular markers associated with yield and yield-contributing traits using precise and accurate genome-wide SSR marker data and trait data on yield and yield-contributing traits collected over 2 years. The knowledge and genetic/genomics resources (candidate genotypes for yield and related traits, associated markers, validated

[1]http://faostat.fao.org/faostat

markers, and major/stable genes/QTLs) generated/developed in this study will be invaluable to the bean breeding programs aimed at improving yield and related traits in common bean throughout the world.

MATERIALS AND METHODS

Plant Materials

The present study comprised a core set of 96 diverse common bean genotypes, including 54 local landraces from 11 hotspots of Jammu and Kashmir and 42 exotic lines belonging to 11 different countries. The 96 lines include 51 Andean types with "T"-type phaseolin and 45 Mesoamerican types having "S"-type phaseolin. Among 54 local landraces, 32 lines are Andean type and 22 lines are Mesoamerican type. Among 42 exotic lines, 19 are Andean type and 23 are Mesoamerican type (**Supplementary Table 1**). The diverse 96 lines have been carefully selected from a set of 428 lines based on the evaluation of qualitative data, such as seed color, seed shape, flower color, and distribution in different regions in Jammu and Kashmir, India and other 11 different countries (for more details about germplasm and selection criteria, see Choudhary et al., 2018b). In short, both quality data and information about their collection sites were kept into consideration while selecting the diverse set of 96 lines. The quality data of 428 lines were used in clustering, and a dendrogram was prepared. The dendrogram in addition to landrace collection site information was used for the selection of the final set of 96 lines (Choudhary et al., 2018b). The local landraces were collected from different common bean growing regions of Jammu and Kashmir, and exotic lines were procured from the National Bureau of Plant Genetic Resources (NBPGR), Shimla, Himachal Pradesh, India.

Trait Phenotyping and Analysis of Data

The diverse set of 96 lines was phenotyped for four important yield-related traits including pods per plant, seeds per pod, 100-seed weight (g), and yield per plant (g). The data on these traits were recorded at two important common bean growing regions in Jammu and Kashmir, i.e., at Bhaderwah (32.980033°N 75.713706°E) located at an elevation of 5292 ft and at SKUAST-Jammu, Chatha-Jammu, India (32.73°N 74.87°E) located at an elevation of 1000 ft. The 96 genotypes were evaluated in an Augmented Block Design (ABD) that consisted of six blocks, each containing 16 genotypes and three local checks allotted to each block randomly. The plots were kept free from weeds, diseases, and pests throughout the cropping cycle. Standard agronomic practices were followed for normal crop growth during both years. Five plants in each genotype were selected for recording the data, and the mean data from two locations were used in statistical analysis. The mean data were analyzed to estimate variability parameters, such as phenotypic coefficient of variation (PCV), genotypic coefficient of variation (GCV), heritability, genetic advance, and correlation coefficient, using the software program "Windostat ver 9.2[2]" developed by Indostat services, Hyderabad, India[3]. The data on these four traits for both the locations were also utilized to identify significant marker–trait associations (MTAs) in GWAS using different software programs.

Genomic DNA Extraction

The genomic DNA of the collection of bean genotypes was extracted using the Qiagen DNeasy DNA extraction kit following standard protocols. More details about checking of quantity and quality are available elsewhere (Choudhary et al., 2018a).

Selection of SSR Markers

A set of markers (91 SSR markers) selected for this work was done based on several parameters and criteria and included: (i) high polymorphic information content (PIC) values (>0.6), (ii) maximum number of alleles detected in earlier studies, (iii) genic SSR vs. random genomic SSR, and (iv) uniform distribution on all the 11 linkage groups. The details of these 91 SSR markers are also available elsewhere (see **Supplementary Tables 2, 3**). Out of these 91 selected markers, 32 were either genic markers associated with different traits or EST-derived SSR markers (**Supplementary Table 3**). The markers and their primer sequences once selected were synthesized on contract from Sigma–Aldrich, Bangalore, India.

PCR and SSR Marker Genotyping

The genotyping of 45 SSR markers was done in the Molecular Breeding Laboratory of the Division of Genetics and Plant Breeding, SKUAST-Jammu, Chatha, Jammu using polyacrylamide gel electrophoresis (PAGE) systems (High-throughput Dual Gel Vertical Electrophoresis System) by Peqlab/CBS Scientific, United States, followed by silver staining before recording the data. For PAGE, the PCR amplifications were done in 10 μl reaction volume using 20 ng of DNA template, 5.0 pmol forward reverse primers, 2.5 mM of each dNTPs, 1 × buffer, 2.0 mM $MgCl_2$, 10 mM Tris–HCl, 50 mM KCl, and 1.0 U of *Taq* polymerase (Sigma/HiMedia). The thermal cycler (Peqlab) was programmed as follows: initial denaturation at 95°C for 5 min, 40 cycles of 94°C for 1 min of denaturation, 50–60°C for 1 min of annealing temperature, 72°C for 1 min, and final extension at 72°C for 8 min. The resulting PCR products were run in 10% PAGE to score the allele polymorphism of various markers.

In addition, the genotyping was also done for 46 SSR markers using an ABI 3730 automatic DNA Sequencer Genotyping Platform (Applied Biosystems, Foster City, CA, United States) at the Centre of Excellence in Genomics and Systems Biology (CEGSB), ICRISAT, Hyderabad, Telangana, India. The genotyping involves PCR amplifications of SSR loci using a thermal cycler (GeneAmp PCR System 9700; Applied Biosystems, Foster City, CA, United States), followed by amplification on 1.2% agarose gel for confirming PCR amplification. Separation of amplified products was done using capillary electrophoresis and GeneMapper software version 4.0 (Applied Biosystems, Foster City, CA, United States).

[2]https://windostat.software.informer.com/

[3]http://www.indostat.org/

SSR Marker Data Analysis

Several parameters of genetic diversity including the most important PIC value and the number of alleles/locus were used to assess the extent of genetic diversity available in the common bean core set. The GenAlEx software program (Peakall and Smouse, 2006) was used to calculate genetic diversity parameters, such as genetic distance, number of alleles, number of effective alleles, number of private alleles, number of common alleles, observed heterozygosity, and expected heterozygosity. The diversity parameters were calculated separately for random genomic SSR markers and genic SSR markers, as well as together on the whole population. The analysis was repeated separately by classifying the core set of 96 lines into exotic vs. local landraces and Mesoamerican vs. Andean gene pool landraces. The PIC value for each SSR was calculated manually using Microsoft Excel following Botstein et al. (1980). DARwin version 5.0 was used to calculate pair-wise genetic distances and to construct the dissimilarity matrix (Perrier et al., 2003). The dissimilarity matrix thus obtained was subjected to cluster analysis using the unweighted neighbor-joining (UNJ) method (Gascuel, 1997), followed by bootstrap analysis with 1000 permutations to obtain a dendrogram (Perrier et al., 2003; Mir et al., 2012a).

Analysis of Molecular Variance (AMOVA)

To test the genetic variation within and between cultivars of exotic and local landraces, analysis of molecular variance (AMOVA) was carried out using the software program GenAlEx (Peakall and Smouse, 2006).

Population Structure Analysis

Population structural analysis, which is a model-based clustering, was done to find out the number of subpopulations in our common bean population of 96 lines, using the software program STRUCTURE version 2.3.4 (Pritchard et al., 2000). We tested the number of subpopulations (K) from 1 to 10, and each was repeated three times. For each run, burn-in was set at 100,000, iteration was set at 200,000, and a model without admixture and correlated allele frequencies was used. The run with maximum likelihood was used to assign our 96 common bean lines into subpopulations. This assignment obtained through maximum-likelihood approach was further confirmed by a modified Delta-K (ΔK) method, which provides the real number of clusters/subpopulations (Evanno et al., 2005). Within a subpopulation, the genotypes with affiliation probabilities (inferred ancestry) $\geq$ 80% were assigned to a distinct subpopulation, and those with $<$ 80% were treated as admixture, i.e., these genotypes seem to have a mixed ancestry from parents belonging to different gene pools or geographical origin (Mir et al., 2012b).

MTAs for Yield and Yield-Contributing Traits

Association mapping was conducted for the identification of significant MTAs for yield and yield-contributing traits. The trait data on 100-seed weight, pods per plant, seeds per pod, and yield per plant for two locations along with SSR marker data were used in the software program TASSEL 3.0[4] to identify significant MTAs. The analysis of MTAs was done using two different models including general linear model (GLM) based on the Q-matrix derived from the STRUCTURE software and mixed linear model (MLM) based on both the Q-matrix and the kinship matrix (K-matrix) derived from the marker data using the TASSEL software program (for details, see Choudhary et al., 2018a). The significance of MTAs was described in terms of *P*-value ($P \leq 0.05$ for significant markers). The Manhattan plot and quantile–quantile (QQ) plot were also prepared using the software program TASSEL.

RESULTS

Trait Variability for Four Yield-Contributing Traits

Yield-contributing traits, such as the number of pods/plant, the number of seeds/pod, 100-seed weight, and grain yield/plant, are important target traits in common bean breeding programs worldwide. During the present study, the analysis of these four important traits in a core set of 96 lines revealed a broad spectrum of variation as indicated by the wide range and high PCV and GCV values. The GCV values were the highest for yield per plant (59.56), followed by 100-seed weight (38.86) and pods per plant (38.12), whereas a lower value was recorded for seeds per pod (16.97). GCV values were lower than PCV values for all traits indicating a significant influence of environment on these traits, underlining the need to test the stability of performance across a range of environments (**Table 1**). A similar trend was observed for broad sense heritability and genetic advance. The highest expected genetic advance (measure of genetic gain while exercising selection) was observed for yield (110.47%), followed by 100-seed weight and pods per plant, whereas the lowest value was recorded for seeds per pod (26.82%) (**Table 1**). The parameters including PCV, GCV, heritability, and expected genetic gain are of paramount importance as they define the limits of genetic gain that can be achieved through selection. In the present study, all the component traits were significantly correlated with grain yield (**Table 1**).

Analysis of Variance (ANOVA)

The analysis of variance (ANOVA) of the field experiment of bean germplasm at two locations (Jammu and Bhaderwah during *Rabi* 2014–15 and *Kharif* 2015) was conducted for four quantitative traits including pods per plant, seeds per pod, 100-seed weight, and yield per plant, and the results of the mean sum of squares (MSS) were calculated separately for both locations. Non-significant difference was found among all four traits at Jammu location (**Table 2**), but among the genotypes, all four traits exhibited significant differences ($P = 0.01$). Similarly,

[4]http://www.maizegenetics.net

TABLE 1 | Trait variability for yield and yield-contributing traits in pooled data over different environments.

Trait	Mean ± SE	Range	PCV	GCV	Heritability	Genetic advance as% of mean	Correlation with yield per plant
Pods per plant	9.61 ± 0.83	5.10–18.59	40.28	38.12	85.32	67.01	0.596**
Seeds per pod	5.07 ± 0.27	3.23–7.31	18.80	16.97	76.52	26.82	0.516**
100-seed weight (g)	32.05 ± 0.29	11.57–65.84	38.90	38.86	98.83	80.56	0.516**
Yield per plant (g)	15.47 ± 1.77	3.61–69.24	61.04	59.56	80.02	110.47	–

*PCV, phenotypic coefficient of variation; GCV, genotypic coefficient of variation; SE, standard error. **indicates significance at p = 0.01.*

TABLE 2 | Analysis of variance for morphological traits of 96 common bean lines at two testing locations (SKUAST-Chatha, Jammu and Bhaderwah, Jammu).

Source of variation	df	Location	Mean squares			
			Pods per plant	Seeds per pod	100-seed weight (g)	Yield per plant (g)
Replications	2	SKUAST-Chatha	0.46	0.44	0.29	1.11
		Bhaderwah	3.32	0.19	0.74	7.02
Genotypes	95	SKUAST-Chatha	36.93**	2.09**	448.3**	253.00**
		Bhaderwah	45.07**	2.51**	483.57**	276.12**
Error	190	SKUAST-Chatha	2.05	0.18	0.18	9.88
		Bhaderwah	2.02	0.25	0.36	8.90
Total	287	SKUAST-Chatha	15.99	-0.08	129.00	279.83
		Bhaderwah	148.63	-0.52	356.84	1937.36

**Significant at 5% level. **Significant at 1% level.*

for data recorded at Bhaderwah, the differences were non-significant for replication, but significant among genotypes ($P = 0.01$).

Trait Correlations

The correlation analysis showed a significant positive correlation between seeds per pod with pods per plant. Yield per plant showed positive and highly significant correlations with three other yield component traits, *viz.*, pods per plant, seeds per pod, and 100-seed weight at both locations. However, 100-seed weight has a significant negative correlation with pods per plant at both locations (for more details, see **Supplementary Table 4**).

Allelic Diversity

Among all the 91 SSR markers tested on a set of 96 common bean lines, only one SSR marker "BMd44" was found to be monomorphic. The remaining 90 SSR markers detected multiple alleles in 96 genotypes. A total of 691 alleles were detected in all the 96 genotypes by 90 polymorphic SSR markers. The number of alleles detected varied from 2 for SSR marker Bmr205 to 21 for SSR marker BM187, with an average of 7.59 alleles/locus (**Supplementary Table 2**). The number of alleles with a frequency $\geq$ 5% was 5.31, and the number of effective alleles was 4.86. Similarly, gene diversity (expected heterozygosity, *He*) varied from 0.31 to 0.93 with an average of 0.73 (**Table 3**). The lowest *He* was recorded for SSR marker GATS54 and the highest for SSR marker BM187.

Allelic Diversity of Local vs. Exotic Beans

Among the 91 SSR markers tested on 54 local and 42 exotic common bean genotypes, we observed a total of 621 alleles in exotic germplasm and 610 alleles in local common bean germplasm. The number of alleles in exotic bean germplasm varied from 2 to 17 with an average of 6.82 alleles/locus. Similarly, the number of alleles in local bean germplasm varied from 2 to 16 with an average of 6.7 alleles/locus. The numbers of alleles with a frequency $\geq$ 5% were 4.92 for exotic and 4.97 for local lines. The numbers of effective alleles were 4.56 and 4.43, respectively, for exotic and local beans (**Figure 1A**). The number of private alleles in exotic beans was 81 against 70 in local common bean landraces with an average of 0.89 in exotic and 0.77 in local beans. The total number of common alleles between the two groups was 540 with an average of 5.94 alleles. Therefore, a set of 81 alleles was present exclusively in exotic beans, and 70 were present exclusively in local germplasm. While comparing gene diversity between the two groups, it was noticed that it does not differ much as the average *He* in exotic beans was 0.73 against 0.72 in local beans (**Table 3** and **Figure 1A**).

Allelic Diversity of Mesoamerican vs. Andean Beans

Among the 96 lines of core set, 51 lines belong to Andean types with "T"-type phaseolin, and the remaining 45 lines were of Mesoamerican type having "S"-type phaseolin. The 91 SSR markers tested during the present study detected 573 alleles in Mesoamerican beans (average: 6.30, range: 2–16) and 577 in Andean beans (average: 6.35, range: 2–15). The average private allele in Mesoamerican beans was 1.25 against 1.29 in Andean beans. We also observed that the average *He* in Mesoamerican beans was 0.67 against 0.65 in Andean beans. The Nei's genetic distance between the two populations was found to be 0.61, and the genetic differentiation (pair-wise population Fst) between these two populations was found to be 0.116 (**Table 3** and **Figure 1B**).

TABLE 3 | Summary of different allelic diversity attributes of all 91 SSR markers, genic markers, and random SSR markers in a single population of 96 common bean lines and in two subpopulation populations (indigenous vs. exotic and Mesoamerican vs. Andean populations).

Allelic diversity attribute	Single population of 96 lines			Exotic	Indigenous	Exotic	Indigenous	Exotic	Indigenous	Mesoamerican	Andean
	All 91 SSR markers	Random 59 markers	Genic 32 markers	All 91 markers		Random 59 markers		Genic 32 markers		All 91 markers	
Na	7.593	7.966	6.906	6.824	6.703	7.220	7.169	6.094	5.844	6.297	6.341
Na freq. ≥5%	5.308	5.797	4.406	4.912	4.967	5.220	5.458	4.344	4.063	4.571	4.418
Ne	4.858	5.336	3.975	4.559	4.426	4.835	4.872	4.050	3.604	4.159	3.950
I	1.611	1.713	1.423	1.550	1.530	1.634	1.633	1.395	1.338	1.413	1.378
No. of private alleles	7.593	7.966	6.906	0.890	0.769	0.797	0.746	1.063	0.813	1.253	1.297
He	0.734	0.766	0.676	0.722	0.715	0.751	0.746	0.670	0.658	0.668	0.648
UHe	0.738	0.770	0.680	0.732	0.722	0.760	0.753	0.678	0.665	0.676	0.655

Na, no. of alleles; Ne, no. of effective alleles; I, information index; He, expected heterozygosity; UHe, unbiased heterozygosity.

Allelic Diversity by Genomic SSR Markers vs. Genic SSR Markers

The 59 polymorphic random genomic SSR markers detected 470 alleles with an average of 7.97 alleles, whereas 31 genic SSR markers detected a total of 221 alleles with an average of 6.90 alleles. The numbers of effective alleles detected were 5.34 and 3.98, respectively, by random and genic markers. While analyzing the data separately for exotic vs. local bean germplasm, it was observed that random markers detected 7.23 alleles in exotic and 7.17 alleles in local beans. The genic markers detected 6.1 alleles in exotic and 5.9 alleles in local beans (**Table 3** and **Figures 1C,D**).

The number of private and common alleles detected was also compared between the random and genic SSR markers. The random SSR markers detected 47 (0.80 average) private alleles in exotic beans vs. 44 (0.75 average) in local beans. The genic SSR markers on the other hand detected 34 (1.1 average) private alleles in exotic vs. 26 (0.82 average) in local beans (**Table 3**). The total number of common alleles between exotic and local beans detected by random markers was 379 against only 161 by genic SSR markers.

The gene diversity (*He*) detected by random SSR markers was 0.77 and that of genic SSR markers was 0.68. While comparing the same separately for exotic and local beans, it was observed that random markers detected 0.75 in exotic and 0.75 in local beans. The genic markers on the other hand detected 0.68 in exotic and 0.66 in local beans (**Table 3**).

Cluster Analysis

The clustering and construction of dendrogram based on 91 SSR markers led to the clustering/distribution of all the 96 lines into two main clusters (cluster I and cluster II). Cluster I was further divided into two sub-clusters (cluster Ia and cluster Ib). Sub-cluster Ia could be further divided into two sub-clusters, i.e., Ia.1 and Ia.2. The main cluster II could be divided into two sub-clusters, i.e., IIa and IIb. Sub-cluster IIa was further divided into two sub-clusters, i.e., IIa.1 and IIb.2 (**Figures 2A–D**). The exotic common bean lines from different countries other than India and indigenous local landraces collected from different hotspots of Jammu and Kashmir clustered together, and there was no clear-cut separation/clustering of local bean landraces from the exotic bean germplasm (**ESM Figure 1**). However, there was clear-cut clustering and assignment of Mesoamerican and Andean lines. All the Mesoamerican lines were clustered in cluster I except two lines (EC-271535 and EC-398494), which were clustered with Andean lines in sub-cluster IIb. On the other hand, all Andean gene pool lines were clustered separately in sub-cluster II (**Figure 2A** and **Supplementary Table 1**). Similar distinct clustering of the Mesoamerican and Andean gene pool lines except one line (WB1189) from the Andean gene pool got clustered with the Mesoamerican gene pool was also obtained by population assignment using GenAlEx ver 6.0 (**Figure 3**).

Structural Analysis

The structural analysis using marker data led to the identification of two (K = 2) genetically distinct subpopulations (although 2–10 subpopulations were tested) in 96 diverse bean lines (**Figure 2B**).

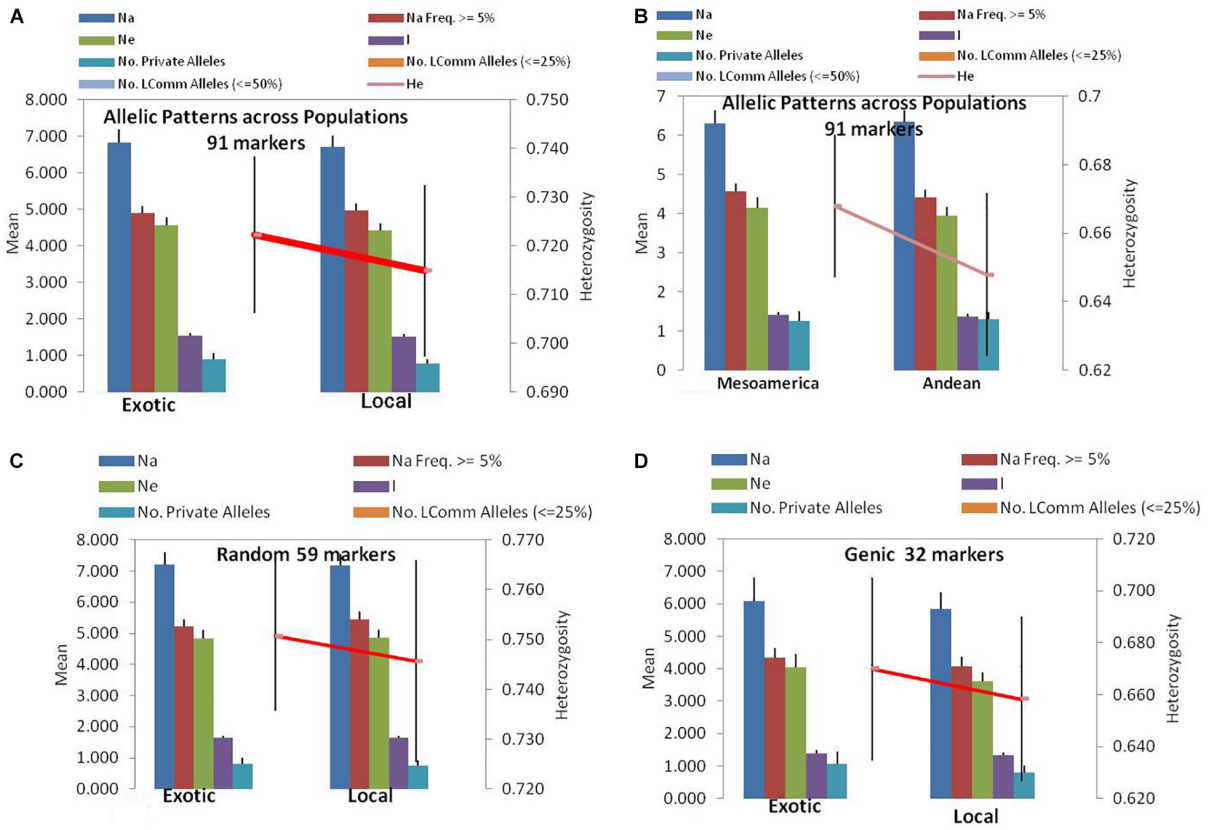


FIGURE 1 | Allelic patterns in common bean by SSR markers during the present study: **(A)** Allelic pattern in exotic and local landraces by all 91 SSR markers. **(B)** Allelic pattern in Mesoamerican and Andean populations by all 91 markers. **(C)** Allelic pattern in exotic and local landraces by only random markers. **(D)** Allelic pattern in exotic and local landraces by only genic SSR markers. The red lines indicate the trend of change in diversity from one population/group to another population/group.

Initially, based on Mean LnP(K), the number of subpopulations could not be predicted since the probability values kept on increasing steadily up to K = 4 and then decreased at K = 5 and then again started increasing up to K = 9 before again started decreasing at K = 10 (**Figure 2C** and **Supplementary Table 5**). Thus, there was no clear trend emerging about the possible number of subpopulations using LnP(K) values. Therefore, the ΔK method by Evanno et al. (2005) was used to infer the correct number of subpopulations in our population of 96 common bean lines. The ΔK method takes the rate of change of the mean probability values (LnP) of each subpopulation into consideration. As per this method, the rate of change was maximum (1,615.97) at K = 2 (**Figure 2D** and **Supplementary Table 5**); therefore, we consider two subpopulations in our sample/population of 96 common bean lines (**Figure 2B**). Both these subpopulations possess equal 48 genotypes each. Subpopulation #1 contains 25 exotic lines and 23 local lines, whereas in subpopulation #2, the number of exotic lines was 17, and the number of local lines was 31. There was no clear trend of the distribution of local (indigenous lines) vs. exotic lines in structural plot (**Figure 2**), but the distribution was largely based on gene pool/phaseolin patterns. Subpopulation #1 possesses 41 individuals from the Mesoamerican gene pool possessing "S"-type phaseolin, and the remaining seven belong to the Andean gene pool with "T"-type phaseolin. On the other hand, subpopulation #2 possesses 44 individuals from the Andean gene pool possessing "T"-type phaseolin, and the remaining four individuals belonging to the Mesoamerican gene pool with "S"-type phaseolin. Further, all the lines in these two subpopulations possess an affiliation probability of >80%, and therefore no line has been declared as admixture between two subpopulations (**Supplementary Table 1** for a structural matrix).

Average distances (expected heterozygosity) between individuals within clusters/subpopulations were also calculated using the software program STRUCTURE, and the analysis revealed that expected heterozygosity is more in the first subpopulation (0.6132) "Mesoamerican gene pool" than in the second subpopulation (0.5543) "Andean gene pool." The allele-frequency divergence among populations (net nucleotide distance), computed using point estimates of P using the software program STRUCTURE, showed a distance of 0.2119 between the two subpopulations.

Analysis of Molecular Variance

Analysis of molecular variance was conducted to test the existence of genetic structure among populations

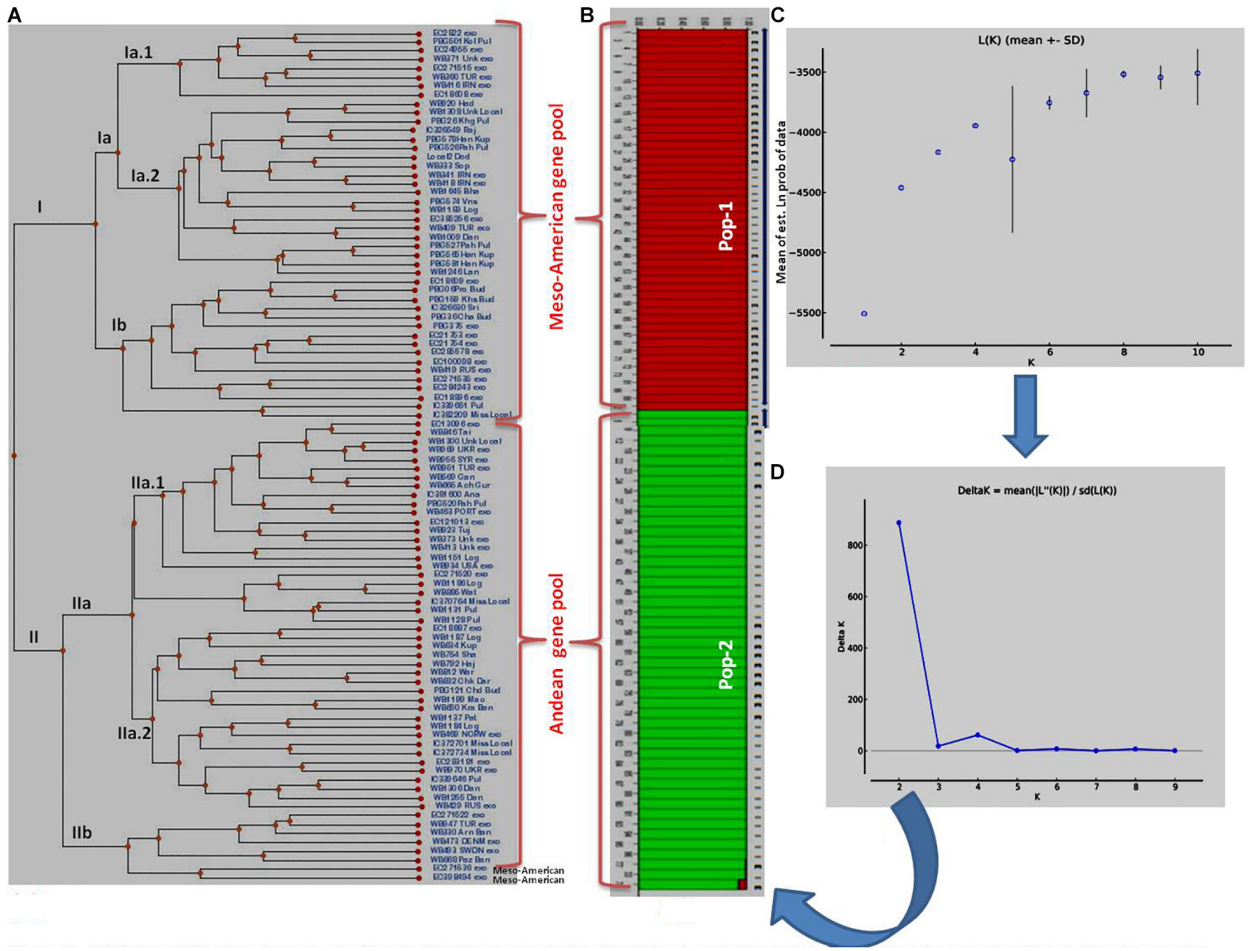


FIGURE 2 | Clustering of 96 common bean genotypes: **(A)** Traditional UPGMA hierarchical clustering where 96 genotypes have been clustered into two groups (Mesoamerican vs. Andean types). **(B)** Bayesian clustering of 96 genotypes in the form of structure plots where two sub-populations could be easily distinguished from each other. The red plots show Mesoamerican sub-population, while green plots show Andean sub-populations. **(C)** Plot of Ln (K) values of different sub-populations from 1 to 10. **(D)** Rate of change of Ln (K) from sub-population 1 to 10 based on Delta-K method.

(bean accessions from Jammu and Kashmir vs. exotic beans from different countries), as well as among and within individuals. This analysis showed that the differences among the two bean populations (indigenous vs. exotic) were significant and explained 2.0% of the total genetic variance (**Table 4**). However, for the whole population, the major source of variance was among individuals and not within individuals (97 vs. 1%), reflecting the predominant self-pollinating reproductive system of the bean.

Discovery of Important QTLs/Genes for Yield and Yield-Contributing Traits

Association mapping identified a total of 53 MTAs (on all the 11 linkage groups) for all the four traits (**Tables 5–9** and **Figures 4**, **5**). The number of significantly associated markers for an individual trait varied from 9 for 100-seed weight to 18 for yield, with an average of 13.25 MTAs/trait. However, several common markers were found to be associated with more than one trait, and therefore the total unique MTAs discovered were 39 for all the four traits (**Tables 5–9**). A set of seven markers was such that influence more than one trait, i.e., these markers influence two to four traits (**Table 9**).

For 100-seed weight, out of the nine MTAs (identified on LG02, 03, 04, 07, 08, 09, 10), four MTAs were declared stable (i.e., identified in both environments), and three MTAs were declared stable and major (i.e., identified in both environments and explaining >20% phenotypic variation for 100-seed weight). Among the nine MTAs, six MTAs were identified by both GLM and MLM, whereas three MTAs were identified by only one model, i.e., GLM (**Table 5**).

For seeds per pod, among the 14 MTAs (identified on LG01–LG09), three MTAs were declared stable, three MTAs were major, and three MTAs were declared both stable and major (i.e., identified in both environments and explaining >20% phenotypic variation for seeds per pod). Among the 14 MTAs

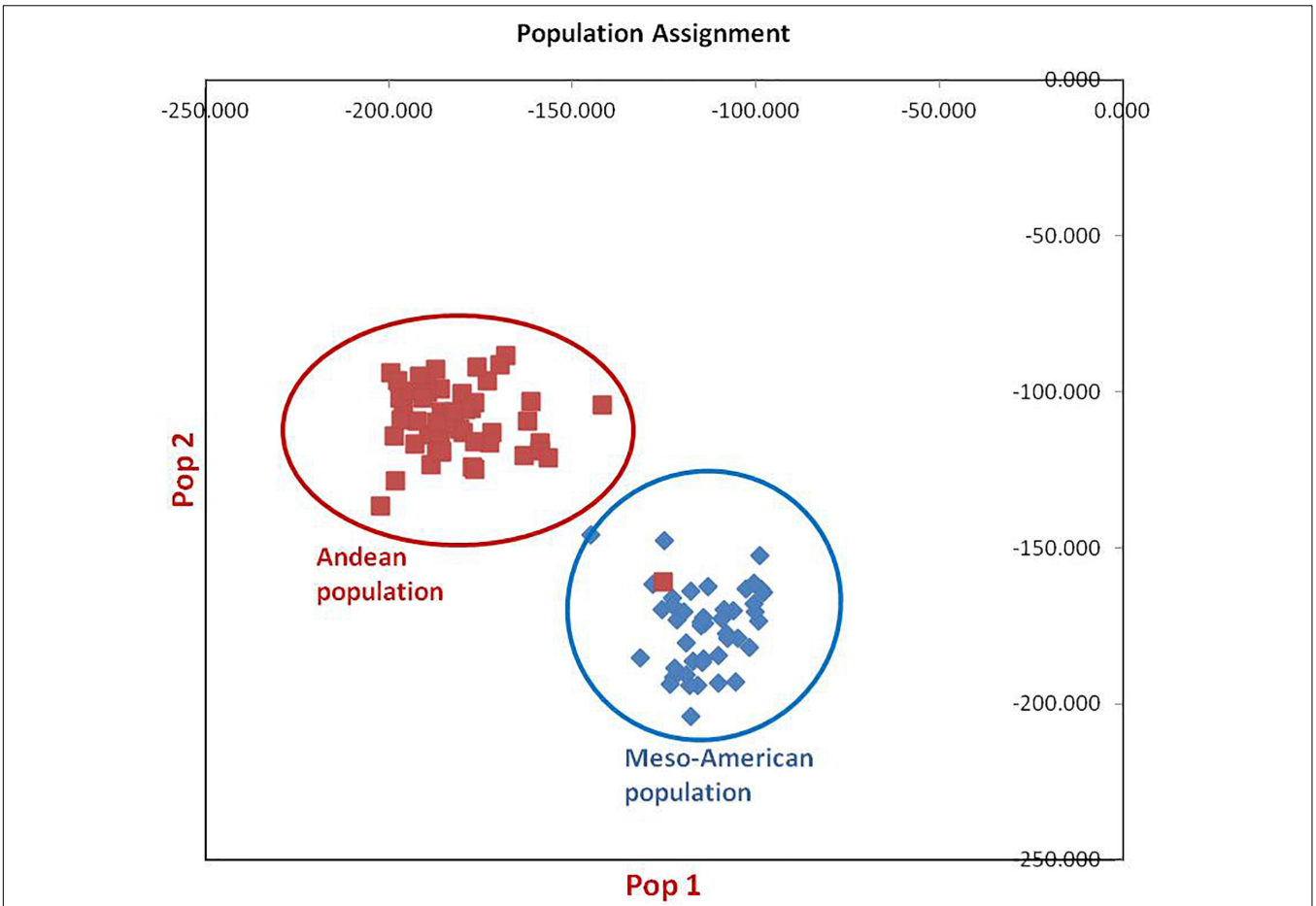


FIGURE 3 | Population assignment of 96 common bean lines into two sub-populations. Population#1 possesses all individuals of Mesoamerican gene pool except one genotype from Andean population, while population#2 possesses all genotypes of Andean gene pool.

TABLE 4 | Analysis of molecular variance (AMOVA) for the partitioning of microsatellite diversity.

Source	df	SS	MS	Est. var.	% Variation
Among pops	1	1,085,721.069	1,085,721.069	5698.423	2%
Among indiv	94	51,438,685.931	547,220.063	272,155.235	97%
Within indiv	96	279,321.000	2909.594	2909.594	1%
Total	191	52,803,728.000		280,763.252	100%

SS, sum of squares; MS, mean squares.

identified for seeds per pod, 10 MTAs were identified by both GLM and MLM, whereas four MTAs were identified by only one model, i.e., GLM (**Table 6**).

For pod per plant, among the 11 MTAs (on LG01, 03, 05, 06, 07, 08, 10, 11), two MTAs were declared stable (i.e., identified in both environments), two MTAs are major (i.e., explaining >20% phenotypic variation for pod per plant), and one MTA is declared both stable and major. Six MTAs were identified by both GLM and MLM, whereas five MTAs were identified by only one model, i.e., GLM (**Table 7**).

For yield per plant, among the 18 MTAs (identified on all the linkage groups except LG05), six MTAs were declared stable, one MTA major, and six MTAs both stable and major. Among the 18 MTAs identified for yield per plant, 12 MTAs were identified by both GLM and MLM (**Table 8**).

It is important to note that 10 MTAs for all the four traits identified during the present study have also been found to be associated with grain yield or yield-contributing traits in earlier studies. Therefore, these 10 MTAs are declared as validated MTAs (**Tables 5–9**). The validated, major, and stable MTAs are considered important and will be recommended for common bean molecular breeding programs aimed at enhancing yield and yield-contributing traits.

TABLE 5 | Marker–trait associations (MTAs) identified for 100-seed weight in two different environments using GLM and MLM approaches of the software program TASSEL.

	Environment-I (Jammu)				Environment-II (Bhaderwah)			
Marker	**Chromosome**	***P*-value**	**PVE (%)**	**Model of detection**	***P*-value**	**PVE (%)**	**Model of detection**	**Nature of MTA**
BM140	4	0.005–0.017	13.1–18.6	GLM, MLM	0.015–0.033	11.7–16.4	GLM, MLM	Stable
BM154	9	0.02–0.04	11.2–14.1	GLM, MLM	0.02	10.6	GLM	Stable and already reported by Blair and Izquierdo (2012). The marker is also one of the flanking markers for seed weight QTL "Sw9.2"
BM164	2	0.0003–0.01	23.4–30.9	GLM, MLM	0.0004–0.01	22.9–29.5	GLM, MLM	Stable and major
BM199	4	0.001–0.01	14.1–18.6	GLM, MLM	0.01	10.8	GLM	Stable
BMb96	10	0.04	6.15	GLM	0.03–0.04	6.7–9.3	GLM, MLM	Stable
BMd25	8	0.0001–0.001	16.7–24.2	GLM, MLM	0.0003–0.002	15.8–22.8	GLM, MLM	Stable and major
BMR048	4	0.0004–0.004	16.4–23.4	GLM, MLM	0.001–0.008	14.8–20.7	GLM, MLM	Stable and major
BM172	3	0.01	13.2	GLM	–	–	–	
BM160	7	0.02	14.1	GLM	–	–	–	Already reported to be associated with DM, EP, PP, SP, and SPL by Galeano et al. (2012)

GLM, general linear model; MLM, mixed linear model; PVE, phenotypic variation explained; DM, days to maturity; PP, pods per plant; SP, seed per pod; SPL, seed per plant; EP, empty pod%.

TABLE 6 | Marker–trait associations (MTAs) identified for seeds per pod in two different environments using GLM and MLM approaches of the software program TASSEL.

	Environment-I (Jammu)				Environment-II (Bhaderwah)			
Marker	**Chromosome**	***P*-value**	**PVE (%)**	**Model of detection**	**P-value**	**PVE (%)**	**Model of detection**	**Nature of MTA**
BM172	3	0.003–0.03	21.9–24.9	GLM, MLM	0.03–0.04	17.8–21.1	GLM, MLM	Stable and major
BMR244	8	0.004–0.02	28.8–29.6	GLM, MLM	0.006–0.04	25.1–27.5	GLM, MLM	Stable and major
BMR269	8	0.02–0.03	11.2–11.7	GLM, MLM	0.002–0.006	14.7–16.1	GLM, MLM	Stable
Pvest008	2	0.02–0.03	20.5–24.3	GLM, MLM	0.0008–0.004	25.4–30.0	GLM, MLM	Stable and major
BMd02	2	0.02	11.5	GLM	0.003–0.02	13.3–14.1	GLM, MLM	Stable
BMd20	5	0.04	9.8	GLM	0.01–0.03	11.1–11.3	GLM, MLM	Stable and already reported by Blair and Izquierdo (2012) linked with seed weight QTL "Sw5.1"
BM160	7	0.03	25.9	GLM	–	–	–	Major and already reported to be associated with DM, EP, PP, SP, and SPL by Galeano et al. (2012)
BM210	7	0.008	18.7	GLM	–	–	–	Already reported to be associated with yield QTL by Asfaw et al. (2012)
BMd45	1	0.006–0.009	14.3–15.3	GLM, MLM	–	–	–	
PVBR113	6	0.01–0.02	15.8–17.6	GLM, MLM	–	–	–	
PVBR94	9	0.02	19.1	GLM	–	–	–	
BM150	7	–	–	–	0.01–0.05	12. 3–12.5	GLM, MLM	
Pvest258	4	–	–	–	0.001–0.007	21.8–24.2	GLM, MLM	Major
PVBR83	3	–	–	–	0.02	22.7	GLM	Major

DM, days to maturity; PP, pods per plant; SP, seed per pod; SPL, seed per plant; EP, empty pod%.

DISCUSSION

Common bean (*P. vulgaris* L.) is one of the important grain legume crops for food and nutritional security in the world. The beans grown in the Himalayan region of Jammu and Kashmir, India possess huge diversity, and sometimes this region in India is considered as the secondary center of diversity for common bean. Common bean germplasm (landraces) grown

TABLE 7 | Marker–trait associations (MTAs) identified for pods per plant in two different environments using GLM and MLM approaches of the software program TASSEL.

	Environment-I (Jammu)				Environment-II (Bhaderwah)			
Marker	**Chromosome**	***P*-value**	**PVE (%)**	**Model of detection**	***P*-value**	**PVE (%)**	**Model of detection**	**Nature of MTA**
BMd28	5	0.009–0.02	14.9–15.9	GLM, MLM	0.04–0.05	10.8–13.3	GLM, MLM	Stable and already reported to be associated with yield QTL "Yld5.1" by Blair and Izquierdo (2012)
BMR269	8	0.05	9.99	MLM	0.03–0.04	8.83–10.77	GLM, MLM	Stable
Pvm021	11	0.02	30.2	GLM	0.03	27.1	GLM	Stable and major
BM137	6	0.03	21.9	GLM	–	–	–	Major
BM151	8	0.01–0.04	17.76–18.82	GLM, MLM	–	–	–	
BM160	7	0.01	25.27	GLM	–	–	–	Major and already reported to be associated with DM, EP, PP, SP, and SPL by Galeano et al. (2012)
BM185	7	0.01–0.02	14.1–15.4	GLM, MLM	–	–	–	
BMd45	1	0.03–0.04	9.1–10.3	GLM, MLM	–	–	–	
BM172	3	–	–	–	0.03	16.3	GLM	
BMb96	10	–	–	–	0.04	8.7	GLM	
BMd01	3	–	–	–	0.03–0.04	7.4–8.8	GLM, MLM	

DM, days to maturity; PP, pods per plant; SP, seed per pod; SPL, seed per plant; EP, empty pod%.

in this Himalayan region possess huge diversity for seed color, shape, size, and flavor (Choudhary et al., 2018b). The insight on the origin and evolution of common bean germplasm grown in this region has been discussed by us in detail in an earlier study (Choudhary et al., 2018b). The study led to the conclusion that both gene pool species of common bean, i.e., Mesoamerican and Andean beans, are grown in the state of Jammu and Kashmir with the prevalence of Mesoamerican beans in the Jammu region and both Mesoamerican and Andean beans in the Kashmir region. These findings indicated multiple introductions of this crop in the hilly state of western Himalayas by travelers from different countries in the Indian subcontinent for trading in the early part of the 16th century *via* the Red and Arabian Sea and by Chinese travelers through the Hindustan Silk Route (Choudhary et al., 2018b). However, there is hardly any report available where this huge diversity has been characterized using sophisticated genomics tools and techniques and trait phenotyping in the field. For instance, earlier studies using germplasm from this region used morphological traits only for characterization (Sofi et al., 2014; Saba et al., 2016) or utilized less reliable RAPD markers (Zargar et al., 2016). In addition, these earlier studies used a very small collection of germplasm from only few hotspot regions. These limitations have been overcome in this study by using very precise genotypic platform (ABI 3730 automatic DNA Sequencer Genotyping Platform; Applied Biosystems, Foster City, CA, United States) using a diverse bean germplasm collection that represented all (11) hotspot regions in Jammu and Kashmir. In addition, exotic bean germplasm from 11 different countries were also included in the preset study. The results of trait analyses revealed desirable values of genetic parameters in the present core set of 96 common bean genotypes. The substantial variability available may provide opportunity to favorably improve yield and related traits through selection. The elucidation of variability in the population is of paramount importance to frame an appropriate breeding strategy for seeking improvement of economically important traits. However, it is very important to mention here that yield is a very complex quantitative trait that is controlled by a network of large number of small effect minor genes/QTLs. The detection of these small effect genes/QTLs may escape detection in a small population using less number of markers. Therefore, there is a scope of using large populations/large germplasm collections with more number of markers in the future to capture more number of small effect minor genes/QTLs. Nevertheless, this study provided a promising insight for the first time into the complex genetic architecture of grain yield in different environments of western Himalayas, and findings may prove useful for common bean improvement programs worldwide.

Germplasm Characterization, Genetic Diversity, and Population Structure Analyses

The study of allelic diversity using all the 91 SSR markers on a diverse set of 96 lines revealed a very high diversity in the common bean germplasm from the state of Jammu and Kashmir, India. This is evident by the detection of up to 21 alleles by SSR marker BM187, very high average number of alleles/locus (7.59), and high average gene diversity (*He* = 0.73) (**Table 3** and **Supplementary Table 2**). The results are very encouraging and may be partly due to the precise ABI sequencing system used for SSR genotyping during the present study. The results also supported the belief that common bean germplasm being grown in north-western Himalayas is very diverse and can be used in gene discovery programs and genetic improvement of common bean. The comparison with few earlier studies revealed that the diversity in our common bean germplasm is more than the Chinese common bean germplasm (Zhang et al., 2008),

TABLE 8 | Marker–trait associations (MTAs) identified for yield per plant in two different environments using GLM and MLM approaches of the software program TASSEL.

	Environment-I (Jammu)				Environment-II (Bhaderwah)			
Marker	**Chromosome**	***P*-value**	**PVE (%)**	**Model of detection**	***P*-value**	**PVE (%)**	**Model of detection**	**Nature of MTA**
BM160	7	0.001	46.1	MLM	0.000007–0.003	41.3–43.5	GLM, MLM	Stable and major and already reported to be associated with DM, EP, PP, SP, and SPL by Galeano et al. (2012)
BM164	2	0.001	32.3	GLM	–	–	–	Major
BM172	3	0.00000006–0.001	34.5–41.5	GLM, MLM	0.000004–0.002	32.8–37.3	GLM, MLM	Stable and major
BM184	11	0.04	14.1	MLM	0.05	12.5	GLM	Stable
BM187	6	0.00002–0.03	30.9–39.5	GLM, MLM	0.0006–0.02	31.5–35.4	GLM, MLM	Stable and major
BMb96	10	0.00002–0.01	13.31–22.6	GLM, MLM	0.0001–0.01	14.9–21.6	GLM, MLM	Stable and major
BMd19	11	0.01	17.49	GLM	0.01	18.35	GLM	Stable and already reported to be associated with seed weight through SMA by Blair and Izquierdo (2012)
BMd41	11	0.002–0.02	17.66–19.7	GLM, MLM	0.004–0.02	18.9–19.8	GLM, MLM	Stable
BMd45	1	0.03–0.04	9.9–11.1	GLM, MLM	0.04–0.05	9.9–11.1	GLM, MLM	Stable
BMR269	8	0.008–0.02	10.4–15.5	GLM, MLM	0.007–0.02	11.3–16.1	GLM, MLM	Stable
Pvest006	2	0.0006–0.009	19.4–21.2	GLM, MLM	0.005–0.01	17.6–17.9	GLM, MLM	Stable
Pvest042	3	0.00000002–0.001	24.4–36.8	GLM, MLM	0.000006–0.002	23.4–31.7	GLM, MLM	Stable and major
Pvest072	6	0.04	27.1	GLM	0.02–0.04	30.8–34.2	GLM, MLM	Stable and major
PVBR251	2	0.01–0.03	16.5–17.1	GLM, MLM	–	–	–	
PVctt001	4	0.04	19.6	MLM	–	–	–	
PVBR112	4	–	–	–	0.02	15.9	GLM	
Pvest030	2	–	–	–	0.01–0.05	14.1–16.3	GLM, MLM	
BM154	9	–	–	–	0.04	14.6	GLM	Already reported to be associated with one of the flanking markers for seed weight QTL "Sw9.2" by Blair and Izquierdo (2012)

DM, days to maturity; PP, pods per plant; SP, seed per pod; SPL, seed per plant; EP, empty pod%.

TABLE 9 | List of co-localized markers/QTLs associated with more than one trait.

Marker	**Chromosome**	**Traits**			
BM154	9	100SW	Yield		
BM160	7	100SW	PPP	SPP	Yield
BM164	2	100SW	Yield		
BM172	3	100SW	PPP	SPP	Yield
BMb96	10	100SW	PPP	Yield	
BMd45	1	PPP	SPP	Yield	
BMR269	8	PPP	SPP	Yield	

PPP, pods per plant; SPP, seed per pod; SW, seed weight.

USDA common bean core collection (McClean et al., 2012), and Portuguese common bean germplasm (Leitao et al., 2017).

The high diversity of common bean from this region can also be predicted by the fact that the local landraces were almost as diverse as exotic common bean germplasm used in the present study. The local landraces got uniformly distributed along with exotic lines during cluster analysis (**ESM Figure 1**). Little difference has been noticed in allelic diversity: the local landraces possess 6.7 avg. no. of alleles/locus against 6.82 avg. no. of alleles/locus in exotic germplasm. Similarly, little difference has been noticed for the number of private alleles, the number of alleles with a frequency $\geq$ 5%, and gene diversity values between exotic and local landraces of common bean from the state of Jammu and Kashmir (**Table 3**).

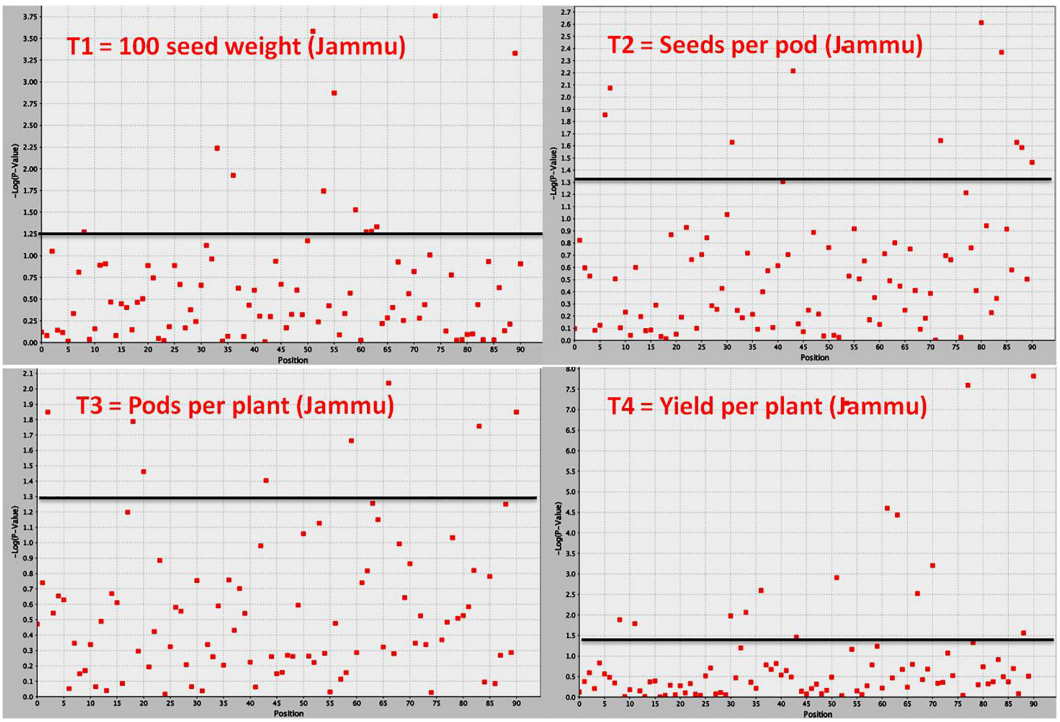


FIGURE 4 | Manhattan plot showing significant MTAs identified using software program TASSEL for yield and yield-contributing traits in common bean. The MTAs have been identified using trait data of SKUAST-Jammu location, and significant MTAs for four traits are depicted above threshold lines.

In our present study, we noticed that the common bean germplasm from the Andean gene pool possess more diversity than the germplasm from the Mesoamerican gene pool. For instance, more total number of alleles and average number of alleles were detected in common bean germplasm belonging to the Andean gene pool than the germplasm belonging to the Mesoamerican gene pool (**Table 3**). However, in earlier studies, an opposite trend, i.e., more number of alleles using genic and genomic SSR markers, has been shown in Mesoamerican beans than in Andean beans (Zhang et al., 2008). These results obtained during the present study may be partly due to more number of private alleles detected in Andean beans (1.29) than in Mesoamerican beans (1.25). The greater diversity in Andean beans than in Mesoamerican beans is considered a feature of SSR marker analysis, and these results got support from some earlier studies (Blair et al., 2006; Zhang et al., 2008; McClean et al., 2012). The gene diversity trends showed that Mesoamerican beans are more diverse (0.67) than Andean beans (0.65). Similar results have been reported earlier as well using isozymes (Koenig and Gepts, 1989), RFLP (Becerra-Velasquez and Gepts, 1994), RAPD (Beebe et al., 2000, 2001), AFLP (Tohme et al., 1996), and DNA sequence data (McClean et al., 2004; McClean and Lee, 2007).

We also observed that the genic markers reveal less diversity than random SSR markers, as has been reported in several earlier studies (Blair et al., 2006; Zhang et al., 2008). However, the diversity revealed by genic markers reflects true diversity of a crop species.

The diverse nature of germplasm collection used during the present study was also evident by the fact that all the 96 lines were clustered uniformly and do not form any specific cluster for local landraces and exotic lines (**ESM Figure 1**). On the other hand, there was clear-cut assignment/clustering of lines based on their phaseolin patterns with the clustering of Andean types separately from the Mesoamerican types in both traditional hierarchical clustering and Bayesian clustering through structural analysis (**Figures 2A,B**). Similar results (only two subpopulations) have also been reported in population structural analysis in earlier studies (Zhang et al., 2008; Leitao et al., 2017), and the two subpopulations corresponded to the Andean and Mesoamerican gene pools (Zhang et al., 2008; Kwak and Gepts, 2009; McClean et al., 2012). The presence of only two subpopulations in the Himalayan beans is typical to most legume crops due to the self-pollinating nature of the legume crops. In summary, both distance and model-based approaches classified our common bean collection into two major subpopulations, and these results are consistent with previous results that recognized two major subdivisions within the cultivated common bean (Gepts et al., 1986; Singh et al., 1991; Becerra-Velasquez and Gepts, 1994; Kwak and Gepts, 2009; McClean et al., 2012).

The information of structure will be useful to avoid spurious association in the study of MTAs through GWAS. The results of structural analysis and UPGMA clustering are in agreement since in both the clustering types, two distinct groups were formed

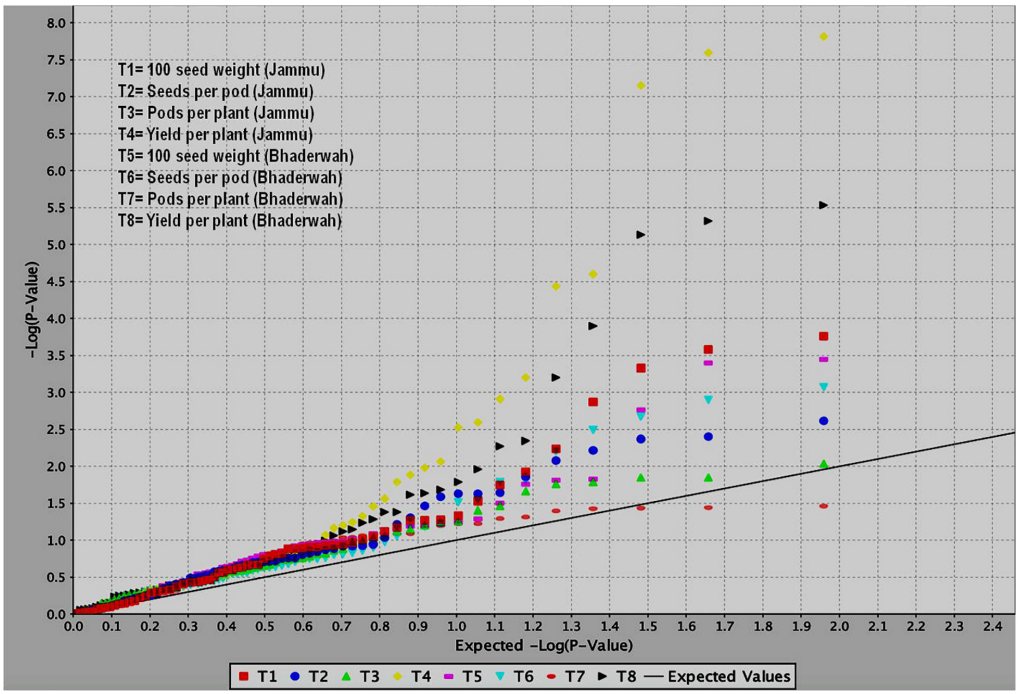


FIGURE 5 | QQ plots obtained during study of marker–trait associations for yield and yield-contributing traits in common bean. The figure shows QQ plots for all the four traits in two different environments (T1–T4 at SKUAST Jammu and T5–T8 at Bhaderwah Jammu).

based on two different gene pools, i.e., Mesoamerican vs. Andean gene pools (Leitao et al., 2017).

Gene Discovery for Yield and Yield-Contributing Traits

In common bean, significant and positive correlations were observed between yield and its component traits including 100-seed weight, pods per plant, and seeds per pod during the present study and in some earlier studies as well (Beebe et al., 2013; Assefa et al., 2015, 2019; Rao et al., 2017). Therefore, yield components could be used as selection criteria for the improvement of yield and the development of next-generation common bean cultivars. In fact, it is well documented that an increase in yield in common bean under favorable environmental conditions has come from improvement in pods per plant, seed per plant, and 100-seed weight (Beebe et al., 2013; for review, see Assefa et al., 2019).

During the present study, a set of 39 significantly associated markers/genes on all the 11 chromosomes has been identified for all the four traits. This includes 15 major MTAs, 15 stable MTAs, and 13 both major and stable MTAs. One of the most important breakthroughs achieved during the present study is the validation of a set of 10 MTAs already identified in earlier studies. Some of the validated markers found correspondence to some important QTLs for yield and yield-contributing traits (**Tables 5–9**). For instance, SSR marker "BM154" associated with 100-seed weight and yield on chromosome 9 has also been reported in an earlier study by Blair and Izquierdo (2012) for seed weight. The marker "BM154" is one of the associated flanking markers for the seed weight QTL "Sw9.2." The marker "BMd20" found to be associated with trait "seeds per pod" during the present study has been earlier identified and found to be linked with seed weight QTL "Sw5.1" (Blair and Izquierdo, 2012). The important marker "BM160" found to be associated with all the four traits (pods per plant, seeds per pod, 100-seed weight, and yield per plant) during the present study has also been found to be associated with a variety of yield-related traits (days to maturity, pods per plant, seed per pod, seed per plant, and empty pod%) in an earlier study (Galeano et al., 2012). The marker "BM210" identified to be associated with seeds per pod during the present study has been found to be associated with yield by Asfaw et al. (2012). Stable QTL-linked marker "BMd28" identified during the present study for "pods per plant" has been already reported to be associated with yield QTL "Yld5.1" by Blair and Izquierdo (2012). Similarly, marker "BMd19" found to be associated with yield per plant during the present study has been found to be associated with seed weight through single marker analysis (Blair and Izquierdo, 2012). The major, stable, and validated MTAs for yield and yield-contributing traits may be used in common bean breeding programs aimed at enhancing yield of common bean.

In common bean, different trait mapping studies have been already conducted using both bi-parental mapping populations and more recent GWASs involving diverse germplasm collections (for review, see González et al., 2018; Assefa et al., 2019). In these earlier studies, several genes with minor effects involved in the genetic control of seed size, pod size, and yield have been identified repeatedly in different genetic backgrounds with increasingly tight genetic bounds (González et al., 2018;

Assefa et al., 2019). For instance, genes for pod size and pod length have been identified in some earlier studies at similar locations on LG01, LG02, and LG04 (Koinange et al., 1996; Yuste-Lisbona et al., 2014; Hagerty et al., 2016). In another study using single-point analysis, a set of 10-positive markers was found to be associated with yield on linkage groups b01, b02, b03, b04, and b09, and 21 markers were found to be associated with seed size (Blair and Izquierdo, 2012). Using composite interval mapping, nine markers were identified for seed weight across four linkage groups (b02, b03, b05, and b09), and one QTL was detected for yield on linkage group b05 (Blair and Izquierdo, 2012). Significant MTAs have also been identified for other yield components including pods per plant (PP), seed per pod (SP), and seed per plant (SPL) through association mapping (Galeano et al., 2012). A number of common bean genes/QTLs for yield and yield-contributing traits have been projected on all the 11 linkage groups except linkage group 01 (LG01) of the consensus reference genetic map developed from genetic maps of three populations[5]. The total number of QTLs for yield and yield-contributing traits projected on 10 linkage groups (LG02 to LG11) is 85 and varies from three QTLs (LG05 and LG11) to 21 QTLs (LG06) with an average of 8.5 QTLs/linkage group. The co-localized markers that influence more than one trait will prove useful in the simultaneous improvement of multiple traits in common bean. The markers BM160 and BM172 that influence all the four traits (pods per plant, seeds per pod, 100- seed weight, and yield per plant) are considered most important markers for breeding programs aimed at enhancing grain yields in common bean.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

[5] https://legumeinfo.org/traits_maps/cmap/?ref_map_set_acc=PvConsensus_GaleanoFernandez2011_a&ref_map_accs=1

AUTHOR CONTRIBUTIONS

RM conceived the idea, conducted the experiments, and wrote the manuscript. NC, VB, and SJ performed the experiments. BS helped in collection and analysis of field data. AK and MAB helped in manuscript writing and analysis of data. RP helped in analysis of data for GWAS. MT, AC, and RV helped in generation of genotypic data and data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

The authors are thankful to the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India, for providing funds under the Young Scientist Award (Grant ID No: SB/FT/LS-355/2012) to RM. NC is also thankful to the SERB, DST, Government of India, for providing their fellowship.

ACKNOWLEDGMENTS

The authors are highly thankful to the Dean, Faculty of Agriculture, SKUAST-Kashmir for providing different facilities during this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.609603/full#supplementary-material

REFERENCES

Angioi, S. A., Rau, D., Attene, G., Nanni, L., Bellucci, E., Logozzo, G., et al. (2010). Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. *Theor. Appl. Genet.* 121, 829–843.

Asfaw, A., Blair, M. W., and Almekinders, C. (2009). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor. Appl. Genet.* 120, 1–12. doi: 10.1007/s00122-009-1154-7

Asfaw, A., Blair, M. W.,and Struik, P. C. (2012). Multi-environment quantitative trait loci analysis for photosynthate acquisition, accumulation, and remobilization traits in common bean under drought stress. *G3: Genes, Genomes, Genetics.* 2, 579–595. doi: 10.1534/g3.112.002303

Assefa, T., Mahama, A. A., Brown, A. V., Cannon, E. K. S., Rubyogo, J. C., Rao, I. M., et al. (2019). A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 39:20. doi: 10.1007/s11032-018-0920-0

Assefa, T., Wu, J., Beebe, S. E., Rao, M. I., Marcomin, D., and Rubyogo, J. C. (2015). Improving adaptation to drought stress in small red common bean: phenotypic differences and predicted genotypic effects on grain yield, yield components and harvest index. *Euphytica* 203, 477–489. doi: 10.1007/s10681-014-1242-x

Beebe, S., Skroch, P. W., Tohme, J., Duque, M. C., Pedraza, F., and Nienhuis, J. (2000). Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci.* 40, 264–273.

Beebe, S. E., Rao, I. M., Blair, M. W., and Acosta-Gallegos, J. A. (2013). Phenotyping common beans for adaptation to drought. *Front. Physiol.* 4:35. doi: 10.3389/fphys.2013.00035

Beebe, S. E., Rengifo, J., Gaita′n-Solis, E., Duque, M. C., and Tohme, J. (2001). Diversity and origin of Andean landraces of common bean. *Crop Sci.* 41, 854–862. doi: 10.2135/cropsci2001.413854x

Benchimol, L. L., de Campos, T., Carbonell, S. A. M., Colombo, C. A., Chioratto, A. F., Formighieri, E. F., et al. (2007). Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. *Genet. Resour. Crop Evol.* 54, 1747–1762. doi: 10.1007/s10722-006-9184-3

Blair, M. W., Astudillo, C., Grusak, M. A., Graham, R., and Beebe, S. E. (2009). Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 23, 197–207. doi: 10.1007/s11032-008-9225-z

Blair, M. W., Giraldo, M. C., Buendía, H. F., Tovar, E., Duque, M. C., and Beebe, S. (2006). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 113, 100–109.

Blair, M. W., and Izquierdo, P. (2012). Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *Theor. Appl. Genet.* 125, 1015–1031. doi: 10.1007/s00122-012-1891-x

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

Broughton, W. J., Hernandez, G., Blair, M. W., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.)-model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611

Buah, S., Buruchara, R., and Okori, P. (2017). Molecular characterisation of common bean (*Phaseolus vulgaris* L.) accessions from Southwestern Uganda reveal high levels of genetic diversity. *Genet. Resour. Crop Evol.* 64, 1985–1998. doi: 10.1007/s10722-017-0490-8

Bukhari, A., Bhat, M. A., Ahmad, M., and Saleem, N. (2015). Examination of genetic diversity in common bean (*Phaseolus vulgaris* L.) using random amplified polymorphic DNA (RAPD) markers. *Afr. J. Biotechnol.* 14, 451–458. doi: 10.5897/AJB2014.14281

Choudhary, N., Bawa, V., Paliwal, R., Singh, B., Bhat, M. A., Mir, J. I., et al. (2018a). Gene/QTL discovery for Anthracnose in common bean (*Phaseolus vulgaris* L.) from North-western Himalayas. *PLoS One* 13:e0191700. doi: 10.1371/journal.pone.0191700

Choudhary, N., Hamid, A., Singh, B., Khandy, I., Sofi, P. A., Bhat, M. A., et al. (2018b). Insight into the origin of common bean (*Phaseolus vulgaris* L.) grown in the state of Jammu and Kashmir of north-western Himalayas. *Genet. Resour. Crop Evol.* 65, 963–977. doi: 10.1007/s10722-017-0588-z

Dıaz, L. M., and Blair, M. W. (2006). Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theor. Appl. Genet.* 114, 143–154. doi: 10.1007/s00122-006-0417-9

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the softwares structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Fisseha, Z., Tesfaye, K., Dagne, K., Blair, M. W., Harvey, J., Kyallo, M., et al. (2016). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L) germplasm of Ethiopia as revealed by microsatellite markers. *Afr. J. Biotechnol.* 15, 2824–2847.

Galeano, C. H., Cortés, A. J., Fernández, A. C., Soler, A., Franco-Herrera, N., Makunde, G., et al. (2012). Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. *BMC Genet.* 13:48. doi: 10.1186/1471-2156-13-48

Gascuel, O. (1997). "Concerning the NJ algorithm and its unweighted version, UNJ," in *Mathematical Hierarchies and Biology. DIMACS Workshop, Series in Discrete Mathematics and Theoretical Computer Science*, eds B. Mirkin, F. R. McMorris, F. Roberts, and A. Rzhetsky (Providence, RI: American Mathematical Society), 149–170. doi: 10.1090/dimacs/037/09

Gepts, P., Osborn, T. C., Rashka, K., and Bliss, F. A. (1986). Phaseolin protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. *Econ. Bot.* 40, 451–468. doi: 10.1007/BF02859659

González, A. M., Yuste-Lisbona, F. J., Fernández-Lozano, A., Lozano, R., and Santalla, M. (2018). "Genetic mapping and QTL analysis in common bean," in *The Common Bean Genome*, eds M. Pérez de la Vega, M. Santalla, and F. Marsolais (Cham: Springer), 69–107.

Gupta, P. K., Varshney, R. K., and Prasad, M. (2002). "Molecular markers: principles and methodology," in *Molecular Techniques in Crop Improvement*, eds M. Jain, D. S. Brar, and B. S. Ahloowalia (Cham: Springer), 9–54.

Hagerty, C. H., Cuesta-Marcos, A., Cregan, P., Song, Q., McClean, P., and Myers, J. R. (2016). Mapping snap bean pod and color traits, in a dry bean × snap bean recombinant inbred population. *J. Am. Soc. Hortic. Sci.* 141, 131–138. doi: 10.21273/JASHS.141.2.131

Khaidizar, M. I., Haliloglu, K., Elkoca, E., Aydin, M., and Kantar, F. (2012). Genetic diversity of common bean (*Phaseolus vulgaris* L.) landraces grown in northeast Anatolia of Turkey assessed with simple sequence repeat markers. *Turk. J. Field Crops* 17, 145–150.

Koenig, R., and Gepts, P. (1989). Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. *Theor. Appl. Genet.* 78, 809–817. doi: 10.1007/BF00266663

Koinange, E., Singh, S., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x

Kwak, M., and Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* 118, 979–992. doi: 10.1007/s00122-008-0955-4

Leitao, S. T., Dinis, M., Veloso, M. M., Šatovic, Z., and Vaz Patto, M. C. (2017). Establishing the bases for introducing the unexplored portuguese common bean germplasm into the breeding world. *Front. Plant Sci.* 8:1296. doi: 10.3389/fpls.2017.01296

McClean, P. E., and Lee, R. K. (2007). Genetic architecture of chalcone isomerase non-coding regions in common bean (*Phaseolus vulgaris* L.). *Genome* 50, 203–214. doi: 10.1139/g07-001

McClean, P. E., Lee, R. K., and Miklas, P. N. (2004). Sequence diversity analysis of dihydro-flavonol 4-reductase intron 1 in common bean. *Genome* 47, 266–280. doi: 10.1139/g03-103

McClean, P. E., Terpstra, J., McConnell, M., White, C., Lee, R., and Mamidi, S. (2012). Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. *Genet. Resour. Crop Evol.* 59, 499–515. doi: 10.1007/s10722-011-9699-0

Metais, I., Hamon, B., Jalouzet, R., and Peltier, D. (2002). Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. *Theor. Appl. Genet.* 104, 1346–1352. doi: 10.1007/s00122-002-0901-9

Mir, R. R., Hiremath, P. J., Riera-Lizarazu, O., and Varshney, R. K. (2013). "Evolving molecular marker technologies in plants: from RFLPs to GBS," in *Diagnostics in Plant Breeding*, eds T. Lübberstedt and R. K. Varshney (New York, NY: Springer), 229–247. doi: 10.1007/978-94-007-5687-8_11

Mir, R. R., Kumar, J., Balyan, H. S., and Gupta, P. K. (2012a). A study of genetic diversity among Indian bread wheat (*Triticum aestivum* L.) cultivars released during last 100 years. *Genet. Resour. Crop Evol.* 59, 717–726. doi: 10.1007/s10722-011-9713-6

Mir, R. R., Kumar, N., Jaiswal, V., Girdharwal, N., Prasad, M., Balyan, H. S., et al. (2012b). Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. *Mol. Breed.* 29, 963–972.

Mir, R. R., and Varshney, R. K. (2013). "Future prospects of molecular markers in plants," in *Molecular Markers in Plants*, ed. R. J. Henry (Oxford: Blackwell Publishing Ltd), 169–190. doi: 10.1002/9781118473023.ch10

Okii, D., Tukamuhabwa, P., Kami, J., Namayanja, A., Paparu, P., Ugen, M., et al. (2014). The genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) germplasm in Uganda. *Afr. J. Biotechnol.* 13, 2935–2949.

Pallottini, L., Garcia, E., Kami, J., Barcaccia, G., and Gepts, P. (2004). The genetic anatomy of a patented yellow bean. *Crop Sci.* 44, 968–977. doi: 10.2135/cropsci2004.9680

Peakall, R. O. D., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Pejic, I., Ajmone-Marsan, P., Morgante, M., Kozumplick, V., Castiglioni, P., Taramino, G., et al. (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. *Theor. Appl. Genet.* 97, 1248–1255. doi: 10.1007/s001220051017

Perrier, X., Flori, A., and Bonnot, F. (2003). "Methods for data analysis," in *Genetic Diversity of Cultivated Tropical Plants*, eds P. Hamon, M. Seguin, X. Perrier, and JC Glazmann (Montpellier: Science Publishers Inc and Cirad), 31–63.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2, 225–238.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multi-locus genotype data. *Genetics* 155, 945–959.

Rana, J. C., Sharma, T. R., Tyagi, R. K., Chahota, R. K., Gautam, N. K., Singh, M., et al. (2015). Characterization of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. *Euphytica* 205, 441–457. doi: 10.1007/s10681-015-1406-3

Rao, I. M., Beebe, S. E., Polania, J., Grajales, M., Cajiao, C., Ricaurte, J., et al. (2017). Evidence for genotypic differences among elite lines of common bean in their ability to remobilize photosynthate to increase yield under drought. *J. Agric. Sci.* 155, 857–875. doi: 10.1017/S0021859616000915

Razvi, S. M., Khan, M. N., Bhat, M. A., Ahmad, M., Khan, M. H., Ganie, S. A., et al. (2013). Genetic diversity studies in common bean (*Phaseolus vulgaris* L.) using molecular markers. *Afr. J. Biotechnol.* 12, 7031–7037.

Saba, I., Sofi, P. A., Zeerak, N. A., Bhat, M. A., and Mir, R. R. (2016). Characterisation of a core set of common bean (*Phaseolus vulgaris* L.) germplasm for seed quality traits. *SABRAO J. Breed. Genet.* 48, 359–376.

Sharma, P., Díaz, L., and Blair, M. (2013). Genetic diversity of two Indian common bean germplasm collections based on morphological and microsatellite markers. *Plant Genet. Resour.* 11, 121–130.

Singh, S. P., Nodari, R., and Gepts, P. (1991). Genetic diversity in cultivated common bean: I. *Allozymes Crop Sci.* 31, 19–23. doi: 10.2135/cropsci1991.0011183X003100010004x

Sofi, P. A., Rana, J. C., and Bhat, N. A. (2014). Pattern of variation in common bean (*Phaseolus vulgaris* L.) genetic resources of Jammu and Kashmir. *J. Food Legum.* 27, 197–201.

Tohme, J., Gonzalez, D. O., Beebe, S., and Duque, M. C. (1996). AFLP analysis of gene pools of a wild bean core collection. *Crop Sci.* 36, 1375–1384. doi: 10.2135/cropsci1996.0011183X003600050048x

Becerra-Velasquez, V. L., and Gepts, P. (1994). RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. *Genome* 37, 256–263. doi: 10.1139/g94-036

Yuste-Lisbona, F., González, A., Capel, C., García-Alcázar, M., Capel, J., De Ron, A. M., et al. (2014). Genetic variation underlying pod size and color traits of common bean depends on quantitative trait loci with epistatic effects. *Mol. Breed.* 33, 939–952. doi: 10.1007/s11032-013-0008-9

Zargar, S. M., Farhat, S., Mahajan, R., Bhakhri, A., and Sharma, A. (2016). Unravelling the efficiency of RAPD and SSR markers in diversity analysis and population structure estimation in common bean. *Saudi J. Biol. Sci.* 23, 139–149. doi: 10.1016/j.sjbs.2014.11.011

Zhang, X., Blair, M. W., and Wang, S. (2008). Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeat markers. *Theor. Appl. Genet.* 117, 629–640. doi: 10.1007/s00122-008-0807-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mir, Choudhary, Bawa, Jan, Singh, Bhat, Paliwal, Kumar, Chitikineni, Thudi and Varshney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 January 2021
doi: 10.3389/fgene.2020.548407

Mapping and Validation of Major Quantitative Trait Loci for Resistance to Northern Corn Leaf Blight Along With the Determination of the Relationship Between Resistances to Multiple Foliar Pathogens of Maize (*Zea mays* L.)

Hosahally Muddrangappa Ranganatha[1], Hirenallur Chandappa Lohithaswa[1] and Anand Pandravada[2]*

[1] Department of Genetics and Plant Breeding, College of Agriculture – Mandya, University of Agricultural Sciences, Bengaluru, Bengaluru, India, [2] Corteva Agriscience Pvt. Ltd., Kallinayakanahalli, India

OPEN ACCESS

Edited by:
Deepmala Sehgal, International Maize and Wheat Improvement Center, Mexico

Reviewed by:
Kumari Neelam, Punjab Agricultural University, India
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, India

***Correspondence:**
Hirenallur Chandappa Lohithaswa
lohithaswa.chandappa@gmail.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *02 April 2020*
Accepted: *14 December 2020*
Published: *29 January 2021*

Citation:
Ranganatha HM, Lohithaswa HC and Pandravada A (2021) Mapping and Validation of Major Quantitative Trait Loci for Resistance to Northern Corn Leaf Blight Along With the Determination of the Relationship Between Resistances to Multiple Foliar Pathogens of Maize (Zea mays L.). Front. Genet. 11:548407. doi: 10.3389/fgene.2020.548407

Among various foliar diseases affecting maize yields worldwide, northern corn leaf blight (NCLB) is economically important. The genetics of resistance was worked out to be quantitative in nature thereby suggesting the need for the detection of quantitative trait loci (QTL) to initiate effective marker-aided breeding strategies. From the cross CML153 (susceptible) × SKV50 (resistant), 344 $F_{2:3}$ progenies were derived and screened for their reaction to NCLB during the rainy season of 2013 and 2014. The identification of QTL affecting resistance to NCLB was carried out using the genetic linkage map constructed with 194 polymorphic SNPs and the disease data recorded on $F_{2:3}$ progeny families. Three QTL for NCLB resistance were detected on chromosomes 2, 5, and 8 with the QTL *qNCLB-8-2* explaining the highest phenotypic variation of 16.34% followed by *qNCLB-5* with 10.24%. QTL for resistance to sorghum downy mildew (SDM) and southern corn rust (SCR) were also identified from one season phenotypic data, and the co-location of QTL for resistance to three foliar diseases was investigated. QTL present in chromosome bins 8.03, 5.03, 5.04, and 3.04 for resistance to NCLB, SDM, and SCR were co-localized, indicating their usefulness for the pyramiding of quantitative resistance to multiple foliar pathogens. Marker-assisted selection was practiced in the crosses CM212 × SKV50, HKI162 × SKV50, and CML153 × SKV50 employing markers linked to major QTL on chromosomes 8, 2, and 10 for NCLB, SDM, and SCR resistance, respectively. The populations were advanced to F_6 stage to derive multiple disease-resistant inbred lines. Out of the 125 lines developed, 77 lines were tested for their combining ability and 39 inbred lines exhibited high general combining ability with an acceptable level of resistance to major diseases.

Keywords: northern corn leaf blight, sorghum downy mildew, southern corn rust, linkage map, QTL, marker assisted selection

INTRODUCTION

Maize (*Zea mays* L.) is a widely cultivated food crop worldwide along with rice and wheat. It also serves as a livestock feed and industrial raw material (Troyer, 2006). Maize was diversified first in the highlands of Mexico where it was domesticated from the wild progenitor teosinte, *Z. mays* spp. *parviglumis* (Matsuoka et al., 2002). Globally, maize (*Z. mays* L.) is cultivated in a wide variety of environments with major cultivation in the warmer parts of temperate regions and in humid–subtropical climate (Dowswell et al., 1996).

Among various biotic constraints, foliar diseases are very important yield-limiting factors worldwide and the prevalence of these diseases varies depending on the region or season (Smith, 1999). About 61 diseases have been recorded on maize in India causing yield losses (Payak et al., 1973; Payak and Sharma, 1985). Northern corn leaf blight incited by *Exserohilum turcicum* (Pass) Leonard and Suggs (Teliomorph = *Setosphaeria turcica* (Luttrell), sorghum downy mildew caused by *Peronosclerospora sorghi* (Weston and Uppal), and southern corn rust caused by *Puccinia polysora* (Underwood) are considered as the most persistent and destructive diseases of field maize (Pratt and Gordon, 2006).

Northern corn leaf blight is prevalent throughout the world and is known to cause more than 50% yield losses (Raymundo and Hooker, 1981; Perkins and Pederson, 1987). Disease is known to appear in regions whenever moderate temperatures and high humidity prevail (Carson, 1999; Smith, 1999). Northern corn leaf blight disease was first reported in India by Butler (1907), and it causes 16–98% reduction in grain and fodder yield (Kachapur and Hegde, 1988; Harlapur et al., 2000).

Among several management options available, cultivation of resistant cultivars is the most practical and cost-effective approach in the management of diseases (Fehr, 1987; Ward et al., 1997). To breed a genotype with a high level of resistance, the inheritance pattern of resistant reaction in the material being handled is a prerequisite. Earlier studies on the genetics of resistance to northern corn leaf blight (Jenkins et al., 1952; Hughes and Hooker, 1971; Hettiarachchi et al., 2009; Chaudhary and Mani, 2010; Ranganatha et al., 2017) suggest that resistance is complex and polygenic in nature. Selection for resistance to foliar diseases is effectively practiced through conventional breeding, where susceptible genotypes under disease pressure can be eliminated before harvest (Ali and Yan, 2012). However, conventional breeding is time-consuming and less feasible due to the complex nature of resistance reaction. This favored the development of molecular tools to assist conventional breeding efforts to breed resistant cultivars. Identification of quantitative trait loci (QTL) is one such tool to help in marker-assisted selection (MAS) of resistant genotypes. In maize, a large amount of valuable information exists with reference to QTL conditioning resistance to foliar diseases like northern corn leaf blight (Welz et al., 1999; Welz and Geiger, 2000; Brown et al., 2001; Ping et al., 2007; Asea et al., 2009; Balint-Kurti et al., 2010; Chung et al., 2010, 2011; Zwonitzer et al., 2010; Poland et al., 2011; Schaefer and Bernardo, 2013; Chen et al., 2016; Shridhar Hegde et al., 2018; Xia et al., 2020). Resistance to multiple diseases conditioned by the same locus is an important consideration in breeding durable resistant genotypes. The information regarding studies on the application of markers linked to QTL for pyramiding quantitative resistance to multiple foliar pathogens in maize is less. Hence, an attempt was made to identify QTL for resistance to northern corn leaf blight (NCLB), to investigate the co-location of QTL for resistance to NCLB with those for sorghum downy mildew (SDM) and southern corn rust (SCR) diseases detected using the same mapping population, and also to develop multiple disease-resistant inbred lines pyramided with major QTL for these three foliar diseases.

METHODOLOGY

Development of Mapping Population

Based on previous experimental data, two inbred lines, viz. CML153 (susceptible inbred, P_1) and SKV50 (resistant inbred, P_2), with contrasting disease reaction against northern corn leaf blight were selected for the development of the mapping population. The selected inbreds were crossed during the rainy season of 2012. The F_1 generation (CML153 × SKV50) was grown during winter 2012–2013 and self-pollinated. The resulting F_2 individuals were planted during summer 2013 and selfed to derive 344 $F_{2:3}$ families. In each $F_{2:3}$ family, the leaves from five randomly selected plants were collected for genotypic analysis. Later, during the rainy season of 2013 and 2014, $F_{2:3}$ families were screened for disease reaction against northern corn leaf blight in the national disease nursery maintained at ZARS, V.C. Farm, Mandya.

Genotyping of $F_{2:3}$ Mapping Population

A set of 768 single nucleotide polymorphisms (SNPs) covering whole maize genome[1] was used for the genotyping of parents, and 199 polymorphic SNPs were identified. The 344 $F_{2:3}$ progenies were genotyped with these polymorphic markers. Leaf samples were pooled from five random plants of each $F_{2:3}$ family and parents and lyophilized in 96-well plates. Samples were loaded to the Illumina BeadXpress Vera Code Reader for genotyping, according to Illumina protocols[2]. The polymorphism detected by SNP marker was scored as A = homozygous maternal genotype (CML153), B = homozygous paternal genotype (SKV50), H = heterozygote genotype, and - = missing samples.

Phenotyping of $F_{2:3}$ Mapping Population

The 344 $F_{2:3}$ families along with the two parental lines were screened against *E. turcicum*, causing northern corn leaf blight of maize. The disease screening was conducted in two seasons during the rainy season of 2013 and 2014 in the national disease screening nursery maintained at Mandya. The experimental area was divided into blocks of 3 m width and 32 m length each accommodating 42 progenies. The $F_{2:3}$ progenies were planted in a single row of 3 m length employing the randomized complete

[1] www.panzea.com

[2] http://www.illumina.com/

block design with two replications. The spacing of 75 cm between rows and 20 cm between plants was provided.

Screening for Northern Corn Leaf Blight (*E. turcicum*)

The susceptible inbred CM202 was planted as the first row and the last row in each block and also after every 10th progeny. The artificial inoculation procedure developed by Shekhar and Kumar (2012) was employed to ensure uniform disease development. The initial inoculum for artificial inoculation of *E. turcicum* was grown in artificial medium under laboratory conditions. The infected leaf tissues were collected from the diseased plants under natural field conditions, washed thrice with sterile water, cultured on potato dextrose agar medium, and then multiplied on sorghum seeds. For this, the sorghum seeds were soaked overnight and transferred to sterilized conical flasks the next day, and the pathogen inoculum was added. The flasks were shaken once in 2 days, and an equal amount of fresh sorghum seeds was mixed after 1 week. The sorghum seeds with pathogen inoculum were ground to a fine powder, and 1–1.5 g of the ground inoculum was added to the leaf whorl of the test entries, followed by a light spray of water to create humidity and initiate infection. Artificial inoculation was made 20 days after sowing between 3:00 and 6:00 p.m., and inoculation was repeated twice after a 1-week interval. The northern corn leaf blight severity was recorded at the flowering stage, i.e., 60th day after sowing by visualizing the leaf area using a standard scale consisting of five broad categories with intermediate ratings (Payak and Sharma, 1983). Based on the disease score, progenies were classified as resistant (<2.5), moderately resistant (2.5 to <3.0), moderately susceptible (3.0 to <3.5), susceptible (3.5 to <4.0), and highly susceptible (4.0–5.0). The disease score data were converted into percent disease severity by using the formula given by Wheeler (1969).

Statistical Analysis

Phenotypic Data Analysis

Transformation of field data

The disease data recorded as percent disease index for northern leaf blight infection ranged from 0 to 100. To make the means and variances independent and normally distributed, the percentage data were subjected to arcsine transformation (Little and Hills, 1978). The analysis of variance was performed on transformed phenotypic data using PROC GLM procedure of SAS package version 9.3. Before pooling the data, Bartlett's test was conducted to test for homogeneity between environments (Gomez and Gomez, 1984). The components of variance in both the seasons were estimated considering various effects (seasons, replicates, and F_3 families) as random in the statistical model. As described by Bohn et al. (1996), transformed entry means were used for the combined analysis of variance. The variance components were estimated as per Searle (1971). The heritability (h^2) was calculated following Hallauer and Miranda (1981).

$$h^2 = \frac{\sigma^2 g}{\frac{\sigma^2 e + \sigma^2 ge}{re} + \frac{\sigma^2 g}{e}}$$

Where r = the number of replications and e = the number of environments.

To understand the nature of distribution of $F_{2:3}$ progenies with respect to disease incidence, skewness and kurtosis were estimated (Snedecor and Cochran, 1994).

Linkage Map Construction Using iMAS (GMendel)

We used 199 SNP marker data on 344 $F_{2:3}$ progenies for linkage map construction. Five markers showed segregation distortion (SD) and the remaining markers showed expected Mendelian segregation ratio of 1:2:1 as revealed by the χ^2 test. The linkage analysis was performed with 194 markers using GMendel program of iMAS software. For determining linkage groups, a minimum logarithm of odds (LOD) of 3.0 and maximum recombination fraction of 0.40 were set as threshold values. The unique feature of GMendel 2.0 is that it performs multipoint linkage analysis on populations with complex genetic structures. It generates two point maximum likelihood estimates for all pairwise markers. Based on probability rules, linkage phases are correctly assigned and gene order is estimated using an advance multipoint mapping algorithm. Using a powerful method called the simulated annealing algorithm (SAA), multipoint gene order was determined by GMendel 2.0. The validation of marker ordering was carried out by Monte Carlo and bootstrap methods. Using the Haldane mapping function, recombination fraction was converted into map distance in centimorgan (cM) and the linkage map was constructed using intermarker distances calculated from the GMendel program.

QTL Location by WinQTL Cartographer Version 2.5

The analysis of QTL controlling the northern corn leaf blight resistance was performed on the arcsine-transformed means of $F_{2:3}$ families within each season as well as over seasons. The phenotypic data (rainy season of 2013 and 2014) and genotypic data of 194 SNP markers across 10 chromosomes were used to identify the QTL associated with disease resistance employing WinQTL Cartographer version 2.5 (Wang et al., 2010). The replicated mean data of 344 F_3 progenies for northern leaf blight were used for QTL mapping in each season. To determine the QTL across the seasons, replicated means of across-season means of 344 $F_{2:3}$ progenies were used. The composite interval mapping method (CIM) was used for QTL analysis (Zeng, 1994) employing WinQTL Cartographer 2.5. The presence of putative QTL in an interval was tested using the Bonferroni χ^2 approximation (Zeng, 1994) corresponding to genome-wise type-I error by using a critical value for the LOD threshold of 2.5. Both additive and dominance models were used for the analysis in this study as the mapping population is comprised of $F_{2:3}$ progenies.

The ratio of the absolute value of dominance effect to the absolute value of additive effect, i.e., $|d|/|a|$, was used to determine gene action with 0–0.20 = additive, 0.21–0.80 = partial dominance, 0.81–1.20 = dominance, and >1.20 = overdominance (Stuber et al., 1987).

Co-localization of QTL for Multiple Foliar Diseases of Maize

The QTL for resistance to SDM and SCR were also identified along with NCLB. These QTL were co-localized to different bins of the chromosomes (Gardiner et al., 1993)[3]. Pearson correlation coefficients between the means of northern leaf blight, sorghum downy mildew, and southern corn rust were calculated to assess the degree of association between the foliar diseases of maize.

$$rp\,(xy) = Covp\,(xy)/\sqrt{\sigma^2 p\,(x).\,\sigma^2 p\,(y)}$$

Where,
$rp\,(xy)$ = Correlation between "x" and "y"
$Covp\,(xy)$ = Covariance between "x" and "y"
$\sigma^2 p\,(x)$ = Variance of "x"
$\sigma^2 p\,(y)$ = Variance of "y"

Development of Inbred Lines With Multiple Disease Resistance and High General Combining Ability

We attempted to develop inbred lines with resistance to three foliar diseases, namely sorghum downy mildew, northern corn leaf blight, and Polysora rust, using the high combining susceptible inbred lines CML153, CM212, and HKI162. These inbreds were crossed to the resistant inbred SKV50 during the 2014 summer season to derive inbred lines. The three F_1s (CM212 × SKV50, HKI162 × SKV50, and CML153 × SKV50) were raised during the rainy season of 2015 and selfed. During the summer season of 2016, F_2 plants were raised and screened for three major QTL present on chromosomes 8 (MZA6428-11 and MZA3856-10) for NCLB, 2 (MZA3668-12 and C00324-01) for SDM, and 10 (MZA15331-16 and MZA3922-32) for SCR resistance. The resistant progenies were advanced to the F_6 stage employing the plant-to-row approach. These progenies were screened in disease nurseries and 125 progenies with acceptable level of disease reaction were identified from three crosses. These progenies were crossed to the open pollinated tester CM500, 77 test cross progenies were obtained and evaluated during the summer season of 2017 in a single row plot of 4 m length with two replications to identify progenies with high general combining ability. The 77 progenies were evaluated for disease reaction during the rainy season of 2018.

RESULTS

Phenotypic Data Analysis

Weather conditions at Mandya favored the development of severe northern corn leaf blight disease. The percent disease severity and percent disease index values during the rainy season of 2013 and 2014 ranged from 0 to 100% which followed a binomial distribution. An attempt was made to make the means and variances independent and normally distributed by subjecting the data for arcsine transformation. The arcsine-transformed percent disease data of 344 $F_{2:3}$ progenies were used for statistical analysis and QTL mapping.

The parents CML153 and SKV50 differed significantly in their reaction to the NCLB disease as indicated by their percent disease incidence (**Table 1**). The parent SKV50 showed resistance reaction and CML153 was susceptible. In F_3 progenies, the mean northern corn leaf blight disease incidence was 52.38% in 2013, 40.71% in 2014, and 46.55% when pooled over seasons. The maximum range of disease incidence (17.14–84.29%) was recorded in 2013 followed by 2014 (16.00–74.00%).

The analysis of variance revealed significant differences among the progenies indicating the presence of genetic variability in the F_3 progenies (**Table 1**). Non-significant Bartlett's χ^2 test indicated the homogeneity of error mean sum of squares between seasons, and hence, data were pooled. Variance due to genotype and genotype × season interaction was significant, indicating that the expression of disease incidence significantly varied among F_3 families and depends upon testing season.

Genetic Variability Studies in the $F_{2:3}$ Population of Maize

The estimates of phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) for northern corn leaf blight were moderate in 2013 and 2014 (**Table 1**). High heritability and moderate GAM were noticed in 2013 (58.82 and 19.68%, respectively) and 2014 (43.67 and 16.57%, respectively), whereas in the pooled data, the heritability value was 65.72% and GAM was 17.42%.

Test for Normality, Skewness, and Kurtosis

The frequency distribution pattern of F_3 families was positively skewed and platykurtic for NCLB disease (**Figure 1**). The results of the test for normality by the Kolmogorov–Smirnov goodness-of-fit test indicated that the distribution of phenotypic means, within and across the seasons, deviated significantly from normal distribution with majority of the progenies skewed toward resistance in northern corn leaf blight.

Construction of the Linkage Map

Out of the 768 SNP markers, 194 SNP markers were found polymorphic and segregating in Mendelian fashion (1:2:1) and, hence, used for linkage map construction. The markers were mapped on 10 linkage groups (LGs) spanning 2,143.02 cM. The number of markers mapped per linkage group ranged from 6 (LG7) to 33 (LG1). The length of the linkage groups ranged from 152.25 (LG9) to 308.76 cM (LG1) with an average interval distance of 10.77 cM, indicating comparatively high-density SNP linkage map. The SNP map constructed was compared to the maize genome database[4,5] and genomically analyzed according to a previous study reported

[3]www.maizegdb.org/bin-viewer

[4]www.panzea.org

[5]www.maizeGDB.org

TABLE 1 | Mean disease incidence, skewness, kurtosis, variance components, and heritability (H) for reaction of $F_{2:3}$ progenies to northern corn leaf blight (NCLB) during the rainy season of 2013 and 2014.

Disease	Season	Susceptible parent mean	Resistant parent mean	$F_{2:3}$ grand mean	Skewness	Kurtosis	KS test (D)	Pr > D	Range in $F_{2:3}$ progenies	Variance components		Heritability (%)
										δ^2_g	δ^2_{ge}	
Northern corn leaf blight	Rainy season, 2013	66.00	5.00	52.38	0.095	−0.397	0.08	<0.01	17.14–84.29	42.56**		58.82
	Rainy season, 2014	75.00	7.00	40.71	0.188	−0.101	0.063	<0.01	16.00–74.00	24.54**		43.67
	Combined	70.50	6.00	46.55	0.125	−0.362	0.072	<0.01	26.50–69.25		23.57**	65.72

*Significant at $p = 0.05$; **significant at $p = 0.01$.

by Jones et al. (2009). This linkage map was used for the identification and mapping of QTL conferring resistance to three foliar diseases of maize.

QTL Analysis

Three QTL positions were identified for northern corn leaf blight resistance during the rainy season of 2013 (**Table 2** and **Figures 2**, **3**). One QTL was located on chromosome 2 (qNCLB-2) flanked by markers C00359-01–MZA13360-13 which explained 3.07% phenotypic variation with LOD of 3.06. Two QTL were located on chromosome 8 (qNCLB-8-1 and qNCLB-8-2) flanked by markers, viz. MZA2487-6 and MZA6428-11–MZA3856-10, and these two QTL showed phenotypic variation of 2.46 and 22.97% with LOD score of 2.77 and 3.44, respectively. A major northern corn leaf blight QTL was mapped on chromosome 8 (qNCLB-8-2) which explained maximum phenotypic variation of 22.97%. These three identified QTL explained a total of 28.50% phenotypic variation. Additive gene effects of these three QTL ranged from 1.71 to 1.92, and the favorable alleles for these QTL were contributed by the resistant parent SKV50. The QTL located on chromosome 2 (qNCLB-2) and on chromosome 8 (qNCLB-8-2) exhibited overdominance gene action, while another QTL on chromosome 8 (qNCLB-8-1) revealed a dominance type of gene action.

During the rainy season of 2014, three QTL regions conferring resistance to northern corn leaf blight were mapped onto chromosome 5 and chromosome 8. Among these, a QTL bracketed by markers MZA6428-11–MZA3856-10 on chromosome 8 explained the highest phenotypic variation of 16.8% with LOD score 3.05 (qNCLB-8-2) followed by a QTL on chromosome 8 present adjacent to the marker MZA2487-6 which explained 2.90% of phenotypic variation with LOD score of 2.80 (qNCLB-8-1). Other QTL on chromosome 5 flanked by markers MZA5359-10–MZA3137-17 (qNCLB-5-1) contributed 2.00% to the phenotypic variation with LOD score of 4.12. The range of additive genetic effects for these QTL was from −0.13 to 3.48, and the total phenotypic variation of 21.60% was explained by the QTL identified. Favorable allele for QTL located on chromosome 8 (qNCLB-8-2) was contributed by susceptible parent CML153, whereas for other QTL, it was by the resistant parent SKV50. The QTL (qNCLB-8-1) on chromosome 8 showed dominance gene action and two QTL (qNCLB-8-2 and qNCLB-5-1) located on chromosomes 8 and 5 showed overdominance gene action.

In the combined QTL analysis, five QTL were detected on chromosomes 2, 5, and 8. Among these QTL, three QTL flanked by markers C00359-01–MZA1336013 (qNCLB-2), MZA2487-6 (qNCLB-8-1), and MZA6428-11–MZA3856-10 (qNCLB-8-2) were found to be consistent across two seasons with a LOD score of 3.22, 3.71, and 3.13. Two novel QTL were located on chromosome 5 flanked by markers MZA3103-47–MZA533-46 (qNCLB-5-2) and MZA5296-6–C00171-01 (qNCLB-5-3). The QTL qNCLB-8-2 explained the highest phenotypic variation of 16.34% followed by qNCLB-5-3 (10.24%). A total phenotypic variation explained by these QTL was 32.27% in the pooled analysis. An additive genetic effect of these QTL ranged from 0.74 to 3.39, and the favorable allele for these QTL was contributed by

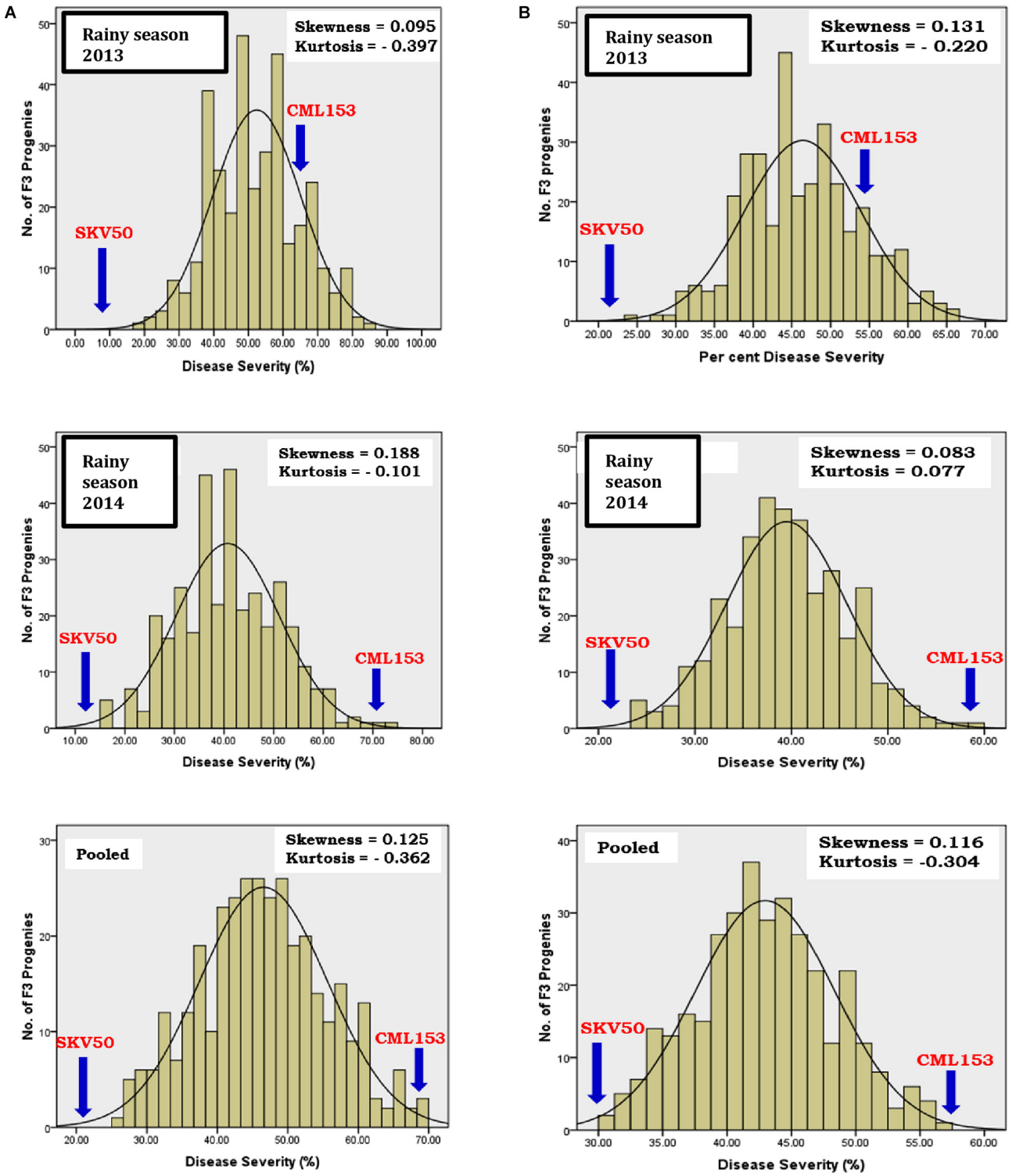


FIGURE 1 | Frequency distribution of mean per cent disease severity of northern corn leaf blight in the F2:3 population derived from the cross CML153 × SKV50 (**A**, original; **B**, Arcsine transformed).

TABLE 2 | QTL detected for northern corn leaf blight resistance during individual seasons and combined over seasons using 344 $F_{2:3}$ families from the cross CML 153 × SKV 50 (threshold LOD score = 2.50).

Season	Chromosome	Bin location	Flanking markers		QTL position (cM)	Maximum LOD score	R^2 (%)	Genetic effect		Gene action	Donor allele	QTL name
			Left	Right				Additive	Dominance			
Rainy season, 2013	2	2.06	C00359-01	MZA13360-13	108.11	3.06	3.07	1.71	3.02	OD	SKV50	*qNCLB-2*
	8	8.03	MZA2487-6		42.91	2.77	2.46	1.72	2.07	D	SKV50	*qNCLB-8-1*
	8	8.06	MZA6428-11	MZA3856-10	80.01	3.44	22.97	1.92	3.62	OD	SKV50	*qNCLB-8-2*
Rainy season, 2014	5	5.03	MZA5359-10	MZA3137-17	49.21	4.12	1.00	3.48	4.30	OD	SKV50	*qNCLB-5-1*
	8	8.03	MZA2487-6		37.81	2.80	1.90	1.65	0.89	D	SKV50	*qNCLB-8-1*
	8	8.06	MZA6428-11	MZA3856-10	81.31	3.05	9.68	−0.13	3.51	OD	SKV50	*qNCLB-8-2*
Combined	2	2.06	C00359-01	MZA13360-13	108.11	3.22	2.28	1.25	2.33	OD	SKV50	*qNCLB-2*
	5	5.03	MZA3103-47	MZA533-46	53.21	4.21	1.77	0.74	2.41	OD	SKV50	*qNCLB-5-2*
	5	5.04–5.05	MZA5296-6	C00171-01	142.71	3.40	10.24	3.39	−1.20	D	SKV50	*qNCLB-5-3*
	8	8.03	MZA2487-6		43.21	3.71	1.64	2.99	1.07	D	SKV50	*qNCLB-8-1*
	8	8.06	MZA6428-11	MZA3856-10	81.31	3.13	16.34	0.56	2.60	OD	SKV50	*qNCLB-8-2*

OD, overdominance; D, dominance.

the resistant parent SKV50. Out of five QTL, two QTL (qNCLB-5-3 and qNCLB-8-1) located on chromosomes 5 and 8 exhibited dominance, and the other three QTL (qNCLB-2, qNCLB-5-2, and qNCLB-8-2) on chromosomes 2, 5, and 8 showed overdominance gene action in the combined analysis.

Co-localization of QTL Conferring Resistance to Multiple Foliar Pathogens

Since QTL analysis for sorghum downy mildew and southern corn rust was performed from one year data, results were not presented. However, the information generated was used for the estimation of pairwise Pearson correlation coefficient between three diseases (northern corn leaf blight, sorghum downy mildew, and southern corn rust) and co-localization of QTL. The highest significant and positive correlation was noticed between northern corn leaf blight and southern corn rust ($r = 0.122$), whereas a positive but non-significant pairwise correlation between sorghum downy mildew and southern corn rust ($r = 0.061$) was observed (**Table 3**). Between sorghum downy mildew and northern corn leaf blight, a negatively significant correlation was revealed ($r = -0.120$). This clearly indicated the presence of genes carrying resistance to multiple diseases in the $F_{2:3}$ population derived from the cross CML153 × SKV50. In the present study, co-localized chromosomal regions harboring QTL for northern corn leaf blight, sorghum downy mildew, and southern corn rust resistance were observed (**Table 4**). In bin 8.03, QTL conferring resistance to all three foliar diseases, viz. northern corn leaf blight (one QTL), sorghum downy mildew (two QTL), and southern corn rust (two QTL), were co-localized at approximately the same map position (<10 cM difference between the QTL peaks for the three diseases) with the common adjacent marker MZA2487-6. The bin 5.03–5.04 was significantly associated with resistance to northern corn leaf blight and sorghum downy mildew, wherein two QTL for northern corn leaf blight and one QTL for sorghum downy mildew co-localized with a map distance of <3 cM. This co-localized QTL exhibited 10.24 and 5.98% of total phenotypic variation for northern corn leaf blight and sorghum downy mildew, respectively. The QTL for southern corn rust and sorghum downy mildew resistance were co-localized in bin 3.04 at a map distance of 1 cM with a common flanking marker MZA1959-26, while QTL conferring resistance to northern corn leaf blight and sorghum downy mildew were co-localized in chromosome bin 2.06.

Development of Inbred Lines With Resistance to Three Diseases With Good General Combining Ability

We developed 125 lines and crossed with the broad genetic base tester CM500 for testing their combining ability. Only 77 lines produced successful test cross progenies. The evaluation of these test cross progenies resulted in the identification of 39 lines with high general combining ability for grain yield with 17 lines showing significant standard heterosis. These 77 lines exhibited resistance/moderate resistance to NCLB, SDM, and Polysora rust diseases (**Supplementary Table 1**).

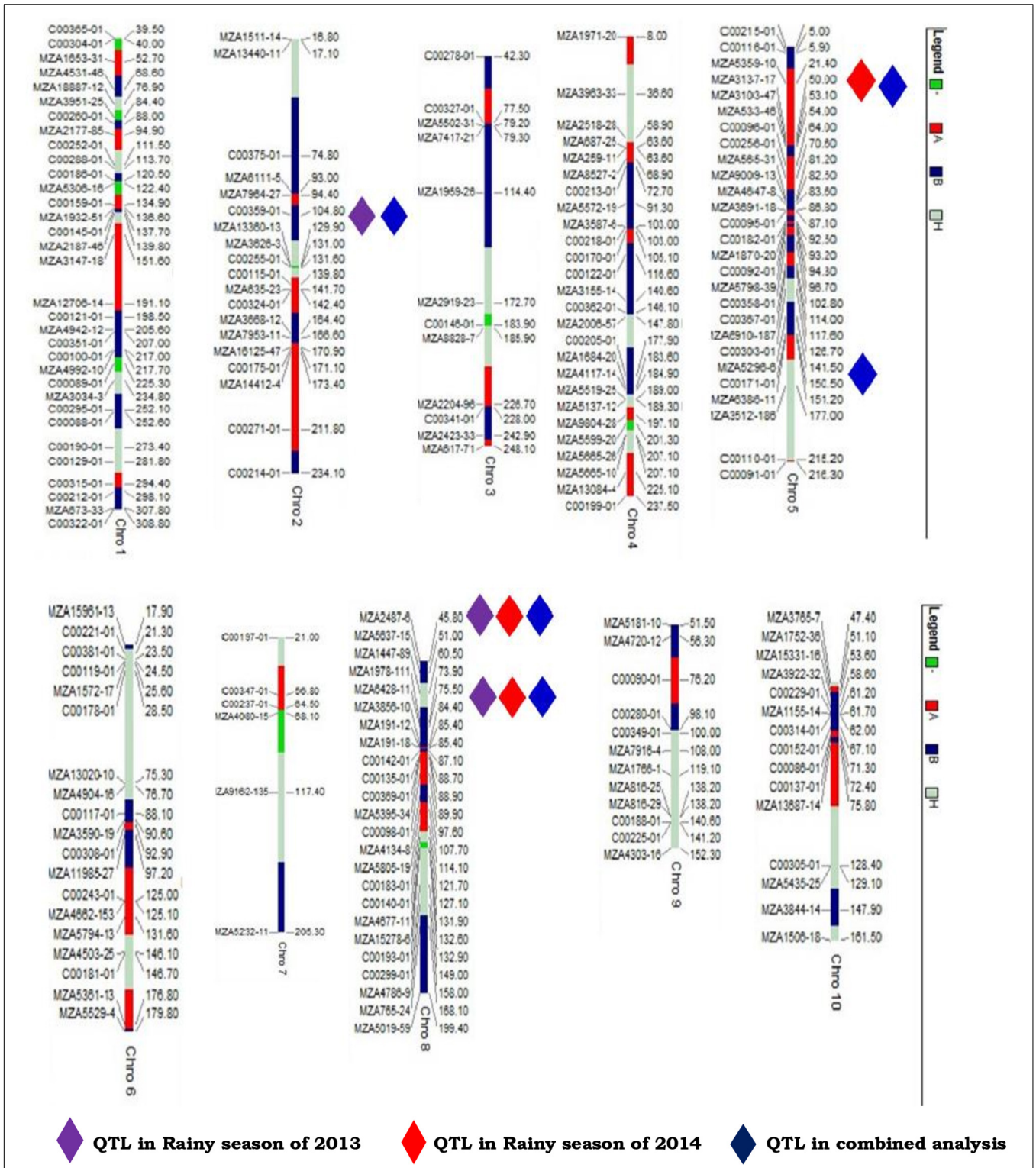


FIGURE 2 | Linkage map and position of the QTL associated with northern corn leaf blight resistance mapped from F2:3 mapping population of the cross CML153 × SKV50.

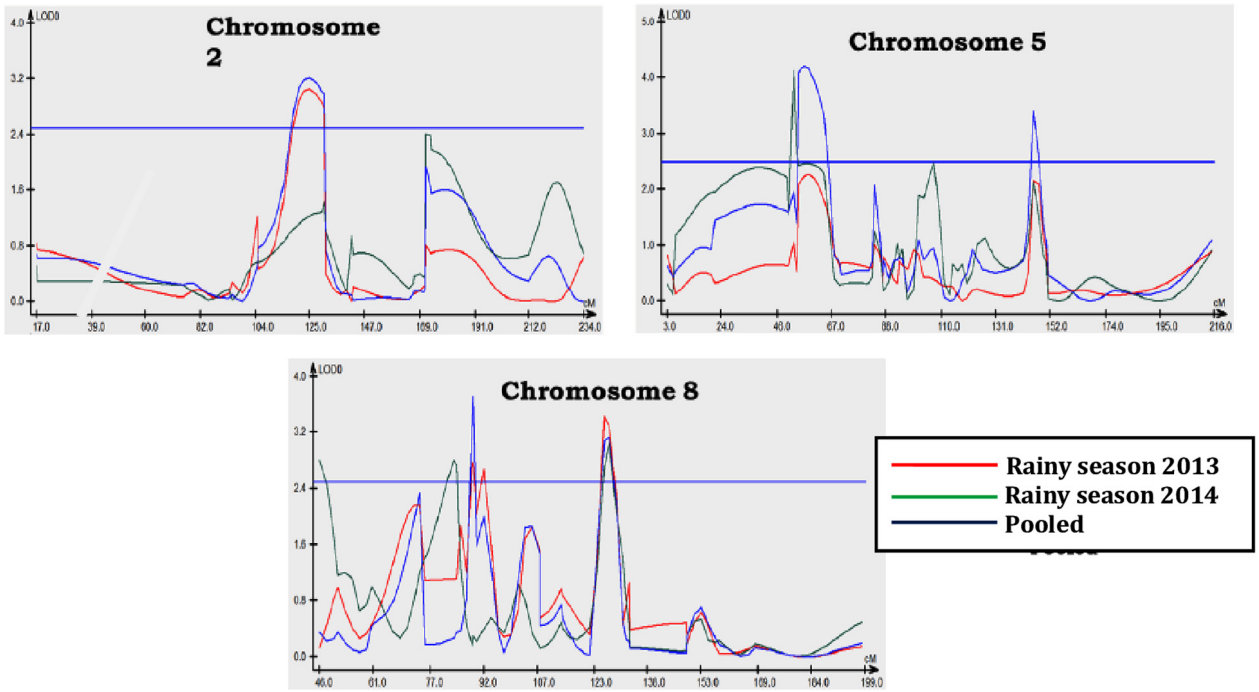


FIGURE 3 | LOD peak for QTL conditioning resistance to northern corn leaf blight on chromosomes 2, 5, and 8 in rainy season *(Kharif)* of 2013, 2014, and pooled analysis over seasons.

TABLE 3 | Pearson correlation coefficient between mean disease severity for three diseases, viz. northern corn leaf blight (NCLB), sorghum downy mildew (SDM), and southern corn rust (SCR).

Disease	NCLB	SDM	SCR
NCLB	1	−0.120*	0.122*
SDM		1	0.061
SCR			1

**Significant at the 5% level of significance.*

DISCUSSION

Highly diverse parents for multiple pathogen infection were used in the development of $F_{2:3}$ progenies which resulted in highly significant differences among the progenies for disease reaction against northern corn leaf blight pathogen. The significant difference among means of F_3 families for northern corn leaf blight indicated the presence of genotypic variability within the population. Assuming a random effects model, Bartlett's test proved the homogeneity of error mean sum of squares for northern corn leaf blight data between 2013 and 2014. Therefore, data from these two seasons were pooled. In the pooled analysis, genotype × season interaction was significant, demonstrating the influence of season on northern corn leaf blight incidence (Chen et al., 2016; Shridhar Hegde et al., 2018). In this study, disease pressure was high in 2013 compared with 2014. A significant seasonal effect was observed which indicated that disease development was highly influenced by weather conditions like rainfall, temperature, and relative humidity. Weather data recorded at the V.C. Farm, Mandya indicated comparatively more number of rainy days coupled with high humidity during 2013 resulting in a higher range of disease expression.

Frequency distribution of 344 F_3 progenies from the cross CML153 × SKV50 revealed non-normal distribution for the incidence of NCLB. Positively skewed distribution was observed for northern corn leaf blight in each season data and pooled data. Skewed distribution was observed toward the resistant parent SKV50 indicating the dominance of resistance (Schechert et al., 1999; Welz et al., 1999; Brown et al., 2001; Asea et al., 2009). However, the distribution was made near normal through arcsine transformation of the percent disease incidence data. Approximately or near normal distribution in phenotypic data was obtained earlier on $F_{2:3}$ populations (Holland et al., 1998; Brunelli et al., 2002; Shridhar Hegde et al., 2018). The resistance to NCLB disease appeared to be controlled by a larger number of genes having decreasing effects with the involvement of dominance-based complementary interaction as evidenced by platykurtic and positively skewed distribution (Pooni et al., 1977; Choo and Reinbergs, 1982). High heritability with high genetic advance over mean indicated the reliability of the estimates of variation between F_3 families, and a reasonable progress in selection is possible for disease resistance in this population. Similar results were reported for northern corn leaf blight by Freymark et al. (1993, 1994), Schechert et al. (1999), Hakiza et al. (2004), Balint-Kurti et al. (2010), and Zwonitzer et al. (2010).

In the present investigation, 194 SNP markers showed the expected Mendelian segregation ratios and were used to construct

TABLE 4 | Co-located QTL conditioning resistance to northern corn leaf blight (NCLB), sorghum downy mildew (SDM), and southern corn rust (SCR).

Co-localized chromosome bin	Disease QTL	Flanking marker interval	QTL position (cM)	Maximum LOD score	R^2 (%)	Genetic effect		Gene action	Donor Allele
						Additive	Dominance		
8.03	NCLB	MZA2487-6	43.21	3.71	1.64	2.99	1.07	D	SKV50
	SDM	MZA2487-6	26.71	5.72	1.20	1.44	9.49	OD	SKV50
	SDM	MZA2487-6	39.61	3.89	1.63	5.23	2.24	PD	SKV50
	SCR	MZA2487-6	22.71	2.92	4.17	1.67	5.23	OD	SKV50
	SCR	MZA2487-6–MZA5637-15	48.21	3.11	4.71	8.19	6.79	D	SKV50
5.03/5.04	NCLB	MZA3103-47–MZA533-46	53.21	4.21	1.77	0.74	2.41	OD	SKV50
	NCLB	MZA5296-6–C00171-01	142.71	3.40	10.24	3.39	−1.20	D	SKV50
	SDM	C00303-01–MZA5296-6	138.71	3.22	5.98	3.44	−2.23	PD	SKV50
3.04	SDM	MZA7417-21–MZA1959-26	100.11	2.56	8.28	8.13	−12.24	OD	SKV50
	SCR	MZA7417-21–MZA1959-26	102.11	9.46	1.01	−16.28	15.09	D	CML153
2.06	NCLB	C00359-01–MZA13360-13	108.11	3.22	2.28	1.25	2.33	OD	SKV50
	SDM	C00324-01–MZA3668-12	147.61	2.57	20.42	3.11	1.09	PD	SKV50
	SDM	C00324-01–MZA3668-12	154.31	2.63	15.43	−2.88	1.35	PD	CML153

PD, partial dominance.

relatively high density linkage map consisting of 10 linkage groups. The total map distance covered about 2,143.02 cM with an average interval length of 10.77 cM. The length of a linkage map is influenced by a number of factors, such as the number of markers, size of the mapping population, and genotyping accuracy. The average distance between markers in our linkage map was 10.77 cM, which was longer than the other saturated linkage maps with SNP markers (729.28–2,236.66 cM in length with an interval of 0.66–10 cM) which were successfully utilized for the identification of QTL for various traits in maize (Zou et al., 2012; Chen et al., 2016; Zhou et al., 2016; Su et al., 2017; Xia et al., 2020). However, Doerge (2002) was of the opinion that QTL mapping of important traits can be practiced with marker density up to 20 cM. The near normal distribution pattern of F_3 progenies suggested the reliability of the $F_{2:3}$ mapping population for the identification of QTL for resistance to northern corn leaf blight.

Mapping of QTL Conferring Resistance to Northern Corn Leaf Blight

The identification of QTL with resistance to diseases could serve as a novel strategy to develop disease-resistant lines. The SNP-based linkage maps have been used to precisely map disease resistance through genetic linkage and extensively helped researchers in understanding the function of the chromosomal region or locus at gene levels (St Clair, 2010). In this study, genomic locations of QTL for northern corn leaf blight were identified on chromosomes 2, 5, and 8. The QTL on chromosomes 5 and 8 were major with a high percentage of the phenotypic variance for resistance. The QTL detected on chromosome 8 in bin 8.06 explained 16.34% of the phenotypic variance followed by the QTL on chromosome 5 which explained 10.24% variation in the combined analysis. When considered together, the total variance explained by these QTL was 26.58%. The UMC reference map of maize (Davis et al., 1999) was used to compare QTL positions (Lin et al., 1995; Tuberosa et al., 2002). The major QTL on chromosome bins 5.04–5.05 and 8.06 were also reported by Freymark et al. (1993, 1994) in bins 5.03–5.05 and 8.03–8.06, Dingerdissen et al. (1996) in bins 5.03 and 8.06, Schechert et al. (1999) in bins 5.03–5.05 and 8.06, Welz et al. (1999) in bins 5.03–5.04 and 8.02–8.06, Welz and Geiger (2000) in bins 5.04 and 8.06, Brown et al. (2001) in bin 8.06, Asea et al. (2009) in bins 5.04 and 8.06, Chung et al. (2011) in bins 5.03 and 8.06, and Chen et al. (2016) in bin 5.04. This supports the suggestion from Chung et al. (2010), who fine-mapped northern corn leaf blight QTL on chromosome bin 8.06, that a major QTL region on chromosome 8 affects the response to northern corn leaf blight. Similarly, an important role of QTL on chromosome 5 at bin 4 was deciphered as an additional QTL to that already known in bins 5.03, 5.04, and 5.05 by Welz et al. (1999), Asea et al. (2009), and Chung et al. (2011). Another QTL detected in chromosome bin 2.06 was also reported earlier by Schechert et al. (1999), Welz et al. (1999), Brown et al. (2001), Ping et al. (2007), Balint-Kurti et al. (2010), Zwonitzer et al. (2010); and Shridhar Hegde et al. (2018). The major QTL identified in our study at bin location 8.06 exhibited overdominance (OD) gene action. It could be confirmed that northern corn leaf blight resistance alleles in SKV50 are a good source for marker-assisted selection supporting maize breeding programs. Similar findings were observed by Welz et al. (1999), Welz and Geiger (2000), Brown et al. (2001), Chung et al. (2010), Zwonitzer et al. (2010), and Chung et al. (2011).

One of the major goals of QTL mapping is to identify and use QTL with little QTL × environment interaction (Stuber et al., 1987). This was substantiated in the present study as QTL in bins 8.03 and 8.06 were detected in both seasons and in the pooled analysis. Earlier workers also reported the identification of QTL in all environments with difference in the level of significance and magnitude of genetic effects (Brown et al., 2001; Asea et al., 2009; Balint-Kurti et al., 2010; Chen et al., 2016; Shridhar Hegde et al., 2018). Conclusively, the markers associated with QTL in bins 2.06 (C00359-01 and MZA13360-13), 5.03 (MZA3103-47 and MZA533-46), 5.04 (MZA5296-6 and C00171-01), and 8.06 (MZA6428-11 and MZA3856-10) are the favorites to be

used to transfer resistance alleles to susceptible lines (Xia et al., 2020). The QTL localized in chromosome bin 5.04 was previously detected in the NAM population (Poland et al., 2011) and is known to contain a candidate gene, GRMZM2G024612 (Xia et al., 2020). The genomic region in chromosome 8 is important as two resistant genes *Ht2* and *Htn1* were identified previously (Zaitlin et al., 1992; Simcox and Bennetzen, 1993; Hurni et al., 2015; Kaefer et al., 2017). Wisser et al. (2006) reported *Histatin-1* (*Htn1*) as a maize disease resistance gene against NCLB and is known to encode a wall-associated receptor-like kinase that acts as an important component of the plant innate immune system by perceiving pathogen or host-derived elicitors (Hurni et al., 2015). The gene *Ht2* has been further delimited to a region of up to 0.46 Mb *via* precise mapping, and other candidate genes were predicted and annotated according to the reference genome sequences of B73 (Chung et al., 2010). Except for chromosomes 5 and 8, NCLB resistance-related QTL seems to be relatively rare at other chromosomes (Xia et al., 2020).

Clearly, there is a lack of commonality between the QTL identified for northern corn leaf blight in different populations. This could be because of multiple reasons such as the use of different types and sizes of mapping populations, involvement of different sets of QTL in different crosses, and epistatic interaction between QTL. Furthermore, as noted in previous studies, it is possible that QTL may not be detected in certain segregating populations if alternate alleles of the QTL are not contributed by both parents (Beavis and Keim, 1996; Kearsey and Pooni, 1996; Bohn et al., 1997). In a previous study, a comparison of QTL for disease resistance in multiple segregating populations revealed only few common QTL (Beavis et al., 1991). This is primarily due to environmental conditions existing in a particular region which might affect the QTL expression. Nonetheless, the detection of different QTL present in diverse resistant genotypes could aid in the pyramiding of multiple QTL in cultivars.

Co-localization of QTL Conferring Resistance to Three Foliar Diseases

Mapping resistance loci for multiple pathogens provides an opportunity for co-localizing resistance loci. Such resistance gene combinations are expected to provide more durable protection (Simmonds, 1985) against a variable number of pathogens. In the present study, QTL for northern corn leaf blight, sorghum downy mildew, and southern corn rust resistance were co-localized based on the bin locations and chromosomal regions where QTL were detected. Associations between resistance to northern corn leaf blight, sorghum downy mildew, and southern corn rust were significant in the $F_{2:3}$ population. These associations were detected most likely because of alleles showing a high level of resistance to one disease and a lower level to another which may not be detected by QTL analysis. Alternatively, it could be because of alleles that confer varied levels of resistance to multiple diseases that are undetectable by QTL analysis. A total of 13 QTL associated with resistance to two or more diseases were detected. These QTL regions may carry a single gene showing multiple disease resistance or due to two closely linked genes causing resistance to different diseases in the $F_{2:3}$ population. The inbred SKV50 with resistance to foliar diseases, viz. northern corn leaf blight, sorghum downy mildew, and southern corn rust, could be used for pyramiding resistance QTL. The study also indicated the potential of using several target QTL present in chromosome bin 8.03 for resistance to northern corn leaf blight, sorghum downy mildew, and southern corn rust: bins 5.03–5.04 and 2.06 for resistance to northern corn leaf blight and sorghum downy mildew and bin 3.04 for resistance to sorghum downy mildew and southern corn rust for marker-assisted selection to pyramid quantitative resistance to multiple foliar pathogens as described by Pratt and Gordon (2006).

The present study resulted in the identification of markers linked to resistance to northern corn leaf blight, and some common QTL for resistance to three important foliar diseases were identified. It was also possible to validate the major QTL linked to NCLB, SDM, and SCR and to develop 39 inbred lines with high general combining ability which could be used in hybrid development programs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

HL and HR designed and carried out the experiments and contributed to manuscript preparation. AP helped in genotyping of the mapping populations and editing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge the financial support from the M/s Pioneer Hi-Bred Pvt. Ltd., India, to HL. HR was supported by the DST under the PURSE program.

ACKNOWLEDGMENTS

The authors are highly indebted to Dr. H. C. Prasanna, Principal Scientist (Plant Breeding), Indian Institute of Horticultural Research, Bengaluru, for carefully editing and providing inputs to improve the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.548407/full#supplementary-material

REFERENCES

Ali, F., and Yan, J. B. (2012). The phenomenon of disease resistance in maize and the role of molecular breeding in defending against global threat. *J. Integr. Plant Biol.* 54, 134–151. doi: 10.1111/j.1744-7909.2012.01105.x

Asea, G., Vivek, B. S., Bigirwa, G., Lipps, P. E., and Pratt, R. C. (2009). Validation of consensus quantitative trait loci associated with resistance to multiple foliar pathogens of maize. *Phytopathology* 99, 540–547. doi: 10.1094/phyto-99-5-0540

Balint-Kurti, P. J., Yang, J., Esbroeck, G. V., Jung, J., and Smith, M. E. (2010). Use of a maize advanced intercross line for mapping of QTL for northern leaf blight resistance and multiple disease resistance. *Crop Sci.* 50, 458–466. doi: 10.2135/cropsci2009.02.0066

Beavis, W. D., Grant, D., Albertsen, M., and Fincher, R. (1991). Quantitative trait loci for plant height in four maize populations and their association with qualitative trait loci. *Theor. Appl. Genet.* 83, 161–165.

Beavis, W. D., and Keim, P. (1996). "Identification of quantitative trait locithat are affected by environment," in *Genotype by Environment Interaction*, eds M. S. Kang and H. G. Gauch Jr. (BocaRaton, FL: CRC Press), 123–149. doi: 10.1201/9781420049374.ch5

Bohn, M., Khairallah, M. M., Gonzalez-de-Leon, D., Hoisington, D., Utz, H. F., Deutsch, J. A., et al. (1996). QTL mapping in tropical maize: I. Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits. *Crop Sci.* 36, 1352–1361. doi: 10.2135/cropsci1996.0011183x003600050045x

Bohn, M., Khairallah, M. M., Jiang, C. Z., Gonzalez de Leon, D., Hoisington, D., Utz, H. F., et al. (1997). QTL mapping in tropical maize. 2. Comparison of genomic regions for resistance to *Diatraea* spp. *Crop Sci.* 37, 1892–1902. doi: 10.2135/cropsci1997.0011183x003700060038x

Brown, A. F., Juvik, J. A., and Pataky, J. K. (2001). Quantitative trait loci in sweet corn associated with partial resistance to Stewart's Wilt, Northern Corn leaf blight, and common rust. *Phytopathology* 91, 293–300. doi: 10.1094/phyto.2001.91.3.293

Brunelli, K. R., Silva, H. P., and Camargo, L. E. A. (2002). Mapping of resistance genes to *Puccinia polysora* in maize. *Fitopatol. Brasil.* 27, 134–140. doi: 10.1590/S0100-41582002000200003

Butler, E. J. (1907). Some diseases of cereals caused by *Sclerospora graminicola*. *Memorandum Dept. Agri. India Bot. Seri.* 2, 1–24.

Carson, M. L. (1999). "*Helminthosporium* leaf spots and blights," in *Compendium of Corn Diseases*, 3rd Edn, ed. D. G. White (St. Paul, MN: The American Phytopathol.), 15–24.

Chaudhary, B., and Mani, V. P. (2010). Genetic analysis of resistance to turcicum leaf blight in semi-temperate early maturing genotypes of maize (*Zea mays*). *Indian J. Genet.* 70, 65–70.

Chen, G., Xiaoming, W., Long, S., Jaqueth, J., Li, B., Yan, J., et al. (2016). Mapping of QTL conferring resistance to northern corn leaf blight using high-density SNPs in maize. *Mol. Breed.* 36, 4. doi: 10.1007/s11032-015-0421-3

Choo, T. M., and Reinbergs, E. (1982). Analyses of skewness and kurtosis for detecting gene interaction in a doubled haploid population. *Crop Sci.* 22, 231–235. doi: 10.2135/cropsci1982.0011183X002200020008x

Chung, C. L., Jamann, T., Longfellow, J., and Nelson, R. (2010). Characterization and fine-mapping of a resistance locus for northern leaf blight in maize bin 8.06. *Theor. Appl. Genet.* 121, 205–227. doi: 10.1007/s00122-010-1303-z

Chung, C. L., Poland, P., Kump, K., Benson, J., Longfellow, J., Walsh, E., et al. (2011). Targeted discovery of quantitative trait loci for resistance to northern leaf blight and other diseases of maize. *Theor. Appl. Genet.* 123, 307–326. doi: 10.1007/s00122-011-1585-9

Davis, G. L., McMullen, M. D., Baysdorf, C., Musket, T., Grant, D., Staebell, M., et al. (1999). A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics* 152, 1137–1172.

Dingerdissen, A. L., Geiger, H. H., Lee, M., Schechert, A., and Welz, H. G. (1996). Interval mapping of genes for quantitative resistance of maize to *Setosphaeria turcica*, cause of northern leaf blight, in a tropical environment. *Mol. Breed.* 2, 143–156. doi: 10.1007/bf00441429

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* 3, 43–52. doi: 10.1038/nrg703

Dowswell, C. R., Paliwal, R. L., and Cantrell, R. P. (1996). *Maize in the Third World*. Boulder: West view press, 1–37.

Fehr, W. R. (1987). *Principles of Cultivar Development. PB*. New York, NY: Macmillan.

Freymark, P. J., Lee, M., Woodman, W. L., and Martinson, C. A. (1993). Quantitative and qualitative trait loci affecting host-plant response to *Exserohilum turcicum* in maize (*Zea mays* L.). *Theor. Appl. Genet.* 87, 537–544. doi: 10.1007/bf00221876

Freymark, P. J., Lee, M., Woodman, W. L., and Martinson, C. A. (1994). Molecular-marker facilitated investigation of host-plant response to *Exserohilum turcicum* in maize (*Zea mays* L.): components of resistance. *Theor. Appl. Genet.* 88, 305–313. doi: 10.1007/bf00223637

Gardiner, J. M., Coe, E. H., Melia-Hancock, S., Hoisington, D. A., and Chao, S. (1993). Development of a core RFLP map in maize using an immortalized F2 population. *Genetics* 134, 917–930.

Gomez, K. A., and Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research*, 2nd Edn. New York: John Wiley & Sons.

Hakiza, J. J., Lipps, P. E., Martin, S. S., and Pratt, R. C. (2004). Heritability and number of genes controlling partial resistance to *Exserohilum turcicum* in maize inbred H99. *Maydica* 49, 173–182.

Hallauer, A. R., and Miranda, J. B. (1981). *Quantitative Genetics in Maize Breeding*. Ames, IA: Iowa State University Press.

Harlapur, S. I., Mruthunjaya, C. W., Anahosur, K. H., and Muralikrishna, S. (2000). Survey and surveillance of maize diseases in North Karnataka. *Karnataka J. Agri. Sci.* 13, 750–751.

Hettiarachchi, K., Prasanna, B. M., Rajan, A., Singh, O. N., Gowda, K. T. P., Pant, S. K., et al. (2009). Generation mean analysis of turcicum leaf blight resistance in maize. *Indian J. Genet.* 69, 102–108.

Holland, J. B., Uhr, D. V., Jeffers, D., and Goodman, M. M. (1998). Inheritance of resistance to southern corn rust in tropical-by-corn-belt maize populations. *Theor. Appl. Genet.* 96, 232–241. doi: 10.1007/s001220050732

Hughes, G. R., and Hooker, A. L. (1971). Gene action conditioning resistance to northern leaf blight in maize. *Crop Sci.* 11, 180–184. doi: 10.2135/cropsci1971.0011183X001100020005x

Hurni, S., Scheuermann, D., Krattinger, S. G., et al. (2015). The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wall associated receptor-like kinase. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8780–8785. doi: 10.1073/pnas.1502522112

Jenkins, M. T., Robert, A. L., and Findley, W. R. (1952). Inheritance of resistance to *Helminthosporium turcicum* leaf blight in populations of F3 progenies. *Agron. J.* 44, 438–442. doi: 10.2134/agronj1952.00021962004400080010x

Jones, E., Chu, W. C., Ayele, M., Ho, J., Bruggeman, E., Yourstone, K., et al. (2009). Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.). *Mol. Breed.* 24, 165–176. doi: 10.1007/s11032-009-9281-z

Kachapur, M. R., and Hegde, R. K. (1988). Studies on turcicum leaf blight of maize caused by *Exserohilum turcicum*. *Plant Pathol.* 6, 33–35.

Kaefer, K. A. C., Schuelter, A. R., Schuster, I., Marcolin, J., and Vendrusco, E. C. G. (2017). Association mapping and genetic control for Northern leaf blight (*Exserohilum turcicum*) resistance in maize lines. *Aust. J. Crop Sci.* 11, 1346–1353.

Kearsey, M. J., and Pooni, H. S. (1996). *The Genetical Analysis of Quantitative Traits*. London: Chapman and Hall.

Lin, Y. R., Schertz, K. F., and Paterson, A. H. (1995). Comparative analysis of QTLQTL affecting plant height and maturity across the *Poaceae,* in reference to an interspecific sorghum population. *Genetics* 141, 391–411.

Little, T. M., and Hills, F. J. (1978). *Agricultural Experimentation Design and Analysis*. New York: John Wiley and Sons.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J. G., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6080–6084. doi: 10.1073/pnas.052125199

Payak, M. M., and Sharma, R. C. (1983). *Disease Rating Scales in Maize in India. Techniques of Scoring for Resistance to Important Diseases of Maize. All India Coordinated Maize Improvement Project*. New Delhi: IARI, 1–4.

Payak, M. M., and Sharma, R. C. (1985). Maize diseases and approaches to their management in India. *Trop. Pest. Manag.* 31, 302–310. doi: 10.1080/09670878509371006

Payak, M. M., Sharma, R. C., and Lilaramani, J. (1973). How to control maize diseases? *Indian Farm*. 23, 20–23.

Perkins, J. M., and Pederson, W. L. (1987). Disease treatment and yield loss associated with northern leaf blight of corn. *Plant Dis.* 71, 940–943. doi: 10.1094/pd-71-0940

Ping, Z. Z., Hong, L. X., Bi, H. Y., Zhong, L., Chuan, H., and Bo, T. Z. (2007). QTL mapping for resistance to northern corn leaf blight in maize. *South West China J. Agric. Sci.* 20, 634–637.

Poland, J. A., Bradbury, P. J., Buckler, E. S., and Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6893–6898. doi: 10.1073/pnas.1010894108

Pooni, H. S., Jinks, J. L., and Cornish, M. A. (1977). The causes and consequences of non-normality in pretending the properties of recombinant inbred lines. *Heredity* 38, 329–338. doi: 10.1038/hdy.1977.95

Pratt, R. C., and Gordon, S. G. (2006). Breeding for resistance to maize foliar pathogens. *Plant Breed. Rev*. 27, 119–173. doi: 10.1002/9780470650349.ch3

Ranganatha, H. M., Lohithaswa, H. C., and Anand, S. P. (2017). Understanding the genetic architecture of resistance to northern corn leaf blight and southern corn rust in maize (*Zea mays* L.). *Indian J. Genet.* 77, 357–363. doi: 10.5958/0975-6906.2017.00048.7

Raymundo, A. D., and Hooker, A. C. (1981). Measuring relationship between northern leaf blight of maize and yield losses. *Plant Dis.* 65, 325–327. doi: 10.1094/pd-65-325

Schaefer, C. M., and Bernardo, R. (2013). Genomewide association mapping of flowering time, kernel composition, and disease resistance in historical minnesota maize inbreds. *Crop Sci*. 53, 2518–2529. doi: 10.2135/cropsci2013.02.0121

Schechert, A. L., Welz, H. G., and Geiger, H. H. (1999). QTL for resistance to *Setosphaeria turcica* in tropical African maize. *Crop Sci*. 39, 514–523. doi: 10.2135/cropsci1999.0011183x003900020036x

Searle, S. R. (1971). Topics in variance components estimation. *Biometrics* 27, 1–76. doi: 10.2307/2528928

Shekhar, M., and Kumar, S. (2012). *Inoculation Methods and Disease Rating Scales for Maize Diseases. Directorate of Maize*. New Delhi: ICAR, 2–16.

Shridhar Hegde, K. M., Mahantesh, G., and Meenakshi, G. N. (2018). Mapping QTLQTL for resistance to northern leaf blight in tropical maize (*Zea mays* L.). *Int. J. Curr. Microbiol. Appl. Sci.* 7, 1940–1946. doi: 10.20546/ijcmas.2018.706.230

Simcox, K. D., and Bennetzen, J. L. (1993). The use of molecular markers to study *Setosphaeria turcica* resistance in maize. *Phytopathol*ogy 83, 1326–1330.

Simmonds, N. W. (1985). A plant breeder's perspective of durable resistance. *FAO Plant Prot. Bullet.* 33, 13–17.

Smith, D. R. (1999). "Global disease assessment of corn," in *Proceedings of the fifty-fourth Annual Corn and Sorghum Research Conference,* December 9-10, (Chicago, IL), 54.

Snedecor, G. W., and Cochran, W. G. (1994). *Statistical Methods*, 5th Edn. Ames, IA: Iowa State University Press.

St Clair, D. A. (2010). Quantitative disease resistance and quantitative resistance Loci in breeding. *Annu. Rev. Phytopathol*. 48, 247–268. doi: 10.1146/annurev-phyto-080508-081904

Stuber, C. W., Edwards, M. D., and Wendel, J. F. (1987). Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Sci*. 27, 639–648. doi: 10.2135/cropsci1987.0011183x002700040006x

Su, C., Wang, W., Gong, S., Zuo, J., Li, S., and Xu, S. (2017). High density linkage map construction and mapping of yield trait QTL in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Front. Plant Sci.* 8:706. doi: 10.3389/fpls.2017.00706

Troyer, A. F. (2006). Adaptedness and heterosis in corn and mule hybrids. *Crop Sci.* 46, 528–548. doi: 10.2135/cropsci2005.0065

Tuberosa, R., Salvi, S., Sanguineti, M. C., Landi, P., Maccaferri, M., and Conti, S. (2002). Mapping QTLQTL regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. *Ann. Bot*. 89, 941–963. doi: 10.1093/aob/mcf134

Wang, S., Basten, C. J., and Zeng, Z. B. (2010). *Windows QTL Cartographer 2.5. Department of Statistics*. Raleigh, NC: NC State University.

Ward, J. M. J., Laing, M. D., and Rijkenberg, F. H. J. (1997). Frequency and timing of fungicide applications for the control of gray leaf spot in maize. *Plant Dis*. 81, 141–148.

Welz, H. G., and Geiger, H. H. (2000). Genes for resistance to northern corn leaf blight in diverse maize populations. *Plant Breed.* 119, 1–14. doi: 10.1046/j.1439-0523.2000.00462.x

Welz, H. G., Xia, X. C., Bassetti, P., Melchinger, A. E., and Lubberstedt, T. (1999). QTLQTL for resistance to *Setosphaeria turcica* in an early maturing dent × flint maize population. *Theor. Appl. Genet*. 99, 649–655. doi: 10.1007/s001220051280

Wheeler, B. E. J. (1969). *An Introduction to Plant Diseases*. London: John Wiley and Sons Ltd., 301.

Wisser, R. J., Balint-Kurti, P. J., and Nelson, R. J. (2006). The genetic architecture of disease resistance in maize: a synthesis of published studies. *Phytopathology* 96, 120–129. doi: 10.1094/phyto-96-0120

Xia, H., Gao, W., Qu, J., Dai, L., Gao, Y., Lu, S., et al. (2020). Genetic mapping of northern corn leaf blight-resistant quantitative trait loci in maize. *Medicine* 99, e21326. doi: 10.1097/MD.0000000000021326

Zaitlin, D., DeMars, S., and Gupta, M. (1992). Linkage of a second gene for NCLB resistance to molecular markers in maize. *Maize Genet. Coop. Newsl.* 66, 69–70.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1466.

Zhou, Z., Zhang, C., Zhou, Y., Hao, Z., Wang, Z., Zeng, X., et al. (2016). Genetic dissection of maize plant architecture with an ultra-high density bin map based on recombinant inbred lines. *BMC Genom.* 17:178. doi: 10.1186/s12864-016-2555-z

Zou, G., Zhai, G., Feng, Q., Yan, S., Wang, A., Zhao, Q., et al. (2012). Identification of QTL for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *J. Exp. Bot*. 63, 5451–5462. doi: 10.1093/jxb/ers205

Zwonitzer, J. C., Coles, N. D., Krakowsky, M. D., Arellano, C., Holland, J. B., McMullen, M. D., et al. (2010). Mapping resistance quantitative trait loci for three foliar diseases in a maize recombinant inbred line population-Evidence for multiple disease resistance? *Phytopathology* 100, 72–79. doi: 10.1094/phyto-100-1-0072

Conflict of Interest: AP is employed by company Corteva Agriscience Pvt. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ranganatha, Lohithaswa and Pandravada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

REVIEW
published: 18 February 2021
doi: 10.3389/fgene.2021.600789

Genomics Armed With Diversity Leads the Way in *Brassica* Improvement in a Changing Global Environment

*Nur Shuhadah Mohd Saad, Anita A. Severn-Ellis, Aneeta Pradhan, David Edwards and Jacqueline Batley**

School of Biological Sciences Western Australia and UWA Institute of Agriculture, University of Western Australia, Perth, WA, Australia

OPEN ACCESS

Edited by:
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Raju Datla, Global Institute for Food Security (GIFS), Canada
Andreas Stahl, Julius Kühn-Institut, Germany
Asif Mohammad Iqbal Qureshi, Sher-e-Kashmir University of Agricultural Sciences and Technology, India

***Correspondence:**
Jacqueline Batley
jacqueline.batley@uwa.edu.au

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *31 August 2020*
Accepted: *15 January 2021*
Published: *18 February 2021*

Citation:
Mohd Saad NS, Severn-Ellis AA, Pradhan A, Edwards D and Batley J (2021) Genomics Armed With Diversity Leads the Way in Brassica Improvement in a Changing Global Environment. Front. Genet. 12:600789. doi: 10.3389/fgene.2021.600789

Meeting the needs of a growing world population in the face of imminent climate change is a challenge; breeding of vegetable and oilseed *Brassica* crops is part of the race in meeting these demands. Available genetic diversity constituting the foundation of breeding is essential in plant improvement. Elite varieties, land races, and crop wild species are important resources of useful variation and are available from existing genepools or genebanks. Conservation of diversity in genepools, genebanks, and even the wild is crucial in preventing the loss of variation for future breeding efforts. In addition, the identification of suitable parental lines and alleles is critical in ensuring the development of resilient *Brassica* crops. During the past two decades, an increasing number of high-quality nuclear and organellar *Brassica* genomes have been assembled. Whole-genome re-sequencing and the development of pan-genomes are overcoming the limitations of the single reference genome and provide the basis for further exploration. Genomic and complementary omic tools such as microarrays, transcriptomics, epigenetics, and reverse genetics facilitate the study of crop evolution, breeding histories, and the discovery of loci associated with highly sought-after agronomic traits. Furthermore, in genomic selection, predicted breeding values based on phenotype and genome-wide marker scores allow the preselection of promising genotypes, enhancing genetic gains and substantially quickening the breeding cycle. It is clear that genomics, armed with diversity, is set to lead the way in *Brassica* improvement; however, a multidisciplinary plant breeding approach that includes phenotype = genotype × environment × management interaction will ultimately ensure the selection of resilient *Brassica* varieties ready for climate change.

Keywords: *Brassica*, plant genomics, crop improvement, molecular breeding, food security

INTRODUCTION

Predictions of exponential increases in the world population and climate change are forcing re-evaluation of efforts in addressing the demand for global food security. Plant crops account for more than 80% of the diet consumed by the human population, and the production of edible crops has thereby dominated almost half of the world's available land mass since the beginning of the twenty-first century (Leff et al., 2004; Herrera and Garcia-Bertrand, 2018). The Industrial and Green Revolutions have furthermore shaped the ways that these commodities are managed for efficiency and commercialization. The next revolution will require crop improvement not just to curb world hunger but also to address sustainability in the face of biotic and abiotic stresses triggered by the impending climate change. It is estimated that staple crop yield must improve by 70–110% to feed the predicted 10 billion population by 2050 (Saini et al., 2020). Crop projections and modeling studies have suggested that climate change may have already been responsible for a small yearly decrease in yield and calories in certain geographic regions (Ray et al., 2019). More recently, the coronavirus disease 2019 (COVID-19) pandemic has emphasized food security in terms of short-term and local supply (Cappelli and Cini, 2020). While food shortages are not a wide set concern as yet, a prolonged crisis could interfere with the current complex food supply network (de Paulo Farias and dos Santos Gomes, 2020; Siche, 2020). Globalized food distribution, though highly profitable, highlighted the critical gap in local production, closer to consumers and less likely affected by international restrictions (Cappelli and Cini, 2020). Meeting local population demand, while remaining sustainable in a shorter food supply chain, may yet have introduced another facet to food security in the post-COVID-19 era. Predicted environmental variation including rising temperatures and increases in carbon dioxide emissions could result in a drier atmosphere and an increase in evapotranspiration (Ficklin and Novick, 2017). In addition to water, most biological processes are temperature sensitive, and climate change is therefore undoubtedly going to affect all crop performance (Dusenge et al., 2019). As a result, resilience to abiotic stresses, such as heat, drought, and salinity, will become traits that are highly desirable in future crop improvement strategies. Reactive nitrogen plays an important role in plant growth, crop yield, and subsequently human nutrition (Dreccer et al., 2000). Alongside their benefits, agricultural practices such as nitrogen fertilizer application and nitrogen fixing crops have the potential to disturb the global nitrogen cycle and adversely affect human health (Townsend et al., 2003; Bodirsky et al., 2012). These practices, if poorly managed, can contribute significantly toward the release of nitrous oxide into atmosphere, which negatively affects the protective ozone layer and advances climate change (Crutzen and Ehhalt, 1977). In an effort to solve the global nitrogen challenge, Houlton et al. (2019) propose, amongst other things, the improvement of nitrogen-use efficiency in crop production. This could be achieved by altering fertilizers and fertilizer application practices, boosting soil health to promote nitrogen uptake, and developing improved crop varieties that efficiently utilize nitrogen (Houlton et al., 2019).

The genus *Brassica* consists of extensively agronomically diverse species. Oilseed canola include *Brassica rapa*, *Brassica napus*, and *Brassica juncea* varieties with internationally defined erucic acid and glucosinolate contents (Sharafi et al., 2015). Vegetable Brassicas include *B. rapa* ssp. *rapa* (turnip), ssp. *oleifera* (turnip rape), ssp. *chinensis* (pak choi/bok choy), and ssp. *pekinensis* (Chinese cabbage); *Brassica oleracea* ssp. *capitata* (cabbages), var. *italica* (broccoli), var. *botrytis* (cauliflower), ssp. *gemmifera* (Brussels sprouts), and ssp. *alboglara* (Chinese kale); and *B. napus* var. *napobrassica* (swede/rutabaga) (Cheng et al., 2016). Mustard *Brassica* include *Brassica nigra* (black mustard), *Brassica carinata* (Ethiopian mustard), and *B. juncea* (Indian mustard). The U triangle summarized the interspecific hybridization events between diploid progenitors *B. rapa* (AA), *B. nigra* (BB), and *B. oleracea* (CC) resulting in polyploidy, *B. juncea* (AABB), *B. napus* (AACC), and *B. carinata* (BBCC) (Nagaharu et al., 1935; Snowdon et al., 2002).

Brassicas are cultivated on a worldwide scale, and it is therefore almost a certainty that forecasted environmental changes would affect the crop (Francisco et al., 2017). Combined with the abiotic stresses, biotic stress is another challenge facing global *Brassica* production. Bebber et al. (2013) combined observation data and mathematical equation to project pest distribution and proposed that pathogen and pests affecting global crops are moving polewards as the temperature rises. Pathogens of *Brassica* oilseed and vegetable crops, such as *Leptosphaeria maculans* (blackleg or stem canker), *Alternaria brassicae* (Alternaria blight), *Albugo candida* (white rust), *Pseudocercosporella capsellae* (white leaf spot), *Plasmodiophora brassicae* (club root), and *Sclerotinia sclerotiorum* (Sclerotinia stem rot), extensively affect yield, seed quality, and crop development (Murray and Brennan, 2012). Breeding *Brassica* varieties that can withstand the pressures of a changing environment is perhaps the best strategy in ensuring sustainability.

Decreasing nutrient content of modern fruit and vegetable cultivars has raised concern in recent studies conducted in the USA and UK, predicting the future need for agricultural bio-fortification (Mayer, 1997; Davis et al., 2005; White and Broadley, 2005; Davis, 2009). Davis coined the term "genetic dilution effect" in 2005 after observing broccoli hybrids (*B. oleracea* var. *italica*) accumulating denser heads without the obvious proportional increase in nutrients (Davis et al., 2005). This is a result of breeding and selection of varieties based on yield and productivity, while overseeing the importance of nutrient content. The World Health Organization considers micronutrient deficiency as a major health challenge especially in developing and poor countries (Khush et al., 2012). Bio-fortification is a cost-effective and sustainable strategy in addressing malnutrition; however, this requires suitable genetic diversity within the genepool to be valuable in breeding (Garg et al., 2018; Kumar et al., 2020).

Available genetic diversity still constitutes the foundation of all breeding efforts. Elite varieties, land races, and crop wild species are important resources of useful variation that can be introgressed, re-introduced, or manipulated to obtain the required biotic and abiotic resilience in *Brassica* crops

(Dwivedi et al., 2017). The identification and exploitation of suitable variation are crucial for crop improvement (Hu et al., 2018) and can be elucidated at the genome scale (Varshney et al., 2018). Genomics can, in addition, contribute toward unraveling the genetic origin and molecular pathways involved in biotic and abiotic stress tolerance traits. Complete and accurate understanding of the ancestry of the *Brassica* species will assist in the tracking and exploitation of genetic inheritance of useful traits (Bancroft et al., 2011).

Crop improvement has always been a co-evolutionary process between humans and edible plants (Harlan, 1992); changes in plants brought about by cultivation allowed changes in human populations to take place. Plant breeding has largely relied on conventional breeding methods based on phenotypic selection. However, it is in doubt whether conventional breeding approaches alone would be adequate in addressing the impending challenges. During the course of the last three decades, genomics has become an integral part of all life sciences. Rapid advances in sequencing tools followed by cost reductions, as well as the development of high-throughput genotyping techniques, have led to advances in trait mapping, functional characterization, and ultimately crop improvement through genomic selection (GS) (Nepolean et al., 2018).

Without available genetic diversity, the introduction of genes may present a suitable solution. Weed and insect control using genetically modified (GM) crops has assisted farmers worldwide in attaining higher yields with fewer resources (Zhang et al., 2016). Cauliflower (Lu et al., 2006), mustard (Hong et al., 2002), and canola (Garg et al., 2018) have been subjected to transgenic bio-fortification efforts for beta-carotene, gamma-linoleic acid, and phytate degradation, respectively. Despite these efforts, so far, only Phytaseed canola has been released by BASF in the USA. Health concerns, consumer skepticism, and long and expensive regulatory processes are restricting the release of transgenic crops (Watanabe et al., 2005). Recent developments in genome editing (GE) provide an option to alter or introduce specific genes in order to obtain the desired trait expression. Oligo-directed mutagenesis (ODM), programmable sequence-specific nucleases (SSNs), and base-editing tools allow the precise creation of insertions/deletions (indels) or even the introduction of a complete sequence at a predetermined target location within the genome (Scheben et al., 2017).

To further address the pressing demands on crop improvement, an accelerated rate of genetic gain is required. The implementation of GS can fast-track the progress in crop breeding (Wang X. et al., 2018). In GS, predicted breeding values based on phenotype and genome-wide marker scores allow the preselection of promising genotypes, thereby substantially quickening the breeding cycle and enhancing genetic gains (Heffner et al., 2009). Further optimization of mating strategies is essential to prevent inbreeding and ensure long-term genetic gain (Allier et al., 2019).

Genomics armed with diversity is currently leading the way in crop improvement. Here, we review aspects of the *Brassica* crop improvement cycle (**Figure 1**) illustrating the importance of genetic diversity, creation of genomic resources, its exploitation in aid of trait discovery, and GE and GS of *Brassica*.

ESTIMATION AND CONSERVATION OF DIVERSITY IN GERMPLASM

Crop breeding is required to be sustainable and to adapt promptly in the face of abiotic and biotic environmental changes (Zhang and Batley, 2020). Most modern crops were developed through repeated cycles of selection, "filtering out" varieties with desirable agronomic traits from ancestral wild species. However, potentially valuable genetic variation is often lost in the process, resulting in lowered trait heritability and increased genetic homogeneity (Rahman, 2013). Comparative population genomics can be used to identify these selective sweeps, or bottlenecks, and potential loci under selection (Slatkin, 2008). An example is the allotetraploid *Brassica napus* (AACC, 2n = 38), which originated from spontaneous interspecific hybridization events between *Brassica rapa* (AA, 2n = 20) and *Brassica oleracea* (CC, 2n = 18) (Chalhoub et al., 2014). Cultivated *B. napus* have been under intense selection over the past decades, which has led to a severe genetic bottleneck in the species (Becker et al., 1995). This is reflected by the extent of highly conserved regions found between the genomes of *B. napus* accessions found in almost all major genepools (Werner et al., 2018). Accordingly, the decline of allelic variation and genetic diversity was reported in Canadian (Fu and Gugel, 2010) and Australian spring canola breeding initiatives (Cowling, 2007).

The key to sustainable crop improvement in the face of climate change, and increased pressure from pests and diseases, is maintaining this diversity. Diversity within germplasm provides breeders with valuable material to enable the selection of parental lines, exploitation of heterosis, and the expansion of breeding pools (Yousef et al., 2018). Heritable variation present within the crop germplasm is therefore essential for efficient breeding programs. Characterizing the genetic diversity allows breeders to select alleles at loci of interest and identify trait-associated markers suitable for introgression into new varieties. The assessment of germplasm diversity furthermore assists in the optimization of conservation strategies of germplasm collections (Rao and Hodgkin, 2002). This was highlighted in the study conducted by Yousef et al. (2018), in which cauliflower (*B. oleracea* var. *botrytis*) accessions originating from 26 countries were grouped into two major groups representing the two genebanks from which the accessions were obtained and not the country of origin. In this case, composition and accession type influenced the level of diversity and contributed toward the differentiation between the genebanks. Routine monitoring of genetic diversity in *ex situ* germplasm collections might therefore be essential to prevent potential loss of genetic diversity.

Wild species related to agricultural crops [crop wild relatives or (CWRs)] and landraces offer an attractive alternative source of variation (Dempewolf et al., 2017; Khan et al., 2020). Many of these unique resources are currently available from genebanks and seed repositories (Tanksley and McCouch, 1997). Introgression from CWRs or landraces can broaden the genetic base for modern breeding programs and can contribute to sought-after characteristics associated with relative tolerance to extreme environments and disease resistance. Despite the vast genetic potential locked up in these resources, their utilization

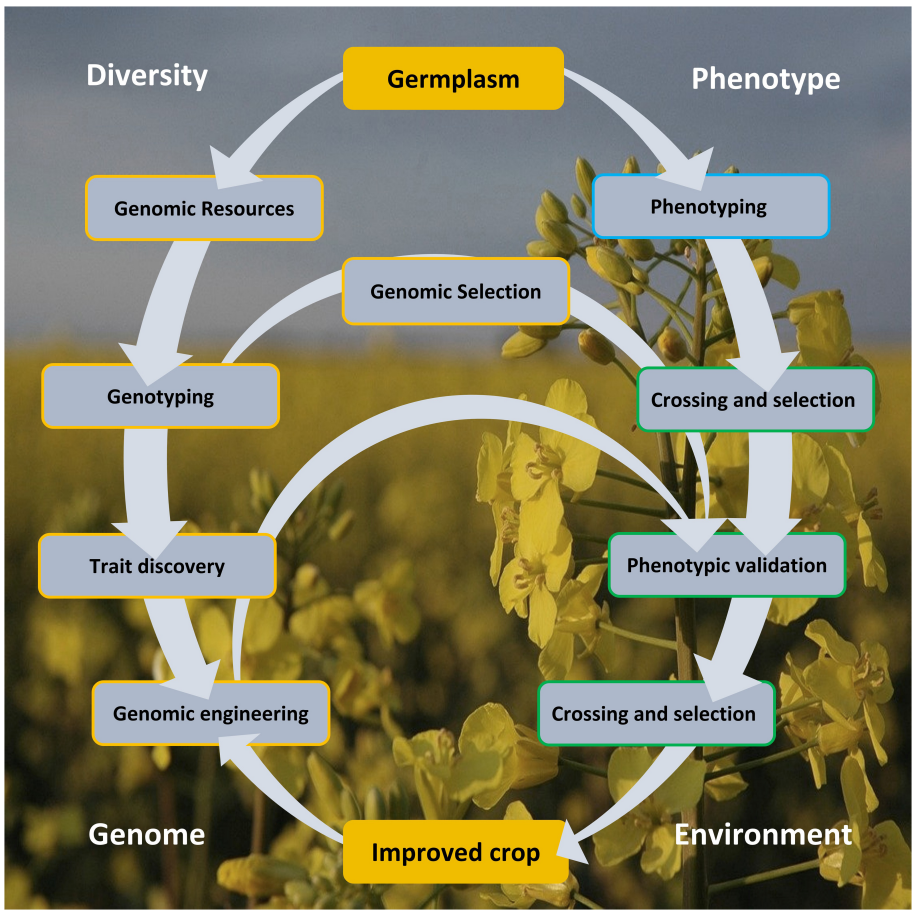


FIGURE 1 | The crop improvement cycle involves the exploitation of genomic diversity and phenotypic variability. These approaches offered a rapid and cost-effective lab-to-field and back loop, centered on arming the genomic arsenal by capitalizing on the variation in the germplasm resource, as defined in the genotype and from trait association. Genomic selection and engineering accelerated crop improvement, circumventing a more laborious and time-consuming conventional crossing and selection approach. Genomic engineering strategies also offer a more precise manipulation for improved crop development. The improved crop must be climate resilient and adaptive to the extreme environmental changes.

is hampered by inconsistent documentation, unintentional duplications, and a lack of available genetic information (Singh N. et al., 2019; Zhang and Batley, 2020). Combining available phenotypic and geographical descriptors with genomic sequence information, perhaps in the form of a universal molecular passport, could facilitate the selection of useful genetic variation and its use in breeding programs (Mascher et al., 2019; Singh N. et al., 2019).

Plant genetic resources, including CWRs in the wild, are considered under threat and in need of conservation (Dempewolf et al., 2017). Climate change, along with habitat fragmentation due to human activity, is predicted to result in drastic plant population declines and even result in extinction (Jump and Peñuelas, 2005). The creation of protected areas, efforts to reduce pollution and legal frameworks to protect endangered species, such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), may aid in the prevention of the decline or loss of biodiversity. Genetic approaches over time have contributed toward the estimation of biodiversity and prioritization of conservation efforts. However, genome-scale data and associated high-density markers can improve estimations of genetic diversity and population structure. Transitioning from a genetics to conservation genomics approach is expected to have a positive impact on future conservation recommendations and policies (Supple and Shapiro, 2018).

CREATION OF GENOMIC RESOURCES IN *BRASSICA* SPECIES

The smaller Brassicaceae genome from *Arabidopsis thaliana* was the first to be sequenced and benefited progress in *Brassica* sequencing given the high degree of genomic conservation between *A. thaliana* and *Brassica* species. In addition, DNA sequencing technology becomes more affordable with longer reads and higher throughput available at a fraction of the cost (Marri et al., 2018). These developments provided the

opportunity to improve *de novo* genome assembly, increase mapping certainty, and identify structural variants (SVs) (Amarasinghe et al., 2020). Furthermore, re-sequencing efforts have increased over time to resolve challenges and gaps in current studies, especially in complex and agronomically significant crops like *Brassica napus*.

As technology progresses, more advanced genomic tools become available, expanding the range of analysis and exploitation possibilities of germplasm resources. The PlabiPD database (https://www.plabipd.de/) maintains an updated list of sequenced plant species, including Brassicas, which can be visualized phylogenetically or temporally. Available genomic resources have played an important role in the advancement of breeding programs throughout the world in cereals, legumes, oilseeds, and even ornamental crops. Hence, here we examine the development and availability of nuclear and organellar genome resources generated in Brassicas.

Nuclear Genome

Draft genomes and pangenome assemblies have been created for five domesticated and a wild *Brassica* species (**Table 1**). More broadly, 27 members of the Brassicaceae family have been sequenced, including three *Arabidopsis* (The Arabidopsis Genome Initiative, 2000; Hu T. T. et al., 2011; Akama et al., 2014; Briskine et al., 2017; Michael et al., 2018), three *Capsella* (Slotte et al., 2013; Kasianov et al., 2017), three *Eutrema* (Yang et al., 2013; Guo et al., 2018), and two *Raphanus* species (Kitashiba et al., 2014; Moghe et al., 2014; Shirasawa et al., 2020). The earlier *Brassica* genome assemblies focused on elite cultivars and combined Illumina paired-end reads with bacterial artificial chromosome (BAC)-end sequences to construct scaffolds and build high-quality assemblies. Markers from genetic maps were used, in addition, to merge assemblies and anchor scaffolds to pseudo-chromosomes. The genomes were either assembled *de novo* or based on a reference genome of the closest relative. The assembly of *B. napus* followed an earlier approach for allopolyploids by sequencing the diploid progenitor genome, similar to the methods used in the assembly of the strawberry and cotton genomes (Tennessen et al., 2014; Ming and Man Wai, 2015; Zhang et al., 2015).

Brozynska et al. (2016) explored the progress of CWR sequencing efforts and noted fewer efforts compared with domesticated relatives, a trend also observed in *Brassica*. Higher levels of heterozygosity in CWRs, which can result in greater assembly difficulties, might be contributing toward the trend (Brozynska et al., 2016). Recently, the genomes of two wild diploid perennial *Brassica* C-genome species were sequenced (Golicz et al., 2016; Kioukis et al., 2020). *Brassica macrocarpa* and *Brassica cretica*, both native to Greece, are potential wild progenitors of *Brassica oleracea*, which was thereby used as reference in the assembly of the two CWRs (Branca et al., 2012).

Next-generation sequencing (NGS) technologies, although extensively and successfully used in genome assembly, are limited by their relatively short read lengths. Shortcomings include misassembly and gaps in long repeat regions, difficulties in detecting larger SVs, transcript isoforms, and haplotype phasing (Van Dijk et al., 2018). Long-read third-generation sequencing (TGS) technologies from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) allow less bias and more homologous coverage of the genome, thereby overcoming the challenges such as polyploidy and frequent repetitive elements. TGS and NGS are often combined with long-range mapping technologies (BioNano Genomics) and chromosome conformation capture (Hi-C) (Van Berkum et al., 2010), to enable the assembly of highly contiguous chromosome-level crop genome assemblies (Hu et al., 2018; Schreiber et al., 2018). Song et al. (2020) created eight high-quality *B. napus* reference genomes by integrating different combinations of Illumina, PacBio, Hi-C, and BioNano data. These high-quality *B. napus* reference genomes allowed the identification of SVs, including copy number variants (CNVs) and presence and absence variations (PAVs), and improve our understanding of the genome structure and genetic basis behind phenotype differentiation in *B. napus*.

Genomic comparison between two *B. napus* reference genomes by Bayer et al. (2017) highlighted the limitations of a single reference assembly, given the wealth of variation between individuals, concluding that the genetic diversity at a species level cannot be sufficiently captured by a single reference genome (Hurgobin and Edwards, 2017). The pangenome was "born" in an effort to capture the totality of diversity in a species or something broader (Vernikos, 2020). *B. oleracea* (Golicz et al., 2016) and *B. napus* (Hurgobin et al., 2018) were the first *Brassica* pangenomes to be published and investigated the diversity within the species. The second *B. napus* pangenome (Song et al., 2020) combined TGS technologies, including PacBio single-molecule real-time (SMRT) and ultralong nanopore sequencing. These pangenome assemblies demonstrated the prospect of uncovering and mining diversity within secondary crop genepools for crop improvement (Voss-Fels and Snowdon, 2016).

Affordability of sequencing is promoting a combination of *de novo* assembly and whole-genome re-sequencing (WGRS) efforts of wider genepools, including close relatives and CWRs in an effort to identify and explore useful genetic variation (Brozynska et al., 2016). While the pangenome represents the genomic makeup of a species (Tettelin et al., 2005), the super-pangenome includes the CWR genetic variability that has been lost due to domestication and breeding selection bottlenecks (Khan et al., 2020). This was well-demonstrated by the discovery of an abundance of unique genes upon the inclusion of the wild *Brassica macrocarpa*, in comparison with the other eight domesticated varieties, in the *B. oleracea* pangenome (Golicz et al., 2016). These genes form part of the pangenome's dispensable genome, and the findings yet again emphasize the value of the genetic variability captured within the wild *Brassica* spp. It is predicted that efforts to catalog and include more lines into pangenomes and super-pangenomes would probably never cease, thereby providing a constant contribution of valuable resources for crop improvement efforts in an ever-changing environment. Future exploration in genomic resources in *Brassica* will likely involve additional WGRS efforts and further pangenome studies to explore breeding histories and identify loci associated with important agronomic traits such as oil content and composition, seed quality, and disease resistance (Wang Y.

TABLE 1 | List of domesticated and wild *Brassica* species nuclear genome assemblies.

Species	Current status	Assembly size (bp)	Number of protein-coding genes	Genomic tools	References
Brassica rapa (2n = 10)	Chinese cabbage ssp. *pekinensis* cv. Chiifu 401–42, *de novo* assembly (version 1.5)	283.80 Mb	41,174	Illumina GAII short, medium, long insert libraries, BAC-end Sanger sequences, EST data, and genetic maps	Wang et al., 2011
	Japanese turnip ssp. *rapa* cv. DH-VT117, reads mapped to *B. rapa* Wang et al. (2011) reference genome	282.93 Mb	40,708	Illumina Hiseq2000 short, medium, long insert libraries	Lin et al., 2014
	Rapid cycling oil-like inbred line cv. RC-144, reads mapped to *B. rapa* Wang et al. (2011) reference genome	282.69 Mb	40,506		
	Chinese cabbage ssp. *pekinensis* cv. Chiifu 401–42, *de novo* assembly (version 2.0/2.5)	389.20 Mb	48,826	Illumina and PacBio sequencing, and genetic maps	Cai et al., 2017
	Chinese cabbage ssp. *pekinensis* cv. Chiifu 401–42, *de novo* assembly (version 3.0)	353.14 Mb	45,985	Illumina HiSeq4000 short reads, PacBio SMRT long reads, BioNano optical mapping, Hi-C sequencing, and genetic map	Zhang L. et al., 2018
	Yellow sarson ssp. *trilocularis* cv. Z1, *de novo* assembly	401.92 Mb	46,721	Illumina HiSeq2500 short, medium, long insert libraries, Oxford Nanopore long reads, BioNano optical mapping, and genetic map	Belser et al., 2018
	Chinese cabbage ssp. *chinensis* cv. NHCC001 (Suzhouqing), *de novo* asssembly	405.33 Mb	48,158	Illumina short reads, PacBio Sequel SMRT long reads, and Hi-C sequencing	Li et al., 2020
Brassica oleracea (2n = 18)	Chinese kale ssp. *capitata* cv. TO1000DH3, *de novo* assembly	488.60 Mb	59,225	Illumina HiSeq2000, Roche 454 GS-FLX Titanium sequencing, and BAC-end sequences, and genetic map	Parkin et al., 2014
	Chinese kale ssp. *capitata* cv. 02-12, *de novo* assembly	539.90 Mb	45,758	Illumina GAII shotgun sequencing combined with Roche 454 GS-FLX Titanium sequencing, BAC-end sequences, and genetic map	Liu et al., 2014
	Pangenome of nine varieties including cabbages, kale, Brussels sprout, Kohlrabi, cauliflowers, broccoli, and wild *Brassica macrocarpa*, with draft genome by Parkin et al. (2014) as reference	587.00 Mb	61,379	Illumina HiSeq2000 and HiSeq2500 short reads	Golicz et al., 2016
	Broccoli ssp. *botrytis italica* cv. HDEM, *de novo* assembly	554.97 Mb	61,279	Illumina HiSeq2500 short, medium, long insert libraries, Oxford Nanopore long reads, BioNano optical mapping, and genetic map	Belser et al., 2018
	Cauliflower var. *botrytis* cv. C-8	584.60 Mb	47,772	Illumina HiSeq and X Ten short reads, and PacBio Sequel SMRT long reads	Sun et al., 2019
	Cabbage var. *capitata* cv. D134 DH (commercially known as Zhonggan 18)	529.92 Mb	44,701	Illumina HiSeq2500 short reads, PacBio Sequel SMRT long reads sequencing, Illumina HiSeq X Ten and GemCode 10x Genomics short reads, and Hi-C sequencing	Lv et al., 2020
Brassica napus (2n = 38)	European winter-type cv. Darmor (version 4.1)	712.30 Mb	101,040	Roche 454-GS-FLX+ Titanium long reads, BAC-end sequences, and Illumina HiSeq short reads	Chalhoub et al., 2014
	European winter-type cv. Darmor (version 8.1)	616.70 Mb	80,382	Illumina HiSeq short, medium, and long insert libraries, and genetic map	Bayer et al., 2017
	European spring-type cv. Tapidor, *de novo* assembly	625.90 Mb	70,162		
	Asian semi-winter cv. ZS11, *de novo* assembly	854.98 Mb	90,731	Illumina HiSeq2500 short, medium, long insert libraries, BAC sequencing, and genetic maps	Sun et al., 2017
	Pangenome of 20 synthetic and 33 non-synthetic accessions based on Darmor v8.1 assembly	1,044 Mb	94,013	Illumina sequencing	Hurgobin et al., 2018
	Asian semi-winter cv. Ningyou 7, *de novo* assembly	994.00 Mb	104,179	Illumina, PacBio RS II, and Hi-C sequencing, and genetic map	Zou et al., 2019
	Pangenome of eight varieties (Westar, No2127, Zheyou7, Gangan, Shengli, Tapidor, Quinta, ZS11) and *de novo* assemblies of each genome	1,001–1,033 Mb	94,586–100,919	Illumina short reads, PacBio SMRT long reads, Hi-C sequencing, and BioNano optical mapping	Song et al., 2020
	European winter-type cv. Express 617	925.00 Mb	89,857	Illumina short reads, PacBio SMRT and Oxford Nanopore long reads, BioNano optical mapping, and genetic map	Lee et al., 2020

(Continued)

TABLE 1 | Continued

Species	Current status	Assembly size (bp)	Number of protein-coding genes	Genomic tools	References
	European winter-type cv. Darmor (version 10)	924.00 Mb	108,190	Illumina HiSeq2500 short reads, PromethION Oxford Nanopore long reads, optical mapping, and genetic map	Rousseau-Gueutin et al., 2020
	Asian semi-winter cv. ZS11	1,160 Mb	106,059	Illumina short reads, PacBio SMRT long reads, Hi-C sequencing, and genetic map	Chen et al., 2020
Brassica juncea (2n = 36)	var. *tumida* cv. T84-66 inbred line, *de novo* assembly	949.60 Mb	80,050	Illumina HiSeq short reads, PacBio SMRT long reads, BioNano optical mapping, and genetic map	Yang et al., 2016a
	cv. Varuna	922.00 Mb	101,959	Illumina HiSeq1000 short reads, PacBio RS II SMRT long reads, BioNano optical mapping, and genetic map	Paritosh et al., 2020
Brassica nigra (2n = 16)	cv. YZ12151 DH, *de novo* assembly	396.90 Mb	49,826	Illumina HiSeq short reads, BAC-end sequences, and genetic map	Yang et al., 2016a
	cv. Sangam (line BnSDH-1)	515.40 Mb	46,227	Illumina HiSeq1000 and MiSeq short reads, Oxford Nanopore long-read sequencing and optical mapping, and genetic map	Paritosh et al., 2020
	cv. Ni100 (short and long reads)	447.00 Mb (SR), 506.00 Mb (LR)	56,331 (SR), 59,877 (LR)	Illumina and Roche 454 short reads, GridION Oxford Nanopore long reads, and genetic map	Perumal et al., 2020
	cv. CN115125 (C2) (long reads)	537.00 Mb	67,030	Illumina HiSeq2500 short reads, MinION Oxford Nanopore long reads, Hi-C sequencing, and genetic map	
Brassica cretica (2n = 18)	Four individuals from isolated Grecian population of subsp. *cretica* and subsp. *nivea* assembled based on related *B. oleracea* Liu et al. (2014) reference	40.02–412.52 Mb	30,360	Illumina HiSeq2500 short reads	Kioukis et al., 2020

et al., 2016; Lu et al., 2019; Dolatabadian et al., 2020; Gabur et al., 2020; Yan et al., 2020; Zhang et al., 2020).

Organelle Genomes

In *Brassica*, the mitochondrial (mt) genome assembly predates that of the whole genome due to its significantly smaller size (∼20 kbp) (Palmer and Herbon, 1988; Handa, 2003; Kode et al., 2005) and high copy number per cell (Lima et al., 2016) (**Table 2**). Due to its small size in comparison with other higher plants, the *Brassica* mt genome was used as an early model in the understanding the plant mt, structure, function, and content (Grewe et al., 2014). Comparative analysis of the mt genome can be used to study interspecific phylogenetic relationships (Darracq et al., 2011) and uncover the plant's evolutionary history (Xue et al., 2020). For example, the mt genomes of *Brassica rapa* subspecies with distinct morphologies were found to be highly conserved (Hatono et al., 2017). Furthermore, the rate of mutation for *Brassica* mt DNA was four times slower than that of the chloroplast (Cp) DNA (Palmer and Herbon, 1988). Xue et al. (2020) compared mt DNA of the six members of the U triangle (Nagaharu et al., 1935) and revealed that *B. oleracea* was undergoing the most mt genomic change, while the B-genomes containing *Brassica carinata* and *Brassica nigra* were identified as maternally more distantly related to the remaining *Brassica* accessions of the U triangle (Xue et al., 2020). It was also suggested that *Sinapis arvensis* could have been misclassified based on both phylogenetic and mt genomic organization, placing it within the *Brassica* species and sister to the *B. nigra*–*B. carinata* lineage (Sang et al., 2020; Xue et al., 2020). Sang et al. (2020) also confirmed that the genome structure and evolutionary analysis of the *S. arvensis* organellar genomes were more similar to those of *B. nigra* and *B. carinata*. In addition, mt genes, encoding putative proteins with transmembrane domains, were discovered, which may explain the alloplasmic male sterility of novel cytoplasmic male sterility (CMS) derived from somatic cell hybridization between *B. napus* and *S. arvensis*.

Cp genomes of most land plants vary between 120 and 160 kbp in size (Wicke et al., 2011). Xiao-Ming et al. (2017) further established that cp gene lengths were proportionally to cp genome size, based on the analysis of 272 species including *B. napus* and 14 other members of the Brassicaceae family. Phylogenetic analysis of the cp genomes of *B. nigra* and *B. oleracea* with those of 10 reported species in the order Brassicales suggested that *B. oleracea* is closely related to *B. rapa* and *B. napus* while *B. nigra* was more diverse than the neighbor species *Raphanus sativus* (Seol et al., 2017). Li et al. (2017) completed the *de novo* assembly of cp genomes of 60 *Brassica* genotypes of the six U triangle species. Subsequent phylogenetic analyses divided the *Brassica* genus into four clades: *B. carinata* and *Brassica juncea*, in accordance with the U-triangle model, shared

TABLE 2 | List of domesticated *Brassica* species organelle genome sequences.

Genome type/species	Data accession	Size (bp)	Number of protein-coding gene	Genomic tools	References
Mitochondrial genome					
Brassica rapa (Chinese cabbage, turnip)	N/A	~218 kb	Mapped 12 genes (61 kb of genome was highly transcribed)	Physical map using *Pst*I, *Sal*I, *Bgl*I, *Kpn*I, and *Nru*I restriction endonucleases and cloning	Palmer and Shields, 1984; Makaroff and Palmer, 1987
B. rapa cv. Purple Top White Globe (turnip)	N/A	~218 kb	N/A	Physical map using *Sal*I and *Pst*I restriction endonucleases and cloning, and comparative mapping with 15 other enzymes	Palmer and Herbon, 1988
B. rapa cv. Suzhouqing	JF920285	219,747	34 (3 rRNA, 18 tRNA)	Roche 454 GS-FLX DNA pyrosequencing, PCR, and Sanger sequencing	Chang et al., 2011
B. rapa var. Oushou hakusai (Chinese cabbage)	AP017996	219,775	34 (3 rRNA, 18 tRNA)	Roche 454 GS-FLX DNA pyrosequencing, PCR, and Sanger sequencing	Hatono et al., 2017
B. rapa var. Chusei shiroguki sensuji kyomizuna (mizuna)	AP017997			Illumina MiSeq, PCR, and Sanger sequencing	
Brassica napus cv. American Purple Top (rutabaga)	N/A	~221 kb	N/A	Physical map using *Sal*I and *Pst*I restriction endonucleases and cloning, and comparative mapping with 15 other enzymes	Palmer and Herbon, 1988
B. napus cv. Westar	AP006444	221,853	34 (3 rRNA, 17 tRNA)	Long range PCR and sequencing	Handa, 2003
B. napus CMS plant (Polima)	FR715249	223,412	34 (3 rRNA, 18 tRNA)	Cycle sequencing, PCR, and Sanger sequencing	Chen et al., 2011
Brassica oleracea (cauliflower)	N/A	~217 kb	N/A	Physical map using *Sal*I, *Kpn*I, and *Bgl*I restriction endonucleases and cloning	Chétrit et al., 1984
B. oleracea cv. Brunswick (cabbage)	N/A	~219 kb	N/A	Physical map using *Sal*I and *Pst*I restriction endonucleases and cloning, and comparative mapping with 15 other enzymes	Palmer and Herbon, 1988
B. oleracea cv. 08C717	JF920286	360,271	56 (4 rRNA, 35 tRNA)	Roche 454 GS-FLX DNA pyrosequencing, PCR, and Sanger sequencing	Chang et al., 2011
B. oleracea var. *botrytis* (cauliflower)	N/A	219,962	33 (3 rRNA, 18 tRNA)	Illumina GAIIx, PCR, and Sanger sequencing	Grewe et al., 2014
B. oleracea var. Fujiwase (cabbage)	AP012988	219,952	34 (3 rRNA, 17tRNA)	Roche 454 GS-FLX DNA pyrosequencing, PCR, and Sanger sequencing	Tanaka et al., 2014
B. oleracea var. *capitata* cv. 4119 (cabbage)	KU831325	219,975	34 (3 rRNA, 19 tRNA)	Illumina HiSeq2000, PCR, and Sanger sequencing	Yang K. et al., 2018
Brassica nigra (black mustard)	N/A	~231 kb	N/A	Physical mapping using BstEII–*Bgl*I and *Pst*I restriction endonucleases and cloning	Palmer and Herbon, 1986
B. nigra cv. Ni-138	AP012989	232,145	33 (3 rRNA, 17 tRNA)	Roche 454 GS-FLX DNA pyrosequencing, PCR, and Sanger sequencing	Yamagishi et al., 2014
B. nigra (black mustard)	KP030753	232,407	33 (3 rRNA, 19 tRNA)	Illumina HiSeq2500, PCR, and Sanger sequencing	Yang et al., 2016b
Brassica juncea cv. Jiangpu-yejiecai	JF920288	219,766	34 (3 rRNA, 18 tRNA)	Roche 454 GS-FLX DNA pyrosequencing, PCR, and Sanger sequencing	Chang et al., 2011
Brassica carinata cv. W29	JF920287	232,241	33 (3 rRNA, 17 tRNA)		
Chloroplast genome					
B. rapa cv. Chiifu-402-41, Z16 and FT	N/A	153,482	89 (4 rRNA, 37 tRNA)	Illumina Solexa sequencing, PCR, and Sanger sequencing	Wu et al., 2012
B. rapa cv. C1.3 (field mustard)	MG717286	153,483	N/A	Illumina HiSeq2500 sequencing	Ferreira de Carvalho et al., 2019
B. rapa cv. Z1 (field mustard)	MG717289	153,502			
B. oleracea cv. C1176	KR233156	153,366	80 (4 rRNA, 30 tRNA)	Illumina HiSeq short-read sequencing	Seol et al., 2017
B. oleracea cv. HDEM (wild cabbage)	MG717287	153,364	N/A	Illumina HiSeq2500 sequencing	Ferreira de Carvalho et al., 2019
B. oleracea cv. RC34 (wild cabbage)	MG717288	153,371			
B. nigra cv. IT119326	KT878383	153,633	80 (4 rRNA, 30 tRNA)	Illumina HiSeq short-read sequencing	Seol et al., 2017
B. napus cv. zy036	GQ861354	152,860	74 (4 rRNA, 30 tRNA)	Illumina Solexa sequencing, PCR and cycle sequencing	Hu Z.-Y. et al., 2011
B. juncea	KT581449	153,483	79 (4 rRNA, 30 tRNA)	Illumina HiSeq2500 sequencing, PCR, and Sanger sequencing	Prabhudas et al., 2016

their cp genome with hybridization donors *B. nigra* and *B. rapa*, respectively. Two types of cp genomes were discovered in *B. rapa*, while the presence of both *B. rapa* cp genomes in *B. napus* strongly suggests two independent hybridization events. These findings were consistent with mt genome findings (Palmer and Herbon, 1988).

Besides photosynthesis, a number of essential metabolic reactions are catalyzed in the cp. These include the biosynthesis of partial amino acids, lipids and fatty acids, vitamins, and isoprenoids, as well as the reduction of nitrites and sulfates (Chen et al., 2018). The cp is furthermore involved in the synthesis defense-related hormones and signaling molecules metabolites associated with disease response and environment changes including heat and light (Lu and Yao, 2018). Genes present in the cp genomes could potentially be explored in *Brassica* to improve yield and resistance to biotic and abiotic stresses. It is clear that the composition and structure of the organellar genome not only hold potential in the elucidation of organellar genome evolution and phylogeny but and understanding thereof may be useful in the identification breeding compatible germplasm resources and CMS and provide opportunities for the introducing of new agronomic and horticultural traits into *Brassica* crops (Daniell et al., 2016). Organellar genome sequences are therefore valuable assets in the future crop improvement efforts.

FUNCTIONAL TRAIT DISCOVERY AND CHARACTERIZATION

The characterization and subsequent exploration of discovered genetic diversity can uncover useful genes linked to adaptation and resistance to abiotic and biotic constraints (Khan et al., 2020). Available *Brassica* reference genomes have provided the foundation to facilitate the fine-scale mapping and elucidation of functionally significant variations in *Brassica* accessions. In this section, we review genome-wide approaches utilized in the identification of alleles associated with desirable traits.

Genome-Wide Single-Nucleotide Polymorphism Discovery Through Whole Genome Re-sequencing

Available reference genomes provide a valuable tool in the study and detection of genetic variation, which can be reliably integrated and reproduced between studies (Malmberg et al., 2018). Alignment and comparison of WGRS data to the reference genome allow the simultaneous detection of large numbers of unbiased genomic single-nucleotide polymorphisms (SNPs), indels, and SVs. WGRS therefore permits a more in-depth interrogation of the genome than complexity reduction methods, resulting in a significant increase in the number of SNP markers detected.

Annotated, high-quality SNPs and SVs set the stage for high-resolution genome-wide association studies (GWASs). Quantitative trait locus (QTL) discovery can lead to the identification of genetic loci and subsequently provide the basis for the functional validation of candidate genes controlling important traits (Lu et al., 2019). Subsequently, by identifying these traits of interest, marker-assisted selection (MAS) could be introduced to advance breeding efforts for easily inherited traits (Harper et al., 2012). For example, in gene pyramiding, multiple genes or major QTLs are introgressed into a single genetic background (Pérez-de-Castro et al., 2012). This approach was recently implemented to introgress multiple genes conveying *Sclerotinia* resistance from wild *Brassica oleracea* into canola (Mei et al., 2020). Furthermore, as long-read sequencing technology advances and improves in accuracy, both PAVs and CNVs are likely to drive the latest source of rich adaptive variation including in crops subjected to biotic and abiotic stresses (Gabur et al., 2018).

In *Brassica* species, WGRS has facilitated the identification of intraspecific and interspecific genetic polymorphisms. Comparative analysis between different *Brassica rapa* morphotypes: Japanese turnip, rapid cycling, and Chinese cabbage cv. Chiifu, revealed, respectively 1,090, 1,118, and 1,464 unique genes in each of the genomes (Lin et al., 2014). Orthologous gene comparison suggested earlier divergence between the three varieties before a more recent domestication event. Gazave et al. (2016) used WGRS to survey the genetic diversity present in a worldwide collection of 782 accessions of *Brassica napus*. A total of 30,881 high-confidence SNP markers were identified, which upon analysis revealed distinct evolutionary histories for the A and C subgenomes. Wu et al. (2019) re-sequenced 991 spring, winter, and semi-winter *B. napus* germplasm accessions, originating from 39 countries. By mapping reads to the "Darmor-bzh" and "Tapidor" reference genomes, a total of 5.56 and 5.53 million SNPs, in addition to 1.86 and 1.92 million indels, were respectively identified. Comparison of SNPs using GWAS revealed a global pattern of genetic polymorphisms as well as paths of allelic drift within the main populations of *B. napus*. Selective sweeps disclosed the genetic basis of divergence between the ecotypes, while SNPs discovered in the promotor regions of FLOWERING LOCUS T and FLOWERING LOCUS C orthologs also corresponded with ecotype groups. Malmberg et al. (2018) furthermore utilized WGRS to develop genomic resources consisting of 4,029,750 high-confidence annotated SNPs with predicted effects, as well as SVs in the form of 10,976 deletions and 2,556 insertions. These valuable genomic resources have the potential to bring together global breeding efforts in the development of locally adapted *B. napus* varieties.

Single-Nucleotide Polymorphism Arrays

Available sequencing data for *Brassica* crops allowed researchers to develop and use high-throughput molecular markers, such as SNPs, more efficiently (Clarke et al., 2013). These markers form an integral part of genomic diversity and, due to their abundance across the plant genome, have become an invaluable tool in crop improvement programs (Scheben et al., 2019). SNP screening can be carried out by WGRS, genotyping-by-sequencing (GBS), or alternatively using SNP arrays. High-density SNP arrays provide an alternative and reproducible genotyping platform, which has been widely used in the characterization of germplasm, GWAS, and QTL studies including the analysis of structural variation (You et al., 2018; Scheben et al., 2019).

A community-driven Brassica 60K (AC genomes) Illumina Infinium™ array (Clarke et al., 2016) was developed and more recently expanded to include the B-genome in the Brassica 90K Illumina Infinium™ array (Scheben et al., 2019). The usefulness of Brassica SNP arrays was demonstrated in the genotyping of resistance genes on chromosome A7 in *B. napus* (Dalton-Morgan et al., 2014), prediction of candidate genes for clubroot disease resistance in *B. napus* (Li et al., 2016), and the assessment of *de novo* homologous recombination events in *B. napus* (Higgins et al., 2018). In addition, several closely linked candidate genes were identified using the 60K Brassica SNP array in the development of functional haplotype markers for the improvement of the oleic acid content in rapeseed (Yao et al., 2020). Several yield associated traits have also been identified using the array platform including branching number (He et al., 2017), ovule numbers (Khan et al., 2019), seed quality (Gajardo et al., 2015), and stem strength (Li H. et al., 2018).

The creation of artificial *Brassica* allohexaploid could potentially result in the development of new oilseed and vegetable crops types with greater inter-subgenomic heterosis. These synthetic tri-genomic hexaploid *Brassica* species are potentially more vigorous and adaptable to a wider range of environmental conditions (Yan et al., 2009; Pradhan et al., 2010; Tian et al., 2010; Geng et al., 2013; Malek et al., 2013; Li et al., 2015; Gupta et al., 2016; Zhou et al., 2016; Mwathi et al., 2020). SNP genotyping was carried out by Gaebelein et al. (2019) using the 90K Brassica SNP array and GWAS to determine the relative impact of genome rearrangement events and inherited allelic variants on meiotic stability. A strong correlation between fertility and meiotic behavior in populations of *Brassica* allohexaploids segregating for alleles from parent allotetraploid species *B. napus*, *Brassica juncea*, and *Brassica carinata* was found. Potential genes of interest were subsequently identified for further investigation into meiotic regulation and future establishment of stable A, B, and C allohexaploids (Gaebelein et al., 2019).

Although SNP arrays can provide vital data to breeders and researchers, public accessibility to genotypes identified can be limited due to the lack of public repositories or databases designed to host crop SNP array data. To address this constraint, Scheben et al. (2019) established the CropSNP database (http://snpdb.appliedbioinformatics.com.au) for SNP array data generated on the Illumina Infinium™ Brassica 60 and 90K array platforms.

Transcriptomics

One of the key challenges in genomics-based breeding remains the complex linking of genotype to phenotype across tissue types, developmental stages, and environmental conditions. Transcriptomics, although part of associated "omics," have emerged as an exceptional tool in the functional inference of genetic variability (Wang et al., 2019) with technological innovation constantly advancing the field (Lowe et al., 2017; Wang et al., 2019).

RNA-Seq technology has been extensively used in the mapping of exon/intron boundaries, improvement of genome annotations, and the detection of rare transcripts and splicing variants (Pérez-de-Castro et al., 2012). For example, the transcriptional regulation of anthocyanin biosynthesis in *B. juncea* was studied to identify differentially expressed genes between the purple and green leaves from a backcrossed BC3 segregation population. Genes associated with phenylpropanoid biosynthesis, phenylalanine metabolism, and flavonoid biosynthesis were differentially expressed, while genes involved with anthocyanin biosynthesis (*BjTT8* and *BjMYC2*) were up-regulated in the purple leaves. Understanding anthocyanin biosynthesis and its regulatory network in *Brassica* is a prerequisite in the development of health-promoting anthocyanin-rich vegetables (Heng et al., 2020).

The complex defense response between a *Sclerotinia sclerotiorum* resistant and susceptible line of *B. napus* was analyzed in a study by Wu et al. (2016). Dynamic transcriptome analyses uncovered differences between susceptibility and resistance associated with the magnitude of expression changes in genes involved in pathogen recognition, MAPK signaling cascade, WRKY transcription regulation, jasmonic acid/ethylene signaling pathways, and biosynthesis of defense-related protein and indolic glucosinolate. Valuable insights gained will assist in the development of effective strategies in *Sclerotinia*-resistance breeding (Lu et al., 2014; Wu et al., 2016).

Limitations such as misassembly, associated with short read RNA-Seq, hampers the full-length assembly of transcripts from highly repetitive regions or analogous gene families. These difficulties are often more pronounced in polyploid plants. TGS technologies such as SMRT by PacBio and Oxford Nanopore single molecule structure sequencing (SMS-seq) provide an opportunity to construct full-length transcripts, with the possibility to capture structural variations, tertiary interactions, and the dynamics of riboswitch ligand binding (Bizuayehu et al., 2020). Full-length transcriptome sequencing (PacBio) was used by Tan et al. (2019) to explore the transcript and splice isoforms expressed during anther development in Chinese cabbage (*B. rapa* ssp. *pekinensis*). In addition to predicted fusion transcripts and poly-A sites, 53 key genes active during anther development were detected, of which eight annotated loci had alternatively spliced isoforms. The transcripts generated provided a valuable resource for the characterization of anther-specific gene expression and improved Chinese cabbage genome annotation (Tan et al., 2019).

An et al. (2019) conducted a comprehensive study comparing the genetic diversity of 183 *B. napus*, 112 *B. rapa*, and 42 *B. oleracea* accessions, along with 20 wild relatives (*Brassica hilarionis*, *Brassica villosa*, *Brassica montana*, *Brassica macrocarpa*, *Brassica rupestris*, *Brassica incana*, and *Brassica insularis*) and five other Brassicaceae species as outgroups in order to improve the understanding in the origin and diversification of *B. napus*. RNA-Seq reads generated from *B. rapa* accessions and *B. oleracea* and other wild C genome species were respectively mapped to the A and C genomes of the *B. napus* Darmor-bzh reference genome to identify SNPs. Six genetic clusters of *B. napus* were identified, which were shown to have undergone different selective pressures in accordance with known breeding histories. Although the multi-origin of *B. napus* remained elusive, the study contributed toward the identification of putative candidate genes to important agronomic traits, which,

along with high-quality SNPs identified, have the potential to facilitate rapeseed improvement and germplasm preservation.

In line with the pangenomics approach, He et al. (2015) assembled the first pan-transcriptome resources for the *Brassica* A and C genomes. The pan transcriptome was established using existing coding DNA sequence (CDS) gene models from the *B. oleracea* TO1000 and *B. napus* Darmor-bzh reference genomes in addition to preliminary CDS models from the *B. rapa* Chiifu genome sequence assembly. The construction of the pan-transcriptome allows, in a similar fashion as the pangenome, the discovery of functional dispensable genes (Jin et al., 2016).

Epi-Genomics

In addition to the identification of novel genes and useful haplotypes, epigenetic variation has the potential to contribute toward crop adaptation and productivity (Dwivedi et al., 2017). These epigenetic variations are plant developmental and adaptation responses to environmental constraints (Gallusci et al., 2017; Tirnaz and Batley, 2019b). Epigenomics encapsulates genotype × environment interactions and their independent influence on the phenotype (Seymour and Becker, 2017) and may contribute as potential phenotypic resources for breeding.

Epigenetic regulation is independent of DNA sequence alteration and stably inherited during mitosis or meiosis (Weigel and Colot, 2012). DNA (de)methylation, histone modification, and chromatin remodeling involves regulatory reprogramming at transcriptional and post-transcriptional levels (Paszkowski and Whitham, 2001; Tirnaz and Batley, 2019a). DNA methylation in plants includes *de novo* methylation, as well as maintenance and demethylation, as a means of regulatory check and balance in gene expression (Elhamamsy, 2016; Tirnaz and Batley, 2019a). Histone modification refers to the methylation or (de)acetylation of histone proteins at the N-terminal. Mono-, di-, or trimethylation of the lysine residue in the former results in various functional responses, while the histone acetylation and deacetylation are associated with gene activation and repression, respectively (Fuchs et al., 2006). Additionally, chromatin remodeling is influenced by histone octomer movement (Perrella and Kaiserli, 2016), ATP-dependent enzyme affecting nucleosome composition (Tariq and Paszkowski, 2004) or histone variants (Rando and Ahmad, 2007) cause the DNA sequence to become inaccessible to transcriptional mechanisms, resulting in transcriptional silencing.

Epialleles, or epivariants, first coined in mammals by Rakyan et al. (2002), refer to genetically identical and stable alleles that are variably expressed due to epigenetic regulations and result in difference in phenotype (Richards, 2006; Dolinoy et al., 2010). Earlier studies on epialleles lack whole-genome information to decipher association with desirable traits (Seymour and Becker, 2017). While more studies have been conducted recently, epigenomics is still considered in its infancy compared with the characterization of genetic sequence- or structural-caused variation.

Epigenomic studies in *Brassica* crops have predominantly involved global or targeted methylation profiling for natural or experimentally induced epivariants as defined by Gallusci et al. (2017). Early *Brassica* methylation studies relied on a chemical demethylation agent treatment using 5-azacytidine (5-AzaC) and on cytological (Solís et al., 2015) as well as targeted allele-specific and methylation-sensitive amplified polymorphism (MSAP) techniques to capture the global DNA methylation pattern (Shiba et al., 2006; Hauben et al., 2009) (refer **Table 3**). Hypomethylated populations created using 5-AzaC treatment could also be mined for epiallelic variation (Amoah et al., 2012). The forward and reverse screening of epigenetic variation was employed in functional and inheritance studies in *B. rapa* var. *trilocularis* and suggested to have potential as an intervention strategy for crop improvement (Amoah et al., 2012).

These techniques were superseded by whole-genome bisulfite sequencing (WGBS) first described in *Arabidopsis* (Cokus et al., 2008; Lister et al., 2008). With the use of the WGBS approach, DNA hypomethylation of the multiallelic *Bnams4* gene associated with male sterility was detected in young floral buds (Wang Z. et al., 2018). The male sterility trait promotes heterosis and hybrid development and hence is favored in crop improvement and breeding strategies (Saxena and Hingane, 2015).

Chromatin immunoprecipitation (ChIP) is another useful strategy for protein–gene interaction studies and, in the case of methylation, for investigating specific histone modifications (Das et al., 2004; Kawanabe et al., 2016). Kawanabe et al. (2016) developed positive and negative control primers to validate ChIP assays. The primers were targeted at histone modifications at H3K4me3 (trimethylation of the 4th lysine of H3), H3K9me2, H3K27me3, and H3K36me3 and used to study the response of 4 *FLC* paralogs to vernalization in *B. rapa* var. *pekinensis*. ChIP combined with WGBS further expanded the ability to investigate complex interactions between genetic and epigenetic factors (Li and Tollefsbol, 2011). These studies assisted in the identification of epigenetic markers, epigenetic QTL, and genes associated with floral and pollen development, self-incompatibility, salt and heat stress, vernalization, disease resistance, and male sterility and in the assessment of methylation profiles of introgression lines for crop improvement resource (**Table 3**). Nonetheless, due to the complexity of epigenetic interaction and its involvement in complex regulatory networks, these epigenomic approaches still require extensive investigation before their application in crop improvement can be implemented.

Reverse Genetics

One of the most cost-effective and quick approaches used to identify genetic variation in crop populations is by Targeting Induced Local Lesions in Genomes (TILLING). This method combines chemical mutagenesis, to create lesions on the genome, and molecular techniques such as PCR and DNA pooling, to identify point mutations within the population. TILLING was first demonstrated in *Arabidopsis* using ethyl methanesulfonate (EMS) (McCallum et al., 2000), which results in base transitions by causing the G residues to alkylate and as a result pair with T instead of C (Rashid et al., 2011). Soon after, the technique was explored in *Brassica* species for various traits of interest including fatty acid content and shatter tolerance (**Table 4**). However, this method is less feasible and becomes more time-consuming if multiple genes are targeted, such as those involved in specific

TABLE 3 | Exploration of epigenetic resources in *Brassica* for natural and experimentally induced epivariants.

Approach	*Brassica* species	Trait of interest	Source of epigenetic variation	References
Allele-specific bisulfite sequencing	*Brassica rapa*	Self-incompatibility	Natural	Shiba et al., 2006
Methylation-sensitive AFLP, cytosine extension assay	*Brassica napus*	Energy use efficiency	Experimental	Hauben et al., 2009
MSAP	*B. rapa* cv. *Osome*	Chromatin remodeling	Experimental	Sasaki et al., 2011
MSAP	*B. rapa* var. *trilocularis* cv. R-o-18	Hypomethylated epialleles	Experimental	Amoah et al., 2012
WGBS using NGS	*B. napus* var. *oleifera*	Salt stress	Natural	Marconi et al., 2013
WGBS using NGS	*Brassica oleracea* cv. TO1000	Cystosine methylation	Natural	Parkin et al., 2014
Reduced representation bisulfite sequencing (RBBS)	*B. rapa* var. *oleifera* cv. 3H120	Hybridization and polyploidy	Natural	Chen et al., 2015
ChIP	*B. rapa* var. *pekinensis* cv. RJKB-T23 and RJKB-T24	Vernalization, high bolting	Natural	Kawanabe et al., 2016
MSAP	*B. oleracea* var. *botrytis* cv. Korso × *Brassica nigra* cv. G1/1 introgression lines	Methylation pattern and heredity	Natural	Wang G.-X. et al., 2016
WGBS using NGS	*B. oleracea* cv. 01-20S and 01-20F	Male sterility	Natural	Han et al., 2017
Methylated DNA immunoprecipitation sequencing (MeDIP-seq)	*B. rapa* ssp. *chinensis* cv. NHCC004	Heat tolerance	Experimental	Liu T. et al., 2017
WGBS using NGS	*B. rapa* cv. Suzhouqing (NHCC001)	Heat stress	Natural	Liu G. et al., 2018
Methylated DNA immunoprecipitation sequencing (MeDIP-seq)	*B. rapa* cv. RJKB-T23 and RJKB-T24	Fusarium disease resistance	Natural	Takahashi et al., 2018a
WGBS using NGS, small RNA-Seq, and ChIP	*B. rapa* var. *pekinensis* cv. RJKB-T24	DNA methylation and expression, and H3K9me2 modification and small RNA expression	Natural	Takahashi et al., 2018b
WGBS using NGS	*B. napus* cv. 7365A and 7365B	Male sterility	Natural	Wang Z. et al., 2018
ChIP sequencing (ChIP-seq) and RNA-Seq	*B. rapa* var. *pekinensis* RJKB-T23 and RJKB-T24 inbred lines	H3K27me3 in *FLC* regulation and vernalization	Natural	Akter et al., 2019
ChIP sequencing (ChIP-seq) and RNA-Seq	*B. rapa* ssp. *trilocularis* cv. R-o-18	H3K27me3 in epigenetic regulation	Natural	Payá-Milans et al., 2019
ChIP sequencing (ChIP-seq) and RNA-Seq	*B. rapa* ssp. *chinensis* cv. *Aijiaohuang*	Pollen wall formation	Experimental	Shen et al., 2019
Targeted enrichment using seq-cap epi	*B. napus* cv. Westar and Sturt	Resistant to *Leptosphaeria maculans*	Natural	Tirnaz et al., 2020
WGBS using ONT	*B. nigra* cv. C2 and Ni100	Active centromeres	Natural	Perumal et al., 2020

AFLP, amplification fragment length polymorphism; MSAP, methylation-sensitive amplified polymorphism; WGBS, whole-genome bisulfite sequencing; NGS, next-generation sequencing; ChIP, chromatin immunoprecipitation; ONT, Oxford Nanopore Technologies.

biopathways (Sashidhar et al., 2020). Examples of canola varieties developed that benefit from this approach include PodGuard trait for In Vigor R5520P and 1H51 RR varieties, which were commercially marketed by Bayer for their resistance to pod shattering (Raman et al., 2019).

TILLING has been further developed to include EcoTILLING (Ecotype) and ORG-EcoTILLING. EcoTILLING involves pooling DNA from only two individuals consisting of a reference and queried genotypes (Backes, 2013); therefore, instead of creating mutant populations, this more recent approach involves a study of allelic variation. EcoTILLING requires re-sequencing efforts to characterize and locate the genotypic polymorphism (Wang et al., 2010). Additionally, this approach is high-throughput and can associate natural variants with gene function, trait association, and phylogenetic relationships (Zeng et al., 2012). ORG-EcoTILLING was first utilized in Brassicas to explore the use of TILLING in organelle genomes by combining CEL1 endonuclease, which cuts specific mismatches in heteroplex DNA, and PCR, for three cp genes and one mt gene. ORG-EcoTILLING consistently confirmed *B. rapa* as *B. napus* maternal progenitors based on phylogenetic analysis. Additionally, it also uncovered the possibility of multiple origins and evolution throughout *B. napus* domestication with the identification of three additional divergences in the accessions.

The TILLING approach is considerably more accessible in terms of cost and time than other specialized reverse genetics approaches like RNA interference (RNAi) or gene silencing, and virus-induced gene silencing (VIGS). RNAi gained traction with the ability to knock down gene functions mediated by small interfering RNA (siRNA) or microRNA (miRNA) (Pe'ery et al., 2003; Limera et al., 2017). Traits introduced using RNAi, such as self-compatibility in *B. rapa*, have been found to be stable even in crosses with commercial variety (Jung et al., 2012). This will help improve likelihood for the seeds to be used in

TABLE 4 | Reverse genetic resource in *Brassica* over the last 10 years.

Approach/Species	Trait of interest (target gene(s))	References
RNAi		
B. rapa cv. Osome	Self-compatibility S-locus (*SRK, SP11/SCR*)	Jung et al., 2012
B. juncea cv. Varuna	Glucosinolate transcriptional regulator (*MYB28*)	Augustine et al., 2013
TILLING		
B. napus cv. Ningyou 7	Seed erucic acid content (*FAE1*)	Wang et al., 2008
B. napus cv. YN01-429 x Express 617	Sinapine synthesis (*SGT, REF1*)	Harloff et al., 2012
B. napus cv. DH12075	Demonstrate mutation density (39 genes)	Gilchrist et al., 2013
B. napus cv. Express 617	Phytic acid (six gene families: *MIPS, MIK, 2-PGK, IPK1, IPK2*, and *MRP5*)	Sashidhar et al., 2019
Three *B. rapa* ssp. *trilocularis* cv. R-o-18 mutants	CAX1 transporter for Calcium homeostasis (*BraA.cax1a*)	Navarro-León et al., 2018
B. oleracea cv. TO1000	Abiotic stress response (15 genes)	Himelblau et al., 2009
B. rapa ssp. *trilocularis* cv. R-o-18	Redundant genes function (three *REPLUMLESS*, one *INDEHISCENT*, two *METHYLTRANSFERASE1*)	Stephenson et al., 2010
EcoTILLING		
101 *B. napus*, nine *B. oleracea* and seven *B. rapa* accessions	Seed erucic acid content (*FAE1*)	Wang et al., 2010
ORG-EcoTILLING		
90 *B. napus*, three *B. rapa*, and three *B. oleracea* accessions	Breeding history and maternal inheritance (*accD*, *matK*, *rbcL*)	Zeng et al., 2012
VIGS		
B. napus (winter varieties)	Flowering (*FLC*)	Álvarez-Venegas et al., 2011
B. rapa ssp. *Chinensis* cv. Wuyueman and 49 Caixin	Flowering (*MAF*)	Huang et al., 2018
B. rapa cv. Sijiucaixin	Carotenoid biosynthetic pathway (*PDS*)	Yu et al., 2018

commercial cultivation. Similar to RNAi, VIGS also involves introduction of dsRNA molecules. A viral vector genome such as *Cabbage Leaf Curl Virus* (CaLCuV) is modified to include the plant target gene fragment (150–800 bp) and to remove the viral inducing host gene, thus forming a recombinant virus (Lu et al., 2003; Ramegowda et al., 2014; Bekele et al., 2019). The recombinant vector introduces infection in the plant, is amplified, and generates dsRNA molecules. These dsRNA molecules triggered post transcriptional gene silencing, once detected by the host plant, causing it to be cleaved into siRNA. The RNAi silencing complex and antisense siRNA strands associates together and begin to target RNAs, which complemented the siRNAs. These target-specific RNAs were screened and destroyed, which subsequently caused the target gene to be silenced. VIGS can also be utilized for tissue-specific gene silencing, helpful in screening stress responses, and induced transcriptional gene silencing by targeting the gene promoter (Kanazawa et al., 2011; Senthil-Kumar and Mysore, 2011; Bekele et al., 2019). VIGS research in *Brassica* includes interest in understanding the vernalization pathway in Brassicas as a means to improve flowering and reproductive development by silencing genes associated with late flowering (Álvarez-Venegas et al., 2011) or floral organ transition (Huang et al., 2018).

GENOME MANIPULATION

Traditional breeding approaches rely on the diversity found in local land races, mutation panels, or even germplasm from related species to introduce desired traits or elite alleles through costly and time-consuming backcrossing programs (Dwivedi et al., 2017). The absence of natural genetic diversity and potential linkage drag introducing closely linked unwanted agronomic characteristics has plagued crop improvement efforts (Holme et al., 2013). Increased availability of genomic resources, identified and well-characterized genes, as well as a deeper understanding of underlying molecular mechanisms has led the way for the introduction of innovative approaches to overcome these limitations and fast-track crop breeding (Scheben et al., 2016: Hickey et al., 2019).

Gene Transformation

Genetic modification through the introduction of transgenes was developed in an attempt to expand the available genepool (Kamthan et al., 2016). *Agrobacterium tumefaciens* and biolistic techniques are widely and efficiently employed to mediate the transfer of selected exogenous genes or regulatory elements from an unrelated species or even non-plant organism (Moloney et al., 1989; Altpeter et al., 2016). Numerous transgenic *Brassica* spp. have been developed in an attempt to introduce traits such as salt tolerance (Kim et al., 2016), disease resistance (Grison et al., 1996; Aghazadeh et al., 2016; Zarinpanjeh et al., 2016), reduced sinapine content (Wolfram et al., 2010; Harloff et al., 2012), and herbicide tolerance (Beversdorf et al., 1988; De Block et al., 1989; Cuthbert et al., 2001). The potential of transgenics in advanced plant metabolic engineering is however best demonstrated in the development of transgenic omega-3 *Brassica napus* varieties by BASF and Cargill, and Nuseed, CSIRO and GRDC (Napier et al., 2019) respectively, as well as *Brassica juncea* by Wu et al. (2005).

Transgenic *B. napus* accumulating long-chain polyunsaturated fatty acids (LC-PUFAs) were engineered by each of the initiatives through the introduction of large multi-transgene cassettes. The BASF cassette (∼44 kbp) contained 12 genes, while the Nuseed initiative's cassette (∼23 kbp) contained six omega-3 LC-PUFA biosynthetic genes. Each gene was under the regulation of a seed-specific promoter. Each cassette also contained a gene for herbicide resistance (Connelly and MacIntosh, 2018; Sottosanto et al., 2018). The LC-PUFA profile produced by the two omega-3 transgenic canola varieties varied and respectively contained ∼7% eicosapentaenoic acid (EPA), ∼3% docosapentaenoic acid (DPA), and ∼1% docosahexaenoic acid (DHA) (LBFLFK); and <0.5% EPA, ∼1% DPA, and ∼10% DHA (NS-B50027-4) respectively (Napier et al., 2019). Very-long-chain (VLC) PUFA accumulating *B. juncea* was on the other hand engineered in a stepwise approach through a series of transformations with increasing numbers of transgenes. The resulting transgenic *B. juncea* yielded up to ∼15% EPA, ∼4% DPA, and ∼1.5% DHA (Wu et al., 2005).

In addition to the transformation of the nuclear genome, modification of the cp genome has also been established. The introduction of foreign genes into the cp can address nuclear transgenic limitations such as low-level transgene expression (Jin and Daniell, 2015) and the potential transgene escape via pollen (Daniell, 2007). The stable integration and expression of more than 40 cp-based transgenes has been reported by Daniell et al. (2016), most of which were aimed at potentially enhancing biotic stress tolerance and consequently yield. An example is the development of a cabbage-plastid transformation system for introduction of the insecticidal *cry1Ab* gene. With the use of a species-specific vector, the expression of the BT-toxin facilitated the control of the diamond moth, an economically important *Brassica* pest (Liu C. W. et al., 2008). Introduced insect resistance using *cry* genes was also reported for rapeseed (Schuler et al., 2004) and collards (Cao et al., 2005).

Improved weed and insect control using GM crops has assisted farmers in attaining higher yields with fewer resources. The analysis by Brookes and Barfoot (2018) estimated that the adoption of commercial herbicide-tolerant GM canola and sugar beet led to a total global income gain of $559 million in 2016 and $6.44 billion cumulatively since 1996. Despite the commercial benefits, growing consumer skepticism (Frewer et al., 2013), as well as potential risks to human health and the environment (Zhang et al., 2016), has encouraged the development of alternative genomic modification technologies.

Cisgenesis and intragenesis, based on the same gene transfer technologies as transgenesis, were consequently developed. Genetic crop modification thereby involved the introduction of target DNA from the same plant species, or a sexually compatible species, for crop improvement. Resulting crop plants are free from any foreign DNA mitigating associated risk to some extent (Espinoza et al., 2013; Holme et al., 2013). Alternative approaches such as reverse breeding exploit the use of transgenes to accelerate initial breeding. The unwanted transgene is eliminated through Mendelian segregation during the later stages of the breeding process (Dirks et al., 2009; Basso et al., 2020).

Genome Editing

Recently, GE technologies have come to the foreground, allowing the precise and permanent modification of specific genes or genomic regions. ODM, programmable SSNs, and base editing provide an opportunity to study gene function and alter crop traits through the mutation of specific genes, reprogramming of epigenetic markers, and the generation of site-specific sequence modifications (Voytas and Gao, 2014; Ran et al., 2017; Jansing et al., 2019).

Oligo-Directed Mutagenesis

Traditional mutation breeding using chemical or irradiation results in random mutations in the genome. The movement toward a more desired and controlled site-specific targeted mutagenesis took shape in the 1970s (Lusser and Davies, 2013). ODM, also known as targeted gene repair, oligonucleotide-directed gene targeting, genoplasty, and chimeraplasty, makes use of 20–100 bp of DNA or RNA oligonucleotides to introduce mutations at the target site. The synthesized oligonucleotides are designed to be homologous to the target site with the exception of 1–4 bp (Lusser et al., 2012). Upon transfection, the oligonucleotides associate with the target site, prompting DNA repair at the sequence mismatch sites resulting in base pair mutations, deletions, or reversal of mutations (Lusser et al., 2012; Lusser and Davies, 2013). With a difference of often just a few nucleotides underlying important traits in plants, the application of ODM held potential as a non-GM organism (GMO) base pair-specific oligonucleotide-directed gene editing platform to augment the genetic diversity of a specific genotype. The use of ODM was furthered by Cibus as part of the commercial Rapid Trait Development System (RTDS^{TM}) introducing novel and commercially valuable traits such as herbicide resistance into a variety of crops including oilseed rape (Gocal et al., 2015). Ruiter et al. (2003) found that spontaneous mutation in plants obscured the intended sequence modifications in *B. napus* mediated through self-complementary RNA–DNA chimeric oligonucleotides or chimeraplasty. Studies by Sauer et al. (2016) confirmed that significant precise gene-editing events in plants could be realized by ODM alone and suggested that ODM efficiency could be further improved in combination with reagents that cause DNA double-stranded breaks (DSBs).

Programmable Sequence-Specific Nuclease

GE using programmable SSNs is generally achieved through the induction of a controlled DSB at a target locus using SSNs. The DBS activates the intracellular DNA-repair pathways and is repaired through either non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Gaj et al., 2013). The imprecise re-joining of the DBS through NHEJ leads to the introduction of indels at the target loci and disruption of gene function. On the contrary, HDR entails the use of an exogenous DNA-repair template to bridge the DSB site. The repair template, a double-stranded DNA vector or a single-stranded DNA oligonucleotide, enables the introduction of a precise mutation or insertion to alter gene function (Zhang et al., 2013). Several engineered nuclease systems have been developed including meganucleases (MN), zinc finger nucleases (ZFNs), transcription activator-like

effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR) coupled with a CRISPR-associated protein (Cas).

Zinc Finger Nucleases

ZFN, first described by Kim et al. (1996), employs novel hybrid site-specific endonucleases created by the linking of two different zinc finger proteins (ZFPs) to the cleavage domain of the bacterial FokI endonuclease. The zinc finger domains are designed to each recognize and bind to a unique 3- to 4-bp DNA sequence adjacent to the target site. The tandem repeats can be constructed to recognize an extended 9- to 18-bp DNA sequence (Lin and Musunuru, 2016). In an attempt to modify seed oil composition in *B. napus*, Gupta et al. (2012) engineered ZFP transcription factors (TFs) to fuse to a conserved region downstream of the transcription start site of two canola KASII genes. The modification resulted in the escalated expression of the KASII mRNA, increasing C18 and lowering palmitic acid levels well as the overall saturated fatty acid content in the seed. Canola oil with a lower saturated fat content is more desirable and potentially poses health benefits (Hyseni et al., 2017). Despite the progress in technology, the engineering of the desired ZFN-binding domain remains challenging and time-consuming, with further limitations presenting in selection of the target site (Cox et al., 2015).

Transcription Activator-Like Effector Nucleases

TALENs are similar in structure to ZFNs, composed of dimeric DNA-binding proteins fused to the nuclease domain FokI (Cermak et al., 2011). The central domain of the TAL effector consists of a polymorphic repeat of ∼34 amino acids with hypervariable di-amino acids at positions 12 and 13. These so-called repeat variable di-residues (RVDs) associate and recognize a corresponding C, T, A, or G nucleotide (Scholze and Boch, 2010). The longer DNA recognition sites promote specificity and reduce potential off-target effects (Li et al., 2012). The TAL effector DNA binding domain is more flexible and can be customized, which broadens its potential application.

Sun et al. (2013) demonstrated the suitability of the TALEN construct to alter the endogenous FRIGIDA (vernalization determinant) gene in *Brassica oleracea* var. *capitata*, further suggesting that the method could be applied to related *Brassica* spp. TALENs with mt localization signals (mitoTALENs) were designed by Kazama et al. (2019) to knock out CMS-associated genes at orf79 and orf125, respectively, of CMS varieties of rice and *B. napus* (SW18). Induced mt modifications restored fertility without causing noticeable phenotypic changes and were found to be stable and maternally inherited. The successful modifications of the mt genome pose the prospect of "mitochondrial breeding" in plants (Kazama et al., 2019), which can play an important role in the study and future conditioning of plant responses toward climate change (Budar and Roux, 2011; Sweetman et al., 2019; Florez-Sarasa et al., 2020). TALENs similar to ZFNs are time-consuming genome manipulation techniques (Razzaq et al., 2019) and require the extensive screening of large numbers of manipulated individuals.

RNA-Guided Nucleases

The second-generation CRISPR/Cas9 system provides an alternative approach in targeted nucleases. In contrast to the ZFN's and TALEN's engineered protein associated DNA-binding systems, the CRISPR/Cas9 system has a single-guide RNA (sgRNA) bound to the Cas9 endonuclease that directs the complex to a specific site in the genome (Jinek et al., 2012; Cong et al., 2013). Recognition is attained through base pairing between the programmable 20-bp-long spacer region at the 5′-end leading sequence of the gRNA and specific DNA target. The Cas9 nuclease uses the CRISPR gRNA–DNA pairing as guide in combination with adjacent the DNA protospacer-adjacent motif (PAM) to cleave the DNA. This simplicity and flexibility in the programming of the CRISPR/Cas9 system have facilitated its adoption and exploitation in plants including *Brassica* (Zhang et al., 2018b). Lawrenson et al. (2015) demonstrated the efficiency of the CRISPR/Cas9 GE tool for the first time in the knockout of target genes in *B. oleracea*. Introduced mutations were stably inherited and transgene-free plants obtained through segregation.

Due to the allotetraploid nature of some *Brassica* species, the observed effect of single gene modification is often limited by its potential redundant function. Modification of all homologous genes is hence required to obtain a reliable altered genotype and phenotype (Sashidhar et al., 2019). Multiple guide sequences can be encoded into a single CRISPR array, allowing the simultaneous editing of several sites (Cong et al., 2013), making the CRISPR/Cas9 system a valuable tool in the knockout of redundant genes or parallel pathways in polyploids. Sashidhar et al. (2019) illustrate a case in point with the CRISPR/Cas9-mediated knockout of multiple paralogs of the key enzyme inositol tetrakisphosphate kinase (ITPK) involved in the synthesis of phytic acid in *B. napus* seed. A noticeable change in the phytic acid content of *B. napus* seed was observed only in triple mutants of the essential *BnITPK* genes.

The CRISPR/Cas9 GE tool has been widely adopted in the manipulation and study of a variety of genes underlying agronomical important traits in *Brassica*. These include traits such as pod shatter resistance (Braatz et al., 2017; Zhai et al., 2019), multi-ocular silique (Yang Y. et al., 2018), increase in oleic acid content in seed (Okuzaki et al., 2018), and seed coat color (Zhai et al., 2020) (**Table 5**). Even though the use of the CRISPR/Cas9 system is popular in gene knockout or knock-in studies, it is limited by the introduction of random indels at the target site in addition to the possibility off-target mutations.

Targeted Base Editing

Targeted base editing is one of the newest additions to GE. The technique is based on the CRISPR/Cas9 system and enables the direct and irremediable conversion of a selected target base without the induction and repair of a DSB (Komor et al., 2016). In the base-editing system, a cytosine or adenosine deaminase domain is fused to the N-terminus of a deactivated Cas9 (dCas9) of Cas9 nickase (nCas9). Although the Cas9 retains the ability to be guided by the gRNA, it instead mediates the direct conversion of cytidine (C) to uridine (U) resulting in a C-to-T or G-to-A substitution resulting in a single controlled point mutation

TABLE 5 | Application of programmable sequence-specific nucleases and targeted base editing tools in genome editing of *Brassica*.

***Brassica* species**	**Target gene/s and modification**	**Editing system**	**Modified/introduced trait**	**References**
Programmable sequence-specific nucleases				
Brassica napus	Introduction of ZFP-transcriptional activators (TF) to activate two canola β-ketoacyl-ACP Synthase II designed (*KASII*) genes	ZFN	Elevated KASII transcript levels, decrease in palmitic acid, increased in total C18, and reduced total saturated fatty acid content	Gupta et al., 2012
Brassica oleracea var. capitata	Targeted cleaving of the endogenous *FRIGIDA* (*FRI*) gene	TALENs	Earlier flowering	Sun et al., 2013
B. napus	Knockout of CMS-associated genes (*orf125*)	mitoTALENs	Restored male fertility	Kazama et al., 2019
B. oleracea	Indel mutations of the *BolC.GA4a*	CRISPR/Cas9	Dwarf phenotype and alterations in pod valve margins	Lawrenson et al., 2015
B. napus	Knockout mutations of *BnaRGA*, *BnaDA2*, and *BnaFUL*	CRISPR/Cas9	Demonstrate CRISPR/Cas9 efficiency in creating targeted genome modifications at multiple loci that are stable and inheritable	Yang et al., 2017
B. napus	Induced mutation in *BnWRKY11* and *BnWRKY70*	CRISPR/Cas9	*BnWRKY70* mutants exhibited enhanced resistance to *Sclerotinia*, while *BnWRKY11* mutants showed no significant difference in *Sclerotinia* resistance	Sun et al., 2018
B. napus	Modification of a fatty acid desaturase 2 gene (*FAD2*)	CRISPR/Cas9	Increase of oleic acid content seed	Okuzaki et al., 2018
B. napus	Knockout of rapeseed homologs of *CLAVATA3* (*CLV3*)	CRISPR/Cas9	Produced more leaves, multilocular siliques, and higher number of seed	Yang Y. et al., 2018
B. napus	Knockout of *INDEHISCENT* (*IND*) and *ALCATRAZ* (*ALC*) gene homologs	CRISPR/Cas9	Association of the *BnIND* gene with pod shatter resistance, whereas the *BnALC* gene appears to have limited potential for rapeseed shatter resistance	Tang et al., 2018
B. napus	Knockout of SPL3 homologous gene copies, *BnSPL3-A5/BnSPL3-A4/BnSPL3-C3/BnSPL3-C4/BnSPL3-Cnn*	CRISPR/Cas9	Developmental delay	Li C. et al., 2018
B. napus	Knockout mutations of the histone lysine methyltransferases *BnaSDG8.A* and *BnaSDG8.C*	CRISPR/Cas9	*BnaSDG8.A* and *BnaSDG8.C* help control the *B. napus* floral transition by directly altering the H3K36m2/3 levels at the *BnaFLC* chromatin loci	Jiang et al., 2018
B. oleracea var. *capitata*	Induced mutations in phytoene desaturase gene *BoPDS*, S-receptor kinase gene *BoSRK*, male-sterility-associated gene *BoMS1*	CRISPR/Cas9 based on endogenous tRNA processing	*BoSRK3* gene mutation suppressed self-incompatibility completely *BoMS1* gene mutation produced a completely male-sterile mutant	Ma et al., 2019
B. napus	Induced mutation in *INDEHISCENT* (*IND*) and *ALCATRAZ* (*ALC*)	CRISPR/Cas9	Establishes that *BnIND* gene is essential for pod shatter and highly conserved in *Brassica*	Zhai et al., 2019
B. napus	Knockout of seven *BnLPAT2* homologous genes and four *BnLPAT5* homologous genes	CRISPR/Cas9	Enlarged oil bodies, disrupted distribution of protein bodies and increased accumulation of starch confirming the role of *BnLPAT2* and *BnLPAT5* in oil biosynthesis	Zhang et al., 2019
B. napus	Knockout of five homoeologs (*BnJAG.A02*, *BnJAG.C02*, *BnJAG.C06*, *BnJAG.A07*, and *BnJAG.A08*) of the *JAGGED* (*JAG*) gene	CRISPR/Cas9	Affected development of the lateral organs in organizing pod shape and size as well as viable seed. *BnJAG.A08-NUB* gene caused significant changes in the pod dehiscence zone	Zaman et al., 2019
B. napus	Knockout mutations of *BnaMAX1* homologs	CRISPR/Cas9	Semi-dwarf and increased branching phenotypes with more siliques, contributing to increased yield per plant.	Zheng et al., 2020
B. napus	Mutation of the M-locus protein kinase (MLPK) *BnaMLPK* genes	CRISPR/Cas9	Revealed that MLPK is a positive regulator of the self-incompatibility response	Chen et al., 2019
B. napus	Knockout mutations of the *BnA10.LMI1* gene	CRISPR/Cas9	Demonstrated *BnLLA10* regulates the development of leaf lobes	Hu et al., 2018
Brassica carinata	Mutation of the *FASCICLIN-LIKE ARABINOGALACTAN PROTEIN 1* (*BcFLA1*) gene	CRISPR/Cas9	Down regulation of *BcFLA1* decreased root growth under Pi-deficient conditions predicting role	Kirchner et al., 2018
B. napus	Knockout mutation of the *APETALA2* (*BnAP2*) genes	CRISPR/Cas9	Loss of *AP2* function severely affects sepal and petal development and generates sepal carpeloid and apetalous mutants	Zhang et al., 2018a
Brassica campestris	Induced *BcPME37c* mutant	CRISPR/Cas9	*BcPME37c* mutation led to the abnormal thickening of the pollen intine	Xiong et al., 2019

(Continued)

TABLE 5 | Continued

Brassica species	Target gene/s and modification	Editing system	Modified/introduced trait	References
B. napus	Targeted knockout of *TRANSPARENT TESTA 2* (*BnTT2*) homologs	CRISPR/Cas9	*BnTT2* homologs had conserved but redundant functions in regulating seed color. Homozygous mutants of *BnTT2* homologs increased oil content and improved fatty acid composition with higher linoleic acid (C18:2) and linolenic acid (C18:3)	Xie et al., 2020
B. napus	Knockout lines of *BnaA9.WRKY47*	CRISPR/Cas9	*BnaA9.WRKY47* contributed to adaptation to B deficiency by up-regulating BnaA3.NIP5;1 expression to facilitate efficient B uptake. Suggested BnaWRKY involvement in adaptation to low B	Feng et al., 2020
B. napus	Knockout mutations of all of *BnaFAD2* homologs	CRISPR/Cas9	Mutation type of *BnaFAD2* affected oleic levels varied, suggesting that fatty acid levels might be manipulated through precise editing of specific regions of a gene	Huang et al., 2020
B. napus	Generated diverse mutants of the male-sterility allele MS5a and MS5c	CRISPR/Cas9	Male fertility dependent on expression levels and protein sequences of MS5a and MS5c	Xin et al., 2020
B. napus	Created gain-of-function mutant of *BnaA6.RGA*, *bnaa6.rga-D*, and the loss-of-function quadruple mutant, *bnarga*	CRISPR/Cas9	Quadruple mutant bnarga significantly decreased the sensitivity of stomatal closure, suggesting that BnaRGA proteins play important roles in plant adaptation to water-deficit stress	Wu et al., 2020b
B. napus	Knockout members of the *BnSFAR4* and *BnSFAR5* gene families	CRISPR/Cas9	Significant increase of seed oil content	Karunarathna et al., 2020
B. napus	Induced calreticulin (*BnCRT1a*) mutants	CRISPR/Cas9	Loss of function of *CRT1a* results in activation of the ethylene signaling pathway. Observed reduced susceptibility to *Verticillium longisporum*	Pröbsting et al., 2020
B. napus	Knockout three functional paralogs of *BnITPK*	CRISPR/Cas9	Reduction of phytic acid and increase of free phosphorus in seed	Sashidhar et al., 2020
B. napus	Mutation of the *BnTT8* genes	CRISPR/Cas9	Yellow seed coat, altered oil content, and fatty acid composition in seeds	Zhai et al., 2020
Targeted base editing				
B. napus	C-to-T substitution at the P197 position of *BnALS 1*	Base editing	Tribenuron-methyl herbicide resistance	Wu et al., 2020a
B. napus	A-to-G substitutions in *BnALS* and *BnPDS*	Base editing	Single amino acid change in the FT protein or mis-splicing of the PDS3 RNA transcript resulted in late-flowering and albino phenotypes	Kang et al., 2018
B. napus	A-to-G substitutions in *ALS*, *RGA*, and *IAA7* genes	Base editing system	Base-edited plants of ALS conferred high herbicide resistance, while base-edited RGA or IAA7 plants exhibited decreased plant height	Cheng et al., 2020

ZFP, zinc finger protein; TF, transcription factor; ZFN, zinc finger nuclease; CMS, cytoplasmic male sterility.

or base correction rather than a random gene disruption. The precise mutations can lead to the introduction of stop codons, changes in amino acids, and regulatory site modification, thereby improving the resolution in the functional analysis of genes and proteins to a single nucleotide or amino acid (Komor et al., 2016). Kang et al. (2018) established an adenosine base editing (ABE) system in *B. napus*, demonstrating the efficiency of ABE in generating A-to-G substitutions at the target BnALS and BnPDS loci. The substitution resulted in a single amino acid change in the FT protein or mis-splicing of the PDS3 RNA transcript generating germline transmissible transgenic *Brassica* plants with late-flowering and albino phenotypes. The nCas9 cytosine base-editing system was employed by Wu et al. (2020a) to introduce a C-to-T conversion at the P197 position of the *BnALS 1* gene in *B. napus*. The P197S substitution conferred tribenuron-methyl resistance generating transgene-free homozygous mutants. Herbicide resistance and dwarfed plant architecture, both important traits in the commercial cultivation of oilseed rape, were reportedly introduced by Cheng et al. (2020) using the A3A-PBE base-editing system. The A3A-PBE base-editing system enabled the substitution of C to T with increased efficiency (>20%) and wider editing window.

Modifications brought upon by genome-editing technologies pose a significantly lower risk than those associated with transgenics. In general, only a few selected nucleotides are altered, rendering changes similar to those that occur in natural populations (Voytas and Gao, 2014). After the genomic-editing agents were segregated out, it is not possible to differentiate between a naturally occurring mutation and the gene edit. GE is therefore a valuable tool to establish rapid and precise changes to aid crop improvement (Zhang et al., 2018b).

GENOMIC SELECTION IN CROP IMPROVEMENT

Plant breeding is founded on the principles of collection, induction, and rearrangement of genetic diversity followed

by phenotypic-driven selection. Conventional breeding success was achieved through the exploitation of natural or mutation-induced variation followed by efficient selection of desirable traits largely based on phenotypic observation (Pérez-de-Castro et al., 2012). This approach has several limitations including long periods required (5–12 years) to develop a crop variety, high environmental noise, and being less effective in the improvement of complex and low heritable traits (Tuberosa, 2012).

Desirable traits often include characteristics such as increased yield, plant architecture, tolerance to environmental stresses, and resistance against pests and diseases. An interlinking network of multiple "minor" genes regulates the expression of these agronomically important features. Phenotypic expression is further shaped by non-genetic factors including genotype–environment interactions (Werner et al., 2018). It is not possible to accurately access these intricate and dynamic interdependencies based on phenotypic observation alone in conventional breeding. Crop development is therefore limited by extended periods of selections of up to 10 years, environmental noise, and low heredity of complex traits in conventional breeding approaches.

With the onset of the genomics era (Nepolean et al., 2018), marker-assisted breeding (MAS) was developed to address the limitations posed by conventional breeding (Collard and Mackill, 2008). In MAS, functional markers linked to QTL are used to detect important traits through linkage mapping or GWAS. As only statistically significant marker–trait associations are retained (Arruda et al., 2016), MAS is restricted to the detection of traits controlled by a limited number of QTL with large contributions to phenotypic variation. MAS has therefore limited value in the selection of traits under complex genetic control and is, as such, outperformed by traditional phenotypic selection (Bernardo, 2001, 2016; Zhao et al., 2014).

Fast-evolving genomic tools and vast amounts of available genomic sources are permitting the establishment of genotype–phenotype relationships, in particular for complex multi-genic traits (Pérez-de-Castro et al., 2012). Genome-wide selection or GS (Meuwissen et al., 2001), contrary to MAS, includes all marker information in the prediction model, reducing marker bias and allowing the potential to explain variance even with small-effect QTL. Predicted marker effects based on phenotype and high-density marker scores are then used to estimate the breeding value of untested genotypes (Zhao et al., 2015). This estimate, applied in the preselection of promising genotypes, can accelerate progress in crop breeding and reduce cost in comparison with conventional breeding (Wang X. et al., 2018).

Hybrid breeding has been widely used in the improvement of crop performance through the exploitation of heterosis (Liu et al., 2020). In heterosis, hybrid offspring created have the potential to outclass agronomic characteristics of the parents. The selection of suitable parental combinations is therefore of essence and can pose a major challenge in the development of hybrids. However, GS has shown the potential to predict hybrid performance (Zhao et al., 2015); resulting hybrid genotypes can be inferred from their inbred parents and potentially reduce genotyping cost and generation interval (Wang X. et al., 2018). GS methods for hybrid canola breeding was evaluated by Jan et al. (2016); genome-wide SNP profiles were used to evaluate the prediction of the best possible parental combination of pollinators crossed with the two tester lines in a testcross performance for a number of important traits in spring canola. Based on genome-wide SNP markers, it was determined that testcross performance prediction in canola breeding could be an effective and efficient method to preselect promising pollinators for combinations with available male-sterile maternal lines, thereby promoting the efficient allocation of breeding resources (Jan et al., 2016).

Würschum et al. (2014) investigated the potential of GS in rapeseed breeding reporting medium-to-high prediction accuracies for several morphological-, quality-, and yield-related traits. Despite lower accuracy in the prediction of some novel families, it was concluded that with increased marker availability, GS will provide a valuable genomic tool in knowledge-based rapeseed breeding. GS has further been applied in winter-type oilseed rape (Werner et al., 2017), spring-sown canola (Jan et al., 2016), and a biparental population based on a cross between a European winter cultivar and a Chinese semi-winter cultivar (Zou et al., 2016; Liu P. et al., 2017). Werner et al. (2018) investigated the value of marker selection approaches in Asian rapeseed and illustrated that high prediction accuracies for polygenic traits are achievable with low marker density, given that the representative markers were selected with regard to the genome-wide linkage disequilibrium (LD) structure in a population.

Increased phenotypic heritability has been shown to have a greater impact on whole-genome prediction accuracies, more so than training set population size and marker density (Zhang et al., 2017). Fikere et al. (2020) reported moderate-to-high genomic prediction accuracies using genomic best linear unbiased prediction (GBLUP) models upon evaluating genetic correlations and genomic prediction accuracies for several agronomic, disease, and seed quality traits in canola. The inclusion of genotype-by-environment interaction in the GBLUP model resulted in further, though slight, improvements in predictions. Koscielny et al. (2020) confirmed these findings, demonstrating higher accuracy in whole-genome predictions within the stress treatment than within the control treatment for the majority of traits evaluated. It is therefore important, even in the genomics era, to link selected phenotypic or demographic models with the underlying processes of genomic variation. As demonstrated in the CWR *Brassica cretica*, if variation is largely selectively neutral, it is not possible to assume that a diverse population will inescapably display the wide-ranging adaptive diversity required for further crop improvement (Kioukis et al., 2020).

FUTURE PROSPECTS

Mechanisms of interaction for stress responses involve complex interactions and traits and are therefore more difficult to investigate than direct interaction (Werner et al., 2018). While genomic advances have exponentially increased during the past

decades, high-throughput phenotyping has not caught up yet. To accelerate plant breeding and improve our understanding of genotype underlying expressed phenotype, dedicated high-throughput phenotyping approaches are required (Singh D. et al., 2019). It is therefore not surprising that high-throughput phenomics has increased in popularity, especially for the management and data collection of *Brassica* oilseed and vegetable crops. Improvements in sensor, drone, and remote sensing technology, as well as high throughput phenotyping techniques, are simplifying and enabling the quantification of complex phenotypic traits without the necessity of destructive sampling (Parmley et al., 2019). *Brassica* physiological studies, for example, plant height and biomass data (Moeckel et al., 2018), flower number (Wan et al., 2018), vegetation and flower fraction (Fang et al., 2016), and nitrogen nutrient studies (Graeff et al., 2008; Liu S. et al., 2018), have been generated using unmanned aerial vehicles. Assimilating large amounts of phenotypic data with the capabilities of machine learning will provide breeders with the analytical tools to optimize cultivar development in relation to target environment and accelerate the rate of genetic gain (Parmley et al., 2019).

Besides phenotypic characteristics, crop breeding requirements are also dictated by an assortment of additional major and minor variables such as environment, cultivation, and management practices and fluctuating consumer needs (Araus et al., 2018). Aligning breeding objectives to an increasing number of critical factors will require cross-disciplinary approaches driven by breeding teams, climate specialist, bioinfomaticians, and crop modelers (Beveridge et al., 2019; Stöckle and Kemanian, 2020). Crop modeling can assist breeders in comprehending the influence and interaction of variable factors in the selection of desirable varieties (Stöckle and Kemanian, 2020).

CONCLUSION

Crop breeding has benefitted from the advancement of genomic tools and associated analysis pipelines. Available genomic resources and lower cost of high-throughput sequencing have contributed toward the increase in WGRS efforts. The vast amount of genomic information created and advances in genomic tools developed will significantly improve capturing the range of genetic diversity estimation and enhance the capturing and exploitation of diversity in *Brassica* germplasm profiles. The genetic libraries of CWRs should be further explored, as quality of available references and assembly methods has improved. The availability of GE tools has improved in precision and specificity; these systems are highly customizable and can be advantageously exploited to fast-track crop improvement. Although genomics is currently taking the center stage, a multidisciplinary plant breeding approach that includes phenotype = genotype × environment × management interaction backed by big data capabilities will ultimately ensure the selection of future-proof *Brassica* crops.

AUTHOR CONTRIBUTIONS

AS-E, NM, and JB conceptualized the manuscript. NM, AS-E, and AP wrote the manuscript, with additions and edits from JB and DE. The tables were prepared by NM and AS-E. AS-E illustrated the figure. All authors read and approved the final manuscript.

FUNDING

This work was funded by Australian Research Council Projects DP1601004497 and LP160100030.

REFERENCES

Aghazadeh, R., Zamani, M., Motallebi, M., Moradyar, M., and Moghadassi Jahromi, Z. (2016). Co-transformation of canola by chimeric chitinase and *tlp* genes towards improving resistance to *Sclerotinia sclerotiorum*. *World J. Microbiol. Biotechnol.* 3:144. doi: 10.1007/s11274-016-2104-6

Akama, S., Shimizu-Inatsugi, R., Shimizu, K. K., and Sese, J. (2014). Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. *Nucleic Acids Res.* 42:e46. doi: 10.1093/nar/gkt1376

Akter, A., Takahashi, S., Deng, W., Shea, D. J., Itabashi, E., Shimizu, M., et al. (2019). The histone modification H3 lysine 27 tri-methylation has conserved gene regulatory roles in the triplicated genome of *Brassica rapa* L. *DNA Res.* 26, 433–443. doi: 10.1093/dnares/dsz021

Allier, A., Lehermeier, C., Charcosset, A., Moreau, L., and Teyssèdre, S. (2019). Improving short- and long-term genetic gain by accounting for within-family variance in optimal cross-selection. *Front. Genet.* 10:1006. doi: 10.3389/fgene.2019.01006

Altpeter, F., Springer, N. M., Bartley, L. E., Blechl, A. E., Brutnell, T. P., Citovsky, V., et al. (2016). Advancing crop transformation in the era of genome editing. *Plant Cell* 28, 1510–1520. doi: 10.1105/tpc.16.00196

Álvarez-Venegas, R., Zhang, Y., Kraling, K., and Tulsieram, L. (2011). Flowering without vernalization in winter canola (*Brassica napus*): use of Virus-Induced Gene Silencing (VIGS) to accelerate genetic gain. *Nova Sci.* 3, 29–50.

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21:30. doi: 10.1186/s13059-020-1935-5

Amoah, S., Kurup, S., Rodriguez Lopez, C. M., Welham, S. J., Powers, S. J., Hopkins, C. J., et al. (2012). A hypomethylated population of *Brassica rapa* for forward and reverse epi-genetics. *BMC Plant Biol.* 12:193. doi: 10.1186/1471-2229-12-193

An, H., Qi, X., Gaynor, M. L., Hao, Y., Sarah, C., Gebken, S. C., et al. (2019). Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. *Nat. Commun.* 10:2878. doi: 10.1038/s41467-019-10757-1

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., Brown-Guedira, G., et al. (2016). Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 36:84. doi: 10.1007/s11032-016-0508-5

Augustine, R., Mukhopadhyay, A., and Bisht, N. C. (2013). Targeted silencing of *BjMYB28* transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*. *Plant Biotechnol. J.* 11, 855–866. doi: 10.1111/pbi.12078

Backes, G. (2013). "TILLING and EcoTILLING," in *Diagnostics in Plant Breeding*, eds T. Lübberstedt and R. K. Varshney (Dordrecht: Springer), 145–165.

Bancroft, I., Morgan, C., Fraser, F., Higgins, J., Wells, R., Clissold, L., et al. (2011). Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nat. Biotechnol.* 29, 762–766. doi: 10.1038/nbt.1926

Basso, M. F., Arraes, F. B. M., Grossi-de-Sa, M., Moreira, V. J. V., Alves-Ferreira, M., and Grossi-de-Sa, M. F. (2020). Insights into genetic and molecular elements for transgenic crop development. *Front. Plant Sci.* 11:509. doi: 10.3389/fpls.2020.00509

Bayer, P. E., Hurgobin, B., Golicz, A. A., Chan, C.-K. K., Yuan, Y., Lee, H., et al. (2017). Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol. J.* 15, 1602–1610. doi: 10.1111/pbi.12742

Bebber, D. P., Ramotowski, M. A. T., and Gurr, S. J. (2013). Crop pests and pathogens move polewards in a warming world. *Nat. Clim.Change* 3, 985–988. doi: 10.1038/nclimate1990

Becker, H. C., Engqvist, G. M., and Karlsson, B. (1995). Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theor. Appl. Genet.* 91, 62–67. doi: 10.1007/BF00220859

Bekele, D., Kassahun, T., and Fikre, A. (2019). Applications of virus induced gene silencing (VIGS) in plant functional genomics studies. *J. Plant Biochem. Physiol.* 7:1. doi: 10.4172/2329-9029.1000229

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Bernardo, R. (2001). What if we knew all the genes for a quantitative trait in hybrid crops? *Crop Sci.* 41, 1–4. doi: 10.2135/cropsci2001.4111

Bernardo, R. (2016). Bandwagons I, too, have known. *Theor. Appl. Genet.* 129, 2323–2332. doi: 10.1007/s00122-016-2772-5

Beveridge, L., Whitfield, S., and Challinor, A. (2019). Crop modelling: towards locally relevant and climate-informed adaptation. *Clim. Change* 147, 475–489. doi: 10.1007/s10584-018-2160-z

Beversdorf, W. D., Hume, D. J., and Daonnelly-Vanderloo, M. J. (1988). Agronomic performance of trianzine-resistant and susceptible reciprocal spring canola hybrids. *Crop Sci.* 28, 932–934. doi: 10.2135/cropsci1988.0011183X002800060012x

Bizuayehu, T. T., Kornel Labun, K., Jefimov, K., and Valen, E. (2020). Single molecule structure sequencing reveals RNA structural dependencies, breathing and ensembles. *bioRxiv*. doi: 10.1101/2020.05.18.101402v1

Bodirsky, B. L., Popp, A., Weindl, I., Dietrich, J. P., Rolinski, S., Scheiffele, L., et al. (2012). N_2O emissions from the global agricultural nitrogen cycle – current state and future scenarios. *Biogeoscience* 9, 4169–4197. doi: 10.5194/bg-9-4169-2012

Braatz, J., Harloff, H. J., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 174, 935–942. doi: 10.1104/pp.17.00426

Branca, F., Argento, S., and Alessandro, T. (2012). "Assessing genetic reserves in Sicily (Italy): The *Brassica* wild relatives case study," in *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*, eds N. Maxted, M. Ehsan Dulloo, B. V. Ford-Lloyd, L. Frese, J. M. Iriondo, and M. A. A. Pinheiro de Carvalho (Wallingford: Centre for Agriculture and Bioscience International), 52–58.

Briskine, R. V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., et al. (2017). Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol. Ecol. Res.* 17, 1025–1036. doi: 10.1111/1755-0998.12604

Brookes, G., and Barfoot, P. (2018). Farm income and production impacts of using GM crop technology 1996-2016. *GM Crop. Food* 9, 59–89. doi: 10.1080/21645698.2018.1464866

Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454

Budar, F., and Roux, F. (2011). The role of organelle genomes in plant adaptation: time to get to work! *Plant Signal. Behav.* 6, 635–639. doi: 10.4161/psb.6.5.14524

Cai, C., Wang, X., Liu, B., Wu, J., Liang, J., Cui, Y., et al. (2017). *Brassica rapa* genome 2.0: a reference upgrade through sequence re-assembly and gene re-annotation. *Mol. Plant.* 10, 649–651. doi: 10.1016/j.molp.2016.11.008

Cao, J., Shelton, A. M., and Earle, E. D. (2005). Development of transgenic collards (*Brassica oleracea* L., var. *acephala*) expressing a *cry*1Ac or *cry*1C Bt gene for control of the diamondback moth. *Crop Prot.* 24, 804–813. doi: 10.1016/j.cropro.2004.12.014

Cappelli, A., and Cini, E. (2020). Will the COVID-19 pandemic make us reconsider the relevance of short food supply chains and local productions? *Trends Food Sci. Tech.* 99, 566–567. doi: 10.1016/j.tifs.2020.03.041

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., et al. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 3:e82. doi: 10.1093/nar/gkr218

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 34599, 950–953. doi: 10.1126/science.1253435

Chang, S., Yang, T., Du, T., Huang, Y., Chen, J., Yan, J., et al. (2011). Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in *Brassica*. *BMC Genome* 12:497. doi: 10.1186/1471-2164-12-497

Chen, F., Yang, Y., Li, B., Liu, Z., Khan, F., Zhang, T., et al. (2019). Functional analysis of M-Locus Protein Kinase revealed a novel regulatory mechanism of self-incompatibility in *Brassica napus* L. *Int. J. Mol. Sci.* 20:3303. doi: 10.3390/ijms20133303

Chen, J., Guan, R., Chang, S., Du, T., Zhang, H., and Xing, H. (2011). Substoichiometrically different mitotypes coexist in mitochondrial genomes of *Brassica napus* L. *PLoS ONE* 6:e17662. doi: 10.1371/journal.pone.0017662

Chen, X., Ge, X., Wang, J., Tan, C., King, G. J., and Liu, K. (2015). Genome-wide DNA methylation profiling by modified reduced representation bisulfite sequencing in *Brassica rapa* suggests that epigenetic modifications play a key role in polyploid genome evolution. *Front. Plant Sci.* 6:836. doi: 10.3389/fpls.2015.00836

Chen, X., Tong, C., Zhang, X., Song, A., Hu, M., Dong, W., et al. (2020). A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. *Plant Biotechnol. J.* doi: 10.1111/pbi.13493

Chen, Y., Zhou, B., Li, J., Tang, H., Tang, J., and Yang, Z. (2018). Formation and change of chloroplast-located plant metabolites in response to light conditions. *Int. J. Mol. Sci.* 19:654. doi: 10.3390/ijms19030654

Cheng, F., Wu, J., Cai, C., Fu, L., Liang, J., Borm, T., et al. (2016). Genome resequencing and comparative variome analysis in a *Brassica rapa* and *Brassica oleracea* collection. *Sci. Data* 3:160119. doi: 10.1038/sdata.2016.119

Cheng, H., Hao, M., Ding, B., Mei, D., Wang, W., Wang, H., et al. (2020). Base editing with high efficiency in allotetraploid oilseed rape by A3A-PBE base editing system. *Plant Biotechnol. J.* 19, 87–97. doi: 10.1111/pbi.13444

Chétrit, P., Mathieu, C., Muller, J. P., and Vedel, F. (1984). Physical and gene mapping of cauliflower (*Brassica oleracea*) mitochondrial DNA. *Curr. Genet.* 8, 413–421. doi: 10.1007/BF00433907

Clarke, W. E., Higgins, E. E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., et al. (2016). A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* 129, 1887–1899. doi: 10.1007/s00122-016-2746-7

Clarke, W. E., Parkin, I. A., Gajardo, H. A., Gerhardt, D. J., Higgins, E., Sidebottom, C., et al. (2013). Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. *PLoS ONE* 8:e081992. doi: 10.1371/journal.pone.0081992

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745

Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Connelly, M., and MacIntosh, S. (2018). *Petition for Determination of Nonregulated Status for DHA Canola*. Available online at: https://www.aphis.usda.gov/brs/aphisdocs/17_23601p.pdf

Cowling, W. A. (2007). Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crop Res.* 104, 103–111. doi: 10.1016/j.fcr.2006.12.014

Cox, D. B. T., Platt, R. J., and Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nat. Med.* 21, 121–131. doi: 10.1038/nm.3793

Crutzen, P. J., and Ehhalt, D. H. (1977). Effects of nitrogen fertilizers and combustion on the stratospheric ozone layer. *Ambio* 6, 112–117.

Cuthbert, J. L., McVetty, P. B. E., Freyssinet, G., and Freyssinet, M. (2001). Comparison of the performance of bromoxynil resistant and susceptible near-isogenic populations of oilseed rape. *Can. J. Plant Sci.* 81, 367–372. doi: 10.4141/P00-115

Dalton-Morgan, J., Hayward, A., Alamery, S., Tollenaere, R., Mason, A. S., Campbell, E., et al. (2014). A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. *Funct. Integr. Genomics* 14, 643–655. doi: 10.1007/s10142-014-0391-2

Daniell, H. (2007). Transgene containment by maternal inheritance: effective or elusive? *Proc. Natl. Acad. Sci. U.S.A.* 104, 6879–6880. doi: 10.1073/pnas.0702219104

Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134–134. doi: 10.1186/s13059-016-1004-2

Darracq, A., Varré, J. S., Maréchal-Drouard, L., Courseaux, A., Castric, V., Saumitou-Laprade, P., et al. (2011). Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. *Genome Biol. Evol.* 3, 723–736. doi: 10.1093/gbe/evr042

Das, P. M., Ramachandran, K., Van Wert, J., and Singal, R. (2004). Chromatin immunoprecipitation assay. *Biotech.* 37, 961–969. doi: 10.2144/04376RV01

Davis, D. (2009). Declining fruit and vegetable nutrient composition: what is the evidence? *Hortic. Sci.* 44, 15–19. doi: 10.21273/HORTSCI.44.1.15

Davis, D. R., Epp, M. D., and Riordan, H. D. (2005). Changes in USDA food composition data for 43 garden crops, 1950 to 1999. *J. Am. Coll. Nutr.* 23, 669–682. doi: 10.1080/07315724.2004.10719409

De Block, M., De Brouwer, D., and Tenning, P. (1989). Transformation of *Brassica napus* and *Brassica oleracea* using *Agrobacterium tumefaciens* and the expression of the *bar* and *neo* genes in the transgenic plants. *Plant Physiol.* 91, 694–701. doi: 10.1104/pp.91.2.694

de Paulo Farias, D., and dos Santos Gomes, M. G. (2020). COVID-19 outbreak: what should be done to avoid food shortages? *Trends Food Sci. Technol.* 102, 291–292. doi: 10.1016/j.tifs.2020.06.007

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885

Dirks, R., van Dun, K., de Snoo, C. B., van den Berg, M., Lelivelt, C. L., Voermans, W., et al. (2009). Reverse breeding: a novel breeding approach based on engineered meiosis. *Plant Biotechnol. J.* 7, 837–845. doi: 10.1111/j.1467-7652.2009.00450.x

Dolatabadian, A., Bayer, P. E., Tirnaz, S., Hurgobin, B., Edwards, D., and Batley, J. (2020). Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* 18, 969–982. doi: 10.1111/pbi.13262

Dolinoy, D. C., Weinhouse, C., Jones, T. R., Rozek, L. S., and Jirtle, R. L. (2010). Variable histone modifications at the A_{vy} metastable epiallele. *Epigenetics*. 5, 637–644. doi: 10.4161/epi.5.7.12892

Dreccer, M. F., Schapendonk, A. H. C. M., Slafer, G. A., and Rabbinge, R. (2000). Comparative response of wheat and oilseed rape to nitrogen supply: absorption and utilisation efficiency of radiation and nitrogen during the reproductive stages determining yield. *Plant Soil* 220, 189–205. doi: 10.1023/A:1004757124939

Dusenge, M. E., Duarte, A. G., and Way, D. A. (2019). Plant carbon metabolism and climate change: elevated CO_2 and temperature impacts on photosynthesis, photorespiration and respiration. *New Phytol.* 221, 32–49. doi: 10.1111/nph.15283

Dwivedi, S. L., Scheben, A., Edwards, D., Spillane, C., and Ortiz, R. (2017). Assessing and exploiting functional diversity in germplasm pools to enhance abiotic stress adaptation and yield in cereals and food legumes. *Front. Plant Sci.* 8:1461. doi: 10.3389/fpls.2017.01461

Elhamamsy, A. R. (2016). DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation. *Cell Biochem. Funct.* 34, 289–298. doi: 10.1002/cbf.3183

Espinoza, C., Schlechter, R., Herrera, D., Torres, E., Serrano, A., Medina, C., et al. (2013). Cisgenesis and intragenesis: new tools for improving crops. *Biol. Res.* 46, 323–331. doi: 10.4067/S0716-97602013000400003

Fang, S., Tang, W., Peng, Y., Gong, Y., Dai, C., Chai, R., et al. (2016). Remote estimation of vegetation fraction and flower fraction in oilseed rape with unmanned aerial vehicle data. *Remote Sens.* 8:416. doi: 10.3390/rs8050416

Feng, Y., Cui, R., Wang, S., He, M., Hua, Y., Shi, L., et al. (2020). Transcription factor BnaA9.WRKY47 contributes to the adaptation of *Brassica napus* to low boron stress by up-regulating the boric acid channel gene *BnaA3.NIP5;1*. *Plant Biotechnol. J.* 18, 1241–1254. doi: 10.1111/pbi.13288

Ferreira de Carvalho, J., Lucas, J., Deniot, G., Falentin, C., Filangi, O., Gilet, M., et al. (2019). Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. *Plant J.* 98, 434–447. doi: 10.1111/tpj.14228

Ficklin, D. L., and Novick, K. A. (2017). Historic and projected changes in vapor pressure deficit suggest a continental-scale drying of the United States atmosphere. *J. Geophys. Res. Atmos.* 122, 2061–2079. doi: 10.1002/2016jd025855

Fikere, M., Barbulescu, D. M., Malmberg, M. M., Maharjan, P., Salisbury, P. A., Kant, S., et al. (2020). Genomic prediction and genetic correlation of agronomic, blackleg disease, and seed quality traits in canola (*Brassica napus* L.). *Plants* 9:719. doi: 10.3390/plants9060719

Florez-Sarasa, I., Fernie, A. R., and Gupta, K. J. (2020). Does the alternative respiratory pathway offer protection against the adverse effects resulting from climate change? *J. Exp. Bot.* 71, 465–469. doi: 10.1093/jxb/erz428

Francisco, M., Tortosa, M., Martínez-Ballesta, M., d,.C., Velasco, P., García-Viguera, C., et al. (2017). Nutritional and phytochemical value of *Brassica* crops from the agri-food perspective. *Ann. Appl. Biol.* 170, 273–285. doi: 10.1111/aab.12318

Frewer, L. J., van der Lans, I., Fischer, A. R. H., Reinders, M. J., Menozzi, D., Zhang, X., et al. (2013). Public perceptions of agri-food applications of genetic modification – A systematic review and meta-analysis. *Trends Food Sci. Technol.* 30, 142–152. doi: 10.1016/j.tifs.2013.01.003

Fu, Y-B. and Gugel, R. K. (2010). Genetic diversity of Canadian elite summer rape (Brassica napus L.) cultivars from the pre- to post-canola quality era. *Can. J. Plant Sci.* 90, 23–33. doi: 10.4141/CJPS09073

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns – from conservation to diversity. *Trends Plant Sci.* 11, 199–208. doi: 10.1016/j.tplants.2006.02.008

Gabur, I., Chawla, H. S., Lopisso, D. T., von Tiedemann, A., Snowdon, R. J., and Obermeier, C. (2020). Gene presence-absence variation associates with quantitative *Verticillium longisporum* disease resistance in *Brassica napus*. *Sci. Rep.* 10:4131. doi: 10.1038/s41598-020-61228-3

Gabur, I., Chawla, H. S., Snowdon, R. J., and Parkin, I. A. P. (2018). Connecting genome structural variation with complex traits in crop plants. *Theor. Appl. Genet.* 132, 733–750. doi: 10.1007/s00122-018-3233-0

Gaebelein, R., Schiessl, S. V., Samans, B., Batley, J., and Mason, A. S. (2019). Inherited allelic variants and novel karyotype changes influence fertility and genome stability in *Brassica* allohexaploids. *New Phytol.* 223, 965–978. doi: 10.1111/nph.15804

Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi: 10.1016/j.tibtech.2013.04.004

Gajardo, H. A., Wittkop, B., Soto-Cerda, B., Higgins, E. E., Parkin, I. A. P., Snowdon, R. J., et al. (2015). Association mapping of seed quality traits in *Brassica napus* L. using GWAS and candidate QTL approaches. *Mol. Breed.* 35:143. doi: 10.1007/s11032-015-0340-3

Gallusci, P., Dai, Z., Génard, M., Gauffretau, A., Leblanc-Fournier, N., Richard-Molard, C., et al. (2017). Epigenetics for plant improvement: current knowledge and modeling avenues. *Trends Plant Sci.* 22, 610–623. doi: 10.1016/j.tplants.2017.04.009

Garg, M., Sharma, N., Sharma, S., Kapoor, P., Kumar, A., Chunduri, V., et al. (2018). Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. *Front. Nutr.* 5:12. doi: 10.3389/fnut.2018.00012

Gazave, E., Tassone, E. E., Ilut, D. C., Wingerson, M., Datema, E., et al. (2016). Population genomic analysis reveals differential evolutionary histories and

patterns of diversity across subgenomes and subpopulations of *Brassica napus* L. *Front. Plant Sci.* 7:525. doi: 10.3389/fpls.2016.00525

Geng, X. X., Chen, S., Astarini, I. A., Yan, G. J., Tian, E., Meng, J. L., et al. (2013). Doubled haploids of novel trigenomic *Brassica* derived from various interspecific crosses. *Plant Cell Tissue Organ Cult.* 113, 501–511. doi: 10.1007/s11240-013-0292-4

Gilchrist, E. J., Sidebottom, C. H. D., Koh, C. S., MacInnes, T., Sharpe, A. G., and Haughn, G. W. (2013). A mutant *Brassica napus* (canola) population for the identification of new genetic diversity via TILLING and next generation sequencing. *plos ONE* 8:e84303. doi: 10.1371/journal.pone.0084303

Gocal, G. F. W., Schöpke, C., and Beetham, P. R. (2015). "Oligo-Mediated Targeted Gene Editing," in *Advances in New Technology for Targeted Modification of Plant Genomes*, eds F. Zhang, H. Puchta, and J. Thomson (New York, NY: Springer), 73–89.

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., et al. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7:13390. doi: 10.1038/ncomms13390

Graeff, S., Pfenning, J., Claupein, W., and Liebig, H.-P. (2008). Evaluation of image analysis to determine the N-fertilizer demand of broccoli plants (*Brassica oleracea* convar. *botrytis var. italica). Adv. Opt. Tech.* 2008, 1–8. doi: 10.1155/2008/359760

Grewe, F., Edger, P. P., Keren, I., Sultan, L., Pires, J. C., Ostersetzer-Biran, O., et al. (2014). Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion* 19, 135–143. doi: 10.1016/j.mito.2014.05.008

Grison, R., Grezes-Besset, B., Schneider, M., Lucante, N., Olsen, L., Leguay, J. J., et al. (1996). Field tolerance to fungal pathogens of *Brassica napus* constitutively expressing a chimeric chitinase gene. *Nat. Biotechnol.* 14, 643–646. doi: 10.1038/nbt0596-643

Guo, X., Hu, Q., Hao, G., Wang, X., Zhang, D., Ma, T., et al. (2018). The genomes of two *Eutrema* species provide insight into plant adaptation to high altitudes. *DNA Res.* 25, 307–315. doi: 10.1093/dnares/dsy003

Gupta, M., Atri, C., Agarwal, N., and Banga, S. S. (2016). Development and molecular-genetic characterization of a stable *Brassica* allohexaploid. *Theor. Appl. Genet.* 129, 2085–2100. doi: 10.1007/s00122-016-2759-2

Gupta, M., DeKelver, R. C., Palta, A., Clifford, C., Gopalan, S., Miller, J. C., et al. (2012). Transcriptional activation of *Brassica napus* β-ketoacyl-ACP synthase II with an engineered zinc finger protein transcription factor. *Plant Biotechnol. J.* 10, 783–791. doi: 10.1111/j.1467-7652.2012.00695.x

Han, F., Zhang, X., Liu, X., Su, H., Kong, C., Fang, Z., et al. (2017). Comparative analysis of genome wide DNA methylation profiles for the genic male sterile cabbage line 01-20S and its maintainer line. *Genes* 8:159. doi: 10.3390/genes8060159

Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): Comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* 31, 5907–5916. doi: 10.1093/nar/gkg795

Harlan, J. R. (1992). *Crops and Man. 2nd Edn*. Madison, WI: American Society of Agronomy and Crop Science Society of America. doi: 10.2135/1992

Harloff, H.-J., Lemcke, S., Mittasch, J., Frolov, A., Wu, J. G., Dreyer, F., et al. (2012). A mutation screening platform for rapeseed (*Brassica napus* L.) and the detection of sinapine biosynthesis mutants. *Theor. Appl. Genet.* 124, 957–969. doi: 10.1007/s00122-011-1760-z

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., et al. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotechnol.* 30, 798–802. doi: 10.1038/nbt.2302

Hatono, S., Nishimura, K., Murakami, Y., Tsujimura, M., and Yamagishi, H. (2017). Complete mitochondrial genome sequences of *Brassica rapa* (Chinese cabbage and mizuna), and intraspecific differentiation of cytoplasm in *B. rapa and Brassica juncea. Breed. Sci.* 67, 357–362. doi: 10.1270/jsbbs.17023

Hauben, M., Haesendonckx, B., Standaert, E., Van Der Kelen, K., Azmi, A., Akpo, H., et al. (2009). Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20109–20114. doi: 10.1073/pnas.0908755106

He, Y., Wu, D., Wei, D., Fu, Y., Cui, Y., Dong, H., et al. (2017). GWAS, QTL mapping and gene expression analyses in *Brassica napus* reveal genetic control of branching morphogenesis. *Sci. Rep.* 7:15971. doi: 10.1038/s41598-017-15976-4

He, Z., Cheng, F., Li, Y., Wang, X., Parkin, I.A., Chalhoub, B., et al. (2015). Construction of Brassica A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. *Data Brief* 4, 357–362. doi: 10.1016/j.dib.2015.06.016

Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512

Heng, S., Wang, L., Yang, X., Huang, H., Chen, G., Cui, M., et al. (2020). Genetic and comparative transcriptome analysis revealed degs involved in the purple leaf formation in *Brassica juncea*. *Front. Genet.* 11:322. doi: 10.3389/fgene.2020.00322

Herrera, R. J., and Garcia-Bertrand, R. (eds.). (2018). "The agricultural revolutions," in *Ancestral DNA, Human Origins, and Migrations* (London: Academic Press), 475–509.

Hickey, L. T., Hafeez, A. N., Robinson, H., Jackson, S. A., Leal-Bertioli, S., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi: 10.1038/s41587-019-0152-9

Higgins, E. E., Clarke, W. E., Howell, E. C., Armstrong, S. J., and Parkin, I. A. P. (2018). Detecting *de novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. *G3* 8, 2673–2683. doi: 10.1534/g3.118.200118

Himelblau, E., Gilchrist, E. J., Buono, K., Bizzell, C., Mentzer, L., Vogelzang, R., et al. (2009). Forward and reverse genetics of rapid-cycling *Brassica oleracea*. *Theor. Appl. Genet.* 118, 953–961. doi: 10.1007/s00122-008-0952-7

Holme, I. B. K., Wendt, T., and Holm, P. B. (2013). Intragenesis and cisgenesis as alternatives to transgenic crop development. *Plant Biotechnol. J.* 11, 395–407. doi: 10.1111/pbi.12055

Hong, H., Datla, N., Reed, D. W., Covello, P. S., MacKenzie, S. L., and Qiu, X. (2002). High-level production of γ-linolenic acid in *Brassica juncea* using a Δ6 desaturase from *Pythium irregulare*. *Plant Physiol.* 129, 354–362. doi: 10.1104/pp.001495

Houlton, B. Z., Almaraz, M., Aneja, V., Austin, A. T., Bai, E., Cassman, K. G., et al. (2019). A world of cobenefits: solving the global nitrogen challenge. *Earths Future* 7, 865–872. doi: 10.1029/2019EF001222

Hu, H., Scheben, A., and Edwards, D. (2018). Advances in integrating genomics and bioinformatics in the plant breeding pipeline. *Agriculture* 8:75. doi: 10.3390/agriculture8060075

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43, 476–481. doi: 10.1038/ng.807

Hu, Z.-Y., Hua, W., Huang, S.-M., and Wang, H.-Z. (2011). Complete chloroplast genome sequence of rapeseed (*Brassica napus* L.) and its evolutionary implications. *Genet. Resour. Crop Evol.* 58, 875–887. doi: 10.1007/s10722-010-9626-9

Huang, F., Liu, T., and Hou, X. (2018). Isolation and functional characterization of a floral repressor, *BcMAF1*, from Pak-choi (*Brassica rapa* ssp. *chinensis). Front. Plant Sci.* 9:290. doi: 10.3389/fpls.2018.00290

Huang, H., Cui, T., Zhang, L., Yang, Q., Yang, Y., Xie, K., et al. (2020). Modifications of fatty acid profile through targeted mutation at *BnaFAD2* gene with CRISPR/Cas9-mediated gene editing in *Brassica napus*. *Theor. Appl. Genet.* 133, 2401–2411. doi: 10.1007/s00122-020-03607-y

Hurgobin, B., and Edwards, D. (2017). SNP discovery using a pangenome: has the single reference approach become obsolete? *Biology* 6:21. doi: 10.3390/biology6010021

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C.-K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867

Hyseni, L., Bromley, H., Kypridemos, C., O'Flaherty, M., Lloyd-Williams, F., Guzman-Castillo, M., et al. (2017). Systematic review of dietary trans-fat reduction interventions. *Bull. World Health Organ.* 95, 821G—830G. doi: 10.2471/BLT.16.189795

Jan, H. U., Abbadi, A., Lücke, S., Nichols, R. A., and Snowdon, R. J. (2016). Genomic prediction of testcross performance in canola (*Brassica napus*). *PLoS ONE* 11:e0147769. doi: 10.1371/journal.pone.0147769

Jansing, J., Schiermeyer, A., Schillberg, S., Fischer, R., and Bortesi, L. (2019). Genome editing in agriculture: Technical and practical considerations. *Int. J. Mol. Sci.* 20:2888. doi: 10.3390/ijms20122888

Jiang, L., Li, D., Jin, L., Ruan, Y., Shen, W.-H., and Liu, C. (2018). Histone lysine methyltransferases BnaSDG 8.A and BnaSDG 8.C are involved in the floral transition in *Brassica napus*. *Plant J.* 95, 672–685. doi: 10.1111/tpj.13978

Jin, M., Liu, H., He, C., Fu, J., Xiao, Y., Wanget, Y., et al. (2016). Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. *Sci. Rep.* 6:18936. doi: 10.1038/srep18936

Jin, S., and Daniell, H. (2015). The engineered chloroplast genome just got smarter. *Trends Plant Sci.* 20, 622–640. doi: 10.1016/j.tplants.2015.07.004

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Jump, A. S., and Peñuelas, J. (2005), Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol. Lett.* 8, 1010–1020. doi: 10.1111/j.1461-0248.2005.00796.x

Jung, H.-J., Jung, H.-J., Ahmed, N. U., Park, J.-I., Kang, K.-K., Hur, Y., et al. (2012). Development of self-compatible *B. rapa* by RNAi-mediated S locus gene silencing. *PLoS ONE* 7:e49497. doi: 10.1371/journal.pone.0049497

Kamthan, A., Chaudhuri, A., Kamthan, M., and Datta, A. (2016). Genetically modified (GM) crops: milestones and new advances in crop improvement. *Theor. Appl. Genet.* 129, 1639–1655. doi: 10.1007/s00122-016-2747-6

Kanazawa, A., Inaba, J-i., Kasai, M., Shimura, H., and Masuta, C. (2011). RNA-mediated epigenetic modifications of an endogenous gene targeted by a viral vector. *Plant Signal. Behav.* 6, 1090–1093. doi: 10.4161/psb.6.8.16046

Kang, B., Yun, J., Kim, S., Shin, Y., Ryu, J., Choi, M., et al. (2018). Precision genome engineering through adenine base editing in plants. *Nat. Plants* 4, 427–431. doi: 10.1038/s41477-018-0178-x

Karunarathna, N. L., Wang, H., Harloff, H. J., Jiang, L., and Jung, C. (2020). Elevating seed oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. *Plant Biotechnol. J.* 18:11. doi: 10.1111/pbi.13381

Kasianov, A. S., Klepikova, A. V., Kulakovskiy, I. V., Gerasimov, E. S., Fedotova, A. V., Besedina, E. G., et al. (2017). High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution. *Plant J.* 91, 278–291. doi: 10.1111/tpj.13563

Kawanabe, T., Osabe, K., Itabashi, E., Okazaki, K., Dennis, E. S., and Fujimoto, R. (2016). Development of primer sets that can verify the enrichment of histone modifications, and their application to examining vernalization-mediated chromatin changes in *Brassica rapa* L. *Genes Genet. Syst.* 91:1. doi: 10.1266/ggs.15-00058

Kazama, T., Okuno, M., Watari, Y., Yanase, S., Koizuka, C., Tsuruta, Y., et al. (2019). Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. *Nat. Plants* 5, 722–730. doi: 10.1038/s41477-019-0459-z

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., and Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 25, 148–158. doi: 10.1016/j.tplants.2019.10.012

Khan, S. U., Yangmiao, J., Liu, S., Zhang, K., Khan, M. H. U., Zhai, Y., et al. (2019). Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in *Brassica napus* L. *Ind. Crop Prod.* 142:111877. doi: 10.1016/j.indcrop.2019.111877

Khush, G. S., Lee, S., Cho, J.-I., and Jeon, J.-S. (2012). Biofortification of crops for reducing malnutrition. *Plant Biotechnol. Rep.* 6, 195–202. doi: 10.1007/s11816-012-0216-5

Kim, J. A., Jung, H. E., Hong, J. K., Hermand, V., McClung, C. R., Lee, Y. H., et al. (2016). Reduction of GIGANTEA expression in transgenic *Brassica rapa* enhances salt tolerance. *Plant Cell Rep.* 35, 1943–1954. doi: 10.1007/s00299-016-2008-9

Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc. Natl Acad. Sci. U.S.A.* 93, 1156–1160. doi: 10.1073/pnas.93.3.1156

Kioukis, A., Michalopoulou, V. A., Briers, L., Pirintsos, S., Studholme, D. J., Pavlidis, P., et al. (2020). Intraspecific diversification of the crop wild relative *Brassica cretica* Lam. using demographic model selection. *BMC Genomics* 21:48. doi: 10.1186/s12864-019-6439-x

Kirchner, T.W., Niehaus, M., Rössig, K.L., Lauterbach, T., Herde, M., Küster, H., et al. (2018). Molecular background of Pi deficiency-induced root hair growth in *Brassica carinata*?a fasciclin-like arabinogalactan protein is involved. *Front. Plant Sci.* 9:1372. doi: 10.3389/fpls.2018.01372

Kitashiba, H., Li, F., Hirakawa, H., Kawanabe, T., Zou, Z., Hasegawa, Y., et al. (2014). Draft sequences of the radish (*Raphanus sativus* L.) genome. *DNA Res.* 21, 481–490. doi: 10.1093/dnares/dsu014

Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313X.2005.02533.x

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. doi: 10.1038/nature17946

Koscielny, C. B., Gardner, S. W., Technow, F., and Duncan, R. W. (2020). Linkage mapping and whole-genome predictions in canola (*Brassica napus*) subjected to differing temperature treatments. *Crop Pasture Sci.* 71, 229–238. doi: 10.1071/CP19387

Kumar, M. S., Mawlong, I., and Rani, R. (2020). "Biofortification of *Brassicas* for quality improvement," in *Brassica Improvement: Molecular, Genetics and Genomic Perspectives*, eds S. H. Wani, A. K. Thakur and Y. Jeshima Khan (Cham: Springer International Publishing), 127–145.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., et al. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome Biol.* 16:258. doi: 10.1186/s13059-015-0826-7

Lee, H., Chawla, H. S., Obermeier, C., Dreyer, F., Abbadi, A., and Snowdon, R. (2020). Chromosome-scale assembly of winter oilseed rape *Brassica napus*. *Front. Plant Sci.* 11:496. doi: 10.3389/fpls.2020.00496

Leff, B., Ramankutty, N., and Foley, J. A. (2004). Geographic distribution of major crops across the world. *Glob. Biogeochem. Cycles* 18:1. doi: 10.1029/2003gb002108

Li, C., Hao, M., Wang, W., Wang, H., Chen, F., and Chu, W., et al. (2018). An efficient CRISPR/Cas9 platform for rapidly generating simultaneous mutagenesis of multiple gene homoeologs in allotetraploid oilseed rape. *Front. Plant Sci.* 9:442. doi: 10.3389/fpls.2018.00442

Li, H., Cheng, X., Zhang, L., Hu, J., Zhang, F., Chen, B., et al. (2018). An integration of genome-wide association study and gene co-expression network analysis identifies candidate genes of stem lodging-related traits in *Brassica napus*. *Front. Plant Sci.* 9:796. doi: 10.3389/fpls.2018.00796

Li, J., Rao, L., Meng, Q., Ghani, M. A., and Chen, L. (2015). Production of Brassica tri-genomic vegetable germplasm by hybridisation between tuber mustard (*Brassica juncea*) and red cabbage (*B. oleracea*). *Euphytica* 204, 323–333. doi: 10.1007/s10681-014-1336-5

Li, L., Liu, Y., Chen, B., Xu, K., Zhang, F., Li, H., et al. (2016). A genome-wide association study reveals new loci for resistance to clubroot disease in *Brassica napus*. *Front. Plant Sci.* 7:1483. doi: 10.3389/fpls.2016.01483

Li, P., Zhang, S., Li, F., Zhang, S., Zhang, H., Wang, X., et al. (2017). A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important *Brassica* species comprising the Triangle of U. *Front. Plant Sci.* 8:111. doi: 10.3389/fpls.2017.00111

Li, T., Lui, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199

Li, Y., Liu, G-F., Ma, L-M., Liu, T-K., Zhang, C-W., Xiao, D., et al. (2020). A chromosome-level reference genome of non-heading Chinese cabbage [*Brassica campestris* (syn. *Brassica rapa*) ssp. chinensis]. *Hortic. Res.* 7:212. doi: 10.1038/s41438-020-00449-z

Li, Y., and Tollefsbol, T. O. (2011). Combined chromatin immunoprecipitation and bisulfite methylation sequencing analysis. *Methods Mol. Biol.* 791, 239–251. doi: 10.1007/978-1-61779-316-5_18

Lima, M. S., Woods, L. C., Cartwright, M. W., and Smith, D. R. (2016). The (in)complete organelle genome: exploring the use and nonuse of available technologies for characterizing mitochondrial and plastid chromosomes. *Mol. Ecol. Res.* 16, 1279–1286. doi: 10.1111/1755-0998.12585

Limera, C., Sabbadini, S., Sweet, J. B., and Mezzetti, B. (2017). New biotechnological tools for the genetic improvement of major woody fruit species. *Front. Plant Sci.* 8:1418. doi: 10.3389/fpls.2017.01418

Lin, J., and Musunuru, K. (2016). Genome engineering tools for building cellular models of disease. *FEBS J.* 283, 3222–3231. doi: 10.1111/febs.13763

Lin, K., Zhang, N., Severing, E. I., Nijveen, H., Cheng, F., Visser, R. G. F., et al. (2014). Beyond genomic variation - comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics* 15:250. doi: 10.1186/1471-2164-15-250

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536. doi: 10.1016/j.cell.2008.03.029

Liu, C. W., Lin, C. C., Yiu, J. C., Chen, J. J., and Tseng, M. J. (2008). Expression of a *Bacillus thuringiensis* toxin (*cry1Ab*) gene in cabbage (*Brassica oleracea* L. *var. capitata* L.) chloroplasts confers high insecticidal efficacy against *Plutella xylostella*. *Theor. Appl. Genet.* 117, 75–88. doi: 10.1007/s00122-008-0754-y

Liu, G., Xia, Y., Liu, T., Dai, S., and Hou, X. (2018). The DNA methylome and association of differentially methylated regions with differential gene expression during heat stress in *Brassica rapa*. *Int. J. Mol. Sci.* 19:1414. doi: 10.3390/ijms19051414

Liu, J., Li, M., Zhang, Q., Wei, X., and Huang, X. (2020). Exploring the molecular basis of heterosis for plant breeding. *J. Integr. Plant Biol.* 62, 287–298. doi: 10.1111/jipb.12804

Liu, P., Zhao, Y., Liu, G., Wang, M., Hu, D., Hu, J., et al. (2017). Hybrid performance of an immortalized F_2 rapeseed population is driven by additive, dominance, and epistatic effects. *Front. Plant Sci.* 8:815. doi: 10.3389/fpls.2017.00815

Liu, S., Li, L., Gao, W., Zhang, Y., Liu, Y., Wang, S., et al. (2018). Diagnosis of nitrogen status in winter oilseed rape (*Brassica napus* L.) using *in-situ* hyperspectral data and unmanned aerial vehicle (UAV) multispectral images. *Comput. Electron. Agric.* 151, 185–195. doi: 10.1016/j.compag.2018.05.026

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., et al. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5:3930. doi: 10.1038/ncomms4930

Liu, T., Li, Y., Duan, W., Huang, F., and Hou, X. (2017). Cold acclimation alters DNA methylation patterns and confers tolerance to heat and increases growth rate in *Brassica rapa*. *J. Exp. Bot.* 68, 1213–1224. doi: 10.1093/jxb/erw496

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS Comput. Biol.* 13:e1005457. doi: 10.1371/journal.pcbi.1005457

Lu, G., Harper, A. L., Trick, M., Morgan, C., Fraser, F., O'Neill, C., et al. (2014). Associative transcriptomics study dissects the genetic architecture of seed glucosinolate content in *Brassica napus*. *DNA Res.* 21, 613–625. doi: 10.1093/dnares/dsu024

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* 10:1154. doi: 10.1038/s41467-019-09134-9

Lu, R., Martin-Hernandez, A. M., Peart, J. R., Malcuit, I., and Baulcombe, D. C. (2003). Virus-induced gene silencing in plants. *Methods* 30, 296–303. doi: 10.1016/S1046-2023(03)00037-9

Lu, S., Van Eck, J., Zhou, X., Lopez, A. B., O'Halloran, D. M., Cosman, K. M., et al. (2006). The cauliflower *Or* gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of β-carotene accumulation. *Plant Cell* 18, 3594–3605. doi: 10.1105/tpc.106.046417

Lu, Y., and Yao, J. (2018). Chloroplasts at the crossroad of photosynthesis, pathogen infection and plant defense. *Int. J. Mol. Sci.* 19:3900. doi: 10.3390/ijms19123900

Lusser, M., and Davies, H. V. (2013). Comparative regulatory approaches for groups of new plant breeding techniques. *New Biotechnol.* 30, 437–446. doi: 10.1016/j.nbt.2013.02.004

Lusser, M., Parisi, C., Plan, D., and Rodríguez-Cerezo, E. (2012). Deployment of new biotechnologies in plant breeding. *Nat. Biotechnol.* 30, 231–239. doi: 10.1038/nbt.2142

Lv, H., Wang, Y., Han, F., Ji, J., Fang, Z., Zhuang, M., et al. (2020). A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. *Sci. Rep.* 10:12394. doi: 10.1038/s41598-020-69389-x

Ma, C., Zhu, C., Zheng, M., Liu, M., Zhang, D., Liu, B., et al. (2019). CRISPR/Cas9-mediated multiple gene editing in *Brassica oleracea* var. *capitata* using the endogenous tRNA-processing system. *Hortic. Res.* 6:20. doi: 10.1038/s41438-018-0107-1

Makaroff, C. A., and Palmer, J. D. (1987). Extensive mitochondrial specific transcription of the *Brassica campestris* mitochondrial genome. *Nucleic Acids Res.* 15, 5141–5156. doi: 10.1093/nar/15.13.5141

Malek, M., Rahman, L., Das, M., Hassan, L., and Rafii, M. (2013). Development of hexaploid 'Brassica' (AABBCC) from hybrids (ABC) of '*Brassica carinata*' (BBCC) x *B. rapa* (AA). *Aust. J. Crop Sci.* 7, 1375–1382.

Malmberg, M. M., Shi, F., Spangenberg, G. C., Daetwyler, H. D., and Cogan, N. O. I. (2018). Diversity and genome analysis of Australian and global oilseed *Brassica napus* L. germplasm using transcriptomics and whole genome re-sequencing. *Front. Plant Sci.* 9:508. doi: 10.3389/fpls.2018.00508

Marconi, G., Pace, R., Traini, A., Raggi, L., Lutts, S., Chiusano, M., et al. (2013). Use of MSAP markers to analyse the effects of salt stress on DNA methylation in rapeseed (*Brassica napus* var. *oleifera*). *PLoS ONE.* 8:e75597. doi: 10.1371/journal.pone.0075597

Marri, P. R., Ye, L., Jia, Y., Jiang, K., and Rounsley, S. D. (2018). "Advances in sequencing and resequencing in crop plants," in *Plant Genetics and Molecular Biology*, eds R. K. Varshney, M. K. Pandey, and A. Chitikineni (Cham: Springer International Publishing), 11–35.

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., and Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51, 1076–1081. doi: 10.1038/s41588-019-0443-6

Mayer, A. (1997). Historical changes in the mineral content of fruits and vegetables. *Br. Food J.* 99, 207–211. doi: 10.1108/00070709710181540

McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123, 439–442. doi: 10.1104/pp.123.2.439

Mei, J., Shao, C., Yang, R., Feng, Y., Gao, Y., Ding, Y., et al. (2020). Introgression and pyramiding of genetic loci from wild *Brassica oleracea* into *B. napus* for improving Sclerotinia resistance of rapeseed. *Theor. Appl. Genet.* 133, 1313–1319. doi: 10.1007/s00122-020-03552-w

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Michael, T. P., Jupe, F., Bemm, F., Motley, S. T., Sandoval, J. P., Lanz, C., et al. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* 9:541. doi: 10.1038/s41467-018-03016-2

Ming, R., and Man Wai, C. (2015). Assembling allopolyploid genomes: no longer formidable. *Genome Biol.* 16:27. doi: 10.1186/s13059-015-0585-5

Moeckel, T., Dayananda, S., Nidamanuri, R. R., Nautiyal, S., Hanumaiah, N., Buerkert, A., et al. (2018). Estimation of vegetable crop parameter by multi-temporal UAV-borne images. *Remote Sens.* 10:805. doi: 10.3390/rs10050805

Moghe, G. D., Hufnagel, D. E., Tang, H., Xiao, Y., Dworkin, I., Town, C. D., et al. (2014). Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. *Plant Cell* 26, 1925–1937. doi: 10.1105/tpc.114.124297

Moloney, M. M., Walker, J. M., and Sharma, K. K. (1989). High efficiency transformation of *Brassica napus* using *Agrobacterium* vectors. *Plant Cell Rep.* 8, 238–242. doi: 10.1007/BF00778542

Murray, G. M., and Brennan, J. P. (2012). *The Current and Potential Costs From Diseases of Oilseed Crops in Australia*. Grains Research and Development Corporation. Available online at: https://grdc.com.au/__data/assets/pdf_file/0021/82641/grdcreportdiseasecostoilseedspdf.pdf.pdf

Mwathi, M. W., Gupta, M., Quezada-Martinez, D., Pradhan, A., Batley, J., Mason, A. S., et al. (2020). Fertile allohexaploid *Brassica* hybrids obtained from crosses between *B. oleracea* and *B. juncea* via ovule rescue and colchicine treatment of cuttings. *Plant Cell Tissue Organ Cult.* 140, 301–313. doi: 10.1007/s11240-019-01728-x

Nagaharu, U., Nagaharu, N., and Nagaharu. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7, 389–452.

Napier, J. A., Olsen, R. E., and Tocher, D. R. (2019). Update on GM canola crops as novel sources of omega-3 fish oils. *Plant Biotechnol. J.* 17, 703–705. doi: 10.1111/pbi.13045

Navarro-León, E., Ruiz, J. M., Graham, N., and Blasco, B. (2018). Physiological profile of CAX1a TILLING mutants of *Brassica rapa* exposed to different calcium doses. *Plant Sci.* 272, 164–172. doi: 10.1016/j.plantsci.2018.04.019

Nepolean, T., Kaul, J., Mukri, G., and Mittal, S. (2018). Genomics-enabled next-generation breeding approaches for developing system-specific drought tolerant hybrids in maize. *Front. Plant Sci.* 9:361. doi: 10.3389/fpls.2018.00361

Okuzaki, A., Ogawa, T., Koizuka, C., Kaneko, K., Inaba, M., Imamura, J., et al. (2018). CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. *Plant Physiol. Biochem.* 131, 63–69. doi: 10.1016/j.plaphy.2018.04.025

Palmer, J. D., and Herbon, L. A. (1986). Tricircular mitochondrial genomes of *Brassica* and *Raphanus*: Reversal of repeat configurations by inversion. *Nucleic Acids Res.* 14, 9755–9764. doi: 10.1093/nar/14.24.9755

Palmer, J. D., and Herbon, L. A. (1988). Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J. Mol. Evol.* 28, 87–97. doi: 10.1007/BF02143500

Palmer, J. D., and Shields, C. R. (1984). Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature* 307, 437–440. doi: 10.1038/307437a0

Paritosh, K., Yadava, S. K., Singh, P., Bhayana, L., Mukhopadhyay, A., Gupta, V., et al. (2020). A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. *Plant Biotechnol. J.* doi: 10.1111/pbi.13492

Parkin, I. A. P., Koh, C., Tang, H., Robinson, S. J., Kagale, S., Clarke, W. E., et al. (2014). Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 15:R77. doi: 10.1186/gb-2014-15-6-r77

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sakar, S., and Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. *Sci. Rep.* 9:17132. doi: 10.1038/s41598-019-53451-4

Paszkowski, J., and Whitham, S. A. (2001). Gene silencing and DNA methylation processes. *Curr. Opin. Plant Biol.* 4, 123–129. doi: 10.1016/S1369-5266(00)00147-3

Payá-Milans, M., Poza-Viejo, L., Martín-Uriz, P. S., Lara-Astiaso, D., Wilkinson, M. D., and Crevillén, P. (2019). Genome-wide analysis of the H3K27me3 epigenome and transcriptome in *Brassica rapa*. *GigaScience* 8:12. doi: 10.1093/gigascience/giz147

Pe'ery, T., Mathews, M. B., and Baulcombe, D. (2003). RNA interference. *Methods* 30, 287–288. doi: 10.1016/S1046-2023(03)00035-5

Pérez-de-Castro, A. M., Vilanova, S., Cañizares, J., Pascual, L., Blanca, J. M., Díez, M. J., et al. (2012). Application of genomic tools in plant breeding. *Curr. Genomics* 13, 179–195. doi: 10.2174/138920212800543084

Perrella, G., and Kaiserli, E. (2016). Light behind the curtain: Photoregulation of nuclear architecture and chromatin dynamics in plants. *New Phytol.* 212, 908–919. doi: 10.1111/nph.14269

Perumal, S., Koh, C. S., Jin, L., Buchwaldt, M., Higgins, E. E., Zheng, C., et al. (2020). A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. *Nat. Plants* 6, 929–941. doi: 10.1038/s41477-020-0735-y

Prabhudas, S. K., Raju, B., Kannan Thodi, S., Parani, M., and Natarajan, P. (2016). The complete chloroplast genome sequence of Indian mustard (*Brassica juncea* L.). *Mitochondrial DNA Part A DNA Mapp. Seq. Anal.* 27, 4622–4623. doi: 10.3109/19401736.2015.1101586

Pradhan, A., Plummer, J. A., Nelson, M. N., Cowling, W. A., and Yan, G. (2010). Successful induction of trigenomic hexaploid *Brassica* from a triploid hybrid of *B. napus* L. and *B. nigra* (L.) Koch. *Euphytica* 176: 87–98. doi: 10.1007/s10681-010-0218-8

Pröbsting, M., Schenke, D., Hossain, R., Häder, C., Thurau, T., Wighardt, L., et al. (2020). Loss of function of CRT1a (calreticulin) reduces plant susceptibility to *Verticillium longisporum* in both *Arabidopsis thaliana* and oilseed rape (*Brassica napus*). *Plant Biotechnol. J.* 18, 2328–2344. doi: 10.1111/pbi.13394

Rahman, H. (2013). Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. *Can. J. Plant Sci.* 93, 363–373. doi: 10.4141/cjps2012-074

Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I., and Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends Genet.* 18, 348–351. doi: 10.1016/S0168-9525(02)02709-9

Raman, H., Uppal, R. K., and Raman, R. (2019). "Genetic solutions to improve resilience of canola to climate change," in *Genomic Designing of Climate-Smart Oilseed Crops*, ed C. Kole (Cham: Springer International Publishing), 75–131.

Ramegowda, V., Mysore, K. S., and Senthil-Kumar, M. (2014). Virus-induced gene silencing is a versatile tool for unraveling the functional relevance of multiple abiotic-stress-responsive genes in crop plants. *Front. Plant Sci.* 5:323. doi: 10.3389/fpls.2014.00323

Ran, Y., Liang, Z., and Gao, C. (2017). Current and future editing reagent delivery systems for plant genome editing. *Sci. China Life Sci.* 60, 490–505. doi: 10.1007/s11427-017-9022-1

Rando, O. J., and Ahmad, K. (2007). Rules and regulation in the primary structure of chromatin. *Curr. Opin. Cell Biol.* 19, 250–256. doi: 10.1016/j.ceb.2007.04.006

Rao, V. R., and Hodgkin, T. (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tissue Organ Cult.* 68, 1–19. doi: 10.1023/A:1013359015812

Rashid, M., He, G., Guanxiao, Y., and Ziaf, K. (2011). Relevance of TILLING in plant genomics. *Aust. J. Crop Sci.* 5, 411–420.

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., and Chatterjee, S. (2019). Climate change has likely already affected global food production. *PLoS ONE* 14:e0217148. doi: 10.1371/journal.pone.0217148

Razzaq, A., Saleem, F., Kanwal, M., Mustafa, G., Yousaf, S., Imran Arshad, H. M., et al. (2019). Modern trends in plant genome editing: an inclusive review of the CRISPR/Cas9 toolbox. *Int. J. Mol. Sci.* 20:4045. doi: 10.3390/ijms20164045

Richards, E. J. (2006). Inherited epigenetic variation — revisiting soft inheritance. *Nat. Rev. Genet.* 7, 395–401. doi: 10.1038/nrg1834

Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., et al. (2020). Long-reads assembly of the *Brassica napus* reference genome, Darmor-bzh. *GigaScience* 9:12. doi: 10.1093/gigascience/giaa137

Ruiter, R., van den Brande, I., Stals, E., Delauré, S., Cornelissen, M., and D'Halluin, K. (2003). Spontaneous mutation frequency in plants obscures the effect of chimeraplasty. *Plant Mol. Biol.* 53, 675–689. doi: 10.1023/B:PLAN.0000019111.96107.01

Saini, P., Saini, P., Kaur, J. J., Francies, R. M., Gani, M., Rajendra, A. A., et al. (2020). "Molecular approaches for harvesting natural diversity for crop improvement," in *Rediscovery of Genetic and Genomic Resources for Future Food Security*, eds R. K. Salgotra and S. M. Zargar (Singapore: Springer), 67–169.

Sang, S., Cheng, H., Mei, D., Fu, L., Wang, H., Liu, J., et al. (2020). Complete organelle genomes of *Sinapis arvensis* and their evolutionary implications. *Crop J.* 8, 505–514. doi: 10.1016/j.cj.2019.12.001

Sasaki, T., Fujimoto, R., Kishitani, S., and Nishio, T. (2011). Analysis of target sequences of DDM1s in *Brassica rapa* by MSAP. *Plant Cell Rep.* 30, 81–88. doi: 10.1007/s00299-010-0946-1

Sashidhar, N., Harloff, H.-J., and Jung, C. (2019). Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large-scale screening of mutant populations through amplicon sequencing. *New Phytol.* 225: 2022–2034. doi: 10.1111/nph.16281

Sashidhar, N., Harloff, H. J., Potgieter, L., and Jung, C. (2020). Gene editing of three *BnITPK* genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnol. J.* 18, 2241–2250. doi: 10.1111/pbi.13380

Sauer, N. J., Narváez-Vásquez, J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Woodward, M. J., et al. (2016). Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol.* 170, 1917–1928. doi: 10.1104/pp.15.01696

Saxena, K. B., and Hingane, A. J. (2015). "Male sterility systems in major field crops and their potential role in crop improvement," in *Plant Biology and Biotechnology*, eds B. Bahadur, M. Venkat Rajam, L. Sahijram, and K. Krishnamurthy (New Delhi: Springer), 639–656.

Scheben, A., Berpaalen, B., Lawley, C. T., Chan, C. K. K., Bayer, P. E., Batley, J., et al. (2019). CropSNPdb: a database of SNP array data for Brassica crops and hexaploid bread wheat. *Plant J.* 98, 142–152. doi: 10.1111/tpj.14194

Scheben, A., Wolter, F., Batley, J., Puchta, H., and Edwards, D. (2017). Towards CRISPR/Cas crops bringing together genomics and genome editing. *New Phytol.* 216, 682–698. doi: 10.1111/nph.14702

Scheben, A., Yuan, Y., and Edwards, D. (2016). Advances in genomics for adopting crops to climate change. *Curr. Plant Biol.* 6, 2–10. doi: 10.1016/j.cpb.2016.09.001

Scholze, H., and Boch, J. (2010). TAL effector-DNA specificity. *Virulence* 1: 428–432. doi: 10.4161/viru.1.5.12863

Schreiber, M., Stein, N., and Mascher, M. (2018). Genomic approaches for studying crop evolution. *Genome Biol.* 19:140. doi: 10.1186/s13059-018-1528-8

Schuler, T. H., Denholm, I., Clark, S. J., Stewart, C. N., and Poppy, G. M. (2004). Effects of Bt plants on the development and survival of the parasitoid *Cotesia plutellae* (Hymenoptera: Braconidae) in susceptible and Bt-resistant larvae of the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). *J. Insect Physiol.* 50, 435–443. doi: 10.1016/j.jinsphys.2004.03.001

Senthil-Kumar, M., and Mysore, K. S. (2011). New dimensions for VIGS in plant functional genomics. *Trends Plant Sci.* 16, 656–665. doi: 10.1016/j.tplants.2011.08.006

Seol, Y.-J., Kim, K., Kang, S.-H., Perumal, S., Lee, J., and Kim, C.-K. (2017). The complete chloroplast genome of two *Brassica* species, *Brassica nigra* and *B. oleracea*. *Mitochondrial DNA Part A DNA Mapp. Seq. Appl.* 28, 167–168. doi: 10.3109/19401736.2015.1115493

Seymour, D. K., and Becker, C. (2017). The causes and consequences of DNA methylome variation in plants. *Curr. Opin. Plant Biol.* 36, 56–63. doi: 10.1016/j.pbi.2017.01.005

Sharafi, Y., Majidi, M. M., Goli, S. A. H., and Rashidi, F. (2015). Oil content and fatty acids composition in *Brassica* species. *Int. J. Food Prop.* 18, 2145–2154. doi: 10.1080/10942912.2014.968284

Shen, X., Xu, L., Liu, Y., Dong, H., Zhou, D., Zhang, Y., et al. (2019). Comparative transcriptome analysis and ChIP-sequencing reveals stage-specific gene expression and regulation profiles associated with pollen wall formation in *Brassica rapa*. *BMC Genomics* 20:264. doi: 10.1186/s12864-019-5637-x

Shiba, H., Kakizaki, T., Iwano, M., Tarutani, Y., Watanabe, M., Isogai, A., et al. (2006). Dominance relationships between self-incompatibility alleles controlled by DNA methylation. *Nat. Genet.* 38, 297–299. doi: 10.1038/ng1734

Shirasawa, K., Hirakawa, H., Fukino, N., Kitashiba, H., and Isobe, S. (2020). Genome sequence and analysis of a Japanese radish (*Raphanus sativus*) cultivar named 'Sakurajima Daikon' possessing giant root. *DNA Res.* 27:2. doi: 10.1093/dnares/dsaa010

Siche, R. (2020). What is the impact of COVID-19 disease on agriculture? *Sci. Agrop.* 11, 3–6. doi: 10.17268/sci.agropecu.2020.01.00

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., et al. (2019). High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Front. Plant Sci.* 10:394. doi: 10.3389/fpls.2019.00394

Singh, N., Wu, S., Raupp, W. J., Sehgal, S., Arora, S., Tiwari, V., et al. (2019). Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Sci. Rep.* 9:650. doi: 10.1038/s41598-018-37269-0

Slatkin, M. (2008). Linkage disequilibrium? understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 9, 477–485. doi: 10.1038/nrg2361

Slotte, T., Hazzouri, K. M., Ågren, J. A., Koenig, D., and Maumus, F., Guo, Y.-L., et al. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831–835. doi: 10.1038/ng.2669

Snowdon, R. J., Friedrich, T., Friedt, W., and Köhler, W. (2002). Identifying the chromosomes of the A- and C-genome diploid *Brassica* species *B. rapa (syn. campestris)* and *B. oleracea* in their amphidiploid *B. napus*. *Theor. Appl. Genet.* 104, 533–538. doi: 10.1007/s00122-001-0787-y

Solís, M. T., El-Tantawy, A. A., Cano, V., Risueño, M. C., and Testillano, P. S. (2015). 5-azacytidine promotes microspore embryogenesis initiation by decreasing global DNA methylation, but prevents subsequent embryo development in rapeseed and barley. *Front. Plant Sci.* 6:472. doi: 10.3389/fpls.2015.00472

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Sottosanto, J., Andre, C., Arias, D. I., Bhatti, M., Breazeale, S., Fu, H., et al. (2018). *Petition for the Determination of Non-regulatory Status for EPA+DHA Canola Event LBFLFK*. Available online at: https://www.aphis.usda.gov/brs/aphisdocs/17_32101p.pdf

Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G. J., et al. (2010). A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biol.* 10:62. doi: 10.1186/1471-2229-10-62

Stöckle, C. O., and Kemanian, A. R. (2020). Can crop models identify critical gaps in genetics, environment, and management interactions? *Front. Plant Sci.* 11:737. doi: 10.3389/fpls.2020.00737

Sun, D., Wang, C., Zhang, X., Zhang, W., Jiang, H., Yao, X., et al. (2019). Draft genome sequence of cauliflower (*Brassica oleracea* L. *var. botrytis*) provides new insights into the C genome in *Brassica* species. *Hortic. Res.* 6:82. doi: 10.1038/s41438-019-0164-0

Sun, F., Fan, G., Hu, Q., Zhou, Y., Guan, M., Tong, C., et al. (2017). The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. *Plant J.* 92, 452–468. doi: 10.1111/tpj.13669

Sun, Q., Lin, L., Liu, D., Wu, D., Fang, Y., Wu, J., et al. (2018). CRISPR/Cas9-mediated multiplex genome editing of the *BnWRKY11* and *BnWRKY70* genes in *Brassica napus* L. *Int. J. Mol. Sci.* 19:2716. doi: 10.3390/ijms19092716

Sun, Z., Li, N., Huang, G., Xu, J., Pan, Y., Wang, Z., et al. (2013). Site-specific gene targeting using transcription activator-like effector (TALE)-based nuclease in *Brassica oleracea*. *J. Integr. Plant Biol.* 55, 1092–1103. doi: 10.1111/jipb.12091

Supple, M. A., and Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biol.* 19:131. doi: 10.1186/s13059-018-1520-3

Sweetman, C., Waterman, C. D., Rainbird, B. M., Smith, P., Jenkins, C. D., Day, D. A., et al. (2019). *AtNDB2* is the main external NADH dehydrogenase in mitochondria and is important for tolerance to environmental stress. *Plant Physiol.* 181, 774–788. doi: 10.1104/pp.19.00877

Takahashi, S., Fukushima, N., Osabe, K., Itabashi, E., Shimizu, M., Miyaji, N., et al. (2018a). Identification of DNA methylated regions by using methylated DNA immunoprecipitation sequencing in *Brassica rapa*. *Crop Pasture Sci.* 69, 107–120. doi: 10.1071/CP17394

Takahashi, S., Osabe, K., Fukushima, N., Takuno, S., Miyaji, N., Shimizu, M., et al. (2018b). Genome-wide characterization of DNA methylation, small RNA expression, and histone H3 lysine nine di-methylation in *Brassica rapa* L. *DNA Res.* 25, 511–520. doi: 10.1093/dnares/dsy021

Tan, C., Liu, H., Ren, J., Ye, X., Feng, H., and Liu, Z. (2019). Single-molecule real-time sequencing facilitates the analysis of transcripts and splice isoforms of anthers in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *BMC Plant Biol.* 1:517. doi: 10.1186/s12870-019-2133-z

Tanaka, Y., Tsuda, M., Yasumoto, K., Terachi, T., and Yamagishi, H. (2014). The complete mitochondrial genome sequence of *Brassica oleracea* and analysis of coexisting mitotypes. *Curr. Genet.* 60, 277–284. doi: 10.1007/s00294-014-0433-2

Tang, T., Yu, X., Yang, H., Gao, Q., Ji, H., Wang, Y., et al. (2018). Development and validation of an effective CRISPR/Cas9 vector for efficiently isolating positive transformants and transgene-free mutants in a wide range of plant species. *Front. Plant Sci.* 9:1533. doi: 10.3389/fpls.2018.01533

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Tariq, M., and Paszkowski, J. (2004). DNA and histone methylation in plants. *Trends Genet.* 20, 244–251. doi: 10.1016/j.tig.2004.04.005

Tennessen, J. A., Govindarajulu, R., Ashman, T.-L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* 6, 3295–3313. doi: 10.1093/gbe/evu261

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692

Tian, E., Jiang, Y., Chen, L., Liu, F., and Menget, J. (2010). Synthesis of a *Brassica* trigenomic allohexaploid (*B. carinata* × *B. rapa) de novo* and its stability in subsequent generations. *Theor. Appl. Genet.* 121, 1431–1440. doi: 10.1007/s00122-010-1399-1

Tirnaz, S., and Batley, J. (2019a). DNA methylation: toward crop disease resistance improvement. *Trends Plant Sci.* 24, 1137–1150. doi: 10.1016/j.tplants.2019.08.007

Tirnaz, S., and Batley, J. (2019b). Epigenetics: potentials and challenges in crop breeding. *Mol Plant* 12, 1309–1311. doi: 10.1016/j.molp.2019.09.006

Tirnaz, S., Merce, C., Bayer, P. E., Severn-Ellis, A. A., Edwards, D., and Batley, J. (2020). Effect of *Leptosphaeria maculans* infection on promoter DNA methylation of defence genes in *Brassica napus*. *Agronomy* 10:1072. doi: 10.3390/agronomy10081072

Townsend, A. R., Howarth, R. W., Bazzaz, F. A., Booth, M. S., Cleveland, C. C., Collinge, S. K., et al. (2003). Human health effects of a changing global nitrogen cycle. *Front. Ecol. Environ.* 1, 240–246. doi: 10.1890/1540-9295(2003)001[0240:HHEOAC]2.0.CO;2

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Front. Physiol.* 3:347. doi: 10.3389/fphys.2012.00347

Van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* 39:e1869. doi: 10.3791/1869

Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The third revolution in sequencing technology. *Trends Genet.* 34, 666–681. doi: 10.1016/j.tig.2018.05.008

Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2018). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* 132, 797–816. doi: 10.1007/s00122-018-3252-x

Vernikos, G. S. (2020). "A review of pangenome tools and recent studies," in *The Pangenome,* eds H. Tettelin, and D. Medini (Cham: Springer), 89–112.

Voss-Fels, K., and Snowdon, R. J. (2016). Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnol. J.* 14, 1086–1094. doi: 10.1111/pbi.12456

Voytas, D. F., and Gao, C. (2014). Precision genome engineering and agriculture: opportunities and regulatory challenges. *PLoS Biol.* 12:e1001877. doi: 10.1371/journal.pbio.1001877

Wan, L., Li, Y., Cen, H., Zhu, J., Yin, W., Wu, W., et al. (2018). Combining UAV-based vegetation indices and image classification to estimate flower number in oilseed rape. *Remote Sens.* 10:1484. doi: 10.3390/rs10091484

Wang, B., Kumar, V., Olson, A., and Ware, D. (2019). Reviving the transcriptome studies: an insight into the emergence of single-molecule transcriptome sequencing. *Front. Genet.* 1:384. doi: 10.3389/fgene.2019.00384

Wang, G.-X., Lv, J., Zhang, J., Han, S., Zong, M., Guo, N., et al. (2016). Genetic and epigenetic alterations of *Brassica nigra* introgression lines from somatic hybridization: a resource for cauliflower improvement. *Front. Plant Sci.* 7:1258. doi: 10.3389/fpls.2016.01258

Wang, N., Shi, L., Tian, F., Ning, H., Wu, X., Long, Y., et al. (2010). Assessment of *FAE1* polymorphisms in three *Brassica* species using EcoTILLING and their association with differences in seed erucic acid contents. *BMC Plant Biol.* 10:137. doi: 10.1186/1471-2229-10-137

Wang, N., Wang, Y., Tian, F., King, G. J., Zhang, C., Long, Y., et al. (2008). A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAE1* point mutations by TILLING. *New Phytol.* 180, 751–765. doi: 10.1111/j.1469-8137.2008.02619.x

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic selection methods for crop improvement: current status and prospects. *Crop J.* 6, 330–340. doi: 10.1016/j.cj.2018.03.001

Wang, Y., Xiao, L., Guo, S., An, F., and Du, D. (2016). Fine mapping and whole-genome resequencing identify the seed coat color gene in *Brassica rapa*. *PLoS ONE* 11:e0166464. doi: 10.1371/journal.pone.0166464

Wang, Z., Wu, X., Wu, Z., An, H., Yi, B., Wen, J., et al. (2018). Genome-wide DNA methylation comparison between *Brassica napus* genic male sterile line and restorer line. *Int. J. Mol. Sci.* 19:2689. doi: 10.3390/ijms19092689

Watanabe, K., Sassa, Y., Suda, E., Chen, C.-H., Inaba, M., and Kikuchi, A. (2005). Global political, economic, social and technological issues on transgenic crops. *Plant Biotechnol.* 22, 515–522. doi: 10.5511/plantbiotechnology.22.515

Weigel, D., and Colot, V. (2012). Epialleles in plant evolution. *Genome Biol.* 13:249. doi: 10.1186/gb-2012-13-10-249

Werner, C. R., Qian, L., Voss-Fels, K. P., Abbadi, A., Leckband, G., Frisch, M., et al. (2017). Genome-ide regression models considering general and specifc combining ability predict hybrid performance in oilseed rape with similar accuracy regardless of trait architecture. *Theor. Appl. Genet.* 131, 299–317. doi: 10.1007/s00122-017-3002-5

Werner, C. R., Voss-Fels, K. P., Miller, C. N., Qian, W., Hua, W., Guan, et al. (2018). Effective genomic selection in a narrow-genepool crop with low-density markers: Asian rapeseed as an example. *Plant Genome* 11:170084. doi: 10.3835/plantgenome2017.09.0084

White, P. J., and Broadley, M. R. (2005). Historical variation in the mineral composition of edible horticultural products. *J. Hortic. Sci. Biotechnol.* 80, 660–667. doi: 10.1080/14620316.2005.11511995

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4

Wolfram, K., Schmidt, J., Wray, V., Milkowski, C., Schliemann, W., and Strack, D. (2010). Profiling of phenylpropanoids in transgenic low-sinapine oilseed rape (*Brassica napus*). *Phytochemistry* 71, 1076–1084. doi: 10.1016/j.phytochem.2010.04.00

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., et al. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol. Plant.* 12, 30–43. doi: 10.1016/j.molp.2018.11.007

Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., et al. (2005). Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat. Biotechnol.* 23, 1013–1017. doi: 10.1038/nbt1107

Wu, J., Chen, C., Xian, G., Liu, D., Lin, L., Yin, S., et al. (2020a). Engineering herbicide-resistant oilseed rape by CRISPR/Cas9-mediated cytosine base-editing. *Plant Biotechnol. J.* 18, 1857–1859. doi: 10.1111/pbi.13368

Wu, J., Liu, B., Cheng, F., Ramchiary, N., Choi, S.R., Lim, Y.P. et al. (2012). Sequencing of chloroplast genome using whole cellular DNA and Solexa sequencing technology. *Front. Plant Sci.* 3:243. doi: 10.3389/fpls.2012.00243

Wu, J., Yan, G., Duan, Z., Wang, Z., Kang, C., Guo, L., et al. (2020b). Roles of the *Brassica napus* DELLA Protein BnaA6.RGA, in modulating drought tolerance by interacting with the ABA signalling component BnaA10.ABF2. *Front. Plant Sci.* 11:577. doi: 10.3389/fpls.2020.00577

Wu, J., Zhao, Q., Yang, Q., Liu, H., Li, Q., Yi, X., et al. (2016). Comparative transcriptomic analysis uncovers the complex genetic network for resistance to *Sclerotinia sclerotiorum* in *Brassica napus*. *Sci. Rep.* 6:19007. doi: 10.1038/srep19007

Würschum, T., Abel, S., and Zhao, Y. (2014). Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. *Plant Breed.* 133, 45–51. doi: 10.1111/pbr.12137

Xiao-Ming, Z., Junrui, W., Li, F., Sha, L., Hongbo, P., Lan, Q., et al. (2017). Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep.* 7:1555. doi: 10.1038/s41598-017-01518-5

Xie, T., Chen, X., Guo, T., Rong, H., Chen, Z., Sun, Q., et al. (2020). Targeted knockout of *BnTT2* homologues for yellow-seeded *Brassica napus* with reduced flavonoids and improved fatty acid composition. *J. Agric. Food Chem.* 68, 5676–5690. doi: 10.1021/acs.jafc.0c01126

Xin, Q., Wang, X., Gao, Y., Xu, D., Xie, Z., Dong, F., et al. (2020). Molecular mechanisms underpinning the multiallelic inheritance of MS5 in *Brassica napus*. *Plant J. Cell Mol. Biol.* 103, 1723–1734. doi: 10.1111/tpj.14857

Xiong, X., Zhou, D., Xu, L., Liu, T., Yue, X., Liu, W., et al. (2019). BcPME37c is involved in pollen intine formation in *Brassica campestris*. *Biochem. Biophys. Res. Commun.* 517, 63–68. doi: 10.1016/j.bbrc.2019.07.009

Xue, J.-Y., Wang, Y., Chen, M., Dong, S., Shao, Z.-Q., and Liu, Y. (2020). Maternal inheritance of U's triangle and evolutionary process of *Brassica* mitochondrial genomes. *Front. Plant Sci.* 11:805. doi: 10.3389/fpls.2020.00805

Yamagishi, H., Tanaka, Y., and Terachi, T. (2014). Complete mitochondrial genome sequence of black mustard (*Brassica nigra*; BB) and comparison with *Brassica oleracea* (CC) and *Brassica carinata* (BBCC). *Genome* 57, 577–582. doi: 10.1139/gen-2014-0165

Yan, C., Huang, Y., Liu, Z., Guo, F., Jiao, Z., Yang, W., et al. (2020). Rapid identification of yellow-flowered gene *Bofc* in cauliflower (*Brassica oleracea* var. botrytis) by bulked segregant analysis and whole-genome resequencing. *Euphytica* 216:26. doi: 10.1007/s10681-020-2560-9

Yan, G., Nelson, M., Pradhan, A., Mason, A., Weerakoon, S., Si, P., et al. (2009). Progress towards the creation of trigenomic brassica hexaploid populations. *SABRAO, J. Breed. Genet.* 41:00274.

Yang, H., Wu, J., Tang, T., Liu, K.-D., and Dai, C. (2017). CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in *Brassica napus*. *Sci. Rep.* 7:7489. doi: 10.1038/s41598-017-07871-9

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., et al. (2016a). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* 48, 1225–1232. doi: 10.1038/ng.3657

Yang, J., Liu, G., Zhao, N., Chen, S., Liu, D., Ma, W., et al. (2016b). Comparative mitochondrial genome analysis reveals the evolutionary rearrangement mechanism in *Brassica*. *Plant Biol.* 18, 527–536. doi: 10.1111/plb.12414

Yang, K., Nath, U. K., Biswas, M. K., Kayum, M. A., Yi, G-e., Lee, J., et al. (2018). Whole-genome sequencing of *Brassica oleracea* var. *capitata* reveals new diversity of the mitogenome. *PLoS ONE* 13:e0194356. doi: 10.1371/journal.pone.0194356

Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, J., et al. (2013). The reference genome of the halophytic plant *Eutrema salsugineum*. *Front. Plant Sci.* 4:46. doi: 10.3389/fpls.2013.00046

Yang, Y., Zhu, K., Li, H., Han, S., Meng, Q., Khan, S. U., et al. (2018). Precise editing of CLAVATA genes in *Brassica napus* L. regulates multilocular silique development. *Plant Biotechnol. J.* 16, 1322–1335. doi: 10.1111/pbi.12872

Yao, M., Guan, M., Zhang, Z., Cui, Y., Chen, H., Liu, W., et al. (2020). GWAS and co-expression network combination uncovers multigenes with close linkage effects on the oleic acid content accumulation in *Brassica napus*. *BMC Genomics* 21:320. doi: 10.1186/s12864-020-6711-0

You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. (2018). Development and applications of a high throughput genotyping tool for polyploid crops: single nucleotide polymorphism (SNP) array. *Front. Plant Sci.* 9:104. doi: 10.3389/fpls.2018.00104

Yousef, E., Müller, T., Börner, A., and Schmid, K. J. (2018). Comparative analysis of genetic diversity and differentiation of cauliflower (*Brassica oleracea* var. *botrytis)* accessions from two *ex situ* genebanks. *PLoS ONE* 13:e0192062. doi: 10.1371/journal.pone.0192062

Yu, J., Yang, X. D., Wang, Q., Gao, L. W., Yang, Y., Xiao, D., et al. (2018). Efficient virus-induced gene silencing in *Brassica rapa* using a turnip yellow mosaic virus vector. *Biol. Plant.* 62, 826–834. doi: 10.1007/s10535-018-0803-6

Zaman, Q. U., Chu, W., Hao, M., Shi, Y., Sun, M., Sang, S. F., et al. (2019). CRISPR/Cas9-mediated multiplex genome editing of JAGGED gene in *Brassica napus* L. *Biomolecules* 9:725. doi: 10.3390/biom9110725

Zarinpanjeh, N., Motallebi, M., Zamani, M. R., and Ziaei, M. (2016). Enhanced resistance to *Sclerotinia sclerotiorum* in *Brassica napus* by co-expression of defensin and chimeric chitinase genes. *J. Appl. Genet.* 57: 417–425. doi: 10.1007/s13353-016-0340-y

Zeng, C.-L., Wang, G.-Y., Wang, J.-B., Yan, G.-X., Chen, B.-Y., Xu, K., et al. (2012). High-throughput discovery of chloroplast and mitochondrial DNA polymorphisms in *Brassica*ceae species by ORG-EcoTILLING. *PLoS ONE* 7:e47284. doi: 10.1371/journal.pone.0047284

Zhai, Y., Cai, S., Hu, L., Yang, Y., Amoo, O., Fan, C., et al. (2019). CRISPR/Cas9-mediated genome editing reveals differences in the contribution of *INDEHISCENT* homologues to pod shatter resistance in *Brassica napus* L. *Theor. Appl. Genet.* 132, 2111–2123. doi: 10.1007/s00122-019-03341-0

Zhai, Y., Yu, K., Cai, S., Hu, L., Amoo, O., Xu, L., et al. (2020). Targeted mutagenesis of *BnTT8* homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotech J.* 18, 1153–1168. doi: 10.1111/pbi.13281

Zhang, A., Wang, H., Beyene, Y., Samagn, K., Liu, Y., Cao, S., et al. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Front. Plant Sci.* 8:1916. doi: 10.3389/fpls.2017.01916

Zhang, C., Wohlhueter, R., and Zhang, H. (2016). Genetically modified foods: a critical review of their promise and problems. *Food Sci. Hum. Wellness* 5, 116–123. doi: 10.1016/j.fshw.2016.04.002

Zhang, F., and Batley, J. (2020). Exploring the application of wild species for crop improvement in a changing climate. *Curr. Opin. Plant Biol.* 56, 218–222. doi: 10.1016/j.pbi.2019.12.013

Zhang, K., Nie, L., Cheng, Q., Yin, Y., Chen, K., Qi, F., et al. (2019). Effective editing for lysophosphatidic acid acyltransferase 2/5 in allotetraploid rapeseed (*Brassica napus* L.) using CRISPR-Cas9 system. *Biotechnol. Biofuels* 12:225. doi: 10.1186/s13068-019-1567-8

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., et al. (2018). Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic. Res.* 5:50. doi: 10.1038/s41438-018-0071-9

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207

Zhang, Y., Huang, S., Wang, X., Liu, J., Guo, X., Mu, J., et al. (2018a). Defective APETALA2 genes lead to sepal modification in *Brassica crops*. *Front. Plant Sci.* 9:367. doi: 10.3389/fpls.2018.00367

Zhang, Y., Massel, K., Godwin, I. D., and Gao, G. (2018b). Applications and potential of genome editing in crop improvement. *Genome Biol.* 19:210. doi: 10.1186/s13059-018-1586-y

Zhang, Y., Thomas, W., Bayer, P. E., Edwards, D., and Batley, J. (2020). Frontiers in dissecting and managing *Brassica* diseases: from reference-based RGA candidate identification to building pan-RGAomes. *Int. J. Mol. Sci.* 21:8964. doi: 10.3390/ijms21238964

Zhang, Z., Zhang, S., Huang, X., Orwig, K. E., and Sheng, Y. (2013). Rapid assembly of customized TALENs into multiple delivery systems. *PLoS ONE* 8:e80281. doi: 10.1371/journal.pone.0080281

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H., and Reif, J. C. (2014). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112, 638–645. doi: 10.1038/hdy.2014.1

Zhao, Y. S., Mette, M. F., and Reif, J. C. (2015). Genomic selection in hybrid breeding. *Plant Breed.* 134, 1–10. doi: 10.1111/pbr.12231

Zheng, M., Zhang, L., Tang, M., Liu, J., Liu, H., Yang, H., et al. (2020). Knockout of two *BnaMAX1* homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotech. J.* 18, 644–654. doi: 10.1111/pbi.13228

Zhou, J., Chen, T., Cheng, C., Xianhong, G., and Zaiyun, L. (2016). Distinct subgenome stabilities in synthesized *Brassica* allohexaploids. *Theor. Appl. Genet.* 129, 1257–1271. doi: 10.1007/s00122-016-2701-7

Zou, J., Mao, L., Qiu, J., Wang, M., Jia, L., Wu, D., et al. (2019). Genome-wide selection footprints and deleterious variations in young Asian allotetraploid rapeseed. *Plant Biotechnol. J.* 17, 1998–2010. doi: 10.1111/pbi.13115

Zou, J., Y., Zhao, P., Liu, L., Shi, X., Wang, M., et al. (2016). Seed quality traits can be predicted with high accuracy in *Brassica napus* using genomic data. *PLoS ONE* 11:e0166624. doi: 10.1371/journal.pone.0166624

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mohd Saad, Severn-Ellis, Pradhan, Edwards and Batley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fgene.2021.615284

Comparative Analyses of Full-Length Transcriptomes Reveal *Gnetum luofuense* Stem Developmental Dynamics

*Chen Hou[1,2], Huiming Lian[1,2], Yanling Cai[1,2], Yingli Wang[1,2], Dongcheng Liang[1,2] and Boxiang He[1,2]**

[1] Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou, China, [2] Guangdong Academy of Forestry, Guangzhou, China

Genus *Gnetum*, of which the majority species are pantropical liana, have broad industrial uses including for string, nets, and paper production. Although numerous studies have investigated anatomical structures during stem development, the underlying molecular mechanisms that regulate this developmental trajectory in *Gnetum* species remain poorly understood. A total of 12 full-length transcriptomes were generated from four stem developmental stages of an arborescent representative of this genus, *Gnetum luofuense*, using Oxford Nanopore Technologies. The results of this analysis reveal a total of 24,151 alternative splicing (AS) and 134,391 alternative polyadenylation events. A remarkably dynamic pattern of AS events, especially in the case of intron retentions, was found across the four developmental stages while no dynamic pattern was found among transcript numbers with varied poly(A) sites. A total of 728 long non-coding RNAs were also detected; the number of *cis*-regulated target genes dramatically increased while no changes were found among *trans*-regulated target genes. In addition, a K-means clustering analysis of all full-length transcripts revealed that primary growth is associated with carbohydrate metabolism and fungi defense, while secondary growth is closely linked with photosynthesis, nitrogen transportation, and leaf ontogenesis. The use of weighted gene co-expression network analysis as well as differentially expressed transcripts reveals that bHLH, GRF, and MYB-related transcription factors are involved in primary growth, while AP2/ERF, MYB, NAC, PLAZ, and bZIP participate in *G. luofuense* stem secondary growth. The results of this study provide further evidence that Nanopore sequencing technology provides a cost-effective method for generating full-length transcriptome data as well as for investigating seed plant organ development.

Keywords: stem development, full-length transcripts, transcription factor, differentially expressed transcripts, gymnosperm

OPEN ACCESS

Edited by:
Pinky Agarwal, National Institute of Plant Genome Research (NIPGR), India

Reviewed by:
Douglas S. Domingues, São Paulo State University, Brazil
Dhruv Lavania, University of Alberta, Canada

***Correspondence:**
Boxiang He
heboxiang@163.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *08 October 2020*
Accepted: *01 February 2021*
Published: *25 March 2021*

Citation:
Hou C, Lian H, Cai Y, Wang Y, Liang D and He B (2021) Comparative Analyses of Full-Length Transcriptomes Reveal Gnetum luofuense Stem Developmental Dynamics. Front. Genet. 12:615284. doi: 10.3389/fgene.2021.615284

INTRODUCTION

Gnetum, within the order Gnetales, is a gymnosperm genus distributed in tropical and subtropical areas globally (Markgraf, 1930, 1951, Kubitzki, 1990). This genus comprises about 40 species (Price, 1996; Hou et al., 2015); most of these are lianoids with just two, *Gnetum gnemon* L. and *Gnetum costatum* K. Schum, reported to be arborescent (Markgraf, 1951; Hou et al., 2015). *Gnetum*

vegetative organs are angiosperm like and species have broad pinnate leaves and swollen nodes at stem connections (Ickert-Bond and Renner, 2015, Hou et al., 2016). Despite the fact that *Gnetum* possesses eudicot-like leaves with paralleled veins, this genus has been characterized by rather low photosynthetic capacity (Feild and Balun, 2008; Deng et al., 2019a). The members of this genus are important economic crops in both Africa and southeast Asia (Markgraf, 1951; Ali et al., 2011, Ingram et al., 2012). Chemical extracts from *Gnetum* leaves and seeds (e.g., stilbenoids and flavonoids) are also of important medicinal value (Ali et al., 2011; Deng et al., 2016, Deng et al., 2017), while stem and bark samples are made into string, nets, and paper (Fu et al., 1999; Isong et al., 1999, Ndam et al., 2001).

Stem anatomical structures and development in this genus have been intensively studied encompassing the entire genus and lineages of species (Carlquist and Robinson, 1995, Carlquist, 1996a,b,c, 2012). It is notable that *Gnetum* stem development was documented in a monograph (Maheshwari and Vasil, 1961); this work notes that the transverse sections of young stem samples are more or less rounded (**Figure 1A**) and that the epidermis in each cases comprises rectangular cells. Four to six cell layers are present beneath this surface that comprises rounded or polygonal units containing chloroplasts. The cortex comprises between 12 and 16 layers comprising thin-walled parenchyma cells. A large number of scattered fibers are then present beneath this surface characterized by a narrow lumen. In slightly older stems, a pronounced sclerenchyma zone rises up from an irregular ring of lignified parenchyma cells within the inner part of the cortex, while collateral and endarch vascular bundles are arranged in a ring shape. These vascular bundles are segregated by medullary rays which comprise mostly tracheids and a few vessels. The phloem comprises sieve and parenchyma cells; in later developmental stages, the cell wall of pith cells becomes lignified and numerous pits emerge.

Although numerous studies have addressed *Gnetum* stem development and anatomical structures, the molecular mechanisms which drive this transition remain poorly understood. Earlier work has emphasized transcription factors (TFs) and has revealed that gene expression of about 70% expressed TFs are upregulated from primary to secondary growth (Dharmawardhana et al., 2010). Another recent study showed that some TFs, e.g., ARF (AUX RESPONSE FACTOR), AP2/ERF, and GRF (GROWTH REGULATION FACTOR), are principally expressed in primary growth, while others, e.g., bZIP (BASIC REGION/LEUCINE ZIPPER), NAC, and PLATS, accumulate during secondary growth in *Populus* (Chao et al., 2019). Among these, AP2/ERF-like TFs have been reported to regulate the expansion and proliferation of *Arabidopsis* cells and shoot architecture (Marsch-Martinez et al., 2006; Mehrnia et al., 2013), while GRFs also determine the formation of shoot meristem and architecture in this genus (Zhang et al., 2018). Similarly, bZIPs are involved in the transition from primary to secondary growth in *Eucalyptus* (Paux et al., 2004) and *Populus* (Dharmawardhana et al., 2010), while the NAC TF has been reported to regulate secondary cell wall development in *Arabidopsis* (Zhong et al., 2006; Hussey et al., 2011). TFs involved in carbohydrate metabolism and plant defense, and protective functions have also been reported during *Populus* primary growth (Dharmawardhana et al., 2010).

In addition to differentially expressed TFs, a growing body of work has also revealed that alternative splicing (AS) and alternative polyadenylation (APA) enrich transcriptome complexity during angiosperm development (e.g., Abdel-Ghany et al., 2016; Liu et al., 2017, Wang et al., 2017; Chao et al., 2018, Chao et al., 2019; Hu et al., 2020) as well as in other gymnosperms (Ye et al., 2019). AS has been proposed as an essential developmental modulator in plants (Wang and Brendel, 2006; Yang et al., 2020). APA promotes the stability, translation, and localization of target RNAs by generating different coding sequences or 3′ UTRs (Elkon et al., 2013). Previous studies on *Gnetum* have shown that various types of AS and APA are involved in *G. luofuense* female strobili and leaf development, but AS and APA events during stem development have never been tested. Previous studies have also shown that *Ginkgo biloba* L. leaves (Wu et al., 2019; Ye et al., 2019) as well as the leaves and female strobili of *G. luofuense* (Hou et al., 2019a) express long non-coding RNAs (lncRNAs) that have *cis*- and *trans*- target genes. LncRNAs are commonly seen in eukaryotic organisms as they are involved in transcriptional and post-transcriptional gene regulation (Mercer et al., 2009; Liu et al., 2015, Karlik et al., 2019). Although full-length transcriptomes have been generated based on merged samples at developmental stages in *Gnetum* (Hou et al., 2019a), dynamic AS and APA event patterns as well as target genes regulated by lncRNAs have been constantly neglected in previous studies.

Oxford Nanopore Technologies (ONT), a new third-generation sequencing technology, is a cost-effective and powerful method. One recent study has shown that this technology generates better quality raw data and estimates transcript levels more accurately than PacBio technology (Cui et al., 2020). ONT has been broadly applied in whole genome sequencing but has rarely been used in full-length transcriptome sequencing. We therefore generated 12 full-length transcriptomes from four developmental stages of *G. luofuense* stems using Nanopore sequencing technology. We first surveyed AS and APA events and identified lncRNAs as well as their dynamic patterns across *G. luofuense* stem development. We then determined deferentially expressed transcripts (DETs) and TFs based on expression levels of each transcript which were then investigated and further validated by qRT-PCR. Our investigation of the molecular mechanisms that regulate vegetative and reproductive organ development provide new insights on *Gnetum* industrial and medical resource exploration.

MATERIALS AND METHODS

Plant Sampling and Anatomic Analysis

Stem material for transcriptomic sequencing was collected from one female and one male individual of *G. luofuense* (voucher nos. CH003 and CH004, respectively, they have been deposited in SYS, Guangzhou, China). Samples were collected from the Bamboo Garden at Sun Yat-sen University, Guangzhou, China,

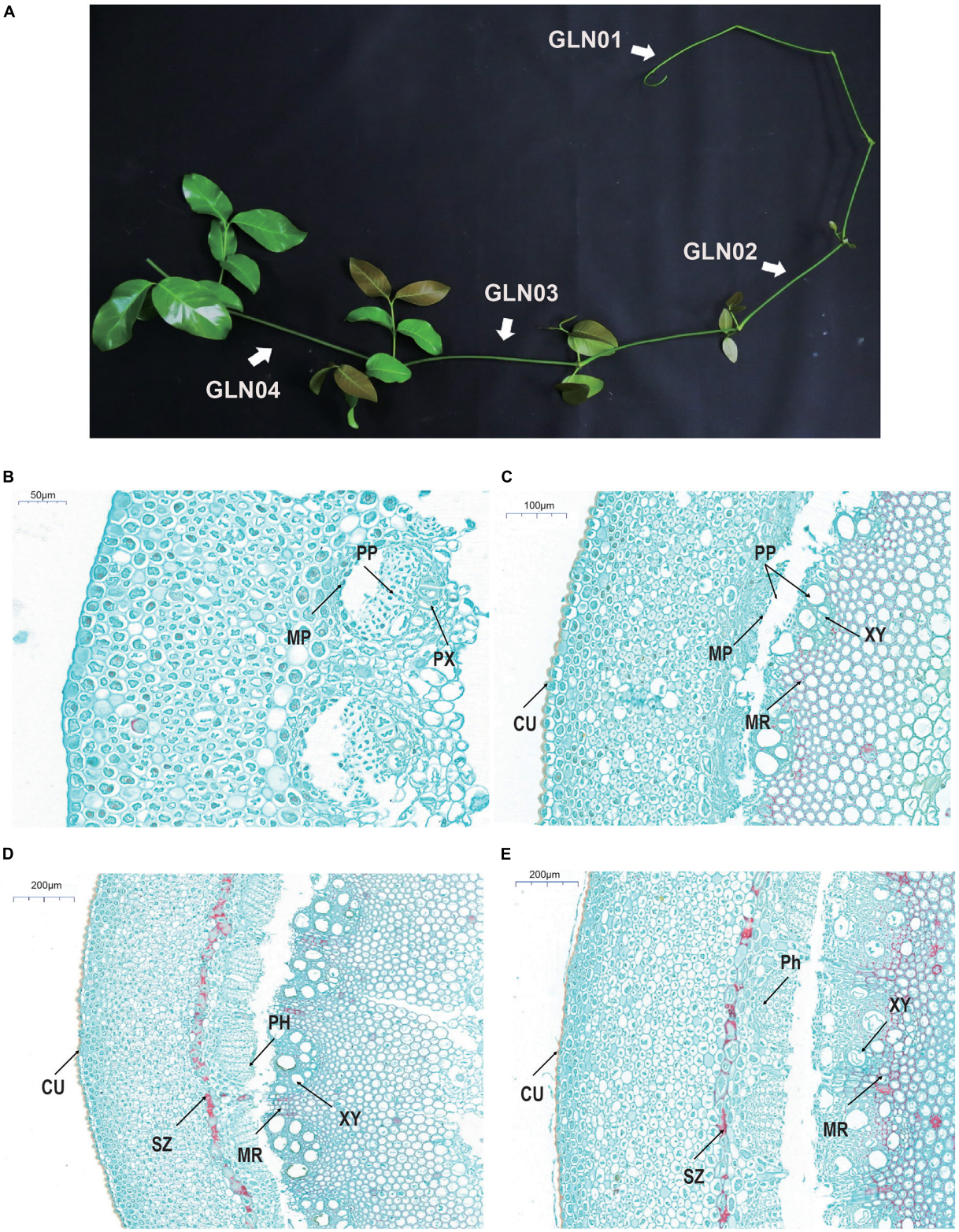


FIGURE 1 | Gross morphology and anatomical structures in *G. luofuense* stems. **(A)** A photograph shows four developmental stages (i.e., GLN01–04) of *G. luofuense* stems identified. **(B–E)** Anatomical structures and histology of *G. luofuense* stems at the developmental stages of GLN01 **(B)**, GLN02 **(C)**, GLN03 **(D)**, and GLN04 **(E)**. Abbreviations: CU, cuticle; MP, metaphloem; MR, medullary ray; PH, phloem; PP, protophloem; PX, protoxylem; SZ, sclerenchymatous zone; XY, xylem.

on October 2, 2019. Plant sampling has been permitted by College of Life Science, Sun Yat-sen University. Four developmental stages, GLN01 (stem apex, diameter = 0.5 to 2 mm), GLN02 (2 to 3 mm), GLN03 (3 to 4 mm), and GLN04 (4 to 5 mm), were identified in this study (**Figure 1**). Three replicate samples from each developmental stage (two from the female individual and one from the male individual) were prepared, resulting in a total of 12 sequencing samples. Identical stem samples were also collected for qRT-PCR analysis.

Gnetum luofuense stem samples were dissected into pieces using a knife blade and incubated in formaldehyde–acetic acid–alcohol (FAA) for 96 h. Softened stem material was then subsequently placed into 70% ethanol and further dehydrated in an ethanol series at concentrations of 70%, 96%, and 100%. Subsequent to dehydration, plant material was embedded in paraffin wax and sectioned into thicknesses between 3 and 4 μm using a rotary microtome (Leica, YGQ-3126F, Germany). Sectioned stem material was then extended in water at 40°C and then adhered to a glass slide at 60°C for 15 min. Glass slides were then stained using a safranin O-fast green staining method (Getzy et al., 1982) and were deposited at the Guangdong Academy of Forestry, Guangzhou, China. Observations were then made using light microscopy (Nikon BD-YG3001).

RNA Extraction and Nanopore Sequencing

All 12 stem samples were incubated in liquid nitrogen and stored at −20°C. Total RNA was then extracted using an RNA kit (Qiagen, Valencia, CA, United States) while relic DNA was removed using RNase-free DNase (Qiagen). We then used 1% agarose gel electrophoresis, a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, United States), and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States) to assess the concentration, purity, and integrity of extracted RNA. Furthermore, before full-length transcriptome sequencing, 1 μg total RNA was prepared for cDNA libraries using cDNA-PCR Sequencing Kit (SQK-PCS109) protocol provided by ONT. Libraries were then created using a sequencing library preparation kit. We added defined PCR adapters directly to both ends of the first-strand cDNA. The establishment of cDNA libraries was subject to 14 circles of PCR amplification with LongAmp Tag (NEB). The PCR products were then subjected to ONT adaptor ligation using T4 DNA ligase (NEB). Agencourt XP beads was used for DNA purification according to ONT protocol. The final cDNA libraries were added to FLO-MIN109 flowcells and libraries were then sequenced using a MinION Mk1B sequencer.

Raw Read Processing and Genome Mapping

Raw data generated from Nanopore sequencing were analyzed using MinKNOW version 2.2 (Oxford, United Kingdom) with a read quality score ≥ 7 and a read length ≥ 500 bp. The Silva rRNA database[1] was then searched to delete ribosomal RNA from sequenced data. Full-length transcripts (FLs) were identified via the presence of primers at both ends of cleaned reads. The clustering of detected FLs was then detected after mapping to the *G. luofuense* reference genome (genome data have been deposited in https://datadryad.org/stash/dataset/doi:10.5061/dryad.0vm37) using minimap2 (Li, 2018). A consensus isoform was obtained and polished using pinfish[2]. Similarly, consensus FLs were mapped to the *G. luofuense* reference genome using minimap2. To remove redundant FLs, all mapped transcripts were collapsed using the cDNA Cupcake package with a minimum coverage of 85% and a minimum identity of 90%. Consensus FLs with sequence differences at 5′ ends were not considered to be redundant isoforms.

Functional Annotation and Classification

Non-redundant FLs were annotated using BLASTX v.2.2.26 searches (E-value $< 1 \times 10^{-5}$) versus gene ontology (GO[3]), Kyoto Encyclopedia of Genes and Genomes (KEGG[4]), Protein Family (Pfam), Clusters of Orthologous Groups of Proteins (KOG/COG[5]), NCBI non-redundant protein sequence (NR[6]), and Swiss-Prot[7] databases as well as on the basis of HMMER v.3.1b2 searches (E-value $< 1 \times 10^{-10}$) based on Pfam[8] databases (Finn et al., 2011; Albert et al., 2013). GO enrichment analysis was performed using the GOseq software package implemented in R (R Core Team, 2018) while KEGG enrichment analysis was performed using the KEGG Orthology Based Annotation System (KOBAS, Xie et al., 2011). The annotation of novel isoforms was also completed following these protocols.

Alternative Splicing and Alternative Polyadenylation Analyses

Subsequent to mapping versus the *G. luofuense* reference genome, AS events were identified using the Astalavista approach (Foissac and Sammeth, 2007). Five types of AS were identified on this basis including an alternative 3′ splice site and an alternative 5′ splice site, as well as exon skipping, intron retention, and mutually exclusive exons. APA events were also analyzed using the TAPIS pipeline (Abdel-Ghany et al., 2016), while FL motifs were detected in the sequence of 50 nucleotides upstream of poly(A) sites using the software MEME (Bailey et al., 2006).

Identification of Coding Sequences and Long Non-coding RNAs

We identified CDS of polished non-redundant isoforms using the software TransDecoder (Haas et al., 2013) based on log-likelihood score, open reading frame lengths, and comparisons of amino acids versus the Pfam database. Prediction of lncRNAs was performed using four methods, the Coding Potential Calculator (CPC) (Kong et al., 2007), the Coding-Non-Coding Index

[1]https://www.arb-silva.de

[2]https://github.com/nanoporetech/pinfish

[3]http://www.geneontology.org

[4]http://www.genome.jp/kegg/

[5]http://www.ncbi.nlm.nih.gov/COG/

[6]http://www.ncbi.nlm.nih.gov/

[7]http://www.expasy.org/sprot/

[8]http://pfam.xfam.org/

(CNCI) (Sun et al., 2013), the Coding Potential Assessment Tool (CPAT) (Wang et al., 2013), and Pfam. Subsequent to filtering protein-coding reads, lncRNAs were identified as reads that possessed at least 200 nt and two exons. Identified lncRNAs were then classified as either lincRNA, antisense-lncRNA, sense-lncRNA, or intronic-lncRNA. Target genes that are regulated by identified lncRNAs were predicted using the software LncTar (Li et al., 2015); thus, two types of *cis*- or *trans*- target genes were defined as regulated by lncRNAs in this analysis (Kung et al., 2013; Yang et al., 2014).

Transcription Factor Identification and Weighted Correlation Network Analysis

We identified TFs and their transcripts which were then assigned to different families using the iTAK software (Zheng et al., 2016). A weighted correlation network analysis (WGCNA) was then conducted using the R package WGCNA version 1.42 (Langfelder and Horvath, 2008). Co-expression networks of TFs across four developmental stages were then constructed with settings defined as CPM values ≥ 1, fold change > 1, minimum module size = 30, and ntop value = 150. A hierarchal clustering tree was also constructed using the Dynamic Tree Cut package in R (Langfelder et al., 2007).

Differentially Expressed Transcripts and K-Means Clustering Analysis

We mapped FLs onto the *G. luofuense* reference genome and saved matched transcripts with coverage above five. Transcript expression was quantified as counts per million (CPM); thus, CPM = reads mapped to transcripts divided by total reads in one sample $\times$ 1,000,000. Differential expression analysis of each pair of developmental stages was then performed using the software DESeq version 1.10.1 (Anders, 2010). Resultant p values were then adjusted using the Benjamini and Hochberg approach (Benjamini and Hochberg, 1995) to produce false discovery rates (FDRs). Transcripts with FDR-adjusted p values < 0.01 and fold changes ≥ 2 were defined as differentially expressed transcripts (DETs). Out of these DETs, transcript expression of three replicates at each developmental stage was averaged, and denoted GLN01, GLN02, GLN03, and GLN04. Thus, based on the averaged transcript expression, all DETs were subjected to K-means clustering analyses. At the same time, we also performed GO and KEGG enrichment analyses with q-values < 0.05.

Validation of Transcript Expression

We selected nine DETs for transcript expression validation using qRT-PCR. Thus, primers were designed using the software Primer Premier 5 (Lalitha, 2000), with information about the qRT-PCR protocol listed in **Supplementary Table 1**. Two micrograms of RNA was extracted from each *G. luofuense* stem developmental stage. The amplification program used here comprised 2 min of initial denaturation at 95°C, followed by 40 cycles of 20 s at 94°C, 20 s at 58°C, and 20 s at 72°C. The series ended with an extension step at 72°C for 5 min. Gene expression was normalized following the *ACTIN* reference (Hou et al., 2019a). Three replicates were performed for each sample with the mean and SD of qRT-PCR gene expression values calculated in each case. The relative expression of target genes was estimated using the $\Delta\Delta$Ct method (Livak and Schmittgen, 2001) with expression of actin gene as the reference. For each sample of qRT-PCR analysis, we applied three replicates to calculate the mean and SD of gene expression.

RESULTS

Anatomic Analysis of *G. luofuense* Stems

Cross-sections of *G. luofusense* stems at each of the four developmental stages considered here were more or less circular. Observations show that at the GLN01 stage, the epidermis is composed of polygonal and rectangular cells (**Figure 1B**). Indeed, beneath the epidermis, between 12 and 16 layers of thin-walled parenchymatous cells are present, while below them, multiple circular portions also occur that are composed of metapholoems, protophloem, and protoxylem. We know that a pronounced cuticle encloses the stem epidermis at the GLN02 stage (**Figure 1C**); it is also the case that relatively well-developed protophloem and xylems were seen, separated further by wedge-shaped bundles of medullary rays with lignified cell walls. Anticlinal and periclinal divisions of the metacambium and cambium also significantly enlarged the girth of stems between the GLN02 and GLN04 stages. The phloem was well developed at the GLN03 stage and the numbers of vessels within xylem samples also increased remarkably. A pronounced sclerenchymatous zone derived from an irregular ring of parenchyma cells was also observed in the inner part of the cortex (**Figure 1D**). We fastened a few cells within the pith and medullary rays with safranin red; this procedure showed that red-stained cells within the last developmental stage, GLN04, became dense and mostly lignified in both medullary stem rays and piths. Four developmental stages were therefore selected to represent the transition from primary to secondary growth on the basis of these anatomic structures.

Raw Data Processing and Genome Mapping

A total of 48,156,451 clean reads were generated using Nanopore sequencing (**Supplementary Figure 1A** and **Supplementary Table 2**); these had a mean length that ranged between 1,209 (GLN023) and 1,345 (GLN012). Among clean reads, a total of 38,635,244 FLs with clear primer sequences at both ends were identified, accounting for between 76.35% (2,541,173; GLN011) and 81.82% (3,839,899; GLN021) of the total, depending on the sample (**Supplementary Figure 1B** and **Supplementary Table 3**). Indeed, among FLs, all samples had a high proportion of mapped transcripts, ranging between 99.43% (3,399,666; GLN023) and 99.74% (3,733,604; GLN013) (**Supplementary Table 4**). A total of 494,859 consensus FLs were determined subsequent to clustering and polishing, ranging between 34,843 (GLN011) and 49,929 (GLN031) (**Supplementary Figure 1C** and **Supplementary Table 5**). A total of 261,306 non-redundant FLs were then finally obtained after remapping and deletion of redundant examples; these had mean lengths ranging between

1,398 bp (GLN023) and 1,551 bp (GLN013) (**Supplementary Figure 1D** and **Supplementary Table 6**).

Functional Annotation and the Detection of Novel Transcripts

A total of 60,082 transcripts were annotated functionally in this analysis by searching against the GO, KEGG, COG, KOG, Pfam, NR, and Swiss-Prot databases (**Supplementary Figure 2A** and **Supplementary Table 7**). Among annotated transcripts, a total of 59,192 transcripts exhibited significant hits versus the NR database; the largest number of hits occurred in *Picea sitchensis* (16,154 genes; 27.29%), *Amborella trichopoda* (3,588 genes; 6.06%), and *Nelumbo nucifera* (2,046 genes; 3.46%) (**Supplementary Figure 2B**). GO analysis revealed that 33,808 transcripts could be classified into three categories: "biological processes," "cellular components," and "molecular functions" (**Supplementary Figure 2C**). The most abundant biological process GO terms were classified as "metabolic processes" (15,768 FLs), "cellular processes" (14,694), and "single-organism processes" (9,404), while the most abundant cellular component GO terms were classified as "cell parts" (15,373), "cell" (15,279), and "organelle" (11,006). Similarly, the most abundant molecular function GO terms were classified as "catalytic activity" (11,271), "binding" (13,958), and "transporter activity" (1.730), while the top three annotated FL KEGG terms were classified as "ribosome" (1,118, ko03010), "spliceosome" (961, ko03040), and "carbon metabolism" (881, ko01200). A total of 41,398 novel transcripts were also identified in this analysis; 35,052 (84.67%) were annotated by searching against the seven databases mentioned (**Supplementary Figure 2A**). The number of annotated transcripts identified using the NR, GO, and KEGG databases are summarized in **Supplementary Figures 2B–D**.

AS and APA Analyses

The results of this analysis showed that a total of 24,151 AS events were detected in the 12 *G. luofuense* stem samples (**Figure 2A**). Intron retention contributed the largest proportion overall (7.793 events; 33.23%), while a mutually exclusive exon added the smallest proportion (319 events; 1.32%). We then surveyed the numbers of various events throughout the four stem developmental stages (**Figure 2B**); results showed that intron retention numbers significantly increased between stage GLN02 and stage GLN04 (i.e., Student's *t*-test with a single tail distribution, p values < 0.05). Specifically, alternative 5′ splice site numbers significantly increased between GLN03 and GLN04 while alternative 3′ splice site and exon skipping numbers significantly increased at just GLN03 (p values < 0.05). No significant change in events was detected regarding the mutually exclusive exon (p values > 0.05).

APA analysis revealed that polyadenylation sites at the 3′ end of FLs scoring more than five encompassed the largest proportion (57,194; 42.56%), followed by those with just one (18,948; 14.02%) or two sites (16,189; 12.05%), respectively (**Figure 2C**). The numbers of APA events also did not differ significantly throughout *G. luofuense* stem development (i.e., Student's *t*-test p values were all > 0.05) (**Figure 2D**). Enrichment in uracil (U) was also detected at the 50 nt position upstream of the cleavage site in 3′ UTR while an enrichment in adenine (A) was found at the 50 nt position downstream; this indicates nucleotide bias across all poly(A) cleavage sites (**Figure 2E**). Three conserved motifs (i.e., AAAUGC, CCAUGC, and CCAUCC) were also detected at the 50 nt position upstream of the poly(A) site in all FLs (**Figure 2F**).

CDS and lncRNAs

We identified a total of 38,108 ORFs in our samples, including 30,323 (79.57%) that were complete and possessed both start and stop codons. In terms of complete ORFs, lengths between 100 and 200 bp (12,855 ORFs), 0 and 100 bp (12,150), and between 200 and 300 bp (3,751) made up the largest groups (**Figure 3A**). A total of 728 lncRNAs were also identified with mean lengths ranging between 298 and 4,362 nt; these were determined using four separate methods, CNCI, CPC, Pfam, and CPAT (**Figure 3B**). These lncRNAs were further classified as comprising 545 lincRNAs (74.86%), 28 antisense lncRNAs (3.85%), 13 intronic lncRNAs (1.79%), and 142 sense lncRNAs (19.50%) (**Figure 3C**). We also found that the *cis*-target genes regulated by these lncRNAs were considerably larger than those that were *trans*-regulated (**Figure 3D**). We found significantly increased numbers of target genes *cis*-regulated by these lncRNAs between GLN01 and GLN03 (Student's *t*-test, $p = 0.002$), and between GLN01 and GLN04 ($p = 0.037$). By contrast, no significant changes (all p values > 0.05) were found throughout stem development with regard to *trans*-target gene numbers regulated by these lncRNAs.

TF and WGCNA Analyses

A total of 4,251 TFs belonging to 208 gene families were identified on the basis of non-redundant FLs. The most abundant of these were AP2/ERF (176), MYB-related (135), and bHLH (133) (**Figure 4A**), while eight modules of highly correlated TFs through *G. luofuense* stem development were also identified using the WGCNA software (**Figures 4B,C**). Turquoise (No. 1, 315 TFs), black (No. 2, 215 TFs), and yellow (No. 7, 75 TFs) were the most abundant of these TF modules; TFs in the turquoise module were largely expressed in GLN01 while those in black and yellow modules were densely expressed in GLN02 and GLN04, respectively (**Figures 4B,C**). It is clear that TFs in the turquoise module were mainly enriched in three KEGG pathways, "starch and sucrose metabolism" (11 TFs, ko00500), "starch and sucrose metabolism" (eight TFs, ko00500), and "pentose and glucoronate interconversions" (six TFs, ko00040) (**Figure 4D**). Similarly, TFs in the black module were primarily enriched in "plant–pathogen interactions" (six TFs, ko04626), "carbon metabolism" (four), and "phenylpropanoid biosynthesis" (four TFs, ko00940), while TFs in the yellow module were mainly enriched in "cyanoamino acid metabolism" (two, ko00460) and "glycolysis/gluconeogenesis" (two TFs, ko00010). Moreover, we found, e.g., bHLH, GRF, and MYB-related were highly expressed in the turquoise module; e.g., AP2/ERF, NAC, and MYB were in the black module; and MYB, bZIP, and PLATZ were in the yellow module (**Figure 4E**).

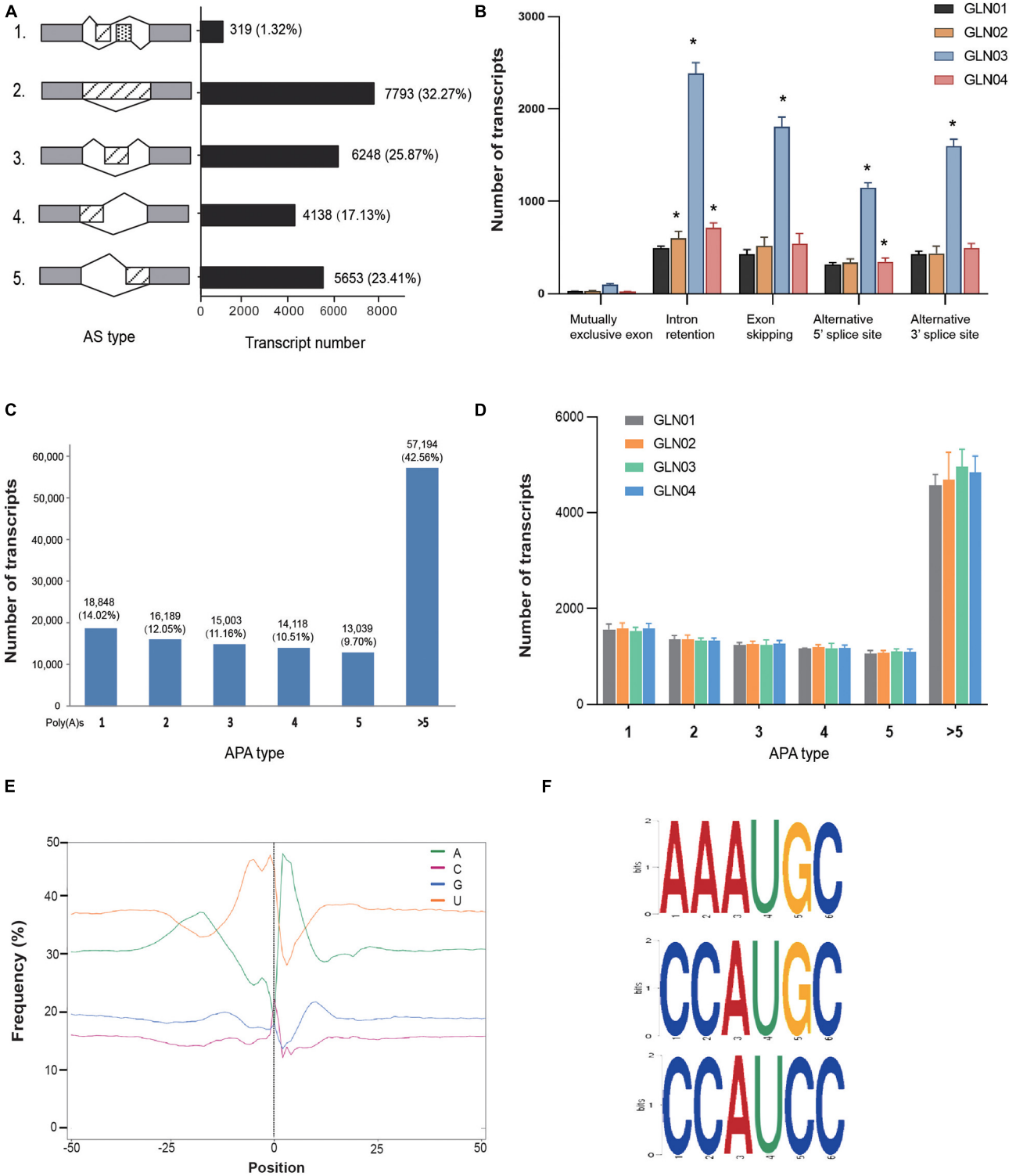


FIGURE 2 | Identification of AS and APA events. **(A)** Numbers of various AS events and their corresponding proportions based on 12 full-length transcripts. **(B)** AS events and corresponding transcript numbers across four developmental stages (GLN01–04) of *G. luofuense* stems. **(C)** Numbers of transcripts that possess various numbers of poly(A) sites based on 12 full-length transcripts. **(D)** APA events and corresponding transcript numbers across four developmental stages (GLN01–04) of *G. luofuense* stems. **(E)** The nucleotide composition around poly(A) cleavage sites, with the relative frequency of each nucleotide on the *y*-axis and the transcriptome composition across all poly(A) cleavage sites on the *x*-axis. **(F)** Three conserved motifs found in the 50 nt upstream region of the poly(A) cleavage site in all full-length transcripts. Symbol * indicates the p values < 0.05 using Student's t-test analysis.

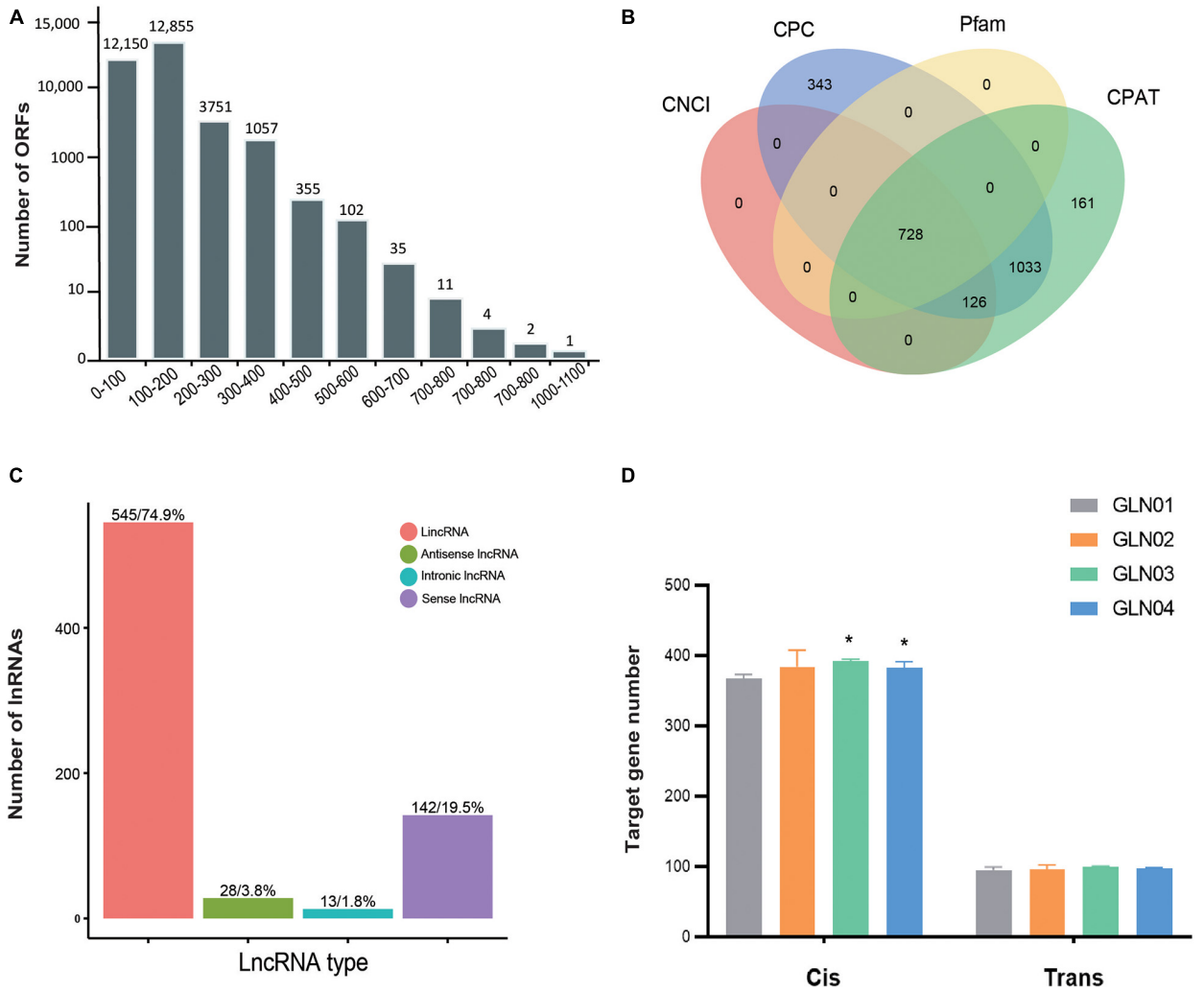


FIGURE 3 | Identification of ORFs and lncRNAs based on 12 full-length transcriptome. **(A)** Length distribution of ORFs detected in all full-length transcripts. **(B)** Venn diagram showing the number of lncRNAs identified using four different approaches: CPC, CNCI, CPAT, and Pfam. **(C)** Functional classification and numbers of four lncRNA types. **(D)** Target genes regulated by lncRNAs in *cis* and in *trans* across four developmental stages. Symbol * indicates the p values < 0.05 using Student's t-test analysis.

DETs and K-Means Clustering

Results showed that there were 492 DETs between GLN01 and GLN04 including 283 that were upregulated and 209 that were downregulated (**Figure 5A**). The second and third largest DET numbers then comprised 468 between GLN01 and GLN02 and 136 between GLN01 and GLN03. The numbers of DETs subsequently annotated by searching against GO, KEGG, and Swiss-Prot databases are presented in **Figure 5B**. Results revealed that the DET set between GLN01 and GLN04 was significantly enriched in the three KEGG pathways (i.e., q-value < 0.05) "phenylpropanoid biosynthesis" (16 DETs), "starch and sucrose metabolism" (10 DETs), and "cyanoamino acid metabolism" (eight DETs) (**Figure 5C**).

All the DETs identified in this analysis were then grouped into seven clusters using the K-means clustering algorithm (**Figure 5D**). Results show that transcripts from the K7 cluster were predominantly expressed at the apex (GLN01) while the remaining clusters, K3 and K4, were predominantly expressed between GLN02 and GLN04. Transcripts of these seven clusters were enriched in multiple GO terms (**Figure 5E**); in one example, K7 cluster transcripts were enriched in "cell wall biogenesis," "plant-type cell wall organization," and "cell wall" terms. In another example, transcripts from the K3 cluster were enriched in "regulation of shoot apical meristem development," "regulation of leaf development," and "response to endogenous stimulus" terms while transcripts from the K4 cluster were enriched in "nitrogen compound transport," "vasculature development," and "photosystem I" terms.

qRT-PCR Analysis

We performed qRT-PCR experiments to validate the nine genes of interest in the results of DET and K-means clustering analyses (**Figure 6**). Our results showed that genes *TnS000170537g04*, *TnS000176189g04*, and *TnS000991505g06*

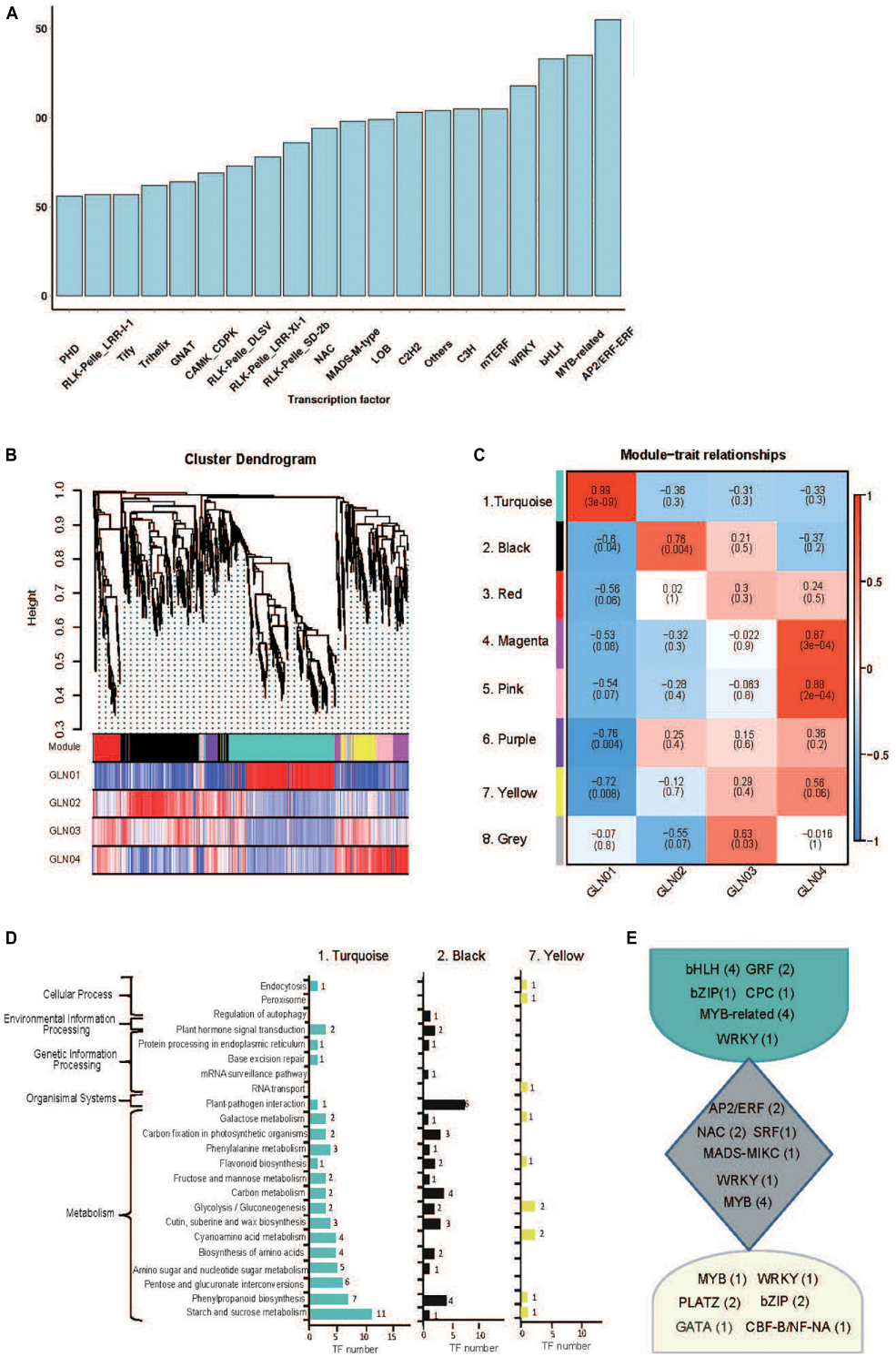


FIGURE 4 | Identification of TFs and results of WGCNA analysis. **(A)** A partial list of TFs (top 20 gene families) identified in 12 full-length transcriptome of *G. luofuense* stems. **(B)** Hierarchical clustering tree showing co-expression modules based on WGCNA analysis. Each branch in the phylogenetic tree corresponds to an individual TF, and highly interconnected TFs are grouped into eight modules. The different colored rows below the phylogeny indicate differentially expressed TFs in *G. luofuense*. **(C)** A heatmap of TF expression patterns in a matrix shows module–trait relationships. The expression patterns of eight modules are shown by the heatmap with a color bar of expression levels from high (red) to blue (red). The correlation coefficient was present for each cell, with the subtended by values of –log(*P*), where *P* is the Fisher's exact test *P* value. **(D)** KEGG pathway annotations of TFs in the largest three modules. **(E)** Highly expressed TFs and numbers in brackets in the largest three modules.

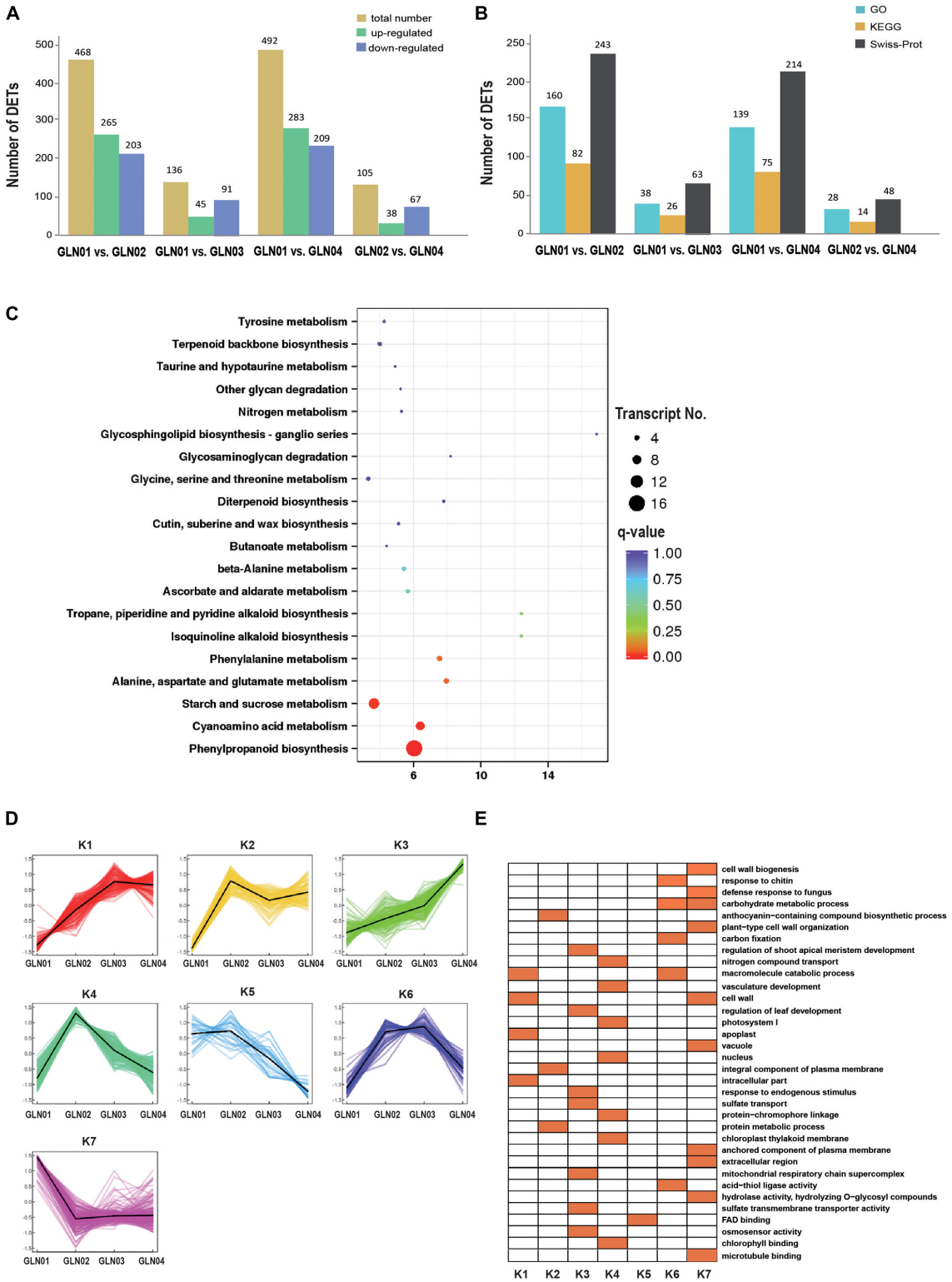


FIGURE 5 | Detected DETs and K-mean clustering analysis of DETs. **(A)** Numbers of DETs between different developmental stages of *G. luofuense* stems. **(B)** Numbers of DETs annotated by the three databases, i.e., GO and KEGG pathway annotations of TFs in the largest three modules, and Swiss-Prot. **(C)** An air bubble graph showing KEGG pathways enriched in the DEGs between GLN01 and GLN04. Gene numbers (circle size) and enrichment *q*-values (circle color) are shown. Warm colors represent *q*-values < 0.05. **(D)** Seven clusters identified based on expression levels of DETs from GLN01 to GLN04. **(E)** GO enrichment among the seven clusters with orange boxes denoting significant enrichment.

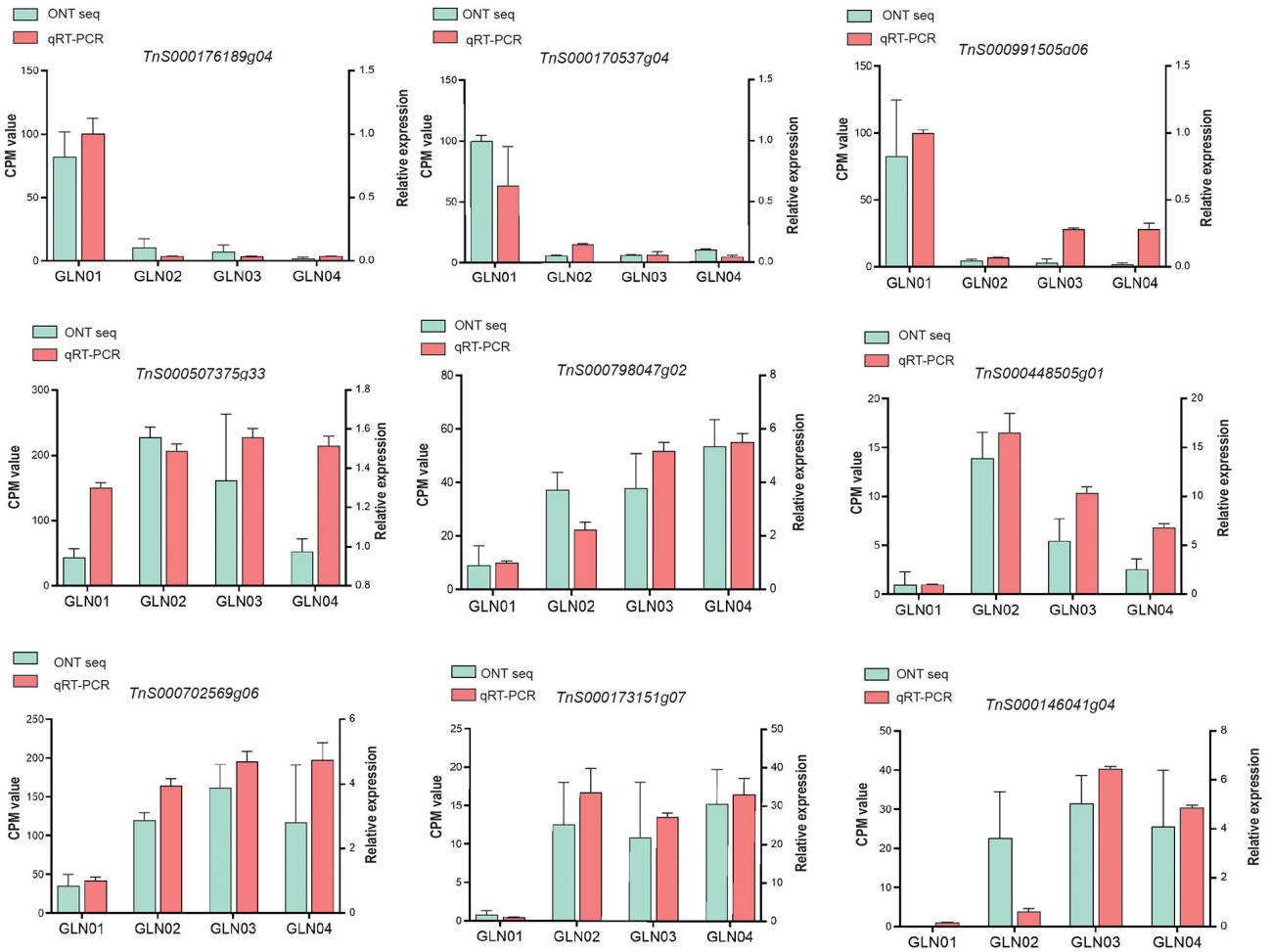


FIGURE 6 | qRT-PCR was used to verify the expression patterns of nine selected genes, using *ACTIN* as an internal control. Counts per million (CPM) values from Nanopore sequencing are indicated on the left *y*-axis and relative qRT-PCR expression levels are indicated on the right *y*-axis.

were predominantly expressed during the GLN01 stage while the remaining six were predominantly expressed between GLN02 and GLN04.

DISCUSSION

Stem Development Transcriptomic Dynamics

In terms of AS event analysis, results show that intron retention accounts for the largest proportion (32.23%) of the total number of events recorded in *G. luofuense* stem FLs (**Figure 2A**). This number turns out to be smaller than detected in the female strobili (46%) and the leaves (41.5%), respectively (Hou et al., 2019a), but larger than that in the seeds (29.9%) of *G. luofuense* (Deng et al., 2020). These AS mode proportions have also been reported in *Ginkgo biloba* L. (Ye et al., 2019) vegetative and reproductive organs as well as in maize and sorghum (Wang et al., 2018). Intron retention numbers also significantly increased throughout *G. luofuense* stem development (**Figure 2B**). It is noteworthy that the results reported here differ from those of a previous study in which the percentages of intron retention at the pre-fertilization stage decreased when compared with the post-fertilization stage (Li et al., 2017). We also found that exon skipping numbers as well as alternative 5′ splice and alternative 3′ splice sites at GLN03 all significantly increased compared with the GLN01 stage. These results suggest that AS events are likely to be essential in regulating *G. luofuense* stem development.

The results of APA analysis differ from those of previous studies where transcript numbers declined from one poly(A) site to five poly(A) sites (**Figure 2C**). At the same time, transcript numbers with more than five poly(A) sites make up a relatively small proportion in *G. luofuense* FLs within the leaves (3.12%), the female strobili (11.14%), and the seeds (3.62%) of *G. luofuense* (Deng et al., 2019b; Hou et al., 2019a, Deng et al., 2020). We further examined transcript numbers and found them consistent regardless of APA mode (**Figure 2D**) and were also able to show that the patterns of these events are more conserved compared with their AS counterparts. Three conserved motifs (i.e., AAAUGC, CCAUGC, and CCAUCC) were also identified

in the 12 *G. luofuense* stem FLs considered in this analysis. Results also differ with regard to the three motifs (i.e., AAAACA, GGGCGC, and CGCCGC) detected in the female strobili of *G. luofuense* (Hou et al., 2019a). Our results suggest that the vegetative and reproductive organs of *G. luofuense* experience variable APA mechanisms during organ development.

Our results show that the short ORFs (length < 200 bp) constituted the largest proportion (82.46%) of the detected ORFs (**Figure 3A**). It might be because short ORFs lead to high substitution rates as an adaptation to tropical habitats. A similar scenario has been reported among the chloroplast genes in Gnetales (Wang et al., 2015). In addition, it is noteworthy that the proportion of lincRNA in stems (74.9%) differs from the proportion seen in either leaves (33.49%) (Hou et al., 2019a) or female strobili (40.80%) (Hou et al., 2019a) in *G. luofuense* as well as in the leaves of *G. biloba* (50.6%) (Ye et al., 2019). Besides, because our results show the different numbers of *cis*- and *trans*-regulated genes across the entire procedure of stem development in *G. luofuense* (**Figure 3D**), we therefore suggest that *cis*-regulation of target genes by lncRNA is more important than *trans*-regulation across *G. luofuense* stem development. It is interesting that 44 lincRNAs in *Gossypium hirsutum* were shown to be differentially expressed under salt stress in a recent study; these lincRNAs are mostly *cis*-regulated (Deng et al., 2018). Additional work will be required to investigate the functions of these lincRNAs as well as the mechanisms by which they regulate genes in *G. luofuense* stem development.

Genes and TFs Related to Primary Growth

The results of this analysis also show that the DETs are enriched in starch and sucrose metabolism (**Figure 5C**), while those belonging to cluster K7 are enriched in "carbohydrate metabolic processes" GO terms (**Figure 5D**). TFs in the turquoise module were also annotated with "starch and sucrose metabolism" and "glycolysis/gluconeogenesis" (**Figure 4D**). These results indicate that the emergence of an apical apex in *G. luofuense* stems requires relatively abundant carbohydrates as a resource for further growth. Our results corroborate those of a previous study that showed the primary growth of *Populus* stems requires carbohydrate metabolism and plant defense (Dharmawardhana et al., 2010). Indeed, another previous study also reported that the regulation of carbohydrate metabolism facilitates cell division and expansion to promote shoot ontogenesis (Morris and Arthur, 1985; Fleming, 2006, Eveland and Jackson, 2012).

The results reported here also show that DETs can be annotated with the KEGG pathway "cyanoamino acid metabolism" (**Figure 5C**). Similarly, DETs in the K7 cluster were annotated with the GO term "defense response to fungus" (**Figure 5D**). These differences might be because apical stem tissues lack a physical barrier between them and so the cuticle is potentially vulnerable to external pathogen and fungi. This scenario is similar to the vulnerability seen in fertile and sterile *G. luofuense* and *G. gnemon* reproductive units during pollination periods (Hou et al., 2019b). In terms of detected TFs, the gene *TnS000176189g04* encodes a MYB-related TF and was shown to be principally expressed at GLN01 using qRT-PCR analysis (**Figure 6**). Another recent study also reported that the *MYB* gene is associated with fungal affection and host cell death induced by jasmonic acid (Lee et al., 2001). This means that MYB-related TFs might trigger a fungi or pathogen-related defense mechanism that aids in *G. luofuense* apical stem development.

Our results reveal that K7 DETs were enriched with the GO terms "cell-wall biogenesis," "plant-type cell-wall organization," "vacuole," and "cell wall" (**Figure 5E**). In addition, TFs were also enriched in "cutin, suberin, and wax biosynthesis" (**Figure 4B**). These results are congruent with the emergence of a cuticle on the stem epidermis as well as with the transition from protophloem and protoxylem to phloem and xylem during stem development (**Figure 1**). In terms of detected TFs, one *bHLH* gene (*TnS000170537g04*) and one *GRF* gene (*TnS000991505g06*) were principally expressed at the GLN01 stage and dramatically declined between GLN02 and GLN04 (**Figure 6**). The results of this analysis further corroborate a previous study which noted that bHLH TFs are involved in early xylem development in *Arabidopsis* roots under the various auxin conditions (Ohashi-Ito et al., 2013). A further example noted here also showed that cell division in root apical meristems is triggered by a bHLH complex via cytokinin action (Ohashi-Ito et al., 2014). It has also been reported that GRF-interacting factor1 determines the emergence of shoot meristem as well as maize architecture (Zhang et al., 2018). Thus, TF bHLH and GRFs are most likely to regulate *G. luofuense* stem apical shoot development.

Genes and TFs Related to Secondary Growth

The results of this analysis show that DETs are mainly enriched in the GO terms "regulation of shoot apical meristems," "response to endogenous stimuli," and "regulation of leaf development" in the K3 cluster (**Figure 5D**). Indeed, the DETs were principally expressed at the GLN02 stage and can be annotated with GO terms related to photosynthesis and nitrogen transportation. Our results suggest that stem growth in *G. luofuense* is accompanied by the strengthening of photosynthetic capability and nutrition transportation. The statement is congruent with a general conclusion that woody tissue photosynthesis facilitates bud and trunk development in young plants (Saveyn et al., 2010). It is noteworthy that *G. luofuense* leaves emerge from the stem stage during GLN02 and gradually develop between GLN03 and GLN04 (**Figure 1A**); *G. luofuense* therefore also probably resorts to the photosynthetic products provided by laterally branched leaves during stem development.

A number of TFs and genes might be involved in *G. luofuense* stem the secondary growth. The gene *TnS000507375g33* that encodes AP2/ERF was upregulated and expressed between GLN02 and GLN03 compared with GLN01 (**Figure 6**); this implies an important role in secondary growth. It has also been reported that a *AP2/ERF*-like TF participates the expansion and proliferation of *Arabidopsis* leaves (Marsch-Martinez et al., 2006), while additional previous work shows that another *AP2/ERF*-like TF is involved in cell proliferation and influences shoot architecture in *Arabidopsis* (Mehrnia et al., 2013). The gene

TnS000507375g33 might therefore trigger cell proliferation and anticlinal/periclinal expansion in *G. luofuense* stems between GLN02 and GLN04; the *PLAZ* gene *TnS000798047g02* and the *NAC* gene *TnS000448505g01* have been shown to be differentially expressed between GLN01 and GLN04 (**Figure 6**). It has also been reported that PLAZ and NAC TFs are likely to be important in *Populus* stem secondary growth (Chao et al., 2019), while other studies have shown that a NAC TF is able to negatively regulate *Arabidopsis* xylem fiber (Ko et al., 2007) and secondary cell wall development (Zhong et al., 2006; Hussey et al., 2011). We therefore assume that NAC TFs might facilitate the lignification of xylem fibers as well as the emergence of a sclerenchymatous zone between GLN03 and GLN04.

The results of this analysis reveal that the expression of one *bZIP* gene *TnS000702569g06* was significantly upregulated between GLN02 and GLN04 (**Figure 6**). This is consistent with the results of a previous study which reported that the morphology of tobacco plants is regulated by a ZIP transcriptional activator via the proportion of the endogenous hormone gibberellin (Fukazawa et al., 2000). Additional studies have also shown that bZIP plays an important role in the transition from primary to secondary growth in *Populus* (Dharmawardhana et al., 2010) and *Eucalyptus* (Paux et al., 2004). Expression levels of the two MYB genes *TnS000173151g07* and *TnS000146041g04* were also both dramatically upregulated between GLN02 and GLN04 (**Figure 6**), and another earlier study also showed that a hybrid aspen used differentially expressed MYB TFs to regulate the vasculature development of secondary stem growth (Karpinska et al., 2004). In a further example, one MYB TF PtrMYB152 has been shown to participate in the biosynthesis of secondary cell walls in *Arabidopsis* (Wang et al., 2014), while a further MYB TF GhMYB25 manipulates the development of fiber and trichome in this genus (Machado et al., 2009). We therefore assume that *bZIP* and *MYB* genes might be involved in the transition from primary to secondary growth by promoting xylem development as well as the emergence of a sclerechymatous zone.

CONCLUSION

We generated 12 full-length transcriptome of *G. luofuense* stems at four developmental stages using Nanopore sequencing technology. The 12 full-length transcriptome of *G. luofuense* stems predicts a total of 24,151 AS events and 134,391 APA events and 728 lncRNAs. WGCNA and K-means clustering analyses revealed that key transcription factors were associated with a series of KEGG pathways including photosynthesis, nitrogen transportation, and leaf ontogenesis. Transcription factors, e.g., bHLH, GRF, and MYB-related transcription factors, participate in primary growth, while others, e.g., AP2/ERF, MYB, NAC, PLAZ, and bZIP, are involved in *G. luofuense* stem secondary growth. These findings provide a valuable the information to the *Gnetum*-related fiber and paper industry, and shed light on the utility of Nanopore sequencing technology for the investigation of full-length transcriptomes in gymnosperms.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA647987/.

AUTHOR CONTRIBUTIONS

CH and BH conceived and designed the experiments. CH, HL, YC, and YW performed the experiments. CH, DL, and YC analyzed the data. CH wrote the manuscript. All authors have read and approved the manuscript.

FUNDING

This work was supported by a grant from the Natural Science Foundation of Guangdong Province, China (No. 2018A0303130163 to CH).

ACKNOWLEDGMENTS

We thank Guan-Shen Liu (Biomarker Technologies, Beijing, China) for providing us with valuable technical and analytical assistance. We acknowledge TopEdit LLC for linguistic editing and proofreading during the preparation of this article. We also thank two reviewers for providing constructive comments and valuable suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.615284/full#supplementary-material

Supplementary Figure 1 | Quality control and information of full-length reads using GLN011 as an example. **(A)** Length distribution of clean reads. **(B)** Length distribution of full-length (FL) reads. **(C)** Length distribution of consensus reads. **(D)** Length distribution of non-redundant FL reads.

Supplementary Figure 2 | All gene and novel gene annotation on the basis of seven databases. **(A)** Gene number annotated by the seven databases. **(B)** NR homologous species distribution diagram of all genes and novel genes. **(C)** Distribution of GO terms for all annotated genes and novel genes. **(D)** KEGG pathways of all genes and novel genes annotated in the full-length transcriptome of *G. luofuense*.

Supplementary Table 1 | Primer information and qRT-PCR systems.

Supplementary Table 2 | Clean reads generated by Nanopore sequencing.

Supplementary Table 3 | Numbers and percentages of FL reads.

Supplementary Table 4 | Mapped FL reads.

Supplementary Table 5 | Information of consensus reads.

Supplementary Table 6 | Statistics of non-redundant consensus transcripts.

Supplementary Table 7 | Annotation and classification of detected ORFs by the seven datasets.

REFERENCES

Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7:11706. doi: 10.1038/ncomms11706

Albert, V. A., Barbazuk, W. B., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., et al. (2013). The Amborella genome and the evolution of flowering plants. *Science* 342:1241089. doi: 10.1126/science.1241089

Ali, F., Assanta, M. A., and Robert, C. (2011). Gnetum africanum: a wild food plant from the african forest with many nutritional and medicinal properties. *J. Med. Food* 14, 1289–1297. doi: 10.1089/jmf.2010.0327

Anders, S. (2010). Analysing RNA-Seq data with the "DESeq" package. *Mol. Biol.* 43, 1–17.

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, 369–373. doi: 10.1093/nar/gkl198

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B (Methodol.)* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Carlquist, S. (1996a). Wood and bark anatomy of lianoid Indomalesian and Asiatic species of *Gnetum*. *Bot. J. Linn. Soc.* 121, 1–24. doi: 10.1111/j.1095-8339.1996.tb00742.x

Carlquist, S. (1996b). Wood, bark and stem anatomy of new world species of *Gnetum*. *Bot. J. Linn. Soc.* 120, 1–19. doi: 10.1111/j.1095-8339.1996.tb00476.x

Carlquist, S. (1996c). Wood, bark, and stem anatomy of Gnetales: a summary. *Int. J. Plant Sci.* 157, 58–76. doi: 10.1086/297404

Carlquist, S. (2012). Wood anatomy of Gnetales in a functional, ecological, and evolutionary context. *J. Syst. Evol. Bot.* 30, 33–47. doi: 10.5642/aliso.20123001.05

Carlquist, S., and Robinson, A. (1995). Wood and bark anatomy of the african species of Gnetum. *Bot. J. Linn. Soc.* 118, 123–137. doi: 10.1111/j.1095-8339.1995.tb00465.x

Chao, Q., Gao, Z. F., Zhang, D., Zhao, B. G., Dong, F. Q., Fu, C. X., et al. (2019). The developmental dynamics of the *Populus* stem transcriptome. *Plant Biotechnol. J.* 17, 206–219. doi: 10.1111/pbi.12958

Chao, Y., Yuan, J., Li, S., Jia, S., Han, L., and Xu, L. (2018). Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing. *BMC Plant Biol.* 18:300. doi: 10.1186/s12870-018-1534-8

Cui, J., Lu, Z., Xu, G., Wang, Y., and Jin, B. (2020). Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the *Arabidopsis* transcriptome. *Plant Methods* 16:85. doi: 10.1186/s13007-020-00629-x

Deng, F., Zhang, X., Wang, W., Yuan, R., and Shen, F. (2018). Identification of *Gossypium hirsutum* long non-coding RNAs (lncRNAs) under salt stress. *BMC Plant Biol.* 18:23. doi: 10.1186/s12870-018-1238-0

Deng, N., Chang, E., Li, M., Ji, J., Yao, X., Bartish, I. V., et al. (2016). Transcriptome characterization of *Gnetum parvifolium* reveals candidate genes involved in important secondary metabolic pathways of flavonoids and stilbenoids. *Front. Plant Sci.* 7:174. doi: 10.3389/fpls.2016.00174

Deng, N., Hou, C., He, B., Ma, F., Song, Q., Shi, S., et al. (2020). A full-length transcriptome and gene expression analysis reveal genes and molecular elements expressed during seed development in *Gnetum luofuense*. *BMC Plant Biol.* 20:531. doi: 10.1186/s12870-020-02729-1

Deng, N., Hou, C., Liu, C., Li, M., Bartish, I., Tian, Y., et al. (2019a). Significance of photosynthetic characters in the evolution of Asian *Gnetum* (Gnetales). *Front. Plant Sci.* 10:39. doi: 10.3389/fpls.2019.00039

Deng, N., Hou, C., Ma, F., Liu, C., and Tian, Y. (2019b). Single-molecule long-read sequencing reveals the diversity of full-length transcripts in leaves of *Gnetum* (Gnetales). *Int. J. Mol. Sci.* 20:6350. doi: 10.3390/ijms20246350

Deng, N., Liu, C., Chang, E., Ji, J., Yao, X., Yue, J., et al. (2017). High temperature and UV-C treatments affect stilbenoid accumulation and related gene expression levels in *Gnetum parvifolium*. *Electron. J. Biotechnol.* 25, 43–49. doi: 10.1016/j.ejbt.2016.11.001

Dharmawardhana, P., Brunner, A. M., and Strauss, S. H. (2010). Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. *BMC Genomics* 11:150. doi: 10.1186/1471-2164-11-150

Elkon, R., Ugalde, A. P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506. doi: 10.1038/nrg3482

Eveland, A. L., and Jackson, D. P. (2012). Sugars, signalling, and plant development. *J. Exp. Bot.* 63, 3367–3377. doi: 10.1093/jxb/err379

Feild, T. S., and Balun, L. (2008). Xylem hydraulic and photosynthetic function of *Gnetum* (Gnetales) species from Papua New Guinea. *New Phytol.* 177, 665–675. doi: 10.1111/j.1469-8137.2007.02306.x

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, 29–37. doi: 10.1093/nar/gkr367

Fleming, A. (2006). Metabolic aspects of organogenesis in the shoot apical meristem. *J. Exp. Bot.* 57, 1863–1870. doi: 10.1093/jxb/erj178

Foissac, S., and Sammeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* 35, 297–299. doi: 10.1093/nar/gkm311

Fu, L. K., Yu, Y. F., and Gilbert, M. G. (1999). "Gnetaceae," in *Fl China*, eds Z. Y. Wu and P. H. Raven (St. Louis: Missouri Botanical Garden Press), 102–105.

Fukazawa, J., Sakai, T., Ishida, S., Yamaguchi, I., Kamiya, Y., and Takahashi, Y. (2000). Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. *Plant Cell* 12, 901–915. doi: 10.1105/tpc.12.6.901

Getzy, L. L., Malemud, C. J., Goldberg, V. M., and Moskowitz, R. W. (1982). Factors influencing metachromatic staining in paraffin-embedded sections of rabbit and human articular cartilage: a comparison of the safranin O and toluidine blue O techniques. *J. Histotechnol.* 5, 111–116. doi: 10.1179/his.1982.5.3.111

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–1512. doi: 10.1038/nprot.2013.084

Hou, C., Deng, N., and Su, Y. (2019a). PacBio long-read sequencing reveals the transcriptomic complexity and Aux/IAA gene evolution in *Gnetum* (Gnetales). *Forests* 10:1043. doi: 10.3390/f10111043

Hou, C., Humphreys, A. M., Thureborn, O., and Rydin, C. (2015). New insights into the evolutionary history of *Gnetum* (Gnetales). *Taxon* 64, 239–253. doi: 10.12705/642.12

Hou, C., Saunders, R. M., Deng, N., Wan, T., and Su, Y. (2019b). Pollination drop proteome and reproductive organ transcriptome comparison in *Gnetum* reveals entomophilous adaptation. *Genes* 10:800. doi: 10.3390/genes10100800

Hou, C., Wikström, N., Strijk, J., and Rydin, C. (2016). Resolving phylogenetic relationships and species delimitations in closely related gymnosperms using high-throughput NGS, Sanger sequencing and morphology. *Plant Syst. Evol.* 302, 1345–1365. doi: 10.1007/s00606-016-1335-1

Hu, H., Yang, W., Zheng, Z., Niu, Z., Yang, Y., Wan, D., et al. (2020). Analysis of alternative splicing and alternative polyadenylation in *Populus alba* var. pyramidalis by single-molecular long-read sequencing. *Front. Genet.* 11:48. doi: 10.3389/fgene.2020.00048

Hussey, S. G., Mizrachi, E., Spokevicius, A. V., Bossinger, G., Berger, D. K., and Myburg, A. A. (2011). SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in *Arabidopsis* fibres and increases fibre cell area in Eucalyptus. *BMC Plant Biol.* 11:173. doi: 10.1186/1471-2229-11-173

Ickert-Bond, S. M., and Renner, S. S. (2015). The Gnetales: recent insights on their morphology, reproductive biology, chromosome numbers, biogeography, and divergence times. *J. Syst. Evol.* 54, 1–16. doi: 10.1111/jse.12190

Ingram, V., Ndumbe, L. N., and Ewane, M. E. (2012). Small scale, high value: *Gnetum africanum* and *buchholzianum* value chains in Cameroon. *Small Scale For.* 11, 539–556. doi: 10.1007/s11842-012-9200-8

Isong, E., Adewusi, S., Nkanga, E., Umoh, E., and Offiong, E. (1999). Nutritional and phytogeriatological studies of three varieties of *Gnetum africanum* ('afang'). *Food Chem.* 64, 489–493. doi: 10.1016/S0308-8146(98)00139-3

Karlik, E., Ari, S., and Gozukirmizi, N. (2019). LncRNAs: genetic and epigenetic effects in plants. *Biotechnol. Biotechnol. Equip.* 33, 429–439. doi: 10.1080/13102818.2019.1581085

Karpinska, B., Karlsson, M., Srivastava, M., Stenberg, A., Schrader, J., Sterky, F., et al. (2004). MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. *Plant Mol. Biol.* 56, 255–270. doi: 10.1007/s11103-004-3354-5

Ko, J. H., Yang, S. H., Park, A. H., Lerouxel, O., and Han, K. H. (2007). ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. *Plant J.* 50, 1035–1048. doi: 10.1111/j.1365-313X.2007.03109.x

Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and

support vector machine. *Nucleic Acids Res.* 35, 345–349. doi: 10.1093/nar/gkm391

Kubitzki, K. (1990). "Gnetaceae," in *The Families and Genera of Vascular Plants*, eds K. Kubitzki and C. Bayer (Berlin: Springer), 383–386. doi: 10.1007/978-3-662-02604-5_71

Kung, J. T., Colognori, D., and Lee, J. T. (2013). Long noncoding RNAs: past, present, and future. *Genetics* 193, 651–669. doi: 10.1534/genetics.112.146704

Lalitha, S. (2000). Primer premier 5. *Biotech. Softw. Internet Rep.* 1, 270–272. doi: 10.1089/152791600459894

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Zhang, B., and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics* 24, 719–720. doi: 10.1093/bioinformatics/btm563

Lee, M.-W., Qi, M., and Yang, Y. (2001). A novel jasmonic acid-inducible rice MYB gene associates with fungal infection and host cell death. *Mol. Plant Microbe Interact.* 14, 527–535. doi: 10.1094/MPMI.2001.14.4.527

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, J. W., Ma, W., Zeng, P., Wang, J. Y., Geng, B., Yang, J. C., et al. (2015). LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform.* 16, 806–812. doi: 10.1093/bib/bbu048

Li, Y., Dai, C., Hu, C., Liu, Z., and Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and illumina-based RNA-seq in strawberry. *Plant J.* 90, 164–176. doi: 10.1111/tpj.13462

Liu, J., Wang, H., and Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* 13, 319–328. doi: 10.1111/pbi.12336

Liu, X. X., Mei, W. B., Soltis, P. S., Soltis, D. E., and Barbazuk, W. B. (2017). Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol. Ecol. Resour.* 17, 1243–1256. doi: 10.1111/1755-0998.12670

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Machado, A., Wu, Y., Yang, Y., Llewellyn, D. J., and Dennis, E. S. (2009). The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J.* 59, 52–62. doi: 10.1111/j.1365-313X.2009.03847.x

Maheshwari, P., and Vasil, V. (1961). *Gnetum. Botanical Monograph No 1*. New Dehli: Council of Scientific & Industrial Research.

Markgraf, F. (1930). Monographie der gattung gnetum ser. 3. *Bull. Jar. Bot. Buitenz.* 10, 407–511.

Markgraf, F. (1951). "Gnetaceae," in *Flora Malesiana Ser 1*, Vol 4, ed. C. G. G. J. Steenis (Batavia: Noordhoff-Kolff), 336–347.

Marsch-Martinez, N., Greco, R., Becker, J. D., Dixit, S., Bergervoet, J. H., Karaba, A., et al. (2006). BOLITA, an *Arabidopsis* AP2/ERF-like transcription factor that affects cell expansion and proliferation/differentiation pathways. *Plant Mol. Biol.* 62:825–843. doi: 10.1007/s11103-006-9059-1

Mehrnia, M., Balazadeh, S., Zanor, M.-I., and Mueller-Roeber, B. (2013). EBE, an AP2/ERF transcription factor highly expressed in proliferating cells, affects shoot architecture in *Arabidopsis*. *Plant Physiol.* 162, 842–857. doi: 10.1104/pp.113.214049

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi: 10.1038/nrg2521

Morris, D., and Arthur, E. (1985). Invertase activity, carbohydrate metabolism and cell expansion in the stem of *Phaseolus vulgaris* L. *J. Exp. Bot.* 36, 623–633. doi: 10.1093/jxb/36.4.623

Ndam, N., Nkefor, J.-P., and Blackmore, P. (2001). Domestication of *Gnetum africanum* and G. buchholzianum (Gnetaceae), over-exploited wild forest vegetables of the central African region. *Syst. Geogr. Plants* 71, 739–745. doi: 10.2307/3668713

Ohashi-Ito, K., Matsukawa, M., and Fukuda, H. (2013). An atypical bHLH transcription factor regulates early xylem development downstream of auxin. *Plant Cell Physiol.* 54, 398–405. doi: 10.1093/pcp/pct013

Ohashi-Ito, K., Saegusa, M., Iwamoto, K., Oda, Y., Katayama, H., Kojima, M., et al. (2014). A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. *Curr. Biol.* 24, 2053–2058. doi: 10.1016/j.cub.2014.07.050

Paux, E., Tamasloukht, M. B., Ladouce, N., Sivadon, P., and Grima-Pettenati, J. (2004). Identification of genes preferentially expressed during wood formation in Eucalyptus. *Plant Mol. Biol.* 55, 263–280. doi: 10.1007/s11103-004-0621-4

Price, R. A. (1996). Systematics of the Gnetales: a review of morphological and molecular evidence. *Int. J. Plant Sci.* 157, 40–49. doi: 10.1086/297402

R Core Team (2018). *R: A Language and Environment for Statistical Computing Version 3.2.0*. Vienna: R Foundation for Statistical Computing.

Saveyn, A., Steppe, K., Ubierna, N., and Dawson, T. E. (2010). Woody tissue photosynthesis and its contribution to trunk growth and bud development in young plants. *Plant Cell Environ.* 33, 1949–1958. doi: 10.1111/j.1365-3040.2010.02197.x

Sun, L., Luo, H. T., Bu, D. C., Zhao, G. G., Yu, K. T., Zhang, C. H., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646

Wang, B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U. S. A.* 103, 7175–7180. doi: 10.1073/pnas.0602039103

Wang, B., Jiang, B., Zhou, Y., Su, Y., and Wang, T. (2015). Higher substitution rates and lower dN/dS for the plastid genes in Gnetales than other gymnosperms. *Biochem. Syst. Ecol.* 59, 278–287. doi: 10.1016/j.bse.2015.02.009

Wang, B., Regulski, M., Tseng, E., Olson, A., Goodwin, S., McCombie, W. R., et al. (2018). A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. *Genome Res.* 28, 921–932. doi: 10.1101/gr.227462.117

Wang, L., Park, H. J., Dasari, S., Wang, S. Q., Kocher, J. P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41:e74. doi: 10.1093/nar/gkt006

Wang, S., Li, E., Porth, I., Chen, J.-G., Mansfield, S. D., and Douglas, C. J. (2014). Regulation of secondary cell wall biosynthesis by poplar R2R3 MYB transcription factor PtrMYB152 in *Arabidopsis*. *Sci. Rep.* 4:5054. doi: 10.1038/srep05054

Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., et al. (2017). Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *Plant J.* 91, 684–699. doi: 10.1111/tpj.13597

Wu, Y., Guo, J., Wang, T., Cao, F., and Wang, G. (2019). Transcriptional profiling of long noncoding RNAs associated with leaf-color mutation in *Ginkgo biloba* L. *BMC Plant Biol.* 19:527. doi: 10.1186/s12870-019-2141-z

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, 316–322. doi: 10.1093/nar/gkr483

Yang, L., Froberg, J. E., and Lee, J. T. (2014). Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem. Sci.* 39, 35–43. doi: 10.1016/j.tibs.2013.10.002

Yang, Z., Yang, Z., Yang, C., Wang, Z., Chen, D., Xie, Y., et al. (2020). Identification and genetic analysis of alternative splicing of long non-coding RNAs in tomato initial flowering stage. *Genomics* 112, 897–907. doi: 10.1016/j.ygeno.2019.06.005

Ye, J., Cheng, S., Zhou, X., Chen, Z., Kim, S. U., Tan, J., et al. (2019). A global survey of full-length transcriptome of *Ginkgo biloba* reveals transcript variants involved in flavonoid biosynthesis. *Ind. Crops Prod.* 139:111547. doi: 10.1016/j.indcrop.2019.111547

Zhang, D., Sun, W., Singh, R., Zheng, Y., Cao, Z., Li, M., et al. (2018). GRF-interacting factor1 regulates shoot architecture and meristem determinacy in maize. *Plant Cell* 30, 360–374. doi: 10.1105/tpc.17.00791

Zheng, Y., Jiao, C., Sun, H. H., Rosli, H. G., Pombo, M. A., Zhang, P. F., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 1667–1670. doi: 10.1016/j.molp.2016.09.014

Zhong, R., Demura, T., and Ye, Z.-H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* 18, 3158–3170. doi: 10.1105/tpc.106.047399

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hou, Lian, Cai, Wang, Liang and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 May 2021
doi: 10.3389/fgene.2021.679162

Screening and Mapping for Head Blast Resistance in a Panel of CIMMYT and South Asian Bread Wheat Germplasm

*Xinyao He[1], Philomin Juliana[1], Muhammad R. Kabir[2], Krishna K. Roy[2], Rabiul Islam[2], Felix Marza[3], Gary Peterson[4], Gyanendra P. Singh[5], Aakash Chawade[6], Arun K. Joshi[7], Ravi P. Singh[1] and Pawan K. Singh[1]**

[1] International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico, [2] Bangladesh Wheat and Maize Research Institute (BWMRI), Nashipur, Bangladesh, [3] Instituto Nacional de Innovación Agropecuaria y Forestal (INIAF), La Paz, Bolivia, [4] United States Department of Agriculture–Agricultural Research Service, Foreign Disease-Weed Science Research Unit, Fort Detrick, MD, United States, [5] ICAR-Indian Institute of Wheat and Barley Research, Karnal, India, [6] Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden, [7] BISA/CIMMYT-India, New Delhi, India

OPEN ACCESS

Edited by:
Reyazul Rouf Mir, Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Reviewed by:
Sanu Arora, John Innes Centre, United Kingdom
Muhammad Sajjad, COMSATS University Islamabad, Pakistan

***Correspondence:**
Pawan K. Singh
pk.singh@cgiar.org

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *11 March 2021*
Accepted: *21 April 2021*
Published: *13 May 2021*

Citation:
He X, Juliana P, Kabir MR, Roy KK, Islam R, Marza F, Peterson G, Singh GP, Chawade A, Joshi AK, Singh RP and Singh PK (2021) Screening and Mapping for Head Blast Resistance in a Panel of CIMMYT and South Asian Bread Wheat Germplasm. Front. Genet. 12:679162. doi: 10.3389/fgene.2021.679162

Wheat blast (WB) is a destructive disease in South America and its first outbreak in Bangladesh in 2016 posed a great risk to food security of South Asian countries. A genome wide association study (GWAS) was conducted on a diverse panel of 184 wheat genotypes from South Asia and CIMMYT. Phenotyping was conducted in eight field experiments in Bolivia and Bangladesh and a greenhouse experiment in the United States. Genotypic data included 11,401 SNP markers of the Illumina Infinium 15K BeadChip and four additional STS markers on the 2NS/2AS translocation region. Accessions with stable WB resistance across experiments were identified, which were all 2NS carriers. Nevertheless, a dozen moderately resistant 2AS lines were identified, exhibiting big variation among experiments. Significant marker-trait associations (MTA) were detected on chromosomes 1BS, 2AS, 6BS, and 7BL; but only MTAs on 2AS at the 2NS/2AS translocation region were consistently significant across experiments. The resistant accessions identified in this study could be used in production in South Asian countries as a preemptive strategy to prevent WB outbreak.

Keywords: GWAS, resistance screening, *Magnaporthe oryzae*, *Pyricularia oryzae*, MoT

INTRODUCTION

Bread wheat grown in tropical and subtropical regions is subjected to a range of diseases, among which is wheat blast (WB). This is an emerging disease with growing impacts on wheat production in South America, South Asia, and Africa, and is a potential threat to other major wheat producers like India, United States and China (Duveiller et al., 2016; Cruz and Valent, 2017; Tembo et al., 2020). The disease had been endemic in four South American countries, Brazil, Bolivia, Paraguay, and Argentina, before its outbreak in Bangladesh in 2016, affecting 15,000 ha wheat fields with yield losses of 5–51% (Islam et al., 2016; Malaker et al., 2016). Soon thereafter, this disease was identified in the Muchinga Province of Zambia during the 2017–2018 rainy season, indicating its introduction into the African continent (Tembo et al., 2020).

The causal agent of WB is the *Triticum* pathotype of the ascomycete fungus *Magnaporthe oryzae* (MoT), which is closely related to the notorious pathogen of rice blast, the *Oryzae* pathotype of *M. oryzae*, but cross infection normally does not happen (Gladieux et al., 2018). Spike infection is the most conspicuous symptom of this disease, showing bleached spikelets above the infection point on rachis, which, together with the often-appeared gray mold on the infected rachis, are typical diagnostic symptoms of WB. Besides spike, MoT can infect all the above ground parts of a wheat plant, especially leaves, where the lesions are elliptical with a white center and a brown margin on the upper side and grayish mold on the underside (Igarashi, 1990). Leaf infection is less prominent than spike infection and usually does not lead to severe yield loss; but the fungal conidia produced on leaves could be an important source of inoculum for spike infection (Cruz et al., 2015).

The disease is favored by hot and humid conditions around anthesis, and according to Cardoso et al. (2008), the optimal temperature for the infection was at 30°C with extended wetting period. WB develops very fast under conducive conditions, and a severe infection might be developed only a few days after the first appearance of WB symptom, which gives no time for fungicidal application (Duveiller et al., 2016). Chemical control has major drawbacks in WB management, including low control efficiencies and fungal resistance (Castroagudín et al., 2015; Cruz et al., 2018). Under such circumstances, varietal resistance to WB is indispensable, as well as being economical and environment friendly. Field resistance of wheat against WB is a quantitative trait, and no immunity has been identified (Cruz et al., 2016). Several varieties have shown moderate disease resistance or tolerance in WB screening experiments or in long-term production in WB epidemic regions, such as BR 18 and CD 116 in Brazil, Montacu and Urubo in Bolivia, and Caninde#1 and Itapua 75 in Paraguay (Kohli et al., 2011). However, it was later found that many such varieties are 2NS/2AS translocation carriers, indicating a very narrow genetic variation (Cruz et al., 2016).

The 2NS chromosomal segment was initially introduced from *Aegilops ventricosa* (Zhuk.) into wheat genome to utilize rust resistance genes *Lr37*, *Sr38,* and *Yr17* (Helguera et al., 2003). Later, additional resistance genes have been identified from this chromosomal segment, like *Cre5* for cereal cyst nematode resistance (Jahier et al., 2001), *Rkn3* for root-knot nematodes resistance (Williamson et al., 2013), as well as WB resistance mentioned above, contributing to 64-81% reduction in head blast severity in Cruz et al. (2016). Recently, Juliana et al. (2019) reported a prominent role of the 2NS/2AS translocation in CIMMYT germplasm, conferring yield advantage, lodging resistance, along with blast and rust resistance. Apart from the 2NS/2AS translocation, there have been very limited blast resistance genes identified for MoT, i.e., *Rmg2*, *Rmg3*, *Rmg7*, *Rmg8*, *RmgTd(t),* and *RmgGR119* (Zhan et al., 2008; Cumagun et al., 2014; Tagle et al., 2015; Cruz et al., 2016; Wang et al., 2018a). However, most of these genes have been overcome by new MoT isolates and only *Rmg8* and *RmgGR119* remained to be validated in field experiments with new MoT isolates.

Since the introduction of MoT from South America to Bangladesh, the affected areas have been increasing, despite the unfavorable weather conditions witnessed in the last few years (Islam et al., 2019), demonstrating the adaptation of the pathogen to South Asian environment. This poses a great risk to the neighboring countries, especially India that shares a very long international border with Bangladesh and relies heavily on wheat production for food security (Mottaleb et al., 2019). Being both air- and seed-borne, MoT could be easily blown by wind or introduced via trade to the neighboring countries. Therefore, it is imperative to screen South Asian wheat germplasm for WB resistance with the hope to identify resistant genotypes that could be used in production/breeding for WB resistance. It will help to prepare for the scenario if WB occurs and spreads in these countries. The objectives of the current study were designed with this aim and includes (1) screen a panel of CIMMYT and South Asian bread wheat germplasm for adult plant WB resistance and (2) identify molecular markers associated with WB resistance via genome-wide association study (GWAS).

MATERIALS AND METHODS

Plant Materials

A panel of 184 spring wheat genotypes was used in the present study. The genotypes were obtained from CIMMYT-Mexico (97 genotypes), India (40), Bangladesh (19), and Nepal (28). In the last three groups, there are genotypes directly introduced from CIMMYT due to extensive traditional utilization of CIMMYT germplasm in these countries.

Field Experiments

The field experiments took place in three locations, Quirusillas and Okinawa in the Department of Santa Cruz, Bolivia, and Jashore in the Division of Khulna, Bangladesh. Field trials were performed in the 2017–2018 and 2018–2019 cropping seasons in Quirusillas, the 2018 cropping season in Okinawa, and the 2017-18 cropping season in Jashore. Two sowings of approx. 14 days apart were adopted in each trial, totally making eight experiments, which were named as per the location ("Quir" for Quirusillas, "Oki" for Okinawa, and "Jash" for Jashore), cropping season ("18" for the 2017–2018 or 2018 cycle, and "19" for the 2018–2019 cycle), and sowing ("a" for the first sowing and "b" for the second). For example, Quir18b represents the second sown experiment in the 2017–2018 cycle conducted in Quirusillas.

The cropping cycle runs from December to April in Quirusillas and Jashore, and from May to September in Okinawa. Field experimental units were 1-m double rows separated by 20-cm spacing in all three locations, and no replication was set within each sowing. A field misting system was equipped at each site to create a micro-environment favorable for WB infection. The system worked from 8am to 7pm in Quirusillas and Okinawa and from 9am to 5pm in Jashore, with a 10-min-misting per hour, during the WB development period. Field inoculation was done twice in all the experiments, with the first at anthesis and the second at two days after anthesis, using a backpack sprayer. Inoculum was a mixture of locally collected MoT isolates

with high pathogenesis, which included isolates QUI1505, QUI1601, QUI1612, OKI1503, and OKI1704 in Quirusillas and Okinawa, and BHO17001, MEH17003, GOP17001.2, RAJ17001, CHU16001.3, and JES16001 in Jashore. The first three letters of the isolate names indicate the place of collection, followed by two digits for the year of collection and the rest digits as isolate identifiers. The MoT isolates were cultured on oatmeal agar medium following the protocol by He et al. (2020b), and the inoculum was adjusted at a concentration of 80,000 spores/mL before field application. Two local checks were used in each experiment, which are Urubo (resistant check) and Atlax (susceptible check) in Bolivia and BARI Gom 33 (resistant check) and BARI Gom 26 (susceptible check) in Bangladesh.

WB evaluation was performed on 10 spikes that had been marked at anthesis at 21 days after the first inoculation, for which the total and infected number of spikelets were recorded. Incidence was derived from the proportion of spikes with blast infection and severity was calculated as the averaged percentage of infected spikelets. WB index was calculated with the formula WB index = incidence × severity, which was used in all subsequent analysis. In addition, days to heading (DH) and plant height (PH) were scored in all experiments.

Greenhouse Experiment

In 2017, a greenhouse experiment was conducted in the biosafety level-3 laboratory at USDA-ARS, Foreign Disease-Weed Science Research Unit, Fort Detrick, MD, United States (nominated as US17 in this study). The accessions were sown into 15-cm-diameter pots containing a commercial potting medium (Metro-Mix 360; Hummert International, Earth City, MO, United States), which were arranged in an incomplete randomized block design with one replication. Wheat varieties Urubo and Glenn were used as resistant and susceptible checks, respectively. A MoT isolate B-71 collected in Okinawa (Bolivia) in 2012 was used. Inoculation production was carried out on homemade oatmeal agar according to Valent et al. (1991), and the spore concentration was adjusted to 20,000 spores/ml for inoculation. Wheat heads were inoculated approximately two days after full head emergence, with approximately 0.75 ml of inoculum applied for each accession using an airbrush (model 69492; Harbor Freight Tools, Camarillo, CA, United States). After inoculation, the heads were covered with black plastic bags (model S-12322BL; ULINE) moistened with water for 48 h to facilitate the infection. Percentage WB severity was evaluated at 12 to 14 days after inoculation, which was determined when the susceptible check Glenn exceeded 90% of WB severity.

Genotyping

The panel was genotyped with Illumina Infinium 15K BeadChip at Trait Genetics GmbH, Germany. Markers with missing data points more than 10% or minor allele frequency less than 5% were excluded from further analysis. Four STS markers in the 2NS/2AS region were applied to evaluate their association with WB index and suitability for marker-assisted selection (MAS), including *Ventriup-LN2* reported by Helguera et al. (2003), *WGGB156* and *WGGB159* by Wang et al. (2018b), and *cslVrgal3* that was derived from a follow-up study of Seah et al. (2001) (E. Lagudah, pers. comm.).

Linkage Disequilibrium and Kinship Analysis

The linkage disequilibrium (LD) parameter R^2 among the markers was calculated using TASSEL 5[1], and LD was visualized with R^2 plotted against the physical distances. A kinship matrix and clusters among individual genotypes was generated with all the SNP markers.

Statistical and GWAS Analysis

Analysis of variance (ANOVA) was performed with the PROC GLM module in SAS program ver. 9.2, and the results were used for calculating the heritability estimates, using the formula $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{g^*y}^2/y + \sigma_{g^*s}^2/s + \sigma_e^2/sy)$ for experiments in Quirusillas and $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2/s)$ for those in Okinawa and Jashore, where σ_g^2 represents genetic variance, $\sigma_{g^*y}^2$ for genotype-by-year interaction, $\sigma_{g^*s}^2$ for genotype-by-sowing interaction, σ_e^2 for error variance, y for the number of years, and s for the number of sowing. Principal component analysis (PCA) on phenotypic data was conducted with the PAST software ver. 3.01 (Hammer et al., 2001).

Marker-trait association (MTA) tests were carried out in TASSEL 5 using the mixed linear model (MLM) (Yu and Buckler, 2006) with the optimum level of compression and the "population parameters previously determined" option. The first two principal components (PC) of the population structure were used as fixed effects, and the kinship among individuals estimated using the centered identity-by-state method (Endelman and Jannink, 2012) was used as a random effect in the mixed linear model. In addition, the multilocus mixed model (MLMM) and fixed and random model circulating probability unification (FarmCPU) models were also conducted using the R software package GAPIT v. 3.5 (Zhang et al., 2010). The *p*-values, additive effects and percentage variation explained by each marker were obtained and Manhattan plots with the −log10 *p*-values of the markers were plotted using the 'R' package CMplot. The Bonferroni method for multiple testing correction at an α level of 0.20 was used to declare significance of the markers.

RESULTS

Phenotypic Evaluation

The WB disease pressure varied greatly among experiments, with Jash18a being the lowest with a grand mean of WB index of 11.8% and Quir19a the highest of 39.3% among the field experiments (**Figure 1**). The resistant and susceptible checks were among the lines of the lowest and the highest infection, respectively, in all field trials (**Supplementary Table 1**). The greenhouse experiment US17, however, exhibited a much higher disease pressure, showing a grand mean of WB severity of 63.6%. A bimodal distribution of the genotypes was observed in most of the experiments, except for the ones with lower disease pressure, like Jash18a and Jash18b (**Figure 1**). ANOVA indicated significant effects of "Genotype" in all three locations, as well as the significant effects of "Year" in Quirusillas and "Sowing" in

[1]http://www.maizegenetics.net

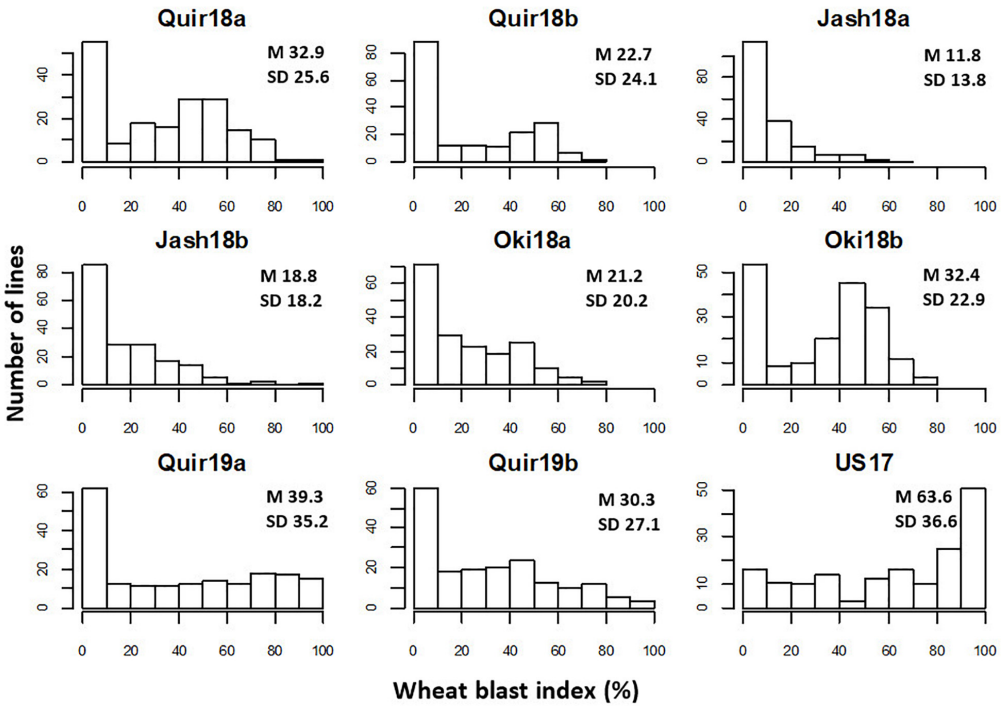


FIGURE 1 | Histograms of wheat blast index in individual experiments. "Quir" stands for Quirusillas, "Jash" for Jashore, and "Oki" for Okinawa, "18" and "19" for the 2017-18 or 2018 cycle and 2018-19 cycle, respectively, and "a" and "b" for the first and second sowing, respectively. "US17" stands for the 2017 greenhouse evaluation in the United States. Grand mean (M) and standard deviation (SD) values are presented for all experiments.

Okinawa and Jashore. A high heritability estimate of 0.84 was obtained for Quirusillas; but only moderate heritability estimates from 0.66 to 0.69 were obtained for the other two locations (**Table 1**). Phenotypic correlations of WB among experiments were all significant, with r values ranging from 0.27 to 0.72. Among the experiments, Jash18a that exhibited the lowest disease pressure showed generally poor correlation with others, whereas other experiments, especially those in Bolivia, exhibited better correlations (**Table 2**). According to the PCA results, however, it was US17 that had the least association with the rest experiments at the dimension of the first two PCs (**Figure 2**).

In both Jashore and Okinawa, the second sown experiments got higher infection than the first sown ones, whereas a reverse trend was found in Quirusillas (**Figure 1**). However, within a same sowing, no significant correlation was found between WB and DH except for Jash18b, with a low r value of 0.17. Similarly, significant correlation between WB and PH was only found in Quir19a and Quir19b, with low r values of around −0.20 (**Table 3**).

Genotyping, Population Structure and Linkage Disequilibrium

This part of results has been reported in our previous study on tan spot resistance of the same panel (Phuke et al., 2020), with the only difference being the incorporation of the four 2NS-associated STS markers. Briefly, 11,405 markers of good quality were selected after filtering for the subsequent analysis, with generally good genomic coverage for A- and B-genome chromosomes but not for D-genome chromosomes (**Supplementary Figure 2**). Two subpopulations were identified in both Kinship and PCA analysis, of which the small sub-population comprised only nine Indian accessions (**Supplementary Figures 3, 4**). The big sub-population could be further divided into two groups, of which the small group is composed of only South Asian lines, whereas the big group comprises CIMMYT lines and some South Asian accessions with CIMMYT parentage (**Supplementary Figure 4**). The average extent of LD was estimated at 25 Mb, a physical distance over which R^2 decayed to a critical value of 0.10 across the genome.

GWAS Results

Based on the Bonferroni method, p = 1.74E-5 was determined as the significant level of MTAs. However, only markers on the 2NS/2AS translocation regions exceeded this threshold regardless of the GWAS algorithm used. When the threshold of p was reduced to 0.001, some significant SNPs on other chromosomes were detected, but none of them could be repeatedly detected across experiments (**Figure 3**). With MLMM and FarmCPU, only three markers outside the 2NS/2AS region could be detected in two of the nine experiments, being located on 1BS, 6BS, and 7BL, with their p values mostly higher than 1.74E-5 and could only be regarded as putative MTA loci (**Table 4**). It is noteworthy that the model MLM did not fit Jash18a; but MLMM and FarmCPU that fitted this experiment were not able to identify a stably expressed MTA locus either. In order to detect potential MTAs that might have been masked by the strong phenotypic

TABLE 1 | Analysis of variance for wheat blast index in different locations and its heritability estimates.

Location	Source	DF	Mean square	*F* value	*P* value	Heritability
Quirusillas	Genotype	183	2273.20	6.18	<0.0001	0.84
	Year	1	9247.88	25.16	<0.0001	
	Sowing (Year)	1	41.40	0.11	0.7375	
	Genotype × Year	183	333.07	0.91	0.7469	
	Genotype × Sowing	183	226.06	0.62	0.9995	
	Error	182	367.56			
Jashore	Genotype	183	387.20	2.92	<0.0001	0.66
	Sowing	1	4453.70	34.39	<0.0001	
	Error	183	132.42			
Okinawa	Genotype	183	715.00	3.26	<0.0001	0.69
	Sowing	1	11663.17	53.23	<0.0001	
	Error	182	219.12			

TABLE 2 | Pearson correlation coefficients of wheat blast index among the 9 environments.

	US17	Quir18a	Quir18b	Jash18a	Jash18b	Oki18a	Oki18b	Quir19a
Quir18a	0.58							
Quir18b	0.50	0.58						
Jash18a	0.35	0.29	0.48					
Jash18b	0.52	0.43	0.52	0.55				
Oki18a	0.52	0.56	0.57	0.41	0.50			
Oki18b	0.63	0.63	0.50	0.30	0.45	0.56		
Quir19a	0.59	0.63	0.68	0.35	0.49	0.62	0.68	
Quir19b	0.55	0.71	0.58	0.27	0.52	0.63	0.66	0.72

All correlations were significant at $P < 0.0001$. "Quir" stands for Quirusillas, "Jash" for Jashore, and "Oki" for Okinawa, "18" and "19" for the 2017–2018 or 2018 cycle and 2018–2019 cycle, respectively, and "a" and "b" for the first and second sowing, respectively. "US17" stands for the 2017 greenhouse evaluation in the United States. Cell shades change from green to red with the increase of correlation coefficients.

effects of 2NS, the non-2NS lines were analyzed independently for GWAS; nevertheless, no repeatable MTA could be found (data not shown).

Effects of the 2NS/2AS Translocation on Blast Resistance

As the only major MTA locus identified in the current study, 2NS/2AS exhibited strong phenotypic effects on WB resistance. In the phenotype-based PCA analysis, PC1 mainly reflected WB resistance and separated most 2NS lines from the 2AS lines, clearly demonstrating the different WB resistance levels of the two groups (**Figure 2**). In total, 55 lines were identified as 2NS carriers, accounting for 29.9% of the entire panel. Majority of the 2NS lines were from CIMMYT, with only nine, three and one accession from India, Nepal and Bangladesh, respectively (**Supplementary Table 1**).

In the field experiments, the 2NS accessions exhibited a grand mean WB index of 3.8%, which was 35.7% for the 2AS lines. In the greenhouse experiment, however, the corresponding values became 26.0% and 79.6%, respectively, much higher than those in the field experiments (**Figure 4**). Despite the general trend, there were several 2NS carriers with higher WB infection, along with some 2AS carriers showing moderate resistance (**Figures 2**, **4**). Unexpectedly, the famous 2NS donor, Milan, exhibited a grand mean of 39.9% in field experiments, probably due to the residual heterozygosity in the 2NS/2AS region (**Supplementary Table 1**).

As expected, the best performers in field experiments were mostly 2NS lines, whereas only four 2AS lines, CIM-5, CIM-11, CIM-32, and IND-29, exhibited a grand mean WB index less than 15%, a tentative threshold for WB resistance. However, in the greenhouse experiment, these four lines showed high levels of infection, with WB severities ranging from 66.6 to 100.0% (**Supplementary Table 1**), implying their vulnerability to WB under high disease pressure. Generally, the best 2NS lines performed more consistently across experiments, whereas the best 2AS lines were more subjective to environmental conditions, with wider ranges of WB index or severity (**Supplementary Table 1**).

DISCUSSION

Wheat blast had been an endemic disease of South America until its large-scale outbreak in Bangladesh in 2016 (Malaker et al., 2016), and its recent incidence in Zambia further corroborated the importance of preemptive actions against this disease in countries where WB is not detected (Tembo et al., 2020). Fortunately, WB has not been officially reported in any South Asian countries except Bangladesh, therefore, identification of WB resistant genotypes from the local wheat germplasm, as well

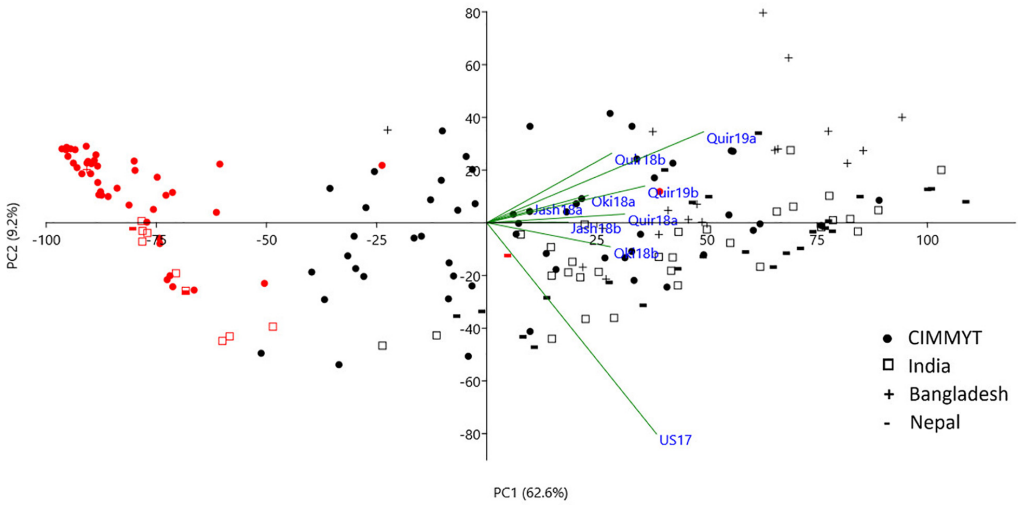


FIGURE 2 | Principal component analysis (PCA) of the 184 accessions on wheat blast index across nice environments. Cosine of the angle between vectors indicates correlation between variables in the dimension of the first two principal components (PCs). *Red* symbols denote accessions with the 2NS/2AS translocation. Refer to **Supplementary Figure 1** for a PCA chart with accession labels.

TABLE 3 | Phenotypic correlation of field wheat blast index with days to heading (DH) and plant height (PH) in individual experiments.

	Quir18a	Quir18b	Jash18a	Jash18b	Oki18a	Oki18b	Quir19a	Quir19b
DH	−0.06	−0.13	0.07	0.17*	−0.06	0.06	−0.01	−0.04
PH	−0.13	−0.11	0.13	0.01	−0.07	−0.04	−0.24**	−0.18**

*"Quir" stands for Quirusillas, "Jash" for Jashore, and "Oki" for Okinawa, "18" and "19" for the 2017–2018 or 2018 cycle and 2018–2019 cycle, respectively, and "a" and "b" for the first and second sowing, respectively. *$p < 0.05$; **$p < 0.01$.*

as CIMMYT genotypes that have been extensively used in South Asian breeding programs, is an important finding to prevent or slow down the possible spread of WB in the region or worldwide.

In the current study, we evaluated 184 CIMMYT and South Asian wheat genotypes for their resistance to WB in eight field experiments and one greenhouse experiment. However, accessions with high and stable WB resistance were all 2NS carriers, in agreement with a previous study (Juliana et al., 2020). There were several 2AS lines showing resistance or moderate resistance under field conditions, though, they turned out to be susceptible in the greenhouse experiment. This implies that the currently identified non-2NS resistance is not as robust as the 2NS resistance under high disease pressure imposed under greenhouse conditions. Although misting systems were equipped for the screening nurseries in Bolivia and Bangladesh to maintain a conducive environmental condition for WB infection, greenhouse condition exerted a stronger disease pressure due to the constant humidity and temperature conditions favorable for WB infection (Cruppe et al., 2020). Several accessions of the current study have been evaluated under greenhouse condition in Bolivia and also exhibited higher WB severity (Marza et al., 2019). It is noteworthy that four 2AS lines, i.e., TEPOCA T89, QUAIU #1, SUPER 152, and SOKOLL/ROLF07, which Cruppe et al. (2020) reported as resistant or moderately resistant, were also included in the current study, but they exhibited wide ranges of WB across experiments, just like other promising 2AS lines (**Supplementary Table 1**). Nevertheless, we must say that such 2AS lines are still valuable in production, especially in terms of diversification of source of WB resistance. Because relying solely on 2NS is risky, due to the emergence of 2NS-virulent isolates that have already been found in South America (Ceresini et al., 2018; Cruppe et al., 2020). In production, these moderately resistant 2AS lines must be able to endure low to medium WB pressure; and under higher disease pressure, other management approaches especially fungicide application could be helpful in reducing yield loss (Cruz et al., 2018).

It is noteworthy that several accessions of the GWAS panel have been released in the WB affected or threatened countries. Two Indian accessions, HD 2967 (IND-1) and HD 3171 (IND-13) have been recommended to the farmers of WB prone areas in West Bengal that borders to Bangladesh[2] (accessed in Jan. 30, 2021). A CIMMYT breeding line Kachu/Solala (BAW-1260, BGD-19) has been released in Bangladesh as BARI Gom 33 and in Bolivia as INIAF Okinawa, while another CIMMYT accession BORLAUG100 F2014 (CIM-35) has been released in Bangladesh as WMRI Gom 3, in Bolivia as INIAF Tropical, and in Nepal as Borlaug 2020. However, all these accessions are 2NS carriers, indicating the limited genetic variation in WB resistant sources, which needs to be complimented with non-2NS resistance. The implication for wheat breeding in the WB vulnerable regions is to combine different resistant sources to achieve a better and

[2]https://www.aicrpwheatbarley.org/wp-content/uploads/2020/07/Crop-Protection-Report-2019-20_compressed.pdf

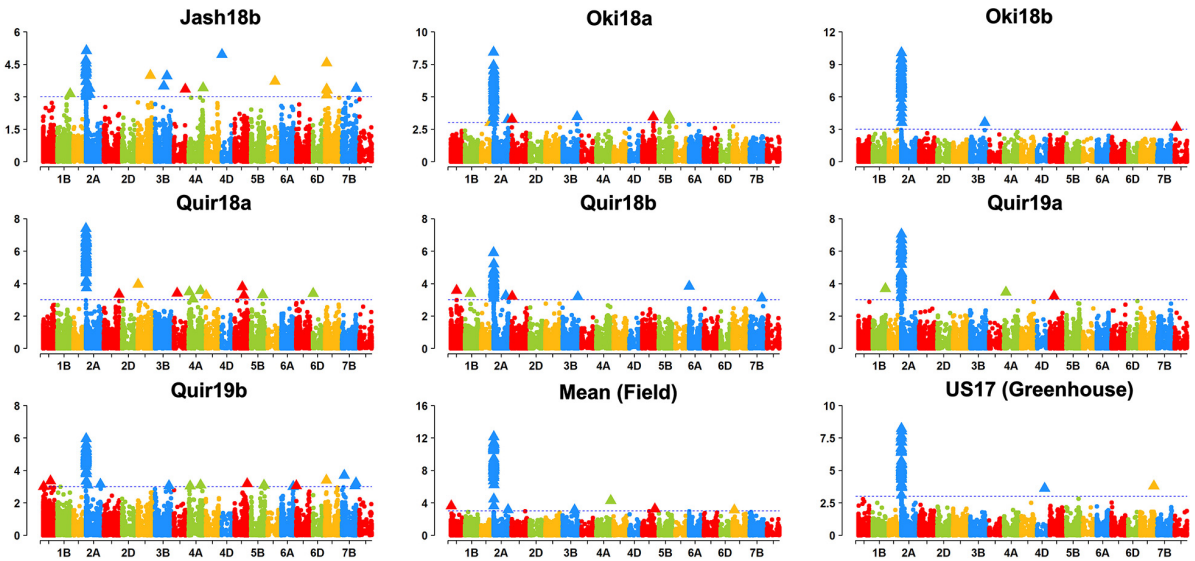


FIGURE 3 | Manhattan plots based on MLM model. "Quir" stands for Quirusillas, "Jash" for Jashore, and "Oki" for Okinawa, "18" and "19" for the 2017-18 or 2018 cycle and 2018-19 cycle, respectively, and "a" and "b" for the first and second sowing, respectively. "US17" stands for the 2017 greenhouse evaluation in the United States. Jash18a did not fit this model and is not shown, instead a plot for mean data across the rest seven field trials is presented.

TABLE 4 | Markers significantly associated with wheat blast resistance.

Algorithm	SNP	Chromosome	Position (Mb)	*P* value	R^2	Experiment
MLM, MLMM, FarmCPU	Multiple SNPs	2NS/2AS	0-35.4	1.87E-13 to 9.35E-4	0.26-0.50	All
MLMM	IAAV2838	1BS	41.6	5.71E-4 to 8.95E-4	0.13	Oki18b, Quir19b
FarmCPU	AX-94523488	6BS	51.3	2.25E-4 to 2.63E-4	0.07	Oki18a, Quir19b
FarmCPU	AX-95215927	7BL	682.9	1.84E-7 to 1.26E-4	0.35	Quir18b, Quir19a

Significant markers on 2NS/2AS are listed in ***Supplementary Table 1****. Physical positions for SNPs are from Chinese Spring RefSeq ver. 1.0.*

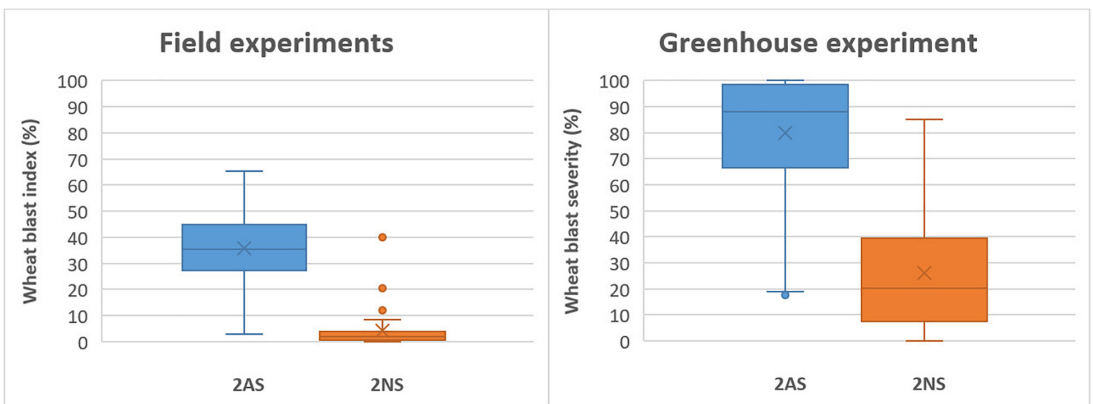


FIGURE 4 | Phenotypic effects of the 2NS/2AS translocation on wheat blast resistance for field experiments in Bolivia and Bangladesh (left) and greenhouse experiment.

durable resistance, just like other wheat diseases (Singh et al., 2016). To achieve this goal, crosses have been made between the 2AS lines mentioned above and elite CIMMYT breeding lines with 2NS. Additional non-2NS resistant sources will be utilized in breeding once identified; but the 2NS resistance will still be a backbone for WB resistance breeding, considering its strong phenotypic effects.

The CIMMYT breeding line Milan (CM-34) is a WB resistance donor that has been widely used in South American, leading to many WB resistant varieties with Milan in their pedigree. Such varieties include Sausal CIAT that was released in Bolivia, CD 116 in Brazil, Caninde#1 in Paraguay, etc. (Kohli et al., 2011). However, according to our results, Milan might have some residual heterozygosity in the 2NS region, as evidenced by the

many heterozygous genotypes of markers in the 2NS region (**Supplementary Table 1**). Indeed, upon genotyping multiple Milan plants with 2NS-associated markers, we did identify a few individuals that harbors 2AS instead of 2NS (data not shown). This is also reflected in the field experiments, where Milan plants with susceptible reaction to WB was often observed, although majority of the plants were resistant. Another explanation to this could be seed mixture, which is less possible since the Milan plants in the field appeared homogeneous apart from WB resistance. Therefore, marker diagnosis is needed to identify the 2NS-bearing Milan for crosses aimed at WB resistance.

Despite the attempts on different GWAS algorithms, MTAs on the 2NS/2AS region were the only stably expressed QTL across experiments, and MTAs in other chromosome regions were hardly repeatable and have not been reported in other genetic studies (Ferreira et al., 2020; Goddard et al., 2020; He et al., 2020b; Juliana et al., 2020). This could be caused by 1) genomic regions with major effects on WB resistance were not present in this panel, 2) they were of very low frequencies below the detection power of the GWAS algorithms employed in this study, or 3) bias introduced by the 15K SNP chip that was developed based on genotypes without WB resistance genes that otherwise would be detected in this study. Our previous study indicated an interval of about 2.3 Mb on 2NS that might harbor the underlying WB resistance gene (He et al., 2020b); however, in the present study, the significant markers on the 2AS/2NS region distributed evenly across the region, and no clear cluster with lower *p* values was identified, probably being caused by the lower mapping resolution. Therefore, larger GWAS panel or specifically designed fine mapping populations will be needed to better map the underlying gene. An additional observation was that STS markers *Ventriup-LN2* and *cslVrgal3* predicted the presence of 2NS better than *WGGB156* and *WGGB159* in Indian germplasm (**Supplementary Table 1**), which was unexpected since our previous results in a biparental population indicated that the latter two markers were closer to the WB resistance QTL (He et al., 2020b). A possible explanation to this could be a founder parent in Indian germplasm, in which the susceptible alleles of *WGGB156* and *WGGB159* were linked with the resistance allele of the WB QTL.

Phenological traits like PH and DH are often associated with disease resistance under field conditions, as frequently reported for Fusarium head blight (FHB, Xu et al., 2020), spot blotch (He et al., 2020a) etc. For field WB resistance, however, such association appeared to be less significant (He et al., 2020b). Based on the prevailing weather conditions, it is a general rule that early sowing in South America and late sowing in Bangladesh should be avoided to prevent severe WB infection (Cruz and Valent, 2017; He et al., 2020b). However, in a specific cropping cycle, this rule may not be true, e.g., the 2018 cycle in Okinawa, where the second sowing exhibited higher infection than the first (**Figure 1**). Similarly, the general trend for the association between DH and WB agreed mostly with that between sowing and WB; but the correlation coefficients were very low and mostly non-significant. Plant height mostly showed a negative correlation with WB index, possibly having a similar underlying mechanism to that in FHB, i.e. spikes of tall wheat plants are well ventilated to become drier compared to those of short plants (Yan et al., 2011), which is unfavorable for WB infection. Despite this, the correlations were mostly not significant, thus the impact of PH on WB infection must be marginal if not absent. Nevertheless, a GWAS or biparental population with low variation in DH and PH is always preferred in field experiments to reduce the influence from these confounding effects to have more consistent WB resistance results.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the Dataverse repository and can be accessed via the link https://hdl.handle.net/11529/10548559.

AUTHOR CONTRIBUTIONS

PS, RS, and AC conceived and designed the experiments. MK, KR, RI, FM, XH, and PS performed field trials. GP conducted the greenhouse experiments. GS, AJ, and RS provided germplasm. XH and PJ analyzed the data. XH wrote the first draft of the manuscript. All authors contributed and approved the final draft of the manuscript.

FUNDING

Financial supports from the CGIAR Research Program on WHEAT, Indian Council of Agricultural Research, India, Vetenskapsrådet (the Swedish Research Council), Sweden, and Australian Centre for International Agricultural Research, Australia, are gratefully acknowledged. USDA contribution was funded by ARS National Program 303, CRIS 8044-22000-046-00D and United States Agency for International Development (USAID) Agreement 021-022310-004-96-R. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

ACKNOWLEDGMENTS

Technical support from Jose Asister, Jhonny Villagomez, Roberto Butron, and Francisco Lopez in field trials in Bolivia, Tanusree Roy and Rokshana Begum in field trials in Bangladesh, Mayela Flores in genotyping work (STS markers) at CIMMYT- Mexico,

and DOE ORISE intern Alyssa Kloos and ARS technician Kelli Keller for greenhouse evaluations is highly acknowledged. The authors are grateful to Dr. Evans Lagudah, CSIRO Plant Industry, Canberra, Australia, for kindly sharing the unpublished *cslVrgal3* marker sequence.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.679162/full#supplementary-material

REFERENCES

Cardoso, C.a.D.A, Reis, E. M., and Moreira, E. N. (2008). Development of a warning system for wheat blast caused by Pyricularia grisea. *Summa Phytopathologica* 34, 216–221. doi: 10.1590/s0100-54052008000300002

Castroagudín, V. L., Ceresini, P. C., De Oliveira, S. C., Reges, J. T. A., Maciel, J. L. N., Bonato, A. L. V., et al. (2015). Resistance to QoI fungicides is widespread in Brazilian populations of the wheat blast pathogen Magnaporthe oryzae. *Phytopathology* 105, 284–294. doi: 10.1094/phyto-06-14-0184-r

Ceresini, P. C., Castroagudín, V. L., Rodrigues, F. Á, Rios, J. A., Aucique-Pérez, C. E., Moreira, S. I, et al. (2018). Wheat blast: Past, present, and future. *Annual Review of Phytopathology* 56, 427–456.

Cruppe, G., Cruz, C. D., Peterson, G., Pedley, K., Asif, M., Fritz, A., et al. (2020). Novel sources of wheat head blast resistance in modern breeding lines and wheat wild relatives. *Plant Disease* 104, 35–43. doi: 10.1094/pdis-05-19-0985-re

Cruz, C. D., Kiyuna, J., Bockus, W. W., Todd, T. C., Stack, J. P., and Valent, B. (2015). Magnaporthe oryzae conidia on basal wheat leaves as a potential source of wheat blast inoculum. *Plant Pathology* 64, 1491–1498. doi: 10.1111/ppa.12414

Cruz, C. D., Peterson, G. L., Bockus, W. W., Kankanala, P., Dubcovsky, J., Jordan, K. W., et al. (2016). The 2NS translocation from Aegilops ventricosa confers resistance to the Triticum pathotype of Magnaporthe oryzae. *Crop Science* 56, 990–1000. doi: 10.2135/cropsci2015.07.0410

Cruz, C. D., Santana, F. M., Todd, T. C., Maciel, J. L. N., Kiyuna, J., Baldelomar, D. F., et al. (2018). Multi-environment assessment of fungicide performance for managing wheat head blast (WHB) in Brazil and Bolivia. *Tropical Plant Pathology* 44, 183–191. doi: 10.1007/s40858-018-0262-9

Cruz, C. D., and Valent, B. (2017). Wheat blast disease: danger on the move. *Tropical Plant Pathology* 42, 210–222. doi: 10.1007/s40858-017-0159-z

Cumagun, C. J. R., Anh, V. L., Vy, T. T. P., Inoue, Y., Asano, H., Hyon, G. S., et al. (2014). Identification of a hidden resistance gene in tetraploid wheat using laboratory strains of Pyricularia oryzae produced by backcrossing. *Phytopathology* 104, 634–640. doi: 10.1094/phyto-04-13-0106-r

Duveiller, E., He, X., and Singh, P. K. (2016). "Wheat blast: An emerging disease in South America potentially threatening wheat production," in *The World Wheat Book*, Vol. 3, eds A. Bonjean and M. Van Ginkel (Paris: Lavoisier), 1107–1122.

Endelman, J. B., and Jannink, J.-L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics* 2, 1405–1413. doi: 10.1534/g3.112.004259

Ferreira, J. R., Torres, G.a.M, Consoli, L., Camilotti, G. A., Scagliusi, S. M. M., Nhani, A., et al. (2020). Quantitative trait loci conferring blast resistance in hexaploid wheat at adult plant stage. *Plant Pathology* 70, 100–109. doi: 10.1111/ppa.13278

Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J. L. N., Nhani, A. Jr., et al. (2018). Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus Magnaporthe oryzae. *mBio* 9, e1219–e1217.

Goddard, R., Steed, A., Chinoy, C., Ferreira, J. R., Scheeren, P. L., Maciel, J. L. N., et al. (2020). Dissecting the genetic basis of wheat blast resistance in the Brazilian wheat cultivar BR 18-Terena. *BMC Plant Biol* 20:398. doi: 10.1186/s12870-020-02592-0

Hammer, Ø, Harper, D., and Ryan, P. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontolia Electronica* 4, 1–9.

He, X., Dreisigacker, S., Sansaloni, C., Duveiller, E., Singh, R. P., and Singh, P. K. (2020a). Quantitative trait loci mapping for spot blotch resistance in two biparental mapping populations of bread wheat. *Phytopathology* 110, 1980–1987. doi: 10.1094/phyto-05-20-0197-r

He, X., Kabir, M. R., Roy, K. K., Anwar, M. B., Xu, K., Marza, F., et al. (2020b). QTL mapping for field resistance to wheat blast in the Caninde#1/Alondra population. *Theoretical and Applied Genetics* 133, 2673–2683.

Helguera, M., Khan, I. A., Kolmer, J., Lijavetzky, D., Zhong-Qi, L., and Dubcovsky, J. (2003). PCR assays for the Lr37-Yr17-Sr38 cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. *Crop Science* 43, 1839–1847. doi: 10.2135/cropsci2003.1839

Igarashi, S. (1990). "Update on wheat blast (Pyricularia oryzae) in Brazil," in *Wheat for the nontraditional warm areas. A proceedings of the international conference*, ed. D. A. Saunders (Foz do Iguaçu, Brazil: CIMMYT), 480–485.

Islam, M. T., Kim, K. H., and Choi, J. (2019). Wheat blast in Bangladesh: The current situation and future impacts. *Plant Pathology Journal* 35, 1–10. doi: 10.5423/ppj.rw.08.2018.0168

Islam, T., Croll, D., Gladieux, P., Soanes, D., Persoons, A., Bhattacharjee, P., et al. (2016). Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. *BMC Biology* 14:84. doi: 10.1186/s12915-016-0309-7

Jahier, J., Abelard, P., Tanguy, M., Dedryver, F., Rivoal, R., Khatkar, S., et al. (2001). The Aegilops ventricosa segment on chromosome 2AS of the wheat cultivar 'VPM1' carries the cereal cyst nematode resistance gene Cre5. *Plant Breeding* 120, 125–128. doi: 10.1046/j.1439-0523.2001.00585.x

Juliana, P., He, X., Kabir, M. R., Roy, K. K., Anwar, M. B., Marza, F., et al. (2020). Genome-wide association mapping for wheat blast resistance in CIMMYT's international screening nurseries evaluated in Bolivia and Bangladesh. *Scientific Reports* 10, 15972.

Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., et al. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat Genet* 51, 1530–1539. doi: 10.1038/s41588-019-0496-6

Kohli, M. M., Mehta, Y. R., Guzman, E., De Viedma, L., and Cubilla, L. E. (2011). Pyricularia blast - a threat to wheat cultivation. *Czech Journal of Genetics and Plant Breeding* 47, S130–S134.

Malaker, P. K., Barma, N. C., Tewari, T. P., Collis, W. J., Duveiller, E., Singh, P. K., et al. (2016). First report of wheat blast caused by Magnaporthe oryzae pathotype Triticum in Bangladesh. *Plant Disease* 100, 2330. doi: 10.1094/pdis-05-16-0666-pdn

Marza, F., Butrón, R., Maydana, A., Singh, P., and He, X. (2019). "Resistance to wheat blast among cultivars under greenhouse conditions," in *Proceedings of the 1st International Wheat Congress* (Canada), 106.

Mottaleb, K. A., Singh, P. K., Sonder, K., Kruseman, G., and Erenstein, O. (2019). Averting wheat blast by implementing a 'wheat holiday': In search of alternative crops in West Bengal, India. *PLoS One* 14:e0211410. doi: 10.1371/journal.pone.0211410

Phuke, R. M., He, X., Juliana, P., Bishnoi, S. K., Singh, G. P., Kabir, M. R., et al. (2020). Association mapping of seedling resistance to tan spot (*Pyrenophora tritici-repentis* race 1) in CIMMYT and South Asian wheat germplasm. *Frontiers in Plant Science* 11:1309. doi: 10.3389/fpls.2020.01309

Seah, S., Bariana, H., Jahier, J., Sivasithamparam, K., and Lagudah, E. S. (2001). The introgressed segment carrying rust resistance genes Yr17, Lr37 and Sr38 in wheat can be assayed by a cloned disease resistance gene-like sequence. *Theoretical and Applied Genetics* 102, 600–605. doi: 10.1007/s001220051686

Singh, R. P., Singh, P. K., Rutkoski, J., Hodson, D. P., He, X., Jørgensen, L. N., et al. (2016). Disease impact on wheat yield potential and prospects of genetic control. *Annual Review of Phytopathology* 54, 303–322. doi: 10.1146/annurev-phyto-080615-095835

Tagle, A. G., Chuma, I., and Tosa, Y. (2015). Rmg7, a new gene for resistance to Triticum isolates of Pyricularia oryzae identified in tetraploid wheat. *Phytopathology* 105, 495–499. doi: 10.1094/phyto-06-14-0182-r

Tembo, B., Mulenga, R. M., Sichilima, S., M'siska, K. K., Mwale, M., Chikoti, P. C., et al. (2020). Detection and characterization of fungus (Magnaporthe oryzae pathotype Triticum) causing wheat blast disease on rain-fed grown

wheat (Triticum aestivum L.) in Zambia. *PLoS One* 15:e0238724. doi: 10.1371/journal.pone.0238724

Valent, B., Farrall, L., and Chumley, F. G. (1991). Magnaporthe grisea genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics* 127, 87–101. doi: 10.1093/genetics/127.1.87

Wang, S., Asuke, S., Vy, T. T. P., Inoue, Y., Chuma, I., Win, J., et al. (2018a). A new resistance gene in combination with Rmg8 confers strong resistance against Triticum isolates of Pyricularia oryzae in a common wheat landrace. *Phytopathology* 108, 1299–1306. doi: 10.1094/phyto-12-17-0400-r

Wang, Y., Zhang, H., Xie, J., Guo, B., Chen, Y., Zhang, H., et al. (2018b). Mapping stripe rust resistance genes by BSR-Seq: YrMM58 and YrHY1 on chromosome 2AS in Chinese wheat lines Mengmai 58 and Huaiyang 1 are Yr17. *Crop Journal* 6, 91–98. doi: 10.1016/j.cj.2017.03.002

Williamson, V. M., Thomas, V., Ferris, H., and Dubcovsky, J. (2013). An Aegilops ventricosa translocation confers resistance against root-knot nematodes to common wheat. *Crop Science* 53, 1412–1418. doi: 10.2135/cropsci2012.12.0681

Xu, K., He, X., Dreisigacker, S., He, Z., and Singh, P. K. (2020). Anther extrusion and its association with Fusarium head blight in CIMMYT wheat germplasm. *Agronomy* 10, 47. doi: 10.3390/agronomy10010047

Yan, W., Li, H. B., Cai, S. B., Ma, H. X., Rebetzke, G. J., and Liu, C. J. (2011). Effects of plant height on type I and type II resistance to Fusarium head blight in wheat. *Plant Pathology* 60, 506–512. doi: 10.1111/j.1365-3059.2011.02426.x

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003

Zhan, S. W., Mayama, S., and Tosa, Y. (2008). Identification of two genes for resistance to Triticum isolates of Magnaporthe oryzae in wheat. *Genome* 51, 216–221.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42, 355–360. doi: 10.1038/ng.546

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 He, Juliana, Kabir, Roy, Islam, Marza, Peterson, Singh, Chawade, Joshi, Singh and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 June 2021
doi: 10.3389/fgene.2021.646712

Genome-Wide Association Study on Root Traits Under Different Growing Environments in Wheat (*Triticum aestivum* L.)

Fengdan Xu[1†], Shulin Chen[1†], Xiwen Yang[1], Sumei Zhou[1], Junsen Wang[1], Ziliang Zhang[1], Yuan Huang[1], Miao Song[1], Jun Zhang[2], Kehui Zhan[1] and Dexian He[1]**

[1] Co-construction State Key Laboratory of Wheat and Maize Crop Science, Collaborative Innovation Center of Henan Grain Crops, College of Agronomy Henan Agricultural University, Henan Agricultural University, Zhengzhou, China, [2] College of Agriculture, Henan University of Sciences and Technology, Luoyang, China

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University, United States

Reviewed by:
Sean Mayes,
University of Nottingham, United Kingdom
Jindong Liu,
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, China

***Correspondence:**
Kehui Zhan
kh486@163.com
Dexian He
hed@henau.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *28 December 2020*
Accepted: *18 May 2021*
Published: *10 June 2021*

Citation:
Xu F, Chen S, Yang X, Zhou S, Wang J, Zhang Z, Huang Y, Song M, Zhang J, Zhan K and He D (2021) Genome-Wide Association Study on Root Traits Under Different Growing Environments in Wheat (Triticum aestivum L.). Front. Genet. 12:646712. doi: 10.3389/fgene.2021.646712

Plant roots are critical for water and nutrient acquisition, environmental adaptation, and yield formation. Herein, 196 wheat accessions from the Huang-Huai Wheat Region of China were collected to investigate six root traits at seedling stage under three growing environments [indoor hydroponic culture (IHC), outdoor hydroponic culture (OHC), and outdoor pot culture (OPC)] and the root dry weight (RDW) under OPC at four growth stages and four yield traits in four environments. Additionally, a genome-wide association study was performed with a Wheat 660K SNP Array. The results showed that the root traits varied most under OPC, followed by those under both OHC and IHC, and root elongation under hydroponic culture was faster than that under pot culture. Root traits under OHC might help predict those under OPC. Moreover, root traits were significantly negatively correlated with grain yield (GY) and grains per spike (GPS), positively correlated with thousand-kernel weight (TKW), and weakly correlated with number of spikes per area (SPA). Twelve stable chromosomal regions associated with the root traits were detected on chromosomes 1D, 2A, 4A, 4B, 5B, 6D, and unmapped markers. Among them, a stable chromosomal interval from 737.85 to 742.00 Mb on chromosome 4A, which regulated total root length (TRL), was identified under three growing environments. Linkage disequilibrium (LD) blocks were used to identify 27 genes related to root development. Three genes *TraesCS4A02G484200, TraesCS4A02G484800*, *TraesCS4A02G493800*, and *TraesCS4A02G493900*, are involved in cell elongation and differentiation and expressed at high levels in root tissues. Another vital co-localization interval on chromosome 5B (397.72–410.88 Mb) was associated with not only RDW under OHC and OPC but also TKW.

Keywords: wheat, root traits, yield traits, genome-wide association study (GWAS), co-localization region

Abbreviations: TRL, Total length of roots; TRV, Total root volume; ARD, Average root diameter; TRA, Total root area; NRT, Number of root tips; RDW, Root dry weight; GY, Grain yield per area; SPA, Spikes per area; GPS, Grains per spike; TKW, Thousand-kernel weight; OHC, Outdoor hydroponic culture; IHC, Indoor hydroponic culture; OPC, Outdoor pot culture; SS, Seedling stage; WP, Wintering period; JS, Jointing stage; MS, Mature stage; GWAS, Genome-wide association study; BLUE, Best linear unbiased estimation; SNP, Single- nucleotide polymorphism; LD, Linkage disequilibrium; QTL, Quantitative trait loci.

INTRODUCTION

As the population increases, the demand for wheat doubles approximately every 20 years; therefore, a continuous yield improvement in wheat production is an urgent issue. Xie et al. (2017) found that grain number per unit area, grain per spike, and plant biomass were lower for wheat lines with shorter root lengths, while lines with longer roots had the highest biomass and grain yield. The root-shoot ratio of plants improves access to residual moisture in deep soils under drought conditions, particularly in the mid- to late-stage (Reynolds et al., 2007). According to another study, the wheat root weight density was positively correlated with grain yield and water used efficiency (Man et al., 2015). Root system facilitates aerial growth and is very critical for yield (Fang et al., 2017), and its morphological characteristics include root length, surface area, volume, diameter, and number of root tips. Root length, surface area, and volume affect the spatial arrangement of roots underground, while root diameter is associated with the ability to penetrate strong soil and drought tolerance, and the root tip is the most active part of the root system (Kabir et al., 2015; Maccaferri et al., 2016; Bai et al., 2019). Previous studies have shown that optimizing root spatial configuration may lead to a higher absorption efficiency of water and nutrients, thus improving yield levels (Osmont et al., 2007; Bishopp and Lynch, 2015). Researchers have adopted indoor cultivation methods combined with digital imaging to determine the root traits of seedlings and elucidate the mechanisms regulating root growth and development through forward genetic strategies. For instance, 46 quantitative trait loci (QTLs) of root morphological traits had been identified on 17 of 21 chromosomes, excluding chromosomes 1D, 4D, 6B, and 6D, and 4.98–24.31% of the phenotypic variation had been explained (Liu et al., 2013). Ren et al. (2012) detected 35 QTLs based on a recombinant inbred line (RIL) population of Xiaoyan 54 × Beijing 411, and these QTLs were primarily distributed on chromosomes 2B, 2D, 4B, 6A, 6B, and 7B. Cao et al. (2014) mapped a root length QTL (*qTaLRO-B1*) flanked by *Xgwm210* and *Xbarc1138.2* on chromosome 2BS using near-isogenic lines (NILs) 178A and 178B. Furthermore, a few genes regulating wheat root morphology have been identified using reverse genetics methods, including *TaZFP34* (Chang et al., 2016), lateral organ boundaries (LOB) family member *TaMOR* (Li et al., 2016), and zinc finger protein *OsCYP2* (Cui et al., 2017).

Due to the enormous size (≈17 Gb) and polyploidy of the wheat genome, research on quantitative traits in hexaploid wheat has lagged behind research in other crops, especially for root traits. GWAS is a powerful tool for identifying loci that are significantly associated with target traits based on LD in panel of accessions. In recent years, with the gradual release of genomic sequences and genotyping by using sequencing data, GWAS has become a rapid and effective method for detecting related QTLs (Neumann et al., 2011). Many types of single-nucleotide polymorphism (SNP) arrays have been successfully developed to identify candidate regions for various traits in GWAS panels, including 9K (Suraj et al., 2014), 55K (Ye et al., 2019), 90K (Sukumaran et al., 2015; Wang et al., 2017), and 660K arrays (Chen et al., 2020).

Previous studies for root primarily involved QTL mapping using biparental populations (RIL, NIL, and doubled haploid (DH) populations) based on a few simple sequence repeat markers for plants grown in greenhouses (Ren et al., 2012; Cao et al., 2014; Ayalew et al., 2016). Thus, the molecular basis of many root traits remains unclear. Furthermore, some root traits have been reported to be closely related to yield traits since QTLs have been found to be co-localized (Atkinson et al., 2015; Xie et al., 2017). Therefore, 196 wheat accessions from the Huang-Huai Wheat Region of Xie et al., 2017). Therefore, 196 wheat accessions from the Huang-Huai Wheat Region of China, were used as a GWAS panel to identify root traits in the seedling stage from three growing environments, RDW from four growth stages under OPC, and yield traits in the field (**Table 1**). GWAS was conducted with the Wheat 660K SNP Array. This work may elucidate the genetic basis of wheat root traits to assist in the genetic improvement of the wheat root system.

MATERIALS AND METHODS

Plant Materials

The 196 wheat accessions used in this study included new bread lines, elite cultivars, and historical varieties from the Huang-Huai Wheat Region of China. These accessions included 158 from Henan Province, nine from Shaanxi, eight from Jiangsu, ten from Shandong, four from Hebei, four from Beijing, one from Anhui, one from Shanxi, and one land variety from Sichuan (Chinese Spring) (**Supplementary Table 1**). Seeds were provided by the Collaborative Innovation Center of Henan Grain Crops, Henan Agricultural University.

Experimental Design

Hydroponic Culture

Individuals were grown in a hydroponic system in an indoor environment (IHC). The indoor environment included day/night temperatures of 20 and 16°C, respectively, and a 16 h day length with a light intensity of 1,000 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (**Figure 1A**). Seeds were germinated in Petri dishes and grown on germination substrate for six days with sterile water. Then, seedlings with uniform growth were transferred to plastic pots containing 18 L

TABLE 1 | Plant growth traits, growing environments and growth stages in this study.

Root and yield trait	Plant growing environment and growth stage
Total length of roots (TRL)	Indoor Hydroponic culture (IHC)
Total root volume (TRV)	Outdoor Pot culture (OPC)
Average root diameter (ARD)	Outdoor Hydroponic culture (OHC)
Total root area (TRA)	Seedling stage (SS)
Number of root tips (NRT)	Wintering period (WP)
Root dry weight (RDW)	Jointing stage (JS)
Grain yield per area (GY)	Mature stage (MS)
Spikes per area (SPA)	
Grains per spike (GPS)	
Thousand-kernel weight (TKW)	



FIGURE 1 | Root experiments on 196 accessions under three growing environments. **(A)** Indoor hydroponic culture (IHC); **(B)** outdoor hydroponic culture (OHC); and **(C)** outdoor pot culture (OPC).

of nutrient solution, and fixed on a foam board with sponge (Length 5 cm; width 1 cm). The nutrient solution composition was the same as that described by Ren et al. (2012) with minor modifications. The nutrient solution was refreshed every 3 days, with the pH maintained at 6.0. Seedlings from fifty accessions were planted in one box, and 12 boxes were used for each hydroponic experiment. The 196 wheat accessions were evaluated in three replicates with two plants per replicate in each hydroponic experiment.

Another hydroponic experiment was arranged outdoors (OHC) with three replicates for three consecutive years, from October 20 to November 20 in 2016, 2017, and 2018 at Zhengzhou (34.7°N, 113.6°E), Henan, China (**Figure 1B**). The planting method was the same as that described above. A full box of nutrient solution was supplied daily, and a movable shelter was placed above the seedlings, especially on rainy days. All boxes for each replicate were fixed together with ropes to prevent outside interference (strong winds and, small animals).

Pot Experiments

All accessions were planted in pots buried in a field in 2017–2018 and 2018–2019 at Zhongmou (34.7°N, 114.0°E), Henan, China to easily extract the complete root system. The 196 wheat accessions were planted in a randomized design with two replicates for each stage. Each pot (height, 15 cm; diameter, 12 cm) was filled with 3 kg of sieved tillage soil, and the top edge of each pot was on the same plane as the ground surface (**Figure 1C**). The tillage soil was described as "white-sand loam" with a classification of sandy soil that contained fertilizer consisting of nitrogen at 2,295 kg·hm^{-1}, phosphorus at 86.445 kg·hm^{-1} and potassium at 302.49 kg·hm^{-1}. Before planting, seeds were germinated as described for hydroponic culture experiments. Uniformly germinating seeds were placed in each pot. Ten days after sowing, each plot was thinned to two seedlings equivalent to 1,800,000 seedlings per hectare in the field. To facilitate management, 1,568 pots representing all accessions were distributed over 20 blocks (length, 150 cm; width, 160 cm), with each block containing 80 pots. A border was planted to prevent the use of plants from around the blocks due to possible edge effects. The plants were regularly watered to 70–80% of field capacity at a 10-day interval after sowing. At jointing stage, we fertilized plants with a total nitrogen content of 3378.00 kg·hm^{-1}, available phosphorus content of 90.00 kg·hm^{-1} and available potassium content of 303.75 kg·hm^{-1}. Other field management procedures were performed in accordance with local agronomic practices. Wheat plants cultivated under OPC were as normal as those grown in the field at different growth stages, with each plant having tillers (2–5) and ears (1–3) (**Supplementary Figure 1**).

Root Phenotype Evaluation

For the hydroponic culture experiments, roots were rinsed with sterile deionized water 30 days after sowing before they were measured. In the pot experiments, the roots were removed from the soil at 30 days (seedling stage, SS), 60 days (wintering period, with a daily mean temperature < 2°C, WP), 150 days (jointing stage, 50% of main culms grew 1.5–2 cm above the ground, JS), and 220 days (mature stage, grain size was normal and water content was less than 20%, MS) after sowing, and rinsed with fresh water before being measured.

Five root system traits, including total root length (TRL), total root volume (TRV), number of root tips (NRT), total root area (TRA), and average root diameter (ARD), were scanned and analyzed with a Win-RHIZO system (Canada Regent Instruments, LA6400XL) at seedling stage. After being scanned, the roots were dried at 80°C for 48 h, and root dry weight (RDW) was measured using an analytical balance (Germany Sartorius, QUINTIX224-ICN). Only RDW was measured at WP, JS and MS under OPC conditions.

Field Experiment of Yield Traits

Field experiments were performed to measure yield traits in Zhengzhou (34.7°N, 113.7°E) in 2014–2015 and 2015–2016, in Shangqiu (33.4°N, 115.4°E) in 2014–2015, and in Zhumadian (33.0°N, 114.1°E) in 2014–2015. The 196 accessions were planted in a randomized block design with two replicates. Each accession was planted in one plot containing four rows with 23-cm of row spacing, and each row was 1.5 m in length with 110 seeds in a north-south orientation. Twenty adjacent spikes in the middle of each plot were randomly selected to measure number of grains per spike (GPS) and then threshed and weighted to calculate thousand-kernel weight (TKW), the middle two rows (measuring 1.0 m in length) were examined to calculate grain yield (GY) and spikes per unit area (SPA).

Statistical Analysis

Joint variance analysis, estimation of broad-sense heritability (H_B^2), and BLUE values of phenotypic values obtained in each environment using in linear mixed models (LMMs) were

performed using IciMapping v4.0 software (Valassi and Chierici, 2014; Lei et al., 2015). Other analyses, including descriptive statistics and correlation and canonical correlation analysis, were conducted using SAS V9.4 software.

The mean values obtained from each environment were used to analyzing the differences in root traits through a comprehensive evaluation of the spatial configuration and *t*-tests. Differences in root traits under different growing environments were compared between IHC and OHC, OHC and OPC, and IHC and OPC. Each root trait was converted into *u*-values using the standardized normal distribution method to facilitate a comprehensive evaluation (Chen et al., 2020).

660K SNP Genotyping, GWAS, and Prediction of Candidate Genes

All 196 accessions surveyed here were genotyped using the Wheat 660K SNP Array, which was provided by Beijing Boao Crystal Code Biotechnology Co., Ltd.[1]. Quality control and filtering of raw data were performed with a missing rate threshold of 20% and a minor allele frequency (MAF) of 5% using TASSEL v5.0 software (Bradbury et al., 2007). Chen et al. (2020) used 197 wheat accessions including the 196 wheat accessions analyzed in this study, and used same as 660K SNP Array for GWAS. Genetic diversity, population structure, and LD analyses were performed using the methods reported by Chen et al. (2020) with minor revisions. Previous reports indicated that the 196 accessions could be divided into two subpopulations, which are largely consistent with the *K*-values, principal component analysis (PCA)-matrix and quantile (Q)-matrix in the present study (Jakobsson and Rosenberg, 2007; Earl and Vonholdt, 2012). Firstly, GWAS was formally performed using five models (generalized linear models (GLMs): Q and PCA; mixed linear models (MLMs): K, PCA + K, and Q + K) to obtain an optimized model using the quantile-quantile (Q-Q) plot of each trait under IHC. The best model was considered when the actual $-\log_{10}$ (*p*-value) was closest to the expected -$\log_{10}$ (*p*) value (Liu et al., 2015). An MLM correcting for both the Q-matrix and K-matrix was confirmed to reduce the population structure and relative kinship errors. Then, we used an MLM (Q + K) to analyze associations among single environment and BLUE values for each trait with Tassel v5.0 software. The threshold for the *p*-value should be set to $\log_{10}$ (1/n) when the number of SNP markers was n. However, the Bonferroni-Holm correction for multiple testing (alpha = 0.05) was too conservative and no significant SNPs were detected. To combine the GWAS results in all of the growing environments Markers with an adjusted–$\log_{10}$ (*P*-value) $\geq$ 3.5 were regarded as significantly associated (Beyer et al., 2019). Lastly, Manhattan and Q-Q plots were generated using the CMplot package in R[2].

LD and haplotype block structure were analyzed using Haploview 4.2 software to obtain physical positions of the identified stable regions. IWGSCv1.0 Chinese Spring[3] was used to retrieve candidate genes ID and sequence located in the haplotype block. Gene sequences were used to perform a BLAST search for orthologous genes in *Arabidopsis* and *Oryza sativa* L. on the NCBI website[4]. Expression pattern of candidate genes were determined by blasting the predicted coding sequences (CDS) to WheatEXP[5] with an *E*-value cutoff 1e-10 and covering five different wheat tissues (spikes, roots, leaves, grain, and stems) each sampled at three different developmental stages.

RESULTS

Phenotypic Variations of Root Traits From Different Growing Environments

The frequencies of most root traits in plants from different growing environments were normally distributed (**Supplementary Figure 2**). Joint variance analysis indicated very significant genotype (variety) and genotype × environment (G × E) effects ($p \leq 0.01$) (**Table 2**). Root traits under OPC exhibited the greatest variation among the growing environments, followed by those of plants grown in OHC and IHC in the seedling period. In addition, the average *u*-value for each accession was used to evaluate its root morphology under different growing environments to obtain extreme root materials (**Supplementary Table 1**). Under IHC, the coefficients of variation (CVs) for TRT, RDW, TRV, TRL, and TRA did not differ substantially and were higher than those of TAD. Yujiao 5, Fanmai 5 and Yandian 9433 displayed poorer root development with lower *u*-values ($u < -1.31$), and Pingan 8, Yunong 416 and Luomai 21 exhibited better root development with higher *u*-values ($u > 1.34$). Under OHC, the CVs for all root traits ranged from 5.57 to 17.21%, and the heritability (H_B^2) ranged from 45.86 to 63.16%. Yumai 2, Yumai 8, and Luomai 23 had poorer root development with lower *u*-values ($u < -1.22$), and Sumai 3, Yake 028 and Nanda 2419 had better root development with higher *u*-values ($u > 1.47$). Under OPC, the CVs for all root traits ranged from 8.80 to 38.23%, and H_B^2 ranged from 42.95 to 51.80%. Kaimai 18, Zhongchuang 805 and San 160 had poorer root development with lower *u*-values ($u < -1.22$), and Zhengpinmai 8, Luohan 6 and Yanke 028 had better root development with higher *u*-values ($u > 1.47$).

As shown in **Table 2**, the RDWs for wheat in pot experiments increased non-linearly over time, ranging from 44.0 (seedling) to 448.5 mg (mature), and the increase at jointing stage was the largest. In terms of individual accession, Zhengpinmai 8, Luohan 6 and Yanke 028 had higher RDWs at seedling stage, Zhoumai 13, Zhengyumai 043 and Chinese Spring had higher RDWs at wintering period, Zhengmai 379, Yunong 9901 and Jinmai 47 had higher RDWs at jointing stage, and Luohan 3, Zhengyumai 043 and Yumai 51 had higher RDWs at mature stage. By contrast, Kaimai 18, Zhongchuang 805 and San 160 had lower RDWs at

[1]http://wheat.pw.usda.gov/ggpages/topics/Wheat_660_SNP_array_developed_by_CAAS.pdf

[2]https://github.com/YinLiLin/R-CMplot

[3]http://plants.ensembl.org/hmmer/index.html

[4]https://www.ncbi.nlm.nih.gov/nuccore/

[5]https://wheat.pw.usda.gov/WheatExp/

TABLE 2 | Phenotypic variation and heritability for the root traits of the 196 wheat accessions under different growing environments.

	Traits	Range	mean	CV (%)	F (G)	F (G*E)	H_B^2 (%)
IHC	TRL (cm)	310.3–806.4	524.8 ± 57.3	22.6	4.4**	–	70.66
	TRA (cm^2)	40.2–125.9	77.4 ± 16.8	23.3	5.3**	–	66.81
	ARD (mm)	0.39–0.62	0.47 ± 0.03	8.3	8.4**	–	78.77
	TRV (cm^3)	0.4–1.6	0.90 ± 0.60	26.8	3.7**	–	66.45
	NRT (count)	196.3–847.0	471.1 ± 15.6	31.2	8.6**	–	79.19
	RDW (mg)	9.8–37.5	21.7 ± 27.6	27.6	3.4**	–	54.24
OHC	TRL (cm)	564.7–1271.3	833.3 ± 14.2	14.2	11.4**	4.6**	63.2
	TRA (cm^2)	77.5–191.1	112.0 ± 17.7	14.4	9.2**	4.9**	59.2
	ARD (mm)	0.32–0.52	0.43 ± 0.1	5.8	7.1**	6.1**	45.9
	TRV (cm^3)	0.6–2.6	1.6 ± 1.3	17.0	6.1**	4.7**	50.8
	NRT (count)	335.8–923.7	606.5 ± 37.5	17.2	9.1**	5.0**	53.9
	RDW (mg)	25.2–53.8	37.8 ± 12.3	12.3	81.7**	72.7**	54.0
OPC-SS	TRL (cm)	224.0–1008.0	524.6 ± 48.9	27.0	7.1**	6.8**	48.0
	TRA (cm^2)	44.1–168.8	95.3 ± 17.3	24.4	6.7**	5.8**	49.9
	ARD (mm)	0.42–0.80	0.58 ± 0.1	8.1	5.9**	6.4**	43.0
	TRV (cm^3)	0.6–3.0	1.5 ± 0.4	28.3	6.6**	5.1**	51.8
	NRT (count)	214.3–1198.8	589.2 ± 45.6	38.2	8.3**	8.1**	47.1
	RDW (mg)	22.6–71.1	44.0 ± 24.4	24.4	131.2**	115.7**	51.0
OPC-WP	RDW (mg)	72.5–185.0	120.0 ± 19.1	19.1	243.6**	200.5**	55.9
OPC-JS	RDW (mg)	200.1–572.8	383.0 ± 45.7	22.7	7.4**	5.1**	57.3
OPC-MS	RDW (mg)	150.2–922.9	448.5 ± 34.8	34.8	72.3**	57.8**	55.0

*TRL, total root length; TRV, total root volume; NRT, number of root tips; TRA, total root area; ARD, average root diameter; RDW, root dry weight; IHC, indoor hydroponic culture; OHC, outdoor hydroponic culture; OPC, outdoor pot culture; SS, seedling stage; WP, wintering period; JS, jointing stage; MS, mature stage; G, genotype; E, environment; H_B^2, broad sense heritability;–, indicated no variance component; **, Significant at $p \leq 0.01$.*

seedling stage, Pu 2056, San 160 and Jinan 17 had lower RDWs at wintering period, Luomai 24, Yumai 52 and Fanmai 11 had lower RDWs at jointing stage, and Yunnong 949, Luomai 24 and Yujiao 5 had lower RDWs at mature stage. In summary, some varieties developed rapidly at earlier stages, while others developed rapidly at later stages.

Differences in Root Traits of Wheat From Different Growing Environments

Large variations in the same root traits were found among different growing environments (**Supplementary Table 2**). Compared to IHC, TRL, TRA, TRV, NRT, and RDW under OHC had a very significant increase but not ARD. Under OPC, TRA, ARD, TRV, NRT, and RDW had a very significant increase, but not TRL. Compared to OHC, TRL, TRA, TRV, and NRT under OPC were reduced very significantly, while ARD and RDW were very significantly increased. Root development in wheat cultivated under IHC was slower than wheat cultivated under OHC and OPC, and root elongation under hydroponic culture was faster than that under pot culture. Compared with that of IHC, root system under OPC showed a higher canonical correlation with that under OHC, and simple correlation coefficients of root traits except for TRL, were very significant between OPC and OHC and larger than those between OPC and IHC (**Table 3**). None of the correlation coefficients of RDW between seedling stage under IHC and each stage under OPC were significant. However, RDW at seedling stage under OHC exhibited a very significant correlation with that of wintering period under OPC, and RDW at seedling stage under OPC had a very significant correlation with that of wintering period and mature stage under OPC (**Table 4**). Based on these results, root traits under OHC might provide a good prediction for those under OPC.

Correlations Between Root Traits

Correlations between root traits were identical in the different growing environments (**Supplementary Table 3**). Except for

TABLE 3 | Correlation coefficients of root traits between the growing environments.

Traits	IHC-vs-OHC	IHC-vs-OPC	OHC-vs-OPC
TRL	0.335**	0.059	0.292**
TRA	0.256**	0.066	0.377***
ARD	0.189*	0.063	0.093
TRV	0.162*	0.103	0.337***
NRT	0.271**	0.078	0.323***
RDW	0.124	0.016	0.378***
V	0.421	0.385	0.515

*V, Canonical correlation coefficient; *, Significant at $p \leq 0.05$; **, Significant at $p \leq 0.01$; ***, Significant at $p < 0.001$.*

TABLE 4 | Correlation coefficients of RDW between SS and other stages.

Environments	OPC		
	WP	JS	MS
SS under IHC	−0.008	−0.068	0.108
SS under OHC	0.185**	−0.036	0.128
SS under OPC	0.335**	0.130	0.168*

**, Significant at $p \leq 0.05$; **, Significant at $p \leq 0.01$.*

ARD, very significant positive correlations were observed between other root traits. Among them, the correlation coefficients between TRL and TRA and between TRA and TRV were very large (approximately 0.9), and the others ranged from 0.367 to 0.834. ARD had a lower correlation with TRL, NRT, and RDW, and a very significant positive correlation with TRV.

Correlations of Root Traits With Yield Traits

Most root traits at seedling stage under different growing environments were negatively correlated with GY, SPA, and GPS, and positively correlated with TKW. Specifically, NRT under three growing environments was significantly negatively correlated with GY. ARD and TRA under OHC were significantly negatively correlated with GPS, and RDW under OPC was also significantly negatively correlated with GPS. TRA and TRV under three growing environments were significantly positively correlated with TKW. Based on the canonical correlation analysis indicated that root traits under OHC and OPC had closer relationships with yield traits than those under IHC. Correlations of RDW at four growth stages under OPC with yield traits showed large differences. RDW at seedling stage and wintering period was significantly positively correlated with TKW, and all correlation coefficients between RDW and grain traits were low at jointing and mature stages (**Table 5**).

TABLE 5 | Correlation coefficients between root traits and yield traits.

Environments	Traits	SPA	GPS	TKW	GY
IHC	TRL	0.016	−0.115	0.190**	−0.02
	TRA	−0.015	−0.113	0.194**	−0.033
	ARD	−0.084	0.009	0.047	−0.023
	TRV	−0.032	−0.104	0.175*	−0.038
	NRT	0.029	0.006	−0.067	−0.189**
	RDW	−0.066	−0.059	0.078	−0.063
	V	0.294			
OHC	TRL	−0.019	−0.09	0.178*	−0.104
	TRA	0.031	−0.137	0.202**	−0.116
	ARD	0.092	−0.145*	0.037	0.065
	TRV	0.064	−0.15*	0.199**	−0.088
	NRT	−0.104	−0.097	0.123	−0.183*
	RDW	−0.078	−0.061	0.134	−0.124
	V	0.407***			
OPC	TRL	0.009	−0.115	0.135	−0.146*
	TRA	−0.033	−0.128	0.198**	−0.157*
	ARD	−0.048	−0.039	0.040	−0.052
	TRV	−0.063	−0.115	0.213**	−0.123
	NRT	−0.017	−0.070	0.055	−0.150*
	RDW	−0.039	−0.141*	0.158*	−0.138
	V	0.406***			
WP at OPC	RDW	−0.069	−0.04	0.163*	−0.009
JS at OPC	RDW	−0.039	−0.02	−0.005	−0.037
MS at OPC	RDW	−0.008	0.108	−0.074	−0.066

*GY, grain yield per area; SPA, spikes per area; GPS, grains per spike; TKW, thousand- kernel weight. *, Significant at $p \leq 0.05$; **, Significant at $p \leq 0.01$; ***, Significant at $p \leq 0.001$.*

GWAS of Wheat Root Traits

After filtering, 390,136 SNP markers were available for GWAS with the optimal mixed linear model of the K + Q matrix. The GWAS panel was comprised of two subpopulations by STRUCTURE software and PCA. The LD decay distance was 15, 10, and 20 Mb in the A, B, D subgenome by Chen et al. (2020) (**Supplementary Figure 3**). A total of 1,329 significantly associated loci were detected for root traits from different environments, and distributed across all 21 chromosomes with an R^2 range of 6.49–18.01% (**Supplementary Table 4**). In view of their respective growing environments, OHC was associated with the most SNP loci, followed by OPC and IHC. Using BLUE values of root traits, 484 SNPs were detected, with an R^2 range of 6.38–19.43%. SNP loci of NRT were the most frequently identified, followed by those of TRL, ARD, TRV, TRA, and RDW (**Supplementary Table 4** and **Supplementary Figure 4**).

Some SNPs for the same trait were different, but they were detected from two or three growing environments and were located at close physical intervals. Taking 10 Mb of LD decay distance as a confidence interval (Liu et al., 2017; Chen et al., 2020), a total of 12 co-localization regions were detected from two or more growing environments, which distributed on chromosomes 1D, 2A, 4A, 4B, 5B, 6D, and unmapped markers, respectively. Among them, three co-localization regions were found for TRL, two for TRA, two for TAD, one for TRV, three for NRT, and one for RDW (**Table 6**). Three co-localization regions associated with TRL were located on chromosomes 1D, 4A, and 5B, explaining 9.10–10.22% of the phenotypic variation (**Figure 2** and **Supplementary Table 5**). In particular, one co-localization region for TRL was found within the 737.85–742.00 Mb on chromosome 4A with 1 SNP each in IHC, OHC, and OPC. Two co-localization regions on chromosomes 1D and 4B for TRA were detected simultaneously in two growing environments, explaining 7.28–10.33% of the phenotypic variation. Two co-localization regions on chromosomes 5B and 6D for TAD

TABLE 6 | Summary of co-localization regions for root traits by GWAS.

Traits	SNP marker (Mb)		IHC	OHC	OPC
TRL	1D	464.37–465.44	2	4	
	4A	737.85–742.00	1	1	1
	5B	696.34–701.69		1	1
TRA	1D	484.26–486.74	1	2	
	4B	657.45–658.78		1	1
TAD	5B	652.10–665.20	1		1
	6D	463.54–472.27	1	12	
TRV	5D	543.09–547.80	3	60	
NRT	2A	769.56–771.34	2	1	
	5B	571.25–572.14	1	1	
	Un	71.06–77.40	1	1	
RDW	5B	397.72–410.88		1	5

Un, unmapped markers.

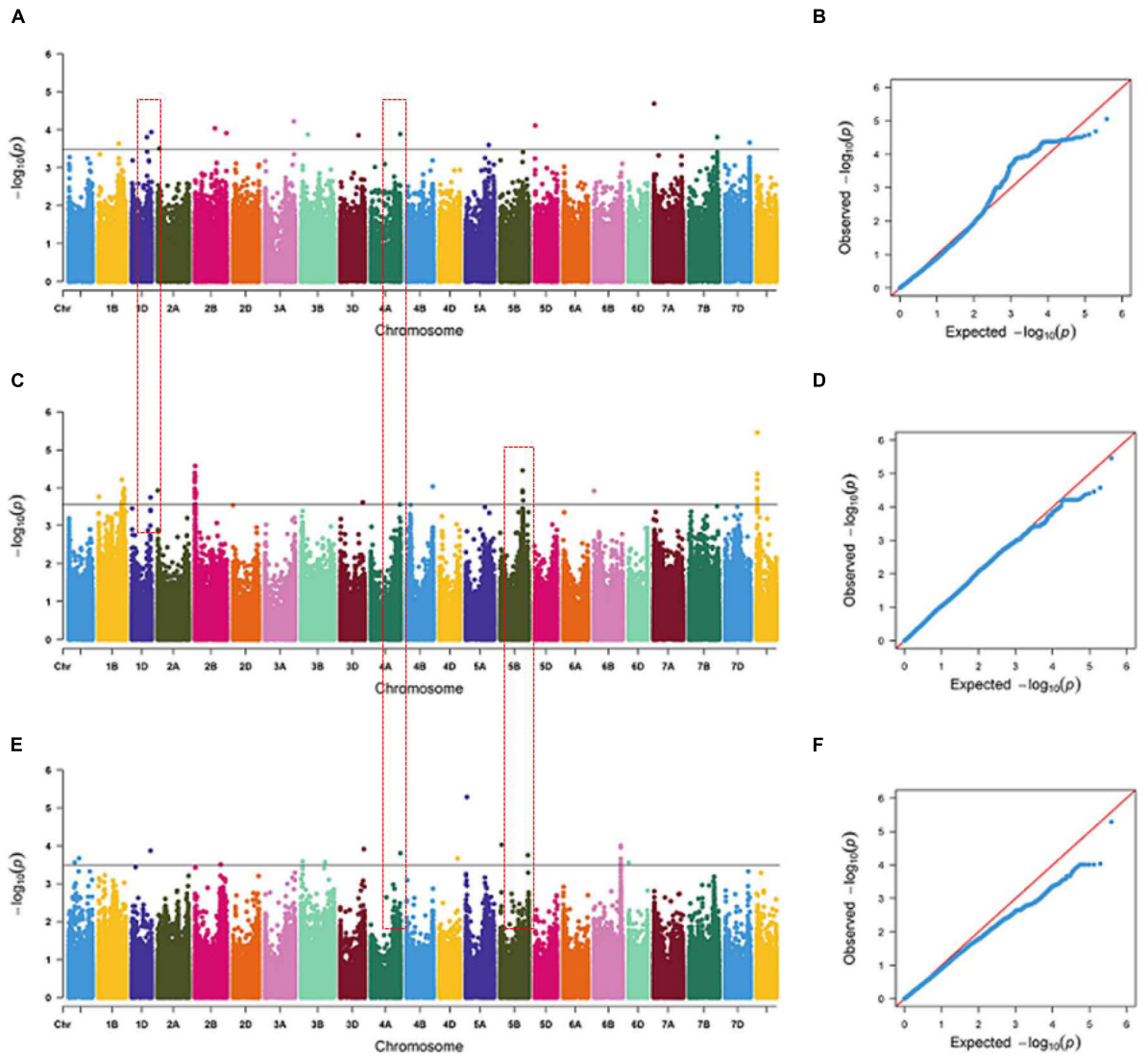


FIGURE 2 | Manhattan and Q-Q plots for total root length (TRL) under indoor hydroponic culture **(A,B)**, outdoor hydroponic culture **(C,D)**, and outdoor pot culture **(E,F)**. The chromosomes on the *x*-axis and $-\log_{10}(p)$ values on the *y*-axis were all based on BLUE values. The gray line indicates the threshold of significance at $-\log_{10}(p) = 3.5$. The red text box represents the co-localization region among different growing environments.

were detected simultaneously in two growing environments, explaining 8.37–9.15% of the phenotypic variation. One co-localization region for TRV in the 543.09–547.81 Mb region on chromosome 5D, with 3 SNPs in IHC and 60 SNPs in OPC, explaining 7.04–11.82% of the phenotypic variation. Three co-localization regions on chromosomes 2A, 5B and unmapped markers for NRT were detected simultaneously in two growing environments, explaining 7.08–10.45% of the phenotypic variation. One co-localization region for RDW was found in the 397.72–410.88 Mb region on chromosome 5B, with 1 SNP in OHC and 5 SNPs in OPC, explaining 6.88–10.30% of the phenotypic variation. Of the 12 co-localization regions for root traits identified, QTL on chromosomes 1D, 4B, and 5D were reported previously (Ma et al., 2008; Kabir et al., 2015; Xie et al., 2017), while nine co-localization regions were first detected here.

Because the root traits were closely associated with one another, some SNPs were simultaneously detected in multiple traits. For example, 9, 29, and 21 SNPs were significantly associated with two or more root traits under IHC, OHC, and OPC, respectively, and they were distributed on all chromosomes except for 1A. Moreover, five SNPs associated with three root traits were detected under OHC, and only one was found under OPC (**Supplementary Table 6**).

Haploblock Analysis and Candidate Genes of TRL

LD and haploblock analyses were performed for the co-localization interval 737.85–742.00 Mb on chromosome 4A, which was associated with TRL across three growing

environments. And accessions with superior alleles at three growing environments showed longer TRL than others. At r^2 = 0.1, two blocks were clearly separated along this 5-Mb region (**Supplementary Figure 5**). Further analysis showed that these two blocks were located at the physical positions 738.86–739.51 and 742.17–742.33 Mb, which contained 18 and nine genes, respectively. Then, 24 and eight orthologous genes in other species were identified in *Oryza sativa* L. and *Arabidopsis*, respectively (**Supplementary Table 7**). These genes encode proteins involved in growth development, carbon metabolism, energy metabolism, and plant resistance. Among them, *TraesCS4A02G484200, TraesCS4A02G484800*, *TraesCS4A02G493800*, and *TraesCS4A02G493900* showed high expression levels in roots at three growth stages (**Figure 3** and **Supplementary Table 8**). *TraesCS4A02G484200* encodes a yippee domain protein, *TraesCS4A02G493900* encodes a member of the leucine-rich repeat domain superfamily, *TraesCS4A02G493800* encodes a member of the F-box-like domain superfamily, and *TraesCS4A02G484800* encodes the glycoside hydrolase family 32 protein. These genes might be involved in regulating root growth and development.

DISCUSSION

Influence of Different Growing Environments on Wheat Root Development

The methods used to acquire root system data for crops have always been a focus in root studies. Currently, approaches are primarily divided into two categories: field and glasshouse identification (Neumann et al., 2009). Lopes and Reynolds (2010) used soil core sampling in the field to confirm a relationship between deep roots and yield across genetic differences in eight wheat sister lines under drought stress. Li et al. (2019) showed that deep-rooted accessions had lower canopy temperatures and higher grain yields than those with shallow roots as indicated by soil column testing. Field identification is labor intensive and time consuming, so it is not suitable for large-population root identification. Researchers have adopted hydroponic culture (Ayalew et al., 2016), gel-filled chambers (Christopher et al., 2013), germination bags (Robertson et al., 1979), and pots (Cao et al., 2014), to identify the seedling root traits in an artificial climate chamber and to find some QTLs or genes important for root development to overcome these challenges. Glasshouse identification may not be reflective of root development in a complex field environment, and it is difficult to apply to the study of root system in growth stages (Fang et al., 2009). Bai et al. (2019) reported a weak that correlation between the laboratory-based root length and field-based root depth across 2 years.

In this study, compared with IHC and OHC, root traits detection under OPC had a higher cost with respect to labor or time, and this system easily caused greater damage to the root system. Root traits under OHC had a very significant correlation with that under OPC and might provide them a good prediction. Considering the integrity and feasibility of root trait collection, hydroponic experiments in an outdoor or climate chamber with a suitable growing environment can be better used to identify root system in a large population.

Relationships Between Root Traits and Yield Traits in Wheat

In wheat, the root system consists of seminal and nodal roots, seminal roots primarily absorb water and nutrients from deep soil, and nodal roots are the primary component of root system after the three-leaf stage (Osmont et al., 2007; Steinemann et al., 2015). An excessive root mass not only prolongs the growth

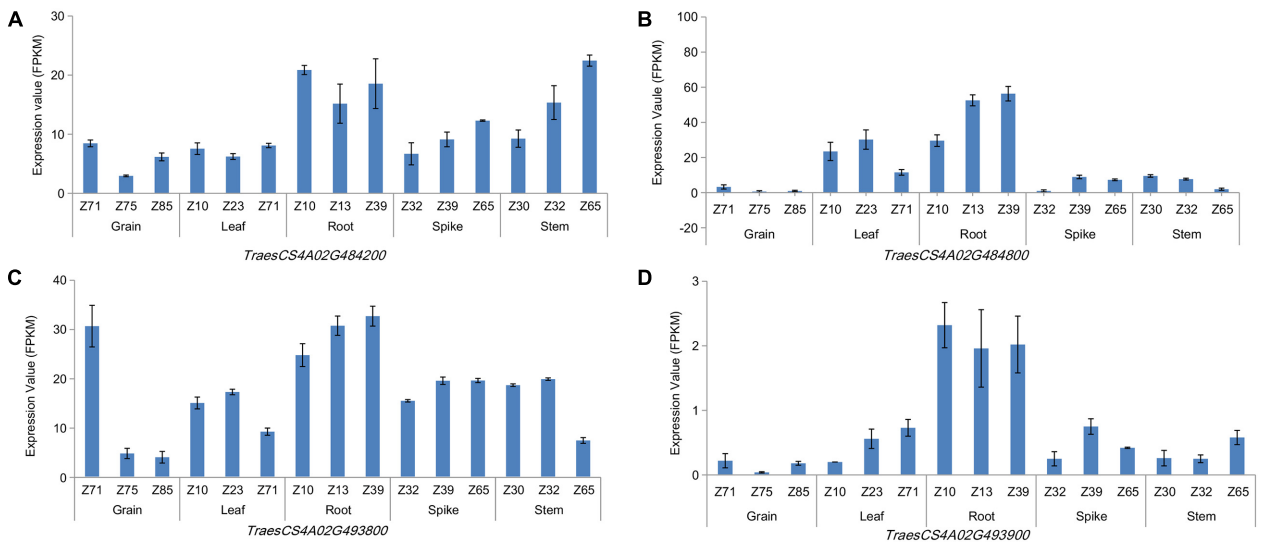


FIGURE 3 | Expression patterns of genes in different wheat organs at three growing stages. *TraesCS4A02G484800* **(A)**, *Traes*CS4A02G494200 **(B)**, TraesCS4A02G 493800 **(C)**, and TraesCS4A02G493900 **(D)**.

period of a crop but also leads to strong competition for effective assimilation within plant (limited sources) (González et al., 2011). Passioura (1983) showed that an excessive outgrowth of roots caused inefficient carbohydrate consumption and yield loss. However, some studies revealed that seminal root number and TRL at seedling stage were positively correlated with GY (Maccaferri et al., 2016; Zheng et al., 2019), but negatively correlated or not correlated with TKW (Neumann et al., 2009). Atta et al. (2013) found that root traits had weak positive correlations with GPS, SPA, GY, and water use efficiency. Reducing nodal root mass can improve water use efficiency after flowering and thus increase harvest index (Ma et al., 2008). In this study, root traits were negatively correlated with SPA, GPS, and GY, and positively correlated with TKW. In most previous studies, the culture cycle was 10–14 days and the roots only consisted of seminal roots without nodal roots, but in our study, the culture cycle was 30 days after sowing with 2–3 nodal roots. Culture time of root system may be the main reason for the different results.

According to previous studies, the root system was closely related to yield from a genetic perspective. Maccaferri et al. (2016) found that some QTLs that not only improved yield traits (GY and TKW) but also increased root number and TRL in tetraploid wheat. Zheng et al. (2019) also found that *QTrl.saw-2D.2* located on chromosome 2D was a major QTL controlling TRL, and was also associated with TKW and kernel number per spike. Atkinson et al. (2015) detected two QTLs on chromosomes 2B and 7Dthat simultaneously controlled root traits at seedling stage, GY and nitrogen uptake. In rice, a QTL associated with an increased root length of 9.6 cm was introduced into a dryland variety, which increased GY by 200 kg·hm^{-2} (Steele et al., 2013).

In this study, five SNPs, namely, *AX-111649489, AX-110419051, AX-109835270, AX-110550045,* and *AX-111547988* on chromosome 5B from 397.72 to 410.88 Mb, were significantly associated with RDW (**Figure 4A**). These five SNPs constituted a haploblock that can be divided into two haplotypes (*Hap1* and *Hap2*; **Figure 4B**). The frequency of the superior *Hap2* was 25.1% in the GWAS panel, showing that it has not been used in many modern wheat varieties. *Hap2* was proposed to significantly increase RDW at every stage, indicating its vital role in supporting root growth (**Figure 4C**). Furthermore, varieties harboring these elite alleles, including Yanke 028, Fengyou 6, Yamai 1, and Mengmai 023, contributed to 49.59 g of TKW on average, which was higher than that of other accessions (42.76 g). This region for RDW may be a promising locus to use and evaluate in future studies.

Prediction of Candidate Genes for Wheat Root Traits

In this study, 12 co-localization regions on chromosomes 1D, 2A, 4A, 4B, 5B, 6D, and unmapped markers were found in wheat

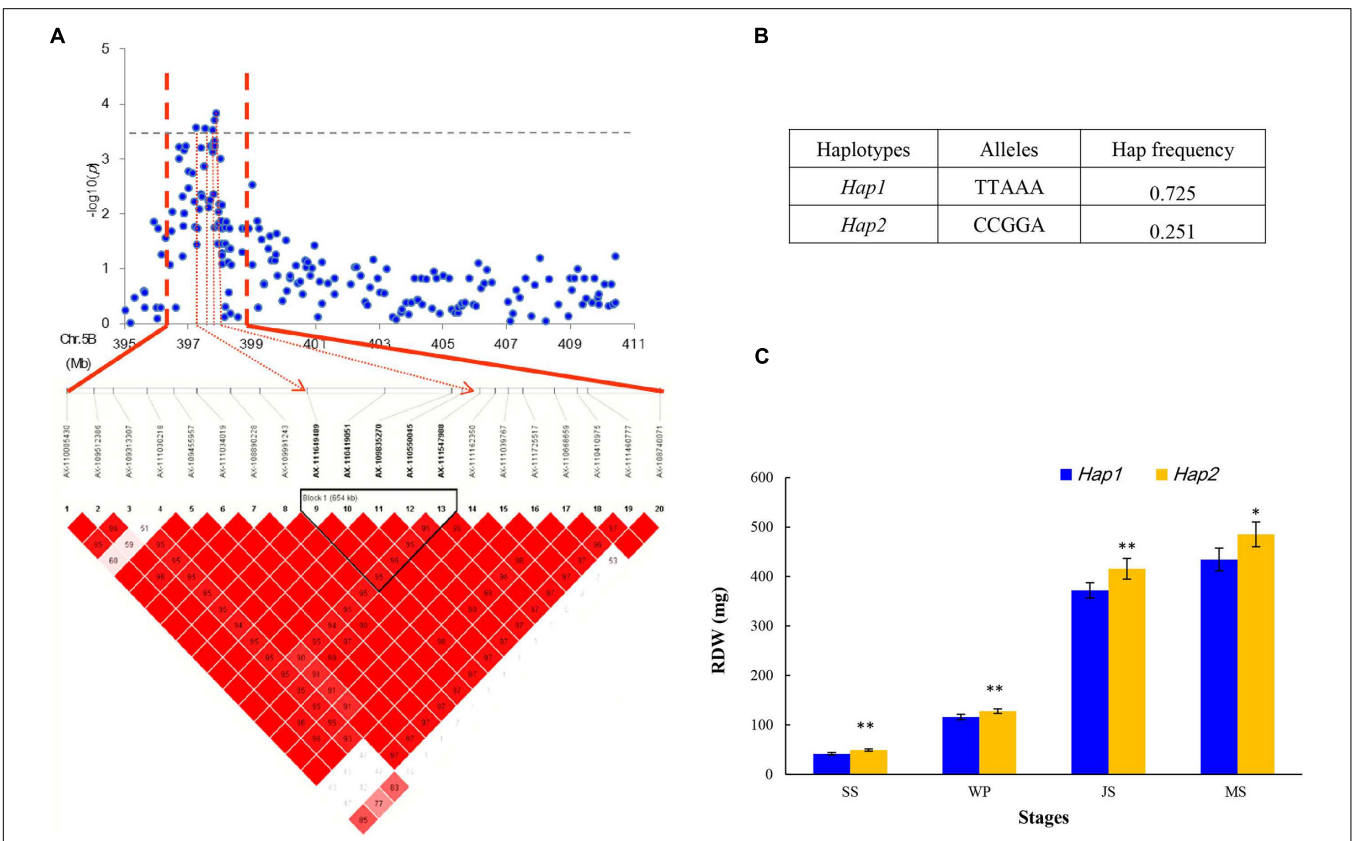


FIGURE 4 | Haplotype analysis of SNPs in the stable region at 397.72–410.88 Mb on chromosome 5B. **(A)** Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peak on chromosome 5B; **(B)** two haplotypes with different alleles; and **(C)** phenotypic effects of different haplotypes.

from different growing environments. Alignment to the reference wheat genome (Chinese Spring) revealed that three regions were located at similar positions or were consistent with those found in previous studies. In the 484.26–486.74 Mb region on chromosome 1D, which was 9 Mb from *Xbarc62_1D* (493.48 Mb) (Ibrahim et al., 2012; Sun et al., 2013), one and two SNPs were significantly associated with TRA under IHC and OHC, respectively. In the 657.45–658.78 Mb region on chromosome 4B, which was approximately 16 Mb from *Xwmc413_4B* (641.40 Mb), one SNP each in OHC and OPC was significantly associated with TRA. And *Xwmc413_4B* was associated with not only root area but also root length, root volume, and root dry weight (Maccaferri et al., 2016). Other pleiotropic SNPs associated with root traits were also detected in previous studies (Ren et al., 2012; Li et al., 2019). We predicted that these pleiotropic SNPs might contain tightly linked genes or encode transcription factors regulating multiple traits.

A co-localization region identified in all three growing environments regulating TRL was located within 737.85–742.00 Mb on chromosome 4A, which has not been reported. 27 candidate genes were detected in this interval, and four of these genes (*TraesCS4A02G484200*, *TraesCS4A02G484800*, *TraesCS4A02G493800,* and *TraesCS4A02G493900*) showed high expression levels and regulated root length based on their annotations and expression patterns. *TraesCS4A02G484200* encoded a protein containing zinc-finger-like metal binding domains (yippee domain). Zinc finger protein *OsCYP2* controlled lateral root development in response to auxin, whereas *OsCYP2* silencing leaded to significantly fewer lateral roots than in the wild-type rice plants (Cui et al., 2017). *TraesCS4A02G484100, TraesCS4A02G493800,* and *TraesCS4A02G494000*, which encode members of the F-box-like protein superfamily, were notably associated with E3 ubiquitin-mediated protein degradation in response to abiotic stress (Hua and Vierstra, 2011). *TaFBA1* encodes a homologous F-box protein in wheat that was translated into tobacco, and its overexpression improved heat tolerance and lengthened the root (Li et al., 2018). *TraesCS4A02G493900* and *TraesCS4A02G494100* encoded leucine-rich repeat domain, which was the important part of the leucine-rich repeat receptor-like kinase (LRR-RLK). Root growth is determined by meristem cell elongation and differentiation (Dello Ioio et al., 2007). One of the LRR-RLK genes contained 21 leucine-rich repeat domains, which were expressed in roots and rosettes of *Arabidopsis* and were involved in cellular signaling in plants (Walker, 1993). Another of the LRR-RLK genes (*OsSERK1*) with high expression levels was detected in rice calluses during somatic embryogenesis, and it promoted cell differentiation and mediated defense signal transduction (Hu et al., 2005). *TraesCS4A02G484800* and *TraesCS4A02G485000* encode the glycoside hydrolase family 32, leucine-rich repeat domain superfamily, and concanavalin A-like lectin/glucanase domain superfamily members. *At1G12240* in *Arabidopsis* encodes glycoside hydrolase family 32, which is involved in the decomposition and transformation of sucrose, which could promote hypocotyl elongation (Sergeeva et al., 2006). Similarly, *OsINV3* in rice is a gene that encodes glycoside hydrolase family proteins. Electron microscopy revealed that the panicles of *OsINV3* mutant plants displayed small and few cells on the inner and outer membranes, indicating that *OsINV3* plays an important role in cell expansion (Morey et al., 2018). *TraesCS4A02G484800* is homologous to *OsINV3* and *AtG12240*. These genes might regulate the elongation and formation of root cells through the zinc-finger domain, F-box-like proteins, leucine-rich repeat domains and glycoside hydrolases can thereby affecting root development.

CONCLUSION

Correlation coefficients of root traits between OHC and OPC were higher than those for either IHC vs. OPC or IHC vs. OHC. And, RDW between OHC and OPC was closely correlated at earlier stages (seedling and wintering). Thus, OHC may be considered a rapid root identification method worthy of promotion. Among the 196 wheat accessions, Luomai 23, Zhongchuang 805, Yandian 9433, Kaimai 18, and Xinmai 19 had poor root systems, while Yumai 54, Luohan 6, Zhengpinmai 8, and Shangmai 156 had large root systems under all three growing environments. Root traits were negatively correlated with GY, SPA, and GPS, but positively correlated with TKW. 12 co-localization regions were detected with an R^2 range of 6.38–19.43%. Among them, a stable region for TRL was detected within 737.85–742.00 Mb on chromosome 4A over three growing environments. In this region, there were 27 genes involved in cell elongation and differentiation, and some genes (*TraesCS4A02G484200, TraesCS4A02G484800*, *TraesCS4A02G493900*, and *TraesCS4A02G494000)* were highly expressed in root tissues. Another vital region on chromosome 5B may contribute to both RDW and TKW.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

FX, KZ, and DH conceived the topic. SC and KZ provided gene chip. ZZ, YH, MS, and JZ collected the phenotype regulation data. SZ, XY, and JW performed the data analysis. FX and SC wrote the manuscript with edits from other co-authors. All authors read and approved the manuscript.

FUNDING

The work was supported the National Key Research and Development Program of China "Science and Technology Innovation of High Grain Production Efficiency" (2018YFD0300701) and National Key Basic Research Program of China "Interaction Network of quality and yield traits and whole genome selection model in wheat" (2014CB138105).

ACKNOWLEDGMENTS

We thank Xuehai Zhang, Jingjing Mei, Chaonan Shi, and Congwei Sun (Henan Agriculture University) for their technical assistance in the field, glasshouse and laboratory work. We also thank American Journal Experts for language editing of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.646712/full#supplementary-material

Supplementary Figure 1 | Wheat under OPC growth at different stages.

Supplementary Figure 2 | Plots of the frequency distributions of root traits under different growing environments and at different stages.

Supplementary Figure 3 | Population structure analysis of the GWAS panel and LD decay. **(A)** Plot of delta K against putative K ranging from 1 to 10; **(B)** plot of first principle component against second principle components; **(C)** plot of clustering based on Archaeopteryx tree; **(D)** over different genetic distances (Mb) for the A,B,D subgenomes and whole hexaploid wheat genome.

Supplementary Figure 4 | Manhattan and quantile-quantile (Q-Q) plots for root traits.

Supplementary Figure 5 | Haploview plot for LD decay within the candidate region on chromosome 4A and TRL phenotypic effect of allele variation.

Supplementary Table 1 | Information on the 196 wheat accessions, phenotypic and the corresponding *u*-values for root under different growing environments.

Supplementary Table 2 | Difference of root traits among OHC, IHC, and OPC at seedling stage.

Supplementary Table 3 | Correlation coefficients of root traits under different growing environments.

Supplementary Table 4 | GWAS results for root traits according to the mixed linear model.

Supplementary Table 5 | Genetic positions of SNPs in co-localization associated with root traits under different growing environments.

Supplementary Table 6 | Physical positions and associated traits of the pleiotropic SNP markers from different growing environments.

Supplementary Table 7 | Gene annotation of the genomic region (737.85–742.00 Mb) on chromosome 4A of Chinese Spring.

Supplementary Table 8 | The expression data (fragments per kilobase of transcript per million reads sequenced, FPKM) in different organs at different stages.

REFERENCES

Atkinson, J. A., Wingen, L. U., Marcus, G., Pound, M. P., Oorbessy, G., John, F. M., et al. (2015). Phenotyping pipeline reveals major seedling root growth QTL in hexaploid wheat. *J. Exp. Bot.* 66, 2283–2292. doi: 10.1093/jxb/erv006

Atta, B. M., Mahmood, T., and Trethowan, R. M. (2013). Relationship between root morphology and grain yield of wheat in north-western NSW, Australia. *Aust. J. Crop Sci.* 7, 2108–2115.

Ayalew, H., Desalegn, T., Hui, L., and Yan, G. (2016). Performance of Ethiopian bread wheat (*Tritium aestivum* L.) genotypes under contrasting water regimes: Potential sources of variability for drought resistance breeding. *Aust. J. Crop Sci.* 10, 370–376. doi: 10.21475/ajcs.2016.10.03.p7230

Bai, C., Ge, Y., Ashton, R. W., Evans, J., Milne, A., Hawkesford, M. J., et al. (2019). The relationships between seedling root screens, root growth in the field and grain yield for wheat. *Plant Soil.* 440, 311–326. doi: 10.1007/s11104-019-04088-9

Beyer, S., Daba, S., Tyagi, P., Bockelman, H., Brown-Guedira, G., Mohammadi, M., et al. (2019). Loci and candidate genes controlling root traits in wheat seedlings—a wheat root GWAS. *Funct. Integr. Genomic.* 19, 91–107. doi: 10.1007/s10142-018-0630-z

Bishopp, A., and Lynch, J. P. (2015). The hidden half of crop yields. *Nat. Plants* 1:1511.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Cao, P., Ren, Y., Zhang, K., Teng, W., Zhao, X., Dong, Z., et al. (2014). Further genetic analysis of a major quantitative trait locus controlling root length and related traits in common wheat. *Mol. Breeding* 33, 975–985. doi: 10.1007/s11032-013-0013-z

Chang, H., Chen, D., Kam, J., Richardson, T., Drenth, J., Guo, X., et al. (2016). Abiotic stress upregulated TaZFP34 represses the expression of type-B response regulator and *SHY2* genes and enhances root to shoot ratio in wheat. *Plant Sci.* 252, 88–102. doi: 10.1016/j.plantsci.2016.07.011

Chen, S., Cheng, X., Yu, K., Chang, X., Bi, H., Xu, H., et al. (2020). Genome-wide association study of differences in 14 agronomic traits under low- and high-density planting models based on the 660k SNP array for common wheat. *Plant Breed.* 139, 272–283. doi: 10.1111/pbr.12774

Christopher, J., Christopher, M., Jennings, R., Jones, S., Fletcher, S., Borrell, A., et al. (2013). QTL for root angle and number in a population developed from bread wheats (*Triticum aestivum*) with contrasting adaptation to water-limited environments. *Theor. Appl. Genet.* 126, 1563–1574. doi: 10.1007/s00122-013-2074-0

Cui, P., Liu, H., Ruan, S., Ali, B., Gill, R. A., Ma, H., et al. (2017). A zinc finger protein, interacted with cyclophilin, affects root development via IAA pathway in rice. *J. Integr. Plant Biol.* 59, 496–505. doi: 10.1111/jipb.12531

Dello Ioio, R., Linhares, F. S., Scacchi, E., Casamitjana-Martinez, E., Heidstra, R., Costantino, P., et al. (2007). Cytokinins determine *Arabidopsis* root-meristem size by controlling cell differentiation. *Curr. Biol.* 17, 678–782. doi: 10.1016/j.cub.2007.02.047

Earl, D. A., and Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Fang, S., Yan, X., and Liao, H. (2009). 3D reconstruction and dynamic modeling of root architecture in situ and its application to crop phosphorus research. *Plant J.* 60, 1096–1108. doi: 10.1111/j.1365-313X.2009.04009.x

Fang, Y., Du, Y., Wang, J., Wu, A., Qiao, S., Xu, B., et al. (2017). Moderate drought stress affected root growth and grain yield in old, modern and newly released cultivars of winter wheat. *Front. Plant Sci.* 8:672. doi: 10.3389/fpls.2017.00672

González, F. G., Miralles, D. J., and Slafer, G. A. (2011). Wheat floret survival as related to pre-anthesis spike growth. *J. Exp. Bot.* 62, 4889–4901. doi: 10.1093/jxb/err182

Hu, H., Xiong, L., and Yang, Y. (2005). Rice SERK1 gene positively regulates somatic embryogenesis of cultured cell and host defense response against fungal infection. *Planta* 222, 107–117. doi: 10.1007/s00425-005-1534-4

Hua, Z., and Vierstra, R. D. (2011). The cullin-ring ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* 62, 299–334. doi: 10.1146/annurev-arplant-042809-112256

Ibrahim, S. E., Schubert, A., Pillen, K., and Léon, J. (2012). QTL analysis of drought tolerance for seedling root morphological traits in an advanced backcross population of spring wheat. *Int. J. Agrc. Stat. Sci.* 2, 619–629.

Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233

Kabir, M. R., Liu, G., Guan, P., Wang, F., Khan, A. A., Ni, Z., et al. (2015). Mapping QTLs associated with root traits using two different populations in wheat (*Triticum aestivum* L.). *Euphytica* 206, 175–190. doi: 10.1007/s10681-015-1495-z

Lei, M., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001

Li, B., Liu, D., Li, Q., Mao, X., Li, A., Wang, J., et al. (2016). Overexpression of wheat gene *TaMOR* improves root system architecture and grain yield in *Oryza sativa*. *J. Exp. Bot.* 67, 4155–4167. doi: 10.1093/jxb/erw193

Li, L., Peng, Z., Mao, X., Wang, J., Chang, X., Reynolds, M., et al. (2019). Genome-wide association study reveals genomic regions controlling root and shoot traits at late growth stages in wheat. *Ann. Bot.* 124:6. doi: 10.1093/aob/mcz041

Li, Q., Wang, W., Wang, W., Zhang, G., Liu, Y., Wang, Y., et al. (2018). Wheat F-box protein gene *TaFBA1* is involved in plant tolerance to heat stress. *Front. Plant Sci.* 9:521. doi: 10.3389/fpls.2018.00521

Liu, X., Li, R., Chang, X., and Jing, R. (2013). Mapping QTLs for seedling root traits in a doubled haploid wheat population under different water regimes. *Euphytica* 189, 51–66. doi: 10.1007/s10681-012-0690-4

Liu, Y., Wang, L., Mao, S., Liu, K., and Zheng, Y. (2015). Genome-wide association study of 29 morphological traits in *Aegilops tauschii*. *Sci. Rep.* 5, 15562. doi: 10.1038/srep15562

Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., et al. (2017). A genome-wide association study of 23 agronomic traits in chinese wheat landraces. *Plant J.* 91:861. doi: 10.1111/tpj.13614

Lopes, M. S., and Reynolds, M. P. (2010). Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. *Funct. Plant Biol.* 37, 147–156. doi: 10.1071/FP09121

Ma, S. C., Xu, B. C., Li, F. M., Liu, W. Z., and Huang, Z. B. (2008). Effects of root pruning on competitive ability and water use efficiency in winter wheat. *Field Crop Res.* 105, 56–63. doi: 10.1016/j.fcr.2007.07.005

Maccaferri, M., El-Feki, W., Nazemi, G., Salvi, S., Canè, M. A., Colalongo, M. C., et al. (2016). Prioritizing quantitative trait loci for root system architecture in tetraploid wheat. *J. Exp. Bot.* 67, 1161–1178. doi: 10.1093/jxb/erw039

Man, J., Shi, Y., Yu, Z., and Zhang, Y. (2015). Dry matter production, photosynthesis of flag leaves and water use in winter wheat are affected by supplemental irrigation in the Huang-Huai-Hai plain of China. *PLoS One* 10:e0137274. doi: 10.1371/journal.pone.0137274

Morey, S. R., Hirose, T., Hashida, Y., Miyao, A., Hirochika, H., Ohsugi, R., et al. (2018). Genetic evidence for the role of a rice vacuolar invertase as a molecular sink strength determinant. *Rice* 11:6. doi: 10.1186/s12284-018-0201-x

Neumann, G., George, T. S., and Plassard, C. (2009). Strategies and methods for studying the rhizosphere—the plant science toolbox. *Plant Soil.* 321, 431–456. doi: 10.1007/s11104-009-9953-9

Neumann, K., Kobiljski, B., Deni, S., Varshney, R. K., and Brner, A. (2011). Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* 27, 37–58. doi: 10.1007/s11032-010-9411-7

Osmont, K. S., Sibout, R., and Hardtke, C. S. (2007). Hidden branches: developments in root system architecture. *Annu. Rev. Plant Biol.* 58, 93–113. doi: 10.1146/annurev.arplant.58.032806.104006

Passioura, J. B. (1983). Roots and drought resistance. *Agr. Water Manage.* 7, 265–280. doi: 10.1016/0378-3774(83)90089-6

Ren, Y., He, X., Liu, D., Li, J., Zhao, X., Li, B., et al. (2012). Major quantitative trait loci for seminal root morphology of wheat seedlings. *Mol. Breed.* 30, 139–148. doi: 10.1007/s11032-011-9605-7

Reynolds, M. P., Pierre, C. S., Saad, A. S., Vargas, M., and Condon, A. G. (2007). Evaluating potential genetic gains in wheat associated with stressadaptive trait expression in elite genetic resources under drought and heat stress. *Crop Sci.* 47, 172–189. doi: 10.2135/cropsci2007.10.002

Robertson, B. M., Waines, J. G., and Gill, B. S. (1979). Genetic Variability for Seedling Root Numbers in Wild and Domesticated Wheats1. *Crop Sci.* 19, 843–847. doi: 10.2135/cropsci1979.0011183X001900060024x

Sergeeva, L. I., Keurentjes, J. J., Bentsink, L., Vonk, J., van der Plas, L. H., Koornneef, M., et al. (2006). Vacuolar invertase regulates elongation of *Arabidopsis thaliana* roots as revealed by QTL and mutant analysis. *Proc. Natl. Acad. Sci. USA.* 103, 2994–2999. doi: 10.1073/pnas.0511015103

Steele, K. A., Price, A. H., Witcombe, J. R., Shrestha, R., Singh, B. N., Gibbons, J. M., et al. (2013). QTLs associated with root traits increase yield in upland rice when transferred through marker-assisted selection. *Theor. Appl. Genet.* 126, 101–108. doi: 10.1007/s00122-012-1963-y

Steinemann, S., Zeng, Z. H., McKay, A., Heuer, S., Langridge, P., and Huang, C. Y. (2015). Dynamic root responses to drought and rewatering in two wheat (*Triticum aestivum*) genotypes. *Plant Soil.* 391, 139–152. doi: 10.1007/s11104-015-2413-9

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* 128, 353–363. doi: 10.1007/s00122-014-2435-3

Sun, J. J., Guo, Y., Zhang, G. Z., Gao, M. G., and Li, S. S. (2013). QTL mapping for seedling traits under different nitrogen forms in wheat. *Euphytica* 191, 317–331. doi: 10.1007/s10681-012-0834-6

Suraj, G., Sujan, M., Michael, B. J., Mai, X., Gina, B. G., Adhikari, T. B., et al. (2014). Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PLoS One* 9:e108179. doi: 10.1371/journal.pone.0108179

Valassi, A., and Chierici, R. (2014). Information and treatment of unknown correlations in the combination of measurements using the BLUE method. *Eur. Phys. J. C.* 74:2717. doi: 10.1140/epjc/s10052-014-2717-6

Wang, S. X., Zhu, Y. L., Zhang, D. X., Shao, H., and Ma, C. X. (2017). Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using snp markers. *PLoS One* 12:e0188662. doi: 10.1371/journal.pone.0188662

Walker, J. C. (1993). Receptor-like protein kinase genes of *Arabidopsis thaliana*. *Plant J.* 3, 451–456. doi: 10.1111/j.1365-313x.1993.tb00164.x

Xie, Q., Fernando, K. M. C., Mayes, S., and Sparkes, D. L. (2017). Identifying seedling root architectural traits associated with yield and yield components in wheat. *Ann. Bot.* 119, 1115–1129. doi: 10.1093/aob/mcx001

Ye, X., Li, J., Cheng, Y., Yao, F., and Chen, G. (2019). Genome-wide association study reveals new loci for yield-related traits in Sichuan wheat germplasm under stripe rust stress. *BMC Genomics* 20:640. doi: 10.1186/s12864-019-6005-6

Zheng, X., Wen, X., Qiao, L., Zhao, J., Zhang, X., Li, X., et al. (2019). A novel QTL *QTrl.saw-2D.2* associated with the total root length identified by linkage and association analyses in wheat (*Triticum aestivum* L.). *Planta* 250, 129–143. doi: 10.1007/s00425-019-0315

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Xu, Chen, Yang, Zhou, Wang, Zhang, Huang, Song, Zhang, Zhan and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fgene.2021.656037

Identification of Main-Effect and Environmental Interaction QTL and Their Candidate Genes for Drought Tolerance in a Wheat RIL Population Between Two Elite Spring Cultivars

S. M. Hisam Al Rabbi[1], Ajay Kumar[1], Sepehr Mohajeri Naraghi[1], Suraj Sapkota[2], Mohammed S. Alamri[3], Elias M. Elias[1], Shahryar Kianian[4], Raed Seetan[5], Ali Missaoui[2,6], Shyam Solanki[7] and Mohamed Mergoum[2,6]**

[1] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [2] Institute of Plant Breeding, Genetics, and Genomics, University of Georgia, Griffin, GA, United States, [3] Department of Food Science and Nutrition, King Saud University, Riyadh, Saudi Arabia, [4] USDA-ARS Cereal Disease Laboratory, University of Minnesota, St. Paul, MN, United States, [5] Department of Computer Science, Slippery Rock University, Slippery Rock, PA, United States, [6] Department of Crop and Soil Sciences, University of Georgia, Griffin, GA, United States, [7] Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States

OPEN ACCESS

Edited by:
Yuan-Ming Zhang,
Huazhong Agricultural University, China

Reviewed by:
Xingwang Yu,
North Carolina State University, United States
Wankui Gong,
Cotton Research Institute (CAAS), China

***Correspondence:**
Mohamed Mergoum
mmergoum@uga.edu
Ajay Kumar
ajay.kumar.2@ndsu.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 20 January 2021
Accepted: 13 May 2021
Published: 17 June 2021

Citation:
Rabbi SMHA, Kumar A, Mohajeri Naraghi S, Sapkota S, Alamri MS, Elias EM, Kianian S, Seetan R, Missaoui A, Solanki S and Mergoum M (2021) Identification of Main-Effect and Environmental Interaction QTL and Their Candidate Genes for Drought Tolerance in a Wheat RIL Population Between Two Elite Spring Cultivars. Front. Genet. 12:656037. doi: 10.3389/fgene.2021.656037

Understanding the genetics of drought tolerance can expedite the development of drought-tolerant cultivars in wheat. In this study, we dissected the genetics of drought tolerance in spring wheat using a recombinant inbred line (RIL) population derived from a cross between a drought-tolerant cultivar, 'Reeder' (PI613586), and a high-yielding but drought-susceptible cultivar, 'Albany.' The RIL population was evaluated for grain yield (YLD), grain volume weight (GVW), thousand kernel weight (TKW), plant height (PH), and days to heading (DH) at nine different environments. The Infinium 90 k-based high-density genetic map was generated using 10,657 polymorphic SNP markers representing 2,057 unique loci. Quantitative trait loci (QTL) analysis detected a total of 11 consistent QTL for drought tolerance-related traits. Of these, six QTL were exclusively identified in drought-prone environments, and five were constitutive QTL (identified under both drought and normal conditions). One major QTL on chromosome 7B was identified exclusively under drought environments and explained 13.6% of the phenotypic variation (PV) for YLD. Two other major QTL were detected, one each on chromosomes 7B and 2B under drought-prone environments, and explained 14.86 and 13.94% of phenotypic variation for GVW and YLD, respectively. One novel QTL for drought tolerance was identified on chromosome 2D. *In silico* expression analysis of candidate genes underlaying the exclusive QTLs associated with drought stress identified the enrichment of ribosomal and chloroplast photosynthesis-associated proteins showing the most expression variability, thus possibly contributing to stress response by modulating the glycosyltransferase (*TraesCS6A01G116400*) and hexosyltransferase (*TraesCS7B01G013300*) unique genes present in QTL 21 and 24, respectively. While both parents contributed favorable alleles to these QTL,

unexpectedly, the high-yielding and less drought-tolerant parent contributed desirable alleles for drought tolerance at four out of six loci. Regardless of the origin, all QTL with significant drought tolerance could assist significantly in the development of drought-tolerant wheat cultivars, using genomics-assisted breeding approaches.

Keywords: drought tolerance, hard red spring wheat, quantitative trait loci, recombinant inbred line, marker-assisted selection

INTRODUCTION

Hard red spring wheat (HRSW), comprising about 25% of the total United States wheat production, is unique for its high protein content (Vocke and Ali, 2013). However, this important crop often experiences drought, which is one of the main natural hazards harming wheat production worldwide (Araus et al., 2008). It regularly affects about 50% of wheat-producing areas (Pfeiffer et al., 2005). Drought refers to reduced accessible water in the soil and atmospheric conditions that cause plants to wilt or even die by losing water through transpiration. However, drought tolerance enables plants to yield satisfactorily under limited or periodic water-deficient conditions (Turner, 1979). Therefore, developing wheat cultivars with improved drought tolerance is the key to reduce yield loss due to water stress.

Drought tolerance in wheat can be achieved through developing cultivars capable of maintaining high water potential under drought conditions (Turner et al., 2001). Also, plants could escape from late-season drought through the development of early wheat cultivars (Araus et al., 2002). Understanding the genetics of drought tolerance in wheat is, therefore, a prerequisite to develop new adapted and drought-tolerant cultivars. Early research indicated that drought tolerance in crop plants is quantitatively inherited, or controlled by many genes or gene complexes (Blum, 1988), which can in turn be traced through quantitative trait loci (QTL) mapping methods.

Breeders have frequent debates over the appropriate phenotypic approaches for QTL analysis (Alexander et al., 2012). Many morphological traits, such as root length, tillering, spike number per m^2, grain number per spike, number of fertile tillers per plant, one thousand grain weight, peduncle length, spike weight, stem weight, awn length, and grain weight per spike, can be affected by drought (Blum, 2005). However, yield (YLD) stability under both drought-stressed and favorable environments has been proposed for the effective selection of drought-tolerant genotypes (Pinter et al., 1990). From a breeder's perspective, YLD and yield-related traits comprise the best morphological traits to screen for drought-tolerant genotypes. Hence, the source of QTL related to drought tolerance and the contribution of favorable alleles to this trait from diverse cultivars including high-yielding but non-drought-tolerant needs to be clarified.

An efficient tool for dissecting the genetics of drought is needed as most of the QTL mapping studies conducted on drought tolerance in wheat have used low-resolution maps composed of only several hundred molecular markers (Kirigwi et al., 2007; Muchero et al., 2009; Peleg et al., 2009; Sayed, 2011; Alexander et al., 2012; Ibrahim et al., 2012b; Kumar et al., 2012; Malik et al., 2015). Because of the size of the bread wheat genome (∼17 Gb), greater marker coverage is needed to generate a dense genetic linkage map, which could help to identify tightly linked markers associated with genes controlling traits of interest (Kumar et al., 2016, 2019). This is very important for the successful introgression of target loci in marker-assisted selection (MAS) and/or genomic selection methods in breeding programs. Precise identification of QTL will also facilitate easier positional cloning of those QTL (Kumar et al., 2016). The Infinium iSelect 90K assay, with 81,587 transcriptome-based single-nucleotide polymorphisms (SNPs) (Wang et al., 2014), can be an excellent tool for investigating the genetic basis of drought tolerance in wheat. Therefore, in this study, the main objective was to decipher the genetics of drought tolerance in wheat in the northern region of United States using a recombinant inbred line (RIL) population derived from a drought-tolerant cultivar 'Reeder' (PI613586) and a high-yielding and non-drought-tolerant cultivar 'Albany'. Additionally, it has been long speculated (particularly, at the International Wheat and Maize Center, CIMMYT with "Shuttle" breeding concept, engineered by Dr. N. Borlaug) that many genes contributing positively to increased yield do so under both stressed and non-stressed conditions, including water stress/drought. Therefore, this study aims to elucidate that concept as well. The knowledge and resources developed using multi-location phenotypic data and high-density genetics map in this study will play an important role in our efforts toward development of drought-tolerant wheat cultivars.

MATERIALS AND METHODS

Plant Materials

The cultivars Reeder and Albany were used to develop a RIL population consisting of 149 lines. Reeder is a drought-tolerant HRSW cultivar released by the North Dakota Agricultural Experiment Station at North Dakota State University (NDSU) in 1999. It is a semi-dwarf cultivar best adapted to western North Dakota (ND), a semiarid region of the state. Reeder has good milling and baking qualities and also possesses resistance to the Upper Midwest races of stem (caused by *Puccinia graminis* f. sp. *tritici*), stripe (caused by *Puccinia striiformis* Westend f. sp. *tritici* Eriks & Henn), and leaf (caused by caused by *Puccinia triticina* Erikss.) rusts. The other drought-sensitive parent, Albany, is a HRSW cultivar developed by Trigen Seed LLC. It is a very high-yielding, semi-dwarf cultivar adapted to high-input management conditions and better adapted to the eastern area of the Northern Plains spring wheat region, where drought conditions are not

prevalent. A single seed descent (SSD) method was used to advance the RIL populations to the F_8 generation. The study also included the checks, 'Glenn' (Mergoum et al., 2006), 'SY Tyra' (AgriPro® wheat variety, United States), 'Faller' (Mergoum et al., 2008), 'Steele-ND' (Mergoum et al., 2005b), 'Alsen' (Frohberg et al., 2006), 'Mott,' 'Elgin' (Mergoum et al., 2016), 'RB07' (Anderson et al., 2009), 'Dapps' (Mergoum et al., 2005a), 'Prosper' (Mergoum et al., 2013), 'ND901CLPlus' (Mergoum et al., 2009) (PI655233), 'Velva' (Mergoum et al., 2014), 'SY Soren' (AgriPro® wheat variety, United States), 'Duclair' (Lanning et al., 2011), 'ND819' (an elite experimental line developed by the NDSU spring wheat breeding program), 'Polaris', 'Saturn', and 'Granite' (PI619072). The checks ND819, Dapps, and Steele-ND are tolerant to drought stress. The genotypes SY Soren, Glenn, Alsen, ND901CLPlus, Saturn, and Velva show moderate tolerance, whereas Granite, Elgin, RB07, Duclair, Prosper, Mott, Faller, and SY Tyra are well adapted to high rainfall regions and therefore are most likely more susceptible to drought.

Field Experiments and Phenotypic Data Collection

The evaluation of agronomic performances of the parents, RIL population, and 18 checks was carried out under non-irrigated field conditions at different locations in ND. The plant material was evaluated at Prosper and Carrington in 2012, 2013, and 2014; Minot in 2012; Williston in 2013; and Hettinger in 2014. Prosper is located in the eastern region of ND (46.9630°N, 97.0198°W). Carrington is located in the east-central region of ND (47.4497°N, 99.1262°W). Minot sits between semiarid grassland in the west and central ND's subhumid grassland (48.2330°N, 101.2923°W). Williston is located in northwestern ND (48.1470°N, 103.6180°W) and Hettinger in southwestern ND (46.0014°N, 102.6368°W). The total rainfall in Prosper during the 2012, 2013, and 2014 growing periods (seed sowing to ripening) was 120.1 mm, 269.9 mm, and 176.8 mm, respectively (**Table 1**). Carrington had total rainfall of 171.2 mm, 159.8 mm, and 190.5 mm during the 2012, 2013, and 2014 growing periods, respectively. Moreover, during the same growing periods, Minot, Williston, and Hettinger had total rainfall of 162.2 mm, 320.4 mm, and 200.3 mm, respectively (**Table 1**; North Dakota Agricultural Weather Network (NDAWN), 2015). The available soil moisture of the experimental sites based on soil types is presented in **Table 1**. Each experiment was conducted in a randomized complete block design (RCBD) with two replicates. In 2012 and 2013, each genotype was planted in a 2.44 m × 1.22 m plot containing seven rows 15.24 cm apart. The plot size was slightly larger in 2014, at 2.44 m × 1.42 m, with the same number of rows (seven), but separation of 17.78 cm between rows.

Each year, the phenotypic data were recorded for days to heading (DH), plant height (PH), YLD, grain volume weight (GVW), and thousand kernel weight (TKW) at each site. The DH were taken when more than 50% of the plants in the plot were heading. The PH was measured from base to tip excluding the awn for plants in the middle of the plot. YLD per plot was converted to yield/ha for further analysis. Similarly, kg/0.5-pint cup was converted to kg/m^3 as the GVW for further analysis. The TKW was measured by counting 1000 kernels using a seed counter (Model U, International Marketing and Design Co., San Antonio, TX, United States) and weighed.

Phenotypic Data Analysis

The statistical analysis system used for analyzing the phenotypic data was ANOVA Proc MIXED (SAS Institute, 2004). The RILs, their parents, and the checks were considered as fixed effects, whereas environments and blocks were considered as random effects. The mean values were separated using the *F*-protected least significant difference (LSD) value at the $P \leq 0.05$ level of significance. Pearson correlations between traits for each environment were calculated using the SAS's CORR procedure (SAS Institute, 2004). Only the locations whose data exhibited a low coefficient of variation (CV) value and a significant difference among entries are reported in this study.

Genotyping and Linkage Map Construction

Genomic DNA from each genotype was isolated from lyophilized young leaves using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, United States, cat. no. 69106). This DNA was run on 0.8% agarose gel to check its quality. The NanoDrop 1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, United States) was used to check the DNA concentration. The RIL population, parents, and checks were genotyped using the Illumina 90K iSelect wheat SNP assay in the Small Grains Genotyping Lab (USDA-ARS, Fargo, ND, United States). The genotyping module GenomeStudio V2011.1[1] was used to analyze the SNP data.

Polymorphic markers between parental genotypes were identified. Out of polymorphic loci, we discarded markers which had (1) an allele frequency of < 0.4 for any of the parental genotypes, (2) inconsistent results in five replicates of each parental genotype, (3) overlapping clusters for RILs, and (4) $> 20\%$ missing data. The remaining markers were used for map construction using a combination of MapMaker 3.0 (Lander and Botstein, 1989) and CartaGène v.1.2.3R (de Givry et al.,

[1]www.illumina.com

TABLE 1 | Soil types, plant-available water (water-holding capacity of soil), and total rainfall for nine environments.

Environments	Soil type	Plant-available water (mm water/30.48 cm soil)	Rainfall (mm)	Total water available (mm)
Prosper 12	Fine silty loam	45.7–63.5	120.1	165.8–183.6
Carrington 12	Coarse loamy	19.1–31.7	171.2	190.3–202.9
Minot 12	Fine sandy loam	31.7–45.7	162.2	193.9–207.9
Prosper 13	Fine silty loam	45.7–63.5	269.9	315.6–333.4
Carrington 13	Coarse loamy	19.0–31.7	159.8	178.8–191.5
Williston 13	Fine sandy loam	31.7–45.7	320.4	325.1–366.1
Prosper 14	Fine silty loam	45.7–63.5	176.8	222.5–240.3
Carrington 14	Coarse loamy	19.0–31.7	190.5	209.5–222.2
Hettinger 14	Fine sandy loam	31.7–45.7	200.3	232.0–246.0

2005) software. At first, five to nine polymorphic markers from each chromosome covering the whole genome were selected as anchors based on available mapping information in multiple populations (Wang et al., 2014). Using MapMaker 3.0 (Lander and Botstein, 1989) and the anchor markers, 10,657 polymorphic markers were placed onto 21 wheat chromosomes using a minimum LOD score of 5.0 and a maximum distance of 40 cM. The linkage maps were then developed using CartaGène V.1.2.3R (de Givry et al., 2005). The details are described elsewhere (Kumar et al., 2012; Seetan et al., 2013, 2014). Briefly, the process starts with removal of identical markers. Then, initial maps are created using the "build" command, starting with the pair of most strongly linked markers. The remaining markers are then inserted incrementally. Then, the map is enhanced using "greedy search." The next step uses genetic and simulated annealing algorithms for local improvement. In the final step, a fixed-length sliding window was applied to try all permutations within the window size to identify the map. Kosambi's mapping function (Kosambi, 1944) was used to determine the genetic distance among markers on the linkage groups.

QTL Mapping and Candidate Gene Identification

Composite interval mapping (CIM) was used to identify QTL for each trait in each environment as well as across environments (AE) using QTL Cartographer V2.5_011 (Wang et al., 2012). In QTL Cartographer, Model 6 (standard model), forward and backward regression, five control markers (co-factors), window size of 10 cM, and walk speed of 1 cM were used. A total of 1,000 permutations were used to determine the LOD threshold for identifying the significant QTL. Confidence intervals (CI) were estimated by the ± 2 LOD (from the peak) method. The QTL with overlapping CIs or QTL located within 10-cM regions were considered as the same QTL. Only the significant QTL detected (those above the threshold LOD score) were included in this study. If any such QTL were identified with an LOD score below the threshold, but > 2.5 in other environments, the QTL were also included in the results as supporting information. The QTL identified in at least two environments or associated with at least two traits were also reported in this study. The QTL regions were drawn using the Mapchart 2.3 program (Voorrips, 2002). Map locations of the associated markers were used to see if the QTL identified in this study have been reported in earlier studies. The sequences of the markers flanking each QTL were obtained from the T3/Wheat database (Blake et al., 2016), and their physical positions were extracted using the BLAST search against Chinese_Spring_IWGSC_RefSeq1.0 (Alaux et al., 2018; Appels et al., 2018). For each QTL, the position of flanking markers was used to determine the underlying block of high-confidence candidate genes and their annotated function[2]. *In silico* expression analysis for drought-specific QTLs (**Table 2**) was carried out in the Wheat expression browser expVIP (Borrill et al., 2016; Ramírez-González et al., 2018) dataset for drought and heat stress and PEG to stimulate drought and identify the expression variation of any repetitive functional class.

RESULTS

Climatic Conditions and Phenotypic Analyses

Climatic conditions that prevailed in 2013 were unusually variable (available water = 159.8 to 366.1 mm) causing high CV and were not conducive to drought, particularly in the western ND region (Williston), where rainfall was unusually high (**Table 1**). Additionally, the genotypes did not show significant differences for the agronomic traits in the 2013 trials (**Supplementary Table 1**). Therefore, given all these challenges and the criteria setup described earlier, only the data from the six environments planted in 2012 and 2014 were used in this study. For these two crop seasons, overall, total moisture available to

[2]https://figshare.com/s/30a0fb5d21315fec1f01

TABLE 2 | Quantitative trait loci for drought tolerance in an RIL population derived from the cross between Reeder and Albany.

QTL	Trait[†]	QTL region	Other associated traits[‡]	Env.[‡]	Position[§]	LOD[¶]	Additive effect	R^2 (%)
QDH.ndsu.2B.1	DH	6	_	1, 2*, 3*	26.81–30.11	3.82	0.78	5.2
QDH.ndsu.4A.2	DH	13	GVW	1, 2, 3, 5, 6	133.91–143.11	9	−1.66	13.44
QDH.ndsu.5A.2	DH	16	_	1*, 2, 3	131.91–142.01	4.09	−0.92	6.22
QDH.ndsu.5A.3	DH	17	YLD, GVW	1, 3, 4, 5, 6	205.71–208.31	20.17	−2.84	38.36
QDH.ndsu.5D2	DH	20	GVW, YLD	2, 3, 4, 5, 6	11.91–20.91	15.16	2.29	29.93
QTW.ndsu.2B	GVW	7	YLD, HD	1, 2, 5, 6, M	84.31–95.61	8.02	−12.25	16.5
QTW.ndsu.7B	GVW	25	DH	1, 2, 3, M	29.11–40.11	8.95	−13.28	14.86
QTKW.ndsu.2D.1	TKW	8	_	2, 3*, M*	110.21–111.21	3.73	0.63	7.69
QTKW.ndsu.6A	TKW	21	_	1, 2, 3, 4*, 5, M	65.41–68.21	5.43	0.89	15.22
QYL.ndsu.2B	YL	7	GVW, HD	1,2, 3*, M	81.31–83.31	7.22	−209.44	13.94
QYL.ndsu.7B	YL	24	PH	1,2*	22.21–25.21	5.87	−178.75	13.6

ǂPH = plant height, DH = days to heading, YLD = yield, GVW = grain volume weight, TKW = thousand kernel weight.
[‡]Env. = environment, 1 = Carrington 2012, 2 = Minot 2012, 3 = Prosper 2012, 4 = Carrington 2014, 5 = Hettinger 2014, 6 = Prosper 2014, 7 = mean across environments.
[§]Position represents the peak point of the QTL interval.
[¶]For log of odds (LOD) score.
*The QTL in that environment was detected above a 2.5 LOD score, but below the threshold score.

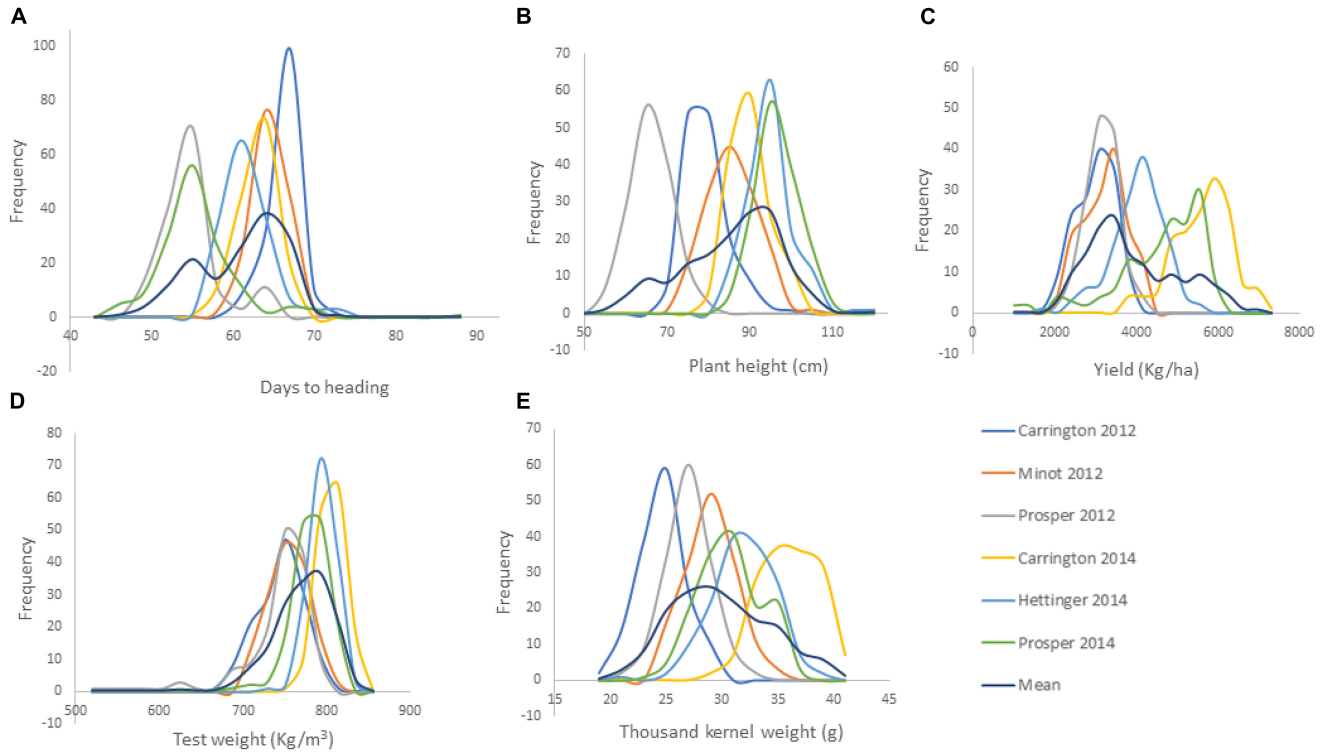


FIGURE 1 | Frequency distribution of the agronomic traits of 149 RILs of the cross of Reeder and Albany (**A.** days to heading; **B.** plant height; **C.** yield; **D.** grain volume weight indicated here as test weight; and **E.** thousand kernel weight).

the wheat crop, including soil residual water from snow, 2012 was considered as dry environment and 2014 was considered as control environment (**Table 1**). Total moisture available for wheat crop varied from 165.8 mm at Prosper in 2012 to 246 mm at Hettinger in 2014. Similarly, the DH data at Carrington in 2014 did not show a significant difference (**Supplementary Table 1**) and thus was discarded for further analysis. However, in the 2012 and 2014 field trials, significant differences among genotypes for most of the agronomic traits were observed.

The RIL population showed continuous variation for all of the agronomic traits (**Figure 1**). Transgressive segregations in both directions were also observed among the RIL population for all traits in each and across environments (**Supplementary Table 2**). This shows that parents Reeder and Albany used to generate the RIL population were diverse and possess different favorable alleles for the studied traits. In particular, YLD means within each location and across the six environments varied significantly among RIL population with transgressive segregation. However, YLD did not differ significantly between the two parents Reeder and Albany with a slight YLD advantage to the later. For the RIL population, the mean YLD across the six environments varied from 2562.57 to 4461.78 kg ha^{-1}. Meanwhile, YLD of parents Reeder and Albany varied from 3577.36 to 3800.93 kg ha^{-1}, respectively. The check YLD mean was 3846.51 kg ha^{-1}. Similar results of transgressive segregation among the RIL population were observed for all other traits, PH, HD, GVW, and TKW. Similarly, the parents Reeder and Albany did not differ for PH and GVW, while Reeder was significantly earlier and had higher TKW (**Supplementary Table 2**).

The heading date had a highly significant negative correlation with YLD, GVW, and TKW in all of the environments. Late-heading plants tended to be taller in two of the environments and also with the mean value of all the environments. Plant height did not show any significant association with any of the traits except DH. The higher-yielding genotypes had higher GVW in every environment. Similarly, higher-yielding genotypes had higher TKW in all environments except at Carrington in 2014. Again, the genotypes with high GVW tended to have high TKW in all environments except at Carrington in 2014 (**Table 3**).

Genetic Linkage Map

Out of 81,587 SNPs markers in the Illumina iSelect 90K assay (Wang et al., 2014), 12,151 SNP polymorphic markers between parental genotypes were identified (**Supplementary Table 3**). After discarding unsuitable markers, 10,760 markers were eventually used for map construction. Out of the 10,760 markers selected for linkage mapping, 10,657 markers were mapped onto 28 linkage groups found on 21 wheat chromosomes (**Table 4**; **Supplementary Table 3**). The 10,657 markers represented 2,057 unique loci (19.3%), and 8,600 markers (80.7%) co-segregated with other loci. The B-genome contained the most number of markers, followed by the A-genome and the D-genome (**Table 4**). The number of markers on individual linkage groups ranged from 5 (1D1, 5D2) to 1,221 (2B), while for

TABLE 3 | Correlation coefficients between five agronomic traits in the RIL population (Reeder × Albany) in six environments (Env.) and the overall mean across environments (M).

Trait† and Env.‡	PH	DH	YLD	GVW
DH				
1	0.04ns	–	- –	- –
2	0.32***	–	- –	- –
3	0.07ns	–	- –	- –
4	.	–	- –	- –
5	0.13ns	–	- –	- –
6	0.31***	–	- –	- –
M	0.24**	–	- –	- –
YLD				
1	0.00ns	−0.58***	- –	- –
2	−0.03ns	−0.47***	- –	- –
3	0.18*	−0.38***	- –	- –
4	0.29***	.	- –	- –
5	0.06ns	−0.44***	- –	- –
6	−0.27***	−0.68***	- –	- –
M	−0.07ns	−0.59***	- –	- –
GVW				
1	−0.02ns	−0.62***	0.62***	- –
2	−0.05ns	−0.57***	0.53***	- –
3	−0.19*	−0.72***	0.51***	- –
4	0.029ns	.	0.18*	- –
5	0.01ns	−0.33***	0.29***	- –
6	−0.21**	−0.55***	0.61***	- –
M	−0.14ns	−0.6***	0.49***	- –
TKW				
1	0.17*	−0.45***	0.56***	0.47***
2	0.2**	−0.33***	0.34***	0.45***
3	0.2*	−0.28***	0.23**	0.29***
4	0.19*	.	0.12ns	0.03ns
5	0.08ns	−0.48***	0.25**	0.36***
6	−0.06ns	−0.5***	0.47***	0.52***
M	0.13ns	−0.4***	0.29***	0.30***

*Significant at $p < 0.05$. **Significant at $p < 0.01$. ***Significant at $p < 0.001$ level
†PH = plant height, DH = days to heading, YLD = yield, GVW = grain volume weight, TKW = thousand kernel weight.
‡1 = Carrington 2012, 2 = Minot 2012, 3 = Prosper 2012, 4 = Carrington 2014, 5 = Hettinger 2014, 6 = Prosper 2014, M = Mean across environments.

TABLE 4 | Distribution of markers across linkage groups in the genetic map developed using the Reeder × Albany RIL population.

Linkage groups	No. of markers	No. of unique loci	Map length	Average map density	Average map density
				cM/marker	cM/locus
1A	567	126	174.90	0.31	1.39
2A	439	101	223.50	0.51	2.21
3A	659	123	213.90	0.32	1.74
4A	560	114	218.90	0.39	1.92
5A	605	163	299.00	0.49	1.83
6A	590	117	176.70	0.30	1.51
7A	905	168	235.30	0.26	1.40
1B	629	86	107.50	0.17	1.25
2B	1221	160	181.80	0.15	1.14
3B	1115	213	250.20	0.22	1.17
4B	244	78	120.90	0.50	1.55
5B1	565	125	209.40	0.37	1.68
5B2	25	8.00	18.00	0.72	2.25
6B	426	101	158.10	0.37	1.57
7B	723	134	213.20	0.29	1.59
1D1	5	2	0.30	0.06	0.15
1D2	254	40	87.80	0.35	2.20
1D3	91	26	126.10	1.39	4.85
2D	653	46	180.40	0.28	3.92
3D	48	18	162.90	3.39	9.05
4D	53	23	129.90	2.45	5.65
5D1	25	8	47.50	1.90	5.94
5D2	5	4	24.90	4.98	6.23
5D3	130	21	31.50	0.24	1.50
6D1	10	5	3.00	0.30	0.60
6D2	23	19	44.50	1.93	2.34
6D3	22	6	4.00	0.18	0.67
7D	65	22	149.00	2.29	6.77
A genome	4,325	912	1,542.20	0.37	1.72
B genome	4,948	905	1,259.10	0.35	1.52
D genome	1,384	240.00	991.80	1.52	3.84
Whole genome	10,657	2,057	3,793.10	0.36	1.84

individual chromosomes, the number of markers ranged from 48 (chromosome 3D) to 1,221 (chromosome 2B) (**Table 4**). The average number of markers mapped per chromosome was 507.48, while the average number of unique loci per chromosome was 97.95.

The 10,657 (2,057 loci) markers mapped in this study covered a total genetic map length of 3,793.1 cM, with an average distance of 0.36 cM between any two markers (**Table 4**). The A-genome chromosomes covered a total length of 1,542.2 cM, with an average distance of 0.37 cM between two markers. The B-genome had a total map length of 1,259.1 cM, with an average distance of 0.35 cM between two markers. The D-genome covered a total map length of 991.8 cM, with an average distance of 1.52 cM between two markers. Individually, chromosome 5A was the longest, with a total map length of 299 cM. Chromosome 6D was the shortest, with a total map length of 51.5 cM. Overall, the marker order was consistent with earlier studies on wheat genetic maps (Wang et al., 2014).

QTL Analysis

YLD

Composite interval mapping for YLD identified six QTL located on six different chromosomes (**Table 5**; **Supplementary Figure 1**). Four of these QTL explained greater than 10% of PV and were considered as major QTL. The major QTL located on chromosome 2B had a phenotypic variation (PV) up to 13.94%; that on 5A had a PV up to 22.35%; and that on 5D had a PV up to 22.83%. All these three QTL were identified in three of the environments and in the overall mean and, thus, could be

considered as consistent or stable QTL. The fourth major QTL on chromosome 7B was identified in one location and in the overall mean, explaining up to 13.6% of PV. The alleles for higher YLD for the QTL on chromosomes 5D, 2B, and 7B were contributed by the parent Albany, whereas the allele for the major QTL on chromosome 5A was contributed by Reeder (**Table 5**).

GVW

Seven QTL located on six different chromosomes were identified for GVW (Table **5**; **Supplementary Figure 1**). Five QTL among them were considered as major QTL as they have explained PV higher than 10%. The QTL with the greatest effect (PV of up to 24.47%) was located on chromosome 5D and identified in two different environments and in the overall mean. The second major QTL, with up to 17.79% PV, was on chromosome 5A and identified in two of the environments. The major QTL on chromosome 2B had the third greatest and consistent effect as it was identified in four different environments, with a PV of up to 16.5%. The fourth major QTL was located on 2A (with a PV of up to 15.93%) and was identified in three of the environments and in the overall mean. A fifth major QTL on chromosome 7B, explaining up to 14.86% of PV, was identified in three different environments and in the overall mean. The alleles for a higher grain volume weight for the major QTL on chromosomes 5D, 2B, 2A, and 7B were contributed by the parent Albany. The allele for the remaining major QTL on chromosome 5A was contributed by Reeder (**Table 5**).

TKW

The eight QTL identified for TKW were located on seven different chromosomes (**Table 5**; **Supplementary Figure 1**). Only three among these with PV more than 10% were considered major QTL. The major QTL with the largest phenotypic effect (with a PV of up to 15.22%) was located on chromosome 6A; it also had a consistent effect as it was identified in five different environments and in the overall mean. The second major QTL was located on chromosome 4A, explaining 14.18% of PV, but it was identified in only a single environment. Another QTL explaining up to 9.66% of PV was located on chromosome 2A and identified in two different environments and in the overall mean. The alleles for increased TKW for the major QTL on 6A were contributed by the cultivar Reeder (**Table 5**).

DH

Nine QTL located on five different chromosomes were identified for DH. These QTL explained from 4.12 to 38.36% of PV (**Table 5**; **Supplementary Figure 1**). Four QTL explained > 10% of PV and, therefore, can be considered as major QTL. The QTL with the greatest and consistent effect for DH was identified on chromosome 5A in all of the environments except one and explained up to 38.36% of PV. The second major QTL was identified on chromosome 5D in all of the environments except one and explained up to 29.93% of PV. The third major QTL explained 17.4% of PV and was identified on 7B in all of the environments. The fourth major QTL was identified on chromosome 4A in all of the environments except one and explained up to 13.44% of PV. The alleles for reduced DH on 5A and 4A were contributed by the parent Reeder, while the alleles for reduced DH on the other two major QTL were contributed by the parent Albany.

PH

Eight QTL identified for PH were located on seven different chromosomes (**Table 5**; **Supplementary Figure 1**). Two of them were considered major QTL (PV > 10%). The QTL found on chromosome 2D had the largest effect, explaining up to 17.2% of PV. This QTL was identified in three different environments and in the overall mean. The second major QTL found on chromosome 6A was also identified in three different environments and explained up to 11.37% of PV. Besides these, three more QTL explained almost 10% of PV. Two of them were identified on chromosome 7B, and another one on chromosome 5B. The QTL in the QTL region 26 of chromosome 7B was identified in three environments and in the overall mean. Another QTL in the QTL region 24 of chromosome 7B was identified in two of the environments and in the overall mean. The QTL on chromosome 5B was identified in two environments only. The alleles for reduced PH for the abovementioned QTL on chromosomes 2D, 6A, and 7B were contributed by the parent Albany. The allele for reduced PH on chromosome 5B was contributed by the parent Reeder (**Table 5**).

Co-localized or Pleiotropic QTL

Co-localized QTL could be used for simultaneous improvement of more than one trait when the desirable alleles come from the same parent. A total of 38 QTL were identified in this study for five agronomic traits (**Table 5**; **Supplementary Figure 1**). Many of those QTL had overlapping confidence intervals (CI). The QTL with overlapping CI or located within 10 cM of each other were considered as the same QTL region. Overall, these 38 QTL were located in 26 different genomic regions on 13 different chromosomes. A total of 21 co-localized or pleiotropic QTL were located in nine genomic regions. Individual genomic regions were associated with two to three traits. Genomic region 7 was associated with DH, YLD, and GVW. The QTL for YLD (*QYL.ndsu.2B*) and GVW (*QTW.ndsu.2B*) had a major effect, whereas that for DH (*QDH.ndsu.2B.2*) had a minor effect. The genomic region 20 located on chromosome 5D also harbored major QTL for the same three traits. The desirable alleles in both regions (7 and 20) were contributed by the parent Albany. Meanwhile, Reeder contributed the favorable alleles for genomic region 17 on chromosome 5A which harbored the major QTL (*QDH.ndsu.5A.3*, *QYL.ndsu.5A*, and *QTW.ndsu.5A*) for the same three traits.

Six genomic regions harbored QTL for two traits. Genomic region 12 harbored QTL for TKW (*QTKW.ndsu.4A*) and DH (*QDH.ndsu.4A.1*). The QTL for DH had a minor effect, whereas the QTL for TKW had a major effect. Reeder contributed the desirable alleles in both cases. Genomic region 13 harbored QTL for DH (*QDH.ndsu.4A.2*) and GVW (*QTW.ndsu.4A*). The QTL for DH was a major QTL, while that for GVW was minor. Desirable alleles for both traits were contributed by Reeder. Genomic region 25 was also associated with DH (*QDH.ndsu.7B*) and GVW (*QTW.ndsu.7B*). Both QTL had major effects, with

TABLE 5 | Quantitative trait loci identified for the agronomic traits in an RIL population derived from the cross between Reeder and Albany.

QTL and trait	QTL region	Other associated traits†	Env.‡	Position§	LOD¶	Additive effect	R^2 (%)
Days to heading (DH)							
QDH.ndsu.2B.1	6	_	1, 2*,3*	26.81–30.11	3.82	0.78	5.2
QDH.ndsu.2B.2	7	YLD, GVW	1	76.11	4.25	0.52	5.74
QDH.ndsu.4A.1	12	TKW	2	47.51	4.56	−0.66	8.24
QDH.ndsu.4A.2	13	GVW	1, 2, 3, 5, 6	133.91–143.11	9	−1.66	13.44
QDH.ndsu.5A.1	15	_	1*, 6	109.51–112.61	3.48	−0.61	4.12
QDH.ndsu.5A.2	16	_	1*,2, 3	131.91–142.01	4.09	−0.92	6.22
QDH.ndsu.5A.3	17	YLD, GVW	1, 3,4, 5, 6	205.71–208.31	20.17	−2.84	38.36
QDH.ndsu.5D2	20	GVW, YLD	2, 3, 4, 5, 6	11.91–20.91	15.16	2.29	29.93
QDH.ndsu.7B	25	GVW	1, 2, 3, 4, 5*, 6	27.41–31.11	10.25	1.43	17.41
Plant height (PH)							
QPH.ndsu.2A	5	_	1, 4*	128.41–133.11	3.60	1.49	7.68
QPH.ndsu.2D	9	TKW	1, 3, 4, M	151.11–165.71	7.31	2.04	17.2
QPH.ndsu.3B	10	_	3, 6*,M*	184.31–187.71	4.33	1.53	8.55
QPH.ndsu.4A	14	_	2*, M*	175.01–176.01	3.23	−1.70	6.73
QPH.ndsu.5B1	18	_	5,6	32.41–33.21	4.5	−1.81	9.01
QPH.ndsu.6A	22	_	2*, 3, 4	85.51–90.61	5.28	1.83	11.37
QPH.ndsu.7B.1	26	_	1, 3*, 6, M	129.41–130.31	4.94	1.54	9.44
QPH.ndsu.7B.2	24	YLD	4*, 5, M*	24.21–26.21	3.69	1.81	9.36
Grain volume weight (GVW)							
QTW.ndsu.2A.1	4	_	1, 2, 5	100.71–104.31	4.53	−7.73	8.16
QTW.ndsu.2A.2	3	TKW	4, 5, 6*,M*	80.11–82.11	7.14	−6.45	15.93
QTW.ndsu.2B	7	YLD, HD	1, 2, 5, 6, M	84.31–95.61	8.02	−12.25	16.5
QTW.ndsu.4A	13	DH	6	139.91	3.79	5.8	7.22
QTW.ndsu.5A	17	YLD, DH	3, 6	207.01	9.43	20.77	17.79
QTW.ndsu.5D2	20	DH, YLD	3, 6, M	11.91	12.38	−25.22	24.47
QTW.ndsu.7B	25	DH	1, 2, 3, M	29.11–40.11	8.95	−13.28	14.86
QTL and trait	**QTL region**	**Other associated traits†**	**Env.‡**	**Position§ (cM)**	**LOD¶**	**Additive effect**	**R^2 (%)**
Thousand kernel weight (TKW)							
QTKW.ndsu.1A	1	_	4*, 6*	87.61–94.01	3.43	−0.77	7.08
QTKW.ndsu.2A	3	GVW	3,4, M	76.51–78.21	4.36	0.82	9.66
QTKW.ndsu.2D.1	8	_	2, 3*, M*	110.21–111.21	3.73	0.63	7.69
QTKW.ndsu.2D.2	9	PH	1, 4	155.31–155.61	4.06	0.72	8.47
QTKW.ndsu.4A	12	DH	3	58.81	6.82	0.84	14.18
QTKW.ndsu.5B1	19	_	1*, 5*	152.01–153.01	2.72	−0.69	5.61
QTKW.ndsu.6A	21	_	1, 2, 3, 4*, 5, M	65.41–68.21	5.43	0.89	15.22
QTKW.ndsu.7A	23	_	1*, 3*	53.71	2.58	0.49	5.36
Yield (YLD)							
QYL.ndsu.1B	2	_	3,5*, M*	64.21–71.91	3.99	−259.69	8.57
QYL.ndsu.2B	7	GVW, HD	1,2, 3*, M	81.31–83.31	7.22	−209.44	13.94
QYL.ndsu.3B	11	_	4*, M*	202.21–213.81	3.17	−189.99	7.3
QYL.ndsu.5A	17	DH, GVW	3, 6, M	198.61–206.51	11.12	192.14	22.35
QYL.ndsu.5D2	20	GVW, DH	3, 5*, 6, M	11.91–14.91	10.49	−466.60	22.83
QYL.ndsu.7B	24	PH	1,2*	22.21–25.21	5.87	−178.75	13.6

†*PH = plant height, DH = days to heading, YLD = yield, GVW = grain volume weight, TKW = thousand kernel weight.*
‡*Env. = environment, 1 = Carrington 2012, 2 = Minot 2012, 3 = Prosper 2012, 4 = Carrington 2014, 5 = Hettinger 2014, 6 = Prosper 2014, M = mean across environments.*
§*Position represents the peak point of the QTL interval.*
¶*For log of odds (LOD) score.*
**The QTL in that environment was detected above a 2.5 LOD score, but below the threshold score.*

the desirable alleles contributed by Albany. The QTL for PH (*QPH.ndsu.2D*) and TKW (*QTKW.ndsu.2D.2*) were associated with genomic region 9. The QTL for PH had a major effect, while that for TKW had a minor effect. Desired alleles from the QTL were contributed by both t parents. Genomic region 24 harbored QTL for PH (*QPH.ndsu.7B.2*) and YLD (*QYL.ndsu.7B*),

where both QTL had major effects and the desired alleles came from Albany. Genomic region 3 harbored QTL for GVW (*QTW.ndsu.2A.2*) and TKW (*QTKW.ndsu.2A*). Both QTL had major effects, and the desired alleles were also contributed by both parents (**Table 5**; **Supplementary Figure 1**).

Drought Tolerance QTL

A total of 11 consistent QTL, important for drought tolerance, were identified. Among these, six QTL were exclusively detected under drought-prone environments and the remaining five were major constitutive QTL (PV > 10%) identified in both water regimes (**Table 2**; **Supplementary Figure 1**). The QTL *QTW.ndsu.7B*, which is also associated with DH, had a major effect on GVW with a LOD score of up to 8.95. The QTL *QYL.ndsu.2B* and *QYL.ndsu.7B* had major effects on grain yield. The desired alleles from these three major QTL were contributed by Albany, the high-yielding and less drought-tolerant parent. The QTL *QDH.ndsu.2B.1*, which had a LOD score of up to 3.82, controlled 5.2% of PV for DH. The desirable allele for this QTL was also contributed by the parent Albany. Another minor QTL for DH, *QDH.ndsu.5A.2*, had an LOD score of up to 4.09; however, the desired allele was contributed by the resistant parent Reeder. The third minor QTL, *QTKW.ndsu.2D.1*, controlled TKW with PV up to 7.69% and a LOD score of up to 3.73; Reeder contributed the desired allele.

Candidate Genes in Identified QTLs

We identified 3,862 genes (**Supplementary Table 4**; see Text footnote 2) with predicted functions in 26 reported QTLs controlling the abovementioned traits, by using the high-confidence annotated genes in Chinese_Spring_IWGSC_RefSeq1.0. Some of these QTLS such as 3, 7, 17, and 20 are consistent in multiple environments (**Table 5**) for multiple traits; thus, underlaying genes in these QTLs possibly control the shared pathway, resulting in drought tolerance and phenotypic responses with associated traits. The candidate genes underlying the drought-specific QTLs were further mined using the wheatexp *in silico* expression analysis, and we identified 104 genes whose expression was reported to modulate during the vegetative and reproductive stage drought stress (**Supplementary Table 5**; **Supplementary Figure 2**). We also identified a glycosyltransferase (*TraesCS6A01G116400*) encoding gene and a hexosyltransferase (*TraesCS7B01G013300*) encoding gene as a single candidate present in QTLs 21 and 24, respectively. An enrichment of genes encoding for the large subunit of cytoplasmic and chloroplast ribosomal proteins and photosynthesis-associated genes were identified (**Supplementary Table 5**) in this expression-sorted 104 gene list, thus indicating the importance of these class of genes in drought stress.

DISCUSSION

Linkage Map Construction

High-density single-nucleotide polymorphism (SNP) genotyping arrays explore genomic diversity and marker–trait associations very efficiently (Wang et al., 2014). The Infinium iSelect 90K assay (Wang et al., 2014) uses > 81,000 gene-associated SNPs to assess polymorphism in allohexaploid and allotetraploid wheat populations (Wang et al., 2014; Wu et al., 2015; Kumar et al., 2016; Liu et al., 2016; Sapkota et al., 2020). Use of this genotyping tool offers higher genome coverage and resolution in the dissection of wheat's agronomic traits than those used in previous studies (Kirigwi et al., 2007; Muchero et al., 2009; Sayed, 2011; Alexander et al., 2012; Ibrahim et al., 2012a; Kumar et al., 2012; Milner et al., 2016). The main results related to marker density (0.36 cM/marker) or unique locus density (1.84 cM/locus) and genetic map length (3,793.1 cM) observed in this study agreed with the previous studies that used the 90K Infinium iSelect assay for genome mapping (Wang et al., 2014; Kumar et al., 2016). The A genome was found to be the longest, while the D genome was the shortest, which is also in agreement with previous studies (Kumar et al., 2016). The marker order strongly corresponded with several linkage maps developed using the Infinium iSelect 90K SNP assay, as well (Desiderio et al., 2014; Russo et al., 2014; Wang et al., 2014; Kumar et al., 2016; Sapkota et al., 2020).

Four of the chromosomes (1D, 5B, 5D, and 6D) had more than one linkage group. Chromosome 5B had two and chromosomes 1D, 5D, and 6D had three linkage groups. The probable reasons for the fragmentation could be the repeated elements that reside between gene-rich regions or the use of stringent mapping parameters (LOD score > 5 and distance < 40 cM) (Kumar et al., 2016). This fragmentation occurred mostly on the D-genome chromosomes as the Infinium iSelect 90K assay had a poor representation of the D genome (Wang et al., 2014). Further, the D genome is the newest inclusion in the hexaploid wheat genome (dating to around 10,000 years ago) and exhibits fewer polymorphisms than the other genomes according to previous studies (Dubcovsky and Dvorak, 2007).

Use of Secondary Data to Assess Drought Conditions

According to Lanceras et al. (2004), drought can be assessed by variables like weather conditions, soil moisture, and crop conditions over a particular growing season. Rainfall data, which impacts soil moisture, was collected to assess drought conditions for this study. It was obtained from the NDAWN database[3]. The total amount of rainfall was collected from the date of planting to the date of plant physiological maturity in addition to soil residual water moister determined at planting (**Table 1**). The date of physiological maturity was calculated by adding 30 days to DH (Simmons et al., 1914). The 2012 crop season had less rainfall than 2014 in all of the environments, and therefore, the 2012 crop season can be considered as dry, whereas 2014 can be considered as normal season. The YLD data also support this categorization as all of the environments in 2012 had a lower YLD than in 2014.

Use of Agronomic Data to Assess Drought Tolerance

Several studies suggested that drought tolerance can most effectively be incorporated into a breeding program by

[3]https://ndawn.ndsu.nodak.edu/

identifying QTL for YLD or YLD-related traits (Lanceras et al., 2004; Alexander et al., 2012). Yield is the trait of ultimate interest to breeders to develop adapted cultivars. In this study, YLD had a positive significant correlation with all studied traits except for DH, which was negatively correlated with YLD. In general, delayed DH gives a plant the opportunity to produce more photosynthates (the product of photosynthesis) and hence a higher YLD. However, in this study, YLD was higher for early (reduced DH) compared to late genotypes. This is usually well known in environments where terminal drought is common. In our study, snow accumulated during winter in the US North Central Plains. It is a major source of soil moisture in this region, and this soil moisture depletes with time. Therefore, the late genotypes (high DH values) were affected by drought, which resulted in reduced YLD. Except for PH, increased values were desirable for the rest of the agronomic traits as they have a positive correlation with YLD. A taller genotype (high PH) has the potential to produce more photosynthates and, therefore, should give more yield, but it often tends to lodge and compromises yield.

QTL for YLD

Grain YLD is considered the most significant trait to plant breeders. It is the result of all the phases of vegetative and reproductive development, therefore reflecting the contribution of all favorable alleles involved directly or indirectly in the formation of wheat kernels. It is therefore influenced by edaphic and aerial environments (Quarrie et al., 2006). QTL for YLD in wheat have been reported in several studies (McCartney et al., 2005; Quarrie et al., 2006; Kirigwi et al., 2007; Li et al., 2007, 2015; Maccaferri et al., 2008; Azadi et al., 2014; Cui et al., 2014; Edae et al., 2014; Gao et al., 2015; Narjesi et al., 2015; Milner et al., 2016). In this study, we revealed six QTL for yield, both major and minor, contributed by both parents of the RIL population, and the confirmation of the quantitative nature of inheritance of YLD. The QTL *QYL.ndsu.2B* on chromosome 2B at 81.31–83.31 cM identified in all the drought-prone environments is close to the QTL (*QGy.ubo-2B*) identified by Milner et al. (2016) in the same region. This QTL can be confirmed as drought-tolerant as it contributed to YLD in all environments with low rainfall. Narjesi et al. (2015) reported a YLD QTL at 8.5 cM on chromosome 5D. Our study also identified a QTL *QYL.ndsu.5D2* on the same chromosome, but at 11.91–12.91 cM on the second linkage group. However, considering the gaps between the linkage groups on the chromosome, the position of the QTL should be around the middle of the chromosome and therefore is most likely different from the one identified by Narjesi et al. (2015). Maccaferri et al. (2008) identified a YLD QTL (*QYld.idw-7B*) at 0 cM on chromosome 7B that could be the same QTL as *QYL.ndsu.7B* identified at 22.21–25.21 cM on the same chromosome. The closest reported QTL of *QYL.ndsu.1B* on chromosome 1B at 64.21–71.91 cM was *QYd-1B.1*, identified on the same chromosome at 23–28 cM (Cui et al., 2014). The QTL *QYld.abrii-3B.4* (Azadi et al., 2014) identified on chromosome 3B at 92.3 cM seemed to be different than the QTL *QYL.ndsu.3B* in this study. Similarly, the QTL *QYL.ndsu.5D2* and *QYL.ndsu.5A* are most likely to be novel QTL as no QTL were earlier reported around these positions.

QTL for TKW

Thousand kernel weight is one of the three major components of YLD. It is also important for grain quality, as larger and uniformly sized kernels are visually attractive, affecting GVW and commanding a higher market price (Ramya et al., 2010). Several studies have reported QTL related to wheat TKW (McCartney et al., 2005; Huang et al., 2006; Breseghello and Sorrells, 2007; Kuchel et al., 2007; Li et al., 2007, 2015; Zhang et al., 2008; Sun et al., 2009; Ramya et al., 2010; Azadi et al., 2014; Simmonds et al., 2014; Wei et al., 2014; Tadesse et al., 2015; Zanke et al., 2015; Kumar et al., 2016, 2019). This study revealed eight QTL having both major and minor effects for TKW, indicating its quantitative nature of inheritance. McCartney et al. (2005) identified the QTL *QGwt.crc-2A* occupying the same position as the QTL *QTKW.ndsu.*2A we identified in this study. The QTL *qTgw2A* (Wei et al., 2014) and *QTgw.abrii-4A.2* (Zhang et al., 2008) also occupied the same location. The QTL *QTgw.abrii-2D1.3* (Azadi et al., 2014) and *QTKW.ndsu.2D.2* identified in this study seemed to be the same QTL, occupying the same position on chromosome 2D. Similarly, the *QTKW.ndsu.4A* we identified on chromosome 4A seems to be in the same location as the QTL *QTgw.abrii-4A.2* (Azadi et al., 2014). The QTL *QTKW.ndsu.6A* was identified in all of the drought-prone environments, indicating its strong relationship with tolerance to drought. This QTL, however, occupied the same location as QTL*qTgw6A2* (Wei et al., 2014). Another QTL, *QTKW.ndsu.7A*, was also identified in the two drought-prone environments and could be comparable to *qTgw7A* (Wei et al., 2014) due to their proximity. The QTL *QTKW.caas-1A.1* (Li et al., 2015) and *QTKW.ndsu.1A* were most likely to be the same QTL since they were found in the same genomic region. However, there were no previous reports on QTL that corresponded with the QTL *QTKW.ndsu.2D.1* and *QTKW.ndsu.5B1*, indicating the probability that they are novel. The QTL *QTKW.ndsu.2D.1* could be very important for drought-tolerance breeding as it was identified in two of the drought-prone environments.

QTL for GVW

Grain volume weight is an important trait to wheat breeders as it impacts flour yield during milling (Rustgi et al., 2013). Quantitative trait loci for GVW were reported in several previous studies (McCartney et al., 2005, 2007; Huang et al., 2006; Narasimhamoorthy et al., 2006; Breseghello and Sorrells, 2007; Kuchel et al., 2007; Zhang et al., 2008; Sun et al., 2009; Rustgi et al., 2013; Hill et al., 2015; Tadesse et al., 2015; Kumar et al., 2016, 2019). In this study, we revealed seven QTL with both major and minor effects, indicating the quantitative nature of inheritance of GVW. The QTL identified in this study on chromosome 7B (*QTW.ndsu.7B*) at 29.11–40.11 cM was identified in all of the drought-prone environments, indicating its potential for drought tolerance. However, this QTL seemed to be close and similar to the QTL (*QTw.sdau-7B*) that Sun et al. (2009) identified. McCartney et al. (2005) identified a QTL, *QTwt.crc-2B*, linked

with the marker *Xbarc183* at 96.7 cM on chromosome 7B that, according to the GrainGenes database, seemed to be the same as the QTL *QTW.ndsu.2B* identified in this study at 84.31–95.61 cM. This QTL was identified in two of the drought-prone environments. In the same study, McCartney et al. (2005) identified another QTL *QTwt.crc-5D*, between SSR markers *Xgdm63–Xwmc765* and positioned between 95 and 214.26 cM, according to the GrainGenes database. The QTL in this study, *QTW.ndsu.5D2*, could be the same as the later QTL as it is also located in the same genomic region. The nearest reported QTL to *QTW.ndsu.5A* was *QTw.hwwgr-5AS* (Li et al., 2016), which seemed to be a different QTL. The QTL *QTw.sdau-2A* (Sun et al., 2009) located between SSR markers *Xwmc181a* and *Xubc840c* seemed to be the same QTL as the QTL *QTW.ndsu.2A.2* identified in this study. However, in our study, we identified the QTL *QTW.ndsu.4A* QTL which does not correspond to any QTL reported previously and therefore is a novel QTL.

QTL for DH

As previously indicated, DH can be critical for drought tolerance, particularly in regions where late drought is prevalent. Therefore, many studies have been conducted and have identified many QTL for DH (Kato et al., 1999; Sourdille et al., 2000; Shindo et al., 2003; Xu et al., 2005; Griffiths et al., 2009; Alexander et al., 2012; Kamran et al., 2013; Bogard et al., 2014; Zanke C. et al., 2014; Guedira et al., 2016; Milner et al., 2016; Kumar et al., 2019). According to these studies, the genetic factors controlling DH are vernalization and photoperiod sensitivities, and earliness *per se* (Shindo et al., 2003). In general, vernalization divides wheat cultivars into two groups. Winter wheat needs cold temperatures (vernalization) to initiate flowering, while spring wheat does not need cold temperatures. Wheat is usually photosensitive and adapted to long-day conditions. Therefore, transitioning from vegetative to reproductive stages (spike emergence) is very late unless genotypes are exposed to long days. However, some genotypes are photoperiod insensitive and therefore can switch to reproductive stage (spike emergence and flowering) even in short days. On the other hand, earliness *per se* is the only environment-independent genetic factor controlling earliness (Shindo et al., 2003).

The present study revealed several major and minor QTL controlling DH, which confirms its quantitative nature of inheritance. Four major QTL (*QDH.ndsu.5A.3*, *QDH.ndsu.5D2*, *QDH.ndsu.7B*, and *QDH.ndsu.4A.2*) were found consistently under all drought conditions. The earliness *per se* QTL *QEet.ocs.5A.2* (Kato et al., 1999) on chromosome 5AL and the QTL identified in our study, *QDH.ndsu.5A.3* at 205.71–208.31 cM, occupy the same location and therefore may represent the same QTL. Similarly, the QTL *QDH.ndsu.4A.2* we identified on chromosome 4A may have corresponded with the QTL reported by McCartney et al. (2005). A relatively minor QTL, *QDH.ndsu.4A.1*, identified at 47.51 cM, could be the same as *QFlt.dms-4A.1* (Kamran et al., 2013). Sourdille et al. (2000) reported a QTL for earliness *per se* on chromosome 7BS, explaining 7.3% to 15.3% of PV. This and the QTL identified in our study on chromosome 7B could represent the same QTL due to their position in the same genomic region. In the same study (Sourdille et al., 2000), a QTL on the long arm of chromosome 5D for earliness *per se* was reported, and this may coincide with the QTL *QDH.ndsu.5D2* identified in this study.

QTL for PH

Plant height is considered crucial in wheat breeding programs as it relates to many important agronomic traits such as lodging resistance and a high harvest index. For example, the dwarfism gene from Nonglin-10 played a vital role in wheat breeding programs during the Green Revolution of the 1960s (Liu et al., 2011). This study showed that PH had a positive correlation with DH, whereas DH had a negative correlation with YLD. Therefore, it could be stated that reduced PH is desirable for higher YLD. Quantitative trait loci for PH have been reported in several previous studies (McCartney et al., 2005; Pushpendra et al., 2007; Liu et al., 2011; Huang et al., 2012; Zanke C.D. et al., 2014; Gao et al., 2015; Li et al., 2015; Narjesi et al., 2015; Milner et al., 2016; Singh et al., 2016). In this study, eight QTL were identified for PH, confirming the findings of Huang et al. (2012) who identified seven QTL for this trait. The QTL they identified on chromosome 2D at 144 cM and chromosome 5B at 64.67 cM could represent the same QTL identified in this study on chromosome 2D at 151.11–165.71 cM (*QPH.ndsu.2D*) and on chromosome 5B at 32.41–33.21 cM (*QPH.ndsu.5B1*), respectively. The QTL *QPH.ndsu.2D* for PH was identified in two drought-prone environments, indicating its potential to tolerate drought. Milner et al. (2016) identified a QTL (*Qph.ubo-7B*) for PH on chromosome 7B at 138.4 cM, which could be the same QTL (*QPH.ndsu.7B.1*) identified in this study on the same chromosome at 129.41–130.31 cM. This QTL was expressed in the drought-prone environments and thus could be useful for drought tolerance. Zanke C.D. et al. (2014) identified a QTL for PH at 93.5 cM on chromosome 6A that could be comparable with this study's *QPH.ndsu.6A* at 85.51–90.61 cM on the same chromosome. This QTL was also identified in the two drought-prone environments. Previously, Zanke C.D. et al. (2014) identified another QTL at 36 cM on chromosome 7B for the same trait that could be comparable to QTL *QPH.ndsu.7B.2* identified in this study on the same chromosome at 24.21–26.21 cM. Similarly, they identified a QTL at 176.5 cM on chromosome 3B for PH, whereas this study identified a QTL at 184.31–187.71 cM for it on the same chromosome. They also identified a QTL at 117.2 cM on chromosome 2A, whereas in this study, we identified the QTL *QPH.ndsu.2A* on the same chromosome but at 128.41–133.11 cM.

Pleiotropic QTL

The associations between traits in correlation studies could be justified by the co-localized or pleiotropic QTL. These co-localized QTL could be of great value to breeders if the desirable alleles come from the same parent. Desirable alleles from three genomic regions (7, 20, and 25) came from parent Albany (**Table 5**; **Supplementary Figure 1**). These QTL primarily have a major effect on YLD and YLD-related traits, making them

even more important to crop improvement. The parent Reeder contributed all of the desirable alleles in three genomic regions (13, 17, and 24) (**Table 5**; **Supplementary Figure 1**). Most of these QTL also had the major effect on YLD and YLD-related traits. The remaining co-localized QTL from three genomic regions did not contain desirable alleles from the same parents.

QTL for Drought Tolerance

The QTL identified on chromosome 7B (*QTW.ndsu.7B*) at 29.11–40.11 cM appears to relate strongly with drought tolerance as it was identified in all environments under drought conditions (**Table 2**). This QTL seems to coincide with a QTL that Sun et al. (2009) identified (*QTw.sdau-7B*) previously. The putative drought-tolerant QTL, *QYL.ndsu.7B*, was also identified very close to another major QTL, *QTW.ndsu.7B*, which is also associated with drought tolerance, indicating the importance of this genomic region to control drought tolerance. This result confirms the previous findings by Alexander et al. (2012), who found a QTL, *Qdt.ksu-7B*, located on chromosome 7B at 34.7 cM with significant effect on drought tolerance. Another putative major QTL, *QYL.ndsu.2B*, was detected in this study and corresponded with the QTL *QCrs*-, which was reported to have a negative effect on the trait of interest under both drought and control conditions (Ibrahim et al., 2012b). In the current study, however, the QTL was identified only in the environments with drought conditions. Similarly, the QTL *QDH.ndsu.5A.2* detected in our study occupied the same location as the previously reported QTL QHea + (Ibrahim et al., 2012a). In the latter study, the QTL *Qhea* + improved the trait of interest in both well-watered and drought conditions. However, in our study, *QDH.ndsu.5A.2* improved the trait of interest only under drought conditions. Ibrahim et al. (2012b) reported four QTL on chromosome 2D around 50 cM that improved the trait of interest under drought conditions. However, none of these reported QTL seemed to correspond with the QTL *QTKW.ndsu.2D.1* identified in this study. Therefore, this QTL can be considered as novel and more studies are needed to elucidate its importance in drought tolerance.

An additional QTL for DH, *QDH.ndsu.5A.3*, could be a constitutive QTL for drought tolerance since it was identified consistently in both drought and non-drought condition environments. This QTL could occupy the same genomic region as the earliness *per se* QTL, *Qeet.ocs.5A.2* (Kato et al., 1999). Similarly, another constitutive QTL for drought tolerance, *QDH.ndsu.5D2*, corresponded with a QTL for earliness *per se* located on the long arm of chromosome 5D (Sourdille et al., 2000). A constitutive QTL for drought tolerance through TKW was identified on chromosome 6A, which most likely represents the QTL *qTgw6A2* (Wei et al., 2014). Also, a constitutive drought-tolerant QTL, *QTW.ndsu.2B*, was identified for GVW, which could be the same QTL as *QTwt.crc-2B* (McCartney et al., 2005). All these QTL appear to play a crucial role directly or indirectly in drought tolerance.

Candidate Gene Analysis

A total of 3,682 candidate genes were identified for the 26 QTLs (**Table 5**; **Supplementary Table 4**; see Text footnote 2) reported in this study. For QTL region 3, controlling the GVW and TKW traits in multiple environments, we identified only 21 candidate high-confidence annotated genes. Among these 21 identified candidate genes, adjacently located MYB transcription factor (*TraesCS2A01G097200*) and hexosyltransferase (*TraesCS2A01G097400*) could be the most important candidates due to the fact that in wheat the *MYB* transcriptional activator-coding gene, i.e., *TaMYB13*, was previously shown to be positively related with sucrose:sucrose 1-fructosyltransferase (*1-SST*) and sucrose:fructan 6-fructosyltransferase (*6-SFT*) mRNA levels, thus controlling the levels of fructan, which serves as a key water-soluble carbohydrate as a carbon source for grain filling (Xue et al., 2011). This locus also harbors the two NLR (nucleotide binding leucine-rich repeats) disease resistance proteins (*TraesCS2A01G096200*, *TraesCS2A01G096300*). NLRs act as immunity receptors in plants (Solanki et al., 2019; Walkowiak et al., 2020), which may indirectly play a role in the overall health of mature grain kernels. Another consistent QTL region was 7, affecting the YLD, GVW, and HD traits and harboring 978 annotated high-confidence genes (**Supplementary Table 4**; see Text footnote 2). YLD, GVW, and HD traits were also found to be associated with the QTL region 17 containing 114 annotated genes. QTL-20 contains 224 annotated genes (**Supplementary Table 4**; see Text footnote 2). However, there were no candidate genes present in the annotated genome between the QTL-25-associated flanking markers; thus, at this point we can only speculate that this region may control the trait of GVW and DH by harboring the trans-acting elements of gene regulation. The QTL region-24 correlated with YLD and PH traits and was found to be consistent in multiple environments. Interestingly, we only found one gene in QTL-24 annotated to encode for a hexosyltransferase (*TraesCS7B01G013300*), a homolog of *TraesCS2A01G097400*, a gene we proposed to be important for GVW and TKW traits controlled by QTL-3. Similarly, only one gene, *TraesCS6A01G116400*, in QTL-21 was found putatively encoding for a glycosyltransferase. Thus, these two genes possibly represent a major effect on drought stress. We also used an *in silico* expression analysis and short-listed 104 candidate genes among the all-underlying genes in drought-related QTL. The most frequent functional class in this differential regulation prioritized 104 candidate genes which associated with the ribosomal proteins and photosynthesis-associated proteins (**Supplementary Table 5**; **Supplementary Figure 2**), indicating a crucial role of this class of genes in drought stress.

CONCLUSION

Understanding the genetic basis of drought tolerance in wheat is of immense value for developing drought-tolerant wheat varieties for world food security. In this study, a population developed from a cross between elite lines was used to elucidate the genetic factors involved in the control of drought tolerance in HRSW in northern United States. Multi-environment phenotypic data on yield-related traits, combined with a high-density Infinium

90K SNP-based genetic map, identified a total of nine QTL for DH, eight QTL for PH, seven QTL for GVW, eight QTL for TKW, and six QTL for YLD. The genetic dissection identified 11 consistent QTL related to drought tolerance in this population. These included six QTL exclusively associated with drought environments and five constitutive QTL (associated with both, drought, and normal conditions). Major QTL for drought tolerance were identified on chromosomes 7B, 2B, 5A, 5D, and 6A. One novel QTL for drought tolerance was identified on chromosome 2D. The ribosomal proteins and chloroplast photosynthesis-associated proteins were the major class found to be abundant in the 104 expression-sorted candidate genes in drought QTLs. Along with the single-candidate genes *TraesCS6A01G116400* and *TraesCS7B01G013300* at QTL 21 and 24, respectively, expression-sorted genes at six drought QTLs provide a valuable resource to breed for drought resistance. More importantly, the desirable alleles for several major loci were contributed by the high-yielding parent that was apparently susceptible to drought. This suggests that the high-yielding cultivars may contribute desirable QTL alleles for drought tolerance. Therefore, exploring high-yielding but seemingly drought-susceptible germplasms in the development of drought-tolerant cultivars is paramount. Although we were successful in identifying many major and minor QTL, future studies could focus on using other approaches (Wang et al., 2016) to detect possible minor effect QTL associated with wheat drought tolerance using different germplasms under different environments and types of drought.

The knowledge gained and closely linked markers associated with the major QTL and candidate genes identified in this study could be of immense value for understanding the genetic control of drought and can be valuable in marker-assisted breeding programs aimed at improving drought tolerance in wheat. The high-density maps that were developed also offer a better starting platform for the fine mapping and ultimately map-based cloning of major and stable loci identified in this study.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SR: data collection, analysis, and major write-up of the manuscript. AK: data analysis and manuscript write-up. SM: data collection and analysis. SSa: data analysis and write-up. MA: data and manuscript review. EE: data analysis and manuscript review. SK: data analysis and manuscript review. RS: data analyses. AM: data and manuscript review. SSo: data analysis and write-up. MM: conceptualized and designed the experiment, phenotypic data collection, and manuscript write-up and review. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Late Dr. J. J. Hammond, Dr. A. El-Fatih ElDolfey, A. Walz, C. Cossette, D. Olsen, K. Dickey, J. Underdahl, K. Beck, K. McMonagle, K. Whitney, M. Abdallah, and T. Selland for their help and support during this project. The authors would also like to thank the technical support team in the NDSU Research Centers at Hettinger, Minot, and Williston, and at the Agronomy Farm at Casselton and Prosper for their help and support in conducting the field trials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.656037/full#supplementary-material

REFERENCES

Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., et al. (2018). Linking the international wheat genome sequencing consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.* 19:111. doi: 10.1186/s13059-018-1491-4

Alexander, L. M., Kirigwi, F. M., Fritz, A. K., and Fellers, J. P. (2012). Mapping and quantitative trait loci analysis of drought tolerance in a Spring wheat population using amplified fragment length polymorphism and diversity array technology markers. *Crop Sci.* 52, 253–261. doi: 10.2135/cropsci2011.05.0267

Anderson, J. A., Linkert, G. L., Busch, R. H., Wiersma, J. J., Kolmer, J. A., and Jin, Y. (2009). Registration of "RB07" wheat. *J. Plant Regist.* 3, 175–180. doi: 10.3198/jpr2008.08.0478crc

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Araus, J. L., Slafer, G. A., Reynolds, M. P., and Royo, C. (2002). Plant breeding and drought in C3 cereals: what should we breed for? *Ann. Bot.* 89, 925–940. doi: 10.1093/aob/mcf049

Araus, J. L., Slafer, G. A., Royo, C., and Serret, M. D. (2008). Breeding for yield potential and stress adaptation in cereals. *Crit. Rev. Plant Sci.* 27, 377–412. doi: 10.1080/07352680802467736

Azadi, A., Mardi, M., Hervan, E. M., Mohammadi, S. A., Moradi, F., and Tabatabaee, M. T. (2014). QTL mapping of yield and yield components under normal and salt-stress conditions in bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol. Rep.* 33, 102–120. doi: 10.1007/s11105-014-0726-0

Blake, V. C., Birkett, C., Matthews, D. E., Hane, D. L., Bradbury, P., and Jannink, J.-L. (2016). The Triticeae Toolbox: combining phenotype and genotype data to advance small-grains breeding. *Plant Genome* 9, 1–10. doi: 10.3835/plantgenome2014.12.0099

Blum, A. (1988). *Plant Breeding for Stress Environments*. Boca Raton, FL: CRC Press.

Blum, A. (2005). Drought resistance, water-use efficiency, and yield potential-are they compatible, dissonant, or mutually exclusive? *Aust. J. Agric. Res.* 56, 1159–1168. doi: 10.1071/ar05069

Bogard, M., Ravel, C., Paux, E., Bordes, J., Balfourier, F., and Chapman, S. C. (2014). Predictions of heading date in bread wheat (*Triticum aestivum* L.) using QTL-based parameters of an ecophysiological model. *J. Exp. Bot.* 65, 5849–5865. doi: 10.1093/jxb/eru328

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170, 2172–2186. doi: 10.1104/pp.15.01667

Breseghello, F., and Sorrells, M. E. (2007). QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. *Field Crops Res.* 10, 172–179. doi: 10.1016/j.fcr.2006.11.008

Cui, F., Fan, X., Zhao, C., Zhang, W., Chen, M., and Li, J. (2014). A novel genetic map of wheat: utility for mapping QTL for yield under different nitrogen treatments. *BMC Genet.* 15:57. doi: 10.1186/1471-2156-15-57

de Givry, S., Bouchez, M., Chabrier, P., Milan, D., and Schiex, T. (2005). CARTHAGENE: multipopulation integrated genetic and radiated hybrid mapping. *Bioinformatics* 21, 1703–1704. doi: 10.1093/bioinformatics/bti222

Desiderio, F., Guerra, D., Rubiales, D., Piarulli, L., Pasquini, M., and Mastrangelo, A. M. (2014). Identification and mapping of quantitative trait loci for leaf rust resistance derived from a tetraploid wheat *Triticum dicoccum* accession. *Mol. Breed.* 34, 1659–1675. doi: 10.1007/s11032-014-0186-0

Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* 316, 1862–1866. doi: 10.1126/science.1143986

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* 127, 791–807. doi: 10.1007/s00122-013-2257-8

Frohberg, R. C., Stack, R. W., Olson, T., Miller, J. D., and Mergoum, M. (2006). Registration of "Alsen.". *Crop Sci.* 46, 2311–2312. doi: 10.2135/cropsci2005.12.0501

Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese Spring. *Front. Plant Sci.* 6:1099.

Griffiths, S., Simmonds, J., Leverington, M., Wang, Y., Fish, L., and Sayers, L. (2009). Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. *Theor. Appl. Genet.* 119, 383–395. doi: 10.1007/s00122-009-1046-x

Guedira, M., Xiong, M., Hao, Y. F., Johnson, J., Harrison, S., and Marshall, D. (2016). Heading date QTL in winter wheat (*Triticum aestivum* L.) coincide with major developmental genes Vernalization1 and Photoperiod1. *PLoS One* 11:e0154242. doi: 10.1371/journal.pone.0154242

Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Langridge, P., Bacic, A., et al. (2015). Detection of QTL for metabolic and agronomic traits in wheat with adjustments for variation at genetic loci that affect plant phenology. *Plant Sci.* 233, 143–154. doi: 10.1016/j.plantsci.2015.01.008

Huang, B. E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., Morell, M. K., et al. (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol. J.* 10, 826–839. doi: 10.1111/j.1467-7652.2012.00702.x

Huang, X. Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D. G., Noll, J. S., et al. (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 113, 753–766. doi: 10.1007/s00122-006-0346-7

Ibrahim, S. E., Schubert, A., Pillen, K., and Léon, J. (2012a). Comparison of QTLs for drought tolerance traits between two advanced backcross populations of spring wheat. *Int. J. AgriSci.* 2, 216–227.

Ibrahim, S. E., Schubert, A., Pillen, K., and Léon, J. (2012b). QTL analysis of drought tolerance for seedling root morphological traits in an advanced backcross population of spring wheat. *Int. J. Agrisci.* 2, 619–629.

Kamran, A., Iqbal, M., Navabi, A., Randhawa, H., Pozniak, C., and Spaner, D. (2013). Earliness per se QTLs and their interaction with the photoperiod insensitive allele Ppd-D1a in the Cutler × AC barrie spring wheat population. *Theor. Appl. Genet.* 126, 1965–1976. doi: 10.1007/s00122-013-2110-0

Kato, K., Miura, H., and Sawada, S. (1999). Detection of an earliness per se quantitative trait locus in the proximal region of wheat chromosome 5AL. *Plant Breed.* 118, 391–394. doi: 10.1046/j.1439-0523.1999.00395.x

Kirigwi, F. M., Van Ginkel, M., Brown-Guedira, G., Gill, B. S., Paulsen, G. M., and Fritz, A. K. (2007). Markers associated with a QTL for grain yield in wheat under drought. *Mol. Breed.* 20, 401–413. doi: 10.1007/s11032-007-9100-3

Kosambi, D. (1944). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Kuchel, H., Williams, K. J., Langridge, P., Eagles, H. A., and Jefferies, S. P. (2007). Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor. Appl. Genet.* 115, 1029–1041. doi: 10.1007/s00122-007-0629-7

Kumar, A., Mantovani, E. E., Seetan, R., Soltani, A., Echeverry-Solarte, M., and Jain, S. (2016). Dissection of genetic factors underlying wheat kernel shape and size in an Elite × Nonadapted cross using a high density SNP linkage map. *Plant Genome* 9, 1–22.

Kumar, A., Mantovani, E. E., Simsek, S., Jain, S., Elias, E. M., and Mergoum, M. (2019). Genome wide genetic dissection of wheat quality and yield related traits and their relationship with grain shape and size traits in an elite × non-adapted bread wheat cross. *PLoS One* 14:e0221826. doi: 10.1371/journal.pone.0221826

Kumar, S., Sehgal, S. K., Kumar, U., Prasad, P. V. V., Joshi, A. K., and Gill, B. S. (2012). Genomic characterization of drought tolerance-related traits in spring wheat. *Euphytica* 186, 265–276. doi: 10.1007/s10681-012-0675-3

Lanceras, J., Pantuwan, G., Jongdee, B., and Toojinda, T. (2004). Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiol.* 135, 384–399. doi: 10.1104/pp.103.035527

Lander, E. S., and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199. doi: 10.1093/genetics/121.1.185

Lanning, S. P., Carlson, G. R., Lamb, P. F., Nash, D., Wichman, D. M., and Kephart, K. D. (2011). Registration of "Duclair" hard red spring wheat. *J. Plant Regist.* 5, 349–352. doi: 10.3198/jpr2011.02.0064crc

Li, C., Bai, G., Chao, S., Carver, B., and Wang, Z. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop J.* 4, 1–11. doi: 10.1016/j.cj.2015.10.002

Li, S., Jia, J., Wei, X., Zhang, X., Li, L., and Chen, H. (2007). An intervarietal genetic map and QTL analysis for yield traits in wheat. *Mol. Breed.* 20, 167–178. doi: 10.1007/s11032-007-9080-3

Li, X., Xia, X., Xiao, Y., He, Z., Wang, D., and Trethowan, R. (2015). QTL mapping for plant height and yield components in common wheat under water limited and full irrigation environments. *Crop Pasture Sci.* 8, 660–670. doi: 10.1071/cp14236

Liu, G., Xu, S. B., Ni, Z. F., Xie, C. J., Qin, D. D., and Li, J. (2011). Molecular dissection of plant height QTLs using recombinant inbred lines from hybrids between common wheat (*Triticum aestivum* L.) and spelt wheat (*Triticum spelta* L.). *Chinese Sci. Bull.* 56, 1897–1903. doi: 10.1007/s11434-011-4506-z

Liu, S., Assanga, S. O., Dhakal, S., Gu, X., Tan, C. T., and Yang, Y. (2016). Validation of chromosomal locations of 90K array single nucleotide polymorphisms in US wheat. *Crop Sci.* 56, 364–373. doi: 10.2135/cropsci2015.03.0194

Maccaferri, M., Sanguineti, M. C., Corneti, S., Ortega, J. L. A., Ben Salem, M., and Bort, J. (2008). Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics* 178, 489–511. doi: 10.1534/genetics.107.077297

Malik, S., Malik, T. A., and Rahman, M. (2015). Genetic mapping of potential QTLs associated with drought tolerance in wheat. *J. Anim. Plant Sci.* 25, 1032–1040.

McCartney, C. A., Somers, D. J., Fedak, G., DePauw, R. M., Thomas, J., and Fox, S. L. (2007). The evaluation of FHB resistance QTLs introgressed into elite Canadian spring wheat germplasm. *Mol. Breed.* 20, 209–221. doi: 10.1007/s11032-007-9084-z

McCartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., and Noll, J. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452x'AC Domain. *Genome* 48, 870–883. doi: 10.1139/g05-055

Mergoum, M., Frohberg, R. C., Miller, J. D., Olson, T., and Rasmussen, J. B. (2005a). Registration of "Dapps" wheat. *Crop Sci.* 45, 420–421.

Mergoum, M., Frohberg, R. C., Miller, J. D., and Stack, R. (2005b). Registration of "Steele-ND" wheat. *Crop Sci.* 45, 1163–1164. doi: 10.2135/cropsci2004.308cv

Mergoum, M., Frohberg, R. C., Stack, R. W., Olson, T., Friesen, T. L., and Rasmussen, J. B. (2006). Registration of "Glenn" wheat. *Crop Sci.* 46, 473–474. doi: 10.2135/cropsci2005.0287

Mergoum, M., Frohberg, R. C., Stack, R. W., Rasmussen, J. W., and Friesen, T. L. (2008). Registration of "Faller" spring wheat. *J. Plant Regist.* 2, 224–229. doi: 10.3198/jpr2008.03.0166crc

Mergoum, M., Frohberg, R. C., Stack, R. W., Simsek, S., Adhikari, T. B., and Rasmussen, J. (2013). "Prosper": a high-yielding hard red spring wheat cultivar adapted to the North Central Plains of the USA. *J. Plant Regist.* 7, 75–80. doi: 10.3198/jpr2012.05.0271crc

Mergoum, M., Frohberg, R., Rasmussen, J., Friesen, T. L., Hareland, G. A., and Simsek, S. (2009). Breeding for CLEARFIELD herbicide tolerance: registration of "ND901CL" spring wheat. *J. Plant Regist.* 3, 170–174. doi: 10.3198/jpr2008.08.0489crc

Mergoum, M., Simsek, S., Zhong, S., Acevedo, M., Friesen, T. L., and Alamri, S. (2016). 'Elgin-ND' spring wheat: an adapted cultivar to north central plains

of the USA with high agronomic and quality performance. *J. Plant Regist.* 10, 130–134. doi: 10.3198/jpr2015.07.0044crc

Mergoum, M., Simsek, S., Zhong, S., Acevedo, M., Friesen, T. L., and Singh, P. K. (2014). "Velva" spring wheat: an adapted cultivar to North-central plains of the United States with high agronomic and quality performance. *J. Plant Regist.* 8, 32–37. doi: 10.3198/jpr2013.06.0026crc

Milner, S. G., Maccaferri, M., Huang, B. E., Mantovani, P., Massi, A., Frascaroli, E., et al. (2016). A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. durum). *Plant Biotechnol. J.* 14, 735–748. doi: 10.1111/pbi.12424

Muchero, W., Ehlers, J. D., Close, T. J., and Roberts, P. A. (2009). Mapping QTL for drought stress-induced premature senescence and maturity in cowpea (*Vigna unguiculata* L.). *Theor. Appl. Genet.* 118, 849–863. doi: 10.1007/s00122-008-0944-7

Narasimhamoorthy, B., Gill, B. S., Fritz, A. K., Nelson, J. C., and Brown-Guedira, G. L. (2006). Advanced backcross QTL analysis of a hard winter wheat x synthetic wheat population. *Theor. Appl. Genet.* 112, 787–796. doi: 10.1007/s00122-005-0159-0

Narjesi, V., Mardi, M., Hervan, E. M., and Azadi, A. (2015). Analysis of quantitative trait loci (QTL) for grain yield and agronomic traits in wheat (*Triticum aestivum* L.) Under normal and salt- stress conditions. *Plant Mol. Biol.* 33, 2030–2040. doi: 10.1007/s11105-015-0876-8

North Dakota Agricultural Weather Network (NDAWN) (2015). Available at: https://ndawn.ndsu.nodak.edu. (accessed November 16, 2015).

Peleg, Z., Fahima, T., Krugman, T., Abbo, S., Yakir, D., Korol, A. B., et al. (2009). Genomic dissection of drought resistance in durum wheat?× wild emmer wheat recombinant inbreed line population. *Plant Cell Environ.* 32, 758–779. doi: 10.1111/j.1365-3040.2009.01956.x

Pfeiffer, W. H., Trethowan, R. M., Ginkel, M. V., Ortiz, M. I., and Rajaram, S. (2005). "Breeding for abiotic stress tolerance in wheat," in *Abiotic Stresses: Plant Resistance Through Breeding and Molecular Approaches*, ed. M. A. Harris (New York, NY: The Haworth Press).

Pinter, J. P. J., Zipoli, G., Reginato, R. J., Jackson, R. D., and Idso, S. B. (1990). Canopy temperature as an indicator of differential water use and yield performance among wheat cultivars. *Agric. Water Manag.* 18, 35–48. doi: 10.1016/0378-3774(90)90034-v

Pushpendra, K. G., Harindra, S. B., Pawan, L. K., Neeraj, K., Ajay, K., Reyazul, R. M., et al. (2007). QTL analysis for some quantitative traits in bread wheat. *J. Zhejiang Univ. Sci. B* 8, 807–814. doi: 10.1631/jzus.2007.b0807

Quarrie, S. A., Pekic Quarrie, S., Radosevic, R., Rancic, D., Kaminska, A., and Barnes, J. D. (2006). Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. *J. Exp. Bot.* 57, 2627–2637. doi: 10.1093/jxb/erl026

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., and Venturini, L. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:6403.

Ramya, P., Chaubal, A., Kulkarni, K., Gupta, L., Kadoo, N., and Dhaliwal, H. S. (2010). QTL mapping of 1000-kernel weight, kernel length, and kernel width in bread wheat (*Triticum aestivum* L.). *J. Appl. Genet.* 51, 421–429. doi: 10.1007/bf03208872

Russo, M. A., Ficco, D. B. M., Laidò, G., Marone, D., Papa, R., and Blanco, A. (2014). A dense durum wheat × T. dicoccum linkage map based on SNP markers for the study of seed morphology. *Mol. Breed.* 34, 1579–1597. doi: 10.1007/s11032-014-0181-5

Rustgi, S., Shafqat, M. N., Kumar, N., Baenziger, P. S., Ali, M. L., Dweikat, I., et al. (2013). Genetic dissection of yield and its component traits using high-density composite map of wheat chromosome 3A: bridging gaps between QTLs and underlying genes. *PLoS One* 8:e70526. doi: 10.1371/journal.pone.0070526

Sapkota, S., Mergoum, M., Kumar, A., Fiedler, J. D., Johnson, J., and Bland, D. (2020). A novel adult plant leaf rust resistance gene Lr2K38 mapped on wheat chromosome 1AL. *Plant Genome* 13:e20061.

SAS Institute (2004). *SAS Online Doc, v. 9.1.2*. Cary, NC: SAS Institute.

Sayed, M. A. E. (2011). *QTL Analysis for Drought Tolerance Related to Root and Shoot Traits in Barley (Hordeum vulgare L.)*. PhD Thesis. Germany: Bonn Universitäts-und Landesbibliothek Bonn.

Seetan, R. I., Denton, A. M., Al-Azzam, O., Kumar, A., Iqbal, M. J., and Kianian, S. F. (2014). Reliable radiation hybrid maps: an efficient scalable clustering-based approach. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 11, 788–800. doi: 10.1109/tcbb.2014.2329310

Seetan, R. I., Kumar, A., Denton, A. M., Iqbal, M. J., Al-Azzam, O., and Kianian, S. F. (2013). "A fast and scalable clustering-based approach for constructing reliable radiation hybrid maps," in *Proceedings of the 12th International Workshop on Data Mining in Bioinformatics*, (Chicago, IL: Association for Computing Machinery).

Shindo, C., Tsujimoto, H., and Sasakuma, T. (2003). Segregation analysis of heading traits in hexaploid wheat utilizing recombinant inbred lines. *Heredity* 90, 56–63. doi: 10.1038/sj.hdy.6800178

Simmonds, J., Scott, P., Leverington-Waite, M., Turner, A. S., Brinton, J., Korzun, V., et al. (2014). Identification and independent validation of a stable yield and thousand grain weight QTL on chromosome 6A of hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 14:191.

Simmons, S. R., Oelke, E. A., and Anderson, P. M. (1914). *Growth and Development Guide for Spring Wheat*. St. Paul, MN: University of Minnesota Agricultural Extension Service.

Singh, A., Knox, R. E., DePauw, R. M., Singh, A. K., Cuthbert, R. D., Kumar, S., et al. (2016). Genetic mapping of common bunt resistance and plant height QTL in wheat. *Theor. Appl. Genet.* 129, 243–256. doi: 10.1007/s00122-015-2624-8

Solanki, S., Richards, J., Ameen, G., Wang, X., Khan, A., and Ali, H. (2019). Characterization of genes required for both Rpg1 and rpg4-mediated wheat stem rust resistance in barley. *BMC Genom.* 20:495.

Sourdille, P., Snape, J. W., Cadalen, T., Charmet, G., Nakata, N., Bernard, S., et al. (2000). Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. *Genome* 43, 487–494. doi: 10.1139/g00-013

Sun, X. Y., Wu, K., Zhao, Y., Kong, F. M., Han, G. Z., and Jiang, H. M. (2009). QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* 165, 615–624. doi: 10.1007/s10681-008-9794-2

Tadesse, W., Ogbonnaya, F. C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., et al. (2015). Genome-wide association mapping of yield and grain quality traits in winter wheat genotypes. *PLoS One* 10:e0141339. doi: 10.1371/journal.pone.0141339

Turner, N. C. (1979). "Drought resistance and adaptation to water de?cits in crop plants," in *Stress Physiology in Crop Plants*, ed. H. Mussell (New York, NY: John Wiley & Sons).

Turner, N. C., Wright, G. C., and Siddique, K. H. M. (2001). Adaptation of grain legumes (pulses) to water-limited environments. *Adv. Agron.* 71, 193–231. doi: 10.1016/s0065-2113(01)71015-2

Vocke, G., and Ali, M. (2013). *U. S. Wheat Production Practices, Costs, and Yields: Variations Across Regions*. Washington, DC: United States Department of Agriculture.

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* 588, 277–283.

Wang, S. B., Wen, Y. J., Ren, W. L., Ni, Y. L., Zhang, J., Feng, J. Y., et al. (2016). Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. *Sci. Rep.* 6:29951.

Wang, S., Basten, C. J., and Zeng, Z. B. (2012). *Windows QTL Cartographer 2.5–011*. Raleigh: North Carolina State University.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., and Huang, B. E. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Wei, L., Bai, S., Li, J., Hou, X., Wang, X., and Li, H. (2014). QTL positioning of thousand wheat grain weight in Qaidam Basin. *Open J. Genet.* 4, 239–244. doi: 10.4236/ojgen.2014.43024

Wu, Q. H., Chen, Y. X., Zhou, S. H., Fu, L., Chen, J. J., and Xiao, Y. (2015). High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 x Beinong6. *PLoS One* 10:e0118144. doi: 10.1371/journal.pone.0118144

Xu, X., Bai, G., Carver, B. F., and Shaner, G. E. (2005). A QTL for early heading in wheat cultivar Suwon 92. *Euphytica* 146, 233–237. doi: 10.1007/s10681-005-9017-z

Xue, G. P., Kooiker, M., Drenth, J., and McIntyre, C. L. (2011). TaMYB13 is a transcriptional activator of fructosyltransferase genes involved in β-2, 6-linked fructan synthesis in wheat. *Plant J.* 68, 857–870. doi: 10.1111/j.1365-313x.2011.04737.x

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., and Korzun, V. (2015). Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. *Front. Plant Sci.* 6:644.

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., and Korzun, V. (2014). Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L). *PLoS One* 9: e0113287.

Zanke, C., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., and Korzun, V. (2014). Genetic architecture of main effect QTL for heading date in European winter wheat. *Front. Plant Sci.* 5:217.

Zhang, W., Chao, S., Manthey, F., Chicaiza, O., Brevis, J. C., Echenique, V., et al. (2008). QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theor. Appl. Genet.* 117, 1361–1377.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Rabbi, Kumar, Mohajeri Naraghi, Sapkota, Alamri, Elias, Kianian, Seetan, Missaoui, Solanki and Mergoum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 June 2021
doi: 10.3389/fgene.2021.649988

Genome-Wide Association Mapping for Yield and Related Traits Under Drought Stressed and Non-stressed Environments in Wheat

S. M. Hisam A. Rabbi[1], Ajay Kumar[1], Sepehr Mohajeri Naraghi[1], Senay Simsek[1], Suraj Sapkota[2], Shyam Solanki[3], Mohammed S. Alamri[4], Elias M. Elias[1], Shahryar Kianian[5], Ali Missaoui[2,6] and Mohamed Mergoum[2,6]**

[1] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [2] Institute of Plant Breeding, Genetics, and Genomics, University of Georgia, Griffin, GA, United States, [3] Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States, [4] Department of Food Sciences and Nutrition, King Saud University, Riyadh, Saudi Arabia, [5] United States Department of Agriculture-The Agricultural Research Service (USDA-ARS) Cereal Disease Laboratory, University of Minnesota, St. Paul, MN, United States, [6] Department of Crop and Soil Sciences, University of Georgia, Griffin, GA, United States

OPEN ACCESS

Edited by:
Vijay Kumar Tiwari, University of Maryland, United States

Reviewed by:
Amita Mohan, University of Pennsylvania, United States
Gautam Saripalli, Chaudhary Charan Singh University, India

***Correspondence:**
Mohamed Mergoum
mmergoum@uga.edu
Ajay Kumar
ajay.kumar.2@ndsu.edu

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *06 January 2021*
Accepted: *28 April 2021*
Published: *22 June 2021*

Citation:
Rabbi SMHA, Kumar A, Mohajeri Naraghi S, Simsek S, Sapkota S, Solanki S, Alamri MS, Elias EM, Kianian S, Missaoui A and Mergoum M (2021) Genome-Wide Association Mapping for Yield and Related Traits Under Drought Stressed and Non-stressed Environments in Wheat. Front. Genet. 12:649988. doi: 10.3389/fgene.2021.649988

Understanding the genetics of drought tolerance in hard red spring wheat (HRSW) in northern USA is a prerequisite for developing drought-tolerant cultivars for this region. An association mapping (AM) study for drought tolerance in spring wheat in northern USA was undertaken using 361 wheat genotypes and Infinium 90K single-nucleotide polymorphism (SNP) assay. The genotypes were evaluated in nine different locations of North Dakota (ND) for plant height (PH), days to heading (DH), yield (YLD), test weight (TW), and thousand kernel weight (TKW) under rain-fed conditions. Rainfall data and soil type of the locations were used to assess drought conditions. A mixed linear model (MLM), which accounts for population structure and kinship (PC+K), was used for marker–trait association. A total of 69 consistent QTL involved with drought tolerance-related traits were identified, with $p \leq 0.001$. Chromosomes 1A, 3A, 3B, 4B, 4D, 5B, 6A, and 6B were identified to harbor major QTL for drought tolerance. Six potential novel QTL were identified on chromosomes 3D, 4A, 5B, 7A, and 7B. The novel QTL were identified for DH, PH, and TKW. The findings of this study can be used in marker-assisted selection (MAS) for drought-tolerance breeding in spring wheat.

Keywords: drought tolerance, hard red spring wheat, association mapping, quantitative trait loci, marker-assisted selection

INTRODUCTION

Wheat (*Triticum aestivum* L.) is a major crop worldwide contributing about 20% of calories to the human population. Current genetic and genomic improvements in wheat have helped increase its production; however, further improvements are essential to increasing wheat productivity to feed the world's population, which is projected to reach over nine billion by 2050 (Hertel, 2011; Sapkota et al., 2019). Wheat production is often reduced by several biotic and abiotic stresses including drought and heat. Plant breeding has improved crop resistance to both biotic and abiotic stresses; nevertheless, the progress is slow and the yield gap between stress-prone areas and favorable production regions of major crops, including wheat, is high (Edae et al., 2014). Therefore, breeding

efforts are being focused on dissecting the genetics of abiotic stresses including drought in wheat to develop knowledge and resources to speed up the development of climate-resilient wheat cultivars.

Drought poses a major threat to crop yield, highlighting the urgent need to develop drought-tolerant cultivars (Ergen and Budak, 2009). The majority of countries worldwide experience drought problems, even those in humid regions as they often have dry spells at some point. Obviously, drought is more severe in arid areas with minimal rainfall (Sun et al., 2006). North Dakota is the biggest producer of hard-red spring wheat (HRSW) in the USA (North Dakota Wheat Commission, 2016). The state, especially the semiarid western half, experiences frequent droughts (Climate Change and the Economy, 2008). Consequently, HRSW, a major cash crop for ND and the USA, is regularly affected by drought in this region. Developing and releasing drought-tolerant HRSW cultivars is critical to countering ND drought conditions, but this cannot be done without understanding the genetics of drought tolerance for HRSW in northern USA.

Quantitative trait locus (QTL) analysis allows genetic dissection, which can be a sound approach for understanding the molecular basis of drought tolerance in HRSW. In the past, several QTL mapping studies for drought tolerance in wheat were conducted (Kirigwi et al., 2007; Alexander et al., 2012; Ibrahim et al., 2012a,b). These studies have used different types of markers, including SSRs, EST-STS, and DArTs. However, almost all of these studies were based on low-resolution molecular maps consisting of only 102–690 markers. The number of markers in the previous studies seems insufficient to saturate the wheat genome due to its large size of 17 gigabase pairs (Brenchley et al., 2012). Also, drought tolerance is a quantitative trait adopting different mechanisms (Blum, 1988) and should have several QTL distributed throughout the whole genome. A high-resolution map can provide a more complete genetic dissection of drought tolerance and also a successful application of associated molecular markers through marker-assisted selection (MAS) programs. The Infinium iSelect 90K assay (Wang et al., 2014), with more than 81,000 gene-associated SNPs to assess polymorphism in bread wheat, provides a better means to identify SNPs tightly linked to drought tolerance.

Biparental QTL mapping, even when using high-density linkage maps, suffers some limitations. The biparental population has fewer recombination events and therefore has low resolution. By comparison, association mapping (AM) exploits a broader population and multiple alleles and has a better resolution of the QTL (Yu and Buckler, 2006). A few AM studies on drought tolerance conducted in the past have used a small number of markers (Dodig et al., 2012; Edae et al., 2013, 2014; Ballesta et al., 2020; Maulana et al., 2020), which seems insufficient to explore the variation in wheat, efficiently. Dodig et al. (2012) used 46 SSR markers, and Edae et al. (2013) used 78 DArT markers. Maulana et al. (2020) used greenhouse for phenotyping for the association study, whereas Ballesta et al. (2020) used field experiments for their study but showed the association only for chromosome 4A. Furthermore, to our knowledge, no genetic studies were conducted on HRSW germplasm to elucidate the QTL associated with drought-related traits in the region. Therefore, the present study was undertaken to identify genes/QTL associated with yield and agronomic traits evaluated under drought conditions in field experiments using an association mapping approach combined with high-density SNP marker assay.

MATERIALS AND METHODS

Plant Materials

In 2012, a germplasm panel comprising of 350 HRSW inbred lines developed by the HRSW breeding program at North Dakota State University (NDSU) and different cultivars with varying drought tolerances was used for this study (**Supplementary Table 1**). Eleven more accessions were added for the experiments conducted in 2013 and 2014 (**Supplementary Table 2**). These lines were developed over time from different crosses, and pedigree selections for different purposes such as drought tolerance, disease resistance, quality, and yield were used for this study.

Field Experiments and Data Collection

The evaluation of agronomic performances of the AM panel was carried out under non-irrigated field conditions at different locations in ND, USA. In 2012, the AM was evaluated at Prosper (46.96300°N, 97.01980°W), Casselton (46.540N, 97.1238°W), and Minot (48°13′59″N 101°17′32″W). In 2013, the evaluation was carried out in Prosper, Minot, and Williston (48°9′23″N 103°37′41″W), while in 2014, it was evaluated in Prosper, Minot, and Hettinger (46°0′3″N 102°38′0″W). Prosper and Casselton are located in eastern ND, at 46.9630° N, 97.0198° W, and 46.9° N, 97.210556° W, respectively. Minot is situated between western ND's semiarid grassland and central ND's subhumid grassland (48.2330° N, 101.2923° W). Williston's location is in northwestern ND (48.1470° N, 103.6180° W), and Hettinger's is in southwestern ND (46.0014° N, 102.6368° W). The total rainfall during the growing period (seed sowing to ripening) in 2012, 2013, and 2014 at Prosper was 119.6, 269.7, and 168.6 mm, respectively (**Table 1**). Minot's rainfall during the growing period was 168 mm in 2012, 159.8 mm in 2013, and 230.9 mm in 2014. In case of Casselton, Williston, and Hettinger, the rainfall was 122.8 mm (2012), 319.3 mm (2013), and 200.3 mm (2014), respectively, during the growing season (North Dakota Agricultural Weather Network, 2015). The total rainfall in the experimental sites during the growing period was considered to assess the drought condition (**Table 1**). The water holding capacity of the experimental sites was achieved from the soil type (Frazen, 2003). In 2012, the experiment was conducted in a randomized complete block design (RCBD) with two replicates, whereas a simple lattice design was used in 2013 and 2014. The plots had an area of 2.44 × 1.22 m and seven rows with a 15.24 cm gap between them in 2012 and 2013. The plot size in 2014 was 2.44 × 1.42 m, but the number of rows was still seven with a larger gap (17.78 cm) between them.

The phenotypic data was collected on DH, PH, YLD, TW, and TKW. The heading date (DH) was recorded when more than 50% of the plants in the plot were starting to flower. Plant height (PH) was measured in the middle of the plot from plant base to tip

TABLE 1 | Soil types, plant-available water (water-holding capacity of soil), and total rainfall for nine environments used in this study.

Environments	Soil type	Plant-available water (mm water/30.48 cm soil)	Rainfall (mm)
Casselton 2012	Fine silty loam	45.72–63.5	120.1
Prosper 2012	Fine silty loam	45.72–63.5	119.6
Minot 2012	Fine sandy loam	31.75–45.72	168
Prosper 2013	Fine silty loam	45.72–63.5	269.7
Minot 2013	Fine sandy loam	31.75–45.72	442.3
Williston 2013	Fine sandy loam	31.75–45.72	319.3
Minot 2014	Fine sandy loam	31.75–45.72	230.9
Prosper 2014	Fine silty loam	45.72–63.5	168.6
Hettinger 2014	Fine sandy loam	31.75–45.72	200.3

excluding the awn. Yield per plot (YLD) was converted to yield/ha for further analysis. Similarly, kg/0.5-pint cup was converted to kg/m^3 as the test weight (TW) for further analysis. A thousand kernels were counted using a seed counter and were weighed for thousand kernel weight (TKW).

Single-Nucleotide Polymorphism Genotyping

Genomic DNA was isolated from lyophilized young leaves of each genotype using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, cat. no. 69106). The quality of the DNA was checked on 0.8% agarose gel. The NanoDrop 1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE) was used to check the DNA concentration. The accessions of the AM panel were genotyped using the Illumina 90K iSelect wheat SNP assay (Wang et al., 2014) in the Small Grains Genotyping Lab at USDA-ARS, Fargo, ND. The Illumina iSelect 90K assay produced data for 81,587 SNPs. The analyses of SNP genotyping, clustering of the SNP alleles, and calling of the genotypes were performed with Genome Studio v2011.1 (www.illumina.com). The minimum number of points used in the cluster was 10 (Wang et al., 2014). Monomorphic SNPs and SNPs having more than 20% missing genotypic data and 10% heterozygosity were excluded. The best linear unbiased prediction (iBLUP) method (Yang et al., 2014b) was used to impute the missing genotypic data for the remaining SNPs. The polymorphic SNPs selected after filtering based on the above mentioned criteria were screened for their positions on the chromosomes based on the wheat consensus genetic map (Wang et al., 2014).

Phenotypic Data Analysis (ANOVA, Descriptive Statistics, and Frequency Distribution)

The ANOVA Proc MIXED procedure was used (SAS Institute, 2004) to analyze the phenotypic data from 2012, whereas for 2013 and 2014, the Proc LATTICE was used. The accessions of the AM panel were considered as fixed effects, and environments and blocks were considered as random effects in the ANOVA Proc MIXED procedure. The mean values were separated using the *F*-protected least significant difference (LSD) value at the $P \leq 0.05$ level of significance. CORR procedure of SAS (SAS Institute, 2004) was used to calculate Pearson correlations between traits for each environment. The phenotypic data with a low coefficient of variance (CV) value and significant differences among entries were used for further analysis. The locations that did not show significant differences for most of the traits and with a high CV were not included for further analysis and reporting.

Marker–Trait Association Analysis

Population structure was calculated using markers with pairwise $R^2 < 0.5$ for all pairwise comparisons. To assign the subpopulation membership for each genotype, STRUCTURE software version 3.2 was used (Pritchard et al., 2000). We used an admixture model with independent allele frequencies, a burn-in of 100,000, and an MCMC replication of 500,000 for $K = 1$–10 with five replications. The delta k calculated from the STRUCTURE software was used to select the optimum number of subpopulations. The number of subpopulations (k) was plotted against the delta k calculated using the STRUCTURE software. Pairwise linkage disequilibrium (LD) between markers in the null model was calculated as the squared allele frequency correlation (R^2) in the R-package (Lipka et al., 2012) after filtering for minor allele frequency (MAF) $\geq$ 5%. Genome-wide LD decay was estimated by plotting R^2 against the corresponding pairwise genetic distance (cM) (Wang et al., 2014). AM analysis was conducted using the software TASSEL v.5.0 (Bradbury et al., 2007). The mixed linear model (MLM) with PC + Kinship (K) was used for AM, where the genotypic data were filtered for minor allele ($\leq$ 5%) frequency. A total of 14,816 filtered SNPs were used for further AM study. The initial cutoff point for marker–trait association (MTA) was considered at $p \leq 0.001$. Then, this cutoff was subjected to Bonferroni correction (Yang et al., 2014a) to get the threshold ($p \leq 3.4 * 10^{-6}$). Only the markers identified to be associated in at least two environments were reported.

Candidate Gene Analysis

For candidate gene analyses, the sequences of the markers showing MTAs were obtained from the T3/Wheat database (Blake et al., 2016) and their physical positions were extracted using the BLAST search against Chinese_Spring_IWGSC_RefSeq1.0 (Appels et al., 2018, Alaux et al., 2018) to identify the most proximal gene. The physical position of each marker was utilized to identify if it represents the

perfect marker in the annotated gene space of Chinese Spring. For markers anchored in the non-gene space, flanking genes were obtained by manual IWGSC RefSeq1.0 (Blake et al., 2019) genome scanning in the GrainGenes Genome Browser (https://wheat.pw.usda.gov/GG3/genome_browser). For each MTA, the linked gene set can be extracted from the GrainGenes database. Finally, annotation of identified genes was included to predict their function. *In silico* expression analysis was carried out in the Wheat Expression Browser expVIP (Ramírez-González et al., 2018, Borrill et al., 2016) dataset for drought and heat stress and PEG to stimulate drought.

RESULTS

Phenotypic Analyses

Significant differences among genotypes were found in the environments of Casselton 2012, Prosper 2012, Minot 2012, Prosper 2013, Prosper and Minot 2013 and 2014, and Hettinger 2014 (**Table 2**). The rest of the phenotypic data was not analyzed further. The seeds of Minot 2013 could not be cleaned due to Fusarium head blight infection, and hence, YLD, TW, and TKW could not be reported for that environment. Also, TKW for Minot 2014 and Hettinger 2014 was not reported. The phenotypic analyses of the data showed that the heading date had a highly significant negative correlation with YLD, TW, and TKW in all environments (**Table 2**). Heading date showed a significant positive correlation with PH in four environments. Plant height had a significantly negative correlation with YLD in four environments including the overall mean. It had a significantly positive correlation with TW in six environments, whereas it did not show any correlation with TKW. The yield had a strong positive association with TW in five environments. Also, it showed a strong positive association with TKW in all the environments including the overall mean. Again, TW had a strong positive correlation with TKW in three environments (**Table 2**). The frequency distributions of the phenotypic data for DH, PH, YLD, TW, and TKW showed a continuous distribution, a characteristic of the typical quantitative traits (**Figure 1**).

TABLE 2 | Correlation coefficients between five agronomic traits in the association mapping panel in different environments (Env.) and overall mean across environments (M).

Trait ǂ and Env ǂ	PH	DH	YLD	TW
DH				
1	0.16**			
2	0.07*ns*			
3	0.15**			
4	0.18***			
5	0.06*ns*			
6	0.09*ns*			
7	0.09*			
8	0.09*ns*			
YLD				
1	−0.13*	−0.21***		
2	0.04*ns*	−0.4***		
3	0.13*	−0.16**		
4	0.13*	0.20***		
5	−0.16**	−0.15**		
6	−0.13*	−0.15**		
7	.	.		
8	−0.36***	−0.24***		
TW				
1	0.24***	−0.16**	0.18**	
2	0.27***	−0.35***	0.41***	
3	0.22**	−0.27***	0.42***	
4	0.12*	−0.03*ns*	−0.03*ns*	
5	0.09*ns*	−0.17**	0.30***	
6	0.25***	−0.34***	0.24***	
7	.	.	.	
8	0.2***	−0.27***	0.08*ns*	
TKW				
1	0.04*ns*	−0.14**	0.27***	0.11*
2	0.04*ns*	−0.15**	0.32***	0.19***
3	0.13*	−0.21***	0.29***	0.39***
4	0.05*ns*	0.05*ns*	0.30***	−0.09*ns*
5	.	.	.	.
6	.	.	.	.
7	.	.	.	.
8	−0.01*ns*	−0.10*ns*	0.34***	0.10*ns*

*Significant at $p < 0.05$, **Significant at $p < 0.01$, ***Significant at $p < 0.001$ level; ǂPH, Plant height; DH, Days to heading; YLD, Yield; TW, Test weight; TKW, Thousand kernel weightǂ1 = Casselton 2012, 2 = Prosper 2012, 3 = Minot 2012, 4 = Prosper 2013, 5 = Prosper 2014, 6 Hettinger 2014, 7 = Minot 2013, 8 = Mean across environments.

Marker Distribution, Population Structure, and LD

A total of 17,514 polymorphic SNPs were selected after the filtering-based criteria mentioned earlier in the material and methods section (**Supplementary Table 3**). An additional 2,756 SNPs were excluded for lacking map positions on the consensus hexaploid wheat maps (Wang et al., 2014). Out of 14,816 SNP markers used in the AM study, 7,848 were located on the B-genome, 5,503 on the A-genome, and 1,465 markers on the D-genome. The D-genome had the lowest density of markers, with an average distance of 0.87 cM between two markers. The number of markers on individual chromosomes ranged from 56 (4D) to 1,433 (2B). The average number of markers per chromosome was 705.52 (**Table 3**). The number of subpopulations (k) was plotted against the delta k calculated using software STRUCTURE. The peak of the broken line graph was observed at $k = 7$, indicating that the natural population can be divided into seven subpopulations (**Figure 2**). The association was analyzed using five principal components (PC), which captured 25% of the variation. Genome-wide LD decay was estimated by plotting the R^2 value against the corresponding pairwise genetic distance (cM) and the LD heat map was created for the whole genome (**Supplementary Figure 1**). The LD pattern varied by chromosome even after controlling for population relatedness. Overall, the A and B genomes showed high LD compared to the D-genome. The LD dropped at an approximate genetic distance of 10 cM, and therefore, ±10 cM was used to establish confidence

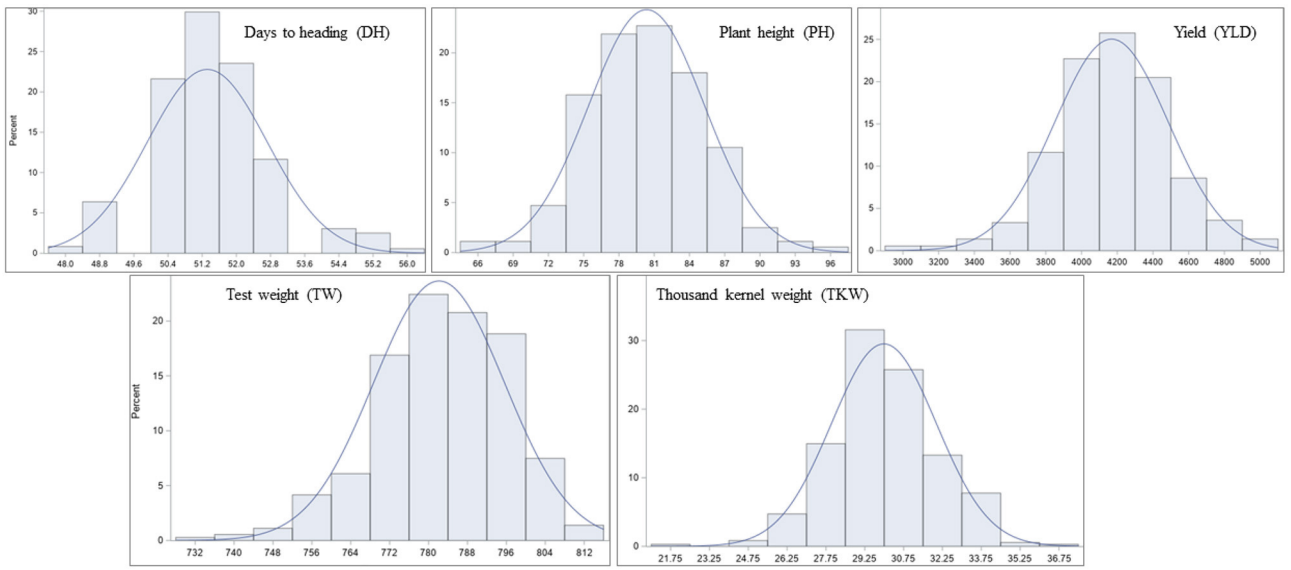


FIGURE 1 | Frequency distribution of least-square means (lsmeans) for the agronomic traits in the hard red spring wheat panel across environments. The panel was evaluated for days to heading (DH), plant height (PH), yield (YLD), test weight (TW), and thousand kernel weight (TKW) in nine different environments of North Dakota in 2012, 2013, and 2014.

intervals for QTL regions. Furthermore, SNP markers with a pairwise $R^2 \geq 0.7$ were considered as a single locus.

Identification of QTL and Associated Candidate Genes

In this study, we detected a total of 69 QTL involved with drought tolerance on all chromosomes except 5D using AM (**Table 4**; **Supplementary Figure 2**). Twenty QTL six were associated with DH. These QTL explained 5.6–11.33% of phenotypic variation (PV). Five of those QTL explained >10% of PV and therefore were considered major QTL. Twelve of the QTL were identified to be constitutive, and eight of the QTL were identified exclusively in drought-prone environments. Similarly, a total of 20 QTL six were associated with PH. These QTL explained 4.54–48.01% of PV, with major effects (>10% PV). Sixteen QTL were identified as constitutive, three were identified in the control environments, and one was identified in the drought environment (**Table 4**).

Seventeen QTL were identified for YLD. These QTL explained 4.11–12.04% of PV (**Table 4**). Only one QTL, located on chromosome 4B, had a major effect. Sixteen QTL were identified as constitutive, and the remaining QTL was identified in the drought-prone experimental sites. Five QTL were associated with TW. All of these QTL had minor effects, explaining 3.7–7.66% of PV. All of the QTL identified were constitutive. Seven QTL were identified for TKW, all of which had minor effects, explaining from 5.2 to 9.2% of PV. One QTL among them was constitutive, and the remaining six were identified in the drought-prone environments (**Table 4**).

The sequence of 69 markers associated with QTL was mined from Wang et al. (2014) to identify their physical position in the Chinese_Spring_IWGSC_RefSeq1.0 (Appels et al., 2018). Out of 69, seven markers (BobWhite_c47948_76, D_contig00840_473, Kukri_c15043_326, Excalibur_c25353_1171, Ex_c3115_2742, Excalibur_rep_c92985_510, Excalibur_c56240_176) were manually anchored on the genome. A total of 47 SNPs were anchored in the predicted gene space, and 22 were found in the intergenic region (**Supplementary Table 4**). For these intergenic SNPs, we identified the most proximal flanking genes representing the possible linked gene contributing to the phenotype. Thus, a total of 91 AM-QTL-associated most proximal genes along with their predicted function and genome orientation (**Table 5**) were mined from the genome. Using the expVIP analysis (Ramírez-González et al., 2018, Borrill et al., 2016), we explored the *in silico* expression of these genes affected by drought stress (**Supplementary Figure 3**) which could be used to prioritize the candidate genes. However, for breeding purposes all the MTAs are valuable.

DISCUSSION

Association Analyses

Studies are conducted to dissect the genetics of drought tolerance in many crops including wheat, and it is well-known that drought tolerance is a complex quantitative trait affected by genetic and environmental factors (Gahlaut et al., 2019). In this study, the iBLUP method (Yang et al., 2014b) was used to impute missing genotypic data as it was reported to tolerate a high rate of missing data especially for rare alleles, compared to the common imputation methods. High-density single-nucleotide polymorphism (SNP) genotyping arrays explore genomic diversity and MTAs very efficiently (Wang et al., 2014). Infinium iSelect 90K assay uses more than 81,000 gene-associated SNPs to reveal polymorphism in allohexaploid wheat populations

TABLE 3 | Chromosome and genome wide distribution of markers in our spring wheat association mapping panel based on the 90k SNP consensus map (Wang et al., 2014).

Chromosome	No. of markers	Map length	Average map density cM/marker
1A	785	156.3	0.2
2A	861	185.47	0.22
3A	661	197.2	0.3
4A	663	166.71	0.25
5A	783	148.3	0.19
6A	852	175.32	0.21
7A	898	244.16	0.28
1B	1,197	173.62	0.15
2B	1,433	188.27	0.13
3B	1,139	154.48	0.14
4B	635	118.91	0.19
5B	1,348	219.77	0.16
6B	1,216	122.92	0.1
7B	880	188.64	0.21
1D	261	199.86	0.77
2D	476	152.84	0.32
3D	207	152.84	0.74
4D	56	170.43	3.04
5D	147	207.33	1.41
6D	170	160.5	0.94
7D	148	226.87	1.53
A genome	5,503	1,273.46	0.23
B genome	7,848	1,166.61	0.15
D genome	1,465	1,270.67	0.87
Whole genome	14,816	3,710.74	0.25

(Wang et al., 2014; Kumar et al., 2016, 2019). Higher genome coverage and resolution in the dissection of wheat's agronomic traits are possible using this genotypic tool (Kirigwi et al., 2007; Alexander et al., 2012; Ibrahim et al., 2012b). The marker density found in this study (0.49 cM/marker) was in agreement with the previous studies using the 90K Infinium iSelect assay (Wang et al., 2014; Ain et al., 2015; Kumar et al., 2016). The MLM model used in this association study has been proven to be very efficient for genome-wide association studies (GWAS) and can be used with either structure (R) or principal component (PC) analyses. This study used five PCs, which captured 25% of the variation. The MLM model, which accounts for both structure and relatedness (PC + K), was used for the marker–trait association study.

Determining the threshold for the *p*-value is crucial. A liberal threshold will declare a false-positive association (a type I error), whereas a too stringent threshold is likely to miss a true association (a type II error). Taking this into consideration, the initial cutoff was chosen as $p \leq 0.001$, which was not very stringent. Then, the threshold ($p \leq 3.4 * 10^{-6}$) was determined using the Bonferroni correction (Yang et al., 2014a), which was very stringent. The MTAs identified at the initial cutoff and the threshold were reported if they were identified in at least two environments. This repetition of the MTA further minimized any false associations.

Use of Secondary Data to Assess Drought Conditions

Drought can be assessed by variable weather conditions, soil moisture, and crop conditions over a particular growing season (Lanceras et al., 2004). Therefore, rainfall data were collected, and the soil types of the experimental sites, which reflect soil moisture, were taken into consideration to assess drought conditions for this study. The total amount of rainfall was collected from planting date to plant physiological maturity. The dates for the physiological maturity of the plants were calculated by adding 30 days to DH (Simmons et al., 1914). Among the experimental locations, Casselton 2012, Prosper 2012, and Minot 2012 were considered to have drought conditions, whereas Prosper 2013, Minot 2013, Prosper 2014, and Hettinger 2014 were considered to have control or normal conditions. Although Minot 2012 and Prosper 2014 had about the same amount of rainfall, the soil in Prosper had a better water-holding capacity. Therefore, Minot 2012 was considered to have drought conditions.

Use of Agronomic Data to Assess Drought Tolerance

Days to Heading

Several major and minor QTL were revealed for DH, which indicated the quantitative nature of the trait. The eight QTL for DH, identified exclusively under drought conditions, could play a vital role in drought tolerance. Also, the constitutive QTL can be used for drought tolerance breeding in wheat. Some of these QTL (exclusively expressed under drought conditions, called constitutive QTL) identified in this study likely correspond with some already reported QTL associated with drought tolerance; however, further studies such as allelism test is warranted to determine their relationship. Malik et al. (2015) identified three adjacent QTL on chromosome 2A for drought tolerance related to the photosynthetic rate, cell membrane stability, and relative water content. The QTL *QDH.ndsu.2A.1* in this study likely represent one of those QTL previously reported (Malik et al., 2015; Gahlaut et al., 2019), but further studies are required to determine their relationship. Two QTL identified in this study on the chromosome 3A which were important for drought tolerance, *QDH.ndsu.3A.1* and *QDH.ndsu.3A.2*, could represent the QTL *QHea.T84-3A* which was earlier found to increase DH under both drought and non-drought conditions (Ibrahim et al., 2012b). Chromosomal arm 3AL also harbors a gene for earliness *per se* (Edae et al., 2014), associated with enhanced response to abscisic acid (*ERA1*), which provides drought tolerance (Edae et al., 2014). The gene *ERA1*, also located on chromosome 3B, could represent the QTL *QDH.ndsu.3B* identified in this study which is closely associated with *TraesCS3B01G356000* (**Table 5**), putatively encoding for inositol-1,4,5-trisphosphate 5-phosphatase ($InsP_3$). $InsP_3$ is reported to be a second messenger in plants responding to many stimuli

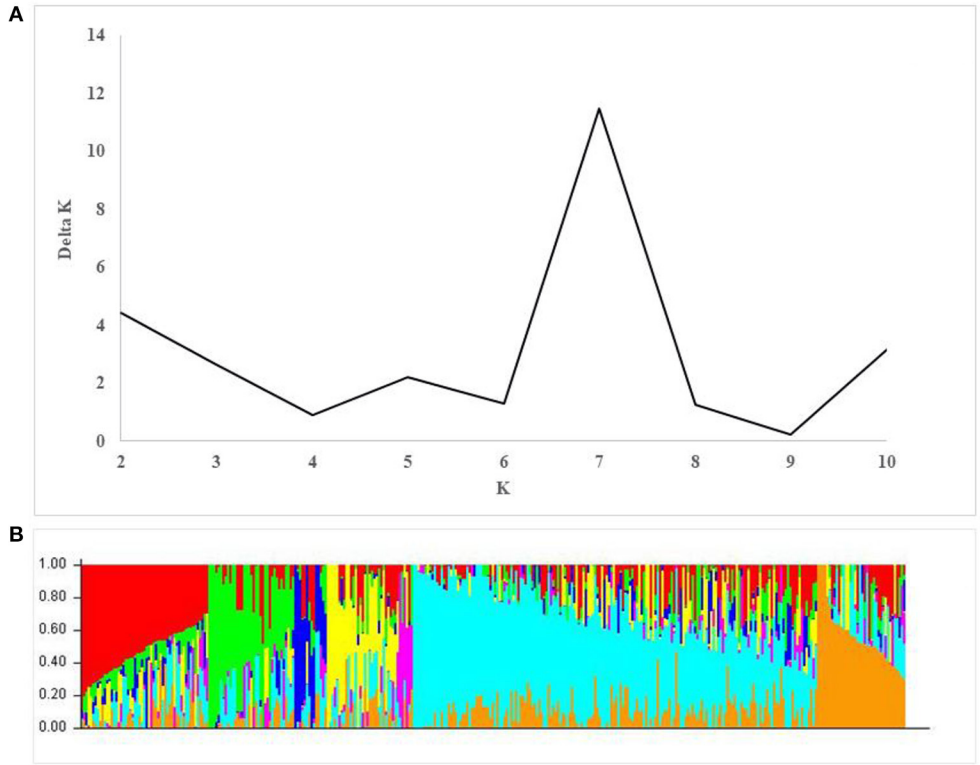


FIGURE 2 | Population structure in the association mapping (AM) panel. **(A)** Estimation of the number of populations by calculating delta K values. **(B)** Estimated population structure of the AM panel ($k = 7$).

and has been shown to affect drought tolerance, carbohydrate metabolism, and phosphate-sensitive biomass increases in tomato (Khodakovskaya et al., 2010). In wheat, differential expression of the *phospholipase C* gene regulating the inositol-1,4,5-triphosphate (IP_3) signal transduction pathway possibly results in the quick sensing of drought stress (Ergen et al., 2009). Kamran et al. (2013) identified a QTL, *QFlt.dms-4A.1*, for reduced DH at 61.2 cM on chromosome 4A, which may represent the constitutive QTL *QDH.ndsu.4A.1* identified in this study. The constitutive QTL *QDH.ndsu.2B.2* is located in the same region as the earlier reported QTL *QCrs-* (Ibrahim et al., 2012a), which deteriorate the number of root crossing in both water regimes. A QTL for drought tolerance on 4AL reported earlier by Alexander et al. (2012) may represent the QTL *QDH.ndsu.4A.2*, which was identified exclusively for drought-prone environments in this study. The constitutive QTL *QDH.ndsu.6B* was located in the same genomic location as the QTL *QHea+*, which was reported to reduce DH under both water conditions (Ibrahim et al., 2012b). Huang et al. (2006) reported a QTL for days to maturity, *QDtm.crc-2D*, which corresponded with the constitutive QTL in this study, *QDH.ndsu.2D*, representing the kinase family protein. However, the SNP markers associated with sucrose-phosphate synthase (*TraesCS3A01G425500*) and vacuolar protein sorting-associated protein (*TraesCS5A01G259200, TraesCS3A01G317000*) in this group (**Table 5**) may represent the important abiotic stress genes controlling the plant height. The QTL *QDH.ndsu.5B.2* and *QDH.ndsu.7B* identified in this study seem to be novel.

Plant Height

The QTL *QPH.ndsu.5B* could represent the ortholog to the GA-insensitive dwarf gene, *GID1L2*, in rice, indicating the synergistic relationship of rice and wheat (Zanke et al., 2014). The major QTL for PH, *QPH.ndsu.6B.1* and *QPH.ndsu.6B.2*, have been identified on wheat chromosome 6B, and several previous studies also reported QTL for PH on a similar location (Zanke et al., 2014; Gahlaut et al., 2019; Abou-Elwafa and Shehzad, 2021). The major QTL *QPH.ndsu.4B* could represent the reduced height gene *Rht-B1* (Wilhelm, 2011), which was reported to be on the short arm of chromosome 4B. This gene encodes the DELLA protein that reduces a plant's sensitivity to gibberellin (GA), thereby reducing stalk length and making the plant semi-dwarf. The QTL *QPH.ndsu.1A*, *QPH.ndsu.2A.1*, *QPH.ndsu.6A.2*, and *QPH.ndsu.3A.3* could represent the QTL for PH reported earlier by Zanke et al. (2014). The QTL *QPH.ndsu.3A.2* and *QPH.ndsu.3D.2*, important for drought tolerance, could be the same as those reported by Ibrahim et al. (2012a). Liu et al. (2011) identified a QTL for PH, *QHt-3B*, which could occupy the same region as the QTL *QPH.ndsu.3B* in the study. The QTL *QPH.ndsu.7A.1* coincided with the earlier reported QTL *QHt.crc-7A* (McCartney et al., 2005). The QTL *QPH.ndsu.7A.2*

TABLE 4 | Marker-trait associations (QTL) identified for yield and related traits under different environments in this study.

QTL and trait	QTL region	Other associated traits ǂ	Env. ǂ	Water regime	Position§	*p*-value¶	R^2 (%)
Days to heading							
QDH.ndsu.1B	5		1*, 2, 4*	Control+drought	90.26	$4.83*10^{-7}$	8.96
QDH.ndsu.2A.2	10		1*, 2, 4*	Control+drought	113.30	$1.92*10^{-6}$	8.07
QDH.ndsu.2B.2	14	PH, TKW	1*, 2, 4*, 6*, 7*	Control+drought	99.8–104.39	$4.86*10^{-8}$	10.44
QDH.ndsu.2D	16		2*, 4	Control+drought	19.03	$2.98*10^{-6}$	7.54
QDH.ndsu.3A.1	17	PH, TW	1*, 2, 6*, 7*	Control+drought	90.55	$1.25*10^{-8}$	11.33
QDH.ndsu.3B	21	TW, YLD	1*, 2, 3*, 4*, 5*	Control+drought	70.09	$1.44*10^{-8}$	11.24
QDH.ndsu.4A.1	25	TKW	1*, 2, 3*, 4*	Control+drought	51.97	$1.83*10^{-6}$	8.10
QDH.ndsu.4B	28	TW, PH, YLD	1*, 2*, 4*, 7*	Control+drought	64.03–75.64	$7.35*10^{-6}$	7.22
QDH.ndsu.4D	29		1*, 2, 4*	Control+drought	94.22	$3.11*10^{-8}$	10.74
QDH.ndsu.5B.2	35		1*, 2*, 3*, 4*, 5*	Control+drought	100.64–110.19	$7.67*10^{-6}$	7.19
QDH.ndsu.6B	39	TKW, PH, YLD	1*, 2*, 6*, 7*	Control+drought	63.14–71.76	$2.97*10^{-5}$	6.34
QDH.ndsu.7B	44	YLD	2*, 3*, 4*	Control+drought	98.30–101.18	$9.18*10^{-5}$	5.60
QDH.ndsu.2A.1	8	TW	1*, 2	Drought	25.02	$4.95*10^{-8}$	10.43
QDH.ndsu.2A.3	11	YLD	1*, 2	Drought	141.66	$1.81*10^{-6}$	8.11
QDH.ndsu.2B.1	13	YLD	1*, 2	Drought	83.80	$8.13*10^{-7}$	8.62
QDH.ndsu.3A.2	18	YLD	1*, 2	Drought	117.73	$1.58*10^{-7}$	9.68
QDH.ndsu.4A.2	26	TKW	1*, 2*, 3*	Drought	99.19–103.03	$3.69*10^{-5}$	6.21
QDH.ndsu.5A.1	30		1*, 2	Drought	55.01	$2.54*10^{-6}$	7.90
QDH.ndsu.5A.2	31		1*, 2*, 3*	Drought	84.13	$3.98*10^{-6}$	7.61
QDH.ndsu.5B.1	33	TKW	1*, 2	Drought	5.70	$8.35*10^{-7}$	8.61
Plant height							
QPH.ndsu.1A	3		1*, 5, 6*	Control+drought	105.74	$9.80*10^{-13}$	16.83
QPH.ndsu.1B	4	YLD	2*, 4*, 7	Control+drought	76.89	$3.2*10^{-7}$	7.73
QPH.ndsu.2A.1	9		2*, 3*, 4*, 7	Control+drought	98.43–101.97	$2.03*10^{-5}$	6.31
QPH.ndsu.2A.2	12		2*, 6*,7*	Control+drought	156.23–162.89	$2.4*10^{-5}$	6.25
QPH.ndsu.2B	14	TKW, DH	2*, 3*, 4*, 5, 7*	Control+drought	109.53	$2.24*10^{-7}$	8.99
QPH.ndsu.3A.1	17	TW, DH	2*, 3*, 4*	Control+drought	77.57	$1.68*10^{-4}$	5.13
QPH.ndsu.3A.2	19		2*, 6	Control+drought	128.64	$6.08*10^{-7}$	8.46
QPH.ndsu.3A.3	20		3*, 4*, 6*	Control+droughte	180.33	$5.42*10^{-5}$	5.72
QPH.ndsu.3B	22		2*, 3*, 5*	Control+drought	102.54–106.73	$5.29*10^{-4}$	4.54
QPH.ndsu.3D.1	23		2*, 5*, 7*	Control+drought	0–4.46	$3.29*10^{-4}$	4.67
QPH.ndsu.3D.2	24		2*, 3*, 4*	Control+drought	66.99	$1.62*10^{-4}$	5.27
QPH.ndsu.4B	28	TW, YLD, DH	1, 2, 3, 4, 5*, 6, 7*	Control+drought	56.19	$3.79*10^{-14}$	19.97
QPH.ndsu.5B	34	YLD	2*, 3*, 4*,5, 6, 7*	Control+drought	63.07	$6.51*10^{-31}$	48.01
QPH.ndsu.6A.2	38		1, 5*	Control+drought	133.74	$8.66*10^{-8}$	10.07
QPH.ndsu.6B.1	39	TKW, DH, YLD	3*, 4*, 5, 6*, 7	Control+drought	56.98	$1.8*10^{-30}$	47.15
QPH.ndsu.7A.2	43		2*, 3*, 4*	Control+drought	212.66	$4.3*10^{-6}$	7.55
QPH.ndsu.6A.1	37		4, 5	Control	82.38	$4.61*10^{-10}$	12.85
QPH.ndsu.6B.2	40	YLD	5, 7*	Control	108.86	$1.07*10^{-30}$	47.60
QPH.ndsu.6D	41		4*, 5, 6*, 7*	Control	22.92	$1.56*10^{-7}$	9.21
QPH.ndsu.7A.1	42		1*, 2 *, 3*	Drought	61.36	$3.38*10^{-4}$	4.83
Thousand kernel weight							
QTKW.ndsu.2B.1	14	PH, DH	1*, 2*, 3*	Drought	106.56–114.57	$7.44*10^{-5}$	5.64
QTKW.ndsu.2B.2	15	YLD	2*, 3	Drought	155.41	$9.33*10^{-7}$	8.55
QTKW.ndsu.4A.1	25	DH	1*, 2*, 3*	Drought	48.98–51.97	$1.74*10^{-4}$	5.22
QTKW.ndsu.4A.2	26	DH	1*, 2*, 3*	Drought	105.87–108.72	$1*10^{-4}$	5.59
QTKW.ndsu.4A.3	27		2, 3*	Drought	154.30	$2.44*10^{-7}$	9.20
QTKW.ndsu.6B	39	PH, DH, YLD	1*, 2*, 3*	Drought	56.64–64.82	$1.79*10^{-4}$	5.20
QTKW.ndsu.5B	33	DH	1*, 3, 4*	Control+drought	17.48	$1.98*10^{-6}$	6.92

(Continued)

TABLE 4 | Continued

QTL and trait	QTL region	Other associated traits ɨ	Env. ǂ	Water regime	Position§	*p*-value¶	R^2 (%)
Test weight							
QTW.ndsu.1A	1		1*, 2*, 6*	Control+drought	29.11–38.11	$4.39*10^{-4}$	4.63
QTW.ndsu.2A	8	DH	1*, 4*, 5*	Control+drought	20.26	$7.12*10^{-4}$	4.07
QTW.ndsu.3A	17	PH, DH	1*, 2*, 5*	Control+drought	85.73	$284*10^{-4}$	3.70
QTW.ndsu.3B	21	DH, YLD	1*, 3*, 4*, 5*	Control+drought	62.31–69.53	$3.83*10^{-6}$	7.58
QTW.ndsu.4B	28	PH, YLD, DH	1, 2*, 3, 4*, 5*, 6*	Control+drought	55.55	$4.66*10^{-7}$	7.66
Yield							
QYL.ndsu.1A	2		1*, 3*, 5*	Control+drought	48.45–56.81	$1.49*10^{-5}$	6.77
QYL.ndsu.1B.1	4	PH	1*, 2*, 3*, 4	Control+drought	70.08	$2.22*10^{-6}$	6.59
QYL.ndsu.1B.2	6		1*, 5*, 6*	Control+drought	112.07	$3.68*10^{-5}$	6.20
QYL.ndsu.1D	7		1, 5*	Control+drought	3.40	$2.89*10^{-6}$	7.80
QYL.ndsu.2A	11	DH	1, 4*, 5*	Control+drought	144.41	$1.86*10^{-6}$	8.08
QYL.ndsu.2B.1	13	DH	1, 4*, 6*	Control+drought	88.93–90.971	$1.31*10^{-5}$	5.57
QYL.ndsu.2B.2	15	TKW	1, 3, 4, 5, 6	Control+drought	157.21	$1.79*10^{-6}$	8.11
QYL.ndsu.3B	21	TW, DH	1*, 4*, 5*, 6*	Control+drought	62.31–69.53	$5.81*10^{-6}$	7.36
QYL.ndsu.4B	28	PH, DH, TW	1, 4*, 5, 6*	Control+drought	56.19	$4.17*10^{-9}$	12.04
QYL.ndsu.5A	32		1*, 4*, 6*	Control+drought	116.35–117.67	$1.7*10^{-4}$	4.11
QYL.ndsu.5B	34	PH	1, 4*	Control+drought	68.36	$1.94*10^{-6}$	6.91
QYL.ndsu.6A	36		1, 3*, 4*	Control+drought	12.48	$1.52*10^{-6}$	8.21
QYL.ndsu.6B.1	39		1*, 4*, 5*	Control+drought	64.08–64.71	$6.08*10^{-6}$	7.33
QYL.ndsu.6B.2	40	PH	1, 3*, 4	Control+drought	115.25	$1.38*10^{-6}$	7.12
QYL.ndsu.7B	44	DH	1*, 3*, 4*	Control+drought	89.82–92.52	$6.64*10^{-4}$	4.26
QYL.ndsu.7D	45		1*, 4*, 5*	Control+drought	128.15–135.55	$3.19*10^{-5}$	5.24
QYL.ndsu.3A	18	DH	1, 2*, 3*	Drought	109.95	$1.52*10^{-6}$	8.21

ɨ *DH, Days to heading; PH, Plant height; YLD, Yield; TW, Test weight; TKW, Thousand kernel weight.*

ǂ *1 = Casselton 2012, 2 = Prosper 2012, 3 = Minot 2012, 4 = Prosper 2013, 5 = Prosper 2014, 6 = Hettinger 2014, 7 = Minot 2013, 8 = Mean across environments.*

§ *Position represents the peak point of the QTL interval. The position is based on consensus map of Wang et al. (2014).*

¶ *$p < 0.001$ but above the threshold level.*

in a receptor-like kinase gene (*TraesCS7A01G540200*) and *QPH.ndsu.3D.1* in a subtilisin-like protease (**Table 5**) in this study did not correspond with any reported QTL and hence could be novel.

YLD

In the past, Edae et al. (2014) reported a QTL for TKW on chromosome 1BL and a QTL for TW on chromosome 2BL that could be in the region of QTL *QYL.ndsu.1B.1* and *QYL.ndsu.2B.2*, respectively. More recently, Tura et al. (2020) detected a main-effect QTL, *QYld.aww-1B.2*, on 1B chromosome which likely represents the same locus as *QYL.ndsu.1B.*1; however, further research is warranted to determine their relationship. Ibrahim et al. (2012a) identified a QTL, *QCrs.D84-2B*, on chromosome 2B at 93.4 cM that deteriorates the number of root crossings under both water regimes and could represent the QTL *QYL.ndsu.2B.1* (*TraesCS2D01G126100*-cellulose synthase) found in this study. Another QTL for YLD *QYld.T84-3Bat* identified earlier (Ibrahim et al., 2012b) occupies the same location as the QTL *QYL.ndsu.3B* identified in this study. Another QTL, *QYld.T84-3Bat* 59.8, which deteriorated YLD under both water regimes, could coincide with the QTL *QYL.ndsu.4B* in (*TraesCS4B01G047900*) identified in this study. The QTL *QYL.ndsu.5B* and *QYL.ndsu.6B.2* also likely correspond to the QTL for TW and TKW identified earlier by Edae et al. (2014). Also, the QTL *QYL.ndsu.5B* corresponded with the QTL *QYld**, which was reported to improve YLD under drought stress (Ibrahim et al., 2012b). The QTL *QYL.ndsu.1B.2* had the same genomic location as the constitutive QTL for the green leaf area reported by Edae et al. (2014). Ibrahim et al. (2012a) reported a QTL, *QTgw+*, which improved thousand-grain weight under both water conditions and could represent the QTL *QYL.ndsu.1D.* The QTL *QYL.ndsu.2A* likely coincides with the YLD QTL *QGY.caas-2A* (Li et al., 2015) or a QTL identified by Mathew et al. (2019) evaluated under drought stress conditions. Huang et al. (2006) identified the QTL *QTgw.crc-6A* for TKW that seems to be present at the same location as the QTL *QYL.ndsu.6A* identified in this study. The QTL *QYL.ndsu.7D* corresponded with the QTL *QHi+*, which was reported to improve the harvest index under both water conditions (Ibrahim et al., 2012b).

TW

The QTL *QTW.ndsu.4B* could be the same QTL earlier reported by Li et al. (2016) on the same chromosome. The QTL *QTW.ndsu.1A* corresponded with two QTL for YLD, *QYld.abrii-1A1.2* (Azadi et al., 2014) and *QGY.caas-1A* (Li et al., 2015). The constitutive QTL *QTW.ndsu.2A* occupied the same genomic region as the QTL for drought tolerance related to the photosynthetic rate reported by Malik et al. (2015). The QTL

TABLE 5 | List of candidate genes associated with QTL identified in this study.

QTL and trait	Associated/proximal flanking gene	Function	Gene space?	Distal flanking gene
Days to heading				
QDH.ndsu.1B	*TraesCS1B01G359800*	Alpha-glucan water dikinase	Yes	
QDH.ndsu.2A.2	*TraesCS2A01G421100*	Myosin heavy chain embryonic smooth protein	Yes	
QDH.ndsu.2B.2	*TraesCS2B01G395600*	Histone H2A deubiquitinase (DUF3755)	Yes	
QDH.ndsu.2D	*TraesCS2D01G076700*	Kinase family protein	Yes	
QDH.ndsu.3A.1	*TraesCS3A01G317000*	Vacuolar sorting-associated protein 28-like protein	No	*TraesCS3A01G316900*
QDH.ndsu.3B	*TraesCS3B01G356000*	Inositol-1,4,5-trisphosphate 5-phosphatase 4 isoform 1	No	*TraesCS3B01G355900*
QDH.ndsu.4A.1	*TraesCS4A01G275500*	Importin-5	Yes	
QDH.ndsu.4B	*TraesCS4B01G293600*	Copper-transporting ATPase	Yes	
QDH.ndsu.4D	*TraesCS4D01G285000*	50S ribosomal protein L18	Yes	
QDH.ndsu.5B.2	*TraesCS5B01G391500*	Evolutionarily conserved C-terminal region 2—RNA binding	Yes	
QDH.ndsu.6B	*TraesCS5B01G416700*	Citrate synthase	Yes	
QDH.ndsu.7B	*TraesCS7B01G369500*	Tegument protein BRRF2	No	*TraesCS7B01G369400*
QDH.ndsu.2A.1	*TraesCS2A01G048900*	Defensin	Yes	
QDH.ndsu.2A.3	*TraesCS2A01G517500*	Chloride channel protein	Yes	
QDH.ndsu.2B.1	*TraesCS2B01G131500*	THUMP domain-containing protein 1	Yes	
QDH.ndsu.3A.2	*TraesCS3A01G425500*	Sucrose-phosphate synthase	Yes	
QDH.ndsu.4A.2	*TraesCS4A01G359500*	Squamosa promoter-binding protein	Yes	
QDH.ndsu.5A.1	*TraesCS5A01G259200*	Vps52, vacuolar protein sorting-associated protein 52	Yes	
QDH.ndsu.5A.2	*TraesCS5D01G381300*	EEIG1/EHBP1 N-terminal domain-containing protein	Yes	
QDH.ndsu.5B.1	*TraesCS5B01G021800*	Glutathione s-transferase	Yes	
Plant height				
QPH.ndsu.1A	*TraesCS1A01G356900*	Choline kinase	Yes	
QPH.ndsu.1B	*TraesCS1B01G309400*	Non-specific serine/threonine protein kinase	Yes	
QPH.ndsu.2A.1	*TraesCS2A01G247000*	D-Lactate dehydrogenase	No	*TraesCS2A01G246900*
QPH.ndsu.2A.2	*TraesCS2A01G559200*	1-Phosphatidylinositol-3-phosphate 5-kinase	Yes	
QPH.ndsu.2B	*TraesCS2A01G452200*	Ribosome biogenesis protein BRX1-like protein	No	*TraesCS2A01G452300*
QPH.ndsu.3A.1	*TraesCS3A01G087800*	Transcription factor GTE9	Yes	
QPH.ndsu.3A.2	*TraesCS3B01G491000*	F-box protein	Yes	
QPH.ndsu.3A.3	*TraesCS3A01G525300*	P-loop-containing nucleoside triphosphate hydrolase superfamily	No	*TraesCS3A01G525200*
QPH.ndsu.3B	*TraesCS3B01G531400*	Voltage-dependent L-type calcium channel subunit	Yes	
QPH.ndsu.3D.1	*TraesCS3B01G004600*	Subtilisin-like protease	Yes	
QPH.ndsu.3D.2	*TraesCS3D01G250900*	ATP synthase subunit beta	Yes	
QPH.ndsu.4B	*TraesCS4B01G047900*	NADH-ubiquinone oxidoreductase subunit	Yes	
QPH.ndsu.5B	*TraesCS5B01G303100*	Hydroxyacylglutathione hydrolase	Yes	
QPH.ndsu.6A.2	*TraesCS6A01G398000*	Protein KINESIN LIGHT CHAIN-RELATED 3	Yes	
QPH.ndsu.6B.1	*TraesCS6B01G126900*	Kinetochore protein nuf2	Yes	
QPH.ndsu.7A.2	*TraesCS7A01G540200*	Receptor-like protein kinase	Yes	
QPH.ndsu.6A.1	*TraesCS6A01G297400*	B-cell receptor-associated like protein	No	*TraesCS6A01G297500*
QPH.ndsu.6B.2	*TraesCS6B01G434900*	ABC transporter G family member	Yes	
QPH.ndsu.6D	*TraesCS6D01G013700*	Zinc finger BED domain-containing protein DAYSLEEPER	Yes	
QPH.ndsu.7A.1	*TraesCS6A01G073600*	DNA topoisomerase 3-alpha	No	*TraesCS6A01G073500*
1K kernel weight				
QTKW.ndsu.2B.1	*TraesCS2B01G491600*	Transmembrane protein	Yes	
QTKW.ndsu.2B.2	*TraesCS2B01G606100*	Receptor kinase 1	No	*TraesCS2B01G606200*
QTKW.ndsu.4A.1	*TraesCS4A01G269200*	TBPIP	No	*TraesCS4A01G269400*
QTKW.ndsu.4A.2	*TraesCS4A01G365500*	Protein PLANT CADMIUM RESISTANCE 2	Yes	
QTKW.ndsu.4A.3	*TraesCSU01G167400*	Receptor kinase	No	*TraesCSU01G167500*

(Continued)

TABLE 5 | Continued

QTL and trait	Associated/proximal flanking gene	Function	Gene space?	Distal flanking gene
QTKW.ndsu.6B	*TraesCS6B01G228100*	RNA-binding protein	No	*TraesCS6B01G228300*
QTKW.ndsu.5B	*TraesCS5B01G015100*	WAT1-related protein	No	*TraesCS5B01G015200*
Test weight				
QTW.ndsu.1A	*TraesCS1D01G020000*	Actin	No	TraesCS1D01G020100
QTW.ndsu.2A	*TraesCS2A01G012200*	Protein transport Sec24-like	Yes	
QTW.ndsu.3A	*TraesCS3A01G149800*	Dual-specificity protein phosphatase	No	*TraesCS3A01G149900*
QTW.ndsu.3B	*TraesCS3B01G153500*	Arginine-tRNA ligase	No	*TraesCS3B01G153600*
QTW.ndsu.4B	*TraesCS4D01G050100*	26S proteasome non-ATPase regulatory subunit	No	*TraesCS4D01G050200*
Yield				
QYL.ndsu.1A	*TraesCS1A01G040600*	Glucose-6-phosphate isomerase	Yes	
QYL.ndsu.1B.1	*TraesCS1B01G252700*	RCC1 family with FYVE zinc finger domain-containing protein	Yes	
QYL.ndsu.1B.2	*TraesCS1B01G253500*	UDP-sulfoquinovose synthase	Yes	
QYL.ndsu.1D	*TraesCS1A01G002300*	Endoribonuclease E-like protein	Yes	
QYL.ndsu.2A	*TraesCS2D01G541500*	Regulatory protein recX	Yes	
QYL.ndsu.2B.1	*TraesCS2D01G126100*	Cellulose synthase	Yes	
QYL.ndsu.2B.2	*TraesCS2A01G567900*	Tetratricopeptide repeat	Yes	
QYL.ndsu.3B	*TraesCS3B01G153500*	Arginine-tRNA ligase	No	*TraesCS3B01G153600*
QYL.ndsu.4B	*TraesCS4B01G047900*	NADH-ubiquinone oxidoreductase subunit	Yes	
QYL.ndsu.5A	*TraesCS7B01G206200*	Calcium-binding family protein	No	*TraesCS7B01G206300*
QYL.ndsu.5B	*TraesCS5D01G326500*	Alpha/beta hydrolase	Yes	
QYL.ndsu.6A	*TraesCS6D01G011300*	DNA-directed RNA polymerase subunit beta	No	*TraesCS6D01G011400*
QYL.ndsu.6B.1	*TraesCS6B01G173000*	Sucrose-phosphate synthase	Yes	
QYL.ndsu.6B.2	*TraesCS6B01G455400*	Tuftelin-interacting protein 11	No	*TraesCS6B01G455500*
QYL.ndsu.7B	*TraesCS7B01G363600*	Geranylgeranyl transferase type-2 subunit alpha 1	Yes	
QYL.ndsu.7D	*TraesCS7D01G287700*	Copia-like polyprotein/retrotransposon	No	*TraesCS7D01G287800*
QYL.ndsu.3A	*TraesCS3A01G406700*	P-loop containing nucleoside triphosphate hydrolases superfamily protein	Yes	

QTW.ndsu.3B corresponded with the YLD QTL *QYld.T84-3Bat* reported by Ibrahim et al. (2012b).

TKW

The QTL *QTKW.ndsu.4A.2* had the same genomic location as the QTL reported by Kirigwi et al. (2007) for YLD and YLD-related traits under drought stress. Ibrahim et al. (2012a) identified the *QTgw-* for thousand-grain weight under both water conditions, which seems to represent the QTL *QTKW.ndsu.6B* identified in this study. The QTL *QTKW.ndsu.2B.1*, *QTKW.ndsu.2B.*2, and *QTKW.ndsu.4A.3* could be the same QTL for thousand-grain weight reported by Zanke et al. (2015). The QTL *QTKW.ndsu.4A.1* and *QTKW.ndsu.5B* seem to be novel QTL as they do not correspond with any reported QTL.

CONCLUSIONS

This study revealed 69 QTL, which included 50 constitutive QTL, three QTL identified for the control water regime, and 16 QTL exclusively under the drought conditions (**Table 4**; **Supplementary Figure 2**). These 16 QTL could be used for developing lines suitable for drought conditions. We also reported the QTL-associated genes, their physical positions, and predicted functions along with *in silico* expression prediction in abiotic stress conditions (**Supplementary Figure 3**). Chromosome 5B, 6B, and 4B seem to be very important for drought tolerance by reducing PH and increasing YLD and YLD-related traits. Several identified QTL occupied genomic regions reported earlier for earliness *per se*, drought tolerance, and reduced height. The consistency of some QTL in the different environments indicated their validity. Overall, this study provides valuable genetic and genomic resources to the breeders to design programs to breed drought-tolerant wheat cultivars, combining traditional and genomics-based approaches.

DATA AVAILABILITY STATEMENT

The list of markers and their genotypic data is available at figshare, doi: 10.6084/m9.figshare.14195348 and also in the **Supplementary Table 3**.

AUTHOR CONTRIBUTIONS

SR: data collection, analysis, and major write-up of the manuscript. SM: data collection and analysis. AK: data analyses, manuscript write-up, and review. EE, SK, and SSi: data analysis and manuscript review. MA and AM: data and

manuscript review. SSo and SSa: data analysis and write-up. MM: conceptualization and design of the experiment, phenotypic data collection, and manuscript write-up and review. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Late Dr. J. J. Hammond, Dr. A. El-Fatih El Dolfey, J. Underdahl, M. Abdallah, A. Walz, T. Selland, D. Olsen, K. McMonagle, C. Cossette, K. Dickey, K. Whitney, and K. Beck for their help and support. We would also like to thank the technical support team in the NDSU Research Centers at Minot, Williston, and Hettinger and at the Agronomy Farm at Casselton and Prosper for their help and supporting to conduct field trials. The authors would also like to thank Dr. Shiaoman Chao and Mary Osenga of Small Grains Genotyping Lab at USDA-ARS, Fargo, ND, for genoyping our samples using 90 K Infinium assay.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.649988/full#supplementary-material

REFERENCES

Abou-Elwafa, S. F., and Shehzad, T. (2021). Genetic diversity, GWAS and prediction for drought and terminal heat stress tolerance in bread wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol.* 68, 711–728. doi: 10.1007/s10722-020-01018-y

Ain, Q. U., Rasheed, A., Anwar, A., Mahmood, T., Imtiaz, M., Xia, X., et al. (2015). Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. *Front. Plant Sci*. 6:743. doi: 10.3389/fpls.2015.00743

Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., et al. (2018). Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.* 19, 1–10. doi: 10.1186/s13059-018-1491-4

Alexander, L. M., Kirigwi, F. M., Fritz, A. K., and Fellers, J. P. (2012). Mapping and quantitative trait loci analysis of drought tolerance in a Spring wheat population using amplified fragment length polymorphism and diversity array technology markers. *Crop Sci*. 52, 253–261. doi: 10.2135/cropsci2011.05.0267

Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Azadi, A., Mardi, M., Hervan, E. M., Mohammadi, S. A., Moradi, F., Tabatabaee, M. T., et al. (2014). QTL Mapping of yield and yield components under normal and salt-stress conditions in bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol. Rep*. 33, 102–120. doi: 10.1007/s11105-014-0726-0

Ballesta, P., Mora, F., and Pozo, A. D. (2020). Association mapping of drought tolerance indices in wheat: QTL-rich regions on chromosome 4A. *Sci. Agric.* 77, 2. doi: 10.1590/1678-992x-2018-0153

Blake, V. C., Birkett, C., Matthews, D. E., Hane, D. L., Bradbury, P., and Jannink, J. L. (2016). The triticeae toolbox: combining phenotype and genotype data to advance small-grains breeding. *Plant Genome* 9:99. doi: 10.3835/plantgenome2014.12.0099

Blake, V. C., Woodhouse, M. R., Lazo, G. R., Odell, S. G., Wight, C. P., Tinker, N. A., et al. (2019). GrainGenes: centralized small grain resources and digital platform for geneticists and breeders. *Database* 2019:baz065. doi: 10.1093/database/baz065

Blum, A. (1988). *Plant Breeding for Stress Environments*. Boca Raton, FL: CRC Press.

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170, 2172–2186. doi: 10.1104/pp.15.01667

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). Genetics and population analysis TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L., D'amore, R., Allen, A. M., et al. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491, 705–710. doi: 10.1038/nature11650

Climate Change and the Economy (2008). *North Dakota Assessing the costs of Climate Change*. Washington, DC: National Conference of State Legislatures. Available online at: http://www.ncsl.org/print/environ/ClimateChangeND.pdf (accessed December 15, 2016).

Dodig, D., M., Zoric, B., Kobiljski, J., Savic, V., Kandic, S., Quarrie, S., and Barnes, J. (2012). Genetic and association mapping study of wheat agronomic traits under contrasting water regimes. *Int. J. Mol. Sci.* 13:6167–6188. doi: 10.3390/ijms13056167

Edae, E. A., Byrne, P. F., Haley, S. D., Lopes, M. S., and Reynolds, M. P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet*. 127, 791–807. doi: 10.1007/s00122-013-2257-8

Edae, E. A., Byrne, P. F., Manmathan, H., Haley, S. D., Moragues, M., Lopes, M. S., et al. (2013). Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. *Plant Genome* 6:13. doi: 10.3835/plantgenome2013.04.0010

Ergen, N. Z., and Budak, H. (2009). Sequencing over 13,000 expressed sequence tags from six subtractive cDNA libraries of wild and modern wheats following slow drought stress. *Plant Cell Env.* 32, 220–236 doi: 10.1111/j.1365-3040.2008.01915.x

Ergen, N. Z., Thimmapuram, J., Bohnert, H. J., and Budak, H. (2009). Transcriptome pathways unique to dehydration tolerant relatives of modern wheat. *Funct. Integr. Genomics* 9, 377–396. doi: 10.1007/s10142-009-0123-1

Frazen, D. W. (2003). *North Dakota Soil and Fertilizer Handbook*. Fargo, ND: North Dakota State University.

Gahlaut, V., Jaiswal, V., Singh, S., Balyan, H. S., and Gupta, P. K. (2019). Multi-locus genome wide association mapping for yield and its contributing traits in hexaploid wheat under different water regimes. *Sci. Rep.* 9:19486. doi: 10.1038/s41598-019-55520-0

Hertel, T. W. (2011). The global supply and demand for agricultural land in 2050: a perfect storm in the making?. *Am. J. Agric. Econ*. 93, 259–275. doi: 10.1093/ajae/aaq189

Huang, X. Q., Cloutier, S., Lycar, L., Radovanovic, N., Humphreys, D. G., Noll, J. S., et al. (2006). Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 113, 753–766. doi: 10.1007/s00122-006-0346-7

Ibrahim, S. E., Schubert, A., Pillen, K., and Léon, J. (2012a). QTL analysis of drought tolerance for seedling root morphological traits in an advanced backcross population of spring wheat. *Int. J. Agri. Sci.* 2, 619–629.

Ibrahim, S. E., Schubert, A., Pillen, K., and Léon, J. (2012b). Comparison of QTLs for drought tolerance traits between two advanced backcross populations of spring wheat. *Int. J. Agri. Sci.* 2, 216–227.

Kamran, A., Iqbal, M., Navabi, A., Randhawa, H., Pozniak, C., and Spaner, D. (2013). Earliness per se QTLs and their interaction with the photoperiod insensitive allele Ppd-D1a in the Cutler × AC Barrie spring wheat population. *Theor. Appl. Genet.* 126, 1965–1976. doi: 10.1007/s00122-013-2110-0

Khodakovskaya, M., Sword, C., Wu, Q., Perera, I. Y., Boss, W. F., Brown, C. S., et al. (2010). Increasing inositol (1, 4, 5)-trisphosphate metabolism affects drought tolerance, carbohydrate metabolism and phosphate-sensitive biomass increases in tomato. *Plant Biotechnol. J.* 8, 170–183. doi: 10.1111/j.1467-7652.2009.00472.x

Kirigwi, F. M., Van Ginkel, M., Brown-Guedira, G., Gill, B. S., Paulsen, G. M., and Fritz, A. K. (2007). Markers associated with a QTL for grain yield in wheat under drought. *Mol. Breed.* 20, 401–413. doi: 10.1007/s11032-007-9100-3

Kumar, A., Mantovani, E., Simsek, S., Jain, S., Elias, E. M., and Mergoum, M. (2019). Genome wide genetic dissection of wheat quality and yield related traits and their relationship with grain shape and size traits in an elite × non-adapted bread wheat cross. *PLoS ONE* 14:e0221826. doi: 10.1371/journal.pone.0221826

Kumar, A., Mantovani, E. E., Seetan, R., Soltani, A., Echeverry-Solarte, M., Jain, S., et al. (2016). Dissection of genetic factors underlying wheat kernel shape and size in an elite × nonadapted cross using a high density SNP linkage map. *Plant Genome* 9:81. doi: 10.3835/plantgenome2015.09.0081

Lanceras, J. C., Pantuwan, G., Jongdee, B., and Toojinda, T. (2004). Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiol.* 135, 384–399. doi: 10.1104/pp.103.035527

Li, C., Bai, G., Chao, S., Carver, B., and Wang, Z. (2016). Single nucleotide polymorphisms linked to quantitative trait loci for grain quality traits in wheat. *Crop J.* 4, 1–11. doi: 10.1016/j.cj.2015.10.002

Li, X. M., Xia, X. C., Xiao, Y. G., He, Z. H., Wang, D. S., Trethowan, R., et al. (2015). QTL mapping for plant height and yield components in common wheat under water limited and full irrigation environments. *Crop Pasture Sci.* 67, 660–670. doi: 10.1071/CP14236

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, G., Xu, S. B., Ni, Z. F., Xie, C. J., Qin, D. D., Li, J., et al. (2011). Molecular dissection of plant height QTLs using recombinant inbred lines from hybrids between common wheat (*Triticum aestivum* L.) and spelt wheat *(Triticum spelta L.). Chin. Sci. Bull.* 56, 1897–1903. doi: 10.1007/s11434-011-4506-z

Malik, S., Malik, T. A., and Engineering, G. (2015). Genetic mapping of potential QTLs associated with drought tolerance in wheat. *J. Anim. Plant Sci.* 25, 1032–1040.

Mathew, I., Shimelis, H., Shayanowako, A. II, Laing, M., and Chaplot, V. (2019). Genome-wide association study of drought tolerance and biomass allocation in wheat. *PLoS ONE* 14:e0225383. doi: 10.1371/journal.pone.0225383

Maulana, F., Huang, W., Anderson, J. D., and Ma, X. (2020). Genome wide association mapping of seedling drought tolerance in winter wheat. *Front. Plant Sci.* 11:573786. doi: 10.3389/fpls.2020.573786

McCartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., Noll, J., et al. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452x'AC Domain'. *Genome* 48, 870–883. doi: 10.1139/g05-055

North Dakota Agricultural Weather Network (NDAWN) (2015). Available at: https://ndawn.ndsu.nodak.edu

North Dakota Wheat Commission (2016). *Building Bigger Better Mark*. Available online at: http://www.ndwheat.com/buyers/default.asp?ID=294 (verified 15 December 2016).

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959 doi: 10.1093/genetics/155.2.945

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361:6089. doi: 10.1126/science.aar6089

Sapkota, S., Hao, Y., Johnson, J., Buck, J., Aoun, M., and Mergoum, M. (2019). Genome-wide association study of a worldwide collection of wheat genotypes reveals novel quantitative trait loci for leaf rust resistance. *Plant Genome* 12:190033. doi: 10.3835/plantgenome2019.05.0033

SAS Institute (2004). *SAS Online Doc, v. 9.1.2*. Cary, NC: SAS Institute.

Simmons, S. R., Oelke, E. A., and Anderson, P. M. (1914). *Growth and Development Guide for Spring Wheat, AG-FO-2547*. Minneapolis, MN: Agricultural Extension at the University of Minnesota.

Sun, Y., Solomon, S., Dai, A., and Portmann, R. W. (2006). How often does it rain? *J. Clim.* 19, 916–934. doi: 10.1175/JCLI3672.1

Tura, H., Edwards, J., Gahlaut, V., Garcia, M., Sznajder, B., Baumann, U., et al. (2020). QTL analysis and fine mapping of a QTL for yield-related traits in wheat grown in dry and hot environments. *Theor. Appl. Genet.* 133, 239–257 doi: 10.1007/s00122-019-03454-6

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Wilhelm, E. (2011). *Genetic Analysis of the Group IV Rht LOCI in Wheat*. (Doctoral thesis) University of East Anglia, Norwich, United Kingdom.

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., et al. (2014a). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet.* 10:e1004573. doi: 10.1371/journal.pgen.1004573

Yang, Y., Wang, Q., Chen, Q., Liao, R., Zhang, X., Yang, H., et al. (2014b). A new genotype imputation method with tolerance to high missing rate and rare variants. *PLoS ONE* 9:1025. doi: 10.1371/journal.pone.0101025

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160 doi: 10.1016/j.copbio.2006.02.003

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2014). Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). *PLoS ONE* 9:e113287. doi: 10.1371/journal.pone.0113287

Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2015). Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. *Front. Plant Sci.* 6:644. doi: 10.3389/fpls.2015.00644

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Rabbi, Kumar, Mohajeri Naraghi, Simsek, Sapkota, Solanki, Alamri, Elias, Kianian, Missaoui and Mergoum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

DATA REPORT
published: 02 July 2021
doi: 10.3389/fgene.2021.651298

Draft Genome Assembly and Transcriptome Dataset for European Turnip (*Brassica rapa* L. ssp. *rapifera*), ECD4 Carrying Clubroot Resistance

Sin-Gi Park[1†], Eonji Noh[1†], SuRyun Choi[2], Boram Choi[1], In-Gang Shin[1], Seung-il Yoo[1], Dong Jin Lee[1], Sumin Ji[1], Hae-Suk Kim[1], Yoon-Jung Hwang[3], Jung Sun Kim[4], Jacqueline Batley[5], Yong Pyo Lim[2], David Edwards[5] and Chang Pyo Hong[1*]*

[1] Theragen Bio Co., Ltd., Suwon, South Korea, [2] Department of Horticulture, College of Agriculture and Life Science, Chungnam National University, Daejeon, South Korea, [3] Department of Chemistry Life Science, Sahmyook University, Seoul, South Korea, [4] Genomics Division, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju, South Korea, [5] School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, WA, Australia

Keywords: ***Brassica rapa*** **L. ssp.** ***rapifera*****, clubroot resistance, draft genome, European turnip, genome assembly, orthologous gene cluster,** ***R*** **gene**

OPEN ACCESS

Edited by:
Ajay Kumar,
North Dakota State University, United States

Reviewed by:
Chenggen Chu,
Edward T. Schafer Agricultural Research Center (USDA-ARS), United States
Ping Lou,
Dartmouth College, United States
Kumar Paritosh,
University of Delhi, India
Sampath Perumal,
University of Saskatchewan, Canada

***Correspondence:**
David Edwards
Dave.Edwards@uwa.edu.au
Chang Pyo Hong
changpyo.hong@theragenbio.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *09 January 2021*
Accepted: *01 June 2021*
Published: *02 July 2021*

Citation:
Park S-G, Noh E, Choi S, Choi B, Shin I-G, Yoo S-i, Lee DJ, Ji S, Kim H-S, Hwang Y-J, Kim JS, Batley J, Lim YP, Edwards D and Hong CP (2021) Draft Genome Assembly and Transcriptome Dataset for European Turnip (Brassica rapa L. ssp. rapifera), ECD4 Carrying Clubroot Resistance. Front. Genet. 12:651298. doi: 10.3389/fgene.2021.651298

INTRODUCTION

The *Brassica* genus consists of economically important oil and leafy vegetable crops, which are cultivated worldwide. The *Brassica* species represent as the "U's triangle" (Nagahara, 1935), which includes the three basic diploid species *Brassica rapa* (A genome), *Brassica nigra* (B genome), and *Brassica oleracea* (C genome), as well as the three amphidiploid species *Brassica juncea* (A and B genomes), *Brassica napus* (A and C genomes), and *Brassica carinata* (B and C genomes). *Brassica rapa* ($2n = 2x = 20$) has been cultivated for specific phenotypic characteristics such as heading (i.e., ssp. *pekinensis*; Chinese cabbage) and non-heading (i.e., ssp. *chinesis*; pak choi) leafy vegetables, tuberized hypocotyl/roots (i.e., ssp. *rapifera*; turnip), and oil rich seedpods (i.e., ssp. trilocularis; yellow sarson) in the oil crops. Although the genomes of its subspecies are very similar (Lin et al., 2014), *B. rapa* demonstrates extreme morphological diversity. Because of the economical value and scientific interest in phenotypic diversity, the genome sequence of mesopolyploid *B. rapa* ssp. *pekinensis* Chiifu, a Chinese cabbage, was the first published *B. rapa* reference genome, and revealed that *B. rapa* evolved via a two-step whole-genome triplication and, as a result, has three syntenic subgenomes (Wang et al., 2011). The polyploidization is hypothesized to have facilitated the diversification of genes as well as gene fractionation, and as a consequence led to the evolution of different morphotypes within and between related *Brassica* species. The study of intraspecific diversity in *B. rapa* offers opportunities to advance our understanding of plant growth, development, and phenotypic evolution (Paterson et al., 2001).

Over the last few years, the genome assembly of *B. rapa* (version 3.0) has been improved using single-molecule sequencing, optical mapping, and chromosome conformation capture technologies (Hi-C), resulting in an approximately 30-fold improvement with a contig N50 size of 1.45 Mb compared with that of previous references (Zhang et al., 2018). The assembly refined the syntenic relationship of genome blocks and centromere locations in the genome, and identified a greater number of annotated transposable elements (TEs) than in previous assemblies. In addition to the whole-genome sequence of Chinese cabbage, chromosome-level genome sequences of other *B. rapa* subspecies, including yellow sarson (Belser et al., 2018) and pak choi (Li P. et al., 2020; Li Y. et al., 2020), have been recently reported. The studies provide insight to the understanding of

genetic drivers underlying the morphological variation among *B. rapa* subspecies. Moreover, a pangenome from Chinese cabbage, rapid-cycling *Brassica*, and Japanese vegetable turnip was constructed using Illumina short-read data (Lin et al., 2014), identifying genomic determinants of morphological variation, especially copy number differences in peroxidases associated with the phenylpropanoid biosynthetic pathway in turnip.

Turnip (*B. rapa* L. ssp. *rapifera*) represent one of the morphotypes in *B. rapa* that forms tubers (hypocotyl/taproot tubers), produces lobed leaves with long petioles, and can be used to study the genetics underlying storage organ formation (Zhang et al., 2014). Brassica species are susceptible to clubroot disease, caused by *Plasmodiophora brassicae* (Schwelm et al., 2015), and forms galls (clubs) on infected root tissues with abnormal proliferation, preventing water and nutrient uptake and retarding the normal growth and development of plants, resulting in significantly reduced yield and quality. Sources of known clubroot resistance genes are derived from European turnip, carrying strong resistance to this disease (Matsumoto et al., 1998; Hirani et al., 2018). The resistance genes of European turnip have been introduced into other *Brassica* crops including vegetables and oilseed rape. In particular, European clubroot differential (ECD) turnips, that exhibited high levels of resistance to clubroot and that consist of four accessions (ECD1–ECD4), were developed and have been helpful for discovering dominant loci conferring clubroot resistance genes including *CRa*, *CRb*, and *CRc* for marker-assisted selection in canola and other *Brassica* species (Hirani et al., 2018).

Here, we report the draft genome assembly of a European turnip, ECD4, which has strong clubroot resistance, that was generated by PacBio single-molecule long-read sequencing technology. This draft genome assembly coupled with transcriptome data derived from various leaf and root tissues will support the discovery of disease resistance genes (*R* genes) especially clubroot resistance genes, enabling the development of allele-specific markers for marker-assisted selection in *Brassica* breeding. Moreover, these data provide a valuable resource for studying the morphological diversity and evolution of turnips.

DATA BRIEFS

Whole-Genome *de novo* Assembly of European Turnip

For the genome assembly of European turnip (ECD4), a total of 70.63 Gb of PacBio long reads (an average sequencing coverage of 136.36x) and 72.16 Gb of Illumina short reads (139.3x) were generated (**Supplementary Table 1**). Based on *k*-mer analysis ($k = 21$), the estimated genome size of European turnip is approximately 518 Mbp (**Figure 1A**). The genome assembly resulted in a 315.8 Mb draft genome with 655 contigs (contig N50 length of 1.45 Mb; the longest contig length was 21.92 Mb) (**Table 1**). Our assembly data covered about 61% of the genome of European turnip, being predicted as mostly euchromatins. The assembly coverage is slightly low compared with those reported in Chinese cabbage (353 Mb) (Zhang et al., 2018), pak choi (370 Mb) (Li P. et al., 2020; Li Y. et al., 2020), and yellow sarson (402 Mb) (Belser et al., 2018). Whole-genome sequences of Chinese cabbage, pak choi, and yellow sarson were assembled at the chromosome-level by using chromosome conformation capture (Hi-C) technology [Chinese cabbage (reference genome version 3.0) and pak choi] and/or BioNano optical mapping [Chinese cabbage (v3.0) and yellow sarson] besides long-read sequencing such as PacBio SMRT sequencing or Oxford Nanopore sequencing. Therefore, such long-range scaffolding technologies are further required to improve the quality of genome assembly in European turnip. While the remained 39% is expected to contain considerably repetitive DNAs, including tandem satellite repeats, rDNAs, and retrotransposons which can disturb assembly (Wang et al., 2011; Perumal et al., 2017; Zhang et al., 2018; Li P. et al., 2020). Of the contigs, 260 were anchored to 10 chromosomes of Chinese cabbage (cv. Chiifu-401) (version 3.0), with a total length of 260.47 Mb (**Figure 1B**; **Supplementary Table 2**).

The quality of the genome assembly was assessed using BUSCO and paired-end read mapping. First, the draft genome assembly captured 99.6% of the complete Benchmarking Universal Single-Copy Orthologs (BUSCOs) with the Virdiplantae_odb10 database; 82.4, 17.2, 0.2, and 0.2% of the BUSCOs were predicted as complete and single-copy, complete and duplicated, fragmented, and missing, respectively. The high completeness of our assembly data seems to be resulted from error correction and curation of heterozygous assemblies by Pilon and Purge_haplotigs with high read depth and coverage (read depth: 139.3x of short reads for Pilon and 136.4x of long reads for Purge_haplotigs). In particular, the method led to a saturation of BUSCO scores with one round of error correction in Pilon (**Supplementary Table 10**). The genome completeness of European turnip is similar to that of Chinese cabbage (**Supplementary Table 3**). Second, 94.58% of the short-insert paired-end reads were successfully realigned to the assembly, with 99.84% assembly coverage. Moreover, the analysis showed that 0.32% of the nucleotides in the genome assembly were heterozygous.

Identification of Genome Rearrangements Between European Turnip and Chinese Cabbage

The whole-genome comparison revealed eight genomic rearrangements between European turnip and Chinese cabbage ("G" layer in **Figure 1B**), with the identification of hotspots (**Supplementary Table 4**; **Supplementary Figure 1**), indicating genomic change since the divergence from a common ancestor of *B. rapa*. For example, the positions 629.5 to 774.9 kb (145.4 kb in length) and positions 11.5 to 579.4 kb (567.9 kb in length) on contig00007136 of European turnip were aligned to positions 18,057.7 to 18,217.2 kb (159.5 kb in length) and positions 18,966.7 to 19,568.7 kb (602 kb in length) on chromosome A10 of Chinese cabbage, respectively. These hotspots were validated by mapping Illumina paired-end reads of European turnip to the corresponding hotspot of Chinese cabbage and then identifying the split of paired-end read mapping in Chinese cabbage (**Supplementary Figure 8**). We further identified five and three genome rearrangements between pak choi and European turnip, and between yellow sarson and European turnip, respectively,

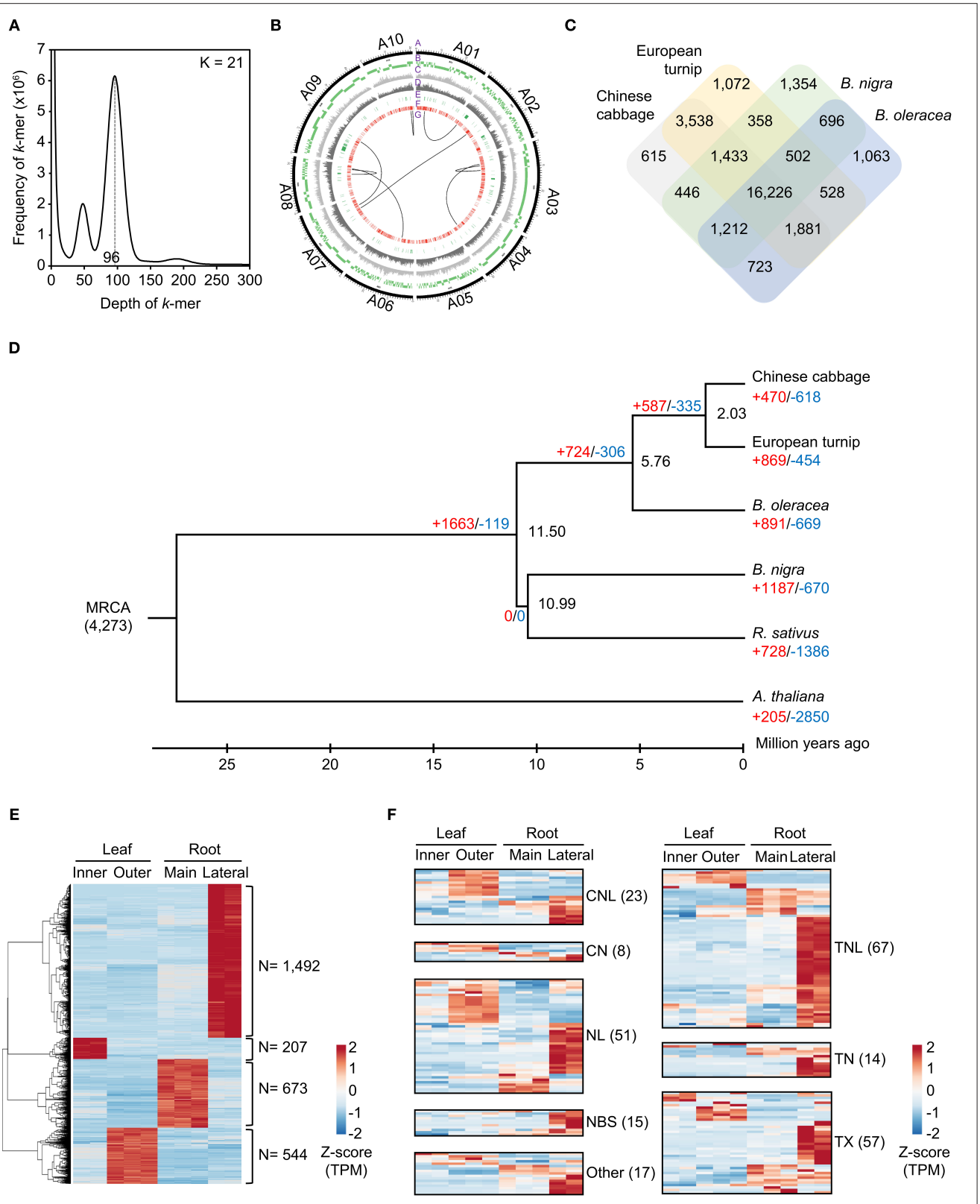


FIGURE 1 | Characterization of the draft genome assembly for the European turnip ECD4. **(A)** Estimation of genome size on the basis of *k*-mer frequency analysis. **(B)** Anchoring genome assembly data of European turnip to 10 chromosomes of Chinese cabbage (A1–A10). In the Circos plot, A, B, C, D, E, F, and G indicate

(Continued)

FIGURE 1 | chromosomes of Chinese cabbage, alignment of genome assembly data of European turnip, SNPs, InDels, NBS-encoding *R* genes, membrane-associated *R* genes, and translocations, respectively. **(C)** Orthologous gene clusters among Chinese cabbage, European turnip, *B. nigra*, and *B. oleracea*. **(D)** Expansion and contraction of gene families based on a time-calibrated phylogeny of four *Brassica* species, *R. sativus*, and the outgroup *A. thaliana*. **(E)** Identification of genes showing tissue-specific expression in the inner and outer leaves and the main and lateral root tissues of European turnip. **(F)** Genome-wide expression of NBS-encoding *R* genes identified in the leaf and root tissues of European turnip.

showing large genome variations among *B. rapa* subspecies (**Supplementary Figure 7**). Additionally, 3,795,367 SNPs, with a transition/transversion ratio (Ts/Tv) of 1.35, and 954,051 InDels were identified ("C" and "D" layers in **Figure 1B**).

Genome Annotation: Identification of TEs and Gene Prediction and Annotation

A total of 112.6 Mb of TEs were identified, accounting for 35.66% of the genome assembly (**Supplementary Table 5**). Of the class I elements, *Ty3/Gypsy* (29.5%) and *Ty1/Copia* (18.7%) for long terminal repeat (LTR) retrotransposons and LINE (L1) for non-LTR retrotransposons were abundant in the genome. The *hAT* family (11.7%) represented the most abundant DNA transposon (**Supplementary Table 5**). A total of 48,349 non-redundant protein-coding genes were predicted with an average number of exons per gene of 5.23 using evidence-driven gene prediction methods coupled with *ab initio* prediction (**Table 1**). The gene models were supported by 93.04% RNA-Seq data derived from two leaves and two root tissues of European turnip (**Supplementary Table 6**). Of the predicted genes, 45,622 (94.36%) were successfully annotated by at least one database, including UniProt/SwissProt, TAIR, NCBI NR, and InterPro. To investigate species-specific and shared genes in European turnip compared with other *Brassica* diploids, including Chinese cabbage (AA), *B. nigra* (BB), and *B. oleracea* (CC), we analyzed orthologous gene clusters. This analysis showed that 24,466 of 31,647 orthologous gene clusters were shared among European turnip and three *Brassica* diploids and that 1,072 orthologous gene clusters were species-specific in European turnip (**Figure 1C**). This highlighted the enrichment of genes with F-box domains that play various roles in developmental processes, including plant hormone signal transduction, floral development, secondary metabolism, senescence, circadian rhythm, and response to biotic and abiotic stress (Xu et al., 2009) (**Supplementary Figure 2**). Of the orthologous gene clusters, five *R* genes (three TIR- and two CC-NBS-LRR-containing domain proteins) were found. We also identified 709, 672, and 463 orthologous gene clusters specific to European turnip, yellow sarson, and Chinese cabbage, respectively (**Supplementary Figure 9A**). Similar to above finding, the enrichment of genes with F-box and leucine-rich repeat (LRR) domains was identified in orthologous gene clusters specific to European turnip (**Supplementary Figure 9B**), reflecting the characteristics of plant disease resistance in European turnip.

TABLE 1 | Genome assembly and gene prediction of European turnip.

Genome assembly	
No. of contigs	655
Total length (Mb)	315.77
Contig N50 length (Mb)	1.45
Longest contig length (Mb)	21.92
GC content (%)	35.58%
Gene prediction	
No. of genes	48,349
Average length of gene (bp)	1,907
Exon	
No. of exons	266,261
Average number of exons per gene	5.23
Average exon length (bp)	215
Intron	
No. of introns	215,413
Average number of introns per gene	4.23
Average intron length (bp)	183
Homology-based functional annotation	45,622 (94.36%)
UniProt/SwissProt	31,538 (65.23%)
TAIR	40,220 (83.19%)
NCBI NR	44,110 (91.23%)
InterPro domains	43,417 (89.80%)

Gene Family Expansions and Contractions

A phylogenetic tree of European turnip was constructed using a total of 3,139 single-copy orthologous genes (**Figure 1D**). Through orthologous gene cluster analysis, we identified that European turnip has 869 expanded and 454 contracted gene families (red- and blue-colored numbers in **Figure 1D**). Of the expanded gene families, protein domains found in genes associated with plant development (no apical meristem-associated C-terminal) and plant defense (haloacid dehalogenase-like hydrolase, xylanase inhibitor C-terminal, tetratricopeptide repeat, and oxidative-stress-responsive kinase 1 C-terminal) were enriched (**Supplementary Figure 3**).

Tissue-Specific Gene Expression in European Turnip

Tissue-specific gene expression was identified in inner leaf (207 genes), outer leaf (544 genes), main root (673 genes), and lateral root (1,492 genes) tissues, with a statistical cut-off of $q < 0.05$, absolute 2-fold change, and at least one tissue with ≥10 FPKMs (**Figure 1E**). Functional annotation for these genes revealed the functions for each of tissues; for example, chloroplast organization ($P = 1.1 \times 10^{-12}$) for inner leaves, phosphorylation ($P = 1.2 \times 10^{-23}$), and defense response ($P = 7.1 \times 10^{-6}$) for outer leaves, cell wall organization ($P = 6.7 \times 10^{-9}$), response to salicylic acid ($P = 1.7 \times 10^{-5}$), and post-embryonic morphogenesis ($P = 5.6 \times 10^{-5}$) for the main root, and transcription ($P = 2.2 \times 10^{-19}$), response to salicylic acid ($P = 3.9 \times 10^{-8}$), wounding ($P = 5.8 \times 10^{-8}$), toxic substance ($P = 4.4 \times 10^{-7}$), abscisic acid ($P = 5.7 \times 10^{-8}$), and jasmonic

acid (2.2×10^{-6}) for lateral roots were significantly represented (**Supplementary Figure 4**).

Identification of Disease Resistance Genes (*R* genes) in European Turnip

We also found a total of 252 plant defense-related *R* genes: 23 CC-NBS-LRRs (CNLs), 8 CC-NBSs (CNs), 67 TIR-NBS-LRRs (TNLs), 14 TIR-NBSs (TNs), 51 NBS-LRRs (NLs), 15 NBSs, 57 TIR-others (TXs), and 17 others (**Supplementary Table 7**). Interestingly, 103 *R* genes exhibited high expression in lateral root tissue (**Figure 1F**). In addition to these *R* genes, a total of 1,135 genes−119 receptor-like proteins (RLPs), 742 receptor-like kinases (RLKs), and 274 TM-CCs—categorized into membrane-associated *R* genes were identified (**Figure 1F**; **Supplementary Table 7**). Of those three gene families, 243 RLKs also showed high expression in the lateral roots (**Supplementary Figure 5**). Additionally, two contigs harboring microsatellite markers linked to clubroot resistance loci *BraA.CR.a* and *BraA.CR.b* (Hirani et al., 2018), which are located on chromosomes A3 and A8, respectively, were found (**Supplementary Table 8**; **Supplementary Figure 6**). Interestingly, the comparison of *R* genes between the linked regions of the two species showed the expansion of *R* genes, mostly *R* genes of TM-CC and TNL classes, in European turnip even though exhibiting their co-linearity in the linked regions (**Supplementary Table 9**). Moreover, three R genes homologous to *Rcr6*, which is a clubroot resistance gene of TNL class identified in *B. nigra* (Chang et al., 2019), were identified in European turnip, but only one was identified in Chinese cabbage (**Supplementary Table 9**).

Summary

We report the first draft genome of European turnip (ECD4) coupled with transcriptome data derived from various tissues. Our data can provide an invaluable genomic resource to study the morphological diversity and evolution of European turnip by applying pan-genomics to *Brassica*. Moreover, our data can help develop molecular markers as a genetic breeding tool to identify plant defense-related genes such as clubroot resistance genes (*R* genes).

MATERIALS AND METHODS

Sample Collection, Library Construction, and Sequencing

Seeds of European turnip ECD4 (*B. rapa* ssp. *rapifera*) (accession no. 25026) were obtained from the Korea Brassica Genome Resource Bank (KBGRB, South Korea). The seeds were germinated in seedling trays containing autoclaved soil in a controlled chamber at 25°C, 60% humidity, and photoperiodic lighting (16 h of light:8 h of dark). From leaf tissue of 3-week-old seedlings, genomic DNA was extracted by a WizPrep Plant DNA Mini Kit (Wizbiosolutions, Seongnam, South Korea) according to the manufacturer's protocol for whole-genome sequencing. A long-read library with high-quality genomic DNA of ≥20 μg was prepared using the SMRTbell Express Template prep kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on a PacBio Sequel System using one SMRT cell. Additionally, DNA sequence libraries were prepared from 1 μg input DNA using a TruSeq Nano DNA Sample Prep Kit according to the manufacturer's instructions (Illumina, Inc., San Diego, CA, USA). The libraries were subjected to paired-end sequencing with a 150-bp read length using the Illumina NovaSeq 6000 platform.

Whole-Genome Assembly

Illumina short reads were processed by Jellyfish (version 2.2.0) (Marcais and Kingsford, 2011) to estimate the genome size of European turnip. The *k*-mer frequency with a *k*-mer size of 21 was counted and plotted as a *k*-mer frequency distribution. PacBio SMRT long reads were assembled using CANU (version 1.8) (Koren et al., 2017) with the following primary parameters; genomeSize 518M, minReadLength 1000, minOverlapLength 500, rawErrorRate 0.3, correctedErrorRate 0.045. For assembly polishing, short reads were aligned to the CANU assembly data using BWA-MEM (version 0.7.17) (Li and Durbin, 2009), and errors in nucleotides in the CANU assembly data were then corrected using Pilon (version 1.22) (Walker et al., 2014) with one round of polishing. The polished contigs were processed with Purge_haplotigs to curate heterozygous diploid genome assemblies (Roach et al., 2018). After filtering of ≤10 kb of contigs, a total of 655 assembled contigs were finally generated. This method was summarized in **Supplementary Figure 10**.

The genome completeness of the draft genome assembly of European turnip was assessed by using BUSCO (version 4.0.5) (Seppey et al., 2019) with 425 single-copy orthologs of the viridiplantae_odb10 database. For estimation of sequence coverage, Illumina short-insert paired-end reads were realigned to the draft genome assembly using BWA-MEM. Furthermore, heterozygous nucleotides were identified by using SAMtools mpileup (Li, 2011).

Genome Annotation

A combination of *ab initio* and evidence-based approaches was employed for gene prediction of European turnip. The assembled genome was premasked for repetitive DNA sequences using RepeatMasker (version 4.0.6) (http://www.repeatmasker.org/). An unsupervised training gene structure was generated using GeneMark-ET (version 4.10) (Lomsadze et al., 2014) by incorporating RNA-Seq data. *De novo* prediction was performed using AUGUSTUS (version 3.3.1) (Keller et al., 2011) with the training gene set and with exon–intron boundary information predicted by RNA-Seq and protein sequence alignments. Here, STAR (version 2.7.1a) (Dobin et al., 2013) was used for RNA-Seq alignment of the species, and GenomeThreader (version 1.7.0) (Gremme et al., 2005) was used for protein sequence alignment with the *B. rapa* protein sequences (version 3.0). Functional annotation for predicted genes was performed on the basis of homology-based searches with TAIR11, UniProt/SwissProt, and NCBI non-redundant (NR) databases using BLASTP (version 2.3.0+) (Altschul et al., 1990) with a cutoff *E*-value of 1*E*-10. Protein domains were also searched by using InterProScan (version 5.19-58.0) (Mulder and Apweiler, 2007). Additionally, tRNA- and miRNA-like sequences were predicted using tRNAscan-SE (version 1.4 alpha) (Schattner et al., 2005) and Infernal (version 1.1.1) (Nawrocki and Eddy, 2013). Resistance gene analogs (RGAs), such as NBS-encoding proteins,

RLKs and RLPs, were predicted using an RGAugury pipeline (Li et al., 2016).

Ortholog and Phylogenetic Analysis

The protein sequences from turnip (AA) and diploid *Brassica* species, including Chinese cabbage (AA; reference genome v3.0), *B. nigra* (BB; v1.1), *B. oleracea* (CC; v1.1), *Raphanus sativus* (v1.1), and the outgroup species *Arabidopsis thaliana* (version TAIR10) were used to predict orthologous gene clusters. The gene set of each species was filtered as follows: First, the genes encoding proteins of fewer than 30 amino acids were filtered out. Second, the similarity relation between the protein sequences of all of the species was obtained through BLASTP (i.e., all-vs.-all BLASTP searches) with a cutoff E-value of 1*E*-05. All of the protein datasets of the representative species were clustered into paralogues and orthologs by OrthoMCL (version 2.0.9) (Li et al., 2003), with an inflation parameter of 1.5.

The single-copy orthologs in these six species were aligned using MAFFT (version 7.123b) (Katoh and Standley, 2013) and then used to extract conserved regions with Gblocks (version 0.91b) (Talavera and Castresana, 2007). The alignment results were combined to create a superalignment matrix. A phylogenetic tree of these six species was constructed using MEGA X (Kumar et al., 2018) with the maximum likelihood method and 1,000 bootstrap values. The divergence time of European turnip was estimated using RelTime of MEGA X.

Gene Family Expansion and Contraction Analysis

The expansion and contraction of the gene families were analyzed by comparing the cluster size differences between the ancestor and each species by using café (De Bie et al., 2006). A random birth and death model was used to study the changes in gene families along each lineage of the phylogenetic tree. A probabilistic graphical model (PGM) was introduced to calculate the probability of transitions in gene family size from parent to child nodes in the phylogeny. With conditional likelihoods as the test statistics, the corresponding *p*-value in each lineage was calculated, and a *p*-value of 0.05 was used to identify the families that were significantly expanded or contracted. The expanded and contracted genes were then subjected to PFAM functional annotation.

Transcriptome Analysis

Three-week-old seedlings were transplanted to pots and grown for 4 weeks in a glasshouse. Total RNA from four tissues (the inner leaf, outer leaf, main root, and lateral root) was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). RNA-Seq libraries from 1 μg of purified total RNA were prepared using the TruSeq Stranded mRNA Sample Prep Kit according to the manufacturer's manual (Illumina, Inc., San Diego, CA). cDNA was synthesized and then subjected to an end repair process, addition of a single "A" base, and ligation of the adaptors. Libraries that were purified and enriched with PCR amplification were subjected to paired-end sequencing with a 150 bp read length using an Illumina NovaSeq 6000 platform.

After removal of adaptor sequences and trimming of low-quality sequences (<Q20) using Cutadapt (version 2.8), clean reads were aligned to the draft genome assembly using STAR (version 2.7.1a). Gene expression quantification was performed using RSEM (version 1.3.1) (Li and Dewey, 2011) with TPM values (version 1.3.1). For analysis of tissue-specific expression in the inner/outer leaves and main/lateral roots, differential expression analysis was performed using DESeq2 (version 1.26.0) (Love et al., 2014) with a cutoff of $q < 0.05$ and absolute ≥2-fold change. Functional annotation for genes showing tissue-specific expression was performed by using DAVID (Huang et al., 2007), and relevant gene ontology (GO) terms were selected with a cutoff EASE score $<1 \times 10^{-4}$. The expression pattern of the analyzed genes was visualized as a heatmap by using ClustVis (Metsalu and Vilo, 2015) with the correlations and the average linkage method.

Comparative Genomics

Syntenic blocks between the chromosomes of Chinese cabbage and contigs of European turnip were detected by using SyMAP (version 4.2) (Soderlund et al., 2006) and are shown in a Circos plot. Genome rearrangement events such translocation between the two genomes were verified by using MUMmer (version 3.9.4) (Delcher et al., 2003).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories . Sequencing data used in this study are available in the NCBI Sequence Read Archive (SRA) database under the following accession PRJNA690160; JAEOXI000000000 (whole genome assembly), SRR13375884 (PacBio SMRT long read sequencing data), SRR13375883 (Illumina DNA-Seq data), SRR13375896 - SRR13375905 (RNA-Seq data derived from four different tissues). In addition, datasets for functional annotation are available from our website http://turnip.genome.theragenbio.com/BRT_1.0.

AUTHOR CONTRIBUTIONS

CH and DE conceived this study. S-GP, BC, I-GS, S-iY, and DL performed the bioinformatics analyses. EN, SJ, H-SK, Y-JH, and JK performed the whole-genome and RNA sequencing. SC and YL prepared the samples and extracted genomic DNA and total RNA. CH, S-GP, JB, and DE wrote the manuscript. All authors read and approved the final manuscript, contributed to the article, and approved the submitted version.

FUNDING

This work was supported by the Cooperative Research Program for National Agricultural Genome Program (Grant No. PJ01338203), Rural Development Administration (RDA), South Korea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.651298/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Chang, A., Lamara, M., Wei, Y., Hu, H., Parkin, I. A. P., Gossen, B. D., et al. (2019). Clubroot resistance gene *Rcr6* in *Brassica nigra* resides in a genomic region homologous to chromosome A08 in *B. rapa*. *BMC Plant Biol.* 19:224. doi: 10.1186/s12870-019-1844-5

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi: 10.1093/bioinformatics/btl097

Delcher, A.L., Salzberg, S.L., and Phillippy, A.M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics* Chapter 10:Unit 10.3. doi: 10.1002/0471250953.bi1003s00

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Gremme, G., Brendel, V., Sparks, M.E., and Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* 47, 965–978. doi: 10.1016/j.infsof.2005.09.005

Hirani, A.H., Gao, F., Liu, J., Fu, G., Wu, C., McVetty, P.B.E., et al. (2018). Combinations of independent dominant loci conferring clubroot resistance in all four turnip accessions (*Brassica rapa*) from the European clubroot differential set. *Front. Plant Sci.* 9:1628. doi: 10.3389/fpls.2018.01628

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., et al. (2007). The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8:R183. doi: 10.1186/gb-2007-8-9-r183

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keller, O., Kollmar, M., Stanke, M., and Waack, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27, 757–763. doi: 10.1093/bioinformatics/btr010

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, L., Stoeckert, C.J. Jr., and Roos, D.S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S., and You, F.M. (2016). RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* 17:852. doi: 10.1186/s12864-016-3197-x

Li, P., Su, T., Zhao, X., Wang, W., Zhang, D., Yu, Y., et al. (2020). Assembly of the non-heading pak choi genome and comparison with the genomes of heading Chinese cabbage and the oilseed yellow sarson. *Plant Biotechnol. J.* 19, 966–976.doi: 10.1111/pbi.13522

Li, Y., Liu, G. F., Ma, L. M., Liu, T. K., Zhang, C. W., Xiao, D., et al. (2020). A chromosome-level reference genome of non-heading Chinese cabbage [*Brassica campestris* (syn. *Brassica rapa*) ssp. *chinensis*]. *Hortic. Res.* 7:212. doi: 10.1038/s41438-020-00449-z

Lin, K., Zhang, N., Severing, E.I., Nijveen, H., Cheng, F., Visser, R.G., et al. (2014). Beyond genomic variation–comparison and functional annotation of three *Brassica rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genomics* 15:250. doi: 10.1186/1471-2164-15-250

Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* 42:e119. doi: 10.1093/nar/gku557

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

Matsumoto, E., Yasui, C., Ohi, M., and Tsukada, M. (1998). Linkage analysis of RFLP markers for clubroot resistance and pigmentation in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Euphytica* 104, 79–86. doi: 10.1023/A:1018370418201

Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* 43, W566–W570. doi: 10.1093/nar/gkv468

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol. Biol.* 396, 59–70. doi: 10.1007/978-1-59745-515-2_5

Nagahara, U. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn. J. Bot.* 7, 389–452.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Paterson, A.H., Lan, T.H., Amasino, R., Osborn, T.C., and Quiros, C. (2001). Brassica genomics: a complement to, and early beneficiary of, the Arabidopsis sequence. *Genome Biol.* 2:REVIEWS1011. doi: 10.1186/gb-2001-2-3-reviews1011

Perumal, S., Waminal, N.E., Lee, J., Lee, J., Choi, B. S., Kim, H. H., et al. (2017). Elucidating the major hidden genomic components of the A, C, and AC genomes and their influence on *Brassica* evolution. *Sci. Rep.* 7:17986. doi: 10.1038/s41598-017-18048-9

Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19:460. doi: 10.1186/s12859-018-2485-7

Schattner, P., Brooks, A.N., and Lowe, T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689. doi: 10.1093/nar/gki366

Schwelm, A., Fogelqvist, J., Knaust, A., Julke, S., Lilja, T., Bonilla-Rosso, G., et al. (2015). The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. *Sci. Rep*, 5:11153. doi: 10.1038/srep11153

Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0_14

Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Res.* 16, 1159–1168. doi: 10.1101/gr.5396706

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Xu, G., Ma, H., Nei, M., and Kong, H. (2009). Evolution of F-box genes in plants: different modes of sequence divergence and their relationships

with functional diversification. *Proc. Natl. Acad. Sci. U. S. A*. 106, 835–840. doi: 10.1073/pnas.0812043106

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., et al. (2018). Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Hortic. Res*. 5:50. doi: 10.1038/s41438-018-0071-9

Zhang, N., Zhao, J., Lens, F., de Visser, J., Menamo, T., Fang, W., et al. (2014). Morphology, carbohydrate composition and vernalization response in a genetically diverse collection of Asian and European turnips (*Brassica rapa* subsp. *rapa*). *PLoS ONE* 9:e114241. doi: 10.1371/journal.pone.0114241

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Park, Noh, Choi, Choi, Shin, Yoo, Lee, Ji, Kim, Hwang, Kim, Batley, Lim, Edwards and Hong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 August 2021
doi: 10.3389/fgene.2021.629267

Identification of Root Rot Resistance QTLs in Pea Using *Fusarium solani* f. sp. *pisi*-Responsive Differentially Expressed Genes

Bruce A. Williamson-Benavides[1], Richard M. Sharpe[2], Grant Nelson[1], Eliane T. Bodah[2†], Lyndon D. Porter[3] and Amit Dhingra[1,2]*

[1] Molecular Plant Sciences, Washington State University, Pullman, WA, United States, [2] Department of Horticulture, Washington State University, Pullman, WA, United States, [3] USDA-ARS, Grain Legume Genetics and Physiology Research Unit, Prosser, WA, United States

OPEN ACCESS

Edited by:
Pinky Agarwal, National Institute of Plant Genome Research (NIPGR), India

Reviewed by:
Ethan Andersen, Francis Marion University, United States
Tom Warkentin, University of Saskatchewan, Canada

***Correspondence:**
Amit Dhingra
adhingra@wsu.edu

†Present address:
Eliane T. Bodah, Pure Line Seeds, Inc., Warden, WA, United States

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *14 November 2020*
Accepted: *06 July 2021*
Published: *05 August 2021*

Citation:
Williamson-Benavides BA, Sharpe RM, Nelson G, Bodah ET, Porter LD and Dhingra A (2021) Identification of Root Rot Resistance QTLs in Pea Using Fusarium solani f. sp. pisi-Responsive Differentially Expressed Genes. Front. Genet. 12:629267. doi: 10.3389/fgene.2021.629267

Pisum sativum (pea) yields in the United States have declined significantly over the last decades, predominantly due to susceptibility to root rot diseases. One of the main causal agents of root rot is the fungus *Fusarium solani* f. sp. *pisi* (*Fsp*), leading to yield losses ranging from 15 to 60%. Determining and subsequently incorporating the genetic basis for resistance in new cultivars offers one of the best solutions to control this pathogen; however, no green-seeded pea cultivars with complete resistance to *Fsp* have been identified. To date, only partial levels of resistance to *Fsp* has been identified among pea genotypes. SNPs mined from *Fsp*-responsive differentially expressed genes (DEGs) identified in a preceding study were utilized to identify QTLs associated with *Fsp* resistance using composite interval mapping in two recombinant inbred line (RIL) populations segregating for partial root rot resistance. A total of 769 DEGs with single nucleotide polymorphisms (SNPs) were identified, and the putative SNPs were evaluated for being polymorphic across four partially resistant and four susceptible *P. sativum* genotypes. The SNPs with validated polymorphisms were used to screen two RIL populations using two phenotypic criteria: root disease severity and plant height. One QTL, *WB.Fsp-Ps* 5.1 that mapped to chromosome 5 explained 14.8% of the variance with a confidence interval of 10.4 cM. The other four QTLs located on chromosomes 2, 3, and 5, explained 5.3–8.1% of the variance. The use of SNPs derived from *Fsp*-responsive DEGs for QTL mapping proved to be an efficient way to identify molecular markers associated with *Fsp* resistance in pea. These QTLs are potential candidates for marker-assisted selection and gene pyramiding to obtain high levels of partial resistance in pea cultivars to combat root rot caused by *Fsp*.

Keywords: root rot, quantitative trait loci, SNP, molecular marker, RNAseq, *Pisum sativum* L., disease resistance

INTRODUCTION

Pea (*Pisum sativum* L.; Family Fabaceae) is an important cool-season, self-pollinating annual diploid crop. A number of cultivars within the species cater to different consumption markets. Green peas, and dry yellow and green peas are sold as food in the fresh and dry markets, respectively, while purple-seeded lines are used for forage and green manure (Miller et al., 2005).

Due to its high protein content (20–30%) and overall high nutritional value, pea has become a major contributor to the plant-based protein market (do Carmo et al., 2016; Peng et al., 2016; Xiong et al., 2018; Wei et al., 2020). The shift to plant-based protein is an environmentally sustainable alternative to animal-based protein because the latter contributes significantly to greenhouse gas emissions (Stehfest et al., 2009). Furthermore, studies have shown that dietary proteins in peas are of great benefit to human health and wellness (Reddy and Yang, 2011; Kudre et al., 2013; Dahiya et al., 2015). Consequently, the pea protein market was projected to reach $313.5 million in 2025 (Grand View Research, 2017; Sim et al., 2019).

The profitable production of pea is threatened by soilborne diseases. These diseases are commonly referred to as the pea root rot complex (PRRC) and are caused by a single or combination of pathogens, including *Aphanomyces euteiches*, *Fusarium* spp., *Mycosphaerella pinodes*, *Pythium* spp., and *Rhizoctonia solani* (Xue, 2003; Kumari and Katoch, 2020). One of the predominant causal agents of PRRC is the fungus *Fusarium solani* f. sp. *pisi* (*Fsp*). *Fsp* occurs in most pea fields throughout the world, and the yields of *P. sativum* cultivars can be reduced up to 15–62% by this pathogen (Seaman, 1976; Grünwald et al., 2003). *Fsp* infects pea seeds during germination, with symptoms of root rot beginning at or near the cotyledon-hypocotyl junction and progressing under the soil and upper region of the taproot (JM, Kraft and Pfleger, 2006). Round or irregular light brown lesions that progress to dark black lesions on below-ground stems have also been reported, along with stunting and death (Jung et al., 1999). *Fsp* can survive in the soil for more than one season and conditions that decrease root growth, such as soil compaction, extreme temperatures, and moisture levels, can increase *Fusarium*-mediated root damage (Shaykh et al., 1977; JM, Kraft and Pfleger, 2006).

The development of pea cultivars with root rot resistance has been considered the best long-term management option among the many root rot control strategies (Conner et al., 2014; Bodah et al., 2016; Wang et al., 2018). However, breeding for *Fsp* resistance is challenging, since resistance to *Fsp* is a quantitative trait (Mukankusi et al., 2011; Román-Avilés et al., 2011; Bodah et al., 2016). Furthermore, routine screening for resistance has proven to be time-consuming, expensive, and highly influenced by the environment (Bodah et al., 2016). Marker-assisted selection (MAS) can help expedite the selection of putative *Fsp* resistant progeny without the need for expensive phenotyping. Several efforts have been made to develop molecular markers associated with resistance to *Fsp* root rot in pea (Feng et al., 2011; Coyne et al., 2015, 2019). However, these studies used a limited number of DNA markers and some of the QTLs identified require further fine mapping to provide informative markers due to large confidence intervals (16.8–28.5 cM).

In a preceding time-course transcriptome study, we utilized a combined genetic and RNAseq approach to identify *Fsp*-responsive differentially expressed genes among four partially resistant and four susceptible genotypes (Williamson-Benavides et al., 2020). These genotypes were selected for their contrasting root severity index phenotype (Bodah et al., 2016). Genes involved in secretion and exocytosis, anthocyanin biosynthesis pathway genes, and a previously-described pathogenesis-related (PR) gene DRR230 were observed to be overexpressed in partially resistant genotypes (Chiang and Hadwiger, 1991; Hadwiger, 2008, 2015; Williamson-Benavides et al., 2020). Since the use of single nucleotide polymorphisms (SNPs) can help to refine genetic mapping studies due to their high abundance in the genome (Deulvot et al., 2010), SNPs mined from differentially expressed genes (DEGs) were utilized to identify QTLs associated with *Fsp* resistance using composite interval mapping in two recombinant inbred line (RIL) populations segregating for root rot resistance as observed in greenhouse evaluations.

MATERIALS AND METHODS

Plant Material

The parental plant material used in this study was the same as described in a preceding study (Williamson-Benavides et al., 2020). Briefly, four genotypes with partial resistance to *Fsp*— 00-5001, 00-5003, 00-5004, and 00-5007— and four susceptible genotypes— "Aragorn," "Banner," "Bolero," and "DSP"—were selected based on their disease resistance to *Fsp* (**Table 1**). These genotypes were previously classified as either partially resistant or susceptible based on phenotyping root disease severity index (RDS), plant height, shoot dry weight, and root dry weight after *Fsp* challenge (Bodah et al., 2016). The 5,000 series pea breeding lines were found to be the most resistant lines among the white-flowered pea lines. The susceptible genotypes are among the most frequently used commercial pea varieties in the United States (**Table 1**).

The 00-5001, 00-5003, 00-5004, and 00-5007 pea breeding lines were developed by Porter et al. (2014) *via* single-seed descent at USDA–ARS, Prosser, WA. The parentage of 00-5001 is PH14-119/M7477// Coquette/3/86-2197/74-410-2 (Kraft, 1989; USDA–ARS NGRP, 2020). The parentage of 00-5003 is 69PH42-691004/Recette//Popet/3/PH14-119/ DL-1/3/B563-429-2/PI257593//DSP TAC (USDA–ARS NGRP, 2020). The parentage of 00-5004 is 79-2022/ICI 1203-1//Menlo/3/PI189171/ DL-2//75-786 (Kraft and Tuck, 1986; USDA–ARS NGRP, 2020). The parentage of 00-5007 is 00-5005/ 00-5006. 00-5005 parentage is B669-87-0/M7477//Blixt B5119/3/ 00-5001/74SN5/3/PH14-119/DL-1//74SN3/Recette/5/ FR-725 (Kraft and Giles, 1976; USDA–ARS NGRP, 2020). The parentage of 00-5006 is 00-5003/00-5004.

Two F_7-derived recombinant inbred line (RIL) populations with 190 individuals each derived from the crosses "Aragorn" × 00-5001 (Population I) and "Banner" × 00-5007 (Population II) were developed by single-seed descent and maintained at ProGene LLC Plant Research, Othello, WA, United States.

Disease Challenge and Greenhouse Evaluations of Disease

For the two populations, a total of 190 individuals with four replicates each were challenged with three *Fsp* isolates: Fs 02, Fs 07, and Fs 09. These isolates were obtained from infected pea roots collected in the Palouse Region of Washington and

TABLE 1 | Selected green-seeded pea genotypes for SNP genotyping [Adapted from Williamson-Benavides et al. (2020)].

Genotype	Source[a]	*Fsp* resistance level[b]	Other disease resistance[c]	100 seed weight (g)	Leaf type[d]	Market class
00-5001	USDA-ARS VFCRU	*	Fop races 1, 2, and 5	22.7	af	Green fresh
00-5003	USDA-ARS VFCRU	*	Fop races 1, 2, and 5	15.9	af	Green fresh
00-5004	USDA-ARS VFCRU	*	Fop races 1, 2, and 5	20.8	af	Green fresh
00-5007	USDA-ARS VFCRU	*	Fop races 1, 2, and 5	22.2	N	Green fresh
"Aragorn"	ProGene	***	Fop races 1, 2; PSBMV	19.5	af	Green dry
"Banner"	ProGene	***	Fop race 2, PM	18.7	af	Green dry
"Bolero"	AsGrow	****	Fop race 1, PM, Pythium, EMV	20.1	N	Green fresh
"DSP"	Canner Seed	***	-	20.9	N	Green fresh

[a]*AsGrow, AsGrow Seed Co., San Juan Bautista, CA, United States; Canner Seed, Canner Seed Co., Idaho Falls, ID, United States; ProGene, ProGene LLC Plant Research, Othello, WA, United States; USDA-ARS VFCRU, USDA-ARS, Vegetable and Forage Crops Research Unit, Prosser, WA, United States.*
[b]*Fsp resistance, Fsp resistance resulted from phenotyping, from most resistance (*) to susceptible (***) to most susceptible (****) (Bodah et al., 2016).*
[c] *Fop, Fusarium oxysporum; PSBMV, Pea Seed-Borne Mosaic Virus, PM, Powdery Mildew; EMV, Enation Mosaic Virus.*
[d]*af (Afila), semi-leafless; N (Normal), normal leaf type. The table shows seed source, Fsp tolerance level, other disease resistance, 100 seed weight, leaf type, and market class per pea genotype.*

Idaho by Dr. Lyndon Porter, USDA-ARS Vegetable and Forage Crops Research Unit, Prosser, WA (United States). The three isolates were single-spored and were identified based on the partial translation elongation factor 1-alpha sequences (Geiser et al., 2004). The pathogenicity of each *Fsp* isolate to pea was also confirmed (Bodah et al., 2016). The three isolates were grown on pentachloronitrobenzene (PCNB) selective media for six days (Nash and Snyder, 1962). Cultures were transferred to KERR's media (Kerr, 1963) and incubated on a shaker at 120 rpm under continuous light for six days at 23 to 25°C. The spore concentration of each isolate was determined with a hemocytometer and diluted to 1×10^6 spores/ml of water. A spore suspension inoculum was created with equal parts by volume from each of the three isolates.

RIL seeds were surface sterilized in a 0.6% sodium hypochlorite solution and rinsed in sterile distilled H_2O. The seeds were then soaked for 16 h in the *Fsp* spore suspension as described previously (Bodah et al., 2016). After the challenge with the spore suspension, seeds were planted in a completely randomized design in plastic planter cones (Conetainer, 0.25 L volume, Stuewe and Sons Inc.) filled with a standard perlite medium in a greenhouse at Crites Seed Inc. (Moscow, ID). Plants were irrigated as needed, generally every 24–36 h, and the perlite was watered to saturation at 100% field capacity. A 12-h photoperiod was maintained using 400-watt metal halide lamps for supplemental light. Plants were grown at temperatures ranging between 21–27°C during the day and 12–18°C at night.

Quantitative evaluation of RDS and plant height were recorded 21 days after planting. RDS was evaluated on a visual scale from 0 to 6, in which 0 = no diseases symptoms; 1 = small hypocotyl lesions; 2 = lesions coalescing around epicotyls and hypocotyls; 3 = lesions starting to spread into the root system with some root tips infected; 4 = epicotyl, hypocotyl and root system almost completely infected and limited white, uninfected tissue visible; 5 = completely infected root; and 6 = plant failed to emerge (Bodah et al., 2016). Plant height is a reliable indication of resistance to *Fsp* and height showed the highest negative correlation among all growth parameters related to RDS (Bodah et al., 2016). RDS and height data across the four replicates were averaged for each RIL for further analyses. Phenotyping was repeated twice. Infected root tissue from three inoculated plants was taken at random to verify the presence of *Fsp* in infected tissue. The root tissue was surface sterilized and plated onto PCNB. Culture morphology and growth were observed under a microscope and compared with the original cultures to verify the presence of *Fsp* in the infected tissue.

Broad-sense heritability was estimated with this equation Va/(Va + Ve), where Va represented the genetic variance, Ve the environmental variance.

DNA Extraction

Leaf tissue was freeze-dried in a lyophilizer. Leaf tissue samples included the eight white-flowered parental genotypes —00-5001, 00-5003, 00-5004, and 00-5007 "Aragorn," "Banner," "Bolero," and "DSP," as well as the 380 RILs from Population I and Population II. DNA was extracted with the BioSprint 96 DNA Plant kit (Qiagen, Mainz, Germany). A Nanodrop ND-8000 Spectrophotometer (ThermoFisher, MA, United States) was used to quantify the extracted DNA.

SNP Mining in DEGs, SNP Validation, and RIL Genotyping

A time-course RNAseq analysis, performed on sets of partially resistant and susceptible genotypes after *Fsp* challenge resulted in the identification of 42,905 differentially expressed contigs (DECs) (Williamson-Benavides et al., 2020). SeqMan Pro (DNASTAR, WI, United States) and custom scripts were utilized to identify single nucleotide polymorphisms (SNPs) within the set of 42,905 DECs. The Assay Design Suite software (Agena Bioscience, CA, United States) and the SNP report generated by SeqMan Pro were used to generate two sets of primers for amplifying SNP containing regions (**Supplementary Table 1**). The high-throughput MassARRAY Technology was used to validate the SNPs. Genotype calling was done from the samples deposited on the chips with the MassARRAY RT v 3.0.0.4 software (Agena Bioscience, CA, United States). Results were analyzed with the MassARRAY Typer v 3.4 software (Agena Bioscience, CA, United States). SNPs were validated across eight

pea genotypes "Aragorn," "Banner," "DSP," "Bolero," 00-5001, 00-5003, 00-5004, 00-5007, which included the four parents of the two segregating populations. Each SNP was screened twice for each individual. SNPs confirmed to be polymorphic between "Aragorn" × 00-5001, and "Banner" × 00-5007 were used -for genotyping of 190 RILs each from Population I and Population II.

Physical Map and QTL Detection

The physical location of the SNPs used in this study was determined using the pea genome (Kreplak et al., 2019). The SNP marker sequence was aligned *via* BLAST against the complete *P. sativum* genome in URGI BLAST[1]. RDS and height averages for each RIL were used to map the QTLs associated with resistance to *Fsp*. QTLs were detected with the composite interval mapping (CIM) function of the R statistical software version 3.0.2 (R core team, Vienna, Austria). CIM default settings were used. The Kosambi map function was applied to impute missing marker genotype data. QTLs were considered significant above the threshold LOD score 3.0. QTLs were named with the prefix *WB.Fsp-Ps* followed by the chromosome number and the QTL number within the chromosome.

Functional Annotation of QTLs Associated With Fsp Resistance in Pea

The QTLs *WB.Fsp-Ps* 5.1, *WB.Fsp-Ps* 5.2, *WB.Fsp-Ps* 5.3, *WB.Fsp-Ps* 2.1, and *WB.Fsp-Ps* 3.1 were annotated using the functional annotation and gene ontology (GO) data generated in a preceding study (Williamson-Benavides et al., 2020). QTL annotation provided the identity of genes and DECs located within the selected genomic regions. The confidence intervals for each of the QTLs were taken into account to identify the genomic sequence of each QTL from the pea genome (Kreplak et al., 2019). The transcriptome data were aligned *via* BLAST against the QTL sequence regions in CLC Bio Genomics Workbench 6.0.1 (CLC Bio, Aarhus, Denmark).

RESULTS

Disease Challenge and Greenhouse Phenotypic Evaluation

The quantitative evaluation of RDS and plant height was averaged per RIL across the four replicates as there was no significant difference between the replicates ($p < 0.05$). Frequency histograms for both traits per population are presented in **Figure 1**. The phenotypic means for the parents for Population I were Aragorn-RDS = 4.5; Aragorn-Height = 7.0; 00-5001-RDS = 2.5; and 00-5001-Height = 10.5. The phenotypic means for the parents for Population II were Banner-RDS = 3.3; Banner-Height = 13.0; 00-5007-RDS = 2.3; and 00-5007-Height = 10.0. The two RIL populations displayed transgressive segregation for both increased susceptibility and resistance over the two parental lines as measured by RDS and height traits (**Figure 1**).

Based on the Shapiro–Wilk test, in Population I, data were not normally distributed for RDS [$W(189) = 0.95$, $p < 0.01$] or for height [$W(189) = 0.91$, $p < 0.01$]. Similarly in Population II, data were not normally distributed for RDS [$W(189) = 0.90$, $p < 0.01$] or for height [$W(189) = 0.97$, $p < 0.01$]. A significant negative correlation was found between the RDS and height values for Population I [$r(188) = -2.76$, $p < 0.01$] and Population II [$r(188) = -4.38$, $p < 0.01$]. For Population I, broad sense heritability was 49.8 and 70.5% for RDS and height, respectively. For Population II, broad sense heritability was 43.1 and 83.4% for RDS and height, respectively.

[1]https://urgi.versailles.inra.fr/blast/

SNP Screening, SNP Validation, and RIL Genotyping

SeqMan Pro (DNASTAR, WI, United States) identified a total of 769 SNPs across DECs (**Figure 2A** and **Supplementary Table 2**). The predicted SNPs were validated in the "Aragorn," "Banner," "DSP," "Bolero," 00-5001, 00-5003, 00-5004, and 00-5007 pea genotypes (**Supplementary Table 3**). A total of 118 SNPs were confirmed for cultivars DSP and Bolero while 256 SNPs were confirmed for 5007 and Banner (**Table 2**). SNPs confirmed to be polymorphic between "Aragorn" × 00-5001 (219 SNPs) (**Figure 2B**) and "Banner" × 00-5007 (256 SNPs) (**Figure 2C**) were used to screen 190 individuals each of RIL Populations I and II, respectively. The screening results of 190 individuals each for both Population I and II are summarized in **Supplementary Tables 4**, **5**, respectively.

Physical Map and QTL Detection

The physical genomic location of all the SNPs used in this study was determined using the pea genome (Kreplak et al., 2019; **Figure 2** and **Supplementary Table 6**). Chromosome 1, 2, 3, 4, 5, 6, and 7 registered a total of 92, 86, 84, 122, 118, 78, 139 SNPs, respectively (**Figure 2** and **Supplementary Table 6**). A total of 47 SNP markers were identified in 42 scaffolds that had not been assigned to any of the seven chromosomes of pea (**Supplementary Table 6**). Three of the predicted SNP markers were not localized on the pea genome.

Prior to QTL mapping for Population I and II, a quality assessment of the genotypic data was performed. Individuals and markers with more than 80% of missing data were omitted in each database. Markers with distorted segregation patterns were also removed from the data for QTL analysis. A total of 190 RILs and 100 markers were used for QTL analysis of Population I (**Figure 2B** and **Supplementary Table 7**). A total of 182 individuals and 154 markers were used for QTL analysis of Population II (**Figure 2C** and **Supplementary Table 8**). Means per RIL for RDS and height were used to map the QTLs associated with resistance to *Fsp*. Five different QTLs were identified in the two RIL populations for RDS and height (**Figure 2D** and **Table 3**). These QTLs explained 5.3 to 14.8% of the phenotypic variance (**Table 3**).

Functional Annotation of QTLs Associated With Fsp Resistance in Pea

The transcriptome data, generated previously (Williamson-Benavides et al., 2020) were aligned *via* BLAST with the 5 QTLs: *WB-Fsp-Ps* 5.1 (**Supplementary Table 9**), *WB-Fsp-Ps*

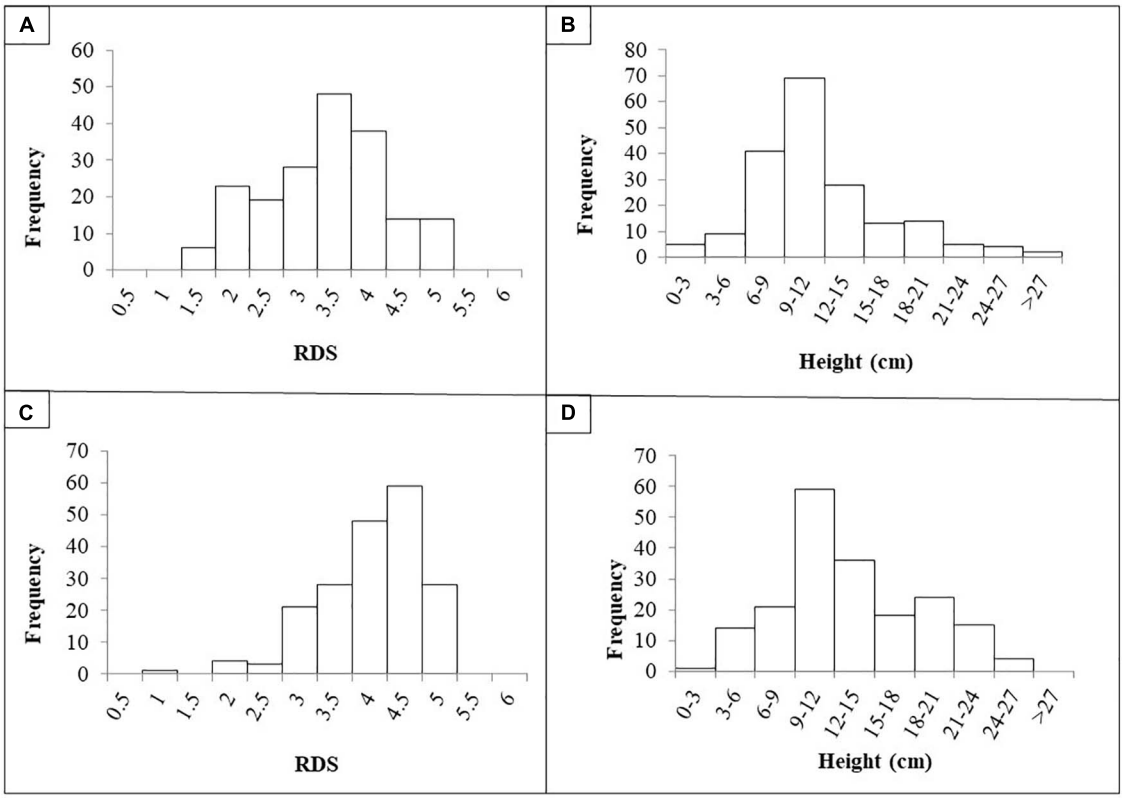


FIGURE 1 | Frequency histograms of root disease severity (RDS) and plant height of recombinant lines (RILs) after challenge with *Fusarium solani* f. sp. *pisi.* RILs were derived from crosses "Aragorn" × 00-5001 **(A,B)** and "Banner" × 00-5007 **(C,D)**.

TABLE 2 | Number of SNPs across eight *Pisum sativum* genotypes from a total of 769 predicted SNPs.

Genotype	Aragorn	Banner	Bolero	DSP	5001	5003	5004	5007
Aragorn	-	162	234	229	219	225	214	227
Banner	162	-	224	239	252	219	231	256
Bolero	234	224	-	118	191	220	191	213
DSP	229	239	118	-	206	223	177	218
5001	219	252	191	206	-	157	166	149
5003	225	219	220	223	157	-	200	181
5004	214	231	191	177	166	200	-	208
5007	227	256	213	218	149	181	208	-

5.2 (**Supplementary Table 10**), *WB-Fsp-Ps* 5.3 (**Supplementary Table 11**), *WB-Fsp-Ps* 2.1 (**Supplementary Table 12**), and *WB-Fsp-Ps* 3.1 (**Supplementary Table 13**). From the total set of aligned genes, 119–133 genes per QTL had previously been classified as differentially expressed (**Supplementary Tables 9–13**). A total of 3–17 DEGs and 6–11 non-DEGs were also identified, predicted as having unknown function or annotated as hypothetical proteins in QTLs *WB-Fsp-Ps* 5.1, 5.2, 5.3, 2.1, and 3.1 (**Supplementary Tables 9–13**).

A total of 7 DEGs associated with disease response were found in *WB-Fsp-Ps* 5.1. These genes are involved in the synthesis of lipids (acetyl-CoA carboxylase)**;** cell signaling (C-type lectin receptor-like tyrosine-protein kinase and MAPK)**;** nodulation (nodulation-signaling pathway 2 protein), and protein degradation (F-box/kelch-repeat). Another ten genes in *WB-Fsp-Ps* 5.1 were associated with disease resistance; however, these genes did not exhibit differential expression (Williamson-Benavides et al., 2020). These set of genes is associated with synthesis of lipids (1 gene)**;** auxin signaling (2), ethylene synthesis (1), pectin synthesis (1), and regulation of transcription (5).

Seventeen genes found in *WB-Fsp-Ps* 5.2 were associated with disease response and also showed differential expression after *Fsp* challenge. This list included three PR (pathogenesis-related) genes (universal stress protein PHOS32-like, endochitinase PR4, protein enhanced disease resistance 2); an anthocyanin 5-aromatic acyltransferase; four receptor-like kinases; and seven

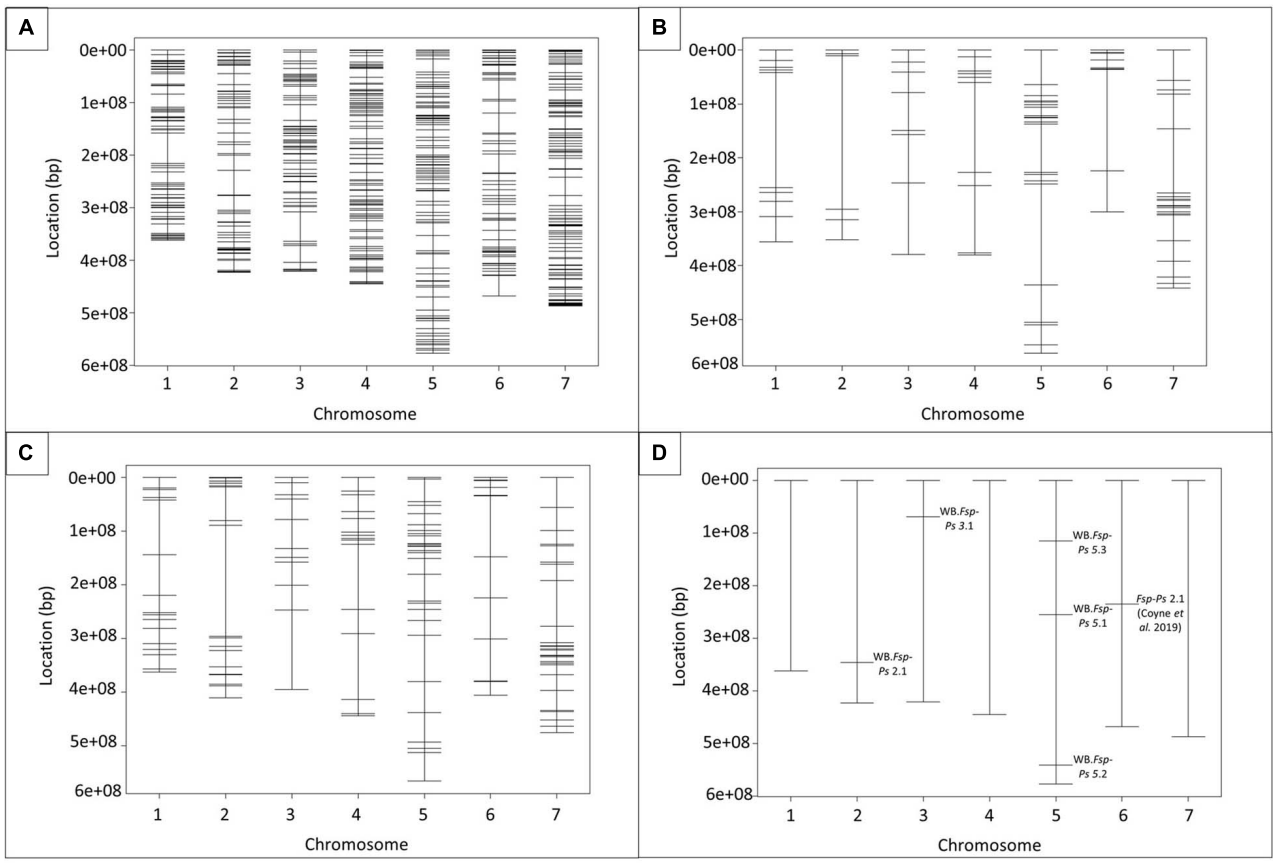


FIGURE 2 | Physical maps of SNPs and QTLs identified in the seven *Pisum sativum* chromosomes. Location of 769 SNPs, mined from *Fsp*-responsive differentially expressed genes (DEGs), in pea genome **(A)**. Total of 100 and 154 SNPs with validated polymorphisms for population I ("Aragorn" × 00-5001) **(B)** and population II ("Banner" × 00-5007) **(C)**, respectively. Five QTLs were identified in association with disease resistance **(D)**. Major QTL *Fsp-Ps* 2.1 (Coyne et al., 2019) was added as a reference **(D)**. Physical distances, represented in base pairs (bp), are shown on the left side of the graphs.

TFs of the GATA, NLP8, C2H2, and scarecrow types. Another set of seventeen contigs were identified as candidate genes but did not show any differential expression. The genes on the latter list are associated with synthesis of lipids (ketoacyl-CoA synthase); three transcription factors (GLABRA and PosF21 type); an endochitinase PR4**;** four receptor like-kinases; and two universal stress proteins PHOS32.

Twenty-two DEGs in *WB-Fsp-Ps* 5.3 were associated with disease response mechanism. This list included drug transporters (ABC transporters); a cluster of seven F-box proteins; genes involved in cell wall biosynthesis and modification (pectinesterase/pectinesterase inhibitor and polygalacturonase); the TFIIS TF; and two PR proteins—protein enhanced disease resistance 4-like and pathogenic type III effector avirulence factor. Three more ABC transporters were found in the *WB-Fsp-Ps* 5.3, but they did not show any differential expression. Other candidate genes found in *WB-Fsp-Ps* 5.3 that did not show differential expression included an autophagy-related protein; a brassinosteroid receptor; another F-box gene; five more receptor kinases; a protein enhanced disease resistance 4-like and pathogenic type III effector avirulence factor; two TFs [CCHC(Zn) family and ERF110]; and a UDP-glucuronate:xylan alpha-glucuronosyltransferase 1.

Fourteen DEGs involved in disease resistance mechanism were associated with *WB-Fsp-Ps* 2.1. These genes are known to participate in cell membrane synthesis and modification (CSC1 protein and sphingolipid transporter)**;** PR gene response (disease resistance protein RPM1 and disease resistance protein RGA3); regulation of transcription (Myb/SANT and ninja-family protein AFP3); and cell signaling (receptor-like protein kinase 2). Three genes involved in cell wall and membrane synthesis/modification (glycerol-3-phosphate acyltransferase, sphingolipid transporter, and 3UDP-arabinopyranose mutase); and nine TFs (Myb/SANT, ninja-family protein AFP3, PLATZ transcription factor family protein) were also identified as potential candidates that contribute to the effect of *WB-Fsp-Ps* 2.1. However, this set of genes did not show differential expression after *Fsp* challenge.

Eleven DEGs associated with disease response were found within *WB-Fsp-Ps* 3.1. These genes were annotated as ethylene response sensors; polygalacturonase inhibitors; phopholipases; receptor kinases; and NDR1/HIN1-like protein 10. A set of fourteen genes were characterized as associated with disease

TABLE 3 | Quantitative trait loci detected for resistance to *Fusarium solani* f. sp. *pisi* root rot in two RIL populations using root disease severity (RDS) and height.

Population	Scoring phenotype	QTL name	Chromosome	Position	Marker	LOD peak	LOD-1.5 support interval (cM)	LOD-1.5 support interval (nt)	R^2
Population I	RDS	*WB.Fsp-Ps* 5.1	5	42.9	1_14714	3.7	40.8–44. 0	236,573,488–254,762,537	8.3
	Height	*WB.Fsp-Ps* 5.2	5	95.4	7401	4.4	89.1–98.2	515,954,961–569,073,935	7.5
Population II	RDS	*WB.Fsp-Ps* 3.1	3	9.9	26703	3.4	6.0–17.9	34,871,998–103,747,090	8.1
		WB.Fsp-Ps 5.3	5	19.2	3_31807	3.1	17.6–22.6	101,835,502–131,088,590	5.3
	Height	*WB.Fsp-Ps* 5.1	5	42.9	1_14714	6.4	40.8–51.2	236,515,561–296,527,837	14.8
		WB.Fsp-Ps 2.1	2	58.2	20326	3.2	57.0–63.4	330,067,516–367,256,590	5.6

The RIL populations were derived from the following crosses: "Aragorn" × 00-5001 (Population I) and "Banner" × 00-5007 (Population II).

response, however they did not show differential expression. This list contains cathepsin B-like protease 2; ethylene-insensitive protein 2; F-box protein PP2-A15 isoform X2; mannan synthase 1-like isoform X1; polygalacturonase inhibitor; nuclear transcription factor Y subunit B-10; and serine/threonine protein receptor genes.

DISCUSSION

Here, we have reported the identification of five QTLs that are associated with *Fsp* resistance in pea (**Figure 2D** and **Table 3**). Each of these QTLs explains 5.3–14.8% of the total phenotypic variation, and together they add up to 33.21% of the variation. These five QTLs were identified by using polymorphisms embedded in *Fsp*-responsive DEGs. These polymorphisms and DEGs were originally identified *via* RNAseq. The identification of DEGs that respond to or are associated with specific biotic or abiotic stimulus, as well as the development of markers embedded in these DEGs is an efficient alternative to genotyping by sequencing for fine mapping.

To the best of our knowledge, this is the only study to report the presence of QTLs associated with *Fsp* resistance on chromosomes 2 and 3 of pea (**Figure 2D** and **Table 3**). A total of three QTLs were identified on chromosome 5. Previously, three QTLs associated with *Fsp* resistance, *Fsp-Ps 3*.1, 3.2 and *3*.3, had been reported on chromosome 5 (Coyne et al., 2015, 2019). QTLs *Fsp-Ps 3*.2 and *Fsp-Ps 3*.3 are located close to each other and adjacent to the newly identified *WB-Fsp-Ps* 5.3. Further fine mapping should be able to determine if these three QTLs are in fact three, two, or only one QTL. The situation is similar in the case of QTLs *WB-Fsp-Ps* 5.1 (**Figure 2D** and **Table 3**) and *Fsp-Ps 3*.1 (Coyne et al., 2015). LOD intervals from these two QTLs do not overlap either, although they are located close to one another on chromosome 5. This proximity can mean that they represent one QTL.

Interestingly, this study did not find any QTLs on chromosome 6. *Fsp-Ps* 2.1 (Coyne et al., 2019; **Figure 2B**) and a QTL identified by Feng et al. (2011) located on chromosome 6 explained 44.4–53.4 and 39% of the phenotypic variance, respectively. These two remain the major QTLs identified so far for *Fsp* resistance in pea. The absence of a major QTL on chromosome 6 in this study could be due to the diversity of the parental source of resistance used in this study (005001 and 00-5007), versus what was used in previous studies (PI 180693, PI 557501, "Carman") (Feng et al., 2011; Coyne et al., 2015, 2019). The large effect shown by *Fsp-Ps* 2.1, seen in previous reports, may explain the bimodal distribution for traits such as the root severity index, plant height, and plant weight (Coyne et al., 2019). However, data presented in this study did not show a bimodal distribution but showed a trend toward normal distribution, which might explain the absence of QTLs with large effects.

The establishment of associations between disease-related genes and resistance, or susceptibility can facilitate the understanding of the possible mechanism(s) involved in the pathogenicity of *Fsp* in pea. *WB-Fsp-Ps* 5.1 was the major QTL identified in this study. Among the DEGs identified in this QTL, an F-box/kelch-repeat protein (DN2516_c0_g1_i1) demonstrated reduced expression at 12 h (FC = −24.4) after *Fsp* challenge in the partially resistant, but not in the susceptible genotypes. Nine F-box protein–coding genes have been found in the region of a highly dominant QTL that provides resistance to *A. euteiches*, a root rot pathogen in pea (Djébali et al., 2009; Pilet-Nayel et al., 2009). F-box proteins are known to be involved in hormone regulation and in plant immunity (Guo and Ecker, 2003; Lechner et al., 2006). A nodulation signaling

gene found in *WB-Fsp-Ps* 5.1 was also found to be upregulated at 0 h (FC = 2.4) in susceptible genotypes when the expression levels were compared against the expression levels in partially resistant genotypes. It has been reported that a central regulator of symbiotic nodule development is determinant of susceptibility toward *A. euteiches* in *Medicago truncatula* (Rey et al., 2013).

Several genes associated with disease resistance were located in *WB-Fsp-Ps* 5.2 QTL region (**Supplementary Table 10**). Contigs DN19556_c0_g1_i1, DN2007_c0_g1_i4, DN77_c0_g1_i6 were identified as anthocyanin 5-aromatic acyltransferase, endochitinase PR4, and protein enhanced disease resistance 2. These genes showed differential expression; their expression was significantly higher in the susceptible genotypes compared to the partially resistant genotypes (Williamson-Benavides et al., 2020). The same pattern was observed for membrane receptors in all five QTLs; TFs present in *WB-Fsp-Ps 5.2*; NDR1/HIN1-like protein 10 and polygalacturonase inhibitors found in *WB-Fsp-Ps* 3.1; the RPM1-like and putative disease resistance protein RGA3 identified in *WB-Fsp-Ps* 2.1; as well as for a cluster of F-box proteins, a pathogenic type III effector avirulence factor, a pectinesterase/pectinesterase inhibitor, and a protein enhanced disease resistance 4-like gene found in *WB-Fsp-Ps* 5.3 (**Supplementary Tables 9, 10,12**). The high expression of any of these genes might be associated with disease susceptibility. However, further reverse genetics analyses will need to be performed to determine the dominant or recessive nature of these genes and QTL(s).

Cell death in the pea-*Fsp* interaction can help in the progression of *Fsp* infection due to the necrotrophic nature of the *Fsp* pathogen (Williamson-Benavides et al., 2020). The contig DN352_c0_g1_i17 located within *WB-Fsp-Ps* 5.3 was identified as CPR-5 protein which is known to negatively regulate the senescence and chlorotic lesions induced by pathogens when controlling programmed cell death (Bowling et al., 1997; Yoshida et al., 2002). This gene is highly suppressed in expression at 12 h (FC = −3.03) after *Fsp* challenge in the susceptible genotype, which might trigger cell death.

Gene DN813_c0_g1_i4, located in *WB-Fsp-Ps* 2.1, was highly overexpressed in the partially resistant genotype when compared against the expression values in the susceptible genotypes under controlled conditions at 6 (FC = −4.76) and 12 h (FC = −2.71) and under *Fsp*-inoculation at 0 (FC = −3.76) and 12 h (FC = −2.29). BLAST search of contig TRINITY_DN813_c0_g1_i4 identified it as *Medicago truncatula* 1-aminocyclopropane-1-carboxylate oxidase homolog 1 (XM_003627980) (e-value: 5e-165, percentage identity: 84.5%). The 1-aminocyclopropane-1-carboxylate oxidase enzyme is involved in the production of ethylene. Jasmonate-induced defense responses, the expected response to counter the presence of necrotrophic pathogens such as *Fsp*, are known to be associated with elevation of 1-aminocyclopropane-1-carboxylate oxidase and also to increase the activity of defense-related enzymes and subsequent control of disease incidence (Yu et al., 2011; Dixit et al., 2016). Another R gene located in *WB-Fsp-Ps* 2.1 is the putative disease resistance protein RGA3; however, differential expression was not observed for this gene (Williamson-Benavides et al., 2020).

We identified transgressive segregation in the two populations in study. Several of the transgressive lines showed enhanced resistance. For instance, a total of 29 and 5 RILs are more resistant than 00-5001 and 00-5007, respectively, based on their RDS scores. Genotypes 00-5001 and 00-5007 were previously characterized as high yielding varieties with important agronomics such as higher resistance to *Fsp*, resistance to Fusarium wilt, semi-leafless leaf type, and anti-lodging characteristics. Therefore, the transgressive lines with enhanced resistance might serve as potential candidate cultivars with good agronomics and even higher resistance to *Fsp*. Yields of these transgressive lines can be compared to elite cultivars under controlled and root rot conditions.

CONCLUSION

The use of polymorphic DEGs for QTL mapping resulted in the identification of a new major QTL *WB.Fsp-Ps* 5.1 as well as other four minor QTLs. This outcome indicates that a combined gene expression and genetics approach is effective in identifying genomic regions that may otherwise remain undetected especially for quantitative traits. Chromosome 5 is a source of several QTLs associated with *Fsp* resistance. Some of these QTLs on chromosome 5 are closely located to each other, which is a sign of resistance islands. Clustered resistance genes within the same genetic locus (resistance islands) can be transferred *en bloc* to new pea varieties through breeding. Selection toward these newly identified QTLs, along with previously identified QTLs, should allow for rapid improvement of resistance to *Fsp* root rot in the commercial pea genotypes. Furthermore, candidate genes, nested in each QTL will be instrumental in furthering our understanding of *Fsp*-pea interactions.

DATA AVAILABILITY STATEMENT

The raw sequencing data used for RNAseq analysis and identification of polymorphic regions is publicly available from the NCBI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under the accession number PRJNA630497.

AUTHOR CONTRIBUTIONS

AD, EB, RS, and BW-B designed the study. BW-B, GN, and EB performed the experiments and generated the data. BW-B analyzed the data. LP provided the tolerant pea genotypes and *Fsp* isolates. AD supervised the study. All authors read and approved the final manuscript.

FUNDING

BW-B acknowledges graduate research assistantship support from Washington State University Graduate School. Work in the Dhingra Laboratory was supported in part by Washington State

University Agriculture Research Center Hatch Grant WNP00011 and United States Dry Pea and Lentil Commission. The authors acknowledge that this study was in part supported by funding from ProGene Plant Research. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

The authors would like to thank Kurt Braunwart, CEO, ProGene Plant Research for critical discussions and his support for the project. The authors would also like to thank Nnadozie Oraguzie for critical reading of the manuscript and their valuable suggestions. This manuscript has been submitted to bioRxiv as a preprint—doi: https://doi.org/10.1101/2020.11.13.382077 (Williamson-Benavides et al., 2021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.629267/full#supplementary-material

REFERENCES

Bodah, E. T., Porter, L. D., Chaves, B., and Dhingra, A. (2016). Evaluation of pea accessions and commercial cultivars for fusarium root rot resistance. *Euphytica* 208, 63–72. doi: 10.1007/s10681-015-1545-6

Bowling, S. A., Clarke, J. D., Liu, Y., Klessig, D. F., and Dong, X. (1997). The cpr5 mutant of *Arabidopsis* expresses both NPR1-dependent and NPR1-independent resistance. *Plant Cell* 9, 1573–1584. doi: 10.2307/3870444

Chiang, C. C., and Hadwiger, L. A. (1991). The *Fusarium solani*-induced expression of a pea gene family encoding high cysteine content proteins. *Mol. Plant Microbe Interact.* 4, 324–331. doi: 10.1094/mpmi-4-324

Conner, R. L., Hou, A., Balasubramanian, P., McLaren, D. L., Henriquez, M. A., Chang, K.-F., et al. (2014). Reaction of dry bean cultivars grown in western Canada to root rot inoculation. *Can. J. Plant Sci.* 94, 1219–1230. doi: 10.4141/cjps2013-416

Coyne, C. J., Pilet-Nayel, M., McGee, R. J., Porter, L. D., Smýkal, P., and Grünwald, N. J. (2015). Identification of QTL controlling high levels of partial resistance to *Fusarium solani* f. sp. pisi in pea. *Plant Breed.* 134, 446–453.

Coyne, C. J., Porter, L. D., Boutet, G., Ma, Y., McGee, R. J., Lesné, A., et al. (2019). Confirmation of *Fusarium* root rot resistance QTL Fsp-Ps 2.1 of pea under controlled conditions. *BMC Plant Biol.* 19:98. doi: 10.1186/s12870-019-1699-9

Dahiya, P. K., Linnemann, A. R., Van Boekel, M., Khetarpaul, N., Grewal, R. B., and Nout, M. J. R. (2015). Mung bean: technological and nutritional potential. *Crit. Rev. Food Sci. Nutr.* 55, 670–688. doi: 10.1080/10408398.2012.671202

Deulvot, C., Charrel, H., Marty, A., Jacquin, F., Donnadieu, C., Lejeune-Hénaut, I., et al. (2010). Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics* 11:468. doi: 10.1186/1471-2164-11-468

Dixit, R., Agrawal, L., Gupta, S., Kumar, M., Yadav, S., Chauhan, P. S., et al. (2016). Southern blight disease of tomato control by 1-aminocyclopropane-1-carboxylate (ACC) deaminase producing *Paenibacillus lentimorbus* B-30488. *Plant Signal. Behav.* 11:e1113363. doi: 10.1080/15592324.2015.1113363

Djébali, N., Jauneau, A., Ameline-Torregrosa, C., Chardon, F., Jaulneau, V., Mathé, C., et al. (2009). Partial resistance of *Medicago truncatula* to *Aphanomyces euteiches* is associated with protection of the root stele and is controlled by a major QTL rich in proteasome-related genes. *Mol. Plant Microbe Interact.* 22, 1043–1055. doi: 10.1094/mpmi-22-9-1043

do Carmo, C. S., Nunes, A. N., Silva, I., Maia, C., Poejo, J., Ferreira-Dias, S., et al. (2016). Formulation of pea protein for increased satiety and improved foaming properties. *RSC Adv.* 6, 6048–6057. doi: 10.1039/c5ra22452g

Feng, J., Hwang, R., Chang, K. F., Conner, R. L., Hwang, S. F., Strelkov, S. E., et al. (2011). Identification of microsatellite markers linked to quantitative trait loci controlling resistance to *Fusarium* root rot in field pea. *Can. J. plant Sci.* 91, 199–204. doi: 10.4141/cjps09176

Geiser, D. M., del Mar Jiménez-Gasco, M., Kang, S., Makalowska, I., Veeraraghavan, N., Ward, T. J., et al. (2004). FUSARIUM-ID v. 1.0: a DNA sequence database for identifying Fusarium. *Eur. J. plant Pathol.* 110, 473–479. doi: 10.1007/978-1-4020-2285-2_2

Grand View Research (2017). *Pea Protein Mark. Size Worth $313.5 Million By 2025*. Available online at: www.grandviewresearch.%0Acom/press-release/global-pea-proteins-market (accessed August 15, 2020).

Grünwald, N. J., Coffman, V. A., and Kraft, J. M. (2003). Sources of partial resistance to *Fusarium* root rot in the Pisum core collection. *Plant Dis.* 87, 1197–1200. doi: 10.1094/pdis.2003.87.10.1197

Guo, H., and Ecker, J. R. (2003). Plant responses to ethylene gas are mediated by SCFEBF1/EBF2-dependent proteolysis of EIN3 transcription factor. *Cell* 115, 667–677. doi: 10.1016/s0092-8674(03)00969-3

Hadwiger, L. A. (2008). Pea–*Fusarium solani* interactions contributions of a system toward understanding disease resistance. *Phytopathology* 98, 372–379.

Hadwiger, L. A. (2015). Anatomy of a nonhost disease resistance response of pea to *Fusarium solani*: PR gene elicitation via DNase, chitosan and chromatin alterations. *Front. Plant Sci.* 6:373. doi: 10.3389/fpls.2015.00373

Jung, Y.-S., Kim, Y.-T., Yoo, S.-J., and Kim, H.-G. (1999). Mycological characteristics of *Fusarium solani* f. sp. pisi isolated from pea, ginseng and soybean in Korea. *Plant Pathol. J.* 15, 44–47.

Kerr, A. (1963). The root rot-*Fusarium* wilt complex of peas. *Aust. J. Biol. Sci.* 16, 55–69. doi: 10.1071/bi9630055

Kraft, J. M. (1989). Registration of 86-638, 86-2197, 86-2231, and 86-2236 PEA Germplasms. *Crop Sci.* 29, 494–495. doi: 10.2135/cropsci1989.0011183x002900020068x

Kraft, J. M., and Giles, R. A. (1976). Registration of 74SN3, 74SN4, and 74SN5 PEA Germplasm1 (Reg. No. GP 15 to GP 17). *Crop Sci.* 16:126. doi: 10.2135/cropsci1976.0011183x001600010042x

Kraft, J. M., and Pfleger, F. L. (2006). *Compendium of Pea Disease and Pests*, 2nd Edn. Saint Paul, MN: APS Press.

Kraft, J. M., and Tuck, J. A. (1986). Registration of 75-786, 84-1638, and 84-1930 Pea Germplasms. *Crop Sci.* 26, 1262–1263. doi: 10.2135/cropsci1986.0011183x002600060053x

Kreplak, J., Madoui, M. A., Cápal, P., Novák, P., Labadie, K., Aubert, G., et al. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* 51, 1411–1422. doi: 10.1038/s41588-019-0480-1

Kudre, T. G., Benjakul, S., and Kishimura, H. (2013). Comparative study on chemical compositions and properties of protein isolates from mung bean, black bean and bambara groundnut. *J. Sci. Food Agric.* 93, 2429–2436. doi: 10.1002/jsfa.6052

Kumari, N., and Katoch, S. (2020). "Wilt and root rot complex of important pulse crops: their detection and integrated management," in *Management of Fungal Pathogens in Pulses*, eds B. Singh, G. Singh, K. Kumar, S. Nayak, and N. Srinivasa (Cham: Springer), 93–119. doi: 10.1007/978-3-030-35947-8_6

Lechner, E., Achard, P., Vansiri, A., Potuschak, T., and Genschik, P. (2006). F-box proteins everywhere. *Curr. Opin. Plant Biol.* 9, 631–638. doi: 10.1016/j.pbi.2006.09.003

Miller, P., McKay, K., Jones, C., Blodgett, S., Menalled, F., Riesselman, J., et al. (2005). *Growing Dry Pea in Montana*. Bozeman, MT: Montana State University Extension Service.

Mukankusi, C., Derera, J., Melis, R., Gibson, P. T., and Buruchara, R. (2011). Genetic analysis of resistance to *Fusarium* root rot in common bean. *Euphytica* 182:11. doi: 10.1007/s10681-011-0413-2

Nash, S. M., and Snyder, W. C. (1962). Quantitative estimations by plate counts of propagules of the bean root rot *Fusarium* in field soils. *Phytopathology* 52, 567–572.

Peng, W., Kong, X., Chen, Y., Zhang, C., Yang, Y., and Hua, Y. (2016). Effects of heat treatment on the emulsifying properties of pea proteins. *Food Hydrocoll.* 52, 301–310. doi: 10.1016/j.foodhyd.2015.06.025

Pilet-Nayel, M.-L., Prospéri, J.-M., Hamon, C., Lesne, A., Lecointe, R., Le Goff, I., et al. (2009). AER1, a major gene conferring resistance to *Aphanomyces euteiches* in *Medicago truncatula*. *Phytopathology* 99, 203–208. doi: 10.1094/phyto-99-2-0203

Porter, L. D., Kraft, J. M., and Grünwald, N. J. (2014). Release of pea germplasm with *Fusarium* resistance combined with desirable yield and anti-lodging traits. *J. Plant Regist.* 8, 191–194. doi: 10.3198/jpr2013.07.0041crg

Reddy, N., and Yang, Y. (2011). Potential of plant proteins for medical applications. *Trends Biotechnol.* 29, 490–498. doi: 10.1016/j.tibtech.2011.05.003

Rey, T., Nars, A., Bonhomme, M., Bottin, A., Huguet, S., Balzergue, S., et al. (2013). NFP, a L ys M protein controlling N od f actor perception, also intervenes in *Medicago truncatula* resistance to pathogens. *New Phytol.* 198, 875–886. doi: 10.1111/nph.12198

Román-Avilés, B., Lewis, J. M., Kelly, J. D., Alves-Santos, F. M., and Diez, J. J. (2011). "Fusarium genetic control: a long term strategy," in *Control Fusarium Diseases*, eds F. Alves-Santos and J. Díez (Thiruvananthapuram: Research Signpost), 65–108.

Seaman, W. L. (1976). Yield loss conversion factors for fusarium root rot of peal. *Can. Plant Dis. Surv.* 56, 25–32.

Shaykh, M., Soliday, C., and Kolattukudy, P. E. (1977). Proof for the production of cutinase by *Fusarium solani* f. pisi during penetration into its host, *Pisum sativum*. *Plant Physiol.* 60, 170–172. doi: 10.1104/pp.60.1.170

Sim, S. Y. J., Karwe, M. V., and Moraru, C. I. (2019). High pressure structuring of pea protein concentrates. *J. Food Process Eng.* 42, e13261.

Stehfest, E., Bouwman, L., Van Vuuren, D. P., Den Elzen, M. G. J., Eickhout, B., and Kabat, P. (2009). Climate benefits of changing diet. *Clim. Change* 95, 83–102. doi: 10.1007/s10584-008-9534-6

USDA–ARS NGRP (2020). *Germplasm Resources Information Network- (GRIN)*. Available online at: https://npgsweb.ars-grin.gov/gringlobal/search.aspx (accessed February 7, 2020).

Wang, W., Jacobs, J. L., Chilvers, M. I., Mukankusi, C. M., Kelly, J. D., and Cichy, K. A. (2018). QTL analysis of *Fusarium* root rot resistance in an Andean× middle American common bean RIL population. *Crop Sci.* 58, 1166–1180. doi: 10.2135/cropsci2017.10.0608

Wei, Y., Cai, Z., Wu, M., Guo, Y., Tao, R., Li, R., et al. (2020). Comparative studies on the stabilization of pea protein dispersions by using various polysaccharides. *Food Hydrocoll.* 98:105233. doi: 10.1016/j.foodhyd.2019.105233

Williamson-Benavides, B., Sharpe, R., Nelson, G., Bodah, E. T., Porter, L. D., and Dhingra, A. (2021). Identification of root rot tolerance QTLs in pea using *Fusarium solani* f. sp. pisi-responsive differentially expressed genes. *bioRxiv* [Preprint]. doi: 10.1101/2020.11.13.382077

Williamson-Benavides, B. A., Sharpe, R. M., Nelson, G., Bodah, E. T., Porter, L. D., and Dhingra, A. (2020). Identification of *Fusarium solani* f. sp. pisi (Fsp) responsive genes in *Pisum sativum*. *Front. Genet.* 11:950. doi: 10.3389/fgene.2020.00950

Xiong, T., Xiong, W., Ge, M., Xia, J., Li, B., and Chen, Y. (2018). Effect of high intensity ultrasound on structure and foaming properties of pea protein isolate. *Food Res. Int.* 109, 260–267. doi: 10.1016/j.foodres.2018.04.044

Xue, A. G. (2003). Biological control of pathogens causing root rot complex in field pea using *Clonostachys rosea* strain ACM941. *Phytopathology* 93, 329–335. doi: 10.1094/phyto.2003.93.3.329

Yoshida, S., Ito, M., Nishida, I., and Watanabe, A. (2002). Identification of a novel gene HYS1/CPR5 that has a repressive role in the induction of leaf senescence and pathogen-defence responses in *Arabidopsis thaliana*. *Plant J.* 29, 427–437. doi: 10.1046/j.0960-7412.2001.01228.x

Yu, M., Shen, L., Zhang, A., and Sheng, J. (2011). Methyl jasmonate-induced defense responses are associated with elevation of 1-aminocyclopropane-1-carboxylate oxidase in *Lycopersicon esculentum* fruit. *J. Plant Physiol.* 168, 1820–1827. doi: 10.1016/j.jplph.2011.05.021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Williamson-Benavides, Sharpe, Nelson, Bodah, Porter and Dhingra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 September 2021
doi: 10.3389/fgene.2021.680457

Molecular Characterization and Expression Analysis of the Na^+/H^+ Exchanger Gene Family in *Capsicum annuum* L.

Xirong Luo[1,2†], Shimei Yang[1†], Yong Luo[2], Huarong Qiu[2], Tangyan Li[1,3], Jing Li[1], Xiaocui Chen[2], Xue Zheng[3], Yongdui Chen[3], Jie Zhang[3], Zhongkai Zhang[3] and Cheng Qin[1,3,4]**

[1] Department of Modern Agriculture, Zunyi Vocational and Technical College, Zunyi, China, [2] Key Lab of Zunyi Crop Gene Resource and Germplasm Innovation, Zunyi Academy of Agricultural Sciences, Zunyi, China, [3] Yunnan Provincial Key Lab of Agricultural Biotechnology, Key Lab of Southwestern Crop Gene Resources and Germplasm Innovation, Biotechnology and Germplasm Resources Institute, Yunnan Academy of Agricultural Sciences, Ministry of Agriculture, Kunming, China, [4] School of Agriculture, Yunnan University, Kunming, China

The Na^+/H^+ exchangers (*NHXs*) are a class of transporters involved in ion balance during plant growth and abiotic stress. We performed systematic bioinformatic identification and expression-characteristic analysis of *CaNHX* genes in pepper to provide a theoretical basis for pepper breeding and practical production. At the whole-genome level, the members of the *CaNHX* gene family of cultivated and wild pepper were systematically identified using bioinformatics methods. Sequence alignment and phylogenetic tree construction were performed using MEGA X software, and the gene functional domain, conserved motif, and gene structure were analyzed and visualized. At the same time, the co-expression network of *CaNHX* genes was analyzed, and salt-stress analysis and fluorescence quantitative verification of the Zunla-1 cultivar under stress conditions were performed. A total of 9 *CaNHX* genes were identified, which have typical functional domains of the Na^+/H^+ exchanger gene. The physical and chemical properties of the protein showed that the protein was hydrophilic, with a size of 503–1146 amino acids. Analysis of the gene structure showed that Chr08 was the most localized chromosome, with 8–24 exons. *Cis*-acting element analysis showed that it mainly contains *cis*-acting elements such as light response, salicylic acid response, defense, and stress response. Transcriptom and co-expression network analysis showed that under stress, the co-expressed genes of *CaNHX* genes in roots and leaves were more obvious than those in the control group, including ABA, IAA, and salt. The transcriptome and co-expression were verified by qRT-PCR. In this study, the *CaNHX* genes were identified at the genome level of pepper, which provides a theoretical foundation for improving the stress resistance, production, development, and utilization of pepper in genetic breeding.

Keywords: *Capsicum annuum*, CaNHX, phylogenetic tree, co-expression, abiotic stress

OPEN ACCESS

Edited by:
Deepmala Sehgal, International Maize and Wheat Improvement Center Mexico, Mexico

Reviewed by:
Ratan Chopra, University of Minnesota Twin Cities, United States
Yufei Zhai, Northwest A&F University, China

***Correspondence:**
Zhongkai Zhang
zhongkai99@sina.com
Cheng Qin
qincheng1001@163.com
orcid.org/0000-0002-8519-4697

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *14 March 2021*
Accepted: *02 August 2021*
Published: *02 September 2021*

Citation:
Luo X, Yang S, Luo Y, Qiu H, Li T, Li J, Chen X, Zheng X, Chen Y, Zhang J, Zhang Z and Qin C (2021) Molecular Characterization and Expression Analysis of the Na^+/H^+ Exchanger Gene Family in Capsicum annuum L. Front. Genet. 12:680457. doi: 10.3389/fgene.2021.680457

INTRODUCTION

Pepper (*Capsicum annuum* L.), also known variously as capsicum, chili pepper, chile, and chili, is an annual or perennial plant belonging to the Solanaceae family. It is one of the most important vegetable crops in the world (Qin et al., 2015). *Capsicum* species were first introduced into China during the Ming Dynasty and today, China has the largest planting area and fresh yield in the world (FAO).[1] It is an important cash crop with many varieties, and is considered also of ornamental value, with considerable genetic diversity for research purposes and breeding (Zhang et al., 2016). Many varieties—including Zunla-1, Yunnan Xiaomi Spicy, and Hainan Bell Pepper—are widely planted in China, and their market share is increasing every year.

Peppers contain many substances of nutritional value including vitamin C and vitamin A. The fruits are not only used for food seasoning, but also in the production of food pigments, medicine, and industrial chemicals (Kantar et al., 2016). In medicine, it is widely used for multiple functions, including antibiosis and the prevention and treatment of disease (Saleh et al., 2018). Three pepper genomes (Zunla-1, Chiltepin, and CM334) have been completely sequenced, and with continuing re-sequencing, transcriptome sequencing, and metabolomics based on the whole genome, an increasing amount of genetic data of various pepper varieties has been mined (Kim et al., 2014; Qin et al., 2014). A key component of peppers is their capsaicin. Peppers produced in northwest China contain higher capsaicin and heme levels, due to the dry climate, low rainfall, high solar radiation, and wide temperature difference between day and night (Liu et al., 2012). This study examines how the nutritional content and capsaicin levels in peppers change when *Capsicum* is stressed by its growing environment. Under drought conditions, the capsaicin content in pepper can be reduced; under certain salt conditions, a significantly higher concentration of salt can promote the yield of capsaicin compared with control and low-salt pepper growth, and photosynthetic efficiency does not necessarily increase with salt (Sarah et al., 2012; Khan et al., 2014).

Plant growth and development depend strongly on environmental factors, such as cold and heat, drought, soil salinity and alkalinity, and other abiotic stresses. When plants are under stress, including some major cash crops, the external environment directly affects plant production. Salt stress is one of the most serious abiotic stresses affecting plant productivity and causes significant crop loss worldwide (Zhang et al., 2018). When plants are in a saline-alkaline soil environment, their ion balance and water balance change significantly. A change in membrane permeability destroys the normal operation of transporters, causing plants to absorb additional sodium ions from the environment, affecting the absorption of other ions and causing nutritional imbalance. Na^+/H^+ antiporters play a key role in plant development and tolerance to salt stress (Akram et al., 2020). In response to the external influence on plants, the ions and water in plants are balanced through their own ion channels. In general, the cytoplasmic pH value is above neutral (pH 7.2–7.6), which is controlled by an array of regulating molecules such as Na^+/K^+ transporters, cation/proton exchangers like Ca^{2+}/H^+, sodium-proton antiporters (NHX), proton/nutrient transporters, and H^+-translocating enzymes (Benèina et al., 2009). Studies indicate that NHX antiporters are involved in regulating the ion balance in plants under salt stress. Their primary physiological functions are the regulation of cytoplasmic pH and expulsion of H^+ generated during metabolism, in exchange for transporting Na^+ or K^+ ions into the cytoplasm and vacuoles of plants and animals (Pedersen et al., 2006). This indicates that studying the salt-tolerance mechanism of plants can improve the growth of plants under salt and alkali stress.

Human *HsNHE* was the first eukaryotic sodium-hydrogen exchanger gene to be identified and cloned, and functions in transport, Na^+/H^+ exchange, and pH regulation (Sardet et al., 1989). The first *NHX* gene identified in plants was the *AtNHX1* gene from *Arabidopsis thaliana*, and the expression of this gene can regulate NaCl in *A. thaliana* and is a salt-tolerance determiner (Gaxiola et al., 1999; Yokoi et al., 2002; Rodríguez-Rosales et al., 2009). Eight *NHX* genes were identified in *A. thaliana*, among which *AtNHX1* and *AtNHX2* were most common in the buds and roots of seedlings, while *AtNHX5* mRNA was expressed in lower abundance in both buds and roots. *AtNHX3* was detected in roots, while *AtNHX4* and *AtNHX6* mRNA were only detected by RT-PCR (Yokoi et al., 2002). To date, *NHX* genes of several plant species have been identified, such as *Vitis vinifera* (6 *VvNHX* genes) (Ayadi et al., 2020), *Medicago truncatula* (*MtNHX1- MtNHX8*) (Sandhu et al., 2018), *Populus trichocarpa* (*PtNHX1- PtNHX8*) (Tian et al., 2017), *Populus euphratica* (*PeNHX1- PeNHX6*) (Ye et al., 2009), *Gossypium hirsutum* (*GhNHX1- GhNHX23*) (Fu et al., 2020), *Morus alba* (*MaNHX1- MaNHX7*) (Cao et al., 2016), and *Beta vulgaris* (*BvNHX1- BvNHX5*) (Wu et al., 2019). This study performed different bioinformatic analyses of *CaNHX* genes in cultivated and wild peppers, and the *CaNHX* gene family of pepper was identified at the genome level, providing a theoretical basis for analyzing the function of the gene under salt stress.

MATERIALS AND METHODS

Material and RNA Extraction

In this study, the whole genome data of pepper (*C. annuum* L. Zunla-1 is hereinafter referred to as pepper) and *C. annuum* var. glabriusculum Chiltepin were taken as the research object.[2] Zunla-1 pepper material was planted in the greenhouse of the Department of modern agriculture, Zunyi vocational and technical college (Zunyi, Guizhou, 107°045 'E, 27°710' N). The pepper was treated with 100 Mmol NaCl for 3, 6, 12, 24, and 72 h. The root material of pepper was stored in liquid nitrogen, and 3 samples were taken at each time point as biological replicates. RNA was extracted from the collected samples using the TianGene RNA Extraction Kit (DP432, Beijing, China). We then added root material with a weight of 50–100 ng for aseptic freezing grinding; 450 μL for oscillation mixing. This was transferred to the CS filter column and centrifuged

[1] http://www.fao.org/

[2] http://peppersequence.genomics.cn

for 3 min (12,000 rpm). The supernatant was transferred from the collection tube with a pipette gun to the Rnase-free centrifuge tube. Then, the supernatant was added and 0.5 times of anhydrous ethanol was mixed into the centrifuge tube and then transferred to the adsorption column CR3 for centrifugation for 30 s (12,000 rpm). A drop of 80 μl DNase I was added to the center of the collecting tube and left at room temperature for 15 min. 250 μL of protein-removing solution RW1 was added to the adsorption column CR3, and left to stand at room temperature for 2 min, before being centrifuged for 30 s (12,000 rpm) (this procedure was repeated once). We then took an enzyme-free centrifuge tube and placed the adsorption column in a new centrifuge tube for several minutes (until the rinsing solution RW was dried). 50 μL Rnase-free ddH_2O was then vertically added to the adsorption column, and the obtained RNA was stored at –80℃ for later use.

Identification of the *CaNHX* Gene in Pepper

C. annuum cv. CM334, *C. chinense* PI159236, *C. annuum* cv. ECW, *C. annuum* SF and *C. baccatum* PBC81 genomes come from PGP (Pepper Genome Platform).[3] *Cuneo*, *Corno di Carmagnola*, *Quadrato di Carmagnola,* and *Tumaticot* genomes comes from RisEPP (Resequencing Piedmontese Pepper Ecotypes).[4] According to the characteristics of the *NHX* gene family in Pfam[5] data, there is an obvious conservative structure of the *NHX* gene family (PF00999) (El-Gebali et al., 2018). The genome-wide protein sequence of capsicum was searched by HMMER V3.3 software and verified with the Hmmer web server,[6] and the sequences with the incomplete conservative structure were removed. meanwhile, the *AtNHX* gene of *A. thaliana* was used for blast comparison, and the *E*-value was maintained at $1e^{-20}$ for comparison. Selecting the intersection of HMMER identification and BLAST alignment, 9 candidate genes of *CaNHX* were finally identified for subsequent analysis (Potter et al., 2018). The physicochemical properties of the pepper *CaNHX* protein were analyzed using the online tool ExPASy[7] (Artimo et al., 2012). Prediction of Plant-mPLoc by subcellular localization of the *CaNHX* gene in pepper was performed using online tools[8] (Chou and Shen, 2010).

Analysis of Phylogeny and Characteristics of the *CaNHX* Genes Family

MEGA X was used to perform multiple sequence alignment analysis on the obtained 9 pepper NHX protein sequences obtained, and the phylogenetic tree (neighbor-joining, bootstrap = 1,000) was constructed, and other parameters were left at default settings (Kumar et al., 2018). The online tool Itol[9] was used for the presentation and form of the pepper *NHX* phylogenetic tree (Letunic and Bork, 2019). The batch CD-search[10] tool in NCBI was used to visually analyze the *NHX* gene structure of the *NHX* gene. The online tool GSDS[11] was used for the visualization of pepper *NHX* gene structure (Hu et al., 2014). The online sequence analysis tool MEME Suite[12] was used for motif analysis, with the motif number set at 10 (Bailey et al., 2009). Collinearity analysis of pepper was performed by BLAST for whole-genome protein levels, and the MCScanX tool was used for collinearity analysis (Wang et al., 2012). TBtools were used for the visualization of gene structure, motifs, and collinearity results (Chen et al., 2020).

Ka/Ks and Promoter Analysis of the *CaNHX* Genes Family

Using BLAST to build a pepper comparison database, and the KaKs Calculator tool to calculate the synonymous substitution rate and nonsensical substitution rate of pepper *CaNHX* genes, the K_a/K_s ratio of genes was obtained, and evolutionary pressure was analyzed (Wang et al., 2010). The upstream 2,000 bp sequences of *NHX* genes were compared and extracted using the Bedtools genome analysis tool.[13] The upstream 2,000 bp sequence was predicted and analyzed using the online tool PlantCARE.[14] Visualization was performed with TBtools, the main action components were discussed (Lescot et al., 2002; Quinlan and Hall, 2010).

Expression Model and Coexpression Analysis

Transcription factors (TFs) in the *C. annuum* genome were identified using the online iTAK Plant Transcription factor and Protein Kinase Identifier and Classifier (Zheng et al., 2016). The expression data[15] obtained from pepper informatic hub were analyzed using temporal and spatial expression patterns and co-expression network associations. The root and leaf tissues of the CM334 pepper cultivar were used for transcriptome and metabolomic analysis, with a total of 574 transcriptome data points. The co-expression results were visualized using Cytoscape 3.7.2 (Liu et al., 2017; Otasek et al., 2019).

cDNA Synthesis and Quantitative RT-PCR Analysis

The First Strand of RNA was synthesized using the Revertaid First Strand cDNA Synthesis Kit (K1622) from Thermo Field (RevertAid First Strand cDNA Synthesis Kit). The fluorescent quantitative primer Actin (GenBank: DQ832719) and Ubiquitin (GenBank: AY496112) were designed by Primer3plus software as the housekeeping gene (**Supplementary Table 1**). The

[3]http://pepper.snu.ac.kr/

[4]https://www.pepper-genomics.unito.it/

[5]http://pfam.xfam.org/

[6]https://www.ebi.ac.uk/Tools/hmmer/

[7]https://web.expasy.org/protparam/

[8]http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/

[9]http://itol.embl.de/

[10]https://www.ncbi.nlm.nih.gov/cdd/

[11]http://gsds.gao-lab.org/

[12]http://meme-suite.org/

[13]https://github.com/arq5x/bedtools2

[14]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

[15]http://pepperhub.hzau.edu.cn/

fluorescence quantitative instrument for 15 samples was 96 Real-time qTOWER3.0 (Analytikjena, Germany), The fluorescence quantitative reaction system consisted of 10 μL SYBR Primix Ex Taq TM II (ZomBio, Beijing, China), and the upstream and downstream primers of each gene were 1 μL. And ddH2O to 20 μL. The PCR reaction procedure was 95°C for 30 s;95°C for 15 s;60°C for 30 s; and 72°C for 1 min, for 40 cycles. The quantitative RT-PCR results were analyzed using the $2^{-\Delta\Delta\ Ct}$ method (Livak and Schmittgen, 2001). GraphPad Prism v8 was used to visualize the fluorescence quantitative results.

RESULTS

Identification and Physicochemical Properties of *CaNHX* Gene Family

According to the characteristics of the *NHX* gene family in the Pfam database, it contains Na_H_Exchanger (PF00999) functional domain. First, a total of 42 *NHX* genes were identified in pepper by the Hmmsearch identification method. Then, 8 *AtNHX* genes of *A. thaliana* were compared with the pepper genome by the BLASTP method. Combined with the two identification methods, the incomplete genes were removed by using the Hmmer online website. Nine *NHX* genes were obtained for subsequent analysis. These 9 *CaNHX* genes sequences were used for subsequent analysis and named *CaNHX1-CaNHX9* in turn (**Supplementary Table 2** and **Table 1**). The physical and chemical properties of the protein showed that the size was 360–1181 aa, the molecular weight was 398.06–129.91 kDa, the isoelectric value was 5.44–8.79, GRAVY was less than 1, and it was a hydrophilic protein. After subcellular localization of pepper *CaNHX* gene, it was found that the subcellular localization of CaNHX2 and CaNHX4 was in the Cell membrane, and the other 7 subcellular localization were all in Vacuole.

Phylogeny Analysis of *CaNHX* Genes With Different Species

The 9 *CaNHX* genes in *C. annuum* identified were compared with 8 *AtNHXs* in *A. thaliana*, 12 *GmNHXs* in *Glycine max*, 8 *PtNHX*s in *P. trichocarpa*, 8 *VvNHXs* in *V. vinifera*, 7 *OsNHX*s in *O. sativa*, 7 *ZmNHXs* in *Zea mays* (**Figure 1** and **Supplementary Table 3**). According to the phylogenetic tree, *NHX* genes were divided into three subgroups, among which subgroup I contained the most genes. Subgroup I and Subgroup III contain four genes, respectively, while subgroup II contains only one gene.

Gene Structure and Conserved Sequence of *CaNHX* Genes

The *NHX* gene subfamily classification, gene structure, and motif analysis maps show the characteristics of *A. thaliana* and pepper gene families (**Figure 2**). Using TBtools to analyze the gene structure of 17 *NHX* genes family members in pepper and *A. thaliana*, the results showed that the exon number of the *CaNHX* gene family was mainly distributed between 8 and 24, while that of *AtNHX* gene family members in *A. thaliana* was between 12 and 23, among which the exon number of Class I subgroup was stable between 12 and 14, while the exon number of Class III was the largest. The NHX protein sequences of *C. annuum* and *A. thaliana* were analyzed by MEME web tool. According to the distribution of motif of *CaNHX* genes family members, the motif number is consistent with the phylogenetic tree. For example, in the Class I subfamily, there are 5 motif sequences, which are motif f1, motif2, motif5, motif6, and motif7. It is consistent with the phylogenetic tree classification of the Class I subgroup. The motif of the *CaNHX* gene in the Class II subfamily all contained motif8 and motif9, of which CaNHX8 was the one with the least motif number. In the Class III subfamily, the motif number of CaNHX1, CaNHX3, CaNHX6, and CaNHX7 remained at 7–8, of which motif8 and motif9 did not exist in the Class III subfamily. Motif7 is a typical amiloride-binding site (LFFIYLPPI), which is a motif contained in the genes of salt-tolerant plants and transgenic *NHX* plants, and contains the motif in CaNHX2, CaNHX3, CaNHX4, and CaNHX5 (**Figures 2**, **3**).

Chromosome Localization, Collinearity Analysis, and Ka/Ks Analysis

Through identification, 9 were identified in cultivated pepper (Zunla-1) and 6 were identified in wild pepper (Chiltepin) (**Supplementary Tables 5, 6**). According to the gene sequence, 9 *CaNHX* gene sequences were mapped to 5 chromosomes, among which one was a simulated chromosome (not assembled to chromosome), 3 *CaNHX* genes were mapped to Chr08, but 2 genes were not mapped to the chromosome, and the other chromosomes Chr01, Chr05, Chr06, and Chr10 were all

TABLE 1 | Family information and subcellular localization of *CaNHX* gene in pepper.

Gene_id	Gene_name	Chr	aa	Kd	pI	GRAVY	Sub. localization
Capana01g003109	*CaNHX1*	Chr01	511	56256.55	6.56	0.474	Vacuole
Capana05g000031	*CaNHX2*	Chr05	947	104045.39	5.44	0.242	Cell membrane.
Capana06g003039	*CaNHX3*	Chr06	527	58411.65	8.79	0.524	Vacuole
Capana08g000122	*CaNHX4*	Chr08	1181	129915.14	6.26	0.090	Cell membrane
Capana08g000123	*CaNHX5*	Chr08	624	67507.96	7.62	0.619	Vacuole
Capana08g001332	*CaNHX6*	Chr08	482	53412.57	6.45	0.590	Vacuole
Capana10g000166	*CaNHX7*	Chr10	512	57296.70	8.59	0.452	Vacuole
Capana00g000346	*CaNHX8*	Chr00	459	51409.16	6.36	0.000	Vacuole
Capana00g004012	*CaNHX9*	Chr00	360	39805.86	6.63	0.720	Vacuole

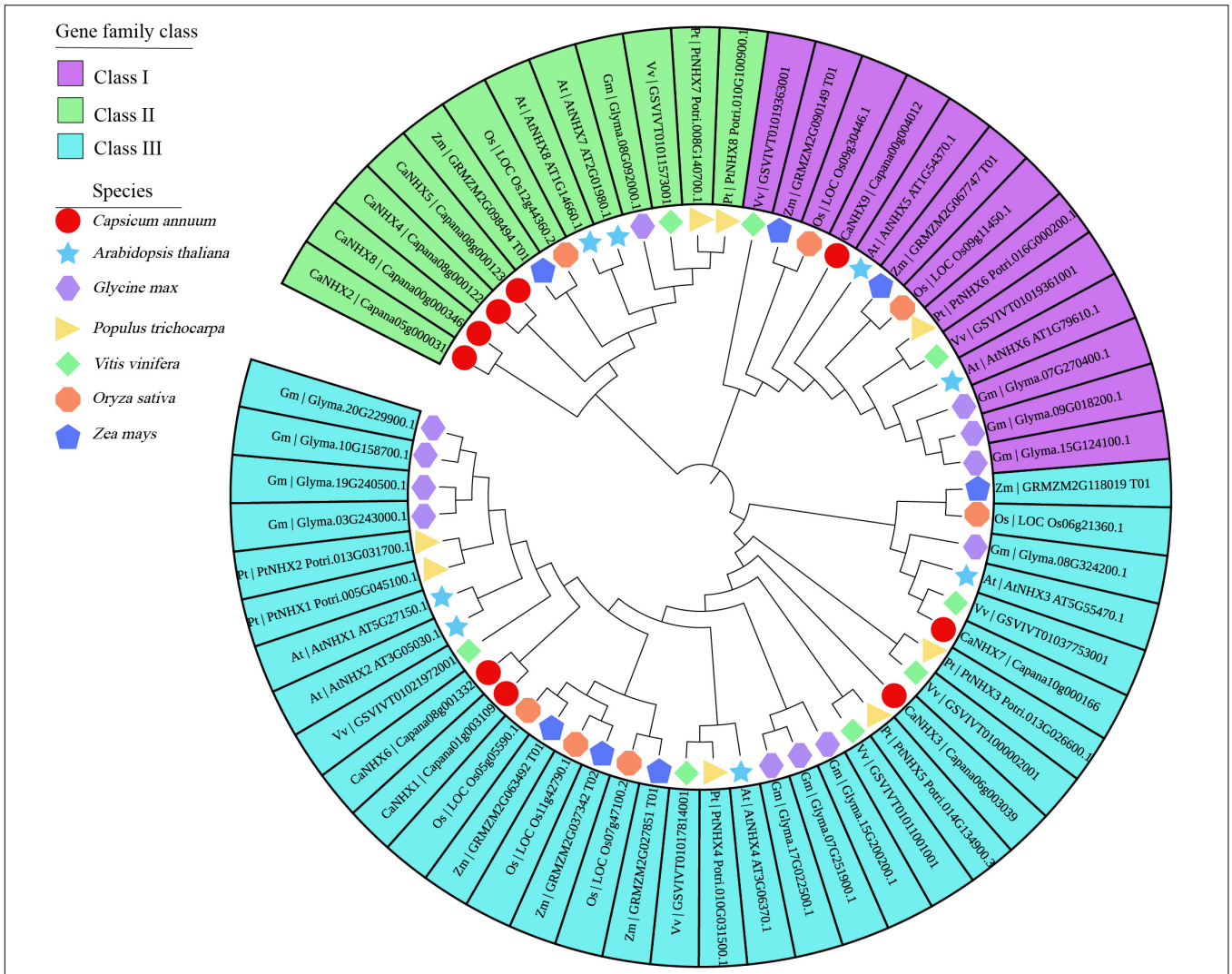


FIGURE 1 | Phylogenetic analysis of *CaNHX* genes in *C. annuum.* Square: *C. annuum;* Star: *A. thaliana;* Horizontal hexagon: *G. max;* Right triangle: *P. trichocarpa;* Rhombus: *V. vinifera;* Octagon: *O. sativa;* Up pointing pentagram: *Z. mays.*

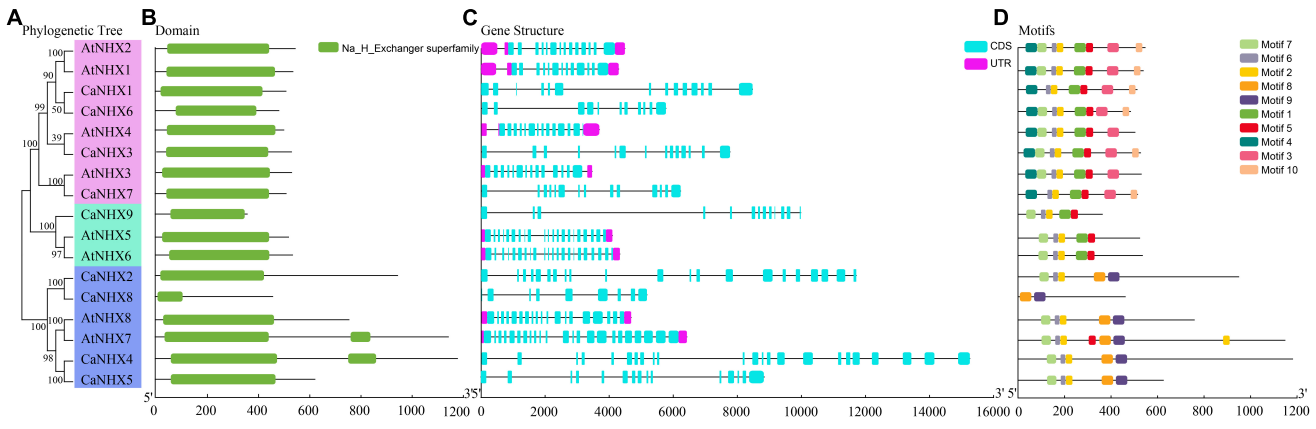


FIGURE 2 | Gene structure and conserved sequence of *C. annuum* and *A. thaliana*. **(A)** Refined relationship between *C annuum* and *A. thaliana*. **(B)** Runctional structure domain. **(B)** NHX genes. **(C)** NHX genes structure. **(D)** Conservative motifs. **(D)** NHX genes. **(E)** Motifs.



FIGURE 3 | Sequence alignment of *CaNHX* Gene in pepper.

one *CaNHX* gene (**Figure 4**). To study the whole genome duplication (WGD) event, 42 cultivars were identified (Contains the confirmed 9 *CaNHX* genes) by Hmmsearch in cultivated pepper and 37 cultivars identified by hmmsearch in wild pepper were analyzed together (**Supplementary Table 3**). There are more collinearity relationships among the 9 pepper genes identified, among which *CaNHX4* and *CaNHX6* are on Chr08 of Zunla-1, while the collinearity block of this gene is on wild type Chr01. There is a collinearity block between *CaNHX7* and wild-type Chr00 on Zunla-1 chromosome Chr10. Among them, *CaNHX1*, *CaNHX5*, and *CaNHX8* have no collinearity block relationship, while *CaNHX2*, *CaNHX3*, and *CaNHX9* also have collinearity, but their chromosomal positions do not change. The results indicated that the cultivated pepper Zunla-1 and the wild pepper were from the same ancestor, and there was a certain gene replication event. At the same time, we performed collinearity analysis on the genome of pepper varieties with Zunla-1and other pepper genomes, including Chiltepin, Corno,

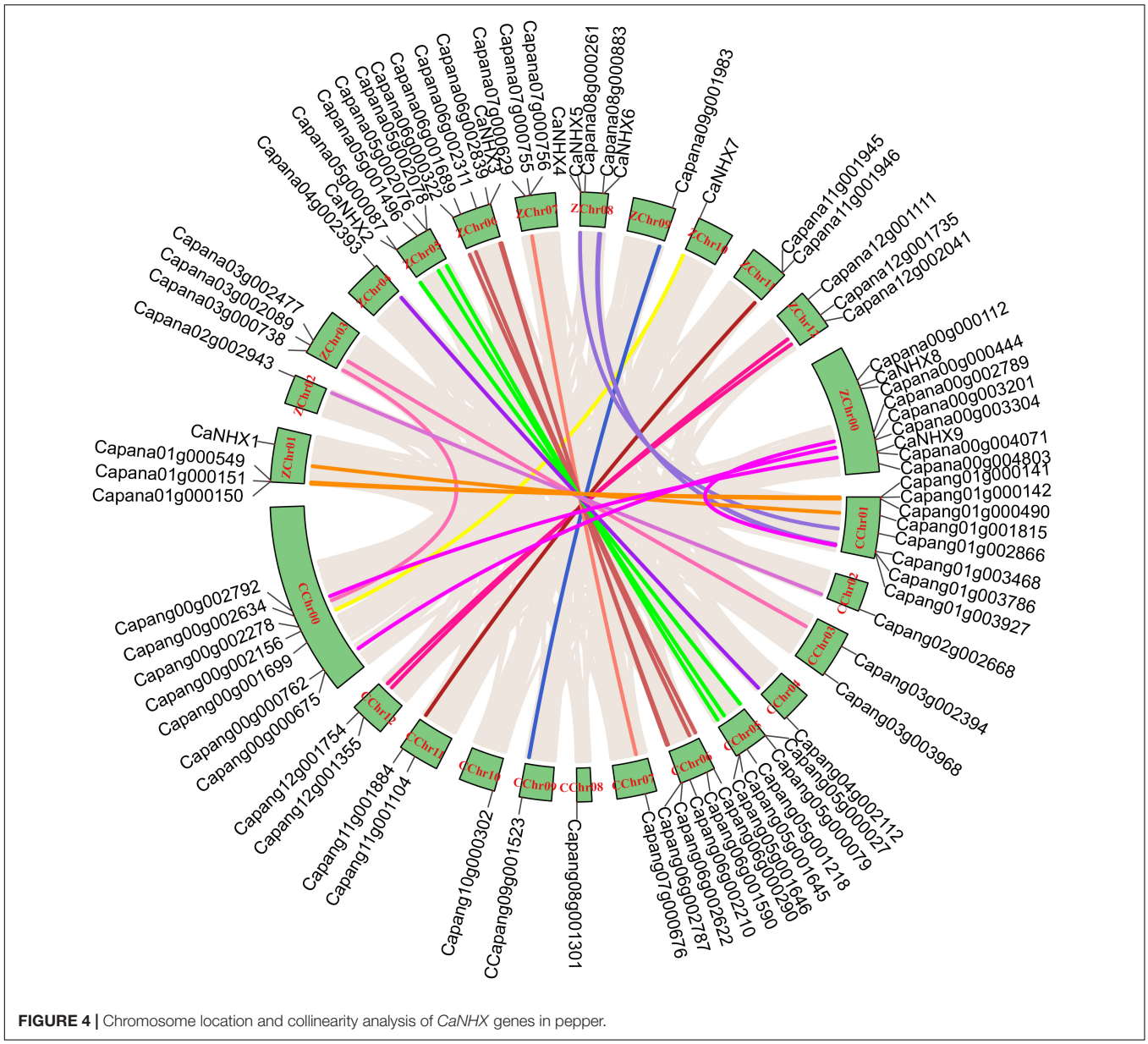


FIGURE 4 | Chromosome location and collinearity analysis of *CaNHX* genes in pepper.

Cuneo, Quadrato, tumaticot, CM334, ECW, PBC81, and SF (**Figure 5**). The collinearity analysis between Zunla-1 and Corno, Cuneo, Quadrato, and Tumaticot showed that the chromosomal position relationship of Corno, Cuneo, Quadrato, and Tumaticot was the same. The results showed that the four pepper varieties were derived from the same ancestor and had less variation during the species evolution. However, the position relationship between the four pepper varieties and Zunla-1 changed greatly, which indicated that they had a far evolutionary relationship with more variation. Zunla-1 showed significant variation with Chiltepin, CM334, ECW, PBC81, and SF, and the changes of gene position were obvious, indicating that Zunla-1 was far related to the other five varieties of pepper. Three pairs of homologous loci were obtained by analyzing the Ka/Ks ratio of the *CaNHX* gene, and their Ka/Ks were less than one, indicating that the gene was mainly purified during the evolution of the *CaNHX* gene in pepper (**Table 2**).

Promoter Analysis of *C. annuum CaNHX* Genes

By analyzing the upstream 2,000 bp sequence of the *CaNHX* genes, the *cis*-acting elements of the gene were predicted. In addition to a large number of basic elements—CAAT-box and TATA-box—there are also G-Box, GAG-motif, chs-CMA1a, TCT-motif, GATA-motif, GT1-motif, and AE-box in the *CaNHX* gene family, and TCA-elements in the salicylic acid reaction. Also present were *cis*-acting elements, TC-rich repeats, meristem expression elements, CAT-box, MYB binding site elements, MBS, MeJA response elements, GCTCA-motif, auxin response elements, TGA-elements, etc. (**Figure 6** and

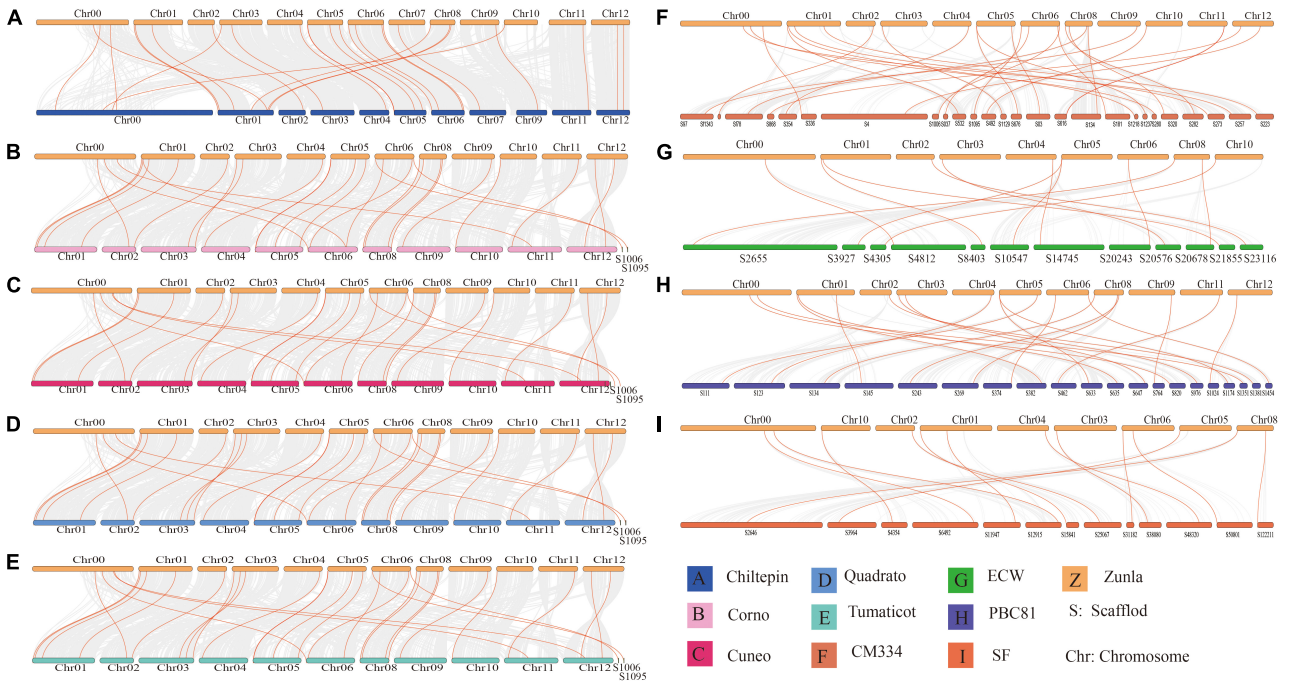


FIGURE 5 | Collinearity analysis of *CaNHX* gene in pepper. **(A)** Chiltepin; **(B)** Corno; **(C)** Cuneo; **(D)** Quadrato; **(E)** Tumaticot; **(F)** CM334; **(G)** ECW; **(H)** PBC81; **(I)** SF. Z: Zunla-1.

Supplementary Table 7). The analysis showed that *CaNHXs* may be regulated by such things as light and salicylic acid, and may participate in defense mechanisms through these *cis*-acting elements, thus playing a role in protecting plant growth.

Co-expression Network of *CaNHX* Gene Under Stress Treatment

According to established methods, the co-expression network related to the *CaNHX* gene was extracted. Under stress, 10 groups of data were obtained: ABA, IAA, GA3, SA, JA, sodium chloride, mannitol, hydrogen peroxide, heat stress, cold stress, plus control groups. The co-expressed gene of the *CaNHX* gene was extracted by script, and the co-expressed gene related to the *NHX* gene under stress was obtained (**Figure 7** and **Supplementary Table 8**). According to the co-expression network, in the control group, the genes co-expressed with the *NHX* gene contained fewer co-expression network genes than the other 10 groups, among which ABA, IAA, GA3, and mannitol were the most abundant. The number of co-expression genes was the highest under heat stress and cold stress, but other co-expression genes were also present under NaCl stress. Under salt stress, a total of 4 *CaNHX* genes were co-expressed with transcription factors, among which *CaNHX9* co-expressed the most with 11 transcription factors, followed by *CaNHX4* with 7 transcription factors, and *CaNHX1* with only 1 NAC co-expressed with *CaNHX1* was the least.

TABLE 2 | Nucleotide substitution rate of Pepper *CaNHX* gene.

Collateral homologous gene site	Non-synonymous substitution rate (Ka)	Synonymous substitution rate (Ks)	Selective pressure ratio (Ka/Ks)
CaNHX1-CaNHX6	0.201701088	2.493573377	0.080888371
CaNHX2-CaNHX8	0.026980027	0.041710828	0.646835085
CaNHX4-CaNHX5	0.155507334	0.327196446	0.475272076

Expression Pattern of Pepper *NHX* Genes Under Hormone and Abiotic Stress

The expression profile of the *CaNHX* gene in pepper was analyzed using an online database (**Figure 8**). The results showed that the expression of *CaNHX3*, *CaNHX4*, and *CaNHX5* in roots was upregulated compared with that in leaves during IR (IAA root stress) and GR (GA3 root stress), while the expression of *CaNHX2*, *CaNHX6*, *CaNHX7*, *CaNHX8*, and *CaNHX9* was lower in leaves. Under hormone stress, *CaNHX1* expression was upregulated. Under abiotic stress, *CaNHX1* and *CaNHX9* genes were upregulated in HR (heat root stress), while *CaNHX2*, *CaNHX3*, and *CaNHX4* were upregulated in RR (H_2O root stress), FR (cold root stress), and MR (mannitol root stress), whereas *CaNHX5* was upregulated in FL (cold leaf stress), *CaNHX6* was upregulated in HL (heat leaf stress), *CaNHX7* was upregulated in RL (H_2O leaf stress), *CaNHX7* was upregulated in RL and *CaNHX8* in FL. At the same time, expression analysis of the *CaNHX* gene family under salt stress showed that *CaNHX1*, *CaNHX3*, *CaNHX4*, *CaNHX5*, and *CaNHX9* were upregulated in roots, and their expression tended to be consistent over time. *CaNHX6*, *CaNHX7*, and *CaNHX8* were upregulated in NL (NaCl

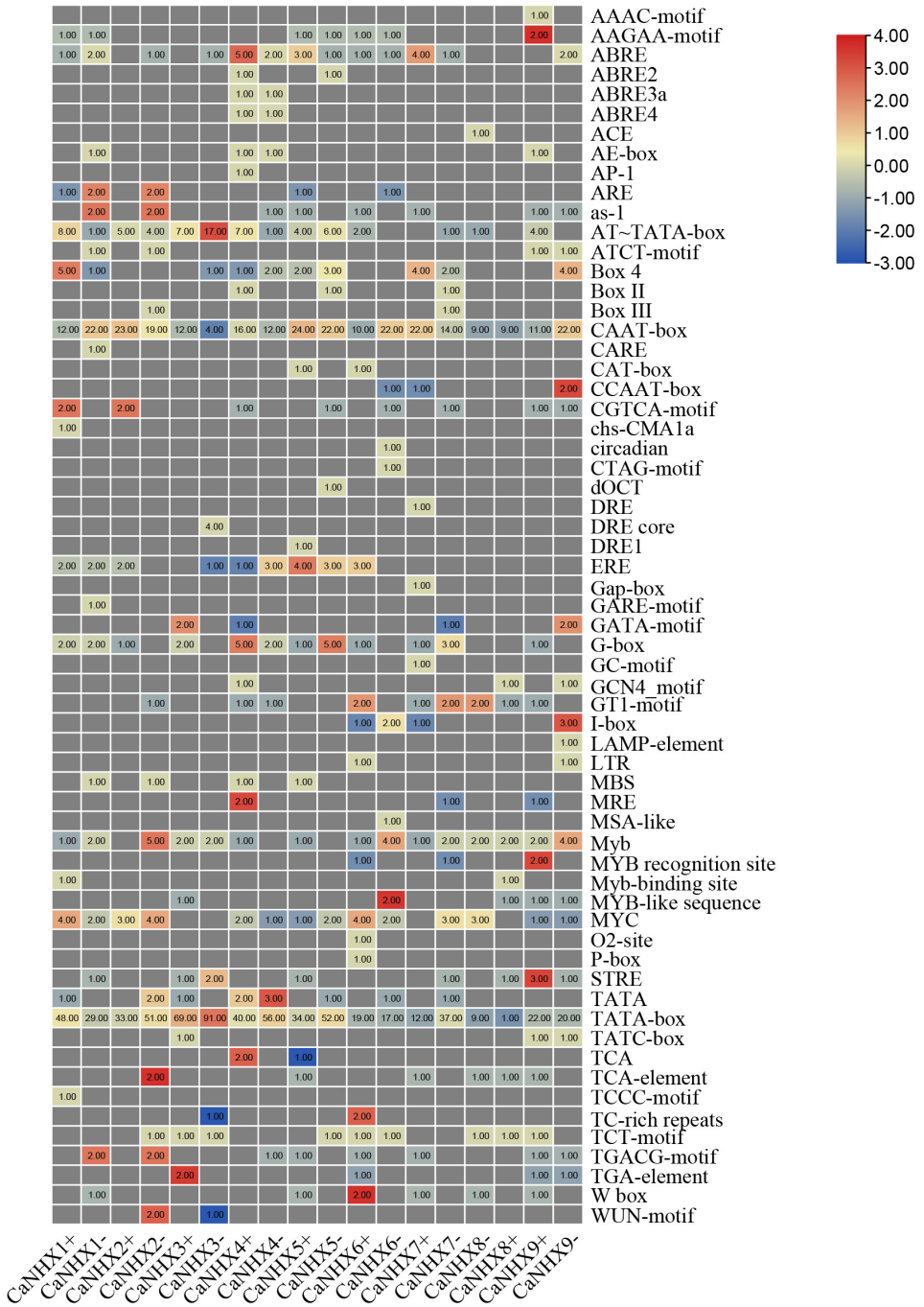


FIGURE 6 | *Cis*-acting elements of *C. annuum CaNHX* genes.

leaf stress), while the root expression of *CaNHX2* was higher in the blank control than in the leaves.

Quantitative RT-PCR Analysis

The pepper at the 6-leaf growth stage was subjected to 100 Mmol salt stress, and the samples at different times (3, 6, 12, 24, and 72 h) were taken for fluorescence quantitative analysis (**Figure 9**). Results show that, under salt stress, when processing *CaNHX1*, *CaNHX2*, *CaNHX6*, *CaNHX9* showed a trend of increased expression, the five time node, *CaNHX1*, *CaNHX9* two genes in a state of relative balance, express no obvious floating. The expression of *CaNHX2* and *CaNHX6* began to be down-regulated over time after initial stress treatment and then began to be up-regulated after 24 h. The expression of *CaNHX3*, *CaNHX4*, *CaNHX5,* and *CaNHX8* genes was less obvious than others. *CaNHX1* with *CaNHX9* fluorescence quantitative results agree

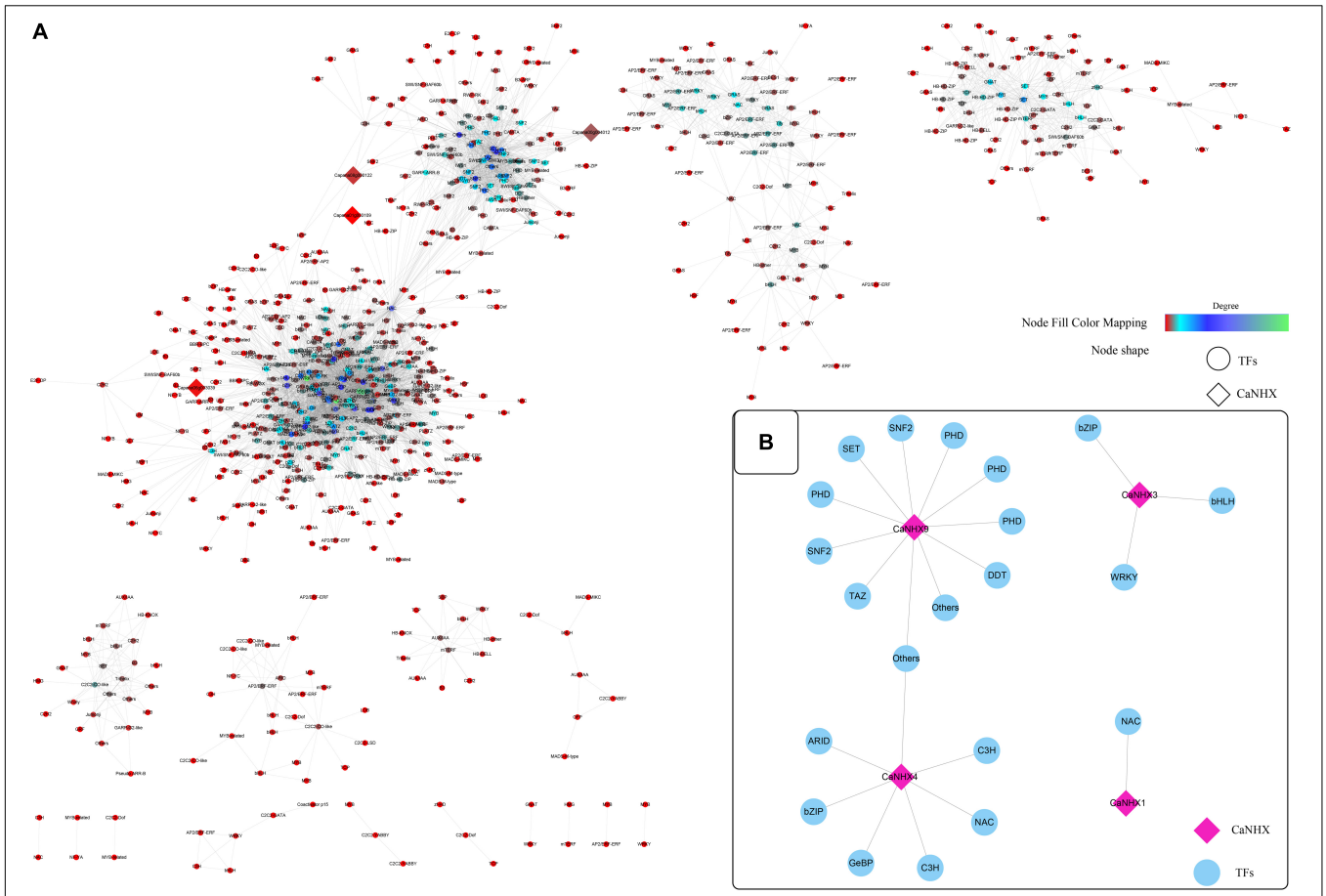


FIGURE 7 | (A,B) Co-expression network of *C. annuum CaNHX* gene under NaCl stress (Diamond: CaNHX genes; Circular: TFs; Node Fill Color Mapping: Degree).

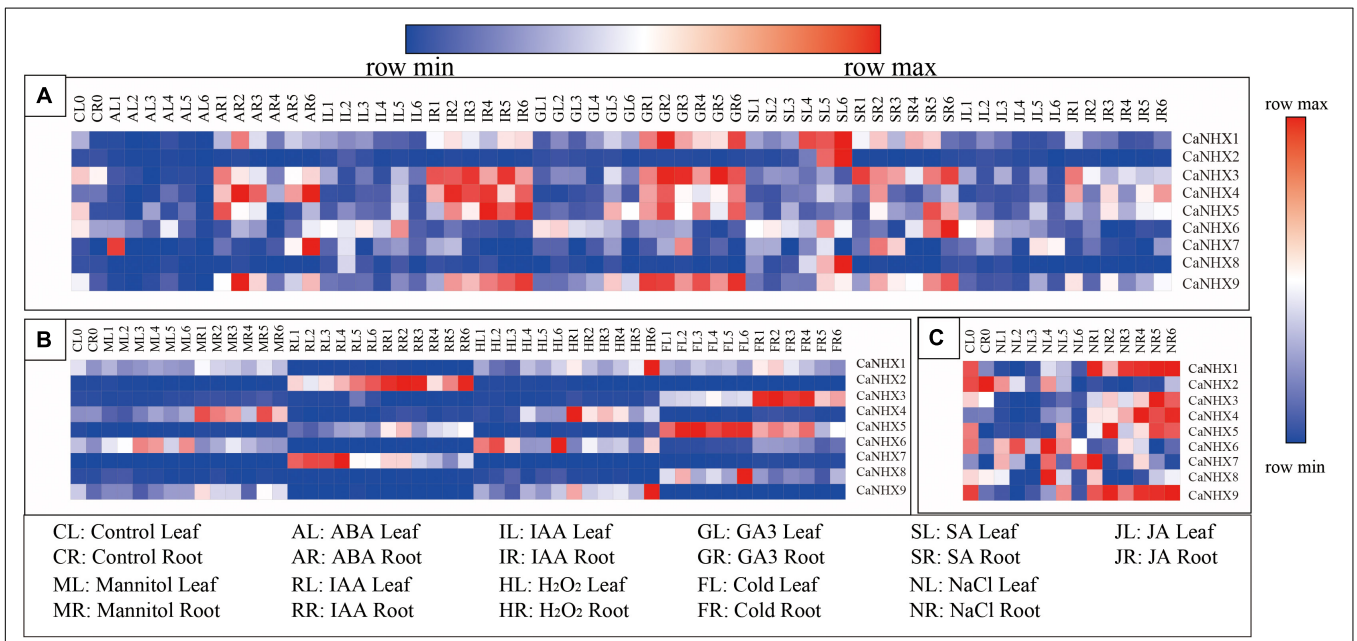


FIGURE 8 | Expression pattern of *CaNHX* gene in pepper under different stress in root and leaf. **(A)** Hormonal stress. **(B)** Abiotic stress. **(C)** Salt stress.

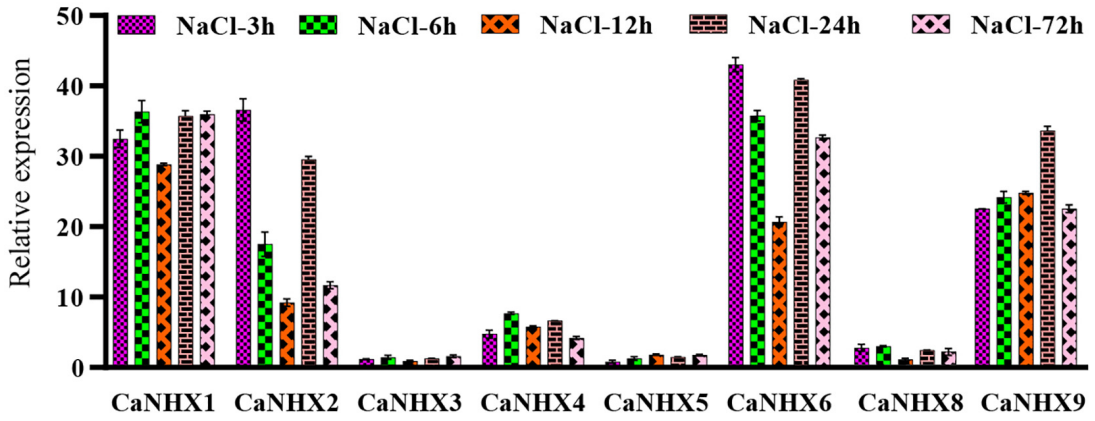


FIGURE 9 | Quantitative RT-PCR analysis of pepper *CaNHX* gene.

with the transcriptome data, in response to salt stress were a higher expressed state, *CaNHX2*, *CaNHX6*, *CaNHX7*, *CaNHX8* increase is not obvious in the transcriptome, which no expression in fluorescence quantitative *CaNHX7* gene, do not make the same amount in the transcriptome. In conclusion, two genes, *CaNHX1* and *CaNHX9*, were stably expressed in pepper under salt stress, which was consistent with transcriptome results. It was speculated that pepper could adjust its own ion balance by up-regulating the expression of *CaNHX1* and *CaNHX9* genes in the process of salt stress, so that pepper could adapt itself to the changes in the environment.

DISCUSSION

NHX Gene Family in Pepper

We identified 9 *CaNHX* genes in the pepper genome (zunla-1), 6 in *C. annuum* chiltepin, 9 in *C. chinense*. PI159236, and 9 in *C. annnuum*. Cv.CM334.There were 9 identified in *C. nnuum*. Cv. ECW, 10 identified in *C. accatum*. PBC81, 8 identified in *C. nnuum*. SF, 8 identified in Corno, 8 identified in Cuneo, 9 identified in Quadrato, and 8 identified in Tumaticot. The *CaNHX* gene was found to contain up to 9 genes in different varieties of pepper. So far, *NHX* genes have been identified in many species, with the most identified in *Gossypium hirsutum* and *G. barbadense* (23 *GhNHX* and 24 *GbNHX*, respectively) (Fu et al., 2020).

However, in subcellular localization, *NHX* genes can be classified into three subgroups according to their subcellular localization according to previous reports which were divided into three categories, namely Vac-Class, Endo-class, and PM-class, among which VAC-class is located in Vacuole, Endo-class in Endoretinal Reticulum, and PM-class in Plasma membrane (Wu et al., 2019). Other *NHX* gene species based on subcellular localization also include *A. thaliana*, *P. trichocarpa*, *G. hirsutum*, *V. vinifera*, *Triticum aestivum*, *Oryza sativa*, *Sorghum bicolor*, *Cucurbita maxima*, *Solanum lycopersicum*, *Panicum virgatum*, *Eutrema halophilum*,*Spinacia oleracea,* and *Hordeum vulgare*. According to the classification of grapes by Ayadi et al. (2020) the *VvNHX* genes of grapes are divided into two categories, namely Group I Vacuolar (*VvNHX1*–*VvNHX5*) and Group II Endosomal (*VvNHX6*). However, *CaNHX* genes in pepper are not classified according to their subcellular location. *CaNHX2* and *CaNHX4* are located in the Cell membrane, while the other subcellular locations of *CaNHX* gene family members are located in Vacuole. There is no PM-Class and Endo-class in pepper, which is quite different from previous studies.

The *CaNHX* gene in Zunla-1 can be divided into three subfamilies, namely Class I, Class II, and Class III. In the identified *NHX* gene families, *NHX* contains a complete functional domain. These *CaNHX* genes can be divided into 3 categories, which are the same as *A. thaliana*, beet, and other plants reported by predecessors (Wu et al., 2019). In addition to *CaNHX*7, *CaNHX*8, and *CaNHX*9, the other six genes also have the typical amilorid-binding site of the *NHX* gene (FFI/LY/FLLPPI), but this structure does not exist in the Class II subgroup. At the same time, the intermediate residues LF/AV/IY, LF in Class I, LA in Class II, and IY in Class III, *PgNHX* gene also had the same motif and residues in pomegranate. Not only pepper but also *A. thaliana* had an *NHX* gene without an amiloride-binding site (Counillon et al., 1993; Dong et al., 2021). In the wheat *TaNHX* gene, it was found that under salt stress, the expression of *TaNHX2* and *TaNHX3* genes were higher in leaves and roots, and the expression of *TaNHX1* was higher only in roots. All three *TaNHX* genes contained an amiloride-binding site (LFFIYLLPPI) (Brini et al., 2005; Yu et al., 2007; Lu et al., 2014). It is speculated that the *NHX* gene containing an amiloride-binding site is more suitable for the growth of salinization conditions.

Plants are affected by the external environment during their growth, such as abiotic stress, drought, high temperature, salt, and alkali, etc. Transcription is particularly important in the response of plants to environmental changes. There are many *cis*-acting elements in the pepper *CaNHX* genes family, such as hormones and stress elements. Studies show that stress-related elements (such as high temperature, low temperature, drought, injury, and defense) and hormone-related elements (such as Auxin, Ethylene, GA, SA, MeJA, and ABA) are identified in the

promoter of *PtNHXs*, and there are also *cis*-acting elements such as ABA and ABRE in sugar beet, which indicate that they can pass through during plant growth (Tian et al., 2017; Wu et al., 2019). In the transcriptional data of pepper, it was found that the co-expression networks of pepper under biotic and abiotic stress had higher gene network abundance than those under untreated conditions. In the co-expression network, *CaNHX*1, *CaNHX*3, *CaNHX*4, and *CaNHX*9 were found to be co-expressed with transcription factors, among which *CaNHX*3 was co-expressed with transcription factor WRKY, indicating that the *cis*-acting element of *CaNHX*3 was G-box, which has been found in studies. The G-box is an element associated with WRKY transcription factors under stress conditions. *CaNHX*3 was upregulated in several periods.

Expression Profiles of *NHX* Genes in Pepper

Up to now, the function of the *CaNHX* gene in pepper has not been analyzed, and no report on the *CaNHX* gene in pepper has been reported. With the transformation of salt-tolerant transgenic plants, the *NHX* gene will provide more benefits for agricultural development in soil salinization. Transgenic technology has become one of the important ways to obtain salt-tolerant plants and verify gene function (Dhankher and Foyer, 2018). *NHX* can improve the salt tolerance of transgenic plants, and the overexpression of *AtNHX5* in rice can improve the salt tolerance and drought tolerance of transgenic rice, and the survival rate is higher (Li et al., 2011). *SbNHXLP* can improve the salt tolerance of tomatoe, and the Na^+ level in tomato is lower, and the Ca^{2+} level is higher, compared with wild-type plants. *SbNHXLP* maintained ion homeostasis in tomato and alleviated NaCl stress (Kumari et al., 2017). When plants are subjected to salt stress, due to the lack of *NHX* expression to maintain homeostasis, the premature apoptosis of plants is caused, and the growth of plants is inhibited, thus affecting the yield of plants. Cao et al. (2011) through the overexpression of *TaNHX2*, improve the survival time of transgenic plants, showing salt tolerance. The number of flowering was more than that of the control group (Cao et al., 2011). In recent years, with the further exploration of the function of *NHX*, it has been found that *NHX* is resistant to cadmium, and the overexpression of *GmNHX1* enhances the antioxidant capacity of plants and reduces the absorption of cadmium (Yang et al., 2017). It was also found that silencing genes in plants had a great influence on plant growth and development. Rodríguez-Rosales et al. (2008) found that tomato seedling growth, fruit, and seed yield had significant inhibitory effects by silencing tomato *LeNHX2*. Overexpression of *LeNHX2* can enhance salt tolerance in plants (Rodríguez-Rosales et al., 2008; Baghour et al., 2019). In conclusion, overexpression of the *NHX* gene can improve ion homeostasis, osmotic regulation, reduce cell membrane damage, improve photosynthetic capacity, and play a role in plant protection and yield increase. The *CaNHX* gene in pepper has never been published and identified before. This study can provide theoretical support for research on the salt tolerance of pepper.

We found that the expression of *CaNHXs* was mainly concentrated in roots under hormone stress, while under abiotic stress, there were up-regulated expressions of *CaNHX2*, *CaNHX3,* and *CaNHX4* genes in roots. In the *cis*-acting elements of the *CaNHX* gene family, it was found that *CaNHX* was expressed under various hormones and stress. However, under salt stress, most of the *CaNHX* genes were up-regulated, which indicated that *CaNHX* genes in Pepper could condition its ion balance by expressing *NHX*. The function of plant vacuole *NHX* antiporter has been identified and expressed in an exogenous system to enhance the salt tolerance of plants. Akram et al. (2020) found that the *NHX* gene was up-regulated under salt stress, and Yokoi et al. (2002) found that *AtNHX1* had higher transcript abundance during salt stress (Yarra, 2019). The results indicated that the expression of *NHX* genes responded to salt stress during plant growth, which played a very important role in plant growth.

Transcriptome analysis found that the *CaNHX* gene had multiple expression patterns under single or multiple stress conditions. Meanwhile, the fluorescence quantitative verification in this study showed that the results were consistent with the transcriptome results, in which *CaNHX1* and *CaNHX9* were up-regulated under salt stress. These results indicated that the *CaNHX* gene provided a guarantee for the normal growth of pepper and the balance of ion channels of plant stress resistance.

CONCLUSION

In the pepper genome, 42 *CaNHX* genes were identified by a Hidden Markov Model database search (hmmsearch). Of these, 9 genes with complete functional domains were identified by BLASTP. We constructed a phylogenetic tree and found that the 9 *CaNHX* genes were divided into three categories: Class I, Class II, and Class III. The exon number of the Class I subgroup was relatively stable, and the genes were distributed on six chromosomes; these were for hydrophilic proteins. There was a motif amilorid-binding site of the *NHX* gene (FFI/LY/FLLPPI) associated with salt tolerance in the pepper *CaNHX* gene. There are many elements in the *CaNHX* gene, such as hormone stress, salt stress, and so on, and it was found that the *CaNHX* gene is associated with many genes in the co-expression process, and salt stress conditions are also associated with many genes. Transcriptome analysis showed that the *CaNHX* gene was up-regulated under various abiotic stresses, which was verified in combination with fluorescence quantification in this study and found to be consistent with transcriptome results. In this study, the whole gene of Pepper was identified at the genome level, which provided a theoretical basis for the genetic breeding of pepper under stress.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XL, SY, ZZ, and CQ conceived and designed the experiments, and drafted the manuscript. YL, HQ, TL, JL, XC, XZ, YC, and JZ performed the experiments. SY and CQ analyzed the data. All authors read and approved the manuscript.

FUNDING

This work was supported by the Special Financial Grant from the China Postdoctoral Science Foundation (No. 2017T100718), the National Natural Science Foundation of China (Nos. 31660575 and 31860498), the Guizhou Province Science and Technology Plan Program of China (Qian Kehe Support Nos. [2019]2258, [2020]1Y088, and [2019]2413), the Government Special Funds for Guiding Local Science and Technology Development of China (Qianke Zhongyindi No. 20174003), the Zunyi Innovative Talent Team Training Project of China (Zunshi Kehe Rencai No. 201904), and the Zunyi Excellent Young Scientific and Technological Innovative Talents Training Project (Zunyouqingke No. 201806).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.680457/full#supplementary-material

REFERENCES

Akram, U., Song, Y., Liang, C., Abid, M. A., Askari, M., Myat, A. A., et al. (2020). Genome-wide characterization and expression analysis of NHX Gene family under salinity stress in gossypium barbadense and its comparison with *Gossypium hirsutum*. *Genes* 11:803. doi: 10.3390/genes11070803

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, W597–W603. doi: 10.1093/nar/gks400

Ayadi, M., Martins, V., Ben Ayed, R., Jbir, R., Feki, M., Mzid, R., et al. (2020). Genome wide identification, molecular characterization, and gene expression analyses of grapevine NHX antiporters suggest their involvement in growth, ripening, seed dormancy, and stress response. *Biochem. Genet.* 58, 102–128. doi: 10.1007/s10528-019-09930-4

Baghour, M., Gálvez, F. J., Sánchez, M. E., Aranda, M. N., Venema, K., and Rodríguez-Rosales, M. P. (2019). Overexpression of LeNHX2 and SlSOS2 increases salt tolerance and fruit production in double transgenic tomato plants. *Plant Physiol. Biochem.* 135, 77–86. doi: 10.1016/j.plaphy.2018.11.028

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Benèina, M., Bagar, T., Lah, L., and Kraševec, N. (2009). A comparative genomic analysis of calcium and proton signaling/homeostasis in *Aspergillus* species. *Fungal Genet. Biol.* 46, S93–S104. doi: 10.1016/j.fgb.2008.07.019

Brini, F., Gaxiola, R. A., Berkowitz, G. A., and Masmoudi, K. (2005). Cloning and characterization of a wheat vacuolar cation/proton antiporter and pyrophosphatase proton pump. *Plant Physiol. Bioch.* 43, 347–354. doi: 10.1016/j.plaphy.2005.02.010

Cao, B., Long, D., Zhang, M., Liu, C., Xiang, Z., and Zhao, A. (2016). Molecular characterization and expression analysis of the mulberry Na+/H+ exchanger gene family. *Plant Physiol. Biochem.* 99, 49–58. doi: 10.1016/j.plaphy.2015.12.010

Cao, D., Hou, W., Liu, W., Yao, W., Wu, C., Liu, X., et al. (2011). Overexpression of TaNHX2 enhances salt tolerance of 'composite' and whole transgenic soybean plants. *Plant Cell Tiss. Org.* 10, 541–552. doi: 10.1007/s11240-011-0005-9

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 5:e11335. doi: 10.1371/journal.pone.0011335

Counillon, L., Franchi, A., and Pouyssegur, J. (1993). A point mutation of the Na+/H+ exchanger gene (NHE1) and amplification of the mutated allele confer amiloride resistance upon chronic acidosis. *Proc. Natl. Acad. Sci. U.S.A.* 90, 4508–4512.

Dhankher, O. P., and Foyer, C. H. (2018). Climate resilient crops for improving global food security and safety. *Plant Cell Environ.* 41, 877–884. doi: 10.1111/pce.13207

Dong, J., Liu, C., Wang, Y., Zhao, Y., Ge, D., and Yuan, Z. (2021). Genome-wide identification of the NHX gene family in *Punica granatum* L. and their expressional patterns under salt stress. *Agronomy* 11:264. doi: 10.3390/agronomy11020264

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2018). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Fu, X., Lu, Z., Wei, H., Zhang, J., Yang, X., Wu, A., et al. (2020). Genome-wide identification and expression analysis of the NHX (Sodium/Hydrogen Antiporter) gene family in cotton. *Front. Genet.* 11:964. doi: 10.3389/fgene.2020.00964

Gaxiola, R. A., Rao, R., Sherman, A., Grisafi, P., Alper, S. L., and Fink, G. R. (1999). The *Arabidopsis thaliana* proton transporters, AtNhx1 and Avp1, can function in cation detoxification in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 96:1480.

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2014). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Kantar, M. B., Anderson, J. E., Lucht, S. A., Mercer, K., Bernau, V., Case, K. A., et al. (2016). Vitamin variation in *Capsicum* Spp. Provides opportunities to improve nutritional value of human diets. *PLoS One* 11:e0161464. doi: 10.1371/journal.pone.0161464

Khan, A. L., Shin, J. H., Jung, H. Y., and Lee, I. J. (2014). Regulations of capsaicin synthesis in *Capsicum annuum* L. by *Penicillium resedanum* LK6 during drought conditions. *Sci. Hortic.* 175, 167–173. doi: 10.1016/j.scienta.2014.06.008

Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* 46, 270–278. doi: 10.1038/ng.2877

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kumari, P. H., Kumar, S. A., Sivan, P., Katam, R., Suravajhala, P., Rao, K. S., et al. (2017). Overexpression of a plasma membrane bound Na+/H+ antiporter-like protein (SbNHXLP) confers salt tolerance and improves fruit yield in tomato by maintaining ion homeostasis. *Front. Plant Sci.* 7:2027. doi: 10.3389/fpls.2016.02027

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, M., Lin, X., Li, H., Pan, X., and Wu, G. (2011). Overexpression of AtNHX5 improves tolerance to both salt and water stress in rice (*Oryza sativa* L.). *Plant Cell Tiss. Org.* 107, 283–293. doi: 10.1007/s11240-011-9979-6

Liu, F., Yu, H., Deng, Y., Zheng, J., Liu, M., Ou, L., et al. (2017). PepperHub, an informatics hub for the chili pepper research community. *Mol. Plant* 10, 1129–1132. doi: 10.1016/j.molp.2017.03.005

Liu, H., Yang, H., Zheng, J., Jia, D., Wang, J., Li, Y., et al. (2012). Irrigation scheduling strategies based on soil matric potential on yield and fruit quality of mulched-drip irrigated chili pepper in Northwest China. *Agric. Water Manage.* 115, 232–241. doi: 10.1016/j.agwat.2012.09.009

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, W., Guo, C., Li, X., Duan, W., Ma, C., Zhao, M., et al. (2014). Overexpression of TaNHX3, a vacuolar Na+/H+ antiporter gene in wheat, enhances salt stress tolerance in tobacco by improving related physiological processes. *Plant Physiol. Biochem.* 76, 17–28. doi: 10.1016/j.plaphy.2013.12.013

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., and Demchak, B. (2019). Cytoscape automation: empowering workflow-based network analysis. *Genome Biol.* 20, 1–15. doi: 10.1186/s13059-019-1758-4

Pedersen, S. F., O'Donnell, M. E., Anderson, S. E., and Cala, P. M. (2006). Physiology and pathophysiology of Na+/H+ exchange and Na+-K+-2Cl-cotransport in the heart, brain, and blood. *Am. J. Physiol. Reg. I* 291, R1–R25. doi: 10.1152/ajpregu.00782.2005

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Qin, C., Cheng, J., Wu, Z., Luo, X., and Hu, K. (2015). Research advances in the genomics of pepper. *J. Agric. Biotechem.* 23, 953–966. doi: 10.1111/j.1475-6773.2012.01428.x

Qin, C., Yu, C., Shen, Y., Fang, X., Min, J., Cheng, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.* 111:5135. doi: 10.1073/pnas.1400975111

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rodríguez-Rosales, M. P., Gálvez, F. J., Huertas, R., Aranda, M. N., Baghour, M., Cagnac, O., et al. (2009). Plant NHX cation/proton antiporters. *Plant Signal. Behav.* 4, 265–276. doi: 10.4161/psb.4.4.7919

Rodríguez-Rosales, M. P., Jiang, X., Gálvez, F. J., Aranda, M. N., Cubero, B., and Venema, K. (2008). Overexpression of the tomato K+/H+ antiporter LeNHX2 confers salt tolerance by improving potassium compartmentalization. *New Phytol.* 179, 366–377. doi: 10.1111/j.1469-8137.2008.02461.x

Saleh, B., Omer, A., and Teweldemedhin, B. (2018). Medicinal uses and health benefits of chili pepper (*Capsicum* spp.): a review. *MOJ Food Proc. Technol.* 6, 325–328. doi: 10.15406/mojfpt.2018.06.00183

Sandhu, D., Pudussery, M. V., Kaundal, R., Suarez, D. L., Kaundal, A., and Sekhon, R. S. (2018). Molecular characterization and expression analysis of the Na+/H+ exchanger gene family in *Medicago truncatula*. *Funct. Integr. Genomic.* 18, 141–153. doi: 10.1007/s10142-017-0581-9

Sarah, A., Todd, P. E., Meekins, J. F., Dale, P., and Marcia, M. (2012). Research article: effects of salt stress on capsaicin content, growth, and fluorescence in a Jalapeño cultivar of *Capsicum annuum* (Solanaceae). *BIOS* 83, 1–7. doi: 10.1893/0005-3155-83.1.1

Sardet, C., Franchi, A., and Pouysségur, J. (1989). Molecular cloning, primary structure, and expression of the human growth factor-activatable Na+/H+ antiporter. *Cell* 56, 271–280. doi: 10.1016/0092-8674(89)90901-X

Tian, F., Chang, E., Li, Y., Sun, P., Hu, J., and Zhang, J. (2017). Expression and integrated network analyses revealed functional divergence of NHX-type Na+/H+ exchanger genes in poplar. *Sci. Rep. UK* 7:2607. doi: 10.1038/s41598-017-02894-8

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinf.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wu, G. Q., Wang, J. L., and Li, S. J. (2019). Genome-wide identification of Na+/H+ antiporter (NHX) genes in sugar beet (*Beta vulgaris* L.) and their regulated expression under salt stress. *Genes* 10:401. doi: 10.3390/genes10050401

Yang, L., Han, Y., Wu, D., Yong, W., Liu, M., Wang, S., et al. (2017). Salt and cadmium stress tolerance caused by overexpression of the Glycine Max Na+/H+ Antiporter (GmNHX1) gene in duckweed (Lemna turionifera 5511). *Aquat. Toxicol.* 192, 127–135. doi: 10.1016/j.aquatox.2017.08.010

Yarra, R. (2019). The wheat NHX gene family: potential role in improving salinity stress tolerance of plants. *Plant Gene* 18:100178. doi: 10.1016/j.plgene.2019.100178

Ye, C. Y., Zhang, H. C., Chen, J. H., Xia, X. L., and Yin, W. L. (2009). Molecular characterization of putative vacuolar NHX-type Na+/H+ exchanger genes from the salt-resistant tree *Populus euphratica*. *Physiol. Plantarum* 137, 166–174. doi: 10.1111/j.1399-3054.2009.01269.x

Yokoi, S., Quintero, F. J., Cubero, B., Ruiz, M. T., Bressan, R. A., Hasegawa, P. M., et al. (2002). Differential expression and function of *Arabidopsis thaliana* NHX Na+/H+ antiporters in the salt stress response. *Plant J.* 30, 529–539. doi: 10.1046/j.1365-313X.2002.01309.x

Yu, J. N., Huang, J., Wang, Z. N., Zhang, J. S., and Chen, S. Y. (2007). An Na+/H+ antiporter gene from wheat plays an important role in stress tolerance. *J. Biosci.* 32, 1153–1161. doi: 10.1007/s12038-007-0117-x

Zhang, X. M., Zhang, Z. H., Gu, X. Z., Mao, Sl, Li, X. X., Chadśuf, J., et al. (2016). Genetic diversity of pepper (*Capsicum* spp.) germplasm resources in China reflects selection for cultivar types and spatial distribution. *J. Integr. Agric.* 15, 1991–2001. doi: 10.1016/S2095-3119(16)61364-3

Zhang, Y., Fang, J., Wu, X., and Dong, L. (2018). Na+/K+ balance and transport regulatory mechanisms in weedy and cultivated rice (*Oryza sativa* L.) under salt stress. *BMC Plant Biol.* 18:375. doi: 10.1186/s12870-018-1586-9

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). ITAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Luo, Yang, Luo, Qiu, Li, Li, Chen, Zheng, Chen, Zhang, Zhang and Qin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership