frontiers | Research Topics

Legume breeding in transition: Innovation and outlook

Edited by
Rafiul Amin Laskar, Aamir Raina, Samiullah Khan, Nasya Borisova Tomlekova, Waltram Ravelombola and Mahendar Thudi

Published in
Frontiers in Genetics



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 978-2-8325-2218-9
DOI 10.3389/978-2-8325-2218-9

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

Legume breeding in transition: Innovation and outlook

Topic editors

Rafiul Amin Laskar — Department of Botany, Bahona College, India
Aamir Raina — Aligarh Muslim University, India
Samiullah Khan — Aligarh Muslim University, India
Nasya Borisova Tomlekova — Maritsa Vegetable Crops Research Institute (MVCRI), Bulgaria
Waltram Ravelombola — Texas A and M University, United States
Mahendar Thudi — Dr. Rajendra Prasad Central Agricultural University, India

Citation

Laskar, R. A., Raina, A., Khan, S., Tomlekova, N. B., Ravelombola, W., Thudi, M., eds. (2023). *Legume breeding in transition: Innovation and outlook*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2218-9

Table of contents

07 Editorial: Legume breeding in transition: innovation and outlook
Aamir Raina, Rafiul Amin Laskar, Samiullah Khan, Nasya Borisova Tomlekova, Waltram Ravelombola and Mahendar Thudi

10 Cloning and Functional Characterization of a Pericarp Abundant Expression Promoter (*AhGLP17-1P*) From Peanut (*Arachis hypogaea* L.)
Yasir Sharif, Hua Chen, Ye Deng, Niaz Ali, Shahid Ali Khan, Chong Zhang, Wenping Xie, Kun Chen, Tiecheng Cai, Qiang Yang, Yuhui Zhuang, Ali Raza and Weijian Zhuang

23 Transcriptome Analysis Reveals Key Pathways and Candidate Genes Controlling Seed Development and Size in Ricebean (*Vigna umbellata*)
Sachin Kumar Verma, Shikha Mittal, Gayacharan, Dhammaprakash Pandhari Wankhede, Swarup Kumar Parida, Debasis Chattopadhyay, Geeta Prasad, Dwijesh Chandra Mishra, Dinesh Chandra Joshi, Mohar Singh, Kuldeep Singh and Amit Kumar Singh

43 Identification of QTLs/ Candidate Genes for Seed Mineral Contents in Common Bean (*Phaseolus vulgaris* L.) Through Genotyping-by-Sequencing
Muslima Nazir, Reetika Mahajan, Sheikh Mansoor, Sheezan Rasool, Rakeeb Ahmad Mir, Ravinder Singh, Vandana Thakral, Virender Kumar, Parvaze A. Sofi, Hamed A. El-Serehy, Daniel Ingo Hefft and Sajad Majeed Zargar

58 High-Density Genetic Variation Map Reveals Key Candidate Loci and Genes Associated With Important Agronomic Traits in Peanut
Huiling Zhao, Ruizheng Tian, Han Xia, Changsheng Li, Guanghui Li, Aiqin Li, Xianying Zhang, Ximeng Zhou, Jing Ma, Huailing Huang, Kun Zhang, Mahendar Thudi, Changle Ma, Xingjun Wang and Chuanzhi Zhao

69 Progress of Genomics-Driven Approaches for Sustaining Underutilized Legume Crops in the Post-Genomic Era
Uday Chand Jha, Harsh Nayyar, Swarup K Parida, Melike Bakır, Eric J. B. von Wettberg and Kadambot H. M. Siddique

93 Introgressing cry1Ac for Pod Borer Resistance in Chickpea Through Marker-Assisted Backcross Breeding
Ajinder Kaur, Urvashi Sharma, Sarvjeet Singh, Ravinder Singh, Yogesh Vikal, Satnam Singh, Palvi Malik, Khushpreet Kaur, Inderjit Singh, Shayla Bindra, Bidyut Kumar Sarmah and Jagdeep Singh Sandhu

108 The bZIP Transcription Factor Family in Adzuki Bean (*Vigna Angularis*): Genome-Wide Identification, Evolution, and Expression Under Abiotic Stress During the Bud Stage
Zhengong Yin, Xianxin Meng, Yifan Guo, Shuhong Wei, Yongcai Lai and Qiang Wang

121 Magnesium- a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm
Faheem Shehzad Baloch, Muhammad Azhar Nadeem, Ferit Sönmez, Ephrem Habyarimana, Zemran Mustafa, Tolga Karaköy, Gönül Cömertpay, Ahmad Alsaleh, Vahdettin Çiftçi, Sangmi Sun, Gyuhwa Chung and Yong Suk Chung

132 Erratum: Magnesium-a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm
Frontiers Production Office

133 *GmARP* is Related to the Type III Effector NopAA to Promote Nodulation in Soybean (*Glycine max*)
Jinhui Wang, Chao Ma, Shengnan Ma, Haiyang Zheng, Haojie Feng, Yue Wang, Jiangxu Wang, Chunyan Liu, Dawei Xin, Qingshan Chen and Mingliang Yang

145 Genome-Wide Association Study and Genomic Prediction for Bacterial Wilt Resistance in Common Bean (*Phaseolus vulgaris*) Core Collection
Bazgha Zia, Ainong Shi, Dotun Olaoye, Haizheng Xiong, Waltram Ravelombola, Paul Gepts, Howard F. Schwartz, Mark A. Brick, Kristen Otto, Barry Ogg and Senyu Chen

158 Gene-Editing Technologies and Applications in Legumes: Progress, Evolution, and Future Prospects
Mehmet Cengiz Baloglu, Yasemin Celik Altunoglu, Pinar Baloglu, Ali Burak Yildiz, Nil Türkölmez and Yelda Özden Çiftçi

178 Genome-wide Association Study for Yield and Yield-Related Traits in Diverse Blackgram Panel (*Vigna mungo* L. Hepper) Reveals Novel Putative Alleles for Future Breeding Programs
Lovejit Singh, Guriqbal Singh Dhillon, Sarabjit Kaur, Sandeep Kaur Dhaliwal, Amandeep Kaur, Palvi Malik, Ashok Kumar, Ranjit Kaur Gill and Satinder Kaur

195 Exploring Chickpea Germplasm Diversity for Broadening the Genetic Base Utilizing Genomic Resourses
Rajesh Kumar Singh, Charul Singh, Ambika, B. S. Chandana, Rohit K. Mahto, Ranjana Patial, Astha Gupta, Vijay Gahlaut, Gayacharan, Aladdin Hamwieh, H. D. Upadhyaya and Rajendra Kumar

224 Multilocation Yield Trials and Yield Stability Evaluation by GGE Biplot Analysis of Promising Large-Seeded Peanut Lines
Nipatcha Pobkhunthod, Jetsada Authapun, Songyos Chotchutima, Sarawut Rungmekarat, Piya Kittipadakul, Jaungjun Duangpatra and Tanapon Chaisan

234 Development of High Yielding Fusarium Wilt Resistant Cultivar by Pyramiding of "Genes" Through Marker-Assisted Backcrossing in Chickpea (*Cicer arietinum* L.)
C. Bharadwaj, J. Jorben, Apoorva Rao, Manish Roorkiwal, B. S. Patil, Jayalakshmi, S. Khayum Ahammed, D. R. Saxena, M. Yasin, J. E. Jahagirdar, P. L. Sontakke, M. S. Pithia, M. K. Chudasama, Indu Swarup, R. K. Singh, S. D. Nitesh, Annapurna Chitikineni, Sarvjeet Singh, Inderjit Singh, Aditya Pratap, G. P. Dixit, A. K. Srivastava and Rajeev K. Varshney

244 Harnessing the potential of bulk segregant analysis sequencing and its related approaches in crop breeding
Aasim Majeed, Prerna Johar, Aamir Raina, R. K. Salgotra, Xianzhong Feng and Javaid Akhter Bhat

259 Comparative transcriptome analysis, unfolding the pathways regulating the seed-size trait in cultivated lentil (*Lens culinaris* Medik.)
Haragopal Dutta, Gyan P. Mishra, Muraleedhar S. Aski, Tejas C. Bosamia, Dwijesh C. Mishra, Jyotika Bhati, Subodh Kumar Sinha, Dunna Vijay, Manjunath Prasad C. T., Shouvik Das, Prashant Anupama-Mohan Pawar, Atul Kumar, Kuldeep Tripathi, Ranjeet Ranjan Kumar, Devendra Kumar Yadava, Shiv Kumar and Harsh Kumar Dikshit

277 Linkage QTL Mapping and Genome-Wide Association Study on Resistance in Chickpea to *Pythium ultimum*
Chiti Agarwal, Weidong Chen, Rajeev Kumar Varshney and George Vandemark

288 Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (*Glycine max* L.)
Javaid Akhter Bhat, Kehinde Adewole Adeboye, Showkat Ahmad Ganie, Rutwik Barmukh, Dezhou Hu, Rajeev K. Varshney and Deyue Yu

307 Genome-wide association mapping for root traits associated with frost tolerance in faba beans using KASP-SNP markers
Ahmed Sallam, Yasser S. Moursi, Regina Martsch and Shamseldeen Eltaher

319 Innovation in the Breeding of Common Bean Through a Combined Approach of *in vitro* Regeneration and Machine Learning Algorithms
Muhammad Aasim, Ramazan Katirci, Faheem Shehzad Baloch, Zemran Mustafa, Allah Bakhsh, Muhammad Azhar Nadeem, Seyid Amjad Ali, Rüştü Hatipoğlu, Vahdettin Çiftçi, Ephrem Habyarimana, Tolga Karaköy and Yong Suk Chung

332 Incorporation of Soil-Derived Covariates in Progeny Testing and Line Selection to Enhance Genomic Prediction Accuracy in Soybean Breeding
Caio Canella Vieira, Reyna Persa, Pengyin Chen and Diego Jarquin

347 Association mapping for important agronomic traits in wild and cultivated *Vigna* species using cross-species and cross-genera simple sequence repeat markers
Gita Kumari, P. S. Shanmugavadivel, G. Roopa Lavanya, Pravin Tiwari, Dharmpal Singh, P. G. Gore, Kuldeep Tripathi, Ramakrishnan Madhavan Nair, Sanjeev Gupta and Aditya Pratap

362 Progress and prospectus in genetics and genomics of *Phytophthora* root and stem rot resistance in soybean (*Glycine max* L.)
Subhash Chandra, Mukesh Choudhary, Pravin K. Bagaria, Vennampally Nataraj, Giriraj Kumawat, Jeet Ram Choudhary, Humira Sonah, Sanjay Gupta, Shabir Hussain Wani and Milind B. Ratnaparkhe

382 Transcriptome profiling of two contrasting pigeon pea (*Cajanus cajan*) genotypes in response to waterlogging stress
Anshika Tyagi, Sandhya Sharma, Harsha Srivastava, Anuradha Singh, Tanvi Kaila, Sajad Ali, Ambika B. Gaikwad, N. K Singh and Kishor Gaikwad

388 Mining legume germplasm for genetic gains: An Indian perspective
Gayacharan, Swarup K. Parida, Nupur Mondal, Rashmi Yadav, Harinder Vishwakarma and Jai C. Rana

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 30 May 2023
DOI 10.3389/fgene.2023.1221551

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Andrew H. Paterson,
University of Georgia, United States

*CORRESPONDENCE
Rafiul Amin Laskar,
rafihkd@gmail.com

RECEIVED 12 May 2023
ACCEPTED 24 May 2023
PUBLISHED 30 May 2023

CITATION
Raina A, Laskar RA, Khan S, Tomlekova NB, Ravelombola W and Thudi M (2023), Editorial: Legume breeding in transition: innovation and outlook.
Front. Genet. 14:1221551.
doi: 10.3389/fgene.2023.1221551

COPYRIGHT
© 2023 Raina, Laskar, Khan, Tomlekova, Ravelombola and Thudi. 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.

Editorial: Legume breeding in transition: innovation and outlook

Aamir Raina[1,2], Rafiul Amin Laskar[3]*, Samiullah Khan[1], Nasya Borisova Tomlekova[4], Waltram Ravelombola[5] and Mahendar Thudi[6]

[1]Mutation Breeding Laboratory, Department of Botany, Aligarh Muslim University, Aligarh, India, [2]Botany Section, Women's College, Aligarh Muslim University, Aligarh, India, [3]Department of Botany, Bahona College, Jorhat, India, [4]Maritsa Vegetable Crops Research Institute, Plovdiv, Bulgaria, [5]Texas A&M University, College Station, TX, United States, [6]Dr. Rajendra Prasad Central Agricultural University, Samastipur, India

KEYWORDS

legumes (Fabaceae), food security and nutrition, biotic and abiotic stress, underutilized legumes/orphan legumes, QTLs (quantitative trait loci), GWAS (genome-wide association study), MAS (marker assisted selection), transcriptome analysis

Editorial on the Research Topic
Legume breeding in transition: innovation and outlook

Introduction

Legumes are important crops that are primarily harvested for their grains, which are rich in proteins, minerals, and other nutrients such as vitamins, fibre, and antioxidants. Legumes are mostly self-pollinated crops, which means they have a narrow genetic base, that poses challenges to crop improvement programs. Still, conventional and modern breeding approaches have contributed significantly to improving the agronomic traits, stress tolerance, and nutritional qualities of legume crops. Conventional breeding involves exposing plant propagules to mutagens and/or crossing two or more plants to generate a new generation with desired traits, while modern breeding approaches include molecular breeding, marker-assisted selection, and genetic engineering techniques. Through these approaches, researchers have been able to develop legume varieties with improved yield, disease resistance, drought tolerance, and nutritional qualities such as higher protein content, iron, zinc, and other essential micronutrients. Both conventional modern breeding approaches have achieved much success in cereal crops and very little attention has been given toward the improvement of legume crops. The genetic improvement of major and underutilized legume crops remains a major challenge in the path leading to the goal of global food security and nutrition. This Research Topic hosted at Frontiers in Genetics entitled "Legume Breeding in Transition: Innovation and Outlook" presents a series of research articles and reviews covering new understandings in the areas of germplasm diversity, transcriptomics, sequencing, genomics, marker assisted backcross breeding, genome wide association study, genome editing, machine learning algorithms and agronomy integrating theoretical, and experimental approaches.

What have we achieved so far in legume breeding?

Recent advances in genomics offer hope for future genetic improvements in important legumes. Completing genome assemblies and resequencing efforts of large germplasm collections have made it possible to identify the underlying genes governing various important traits, which can enhance genetic gain and help develop more climate-resilient cultivars. Adzuki bean, cluster bean, horse Gram, lathyrus, red clover, urd bean, and winged bean are among the underutilized legumes that can benefit from these advancements in genomics. Gayacharan et al., reported that genetic gain in legumes can be enhanced by mining approximately 0.4 million *ex-situ* collections of legumes in gene banks. This would facilitate the identification of ideal donors for various agronomic traits. Jha et al., highlighted the significant advances made in developing genomic resources for underused legumes. These included genome-wide molecular markers, genome sequencing, genetic linkage maps, and trait mapping played a vital role in increasing legume production, which can contribute to global food security. Integrating genomic resources with unique breeding expertise and good seed system techniques can help increase the production of underused legumes and contribute to global food security.

In legume breeding, yield stability and adaptability are of prime importance, which requires yield trials at different geographic conditions to assess the impact of environmental factors. Pobkhunthod et al., identified KUP12BS029-1-1-3 large-seeded peanut genotype with significantly stable yield potential through multilocation yield trial using GGE biplot analysis in Thailand.

The seed size and shape are directly correlated to the overall quantity and quality of the lentil production. By controlling cell division via cell expansion and overall seed growth, Dutta et al., using a transcriptomic approach, demonstrated how essential genes, including kinases, transcription factors, cell wall-building enzymes, and hormone production pathways, are involved in lentil seed size regulation.

Bhat et al., identified 57 significant SNPs and six stable QTL regions using GWAS and haplotype alleles for improvement of yield and yield-related traits in soybean. In legume breeding, plant traits and soil attributes also play a role in achieving the desired production goals. While working on 797 soybean lines they indicated that the availability of soil texture information prior to the growing season might maximize the efficiency of a breeding program by allowing the reconsideration of experimental field design, allocation of resources, reduction of preliminary trials, and shortening of the breeding cycle (Vieira et al.).

Biotic and abiotic stresses are the main obstacles in achieving the desired goals of legume production

Legumes are susceptible to biotic stresses that can adversely affect their growth and production, leading to significant yield losses. Biotic stresses include diseases caused by pathogens such as fungi, bacteria, viruses, and nematodes, as well as pests like insects and mites. Kaur et al., introgressed cry1Ac from transgenic chickpea lines into commercial cultivars using Marker-Assisted Backcrossing (MABC) breeding for pod borer resistance, which caused 100% *H. armigera* larval mortality. Using MABC, Bharadwaj et al., have developed a high yielding *Fusarium* wilt resistant chickpea cultivar BGM20211 by gene pyramiding and released it as Pusa Manav/Pusa Chickpea 20211 for commercial cultivation. One of the most common diseases that severely influence soybean output worldwide is *Phytophthora* rot and stem rot (PRSR) caused by *Phytophthora sojae*. Chandra et al., highlighted the developments made in understanding the genetic basis of PRSR resistance, genomic developments, and prospective uses of PRSR resistance in soybean for long-term control. Another important pulse crop, peanut is affected by many soil borne diseases and pathogens and substantially reduces yield. Sharif et al., identified pericarp abundant promotor AhGLP17-1P. Such promotors could drive the expression of defence-related genes in the pericarp and improve disease tolerance. Bacterial wilt is one of the primary diseases that cause a substantial decline in the common bean production. Zia et al., evaluated 168 accessions for resistance against bacterial wilt and identified 14 single nucleotide polymorphism (SNP) markers associated with the bacterial wilt resistance that can be utilized in developing new ideotypes of common bean with improved tolerance to bacterial wilt. Agarwal et al., identified a major QTL, "qpsd4-1," on LG 4 and a minor QTL, "qpsd8-1," on LG8 that explained 41.8% and 4.5% of phenotypic variance, respectively, associated with resistance to *Pythium ultimum* in chickpea.

Besides biotic stress, legumes are susceptible to various abiotic stresses such as drought, salinity, heat, cold, and heavy metal toxicity that significantly impact their production. Yin et al., identified 72 basic leucine zipper (bZIP) family transcription factors in the adzuki bean and concluded that tissue-specific bZIP might play a role in conferring tolerance to abiotic stress such as drought, cold, salt, and heavy metal stress. Frost is an important abiotic stress that reduces production, destroys nitrogen-fixing bacteria, and reduces diet value of legumes. Sallam et al., evaluated 185 genotypes of winter faba bean for frost tolerance and identified two frost tolerant genotypes *viz.*, S_028 and S_220. Using KASP markers and GWAS genotypes two markers *viz.*, Vf_Mt1g072640_001 and Vf_Mt7g073970_001 showed pleiotropic effects on root fresh and dry mass in both the genotypes. The two markers can be used for isolating frost tolerant genotypes and the two genotypes may be used as a donor of alleles for improving frost tolerance. Besides frost, floods also cause a substantial reduction in agricultural production. Floods are more detrimental to legumes, especially in pigeon pea, causing substantial crop loss. Using *De-novo* transcriptome assembly Tyagi et al., tried to decode the flood tolerance and identify the candidate genes that could help develop climate-resilient pigeon pea genotypes. Thus, effectively managing abiotic stresses is crucial for sustainable legume production and food security.

Omics approaches in improving the nutritional profile and agronomic traits of legumes

Micronutrients are essential for human and plant growth and development, and legumes are a rich source of many important micronutrients. Some of the crucial micronutrients found in legumes include iron (Fe), zinc (Zn), copper (Cu), manganese (Mn), magnesium (Mg), calcium (Ca), and molybdenum (Mo). Any essential micronutrient deficiency impairs the correct operation of cellular systems and has several metabolic and physiological ramifications. Nazir et al., identified 113 SNPs through GBS associated with most of the seed micronutrients on

chromosome 3 and chromosome 11 in common bean, which showed significant phenotypic variance ranging from 13.50% to 21.74%. Baloch et al. reported that the DArT-3367607 marker on chromosome Pv03, among the six markers identified, showed the highest phenotypic variation (7.5%) with the significant association for seed Mg contents in Turkish common bean germplasm.

To address an escalating global food demand, it is essential to produce superior crop types with high yield, increased nutrition, disease, and insect resistance. In comparison to other crops, chickpea and other legume crops have much lower genetic gains due to their limited genetic base. We must quicken genetic gains—a cyclical process of finding new variations, applying selection, and fixing good traits—to fulfill future demand. Additionally, modern variations of *Cicer* must be infused with genetic diversity from landraces and wild *Cicer* species to sustain greater genetic gain for a longer period of time. Singh et al. reviewed the current status of the narrow chickpea genetic base and the scope of modern "Omics" technologies in breaking this bottleneck overcoming the yield limits and achieving higher genetic gains. Underutilized crops represent a treasure of genes that may be used to improve the agronomic traits of widely used legume crops. Verma et al., working on seed development of lesser-known pulse crops *viz.*, ricebean, reported 6,928 differentially expressed genes in bold and small-seeded genotypes and identified several genes for seed development related traits. Zhao et al., carried out a genome-wide association study (GWAS) of 178 peanut cultivars and reported several marker-trait associations and candidate genes associated with hundred seed weight, total number branches, and pod shape. High yielding genotypes are considered a promising way of achieving food security for a burgeoning population. Singh et al., evaluated a panel of 100 blackgram genotypes at two locations and identified 49 significant SNP associations representing 42 QTLs related to yield and yield attributing traits. Wang et al., while working on QTL analysis in soybean, reported genes associated with the regulation of symbiosis. The group reported 10 QTLs associated with type III effector NopAA in *Sinorhizobium fredii* HH103 which functions as glycosyl hydrolase and plays a critical role in nodulation. Such studies are important in gaining more insights into the underlying mechanism of nitrogen fixation, making nitrogen fixation, and alleviating the harmful effects of chemical fertilizers on human health and the environment. Kumari et al., evaluated 98 wild and cultivated *Vigna* accessions and identified marker-trait associations (MTAs) for traits such as days to first flowering, days to maturity plant height and hundred seed weight that may be utilized in *Vigna* improvement programs. These MTAs may also help gain insights into understanding the underlying mechanism controlling the expression of these traits in various *Vigna* species.

Recent advancements in gene editing, genome sequencing, and machine learning algorithms in the field of legume breeding

Even though conventional breeding strategies have contributed in improving thousands of crop cultivars, however, these approaches are laborious and tedious and are not enough to deliver the improved products at the required pace to sustain a growing world population. However, the huge advancements made in all fields of science, including plant breeding overcame all the limitations of conventional breeding approaches. One such example is gene editing technology in legumes reviewed by (Baloglu et al.). This review article provides details of the comparative governmental regulatory restrictions on gene-edited crops in European Union and United States of America. Besides gene editing, artificial intelligence is an emerging technology for crop improvement in analyzing big data in phenomics and genomics. Aasim et al. reported using machine learning models for predicting and optimizing the tissue culture protocols in common bean, a recalcitrant crop. The main limitations of linkage mapping and positional cloning in mapping genomic loci controlling agronomic traits are their low resolution, low-throughput, and time requirement. With recent developments in genomics and sequencing techniques, Bulk segregant analysis sequencing (BSA-seq) and its related approaches, *viz.*, quantitative trait locus (QTL)-seq, bulk segregant RNA-Seq (BSR)-seq, and MutMap, helped breeders in rapid identification of genetic loci/QTLs controlling agronomic traits at high resolution, accuracy, reduced time span, and in a high-throughput manner. Majeed et al., reviewed the BSA-seq and its related approaches in crop breeding, along with their merits and demerits in trait mapping. Overall genomic tools such as molecular markers, gene editing, and transcriptomics can be utilized to accelerate the breeding process and improve the efficiency of crop improvement programs. In conclusion, the recent advancements in genomics provide an opportunity to improve important crop legumes and develop more climate-resilient, high-yielding, and nutritionally rich cultivars. These advancements can contribute significantly to global food security and help meet the increasing demand for high-quality food.

Author contributions

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

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 January 2022
doi: 10.3389/fgene.2021.821281

Cloning and Functional Characterization of a Pericarp Abundant Expression Promoter (*AhGLP17-1P*) From Peanut (*Arachis hypogaea* L.)

*Yasir Sharif[1,2], Hua Chen[2,3], Ye Deng[1,2], Niaz Ali[2,3], Shahid Ali Khan[2,3], Chong Zhang[2,3], Wenping Xie[1,2], Kun Chen[1,2], Tiecheng Cai[2,3], Qiang Yang[2,3], Yuhui Zhuang[2], Ali Raza[2,3] and Weijian Zhuang[1,2,3]**

[1]College of Plant Protection, Fujian Agriculture and Forestry University (FAFU), Fuzhou, China, [2]Fujian Provincial Key Laboratory of Plant Molecular and Cell Biology, Oil Crops Research Institute, Fujian Agriculture and Forestry University (FAFU), Fuzhou, China, [3]Center of Legume Crop Genetics and Systems Biology, College of Agriculture, Fujian Agriculture and Forestry University (FAFU), Fuzhou, China

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Farhat Abbas,
South China Agricultural University, China
Muhammad Azhar Nadeem,
Sivas University of Science and Technology, Turkey

***Correspondence:**
Weijian Zhuang
weijianz@fafu.edu.cn

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

Received: 24 November 2021
Accepted: 31 December 2021
Published: 20 January 2022

Citation:
Sharif Y, Chen H, Deng Y, Ali N, Khan SA, Zhang C, Xie W, Chen K, Cai T, Yang Q, Zhuang Y, Raza A and Zhuang W (2022) Cloning and Functional Characterization of a Pericarp Abundant Expression Promoter (AhGLP17-1P) From Peanut (Arachis hypogaea L.). Front. Genet. 12:821281. doi: 10.3389/fgene.2021.821281

Peanut (*Arachis hypogaea* L.) is an important oil and food legume crop grown in tropical and subtropical areas of the world. As a geocarpic crop, it is affected by many soil-borne diseases and pathogens. The pericarp, an inedible part of the seed, acts as the first layer of defense against biotic and abiotic stresses. Pericarp promoters could drive the defense-related genes specific expression in pericarp for the defense application. Here, we identified a pericarp-abundant promoter (*AhGLP17-1P*) through microarray and transcriptome analysis. Besides the core promoter elements, several other important *cis*-elements were identified using online promoter analysis tools. Semiquantitative and qRT-PCR analyses validated that the *AhGLP17-1* gene was specifically expressed only in the pericarp, and no expression was detected in leaves, stem, roots, flowers, gynophore/peg, testa, and embryo in peanut. Transgenic *Arabidopsis* plants showed strong GUS expression in siliques, while GUS staining was almost absent in remaining tissues, including roots, seedlings, leaf, stem, flowers, cotyledons, embryo, and seed coat confirmed its peanut expressions. Quantitative expression of the *GUS* gene also supported the GUS staining results. The results strongly suggest that this promoter can drive foreign genes' expression in a pericarp-abundant manner. This is the first study on the functional characterization of the pericarp-abundant promoters in peanut. The results could provide practical significance to improve the resistance of peanut, and other crops for seed protection uses.

Keywords: cis-elements, GUS staining, pathogens, tissue-specific expression, transgenic arabidopsis

INTRODUCTION

Peanut (*Arachis hypogaea* L.), also known as groundnut, is an important legume crop (Fabaceae family) that is mainly cultivated for its edible seeds. Botanically peanut is a legume, but it is frequently eaten as a "nut", and its nutritional value is comparable to other nuts (Settaluri et al., 2012). It is a major nutrition source for humanity, providing edible oil and proteins (Khan et al., 2020) after soybean and rapeseeds. Peanut is full of nutrients, including carbohydrates, vitamins, lipids, minerals, etc. (Akhtar et al., 2014). Peanut is widely cultivated globally in tropical and subtropical regions, producing 48.75 million tons of shells annually. China is the leading producer with 17.52 million tons of shells and shares 36% of the Worlds' total production, followed by India (14%), Nigeria, Sudan, and the USA (FAOSTAT, 2019).

Like other crops, the peanut is affected by several bacterial, fungal, and viral pathogens (Ali et al., 2020). To manage these pathogens, nature has provided peanuts with several defense mechanisms. As peanut seeds are of primary importance, they have evolved to form non-edible outer coverings (pericarp and testa) to protect the seeds from insects, pathogens, and physical damage and maintain seed viability from generation to generation (Souza and Marcos-Filho, 2001). The shape, color, thickness, etc., of the seed coat, may differ among species during the evolutionary process and due to different growing environments. Biologically active chemicals in seed coats provide the right solution for infections, especially seed coat-based pathogen resistance attributed to hydrophobic molecules like lignin and tannins (Dardick and Callahan, 2014). Seed inner pericarp tissues also accumulate different antifungal and antibacterial metabolic by-products and flavonoids (Dixon and Paiva, 1995). Plenty of work is available on the genetic basis of seed coat development in the model dicot plant, i.e., *Arabidopsis*, and many genes have been worked out for their functions in seed coat development (Windsor et al., 2000; Haughn and Chaudhury, 2005; Dardick and Callahan, 2014). Similarly, detailed knowledge about the genes involved in seed coat development in peanut can help manipulate specific genes to get resistance against specific biotic or abiotic stresses. Paik-Ro and his team have reported that the cDNA of *PSC33I* is specifically expressed in peanut seed tissues without showing any expression in other tissues (Paik-Ro et al., 2002).

The physical appearance of any plant and its response to the surrounding environment is mainly controlled by a cascade of genes. Expression of any gene depends upon the binding of specific transcription factors (TFs) or proteins with unique upstream elements (*cis*-elements) and regulation at the transcription site. Promoters are the non-coding regulatory DNA sequences present upstream of a genes' coding region and contain specific *cis*-elements that regulate the spatio-temporal expression of a gene. A promoter has multiple binding sequences for TFs and RNA polymerase for activation and regulation of functions of the downstream gene (Ong and Corces, 2011; Gupta et al., 2012; Hernandez-Garcia and Finer, 2014). Promoters are important tools for molecular research to study the functions of a gene (Xu et al., 2016). Promoters can be divided into constitutive, spatio-temporal/tissue-specific, inducible, and synthetic promoters based on regulating the gene functions. Constitutive promoters drive the expression of a gene in all tissues, such as maize ubiquitin promoter (Christensen et al., 1992) and cauliflower mosaic virus promoter (CaMV35S) (Odell et al., 1985), are widely used for functional gene studies in plants.

Similarly, Figwort mosaic virus sub-genomic transcript promoter *FMV Sgt* is another example of constitutive promoters (Bhattacharyya et al., 2002). The constitutive expression of an exogenous gene can cause adverse effects on a plants' normal growth and functioning as it causes an extra burden on plant metabolism for expressing a gene in tissues where it is not required, and sometimes it results in undesired phenotypes (Nakashima et al., 2007). Tissue-specific and inducible promoters are more important as they can drive a genes' expression in a tissue-specific manner or under specific stress; hence, avoiding the adverse effects of constitutive expression. The selection of a suitable promoter is the key step for developing transgenic plants since it influences the cell, tissue, or stage specificity and determines the expression level of the transgene (Koyama et al., 2005). Therefore, comprehensive knowledge of promoter activity is a prerequisite for transgenic breeding.

In this paper, we studied the promoter of a peanut pericarp abundant expression gene belonging to the Germin-like protein (GLPs) family. Germin-like proteins were first identified in germinating wheat seedlings as germination specific markers during the 1980s (Lane et al., 1993). These proteins belong to a group of water-soluble glycoproteins with diverse functions. GLPs are ubiquitously found in gymnosperms, and angiosperms (Lu et al., 2010) and are broadly involved in plants' defense against biotic and abiotic stresses (Godfrey et al., 2007). Germin-like proteins are a group of "Cupin superfamily" containing the Cupin domain (PF00190). Structurally these proteins are composed of the β-sheet barrel (jellyroll beta-barrel structure), and the protein's C-terminus contains a metal ion binding site (Agarwal et al., 2009). GLPs contain two conserved motifs known as "germin-box". These motifs G(x)5HxH(x)3,4E(x)6G and G(x)5PxG(x)2H(x)3N are packed in a classic jellyroll beta-barrel structural domain (Yamahara et al., 1999). Thus far, there is no report available which investigated its specific expression and characterized its upstream promoter in a plant pericarp.

Here, this study identified and characterized a peanut *GLP* gene, "Germin-like protein subfamily 1 member 7 (*AhGLP17-1*)," showing abundant expression in seed pericarp. This gene was selected based on the transcriptome and microarray expression data, and its pericarp abundant expression was verified by semiquantitative and qRT-PCR. Further, the *cis*-regulatory elements of the promoter were analyzed, and their expression was characterized in transgenic *Arabidopsis* plants.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Seeds of peanut cultivars Minhua-6 (M-6), and Xinhuixiaoli (XHXL), and *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) were obtained from the Institute of Oil Crops, Fujian Agriculture and Forestry University, Fuzhou, China. Peanut plants (M-6) were

grown in research fields at Yangzhong, Sanming county of Fujian province during the summer season (April-August, 2018), and samples from different tissues, including leaves, stem, flower, peg, pericarp, testa, embryo, and roots were collected at different growing stages for semiquantitative and qRT-PCR-based confirmation of *AhGLP17-1* gene. For isolation of promoter, peanut plants (XHXL) were grown in the greenhouse in small pots (10 cm diameter) filled with compost, and leaf samples were collected from one-week-old plants. Arabidopsis plants were germinated on MS medium (Murashige and Skoog, 1962), transferred into 8-cm diameter plastic pots containing compost after 2 weeks, and further grown in the greenhouse, where 25°C temperature and 16/8 h photoperiod for day/night cycle was maintained for *Arabidopsis* seedlings. All samples were washed with 75% ethanol and sterilized water, packed in previously labeled plastic bags, and immediately frozen in liquid nitrogen and stored at −80°C for further use.

Selection and Verification of Pericarp-Abundant *AhGLP17-1* Gene

A pericarp abundant gene with Peanut Genome Resource (PGR) ID AH06G08990.1 was identified by the microarray and transcriptome expression data, which is available at the PGR database http://peanutgr.fafu.edu.cn/ (accessed on 20th April 2018) (Zhuang et al., 2019). This gene belongs to the germin-like protein family viz. "germin-like protein subfamily 1 member 7". Due to the polyploid nature of the peanut genome, peanut contains nine different copies of germin-like protein subfamily 1 member 7, and we named this gene *AhGLP17-1*. The protein 3D structure of *AhGLP17-1* was constructed using 3D Ligand Site https://www.wass-michaelislab.org/3dlig/index.html (accessed on 15th June 2018) (Wass et al., 2010), physio-chemical properties were predicted by Expasy server https://web.expasy.org/protparam/ (accessed on 20th June 2018) (Gasteiger et al., 2005), and subcellular localization was predicted by CELLO V2.5 http://cello.life.nctu.edu.tw/ (accessed on 22nd June 2018) (Yu et al., 2006). Gene functional annotation "gene ontology" information was retrieved from the PGR database (http://peanutgr.fafu.edu.cn/). Semiquantitative and qRT-PCR analysis were used to confirm its expression in different tissues.

Selection of Promoter Region and Identification of *Cis*-regulatory Elements

A 2296 bp region upstream of the start codon of the *AhGLP17-1* gene was selected for promoter analysis. The promoter was named *AhGLP17-1P*. *Cis*-regulatory elements of promoter region were predicted by PLACE https://www.dna.affrc.go.jp/PLACE/?action=newplace (accessed on 30th June 2018) (Higo et al., 1998) and PlantCARE databases http://bioinformatics.psb.ugent.be/webtools/plantcare/html/ (accessed on 30th June 2018) (Rombauts et al., 1999).

RNA Extraction, cDNA Synthesis, Semiquantitative, and qRT-PCR Analysis

RNA from different tissues (leaves, stem, flowers, peg, pericarp, testa, embryo, and roots) of peanut cultivar Minhua-6 was extracted by the Cetyl Trimethyl Ammonium Bromide (CTAB) (200 mM Tris HCL, 20 mM EDTA, 2% CTAB, 1.4 M NaCl, and 2% PVP-40, pH = 8.0, 0.2% β-Mercapto Ethanol added before use) method with some modification (Chen et al., 2016). First-strand cDNA was synthesized using 1 μg RNA with the PrimeScript 1st strand cDNA Synthesis Kit (Cat# 6110A) (Takara, Dalian, China) according to manufacturer guidelines. Semiquantitative PCR analysis was performed to check the expression of the *AhGLP17-1* gene in different tissues using the peanut *Actin* gene and gene-specific primers. The semiquantitative PCR results were viewed on 1% agarose gel. For qRT-PCR analysis, three different cDNA preparations for each tissue were used. Quantitative real-time polymerase chain reaction (qRT-PCR) was performed by MonAmp™ ChemoHS qPCR Mix (Cat# 160431) (Monad Biotech, Wuhan, China). The PCR reaction mixture contained 10 μL MonAmp Master Mix, 1 μL forward and reverse primer (diluted at 10 mM concentration), and 1 μL cDNA template. The peanut *Actin* gene (Chi et al., 2012) was used as an internal control to normalize the cDNA. The qRT-PCR reaction was performed using Applied Biosystems 7,500 real-time PCR system (ThermoFisher Scientific, USA). The cycling program was as follows: 94°C (1 min), 60°C (1 min), and 72°C (1 min) for 40 cycles. Primers used for semiquantitative and qRT-PCR analysis are given in (**Supplementary Table S1**).

DNA Extraction and Isolation of Promoter

Genomic DNA from peanut (XHXL variety) leaves was extracted using the CTAB (1 M Tris-HCl, 0.5 M Na-EDTA, 2% CTAB, 1.4 M NaCl, 3% PVP-40, and 0.2% β-Mercapto Ethanol added before use) method (Li et al., 2013) with some modifications. Promoter-specific primers (**Supplementary Table S1**) covering 2,296 bp upstream region before start codon were used to clone *AhGLP17-1P*. The promoter region was amplified with PrimeSTAR® Max DNA polymerase (Cat# Ro45B) (Takara, Dalian, China), according to manufacturers' guidelines. The PCR amplified product was visualized on 2% agarose gel and purified by TIANGEN Universal DNA Purification Kit (Cat# DP103-03) (Tiangen Biotech Beijing, China). Purified PCR products were sub-cloned into pMD19T vector (Cat# 3271) (Takara, Dalian, China) and sequenced by Beijing Genomics Institute (BGI, Shenzhen, China).

Vector Construction

The sequence-verified PCR clones were used to construct the expression vector using Two-step Gateway cloning. First, the promoter was amplified by primers containing universal overlapping sequences for gateway vectors (**Supplementary Table S1**) using the TA-cloning product as template and purified through TIANGEN Universal DNA Purification Kit (Cat# DP214-03) (Tiangen Biotech Beijing, China). The purified PCR products were ligated by Gateway BP reaction using BP Clonase enzyme (Cat# 11789020) (Invitrogen, ThermoFisher Scientific USA) into entry vector pDONR-207 between attP sites. Vector pDONR-207 containing *AhGLP17-1P* was transferred to *E. coli* (DH5α), and positive clones were selected for sequence confirmation. After sequence confirmation,

AhGLP17-1P was ligated into destination vector pMDC164 by Gateway LR reaction using LR Clonase enzyme (Cat# 11791020) (Invitrogen, ThermoFisher Scientific USA). Expression vector pMDC164 contains the hygromycin resistance gene for positive plants' selection and the *GUS* reporter gene. The vector was named pMDC164-*AhGLP17-1P*.

Transformation Into Arabidopsis

Expression vector pMDC164-*AhGLP17-1P* was transferred into *Agrobacterium tumefaciens* strain GV3101, and positive colonies were selected on yeast extract beef (YEB) selection medium plates supplemented with 50 μg ml^{-1} kanamycin. Positive *A. tumefaciens* cells harboring the expression vector were grown overnight at 28°C, 220 rpm to get logarithmic growth phase (OD600 = 1.0-1.5) in liquid YEB medium supplemented with kanamycin 50 μg ml^{-1}, and 75 μg ml^{-1} rifampicin. Bacterial cells were harvested by centrifugation at 4,000 rpm for 10 min and resuspended in 5% sucrose solution containing 0.02% Silwet L-77 and 100 μg ml^{-1} acetosyringone was also added to achieve higher transformation efficiency (Sheikholeslam and Weeks, 1987). Mature *Arabidopsis* plants were used for transformation by floral dip method (Clough and Bent, 1998) by dipping the unopened flowers into the prepared solution for 10–15 s, and then placed under dark for 24 h. Siliques and opened flowers were removed before the transformation, and the floral dipping was repeated after 5 days. After that, plants were grown under optimum growth conditions until seeds were ready to harvest, and finally, the T0 seeds were harvested.

Screening of Positive Transgenic Plants

To identify positively transformed plants, T0 transgenic seeds were screened on MS medium containing 50 μg ml^{-1} hygromycin. First, seeds were surface sterilized with 75% ethanol for 2 min and then treated with 10% H_2O_2 for 2 min, followed by 4–5 times washing with distilled water. Then seeds were spread over MS medium containing plates supplemented with Hygromycin antibiotic. Eight randomly selected hygromycin-resistant plants were verified by PCR amplification with promoter-specific forward and *GUS* gene specific reverse primers (**Supplementary Table S1**). The selected transgenic plants were covered with plastic sheets to avoid the chances of cross-pollination, and T1 transgenic seeds were obtained. Transgenic seeds were grown to further generations to get homozygous T3 lines.

Histochemical GUS Staining and Expression Analysis of *GUS* Gene

For the GUS staining assay (Jefferson et al., 1987), tissue samples were incubated in 2 mM 5-Bromo-4-chloro-3-indolyl β-D-glucuronide (X-Gluc) solution prepared in 0.1% Triton X-100, 50 mM sodium phosphate buffer, 10 mM EDTA, 2 mM potassium ferricyanide, and 2 mM potassium ferrocyanide. Plant samples were incubated in the above-prepared staining solution at 37°C for 12 h. After staining, samples were washed with 50, 75, and 100% ethanol for 5 min separately. Finally, samples were decolorized by incubating in 75% ethanol at 37°C until all green color was removed, while ethanol was changed every 4 hours. Samples were photographed with the digital camera and Olympus microscope Model BX3-CBH with attached Olympus DP80 digital camera (Olympus, Tokyo, Japan). Quantitative expression of the GUS gene in different tissues of transgenic *Arabidopsis* plants was analyzed by qRT-PCR with *GUS* gene-specific primers (**Supplementary Table S1**).

For pericarp sampling, young siliques were carefully opened by sharp needles and forceps, and seeds were removed. Siliques without seeds were stored for RNA extraction. Similarly, seeds were used to examine testa, cotyledon, and embryo expression. It was impossible to separate the testa from cotyledons and get enough samples for RNA extraction from Arabidopsis seeds, so whole seeds were used to extract the RNA. RNA from different tissues of transgenic Arabidopsis was extracted using TriQuick Reagent (Cat# R1100) (Solarbio, Beijing, China), following the manufacturers' instructions.

Cryostat Sectioning of Transgenic *Arabidopsis* Seeds

The seeds of transgenic Arabidopsis plants were ruptured and incubated in the GUS staining solution to confirm whether GUS staining is present in testa, cotyledons, and embryo or not. After overnight incubation in GUS solution, seeds were further processed for cryostat sectioning in Leica CM1950 Cryostat Microtome (Leica Biosystems, Germany). The cryostat microtome was turned on for 5 hours before use, and the temperature was set at −20°C. Specimen discs, brushes, and forceps were put inside the cooling chamber. The freezing compound was applied on specimen discs, and seeds were gently placed on specimen discs containing the freezing compound. Specimen discs containing the seeds were kept at −20°C for 30 min. After that, 50 μm sections were made and placed on glass slides. Later, images of sectioned specimens were taken by Olympus IX73 microscope with attached Olympus DP80 digital camera (Olympus, Tokyo, Japan).

RESULTS

Selection and Characterization of Pericarp Abundant Gene

We searched the Peanut Genome Resource database (http://peanutgr.fafu.edu.cn/) (Zhuang et al., 2019) for the candidate gene with high transcriptome and microarray expression in pericarp and with no or very low expression in other tissues for cloning of promoter. A member of the peanut germin-like protein family (*AhGLP*) named germin-like protein subfamily 1 member 7 (*AhGLP17-1*) with the PGR gene ID AH06G08990 and mRNA ID AH06G08990.1 was found to be specifically showing high expression in pericarp as compared to other tissues (**Figure 1A**). Although microarray expression was found in root tissues and gynophore/peg (**Supplementary Table S2**), some transcriptome expression was also present in roots and peg (**Supplementary Table S3**). Still, the tendency of expression of the *AhGLP17-1* gene was higher in the

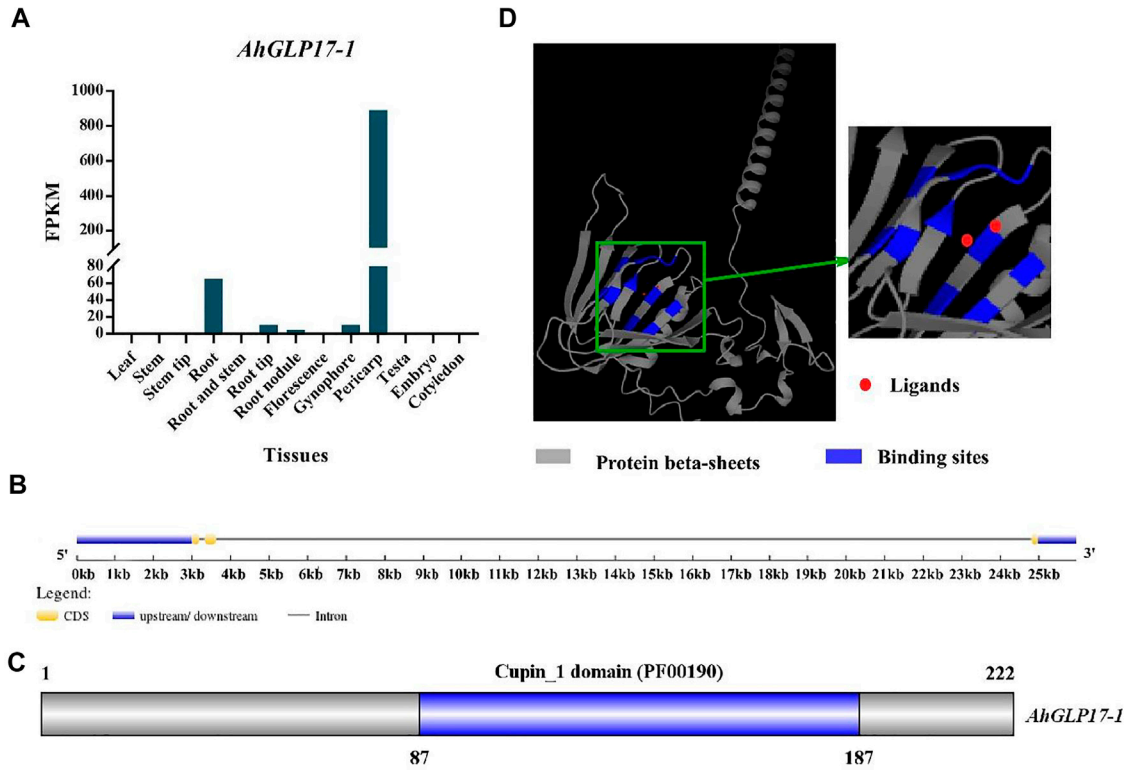


FIGURE 1 | Expression and characterization of *AhGLP17-1* gene. **(A)** Transcriptome expression (FPKM values) of the *AhGLP17-1* gene in different tissues of peanut (average values of the pericarp, testa, and embryo transcriptome expression are used). **(B)** Gene structure of *AhGLP17-1*. **(C)** The position of cupin_1 domain. **(D)**The 3D protein structure of *AhGLP17-1*. Where grey color shows the protein β-sheets, blue color shows binding sites, and red dots show ligands.

pericarp (**Figure 1A**). This gene is present on the 6th chromosome of sub-genome A at 12182490-12204462 position on the negative strand and has a CDS length of 666 base pairs and 21973 base pairs genomic length. The protein, CDS, and promoter sequences of the *AhGLP17-1* gene are given in **Supplementary file S1**. It consists of three exons of almost the same size and two introns, one of which is 21170 bp and the second intron is 134 bp long (**Figure 1B**). *In silico* subcellular localization showed that *AhGLP17-1* is localized in extracellular spaces and plasma membrane. Protein comprises 222 amino acid residues with a molecular weight of 24.49 KDa and a theoretical isoelectric point of 9.36. Further, it contains the Cupin_1 domain (PF00190) at 87-187 aa position (**Figure 1C**). The protein 3D structure prediction showed that *AhGLP17-1* is composed of the β-sheet barrel (jellyroll beta-barrel structure) with the ligands (**Figure 1D**). GO functional annotation revealed that the *AhGLP17-1* gene participates in three categories, including biological process in oxalate metabolic process (GO:0033609), molecular functions in nutrient reservoir activity (GO:0045735), oxalate decarboxylase activity (GO:0046564), manganese ion binding (GO:0030145), and cellular components apoplast (GO: 0048046), cell wall (GO:0005618), and extracellular region (GO: 0005576) (**Supplementary Table S4**). Other related information (orthologues in other plant species) is given in **Supplementary Table S5**.

Validation of Pericarp Abundant Expression by Semiquantitative and qRT-PCR

Pericarp-abundant expression of *AhGLP17-1* among different tissues was confirmed by semiquantitative PCR and qRT-PCR. The peanut *Actin* gene was used as an internal control for both semiquantitative and qRT-PCR analysis. The peanut actin gene showed a bright band in RNA samples of all tissues, while the *AhGLP17-1* gene showed a bright band in the pericarp samples, but no expression was detected in all other tissues (**Figure 2A**). Results of semiquantitative PCR showed that *AhGLP17-1* was preferentially expressed in the pericarp. On the other hand, the qRT-PCR results clearly showed that a high level of transcripts of *AhGLP17-1* was present in the pericarp with minute expression in all other tissues (**Figure 2B**). Both semiquantitative PCR and qRT-PCR results showed that the *AhGLP17-1* gene was specifically expressed in the pericarp. It showed very minute expression in all other tissues.

Analysis of *Cis*-regulatory Elements of *AhGLP17-1* Promoter

A 2296 bp upstream sequence of *AhGLP17P-1* contained the basic promoter elements, including the TATA box, the key element for precise transcription initiation (Grace et al., 2004),

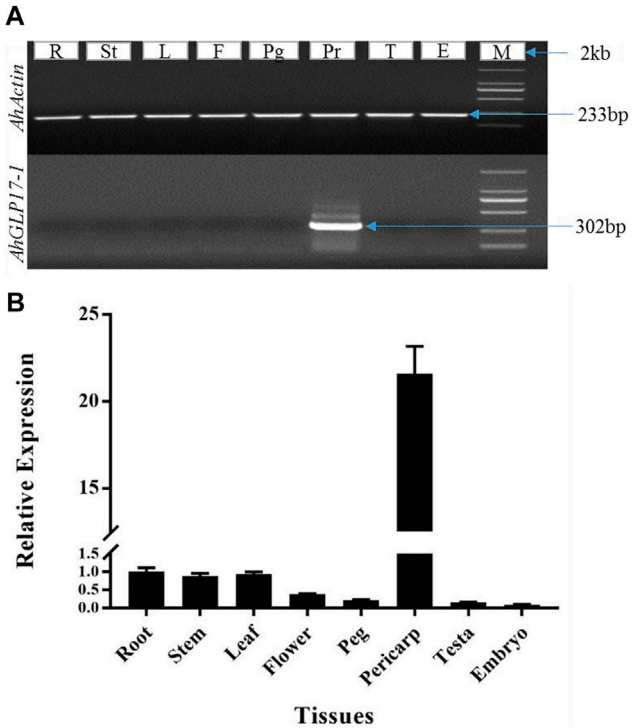


FIGURE 2 | Expression analysis of *AhGLP17-1* gene expression in different tissues. **(A)** Semiquantitaive PCR-based expression analysis, **(B)** qRT-PCR-based expression analysis of *AhGLP17-1* gene expression. Both semiquantitative and qRT-PCR results are consistent with transcriptome and microarray expression data (as shown in **Figure 1A**). L = leaf, St = stem, Fl = flower, Em = embryo, Ts = testa, peg = peg/gynophore, peri = pericarp, and R = root. Root expression was used as a control to analyze the data.

and the CAAT box required for tissue-specific activity (Shirsat et al., 1989). Many other important regulatory elements, including light-responsive elements (ATCT-motif, Box 4, G-Box, GA-motif, GATA-motif, GT1-motif, and Gap-box); hormones-responsive elements including salicylic acid (TCA-element), gibberellin (TATC-box), ethylene (ERE), and abscisic acid (ABRE) were also predicted in the *AhGLP17-1P*. Moreover, defense-related elements (TC-rich repeats and MYB binding sites), wound responsive element (WUN-motif), and anaerobic induction responsive element (ARE) were also found inside the promoter region. Further information on *cis*-regulatory elements and their position in *AhGLP17-1P* is presented in **Figure 3**.

Analysis of *cis*-elements by the PLACE database revealed the presence of a number of important elements, including seed-specific elements (RY-element) and transcription factor binding sites. The details of transcription factor binding sites and other elements are given in **Supplementary Table S6**. The presence of these regulatory elements strongly suggests that this promoter can be suitable substitute for a genes' native promoter. Except for these already reported *cis*-elements, some unknown elements were also found in the promoter region of the *AhGLP17-1* gene (**Figure 3**).

Cloning of Promoter, Vector Construction, and Transformation

A 2296bp region for *AhGLP17-1P* was PCR amplified (**Figure 4A**) from the DNA template of a high yielding and fungal pathogens resistant peanut variety XHXL (Khan et al., 2020) by promoter-specific primers (**Supplementary Table S1**). After confirmation of sequence, the amplified promoter fragment was again amplified with gateway primers containing gateway adapter sequences and then ligated into the *attP* sites of entry vector pDONR207 by Gateway BP-cloning reaction (**Figure 4B**). The sequence was confirmed again after BP-cloning, and the promoter fragment was ligated into the *attR* sites of expression vector pMDC164 by Gateway LR-cloning reaction. The resulting expression vector *pMDC164-AhGLP17-1P* (**Figure 4C**) was transformed to *Agrobacterium tumefaciens* competent cells by heat shock method. Positive Agrobacterium colonies harboring the expression vector were selected on selection medium (YEB plates containing kanamycin and rifampicin antibiotics). Positive agrobacterium colonies were used to transform the *Arabidopsis* plants through the floral dip method, and hygromycin resistant T0 transgenic plants were screened on MS plates containing 50 $\mu g\,ml^{-1}$ hygromycin antibiotic. Non-transformed plants turned yellow on hygromycin selection medium, while positively transformed plants were dark green and healthy, and these plants were transplanted into plastic pots containing compost. These hygromycin-resistant plants were also verified by PCR amplification.

Characterization of the Promoter in Transgenic Plants

Hygromycin resistant positive transgenic plants were confirmed by PCR amplification using DNA as the template with promoter-specific forward and *GUS* gene specific reverse primers. While *Arabidopsis* Col-0 plants were used as a negative control, and Gateway LR constructs were used as a positive control for PCR confirmation. Eight hygromycin-resistant plants were confirmed by PCR amplification (**Figure 5A**). Seeds of eight positively transformed plants were sown to get the T1 generation. In T1 generation again, eight plants were selected based on hygromycin resistance and PCR confirmation. Eight selected plants of the T1 generation were covered to avoid cross-pollination, and in this way, homozygous T3 generation was obtained. Histochemical GUS expression was checked in different tissues at different growth stages in the T3 generation.

The quantitative expression of the *GUS* gene was analyzed in different tissues of transgenic plants by qRT-PCR (**Figure 5B**). qRT-PCR results showed a relatively higher transcript level of the *GUS* gene in the pericarp of transgenic plants and a very low transcript level in other tissues (**Figure 5B**). GUS staining was not detected in all vegetative tissues and young seedlings (**Figure 6**). Among reproductive organs, a moderate level of GUS staining was present in siliques, and mild staining was also present in flowers in some cases; staining was not present in seeds. To confirm that GUS staining is not present in seed coat/testa, cotyledons, and embryo, the ruptured seeds incubated in GUS staining solution were further processed for cryostat sectioning. Staining was not found in any of the seed tissues (**Figure 7**). Non-transformed *Arabidopsis* Col-0



FIGURE 3 | Sequence analysis of *AhGLP17-1P* promoter. Presence of *cis*-elements in promoter sequences predicted by the PlantCARE database.

plants were used as a control to compare the GUS staining. Staining results showed dark blue color only in the pericarp (outer covering of siliques). In all other tissues, staining was not present except a minute staining in flowers in rare cases. These results clearly showed that *AhGLP17-1P* is abundantly expressed in the pericarp and almost no expression in other tissues. Overall, the results strongly suggest that this promoter is a suitable candidate to guide the expression of a gene in a pericarp-specific manner.

DISCUSSION

Constitutive expression of a gene in transgenic plants results in an additional metabolic burden on the plant system, and constitutive promoters can produce undesired phenotypes (Yuan et al., 2019b) and reduced production. Plants need to direct the valuable resources to the target areas for survival and smooth growth under normal and stressed conditions (Divya et al., 2019). Therefore, tissue-specific or stress-inducible promoters are ideal for altering the plants' genetic architecture to perform better according to a researchers' desired ideotype. Previous crop biofortification programs that resulted in present-day purple embryo maize (Liu et al., 2018), purple endosperm rice (Zhu et al., 2017), golden rice (Paine et al., 2005), and *Brassica juncea* for fish oil docosahexaenoic acid (DHA) production (Wu et al., 2005) were carried out to introduce new metabolic pathways in endosperm and seeds of these crop species by employing endosperm and seed-specific

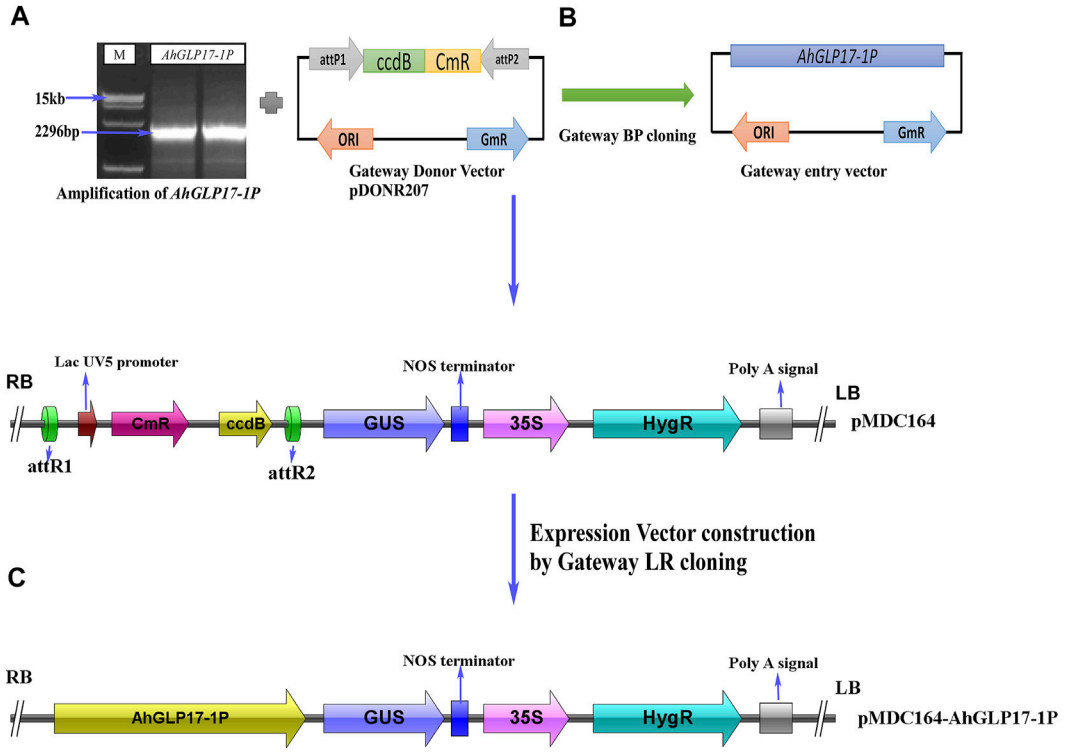


FIGURE 4 | Construction of vectors using the backbone of pMDC164 vector by Gateway cloning. **(A)** Amplification of *AhGLP17-1P* promoter, **(B)** construction of Gateway entry vector by Gateway BP-cloning, **(C)** construction of Gateway expression vector using the backbone of binary vector pMDC164.

promoters. For example, a strong endosperm-specific rice glutenin *GluT01* promoter (*Glu*) was used to drive a novel rice phytoene synthase (*psy*) gene and *Erwinia uredovora crtI* gene fused with pea Rubisco small subunit plastid peptide to produce the high amount of β-carotene in Golden Rice (Paine et al., 2005). As peanut is an important oil and protein providing crop and primary source of nutrition in many Asian and African countries, it is prone to many biotic and abiotic stresses (Zhang et al., 2017). Changing its genetic makeup is key for its better performance under stressed conditions. The pericarp is a non-edible part of peanut seeds and serves as the first layer of defense against pathogens. Using pericarp-specific promoters to drive resistance-related genes is ideal for improving its fighting ability against soil and seed-borne pathogens and diseases. Although, in recent years, some studies have reported the identification and functional characterization of seed-specific promoters and genes in peanut (Yuan et al., 2019a; Yuan et al., 2019b; Tang et al., 2021), but no detail is viable for pericarp-specific promoters.

Therefore, the current study was based on identifying and functionally characterizing the promoter of a gene with unique expression in peanut pericarp and no/minimum expression in all other tissues. We identified a pericarp-specific gene (germin-like protein subfamily 1 member 7) by available microarray and transcriptome expression data. Germin-like proteins GLPs are a group of well-known proteins ubiquitously found in the plant kingdom. GLPs are "cupin superfamily" domain-containing proteins, composed of β-sheet barrel structure and metal ion binding site at their C-terminus (Dunwell et al., 2008; Agarwal et al., 2009). GLPs actively participate in plant defense against various fungal, bacterial, and viral pathogens (Godfrey et al., 2007; Knecht et al., 2010; Guevara-Olvera et al., 2012). We checked the expression specificity of the *AhGLP17-1* gene in different tissues by semiquantitative and qRT-PCR using a widely grown peanut cultivar, "Minhua-6". Further, we cloned the promoter region of the *AhGLP17-1* gene from a high-yielding and disease-resistant cultivar, "Xinhuxiaoli". We used two different peanut varieties as Minhua-6 is a largely cultivated variety, and our microarray expression is based on this cultivar. Its samples were easily available for RNA extraction and expression verification. At the same time, Xinhuixiaoli cultivar (disease-resistant; fungal and bacterial diseases), was used for cloning of promoter based on its possible future use to develop transgenic peanut for disease resistance. Online databases for promoter analysis predicted many *cis*-acting elements, including core promoter element as TATA Box and proximal control elements as CAAT Box, GC Box (Muthusamy et al., 2017), and many other light-responsive, hormones responsive, growth and regulation responsive, and stress-responsive elements. The number and types of *cis*-regulatory elements are important determinants of promoter strength and specificity (Stålberg et al., 1993; Tang et al., 2015). RY-repeat elements, known for seed-specific expression (Fujiwara and Beachy, 1994), were also present in the promoter region of the

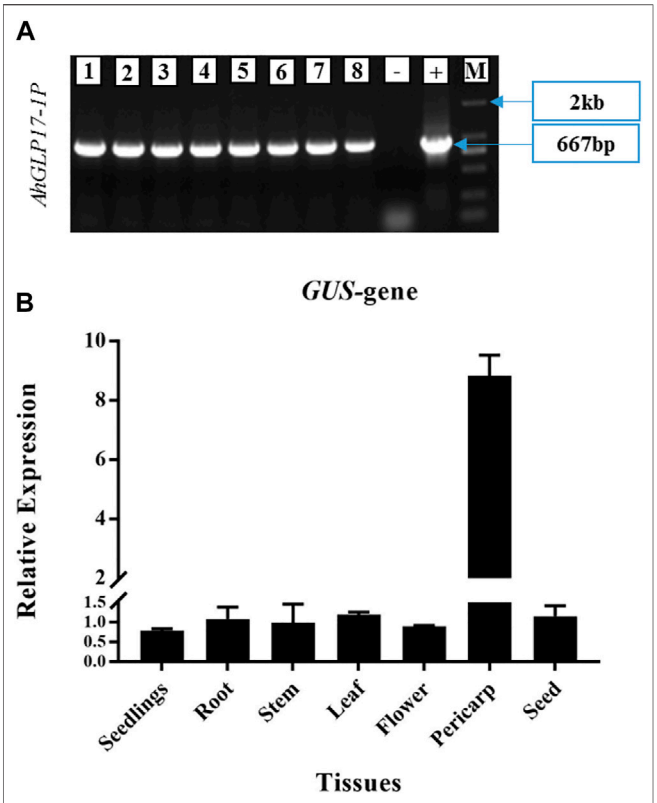


FIGURE 5 | (A) Confirmation of T0 transgenic *Arabidopsis* plants transformed with *AhGLP17-1P* (667 bp fragment). Eight hygromycin-resistant plants verified by PCR amplification with promoter-specific forward and *GUS* gene specific reverse primer. *Arabidopsis* Col-0 was used as –ve control, and Gateway LR constructs were used as + ve control for PCR verification. M shows 2 kb marker, **(B)** Quantitative expression of *GUS* gene driven by *AhGLP17-1P* in different tissues of transgenic *Arabidopsis* plants. Root expression was used as control to analyse the data.

AhGLP17-1 gene. One copy of (CA)n element "CNAACAC" was also found in the promoter region. (CA)n is known to be involved in the embryo and endosperm-specific transcription (Ellerström et al., 1996). But, to date, not a single element has been reported to be involved in the pericarp-specific expression. From *in silico* analysis of *AhGLP17-1P*, some unnamed elements were found that include Unnamed_1 (GAATTTAATTAA), Unnamed_2 (AACCTAACCT), Unnamed_4 (CTCC), and Unnamed_6 (taTAAATATct). There is a possibility that some of these elements or some other element have role in pericarp-specific activity.

Peanut is a suitable crop for genomic studies, but the main bottleneck is difficulties in peanut transformation, so *Arabidopsis* becomes an ideal alternative for functional studies of genes and promoters (Grace et al., 2004; Zavallo et al., 2010; Sunkara et al., 2014). Here, we verified the pericarp-abundant behavior of a germin-like protein family gene (*AhGLP17-1*) of peanut. Peanut is affected by several biotic and abiotic stresses, and the pericarp is the primary defense organ that protects peanut seeds from stresses and harsh conditions. Hence, altering the composition of this organ can result in new peanut cultivars with enhanced defense capabilities. If stress-related genes under the control of pericarp-specific promoters are successfully transformed into peanuts, it will be a milestone achievement in peanut breeding.

Pod rot is a complex disease associated with several *Pythium species,* deteriorates young pods and seeds (Yu et al., 2019). *Aspergillus flavus* is a serious threat to food safety which causes crop yield and quality deterioration by producing aflatoxins (Deng et al., 2018). Similarly, gray mold disease of peanut caused by *Botrytis cinerea* (Alam et al., 2019), web blight disease of groundnut caused by *Rhizoctonia solani* (Ganesan and Sekar, 2011), and a huge seed-borne fungal microflora attack peanut seeds and deteriorates yield and quality. Their first target is the pericarp or pod; after that, these pathogens invade edible seeds. Changing the genetic makeup of peanut pods is an ideal solution to avoid the damages of these pathogens. Chitinases, stilbene/resveratrol synthase, glucanases are well-known genes showing resistance to bacterial and fungal diseases (Medeiros et al., 2018; Vestergaard and Ingmer, 2019; Loc et al., 2020; Ueki et al., 2020). These genes can be derived by pericarp abundant/specific promoter to show high expression in pericarp tissues. In

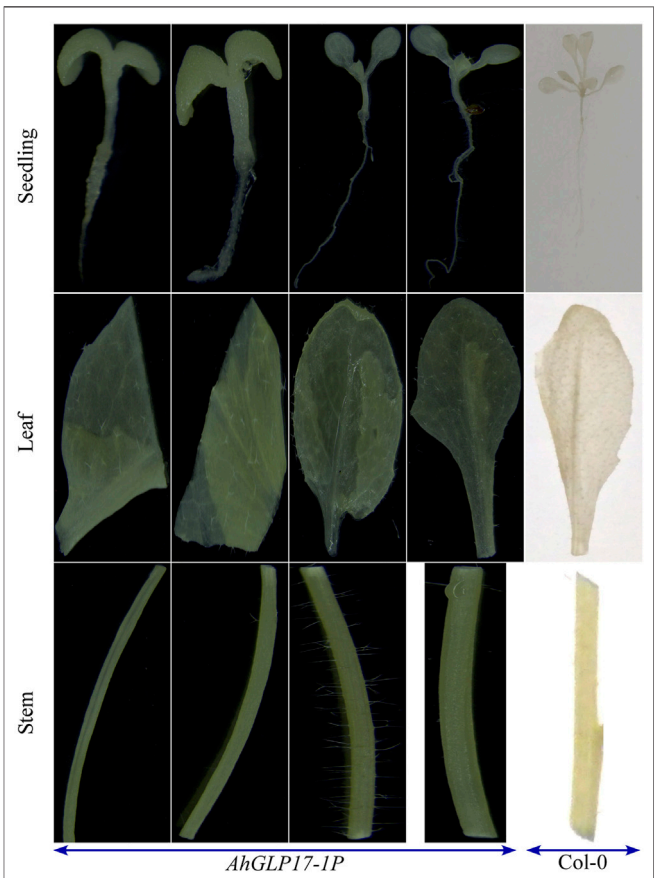


FIGURE 6 | GUS staining of different vegetative tissues of *Arabidopsis* transgenic plants. *AhGLP17-1P* plants showed no staining in any vegetative tissue (seedlings, roots, leaf, and stem). Different vegetative tissues of wild-type (Col-0) plants were also used for GUS staining to compare the results.

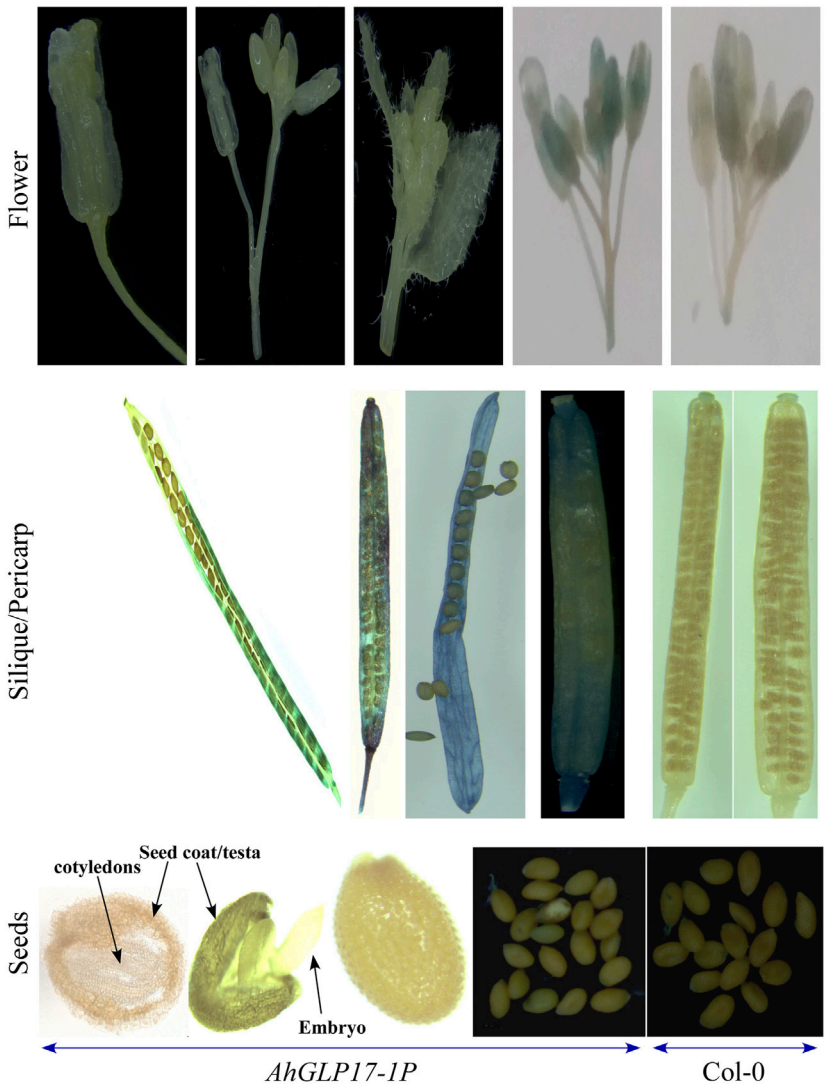


FIGURE 7 | GUS staining of different reproductive tissues/organs of transgenic *Arabidopsis* plants. *AhGLP17-1P* plants showed no staining in flowers (a minute staining in some cases). Seed outer covering (pericarp) showed good staining. While staining was not present in any seed tissue (testa, cotyledons, and embryo).

this way, the defensive ability of this tissue can be enhanced to protect the edible seeds.

Similarly, there are several other seed-borne diseases and pathogens that attack the growing peanut kernels. The pericarp is a potential barrier against these diseases and pathogens (Cobos et al., 2018). Transformation of these genes under the control of pod-specific promoters into high-yielding varieties can improve their disease resistance. Our results showed that *the AhGLP17-1* promoter showed expression specificity in pericarp tissues. Although there are some variations in expression patterns like in peanut, this gene also showed some expression in roots and no expression in flowers. But in transgenic Arabidopsis plants, the *GUS* gene under *AhGLP17-1P* showed some staining in flowers in some cases and no staining in roots. These variations are possibly attributed to diverse species. We are fully convinced that the *AhGLP17-1P* promoter investigated in this study could potentially drive the resistance-related genes in pericarp specific manner and alter the peanut pericarp genetic architecture to protect the edible seed from biotic stresses and environmental stresses.

CONCLUSION

In this study, we cloned and functionally characterized a novel pericarp-specific promoter (*AhGLP17-1P*) of peanut for the first time. This specifically expressed gene was cloned based on the microarray and transcriptome expression data. Both semiquantitative and qRT-PCR confirmed its pericarp-specific and abundant expressions. The GUS staining and qPCR analysis of the *GUS* gene under *AhGLP17-1P* in different vegetative and

reproductive tissues/organs of transgenic *Arabidopsis* plants clearly showed its expression in pericarp tissues and no expression in all other tissues including, roots, seedlings, stem, leaf, seeds, except minute expression in flowers in some cases. Our studied promoter can potentially improve disease/pathogen resistance in transgenic peanuts and other agronomically important crops by employing resistance-related genes.

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.

AUTHOR CONTRIBUTIONS

YS, HC, and WZ conceived the idea. YS analyzed the data and wrote the manuscript. YD, NA, SK, CZ, WX, KC, TC, QY, YZ, and AR helped in literature search, data analysis, and provided technical guidance. HC, CZ, and WZ supervised the work and edited the final version. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by grants from the National Science Foundation (NSF) of China (U1705233, 31601337, and 32072103), the Science and Technology Foundation of Fujian Province of China (2017N0006, 2018N0004, and 2021N5007), and Fujian Agriculture and Forestry University, Fuzhou, China.

ACKNOWLEDGMENTS

YS wants to say special thanks to Fujian Agriculture and Forestry University, and Chinese Government Scholarship Council, China, for providing Ph.D. scholarship and research environment. Further, we are also thankful to all lab/institute members for their support and guidance in conducting this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.821281/full#supplementary-material

REFERENCES

Agarwal, G., Rajavel, M., Gopal, B., and Srinivasan, N. (2009). Structure-based Phylogeny as a Diagnostic for Functional Characterization of Proteins with a Cupin Fold. *PLoS ONE* 4, e5736. doi:10.1371/journal.pone.0005736

Akhtar, S., Khalid, N., Ahmed, I., Shahzad, A., and Suleria, H. A. R. (2014). Physicochemical Characteristics, Functional Properties, and Nutritional Benefits of Peanut Oil: a Review. *Crit. Rev. Food Sci. Nutr.* 54, 1562–1575. doi:10.1080/10408398.2011.644353

Alam, A., Gupta, S., and Sharma, V. J. M. (2019). *Biochemical Screening of Ten Peanut Cultivars for Tolerance to Botrytis Cinerea*, 15.

Ali, N., Chen, H., Zhang, C., Khan, S. A., Gandeka, M., Xie, D., et al. (2020). Ectopic Expression of AhGLK1b (GOLDEN2-like Transcription Factor) in Arabidopsis Confers Dual Resistance to Fungal and Bacterial Pathogens. *Genes* 11, 343. doi:10.3390/genes11030343

Bhattacharyya, S., Dey, N., and Maiti, I. B. (2002). Analysis of Cis-Sequence of Subgenomic Transcript Promoter from the Figwort Mosaic Virus and Comparison of Promoter Activity with the Cauliflower Mosaic Virus Promoters in Monocot and Dicot Cells. *Virus. Res.* 90, 47–62. doi:10.1016/s0166-0934(02)00146-5

Chen, H., Zhang, C., Cai, T. C., Deng, Y., Zhou, S., Zheng, Y., et al. (2016). Identification of Low Ca2+stress-Induced Embryo Apoptosis Response Genes inArachis Hypogaeaby SSH-Associated Library Lift (SSHaLL). *Plant Biotechnol. J.* 14, 682–698. doi:10.1111/pbi.12415

Chi, X., Hu, R., Yang, Q., Zhang, X., Pan, L., Chen, N., et al. (2012). Validation of Reference Genes for Gene Expression Studies in Peanut by Quantitative Real-Time RT-PCR. *Mol. Genet. Genomics* 287, 167–176. doi:10.1007/s00438-011-0665-5

Christensen, A. H., Sharrock, R. A., and Quail, P. H. (1992). Maize Polyubiquitin Genes: Structure, thermal Perturbation of Expression and Transcript Splicing, and Promoter Activity Following Transfer to Protoplasts by Electroporation. *Plant Mol. Biol.* 18, 675–689. doi:10.1007/bf00020010

Clough, S. J., and Bent, A. F. (1998). Floral Dip: a Simplified Method forAgrobacterium-Mediated Transformation ofArabidopsis Thaliana. *Plant J.* 16, 735–743. doi:10.1046/j.1365-313x.1998.00343.x

Cobos, C. J., Tengey, T. K., Balasubramanian, V. K., Williams, L. D., Sudini, H. K., Varshney, R. K., et al. (2018). *Employing Peanut Seed Coat Cell Wall Mediated Resistance against Aspergillus flavus Infection and Aflatoxin Contamination*.

Dardick, C., and Callahan, A. M. (2014). Evolution of the Fruit Endocarp: Molecular Mechanisms Underlying Adaptations in Seed protection and Dispersal Strategies. *Front. Plant Sci.* 5, 284. doi:10.3389/fpls.2014.00284

Deng, Y., Chen, H., Zhang, C., Cai, T., Zhang, B., Zhou, S., et al. (2018). Evolution and Characterisation of the AhRAF4 NB-ARC Gene Family Induced by Aspergillus flavus Inoculation and Abiotic Stresses in Peanut. *Plant Biol. J.* 20, 737–750. doi:10.1111/plb.12726

Divya, K., Kavi Kishor, P. B., Bhatnagar-Mathur, P., Singam, P., Sharma, K. K., Vadez, V., et al. (2019). Isolation and Functional Characterization of Three Abiotic Stress-Inducible (Apx, Dhn and Hsc70) Promoters from Pearl Millet (Pennisetum Glaucum L.). *Mol. Biol. Rep.* 46, 6039–6052. doi:10.1007/s11033-019-05039-4

Dixon, R. A., and Paiva, N. L. (1995). Stress-induced Phenylpropanoid Metabolism. *The plant cell* 7, 1085. doi:10.2307/3870059

Dunwell, J. M., Gibbings, J. G., Mahmood, T., and Saqlan Naqvi, S. M. (2008). Germin and Germin-like Proteins: Evolution, Structure, and Function. *Crit. Rev. Plant Sci.* 27, 342–375. doi:10.1080/07352680802333938

Ellerström, M., Stålberg, K., Ezcurra, I., and Rask, L. J. P. m. b. (1996). Functional Dissection of a Napin Gene Promoter: Identification of Promoter Elements Required for Embryo and Endosperm-specific Transcription. 32, 1019–1027.

FAOSTAT (2019). *Food and Agriculture Organization of the United Nations*. FAO.

Fujiwara, T., and Beachy, R. N. (1994). Tissue-specific and Temporal Regulation of a ?-conglycinin Gene: Roles of the RY Repeat and Other Cis-Acting Elements. *Plant Mol. Biol.* 24, 261–272. doi:10.1007/bf00020166

Ganesan, S., and Sekar, R. J. P. I. T. M. W. (2011). *Screening of Biocontrol Agents against Rhizoctonia solani Causing Web Blight Disease of Groundnut*. Arachis hypogaea L., 127–140.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., et al. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*, 571–607. doi:10.1385/1-59259-890-0:571Protein Identification and Analysis Tools on the ExPASy Server

Godfrey, D., Able, A. J., and Dry, I. B. (2007). Induction of a grapevine Germin-like Protein (VvGLP3) Gene Is Closely Linked to the Site of Erysiphe Necator Infection: a Possible Role in Defense. *Mpmi* 20, 1112–1125. doi:10.1094/mpmi-20-9-1112

Grace, M. L., Chandrasekharan, M. B., Hall, T. C., and Crowe, A. J. (2004). Sequence and Spacing of TATA Box Elements Are Critical for Accurate Initiation from the β-Phaseolin Promoter. *J. Biol. Chem.* 279, 8102–8110. doi:10.1074/jbc.m309376200

Guevara-Olvera, L., Ruíz-Nito, M. L., Rangel-Cano, R. M., Torres-Pacheco, I., Rivera-Bustamante, R. F., Muñoz-Sánchez, C. I., et al. (2012). Expression of a Germin-like Protein Gene (CchGLP) from a Geminivirus-Resistant Pepper (Capsicum Chinense Jacq.) Enhances Tolerance to Geminivirus Infection in Transgenic Tobacco. *Physiol. Mol. Plant Pathol.* 78, 45–50. doi:10.1016/j.pmpp.2012.01.005

Gupta, N. C., Jain, P. K., Bhat, S. R., and Srinivasan, R. (2012). Upstream Sequence of Fatty Acyl-CoA Reductase (FAR6) of *Arabidopsis thaliana* Drives Wound-Inducible and Stem-specific Expression. *Plant Cel Rep* 31, 839–850. doi:10.1007/s00299-011-1205-9

Haughn, G., and Chaudhury, A. (2005). Genetic Analysis of Seed Coat Development in Arabidopsis. *Trends Plant Science* 10, 472–477. doi:10.1016/j.tplants.2005.08.005

Hernandez-Garcia, C. M., and Finer, J. J. (2014). Identification and Validation of Promoters and Cis-Acting Regulatory Elements. *Plant Sci.* 217-218, 109–119. doi:10.1016/j.plantsci.2013.12.007

Higo, K., Ugawa, Y., Iwamoto, M., and Higo, H. (1998). PLACE: a Database of Plant Cis-Acting Regulatory DNA Elements. *Nucleic Acids Res.* 26, 358–359. doi:10.1093/nar/26.1.358

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS Fusions: Beta-Glucuronidase as a Sensitive and Versatile Gene Fusion Marker in Higher Plants. *EMBO J.* 6, 3901–3907. doi:10.1002/j.1460-2075.1987.tb02730.x

Khan, S. A., Chen, H., Deng, Y., Chen, Y., Zhang, C., Cai, T., et al. (2020). High-density SNP Map Facilitates fine Mapping of QTLs and Candidate Genes Discovery for Aspergillus flavus Resistance in Peanut (*Arachis hypogaea*). *Theor. Appl. Genet.* 133, 2239–2257. doi:10.1007/s00122-020-03594-0

Knecht, K., Seyffarth, M., Desel, C., Thurau, T., Sherameti, I., Lou, B., et al. (2010). Expression of BvGLP-1 Encoding a Germin-like Protein from Sugar Beet in *Arabidopsis thaliana* Leads to Resistance against Phytopathogenic Fungi. *Mpmi* 23, 446–457. doi:10.1094/mpmi-23-4-0446

Koyama, T., Ono, T., Shimizu, M., Jinbo, T., Mizuno, R., Tomita, K., et al. (2005). Promoter of *Arabidopsis thaliana* Phosphate Transporter Gene Drives Root-specific Expression of Transgene in rice. *J. Biosci. Bioeng.* 99, 38–42. doi:10.1263/jbb.99.38

Lane, B. G., Dunwell, J. M., Ray, J. A., Schmitt, M. R., and Cuming, A. C. (1993). Germin, a Protein Marker of Early Plant Development, Is an Oxalate Oxidase. *J. Biol. Chem.* 268, 12239–12242. doi:10.1016/s0021-9258(18)31377-2

Li, J., Wang, S., Yu, J., Wang, L., and Zhou, S. J. C. B. o. B. (2013). A Modified CTAB Protocol for Plant DNA Extraction. 48, 72.

Liu, X., Yang, W., Mu, B., Li, S., Li, Y., Zhou, X., et al. (2018). Engineering of 'Purple Embryo Maize' with a Multigene Expression System Derived from a Bidirectional Promoter and Self-Cleaving 2A Peptides. *Plant Biotechnol. J.* 16, 1107–1109. doi:10.1111/pbi.12883

Loc, N. H., Huy, N. D., Quang, H. T., Lan, T. T., and Thu Ha, T. T. (2020). Characterisation and Antifungal Activity of Extracellular Chitinase from a Biocontrol Fungus, Trichoderma Asperellum PQ34. *Mycology* 11, 38–48. doi:10.1080/21501203.2019.1703839

Lu, M., Han, Y. P., Gao, J. G., Wang, X. J., and Li, W. B. (2010). Identification and Analysis of the Germin-like Gene Family in Soybean. *BMC Genomics* 11, 620–715. doi:10.1186/1471-2164-11-620

Medeiros, S. C. d., Monteiro-Júnior, J. E., Sales, G. W. P., Grangeiro, T. B., and Nogueira, N. A. P. (2018). *Chitinases as Antibacterial Proteins: A Systematic Review*.

Murashige, T., and Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant* 15, 473–497. doi:10.1111/j.1399-3054.1962.tb08052.x

Muthusamy, S., Sivalingam, P., Sridhar, J., Singh, D., Haldhar, S., and Kaushal, P. (2017). Biotic Stress Inducible Promoters in Crop Plants-A Review. *Jae* 04, 14–24. doi:10.53911/jae.2017.4202

Nakashima, K., Tran, L.-S. P., Van Nguyen, D., Fujita, M., Maruyama, K., Todaka, D., et al. (2007). Functional Analysis of a NAC-type Transcription Factor OsNAC6 Involved in Abiotic and Biotic Stress-Responsive Gene Expression in rice. *Plant J.* 51, 617–630. doi:10.1111/j.1365-313x.2007.03168.x

Odell, J. T., Nagy, F., and Chua, N.-H. (1985). Identification of DNA Sequences Required for Activity of the Cauliflower Mosaic Virus 35S Promoter. *Nature* 313, 810–812. doi:10.1038/313810a0

Ong, C.-T., and Corces, V. G. (2011). Enhancer Function: New Insights into the Regulation of Tissue-specific Gene Expression. *Nat. Rev. Genet.* 12, 283–293. doi:10.1038/nrg2957

Paik-Ro, O. G., Seib, J. C., and Smith, R. L. (2002). Seed-specific, Developmentally Regulated Genes of Peanut. *Theor. Appl. Genet.* 104, 236–240. doi:10.1007/s001220100763

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., et al. (2005). Improving the Nutritional Value of Golden Rice through Increased Pro-vitamin A Content. *Nat. Biotechnol.* 23, 482–487. doi:10.1038/nbt1082

Rombauts, S., Dehais, P., Van Montagu, M., and Rouze, P. (1999). PlantCARE, a Plant Cis-Acting Regulatory Element Database. *Nucleic Acids Res.* 27, 295–296. doi:10.1093/nar/27.1.295

Settaluri, V., Kandala, C., Puppala, N., and Sundaram, J. (2012). *Peanuts and Their Nutritional Aspects—A Review*.

Sheikholeslam, S. N., and Weeks, D. P. (1987). Acetosyringone Promotes High Efficiency Transformation of *Arabidopsis thaliana* Explants by Agrobacterium Tumefaciens. *Plant Mol. Biol.* 8, 291–298. doi:10.1007/bf00021308

Shirsat, A., Wilford, N., Croy, R., and Boulter, D. (1989). Sequences Responsible for the Tissue Specific Promoter Activity of a Pea Legumin Gene in Tobacco. *Mol. Gen. Genet.* 215, 326–331. doi:10.1007/bf00339737

Souza, F. H. D. D., and Marcos-Filho, J. (2001). The Seed Coat as a Modulator of Seed-Environment Relationships in Fabaceae. *Rev. Bras. Bot.* 24, 365–375. doi:10.1590/s0100-84042001000400002

Stålberg, K., Ellerström, M., Josefsson, L.-G., and Rask, L. (1993). Deletion Analysis of a 2S Seed Storage Protein Promoter of Brassica Napus in Transgenic Tobacco. *Plant Mol. Biol.* 23, 671–683.

Sunkara, S., Bhatnagar-Mathur, P., and Sharma, K. K. (2014). Isolation and Functional Characterization of a Novel Seed-specific Promoter Region from Peanut. *Appl. Biochem. Biotechnol.* 172, 325–339. doi:10.1007/s12010-013-0482-x

Tang, G., Xu, P., Li, P., Zhu, J., Chen, G., Shan, L., et al. (2021). Cloning and Functional Characterization of Seed-specific LEC1A Promoter from Peanut (*Arachis hypogaea* L.). *PLoS ONE* 16, e0242949. doi:10.1371/journal.pone.0242949

Tang, G., Xu, P., Liu, W., Liu, Z., and Shan, L. (2015). Cloning and Characterization of 5′ Flanking Regulatory Sequences of AhLEC1B Gene from Arachis Hypogaea L. *PloS one* 10, e0139213. doi:10.1371/journal.pone.0139213

Ueki, A., Takehara, T., Ishioka, G., Kaku, N., Ueki, K., and biotechnology (2020). β-1,3-Glucanase Production as an Anti-fungal Enzyme by Phylogenetically Different Strains of the Genus Clostridium Isolated from Anoxic Soil that Underwent Biological Disinfestation. *Appl. Microbiol. Biotechnol.* 104, 5563–5578. doi:10.1007/s00253-020-10626-8

Vestergaard, M., and Ingmer, H. (2019). Antibacterial and Antifungal Properties of Resveratrol. *Int. J. Antimicrob. Agents* 53, 716–723. doi:10.1016/j.ijantimicag.2019.02.015

Wass, M. N., Kelley, L. A., and Sternberg, M. J. E. (2010). 3DLigandSite: Predicting Ligand-Binding Sites Using Similar Structures. 38, W469–W473.doi:10.1093/nar/gkq406

Windsor, J. B., Symonds, V. V., Mendenhall, J., and Lloyd, A. M. (2000). Arabidopsis Seed Coat Development: Morphological Differentiation of the Outer Integument. *Plant J.* 22, 483–493. doi:10.1046/j.1365-313x.2000.00756.x

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

Xu, N., Wang, R., Zhao, L., Zhang, C., Li, Z., Lei, Z., et al. (2016). The Arabidopsis NRG2 Protein Mediates Nitrate Signaling and Interacts with and Regulates Key Nitrate Regulators. *Plant Cell* 28, 485–504. doi:10.1105/tpc.15.00567

Yamahara, T., Shiono, T., Suzuki, T., Tanaka, K., Takio, S., Sato, K., et al. (1999). Isolation of a Germin-like Protein with Manganese Superoxide Dismutase

Activity from Cells of a Moss, Barbula Unguiculata. *J. Biol. Chem.* 274, 33274–33278. doi:10.1074/jbc.274.47.33274

Yu, C.-S., Chen, Y.-C., Lu, C.-H., and Hwang, J.-K. (2006). FunctionPrediction of Protein Subcellular Localization. *Proteins* 64, 643–651. doi:10.1002/prot.21018

Yu, J., Xu, M., Liang, C., Zhang, X., Guo, Z., Wu, J., et al. (2019). First Report of Pythium Myriotylum Associated with Peanut Pod Rot in China. *Plant Dis.* 103, 1794. doi:10.1094/pdis-02-19-0321-pdn

Yuan, C., Li, C., Yan, C., Zhao, X., Wang, J., Sun, Q., et al. (2019a). Isolation and Characterization of a Novel Seed-specific Promoter from Peanut (*Arachis hypogaea* L.). *Mol. Biol. Rep.* 46, 3183–3191. doi:10.1007/s11033-019-04775-x

Yuan, C., Sun, Q., and Kong, Y. (2019b). Genome-wide Mining Seed-specific Candidate Genes from Peanut for Promoter Cloning. *PLoS ONE* 14, e0214025. doi:10.1371/journal.pone.0214025

Zavallo, D., Lopez Bilbao, M., Hopp, H. E., and Heinz, R. (2010). Isolation and Functional Characterization of Two Novel Seed-specific Promoters from sunflower (Helianthus Annuus L.). *Plant Cel Rep* 29, 239–248. doi:10.1007/s00299-010-0816-x

Zhang, C., Chen, H., Cai, T., Deng, Y., Zhuang, R., Zhang, N., et al. (2017). Overexpression of a Novel Peanut NBS-LRR geneAhRRS5enhances Disease Resistance toRalstonia Solanacearumin Tobacco. *Plant Biotechnol. J.* 15, 39–55. doi:10.1111/pbi.12589

Zhu, Q., Yu, S., Zeng, D., Liu, H., Wang, H., Yang, Z., et al. (2017). Development of "Purple Endosperm Rice" by Engineering Anthocyanin Biosynthesis in the Endosperm with a High-Efficiency Transgene Stacking System. *Mol. Plant* 10, 918–929. doi:10.1016/j.molp.2017.05.008

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The Genome of Cultivated Peanut Provides Insight into Legume Karyotypes, Polyploid Evolution and Crop Domestication. *Nat. Genet.* 51, 865–876. doi:10.1038/s41588-019-0402-2

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.

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, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sharif, Chen, Deng, Ali, Khan, Zhang, Xie, Chen, Cai, Yang, Zhuang, Raza and Zhuang. 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 January 2022
doi: 10.3389/fgene.2021.791355

Transcriptome Analysis Reveals Key Pathways and Candidate Genes Controlling Seed Development and Size in Ricebean (*Vigna umbellata*)

Sachin Kumar Verma[1‡], *Shikha Mittal*[1‡], *Gayacharan*[1], *Dhammaprakash Pandhari Wankhede*[1], *Swarup Kumar Parida*[2], *Debasis Chattopadhyay*[2], *Geeta Prasad*[1], *Dwijesh Chandra Mishra*[3], *Dinesh Chandra Joshi*[4], *Mohar Singh*[1], *Kuldeep Singh*[1†] *and Amit Kumar Singh*[1*]

[1]*ICAR-National Bureau of Plant Genetic Resources, New Delhi, India,* [2]*National Institute of Plant Genome Research, New Delhi, India,* [3]*ICAR- Indian Agricultural Statistics Research Institute, New Delhi, India,* [4]*ICAR-Vivekananda Parvatiya Krishi Anusandhan Sansthan, Almora, India*

Ricebean (*Vigna umbellata*) is a lesser known pulse with well-recognized potential. Recently, it has emerged as a legume with endowed nutritional potential because of high concentration of quality protein and other vital nutrients in its seeds. However, the genes and pathways involved in regulating seed development and size are not understood in this crop. In our study, we analyzed the transcriptome of two genotypes with contrasting grain size (IC426787: large seeded and IC552985: small seeded) at two different time points, namely, 5 and 10 days post-anthesis (DPA). The bold seeded genotype across the time points (B5_B10) revealed 6,928 differentially expressed genes (DEGs), whereas the small seeded genotype across the time point (S5_S10) contributed to 14,544 DEGs. We have also identified several candidate genes for seed development–related traits like seed size and 100-seed weight. On the basis of similarity search and domain analysis, some candidate genes (*PHO1*, *cytokinin dehydrogenase*, A-type cytokinin, and *ARR* response negative regulator) related to 100-seed weight and seed size showed downregulation in the small seeded genotype. The MapMan and KEGG analysis confirmed that auxin and cytokinin pathways varied in both the contrasting genotypes and can therefore be the regulators of the seed size and other seed development–related traits in ricebeans. A total of 51 genes encoding $SCF^{TIR1/AFB}$, *Aux/IAA*, *ARFs*, *E3* ubiquitin transferase enzyme, and *26S* proteasome showing distinct expression dynamics in bold and small genotypes were also identified. We have also validated randomly selected SSR markers in eight accessions of the *Vigna* species (*V. umbellata*: 6; *Vigna radiata*: 1; and *Vigna mungo*: 1). Cross-species transferability pattern of ricebean–derived SSR markers was higher in *V. radiata* (73.08%) than *V. mungo* (50%). To the best of our knowledge, this is the first transcriptomic study conducted in this crop to understand the molecular basis of any trait. It would provide us a comprehensive understanding of the complex transcriptome dynamics during the seed development and gene regulatory mechanism of the seed size determination in ricebeans.

Keywords: ricebean, seed size, hormone signaling, MapMan, SSR

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Mahesh Rao,
Indian Council of Agricultural Research, India
Suresha Giriyapura Shivalingamurthy, Indian Council of Agricultural Research, Coimbatore, India
Khela Ram Soren,
Indian Institute of Pulses Research (ICAR), India

***Correspondence:**
Amit Kumar Singh
amit.singh5@icar.gov.in

***Present address:**
Kuldeep Singh,
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

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: *08 October 2021*
Accepted: *23 December 2021*
Published: *21 January 2022*

Citation:
*Verma SK, Mittal S, Gayacharan, Wankhede DP, Parida SK, Chattopadhyay D, Prasad G, Mishra DC, Joshi DC, Singh M, Singh K and Singh AK (2022) Transcriptome Analysis Reveals Key Pathways and Candidate Genes Controlling Seed Development and Size in Ricebean (*Vigna umbellata*). Front. Genet. 12:791355. doi: 10.3389/fgene.2021.791355*

INTRODUCTION

A rapid increase in the human population, which is expected to reach 9.7 billion by 2050, is one of the biggest challenges of this world (Gu et al., 2021). To ensure food and nutritional security to the ever-growing human population, it is extremely important to bring underutilized and neglected crops into mainstream agriculture. Owing to its short growth duration and ability to thrive well in stress conditions and various soil types, ricebean (*Vigna umbellata*) is one such crop which has huge potential to sustain food and nutritional security in most parts of the world (Pattanayak et al., 2019). It is a diploid ($2n = 2\times = 22$), warm-season annual legume with a genome size of approximately 440 Mb (Kaul et al., 2019). Ricebean is mainly cultivated in Nepal, Bhutan, Northeast India up to Myanmar, Southern China, Northern Thailand, Laos, Vietnam, Indonesia, and East Timor (Tian et al., 2013), where it constitutes an important source of protein for the sizable population and contributes to household food and nutritional security. The observed protein content in ricebean is 25.57% with high concentration of various essential amino acids. Besides protein, ricebean grains also contain a significant amount of other nutrients such as carbohydrates, fiber, minerals, vitamins, and fatty acids. Moreover, ricebean is a rich source of unsaturated fatty acids like linoleic and linolenic acids (Katoch, 2013).

Among various productivity traits, pod length, seed size, and seed weight have major emphasis on ricebean genetic improvement programs because of their direct impact on the total grain yield. Furthermore, the seed is the key reservoir of proteins, essential amino acids, unsaturated fatty acids, and minerals in ricebean. Therefore, it is of great importance to decipher the molecular mechanism underlying seed development and size determination process in this minor but potential pulse crop. In recent years, with the advent of next-generation sequencing technology, key gene regulatory networks governing pod and seed development have been well characterized in both model plants like rice (Herridge et al., 2011), *Arabidopsis* (Herridge et al., 2011; Mahto et al., 2017), and also in non-model legume species like black gram (Souframanien and Reddy, 2015), cowpea (Lonardi et al., 2019), chickpea (Pradhan et al., 2014), mungbean (Tian et al., 2016a), and soybean (Jones and Vodkin, 2013; Qi et al., 2018; Peng et al., 2021). These studies revealed that seed development in higher plants is a highly complex process and governed by phytohormone signaling including cytokinins (CKs), gibberellins (GAs), brassinolides (BRs), ethylene (ET), and their associated genes and transcription factors. In all these phytohormones, genes related to auxin pathways including indole-3-acetic acid (*IAA*), auxin-responsive protein (*IAA12*, *IAA*), auxin response factors (*ARFs*), *SAUR*-like auxin superfamilies, auxin-related *Aux/IAA*, *OsIAA18*, and *AP2/ERF*, along with other genes such as *ARR-B* (cytokinin signaling), ethylene-responsive transcription factor (*ERF084*-like, *ERF4*, *ERF061*), ethylene-insensitive protein 3 (*EIN3*), ethylene receptor (*ETR*), ethylene-insensitive protein 4 (*EIN4*), serine/threonine-protein kinase (*CTR1*), ethylene responsive *APATELA2* (*AP2*), ethylene-responsive element binding protein (*EREBP*), and many more genes were reported during seed development (Garg et al., 2011; Jones and Vodkin, 2013; Tian et al., 2016a; Nelson et al., 2017; Geng et al., 2018; Li et al., 2019a; Lonardi et al., 2019; Yi et al., 2019; Raizada and Jegadeesan, 2020; Zhu et al., 2020).



FIGURE 1 | Two contrasting genotypes of ricebean, that is, IC426787 (bold size) and IC552985 (small size), selected for the transcriptome analysis on the basis of their seed size.

The aforementioned transcriptome-based gene expression analysis has provided a robust functional genomics resource for deciphering gene networks and candidate genes regulating various biological processes in crop plants. For minor crops with poorly characterized genomes, like ricebean, such detailed transcriptome analysis will provide comprehensive information about expression patterns of genes and molecular mechanisms governing traits of economic importance. This valuable information can further be employed for the development of functional markers for gene and QTL mapping. Therefore, in the present study, we conducted transcriptome analyses to investigate gene expression networks and identify the candidate genes controlling seed size variation in ricebean. RNA sequencing of two contrasting ricebean genotypes was performed at early development stages (i.e., 5 and 10 DPA). The study provides detailed insights into various gene networks and their potential roles in determining seed size. Furthermore, the study also identified simple sequence repeat (SSR) motifs that could be used for molecular mapping of seed size/weight and other related traits.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of two contrasting ricebean genotypes, namely, IC426787 (bold seeded) and IC552985 (small seeded) were obtained from ICAR-National Bureau of Plant Genetic Resources (NBPGR), New Delhi (**Figure 1**). On the basis of the 2-year trial (2018 and 2019), the average 100-seed weight of IC426787 and IC552985 was 13.20 and 3.87 gm, respectively. Plants were grown in a net-house at ICAR-NBPGR, New Delhi (latitude: 28°38′56″N, longitude: 77°9′8″E, altitude: 228 mean sea level (msl)), during *Kharif* (rainy) season 2020. During pod filling, the minimum

temperature ranged from 10.8 to 23°C, maximum temperature ranged from 30.4 to 36°C, and average RH% varied from 53 to 56. The ricebean pod filling duration varied from 20 to 30 DPA depending upon the genotype. Genotypes with smaller grain size took comparatively less pod filling time than the genotypes having larger grain size. Three biological replicates of pod samples were harvested from three full-grown plants of both genotypes at 5 and 10 DPA each. The seeds were separated and immediately frozen in liquid nitrogen and stored at −80°C for future use. A total of 12 samples were prepared for the construction of RNA libraries.

RNA Extraction, Library Preparation, and Sequencing

The Pure Link RNA Mini Kit (Ambion, United States) was used to extract RNA from the frozen samples. The total RNA quality was checked using the RNA 6000 Nano Kit (Agilent Technologies, United States) on a 2100 Bioanalyzer (Agilent Technologies, United States), with a minimum RNA integrity number (RIN) of 7. RNA concentrations were determined with a NanoDrop ND-8000 spectrophotometer (Nano-Drop Technologies, Thermo scientific, Wilmington, DE). RNA-Seq libraries for all samples were prepared using the NEBNext UltraII RNA library preparation kit for Illumina; Cat no: E7770 (New England Biolabs), according to manufacturers recommended protocol, and sequencing was done in a single HiSeq 4000 lane using 150 bp paired-end chemistry. Briefly, total RNA was used to purify poly (A) messenger RNA (mRNA) using oligo-dT labeled magnetic beads. Then, the isolated mRNA was fragmented into 200 to 500 bp pieces in the presence of divalent cations at 94°C for 5 min using an ultrasonicator. The cleaved RNA fragments were copied into first-strand cDNA using SuperScript-II reverse transcriptase (Life Technologies, Inc.) and random primers. After second-strand cDNA synthesis, fragments were end-repaired and A-tailed, and indexed adapters were ligated. The products were purified and enriched with PCR to create the final cDNA library. The tagged cDNA libraries were pooled in equal ratios and used for 2 × 150 bp paired-end sequencing on a single lane of the Illumina HiSeq 4000. Illumina clusters were generated and loaded onto the Illumina Flow Cell on the Illumina HiSeq 4000 instrument, and sequencing was carried out using 2 × 150 bp paired-end chemistry. After sequencing, the samples were demultiplexed, and the indexed adapter sequences were trimmed using CASAVA v1.8.2 software (Illumina Inc.).

Read Quality and Adapter Removal

Raw reads of ricebean were evaluated for their quality using FASTQC v0.11.8 package (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Four parameters were considered: base quality score distribution, sequence quality score distribution, average base content per read, and GC distribution in the reads. Trimmomatic v0.36 was applied to remove the adapter and trim the low-quality reads (trimming includes reads with or without ambiguous sequence "N") using default parameters. To correct the random sequencing errors in Illumina RNA-Seq reads, Rcorrector v1.0.3 was used. Clean reads were also checked for their quality using FASTQC only.

RNA-Seq *De Novo* Assembly and Transcriptome Assessment

The obtained clean reads of all 12 samples were assembled using Trinity v2.4.0 with the paired-end model and default K-mer value of 25. The *de novo* assembly was merged and clustered using CDHIT v4.0 to get non-redundant sequences. Furthermore, these non-redundant sequences were made transcripts using the trinity in-built script. The clean reads of each sample were mapped back to the *de novo* assembled genome through BWA-mem software with default parameters. The BAM files were handled by samtools. The number of reads mapped to genes was calculated using samtools v0.1.19. The expression difference of each transcript between different samples was calculated using DESeq2 R package. False discovery rate (FDR) values less than 0.01 and $|\log2 \text{ (fold change)}| \geq 2$ were considered significant differences at the expression level. The transcript abundance was normalized by the fragments per kilobase of transcript per million mapped reads (FPKM) value.

Gene Functional Analysis

To annotate the assembled transcripts, sequences were aligned by BLASTX (*e*-value $<1e^{-5}$) to protein databases, including the non-redundant protein (NR) database, Swiss-Prot, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. A GO enrichment analysis was conducted for the transcripts according to biological process, cellular component, and molecular function ontologies using Blast2GO software (Liu et al., 2013; Calzadilla et al., 2016). The GO annotation functional classifications were determined using WEGO software for the distribution of gene functions (Ye et al., 2006). GO functional enrichment and KEGG pathway enrichment analysis were also tested at a significance cutoff of *p*-value. All the *p*-values were adjusted with the criterion of Bonferroni correction. We selected the corrected *p*-value of 0.05 as the threshold to determine significant enrichment terms of the gene sets. The MapMan analysis was also conducted to provide a graphical overview of the metabolic and regulatory pathways for the detected genes using the MapMan tool, and the mapping file of ricebean for all the samples was generated using the Mercator tool.

Candidate Gene Identification and Their Domain Analysis

The candidate genes for seed development-related traits were identified on the basis of similarity (BLASTX with similarity >80% and *e*-value <0.001) with genes responsible for similar traits in other species, including *Arabidopsis*, *Phaseolus vulgaris*, and *Vigna* species (*V. radiata* and *V. angularis*). Furthermore, the candidate genes were also validated *in silico* on the basis of their domain analysis. The amino acid sequences of the identified candidate genes were predicted and compared against the Pfam protein database using HMMER 3.0 (*e*-value $\leq 1e^{-10}$) to obtain candidate gene domain/family annotation information. A

heatmap was also generated for the candidate genes on the basis of their expression in both the genotypes at different times of development. The heatmap was made using an in-house R script.

Simple Sequence Repeats Identification and Primer Design

The MIcroSAtellite (MISA) search engine (http://pgrc.ipk-gatersleben.de/misa) was employed for the identification of SSRs. The minimum numbers of repeats used for selecting the SSRs were ten for mononucleotide-based loci, six for dinucleotide loci, five for trinucleotide loci, and three for all larger repeat types (tetra- to hexanucleotide motifs). For validation, 50 SSR motifs were randomly selected, that is, 25 for dinucleotide and trinucleotide each. The primers for these selected SSR motifs were designed based on flanking sequences using Primer3 software (http://sourceforge.net/projects/primer3) with targeted size of PCR products in the range between 100 and 300 bp, primer length between 18 and 22 bp, GC content between 40 and 70, and melting temperature of 50–60°C.

Simple Sequence Repeats Validation

DNA was isolated from young leaves of eight accessions of *Vigna* species including *V. umbellata* (6), *V. mungo* (1), and *V. radiata* (1) by following the protocol described in the DNeasy plant mini kit (QIAGEN, Hilden, Germany). DNA concentration was measured using a NanoDrop™ 2000 spectrophotometer (Thermo, United States), and DNA quality was analyzed using 0.8% agarose gel. A working stock of DNA was (10 ng/μl) prepared with nuclease-free water for polymerase chain reaction (PCR) for SSR amplification.

For the SSR amplification, 20 μl reaction mixture containing 4 μl genomic DNA (40 ng), 10 μl Taq Polymerase 2X Master Mix (United States), 0.8 μl primers (10 pM), and 5.2 μl nuclease-free water were used. For the amplification, the following thermal conditions were carried out: initial denaturation of 94°C for 3 min, then 35 cycles of 94°C for 30 s, primer annealing at 55°C for 45 s, extension at 72°C for 1 min, and final extension for 10 min. PCR products were separated using high-resolution metaphor agarose gel (3%) electrophoresis. Furthermore, the dendrogram of the genotypes was generated using the hierarchical clustering algorithm in DARwin v6.0.21 software (https://darwin.cirad.fr/).

RESULTS AND DISCUSSION

Transcriptome Sequencing and *De Novo* Assembly

To obtain a comprehensive transcriptome profile of ricebean 12 RNA libraries were sequenced, and a total of 94.35 Gb raw data were generated. For these 12 samples, approximately, 98.50–99.80% of reads passed the quality control, and 98.60–99.60% of the clean reads were mapped back to the *de novo* assembled ricebean genome. On average, raw data of the seed transcriptome at 5 DPA and 10 DPA had 50.33 and 48.66% GC content, respectively, while after trimming, the GC content of clean data at 5 DPA and 10 DPA was 48.66 and 49.33%, respectively, which is similar to the GC content reported in the previous study of ricebean (Chen et al., 2016; **Table 1**).

The obtained clean reads of all 12 samples were assembled using Trinity (v2.4.0) with default parameters. The assembled transcriptome consists of 218,486 super transcripts with an N50 value of 1,041. The number of transcripts generated in the current study is comparable to previous studies. In terms of N50, the ricebean had a higher N50 value than field pea (781) and chickpea (441) (Pradhan et al., 2014; Sudheesh et al., 2015) and less value than mungbean, common bean, and adzuki bean (Hiz et al., 2014; Chen et al., 2015b; Chen et al., 2016). These results indicate the good quality of ricebean transcriptome.

The lengths of the transcripts ranged from 201 to 15,828 bp, with an average length of 669 bp, which is less than other *Vigna* species like cowpea (871 bp) and mungbean (874 bp) but more than that of black gram (443 bp) (Chen et al., 2015b, 2017; Souframanien and Reddy, 2015). Of these transcripts, 146,622 (67.11%) were 201–500 bp; 39,620 (18.13%) were 501–1,000 bp; 12,654 (5.79%) were 1,001–1,500 bp; 6,511 (2.98%) were 1,501–2,000 bp; 3,986 (1.82%) were 2,001–2,500 bp; 2,567 (1.17%) were 2,501–3,000 bp; and 6,526 (2.99%) were more than 3,000 bp in length (**Figure 2**). The developed assembly showed ~100% back mapping of total and important reads, and this shows that our assembly had vast and proper mapping quality for the generated reads. The high percentage of reads mapping back to the *de novo* assembled transcriptome is a quality metric that provides an assessment of assembly entirety (Hornett and Wheat, 2012).

Differential Expression Analysis

In this study, a comprehensive transcriptome analysis has been performed with the aim to reveal those gene expression changes that, independently of the genotype diversity, are involved in controlling seed size in ricebean. Comparative transcriptome analysis was performed between two genotypes with contrasting seed size at two time points, namely, 5 and 10 DPA. A similar type of study using two genotypes with a contrasting seed size has also been done in the peanut (Li et al., 2019b). The expression profile was checked for the individual genotypes across the time points (B5_B10, S5_S10) as well as between the genotypes at each time point (B5_S5, B10_S10) (**Supplementary Table S1**). False discovery rate (FDR) values less than 0.01 and |log2 (fold change) |≥2 were considered significant differences at the expression level.

While evaluating the expression difference individually for the bold genotype across the time point (B5_B10), 6,928 differentially expressed genes were identified. In B5_B10, the number of upregulated genes (6,284) were higher than downregulated genes (644), suggesting that these upregulated DEGs might be responsible for the increase in seed size. Similarly, a small genotype across the time point (S5_S10) contributed to 14,544 DEGs (**Figure3A**; **Table 2**). In contrast to B5_B10 expression results, S5_S10 had a high number of downregulated genes (7,862) in comparison with the upregulated genes (6,682), indicating that these downregulated genes might be repressing any transcriptional activity or downstream pathways resulting in the small size of ricebean seeds (Li et al., 2019b). To gain a better understanding of molecular processes/regulatory networks

TABLE 1 | Summary of RNA-Seq data for 12 samples of ricebean at 5 DPA and 10 DPA.

Genotype	Replicate	Time point	Read before quality control	Read after quality control	GC%
Bold (IC426787)	Replicate 1	5 DPA	28,178,488	28,081,924	49
		10 DPA	32,397,064	32,223,234	52
	Replicate 2	5 DPA	20,042,440	19,897,123	51
		10 DPA	23,598,201	23,252,956	48
	Replicate 3	5 DPA	32,043,486	31,869,247	51
		10 DPA	26,455,283	26,137,490	46
Small (IC552985)	Replicate 1	5 DPA	27,160,408	27,037,555	49
		10 DPA	21,272,208	21,053,611	50
	Replicate 2	5 DPA	24,589,394	24,425,550	49
		10 DPA	25,209,773	25,053,008	49
	Replicate 3	5 DPA	26,969,114	26,850,966	48
		10 DPA	24,511,033	24,368,871	49

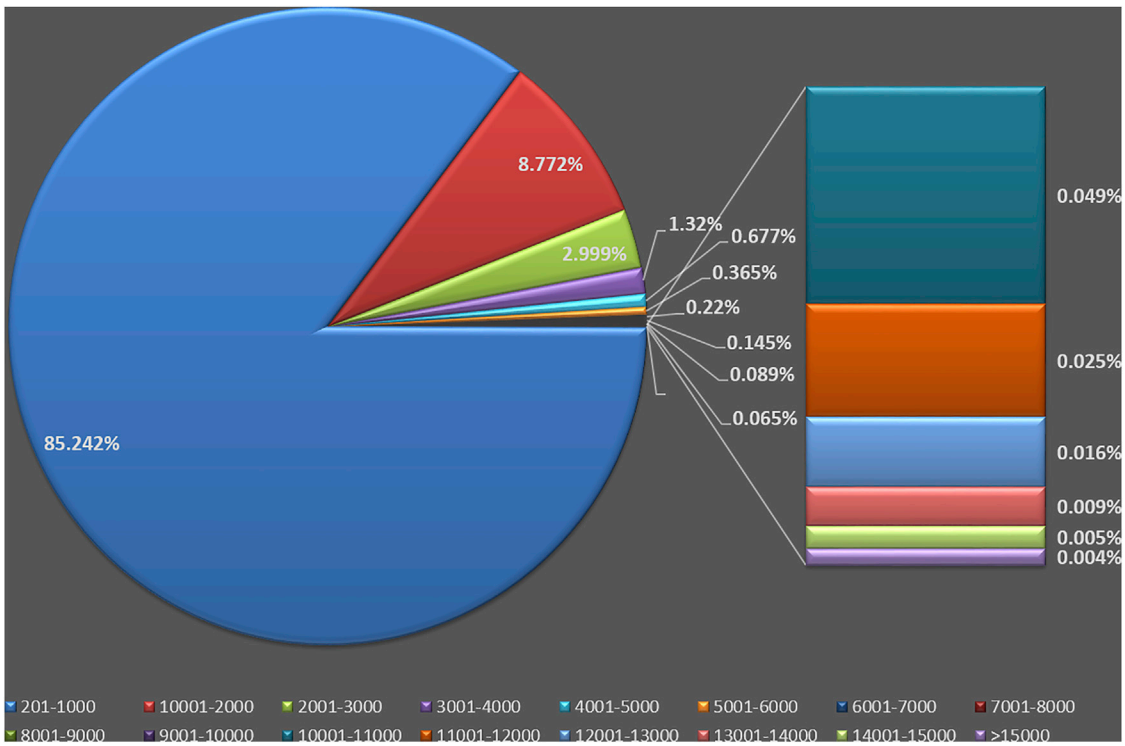


FIGURE 2 | Sequence length distribution of the assembled transcripts.

associated with the seed size in ricebeans, the pattern of differentially expressed genes was analyzed between genotypes in each time point and across the time point using a Venn diagram (**Figure 3B**).

We also identified common genes between the individual genotype across time points (i.e., B5_B10 and S5_S10) as well as between the genotypes at each time point (B5_S5 and B10_S10). In case of B5_B10 and S5_S10, in total, 2091 DEGs were common. On the other hand, 850 DEGs were common between B5_S5 and B10_S10 (**Supplementary Table S2**). The comparative gene expression analysis indicated that a relatively large amount of the transcriptional program operating during seed development or maturation is shared between both the genotypes. The same results have been observed in the case of common bean, where 2,487 DEGs were shared by two contrasting genotypes (González et al., 2021).

Gene Ontology Analysis of the Transcriptome

To infer the biological processes and the functions related to seed development stages, gene ontology analysis was conducted for differentially expressed genes in terms of their biological involvement, target cellular component, and molecular

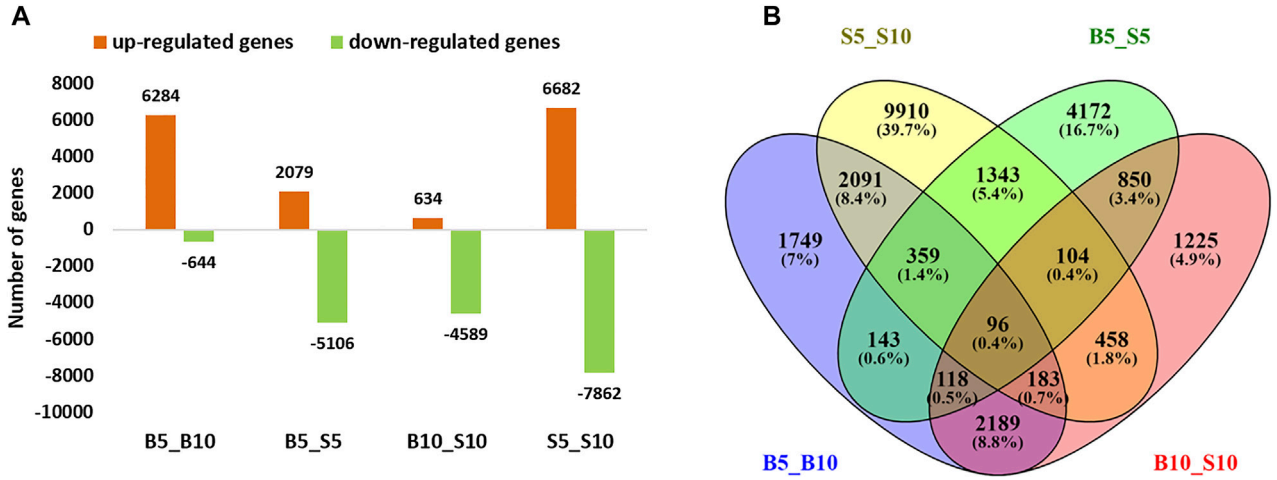


FIGURE 3 | Differentially expressed genes in the two genotypes of ricebean at two time points, i.e., 5 and 10 DPA. **(A)** Comparison of DEGs representing the share of overlapped and non-overlapped transcripts in bold and small genotypes at 5 and 10 DPA. **(B)** Number of upregulated and downregulated significant genes in bold and small genotypes.

TABLE 2 | Summary of significant DEGs identified in ricebean.

Comparison	Total DEG	Total significant DEG	Significantly upregulated DEG	Significantly downregulated DEG
B5_B10 DPA	276,372	6,928	6,284	644
B5_S5 DPA	227,479	7,185	2,079	5,106
B10_S10 DPA	264,964	5,223	634	4,589
S5_S10 DPA	220,089	14,544	6,682	7,862

function using Blast2GO. Out of total 33,880 DEGs, 16,002 DEGs contributed to GO terms. In core GO annotation, 7,002 (25.37%) genes annotated for biological process (BP), 12,069 (43.72%) for molecular function (MF), and 8,533 (30.91%) for cellular components (CC). The highest number of GO terms were observed in the case of S5_S10 (44.10%), followed by B5_B10 (23.89%), B10_S10 (16.91%), and B5_S5 (15.09%).

In case of the bold genotype across the time point (B5_B10), out of 6,928 DEGs, only 3,764 were annotated, constituting 1,696, 2,046, and 2,864 GO terms for BP, CC, and MF, respectively. However, in S5_S10, we observed 7,158 annotated DEGs from 14,544 DEGs and 2,954, 3,909, and 5,312 GO terms for BP, CC, and MF, respectively. On the other hand, in the case of between the genotype at the first time point (B5_S5), 2,537 DEGs were found to be annotated as compared B10_S10, in which 7,158 DEGs were annotated. In case of BP, 1,075 and 2,954 GO terms were identified in B5_S5 and S5_S10. Similarly, 2,046 and 3,909 GO terms were found for the cellular component function in B5_S5 and S5_S10, respectively, whereas in the case of molecular function, B5_S5 and S5_S10 consisted of 1,875 and 5,312 GO terms, respectively (**Supplementary Table S3**).

We have also illustrated the top or enriched functions in terms of BP, MF, and CC for both the genotypes. For example, the top biological activities include "cellular process," "nitrogen compound metabolic process," "small molecule metabolic process," "cellular component organization," "regulation of metabolic process," "response to stress," "cell wall organization," cellular response to stimulus," and developmental process. All these results indicated the biological process of DEGs vary over a broad range of terms. These enriched GO terms for BP indicate that hormone and environment stimuli played a vital role in ricebean seed/pod development. A similar type of results was also found in peanut pod development (Zhu et al., 2014).

Similarly, in the case of MF, bold and small genotypes were identified to be involved in "binding," "metabolic processes," "organic cyclic compound binding," "heterocyclic compound binding," "ion binding," "transferase activity," and "biosynthetic processes." However, on the other hand, cellular component activities include "catalytic activity," "membrane," "membrane part," "intrinsic component of the brain," and "intracellular" and cellular activities" (**Figure 4**). Similar results for MF and CC were observed in the pod development of peanuts (Zhu et al., 2014).

MapMan Analysis

For comprehensive assessment of gene expression network dynamics in a developing seed of bold and small genotypes, identified DEGs were mapped onto metabolic maps using the MapMan tool and categorized into BINS on the basis of their functional groups. We could observe various functional groups of genes activated at different stages of seed development.

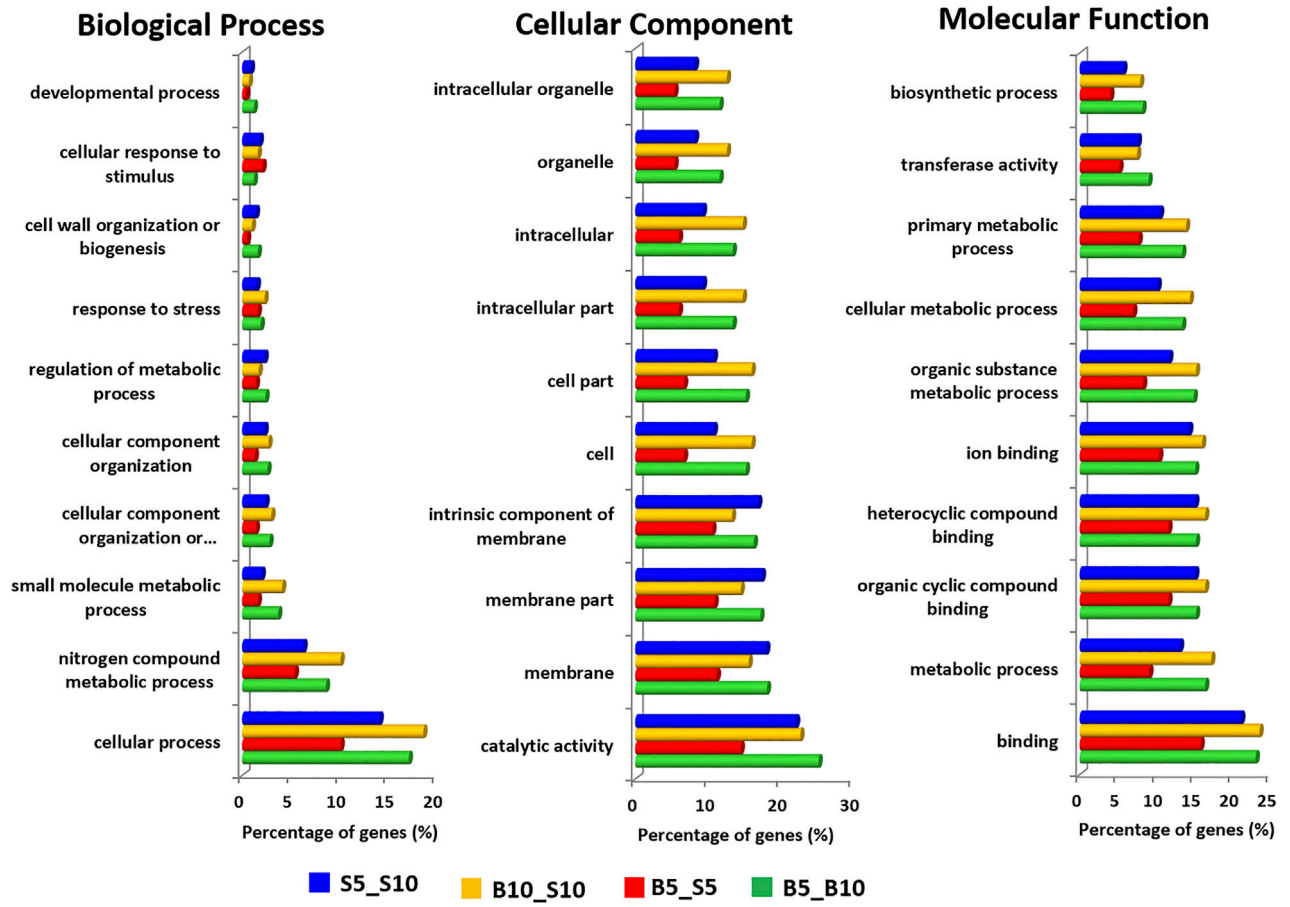


FIGURE 4 | Gene ontology (GO) annotation of differentially expressed genes in ricebean summarized in three main categories: biological process, cellular component, and molecular function.

Interestingly, we noticed a major variation between the bold and small genotypes with respect to genes related to important functions like those involved in different aspects of metabolism and signaling or regulation. A detailed analysis of genes expressed in these categories that actually distinguish the two genotypes was considered relevant, and a major emphasis was therefore given to the BINS in which the genotypes were found to be involved. This analysis allowed exploration of the global activation of specific metabolic pathways and gene regulatory networks activated during ricebean seed development.

For the whole ricebean transcriptome, we annotated 13,759 transcripts with MapMan BINS of known function after running the Mercator web tool. In total, these transcripts were classified into 29 BINS. The transcripts were expressed mainly in the following categories: carbohydrate metabolism (major and minor CHO metabolism), amino acid turnover, photosynthesis, secondary metabolism, and cell wall organization (**Supplementary Table S4**). In the former categories, most of the transcripts were highly expressed in B5_B10, while downregulated in the case of S5_S10 (**Figure 5A**). B5_B10 and S5_S10 shared 17 pathways, but only two pathways were found in B5_B10 such as RNA processing and polyamine metabolism, indicating that these two pathways triggered after 5 DPA. Similarly, while comparing expressed transcripts between the genotype at the same time points (i.e., B5_S5 and B10_S10), 18 categories were the same, except the polyamine metabolism which was detected only at the second time point, that is, 10 DPA, which also confirms our previous result that polyamines activate only in the case of bold genotype after 5 DPA of seed development (**Figure 6A**).

In our study, photosynthesis-related genes were highly enriched in bold genotypes in comparison with the small genotype which is in similarity with the previously published reports (Zhu et al., 2014; Clevenger et al., 2016; Sinha et al., 2020). The main role of photosynthesis in seed development is reported to increase the internal oxygen content and to control biosynthetic fluxes by improving the energy supply (Borisjuk et al., 2004), and it can also affect the metabolism in a number of distinct ways (Ruuska et al., 2004). Our results indicate that many metabolic genes are most active during ricebean seed filling, which aligns with previous studies on *M. truncatula* and *P. sativa* where approximately half of the seed-regulated genes were assigned to metabolic pathways (Benedito et al., 2008; Liu et al., 2015).

Furthermore, some DEGs are also mapped to hormone metabolic pathways. Majority of the genes associated with

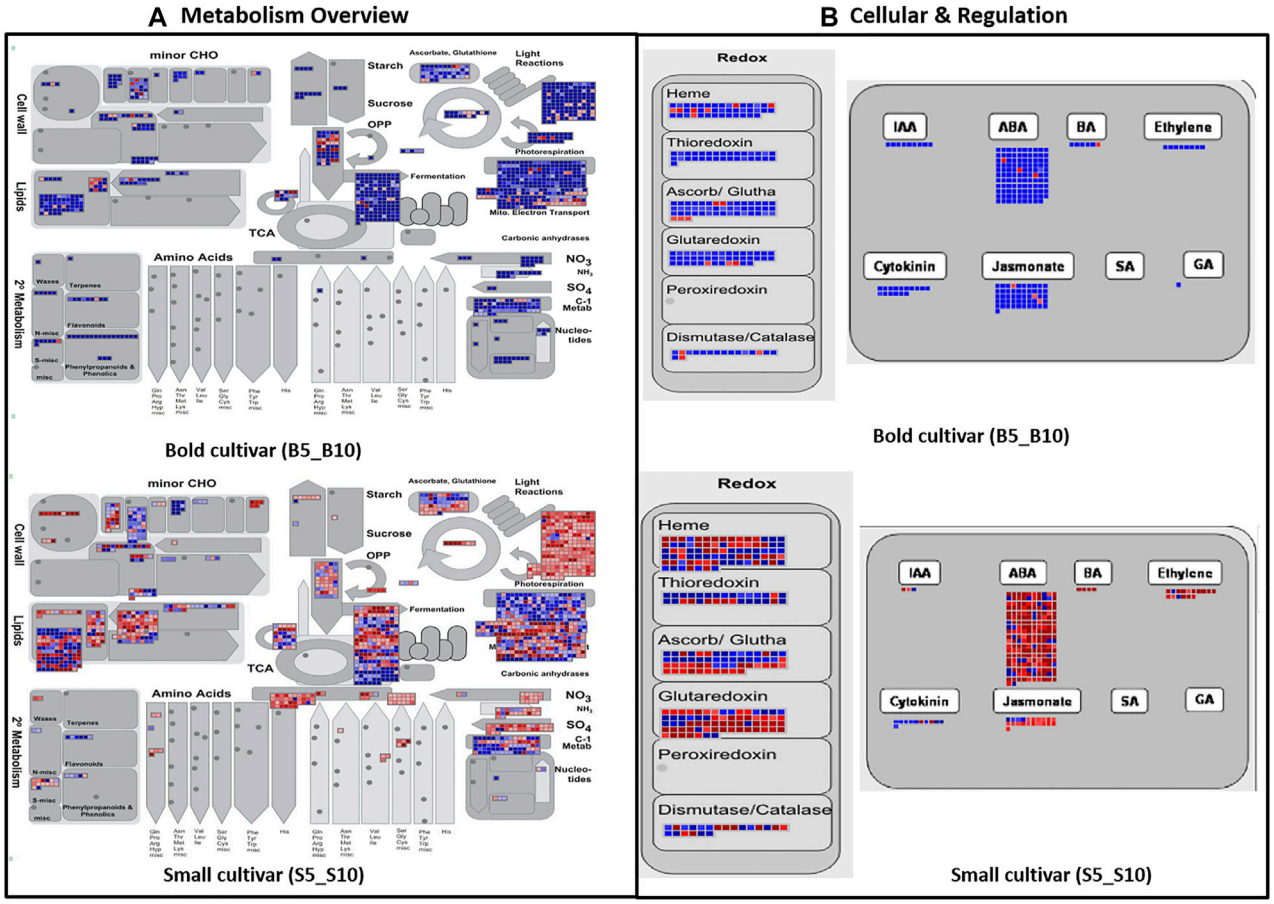


FIGURE 5 | MapMan pathway representing the differential expression of genes across the time point involved in **(A)** metabolism **(B)** cellular and regulation pathway in bold and small genotypes.

biosynthesis and response of many phytohormones like IAA, ABA, BAP, ethylene, cytokinin, jasmonate, and gibberellic acid were upregulated in the case of bold genotypes (B5_B10) as compared to small genotypes (S5_S10), in which most of the genes were downregulated (**Supplementary Table S5**; **Figure 5B**), whereas in case of B5_S5 and B10_S10, mixed expression of phytohormones was observed (**Figure 6B**); a complex regulatory network triggers the initiation of seed development, maturation, and accumulation of storage products. Several studies suggested the vital role of phytohormones in pod and seed development (Zhu et al., 2014; Huang et al., 2017; Wan et al., 2017; Kumar et al., 2019; Sinha et al., 2020). In 2017, a study demonstrated the role of phytohormones in various aspects of plant hormone homeostasis including biosynthesis, metabolism, receptor, and signal transduction (Xu and Huang, 2017).

Kyoto Encyclopedia of Genes and Genomes Pathway Analysis

The KEGG pathway enrichment analysis was conducted for two contrasting genotypes at both time points (i.e., 5 DPA and 10 DPA) at a p-value <0.05 using the KEGG database server. The KEGG pathway enrichment analysis indicated that 7,178 transcripts obtained hits in the KEGG database, and those transcripts were associated with 106 unique pathways. The 7,178 transcripts included 3,112, 434, 2,103, and 1,529 transcripts with respect to B5_B10, B5_S5, B10_S10, and S5_S10, respectively. The pathway enrichment analysis of DEG conducted between different combinations, B5_B10, B5_S5, B10_S10, and S5_S10, revealed involvement the of 7, 52, 458, and 35 pathways, respectively. In case of B5_B10 and S5_S10, from the top 10 pathways, four pathways, namely, biosynthesis of secondary metabolites, protein processing in the endoplasmic reticulum, plant–pathogen interaction, and starch and sucrose metabolism were common. On the other hand, between the genotypes at both the time points (i.e., B5_S5 and B10_S10), only one pathway i.e., metabolic pathway—was common. The top 10 pathways among the time points for both genotypes as well as between the genotypes at both the time points are represented in **Figure 7**.

In KEGG pathway–based analysis, we observed a clear difference in the expression of some phytohormones which regulates seed development, including auxin, cytokinin, gibberellin, and ethylene. The differential expression of these

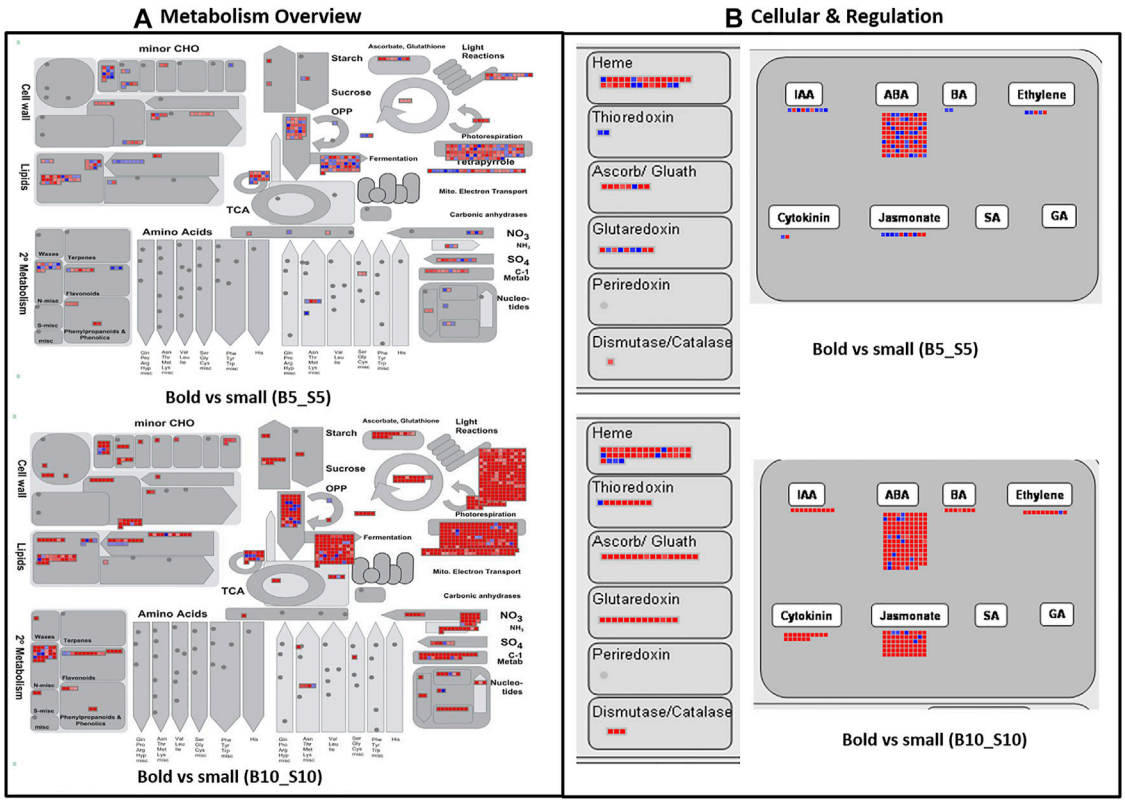


FIGURE 6 | MapMan pathway representing the differential expression of genes across the genotype involved in **(A)** metabolism **(B)** cellular and regulation pathway in bold and small genotypes.

phytohormones was also observed in our MapMan analysis. This was not surprising since phytohormones control or influence all aspects of plant growth and reproduction, including seed germination, growth of roots, stems and leaves, plant flowering, seed development, seed fill, and seed dormancy. The expression pattern of key genes involved in biosynthesis and signaling of important phytohormones was compared between small and bold seeded genotypes for their possible role in determining seed size.

Auxin Pathway

Auxin regulates many aspects of plant growth and development, including embryogenesis (Möller and Weijers, 2009), the architecture of the root system (Benková et al., 2003), gravitropism (Rashotte et al., 2003), phototropism (Blakeslee et al., 2004), initiation and radial positioning of plant lateral organs, and cell elongation (Reinhardt et al., 2000; Christian et al., 2006). Auxin is sensed by its receptor protein such as *TRANSPORT INHIBITOR RESPONSE 1/AUXIN-SIGNALING F-BOX* proteins (*TIR1/AFBs*) which mediate the auxin signaling pathway and centered on a ubiquitin-dependent *Skp1-Cullin-F-box* (*SCF*)$^{TIR1/AFBs}$ protein complex to regulate the *Aux/IAAs-ARFs* flow (Leyser, 2003; **Figure 8A**). The TIR receptor protein confers substrate specificity and target-specific *Aux/IAA* proteins for degradation via the *SCF*$^{TIR1/AFBs}$ protein complex, in the presence of auxin. The degradation of *Aux/IAA* leads to switching on transcriptional expression of a range of genes including auxin responsive factors (*ARFs*) which in turn regulate the expression of several other genes that have a role in auxin-mediated plant growth and development.

The KEGG pathway expression–based analysis revealed a clear difference in the auxin signaling pathway in two contrasting ricebean genotypes, which is also in accordance with our MapMan results where auxin signaling related genes showed higher expression in the bold genotype than the small genotype. We found approximately 51 genes encoding *SCF*$^{TIR1/AFB}$, *Aux/IAA*, *ARFs*, *E3* ubiquitin transferase enzyme, and *26S* proteasome, showing distinct expression dynamics in bold (B5_B10) and small (S5_S10) genotypes (**Supplementary Table S6**). The three key signaling elements *TIR1/AFBs*, *Aux/IAAs*, and *ARFs* have also been identified in different species including *Arabidopsis* (Chapman and Estelle, 2009), populus (Kalluri et al., 2007), and rice (Jain et al., 2005; Parry et al., 2009; Shen et al., 2010). Similarly, several studies focused on the role of *AUX/IAA* in determining the seed size with the influence of the expression of a gene in *AUX* biosynthesis (*ZmTar3*, *ZmTar1*, and *ZmYuc1*) and signaling (auxin efflux carriers, *PIN*, and *ARF2*) (Schruff et al., 2006; Bernardi et al., 2016). Homologs of *ZmYuc1*, *PIN*, and *ARF2* were significantly differentially expressed during tartary buckwheat seed development (Huang et al., 2017). The high expression of DEGs in bold genotypes corresponds to cell division, and expansion is faster to form larger size seeds at these stages.

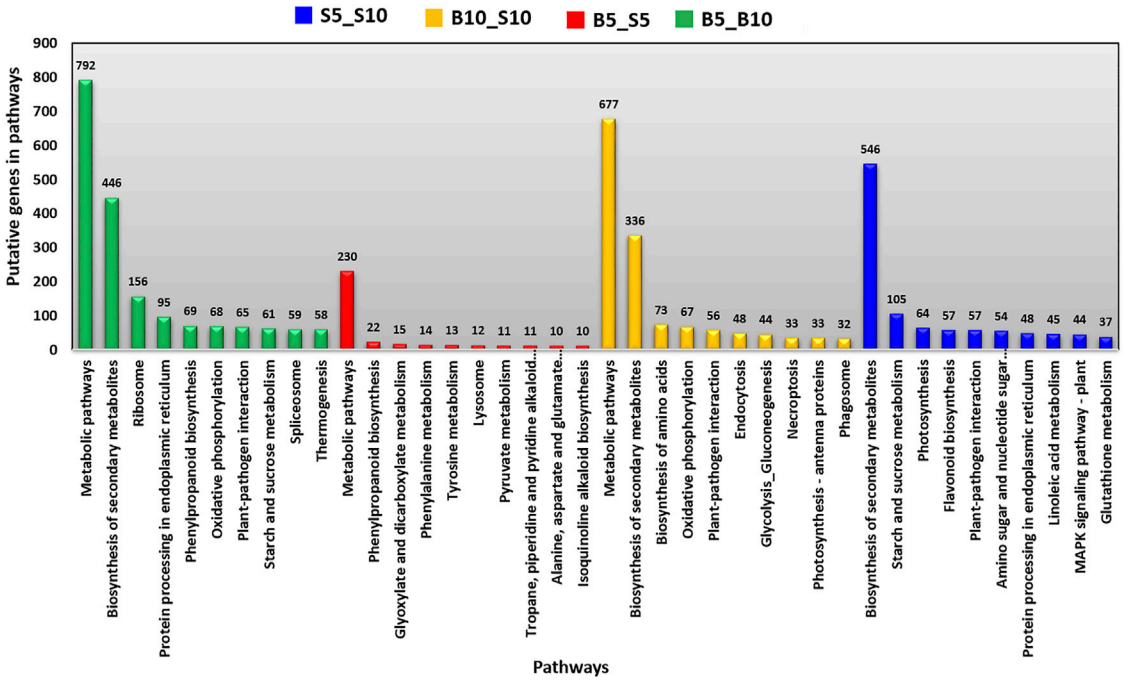


FIGURE 7 | List of top 10 pathways revealed by KEGG enrichment analysis.

The upregulation of SCF^{TIR1}, *E3* ubiquitin transferase enzyme, and *26S* proteasome was found in the bold genotypes, indicating the degradation of *Aux/IAA* and release of *ARFs* to modulate the expression of their target genes including *SMALL AUXIN UP RNA* (*SAUR*), Gretchen Hagen 3 (*GH3*), and indole-3-acetic acid–inducible gene (*Aux/IAA*), while in case of small genotypes, SCF^{TIR1} was not expressed, but *TOPLESS* (*TPL*) gene was upregulated, suggesting that Aux/IAA might have formed the complex with *ARFs* to block the transcriptional activity (**Figure 8A**; Hayashi, 2012).The induction of auxin-inducible acyl amidosynthetases, *GH3*, by the *ARF* family is the early event of auxin signaling cascade (Zhang et al., 2016). The expression of *GH3* gene was upregulated in the case of bold genotypes, while it was downregulated in the small genotypes. *SAUR* expression was upregulated in small genotypes, while downregulated in bold genotypes. The aforementioned results clearly inferred that the differential regulation of the auxin signaling pathway in bold and small genotypes might be the main factor contributing to the variation in ricebean seed size (Bai et al., 2019).

Cytokinin Pathway

Similar to auxin, cytokinin is another important plant hormone regulating many aspects of plant growth (Tarkowski et al., 2006; Werner et al., 2008). In plants, the regulation of cytokinin is facilitated by the two-component system (TCS) which consists of four groups of proteins: histidine kinases (*AHKs*; *AHK2*, *AHK3*, and *AHK4*/*WOL1*/*CRE1*), histidine-containing phosphotransfer proteins (*AHPs*; *AHP1-AHP5*), type-B response regulators (type-B *ARRs*; *ARR1*, *ARR2*, *ARR10-ARR14*, and *ARR18-ARR21*), and type-A ARRs (*ARR3-ARR9* and *ARR15-ARR17*). In *Arabidopsis*, *AHK2*, *AHK3*, and *CRE1* were found to be involved in seed size (Riefler et al., 2006; Heyl et al., 2012).

Cytokinins have been reported to function in seed development, such as seed size, seed yield, embryonic growth, with the involvement of genes encoding *isopentenyl transferase* (*IPT*), *cytokinin oxidase/dehydrogenase* (*CKX*), and *histidine kinase* (*HK*) (Bartrina et al., 2011). In our study, we have also found the expression of genes such as *IPT*, *CKX*, and *HK*. *IPT* upregulation was observed only in the case of small genotypes, whereas *CKX* was upregulated in bold genotypes, and mixed expression of *HK* was noticed in both the genotypes (**Figure 8B**; **Supplementary Table S7**). The upregulation of *CKX* in bold seed genotypes hints at its possible role in determining the seed size. The *CKX* proteins are widely distributed in plants and implicated in various plant growth and developmental processes by maintaining the endogenous cytokinin level via irreversible degradation. In plant tissues, the expression of the *CKX* genes is primarily regulated by the endogenous cytokinin level. Various past studies have shown the role of *CKX* genes in the regulation of the seed size and grain yield in different plant species. In *Arabidopsis*, a *CKX* family gene–encoded enzyme *CYTOKININ OXIDASE 2* (*CKX 2*) has been demonstrated to be associated with large seed size *via* catalyzing irreversible degradation of cytokinin. Similarly, in rice, a Gn1a locus encoding for *cytokinin oxidase/dehydrogenase* (*OsCKX2*) is shown to be responsible for high grain yield (Ashikari et al., 2005). On the other hand, the expression of type-A *Arabidopsis* response regulator (type-A *ARRs*) genes that negatively regulates the cytokinin signaling was majorly

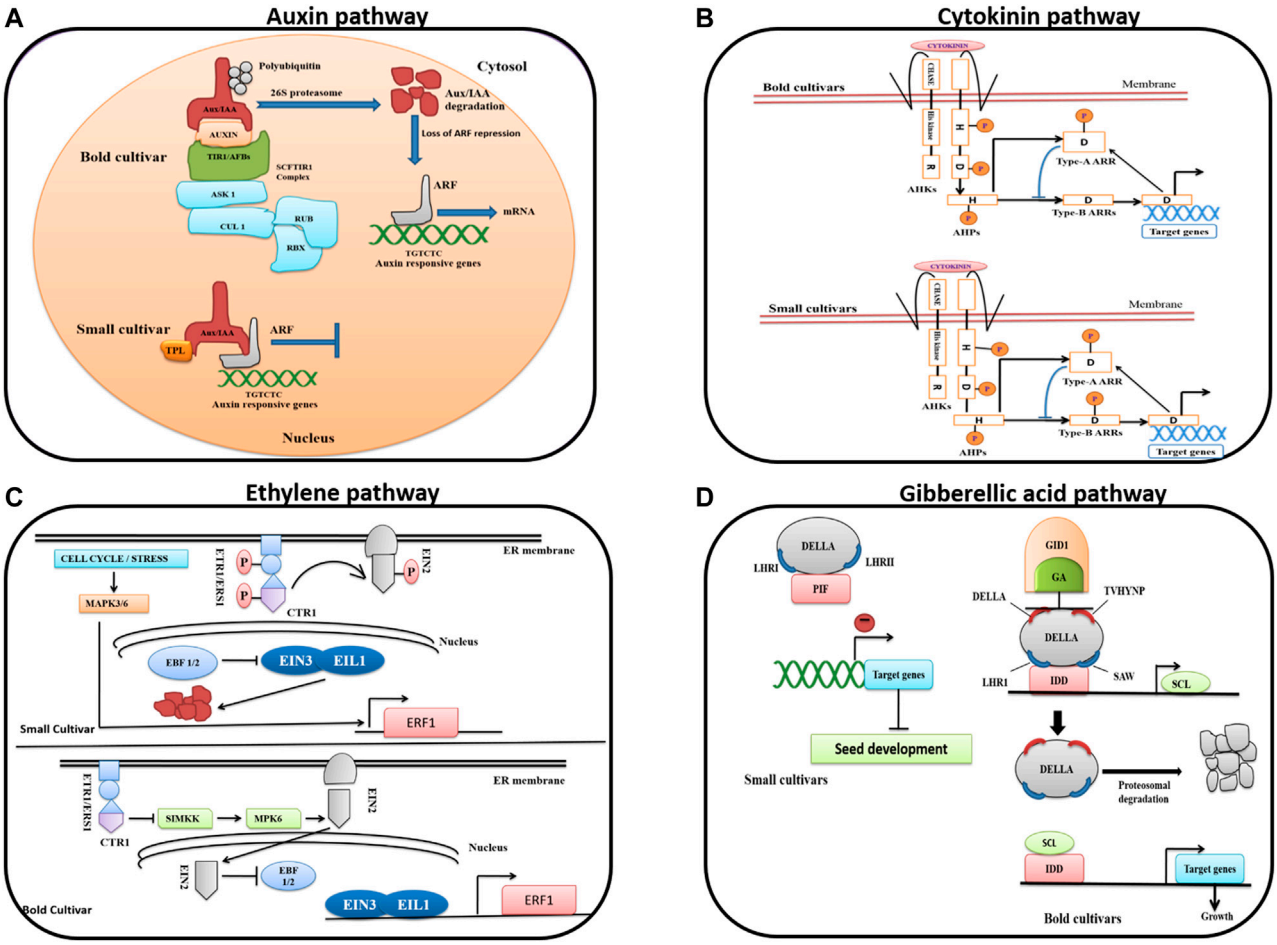


FIGURE 8 | Phytohormone pathways important for seed development are represented in two contrasting genotypes of ricebeans on the basis of their expression and involvement in the enriched KEGG pathways. **(A)** Auxin signaling pathway. **(B)** Cytokinin pathway. **(C)** Ethylene pathway. **(D)** Gibberellic acid pathway.

detected in small genotypes. This suggested that type-A *ARR* genes may be repressing the cytokinin signaling pathway (Heyl and Schmülling, 2003; Lohar et al., 2004; Desbrosses and Stougaard, 2011). The inhibition of the cytokinin signaling pathway may contribute to plant and bacterial cell differentiation (Bromley et al., 2014). Mixed expression of *AHP* and type-B *ARRs* was found in both the genotypes. Phosphate transfer to type-B *ARR* proteins modulates the transcriptional changes in the nucleus and causes the expression of primary cytokinin response genes including the type-B *ARRs*.

Ethylene Pathway

Ethylene, an "aging" hormone, has been reported to control the development of plant seeds and grains in various species (Zhong et al., 2002; Hentrich et al., 2013; Huang et al., 2013; Guo et al., 2016).Molecular evidence demonstrated ethylene's role in the regulation of seed size and seed shape, in which genes in ethylene biosynthesis (*EIN2*, *ERS1*, and *ETR1*), signaling (*CTR1*, *ETO1*, *ETR1*, and *EIN2*), and catabolism (*ACC deaminase*) were involved (Robert et al., 2008; Walton et al., 2012). According to our results, the expression of ethylene receptors (*ERS1/2*) was higher in bold genotypes than small genotypes, whereas *CTR1*, a negative regulator of ethylene hormone showed contrasting expression with upregulation in small and downregulation in bold genotypes (**Figure 8C**; **Supplementary Table S8**). In buckwheat, the differential expression of *ERS1*, *ETO1*, *ETR1*, etc. was observed (Huang et al., 2017). In case of bold genotypes, we have noticed the high expression of *SIMKK*, *MPK6*, *EIN3*-like transcription factors, and *EIN2*, indicating positive regulation of transcriptional response in the bold genotype. In case of small genotypes, the upregulation of *ERFs* depicted that *ERF* might have shown activity after the phosphorylation via the *MPK3/6*-cascade, which regulates the ethylene biosynthesis, and the expression of EIN3/EIL1 was not found which possibly indicates its degradation by ubiquitination. In our samples, we found a full cascade of gene expressions in bold genotypes, while in small genotypes, the expression of genes detour from the normal expression and opted a new route for the ethylene-inducible gene expression.

Gibberellin Pathway

Gibberellins (*GAs*) are well-known plant hormones that are widely involved in the growth and development processes. *GAs*, auxin, *ABA*, and ethylene have been involved in the regulation of seed development and pod maturation (Ziv and Kahana, 1988; Shlamovitz et al., 1995; Ozga et al., 2003). In case of bold genotypes, the expression of *GA*, *DELLAs*, and *SCF*-complex protein is upregulated, which indicates *DELLA* proteolysis; simultaneously, the upregulation of protein indeterminate domains (*IDDs*) and *scarecrow-like* proteins (*SCLs*) were also observed, which supports the feedback loop mechanism which regulate the *GA* signaling (**Figure 8D**). According to the feedback loop mechanism, *DELLA* initiates the expression of downstream genes, including *SCLs* by *IDD*-mediated interaction with their promoters. The subsequent increased concentration of *SCLs* enhances the *SCL3/IDD* complex synthesis while decreasing the formation of the *DELLA/IDD* complex and consequent suppression of the expression of *SCLs*, which mediates the homeostatic regulation of the downstream genes, including positive regulation of *SCLs* and *GA* signaling. In case of small genotypes, the expression of *SCF complex* protein was absent, while the expression of *DELLAs* was unregulated. Consecutively, we observed the *SCL* protein script, while *IDD* protein was completely absent. The expression of the phytochrome-interacting factor (*PIF*) protein was found, which indicates the *DELLA*-mediated inhibition of hypocotyls elongation (**Supplementary Table S9**).

Previous studies have revealed that genes encoding *GA2 oxidase* and *GA3 oxidase* in the GA biosynthesis pathway can affect seed development, starch biosynthesis, embryo, and seed coat development (Nakayama et al., 2002; Singh et al., 2002). The downregulation of *GA2-oxidase* was observed in our results similar to a study of the tartary buckwheat in which the downregulation of *GA2-oxidase* was also depicted during seed development (Huang et al., 2017).

The KEGG pathway and the MapMan analysis suggested the differential expression of phytohormone biosynthesis or response genes. According to the MapMan analysis, auxin, cytokinin, ethylene, and gibberellin showed contrasting expressions in both the genotypes (**Figure 5B**). Similarly, in terms of the KEGG pathway, we have observed how the signaling pathways of these phytohormones were different. The present work confirms that auxin, cytokinin, ethylene, and gibberellin are the important regulators of the seed size in ricebean. Our results are also in accordance with those of previous studies in other species (Riefler et al., 2006; Heyl et al., 2012).

Candidate Gene Identification for Seed Development–Related Traits

The expression of a number of genes starting from the anthesis to early stages of maturity may have a crucial role in determining grain size and various other pod-related traits in pulses (Pazhamala et al., 2016). In this study, candidate genes for various traits such as days of flowering, pod shattering, seed per pod, seed size, 100-seed weight, and pod length were identified from the assembled transcriptome on the basis of sequence similarity search. In total, we identified 142 genes in ricebean belonging to development-related traits on the basis of similarity search (BLASTX) and e-value. Furthermore, the candidate genes were also characterized *in silico* on the basis of their domain analysis using Pfam software. Out of 142 genes, only 120 genes showed domain similarity with their hits. Therefore, we discarded 22 genes whose domain was not matched. Hence, according to our study, we found 120 candidate genes of ricebean belonging to different development-related traits (days of flowering, pod shattering, seed per pod, seed size, 100-seed weight, and seed length) (**Figure 9**; **Supplementary Table S10**).

In terms of pod development, seed size is a key determinant for the seed or grain yield in legume crops (Amkul et al., 2020). In ricebean, we found four candidate genes for seed size encoding: *histidine kinase 2*, *delta sterol reductase*, *phosphate transporter* (*PHO1*), and *WRKY* domain–containing protein (*WRKY 40*). These genes have already been reported to be involved in seed size. For example, *Vigun05g039600* (*PHO1*) has been reported to be a positive regulator of seed development that affects both the cell size and cell number (Lo et al., 2019). Similarly, *Vigun08g217000* which codes for *histidine kinase 2* has been identified as a potential candidate gene for improved organ size during cowpea domestication (Lonardi et al., 2019), and its *Arabidopsis* ortholog *AHK2* has been shown to regulate the seed size (Riefler et al., 2006; Bartrina et al., 2017). *Vigun11g191300* encoding a *delta (24)-sterol reductase* is an ortholog of the *Arabidopsis DIMINUTO* gene which has been shown to regulate cell elongation (Takahashi et al., 1995). In foxtail millet, *Loose Panicle1*–encoded *WRKY* transcription factor regulates the seed size by increasing the length and width of the seed (Xiang et al., 2017). Hence, these genes are the strong candidates as seed size is affected by multiple pathways.

On the other hand, for 100-seed weight, 29 candidate genes were identified corresponding to *expansin*, *cytokinin dehydrogenase*, *cytochrome P450*, and response regulatory domain containing protein. The significance of these genes as candidate loci related with the 100-seed weight is supported by the work done on *Arabidopsis*, where orthologs of the candidate genes in the cytokinin pathway have been shown, in transgenic studies, to regulate seed size and/or weight (Daele et al., 2012). Our findings are also in accordance with the common bean in which type-B regulators were found to be involved in the activation of downstream genes in the cytokinin pathway, and the genes encoding cytokinin dehydrogenase regulates the pathway by degrading active cytokinin (Hwang et al., 2012; Schmutz et al., 2014). Likewise, in *Arabidopsis*, *expansins* increased grain size and also improved grain production (Bae et al., 2014). Recent studies have also associated *expansins* with grain size and weight in wheat and tomato (Muñoz and Calderini, 2015; Brinton et al., 2017). *TaCYP78A3* in wheat and *CYP78A5* in *Arabidopsis* encodes the *cytochrome P450*, which positively correlates with seed size and seed weight (Ma et al., 2015; Tian et al., 2016b).

Like other development-related traits, flowering time is also an important trait because several agronomical traits such as quality of the grain and grain yield depend on flowering time. For days to flowering trait, we identified 21 candidate genes in our dataset encoding protein *Flowering Locus T-like*, *GIGANTEA-like*, *cryptochrome*, and transcription factors

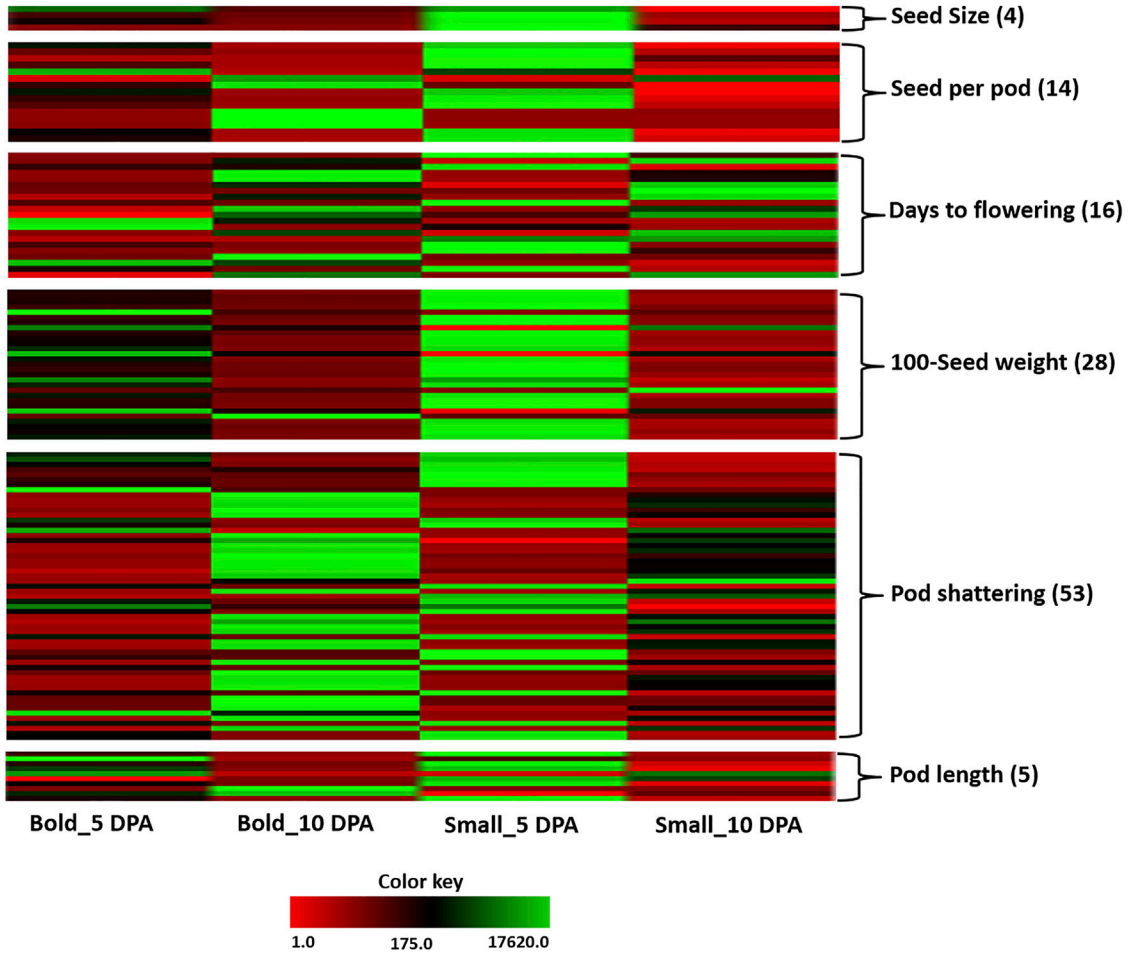


FIGURE 9 | Heat map representing the differential gene expression of the identified candidate genes for six traits including seed size, 100-seed weight, seed/pod, days to flowering, pod shattering, and pod length in bold and small genotypes at 5 DPA and 10 DPA.

such as *bHLH*, *ERF*, and *PIF-3*. Most of the candidate genes of days to flowering had high expression in case of 10 DPA, instead of 5 DPA. In rice, *florigen* is encoded by *RICE FLOWERING LOCUS T 1* (*RFT1*) and the orthologs of *Arabidopsis* FT and plays important role in heading date, influencing yield traits in rice (Tamaki et al., 2007; Komiya et al., 2009), whereas *GIGNANTEA*-like genes observed in the regulation of many genes which influence the circadian clock, blue light photoreceptor, and flowering time have also been reported in *Arabidopsis* (Hayama et al., 2003; Fornara et al., 2009). Similar to rice results, the *Flowering Locust T-like* in ricebean might help in the improvement of yield.

On the other hand, the number of seeds per pod might be useful for increasing the seed yield of ricebean. We identified 15 candidate genes for seeds per pod trait having annotations like *MAPK*, *NAC*, *MALE STERILE 5*, and *ABSCISIC ACID-INSENSITIVE 5*-like protein. *Vigun03g187300* (*ABA-insensitive 5-like protein 6*) is an ABA-responsive element (*ABRE*)–binding factor that regulates *ABRE*-dependent gene expression (Nakashima and Yamaguchi-Shinozaki, 2013). In *Arabidopsis*, ABA deficiency reportedly decreases the number of seeds per siliqua (Cheng et al., 2014). Hence, the higher expression of this gene in bold genotypes implies an increase in the number of seeds per pod that could result in the improvement of the ricebean yield. The *Vigun05g126900* gene, encoding *MALE STERILE 5*, was selected as a candidate gene in zombie pea (Amkul et al., 2020). In a previous study on *Arabidopsis*, mutations to *MALE STERILE 5* resulted in the development of "polyads" (i.e., tetrads with more than four pools of chromosomes following male meiosis) (Glover et al., 1998). Plants that are homozygous for the *MS5* recessive allele apparently revealed arrested growth and harvested empty siliques, whereas in plants that are heterozygous for *MS5*, siliqua elongation and seed set are less repressed (Glover et al., 1998). In case of the pod length, five candidate genes in ricebeans have been identified, mostly corresponding to the auxin response factor. *Glyma.07G134800*, an ortholog of *Arabidopsis*, was also associated with the auxin pathway (Jiang et al., 2018).

Furthermore, we have also identified a few candidate genes associated with pod shattering which is considered to be an undesirable agronomical trait. We identified maximum candidate genes (i.e., 57) for this trait in our ricebean study. Out of the 57 candidate genes, 18 genes encode transcription factors like *AP2/ERF*,

TABLE 3 | List of candidate genes matched with our MapMan results.

MapMan category	Candidate ricebean gene ID	Description	B5_B10	S5_S10	B5_S5	B10_S10	Trait
Amino acid metabolism	*cs5dtrinity_dn11557_c0_g1_i4*	Histidinol-phosphate aminotransferase	—	—	2.23	—	Seeds/pod
Cell wall organization	*bs10d1trinity_dn10624_c1_g9_i1*	Catalytic component CesA of cellulose synthase complex	2.34	—	—	—	Pod shattering
	ab10dtrinity_dn33078_c0_g1_i1	Alpha-class *expansin*	7.38	—	—	—	Seed weight
	ab4dtrinity_dn10598_c3_g1_i1	Alpha-class *expansin*	—	–3.88	—	—	Seed weight
	ab10dtrinity_dn30367_c2_g3_i1	Alpha-class *expansin*	—	–3.22	—	—	Seed weight
	bb4dtrinity_dn13171_c6_g7_i1	Alpha-class *expansin*	—	–3.13	—	—	Seed weight
	bb10dtrinity_dn16316_c5_g6_i2	Alpha-class *expansin*	—	–2.76	—	—	Seed weight
	bs5dtrinity_dn13044_c1_g3_i3	Alpha-class *expansin*	—	–2.82	—	—	Seed weight
	cb4dtrinity_dn14045_c9_g6_i1	Alpha-class *expansin*	—	–3.33	—	—	Seed weight
	cb4dtrinity_dn14247_c0_g4_i1	Alpha-class *expansin*	—	–4.18	—	—	Seed weight
	cs5dtrinity_dn11417_c0_g2_i1	Alpha-class *expansin*	—	–3.43	—	—	Seed weight
	cs5dtrinity_dn12484_c15_g1_i1	alpha-class *expansin*	—	–3.3	—	—	Seed weight
Lipid metabolism	*cs5dtrinity_dn11621_c1_g7_i3*	Sterol delta24 reductase	—	–2.09	—	—	Seed size
	ab10dtrinity_dn14585_c0_g1_i2	Dihydrolipoamide acetyltransferase component E2	5.3	—	—	—	seeds/pod
Nucleotide metabolism	*ab10dtrinity_dn30303_c0_g10_i1*	Uracil phosphoribosyltransferase (UPP)	—	—	–3.05	—	days to flowering
Nutrient uptake	*cs5dtrinity_dn18119_c0_g1_i1*	Phosphate transporter (PHO1)	—	–2.93	—	—	Seed size
Phytohormone action	*cs5dtrinity_dn10996_c2_g4_i3*	Receptor protein (AHK)	—	—	2.2	—	Seed size
	ab4dtrinity_dn16153_c0_g1_i2	Cytokinin dehydrogenase	—	–4.66	—	—	Seed weight
	cs5dtrinity_dn9587_c0_g1_i4	Steroid 22-alpha-hydroxylase (DWF4)	—	–3.24	—	—	Seed weight
	as10dtrinity_dn8390_c0_g1_i1	A-type cytokinin ARR response negative regulator	—	3.93	—	—	Seed weight
	cs5dtrinity_dn5774_c0_g1_i1	Cytokinin dehydrogenase	—	–3.51	—	—	Seed weight
Protein homeostasis	*bb4dtrinity_dn32604_c0_g1_i1*	Matrixin-type metalloprotease	—	–2.66	—	—	Pod Shattering
Redox homeostasis	*ab4dtrinity_dn10550_c1_g1_i11*	GDP-D-mannose-epimerase (GME)	—	–2.78	—	—	Seeds/pod

WRKY, and *NAC*, whereas the rest of the genes were involved in cellulose synthase and serine/threonine protein kinase. The candidate genes for pod shattering have also been identified in other legumes including *Vigun02g095200* (cellulose synthase), *Vigun03g306000* (*NAC* domain transcription factor), and zombi pea (Suanum et al., 2016; Lo et al., 2018; Takahashi et al., 2019; Amkul et al., 2020; Watcharatpong et al., 2020). In *Sorghum propinquum*, *WRKY* modulates the flower and seed development and lignin deposition, and it is also found to be involved in pod shattering (Tang et al., 2013). Recently, in rice, *AP2* transcription factor–coding gene *SHATTERING ABORTION1* (*SHAT1*) was observed having a crucial role in pod shattering. Two genes encoding *NAC* in *Vigna unguiculata* were found to be involved in cell wall biosynthesis and hence influencing the pod shattering (Zhou et al., 2012; Lo et al., 2018). The identification of pod shattering genes may reduce preharvest yield damages in ricebean, resulting in a more efficient yield. Thus, pod indehiscence may be a valuable trait during seed harvesting, making it a main concern during crop domestication (Amkul et al., 2020).

To support our findings related to candidate genes, we performed a comparative analysis of the identified candidate genes with our MapMan and KEGG pathway results. Out of 120 candidate genes, 23 genes matched with the MapMan results (**Table 3**). For example, the expression of eight candidate genes of 100-seed weight and seed size were only shown in the small genotype encoding *PHO1*, *cytokinin dehydrogenase*, A-type cytokinin *ARR* response negative regulator, etc. Similarly, for bold genotypes, only one gene, *aB10dtrinity_dn14585_c0_g1_i2*, for a seed was upregulated, revealing a high number of pods in bold genotypes as compared to the small genotype. On the other hand, in terms of time point, three genes (*cS5dtrinity_dn10996_c2_g4_i3*: seed size; *cS5dtrinity_dn11557_c0_g1_i4*: seeds/pod; and *aB10dtrinity_dn30303_c0_g10_i1*: days to flowering) were detected only at the first time point, that is, 5 DPA. Two genes (*aB10dtrinity_dn33078_c0_g1_i1*: 100-seed weight and *bS10d1trinity_dn10624_c1_g9_i1*: pod shattering) were found to be highly expressed only in bold genotypes, whereas nine genes encoding alpha class *expansins* were found to be downregulated, specifically in the small genotype.

Similarly, 16 candidate genes (auxin: 2; cytokinin: 8; ethylene: 5; GA: 1) were matched with the KEGG pathway results (**Table 4**). The matched genes were found to be associated with several seed development–related traits like pod length, days to flowering, 100-seed weight, seeds/pod, and pod shattering. All the genes were expressed in the small genotype, except two (*ab10dtrinity_dn29885_c1_g2_i2* and *bs10dtrinity_dn12088_c3_g6_i6*) which were expressed in bold genotypes corresponding to *MAPK*.

Simple Sequence Repeat Identification

In this study, we used the MISA Perl script (http://pgrc.ipk-gatersleben.de/misa) to detect the microsatellites. Of the 288,393 transcripts generated in this study, 14,663 contained an SSR totaling 201,517,181 bp. Out of these 14,663 sequences, 2,317 sequences had more than a single SSR, and 1,487 had SSRs of

TABLE 4 | List of candidate genes matched with our KEGG pathway results.

KEGG pathway	Ricebean candidate gene ID	Description	B5_B10	S5_S10	Trait
Auxin	*cb10dtrinity_dn16977_c3_g12_i1*	Auxin response factor	—	–2.09	Pod length
	cs5dtrinity_dn10719_c0_g1_i4	Auxin response factor	—	–2.15	Pod length
Cytokinin	*ab4dtrinity_dn16153_c0_g1_i2*	Cytokinin dehydrogenase 6–like	—	–4.66	Seed weight
	cs5dtrinity_dn10983_c0_g1_i2	Two-component response regulator–like APRR1 isoform X4 (CCT motif, rec)	—	–2.19	Days to flowering
	as5dtrinity_dn1065_c0_g1_i1	HPt domain–containing protein	—	2.17	Seed weight
	as10dtrinity_dn8390_c0_g1_i1	Response regulatory domain–containing protein (type A)	—	3.93	Seed weight
	bs10d1trinity_dn9417_c0_g2_i1	HPt domain–containing protein	—	2.26	Seed weight
	cb10dtrinity_dn35628_c0_g1_i1	HPt domain–containing protein	—	2.71	Seed weight
	cs5dtrinity_dn4103_c0_g1_i1	Cytokinin hydroxylase–like	—	–2.85	Seed weight
	cs5dtrinity_dn5774_c0_g1_i1	Cytokinin dehydrogenase 6–like	—	–3.51	Seed weight
Ethylene	*bb10dtrinity_dn16202_c0_g2_i1*	Ethylene-responsive transcription factor *RAP2-7–like* isoform *X2*	—	2.52	Days to flowering
	cb4dtrinity_dn13190_c1_g1_i4	*AP2/ERF* domain–containing protein	—	3.14	Pod Shattering
	ab10dtrinity_dn29885_c1_g2_i2	Mitogen-activated protein kinase	9.72	—	Seed/pod
	bs10dtrinity_dn12088_c3_g6_i6	Mitogen-activated protein kinase	4.21	—	Seed/pod
	bs10dtrinity_dn12088_c3_g6_i6	Mitogen-activated protein kinase	—	5.55	Seed/pod
GA	*ab4dtrinity_dn7670_c0_g1_i4*	Transcription factor *PIF3*-like isoform	—	2.06	Days to flowering

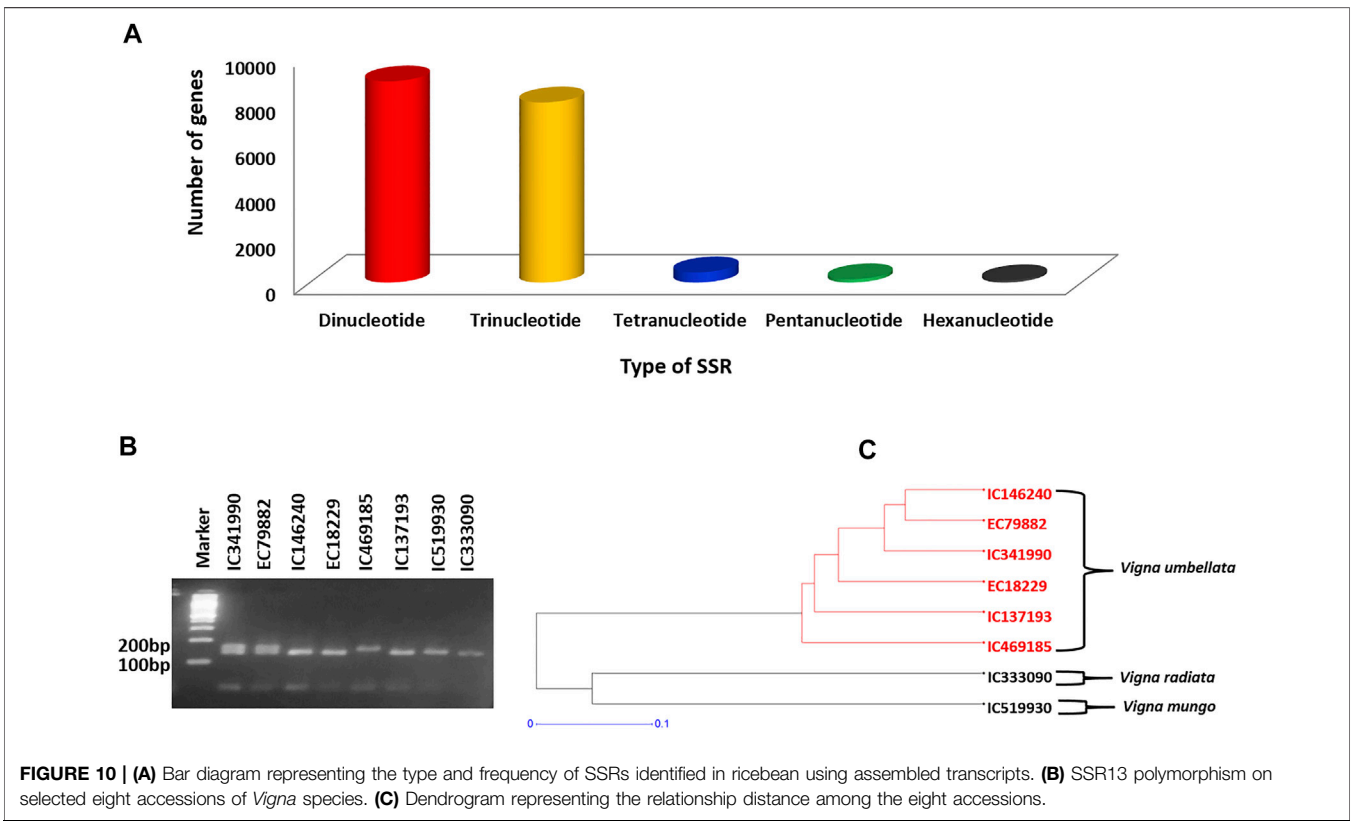


FIGURE 10 | (A) Bar diagram representing the type and frequency of SSRs identified in ricebean using assembled transcripts. **(B)** SSR13 polymorphism on selected eight accessions of *Vigna* species. **(C)** Dendrogram representing the relationship distance among the eight accessions.

different motifs (compound SSR). Dinucleotide repeat motifs were the most abundant among the five types of motifs, totaling 8,866 (50.67%). The second most abundant were trinucleotides totaling 7,938 (45.36%), followed by 448 tetranucleotides (2.56%), 145 pentanucleotides (0.82%), and 100 hexanucleotide motifs (0.57%) (**Figure 10A**). Similar results have been reported in the previous transcriptome published for ricebean varieties (Chen et al., 2016) as well as for other legume species including mungbean (Tian et al., 2016b), adzuki bean (Chankaew et al., 2014; Chen et al., 2015b), cowpea (Gupta et al., 2010), and chickpea (Choudhary et al., 2008).

TABLE 5 | List of 26 SSR markers that showed polymorphism in a set of eight accessions of *Vigna* species.

Transcript ID	Marker	Forward primer (5'->3')	Reverse primer (5'->3')	Annealing temp (°C)	Repeat motif	Allele size (bp)	No. of alleles
BB7DTRINITY_DN9136_c0_g1_i1	SSR2	ATGATCGGACACTAGGAGAC	TTGGCCAATGTCTATTTGA	54	ATT(18)	150–160	2
BS7D1TRINITY_DN10554_c1_g1_i1	SSR3	ACGCACAGTTTCATGGTTA	ACAATCTTCAACCACACTCC	55	GAA(19)	100–130	4
BS4DTRINITY_DN12307_c0_g4_i3	SSR4	CAAACCCACTAACCCAAGTA	ATGAAAATGCAAACACACTG	55	TAA(17)	140–150	2
CB4DTRINITY_DN13434_c0_g3_i2	SSR7	ATTCCCAGCTTAGGAGAAAC	TGGATTTGTTCTTAATGGTG	55	ATA(18)	140–170	2
AB4DTRINITY_DN10798_c2_g4_i2	SSR8	GTTATTGGAATGGAAGAGCA	CTTCCGACAACAATTCCTT	55	GAA(16)	120–140	3
AS4DTRINITY_DN8490_c0_g1_i1	SSR9	CAACCGGGTAGAGAAAAGTA	CTACCAAGTTGCTTGCTTCT	54	AAT(22)	210–220	2
CB7DTRINITY_DN17452_c5_g1_i3	SSR11	ATGGGTTTCCTATGAATTTG	GCTAATGACTCTGCTGTTCC	55	TAA(11)	140–150	2
AB4DTRINITY_DN10727_c2_g3_i1	SSR12	GCTAATGACTCTGCTGTTCC	ATGGGTTTCCTATGAATTTG	55	TTA(11)	140–150	2
AS4DTRINITY_DN11798_c1_g2_i4	SSR13	GGGAAAATGTTACGGAGTTC	GTTTTCCCACCACAACTAAC	56	TGG(12)	120–150	2
BS4DTRINITY_DN12006_c0_g8_i4	SSR14	CTGGGAAACTGAGCAGATAG	CAGATAGTTGCAATAGCTTGAA	55	TAT(12)	170–190	3
CB7DTRINITY_DN16904_c1_g1_i5	SSR15	TTAGAATTTCCGTTGCTACC	CCCTGAAAGAAGTTTGGAAT	55	TAT(12)	170–180	2
BB7DTRINITY_DN17092_c1_g1_i3	SSR16	TTCACCTCTGACTGATCACA	CAAGTCTAATGCATCCACCT	55	GAT(13)	160–180	3
BS7DTRINITY_DN11614_c2_g22_i1	SSR18	CTGGGAAACTGAGCAGATAG	CAGATAGTTGCAATAGCTTGAA	55	TAT(12)	190–200	2
AS4DTRINITY_DN11028_c4_g1_i3	SSR24	CTGGGAAACTGAGCAGATAG	CAGATAGTTGCAATAGCTTGAA	55	TAT(12)	190–200	2
BS4DTRINITY_DN11934_c2_g1_i5	SSR28	TTCCACGTTCTCACTCTCTT	GGAATCCATTACTGTGAACG	55	TC(37)	100–130	4
BS7DTRINITY_DN11790_c6_g2_i8	SSR30	CTCTTCTTAGAGCCAAACCA	ACGCCATGTGTATGAAGATT	55	CT(36)	100–120	3
BS7DTRINITY_DN3955_c0_g1_i2	SSR31	CGTTTCCTAAGCTTCCTTTA	GAGAAGCGAAGAAGAAAGGT	55	TC(35)	100–130	4
AB7DTRINITY_DN28298_c2_g1_i4	SSR32	CTACCAGTGGGTTCGTTTAC	TCTCTCTTCTCCCCTTAACC	55	GA(32)	130–160	4
AB4DTRINITY_DN10313_c1_g2_i7	SSR35	CACCCTAACCTCATTCTCAG	GACAGCAAGAAGGAGAGAGA	54	CT(48)	100–110	2
CB4DTRINITY_DN14022_c6_g1_i1	SSR37	TCACAAAACCCTAAAACTCG	GGCAGTGTGAAAGAAAGAGA	55	TC(28)	200–220	3
CS4DTRINITY_DN11268_c2_g5_i1	SSR38	AATGTGCTCTTCTTGTTGCT	ACCGATGGAATAACCAAAC	55	TC(28)	100–110	2
BB7DTRINITY_DN13571_c0_g1_i4	SSR39	TTGTGGATATAAACCCAACC	GCTCCTCCGCTCTTCTATTA	56	AG(28)	120–150	4
BB7DTRINITY_DN16958_c4_g1_i5	SSR40	TGATTAACTGGGTTCTCTGC	TTCTACAACCACCCAATCTC	55	AT(28)	110–130	3
BS7DTRINITY_DN11905_c0_g1_i6	SSR41	GGGAGTATCCAAAGAAACAA	AATCCACACACAAATGTGAA	54	TC(30)	110–120	2
BB4DTRINITY_DN12047_c0_g1_i7	SSR42	GGAATCCATTACTGTGAACG	TTCCACGTTCTCACTCTCTT	55	GA(30)	110–140	4
CB7DTRINITY_DN16920_c1_g4_i9	SSR45	GTGGGTAACTATGCCCTAAGT	GGTGAGTGGATGTGAGAAAG	55	TC(27)	110–120	2

The number of the given repeat unit of SSRs ranged from 5 to >10, and as the number of repeat units increased, the frequency of the given SSR structure progressively decreased (**Supplementary Table S11**). As for the two most abundant repeat motif types (di- and trinucleotides), the frequency of the AG/CT motif type accounted for 17.41% in dinucleotide repeat motifs, and the frequency of GAA/TTC was the most abundant motif type in the trinucleotide, accounting for 6.3%. A previous study on adzuki bean also showed a high frequency of AG motifs in dinucleotides (Chankaew et al., 2014).

Simple Sequence Repeat Validation

To determine the polymorphism level of the identified EST-SSRs, the randomly selected 50 SSRs were evaluated in eight accessions of *Vigna* species including *V. umbellata* (6), *V. mungo* (1), and *V. radiata* (1) (**Supplementary Table S12**). From 50 pairs, 43 were successfully

amplified, while seven pairs were not able to generate a PCR product (**Supplementary Table S13**). More than 85% of the SSR markers were successfully amplified, suggesting that the quality of our assembled transcripts was very high. The annealing temperatures of the primers ranged between 54 and 56°C. Out of these 43 SSR primer pairs, 26 pairs showed polymorphism (dinucleotide: 12, trinucleotide: 14) and the rest were monomorphic (**Figure 10B**; **Table 5**). A high polymorphism level (60.46%) of ricebean EST-SSRs was observed in the selected set of eight accessions which was higher than that from previous reports in other legume species including the chickpea (47.3%) (Nayak et al., 2010), mungbean (33%) (Chen et al., 2015b), black gram (58.2%) (Souframanien and Reddy, 2015), and adzuki bean (7.6%) (Chen et al., 2015a) while lower than common bean (71.3%) (Hanai et al., 2007), whereas when we considered only ricebean genotypes (six accessions), only 34.88% SSR markers were found polymorphic. We have also checked the cross-species transferability pattern and found that the transferability of ricebean–derived SSR markers was higher in *V. radiata* (73.08%) than in *V. mungo* (50%). Various studies depicted the importance of SSR cross-transferability in *Vigna* species including ricebean, mungbean, and cowpea (Pattanayak et al., 2019). Furthermore, the genetic distance among the accessions was determined, and we found two clusters, with six (*V. umbellata*) in the first cluster and two (*V. mungo* and *V. radiata*) in the second cluster, respectively (**Figure 10C**).

SSR molecular markers on the basis of transcriptomes have become more promising and useful because of their high cross-species transferability, their high rate of amplification, and being reasonably cheap as compared with the SSR markers of non-transcribed regions (Hansen et al., 2008; Rai et al., 2013). Moreover, since they can easily expose variance in the expressed portion of the genome, it is possible to evaluate marker–trait association (MTA) and specific genomic regions stating important physio-agronomic traits (Kalia et al., 2011).

CONCLUSION

The transcriptomic analysis in our study provided detailed insights into molecular processes and candidate genes controlling seed size and other seed development–related traits in ricebean. The MapMan and KEGG analyses confirmed that the phytohormone signaling pathways varied in both the contrasting expressions taken in this study and can therefore be the regulators of seed size as well as other seed development–related traits in ricebean. We hypothesize that the auxin, cytokinin, ethylene, and gibberellin signaling pathways interact cooperatively with one another, thereby modulating the expression of genes of seed development–related traits. Further research is required to identify key regulators/genes in determining seed size.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI under accession number PRJNA765494.

AUTHOR CONTRIBUTIONS

SV contributed to investigation, formal analysis, and writing—original draft. SM helped with data curation, investigation, formal analysis, writing—original draft, and writing—review, and editing. GC contributed to resources and writing—review and editing. DW assisted with analysis and writing—review and editing. SP, DC, GP, DM, DJ, MS, and KS assisted with writing—review and editing. AS involved in conceptualization, supervision, resources, and writing—review and editing.

FUNDING

The financial support was received from the Department of Biotechnology under project: BT/Ag/Network/Pulses-1/2017-18.

ACKNOWLEDGMENTS

The authors are also thankful to the Head of Division, Genomic Resources and Director, ICAR-NBPGR, New Delhi, for extending laboratory facilities for this work. The authors would also like to acknowledge the HPC facility at IACR-NBPGR for this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.791355/full#supplementary-material

REFERENCES

Amkul, K., Somta, P., Laosatit, K., and Wang, L. (2020). Identification of QTLs for Domestication-Related Traits in Zombi Pea [*Vigna vexillata* (L.) A. Rich], a Lost Crop of Africa. *Front. Genet.* 11, 803. doi:10.3389/FGENE.2020.00803

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

Bae, J. M., Kwak, M. S., Noh, S. A., Oh, M. J., Kim, Y. S., and Shin, J. S. (2014). Overexpression of Sweetpotato Expansin cDNA (IbEXP1) Increases Seed Yield in Arabidopsis. *Transgenic Res.* 23, 657–667. doi:10.1007/S11248-014-9804-1

Bai, L., Chen, Q., Jiang, L., Lin, Y., Ye, Y., Liu, P., et al. (2019). Comparative Transcriptome Analysis Uncovers the Regulatory Functions of Long Noncoding RNAs in Fruit Development and Color Changes of *Fragaria pentaphylla*. *Hortic. Res.* 61 6, 1. doi:10.1038/s41438-019-0128-4

Bartrina, I., Jensen, H., Novák, O., Strnad, M., Werner, T., and Schmülling, T. (2017). Gain-of-Function Mutants of the Cytokinin Receptors AHK2 and AHK3 Regulate Plant Organ Size, Flowering Time and Plant Longevity. *Plant Physiol.* 173, 1783–1797. doi:10.1104/PP.16.01903

Bartrina, I., Otto, E., Strnad, M., Werner, T., and Schmülling, T. (2011). Cytokinin Regulates the Activity of Reproductive Meristems, Flower Organ Size, Ovule Formation, and Thus Seed Yield in *Arabidopsis thaliana*. *Plant Cell* 23, 69–80. doi:10.1105/TPC.110.079079

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A Gene Expression Atlas of the Model Legume *Medicago truncatula*. *Plant J.* 55, 504–513. doi:10.1111/J.1365-313X.2008.03519.X

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertová, D., Jürgens, G., et al. (2003). Local, Efflux-Dependent Auxin Gradients as a Common Module for Plant Organ Formation. *Cell* 115, 591–602. doi:10.1016/S0092-8674(03)00924-3

Bernardi, J., Li, Q.-B., Gao, Y., Zhao, Y., Battaglia, R., Marocco, A., et al. (2016). The Auxin-Deficient Defective Kernel18 (dek18) Mutation Alters the Expression of Seed-Specific Biosynthetic Genes in Maize. *J. Plant Growth Regul.* 353 (35), 770–777. doi:10.1007/S00344-016-9581-6

Blakeslee, J. J., Bandyopadhyay, A., Peer, W. A., Makam, S. N., and Murphy, A. S. (2004). Relocalization of the PIN1 Auxin Efflux Facilitator Plays a Role in Phototropic Responses. *Plant Physiol.* 134, 28–31. doi:10.1104/PP.103.031690

Borisjuk, L., Rolletschek, H., Radchuk, R., Weschke, W., Wobus, U., and Weber, H. (2004). Seed Development and Differentiation: A Role for Metabolic Regulation. *Plant Biol.* 6, 375–386. doi:10.1055/s-2004-817908

Brinton, J., Simmonds, J., Minter, F., Leverington-Waite, M., Snape, J., and Uauy, C. (2017). Increased Pericarp Cell Length Underlies a Major Quantitative Trait Locus for Grain Weight in Hexaploid Wheat. *New Phytol.* 215, 1026–1038. doi:10.1111/NPH.14624

Bromley, J. R., Warnes, B. J., Newell, C. A., Thomson, J. C. P., James, C. M., Turnbull, C. G. N., et al. (2014). A Purine Nucleoside Phosphorylase in *Solanum tuberosum* L. (Potato) with Specificity for Cytokinins Contributes to the Duration of Tuber Endodormancy. *Biochem. J.* 458, 225–237. doi:10.1042/BJ20130792

Calzadilla, P. I., Maiale, S. J., Ruiz, O. A., and Escaray, F. J. (2016). Transcriptome Response Mediated by Cold Stress in *Lotus japonicus*. *Front. Plant Sci.* 0, 374. doi:10.3389/FPLS.2016.00374

Chankaew, S., Isemura, T., Isobe, S., Kaga, A., Tomooka, N., Somta, P., et al. (2014). Detection of Genome Donor Species of Neglected Tetraploid Crop Vigna Reflexo-Pilosa (Créole Bean), and Genetic Structure of Diploid Species Based on Newly Developed EST-SSR Markers from Azuki Bean (*Vigna angularis*). *PLoS One* 9, e104990. doi:10.1371/JOURNAL.PONE.0104990

Chapman, E. J., and Estelle, M. (2009). Mechanism of Auxin-Regulated Gene Expression in Plants. *Annu. Rev. Genet.* 43, 265–285. doi:10.1146/annurev-genet-102108-134148

Chen, H., Chen, X., Tian, J., Yang, Y., Liu, Z., Hao, X., et al. (2016). Development of Gene-Based SSR Markers in Rice Bean (*Vigna umbellata* L.) Based on Transcriptome Data. *PLoS One* 11, e0151040. doi:10.1371/JOURNAL.PONE.0151040

Chen, H., Liu, L., Wang, L., Wang, S., Somta, P., and Cheng, X. (2015a). Development and Validation of EST-SSR Markers from the Transcriptome of Adzuki Bean (*Vigna angularis*). *PLoS One* 10, 1–14. doi:10.1371/journal.pone.0131939

Chen, H., Wang, L., Liu, X., Hu, L., Wang, S., and Cheng, X. (2017). De Novo transcriptomic Analysis of Cowpea (*Vigna unguiculata* L. Walp) for Genic SSR Marker Development. *BMC Genet.* 18, 1–12. doi:10.1186/S12863-017-0531-5

Chen, H., Wang, L., Wang, S., Liu, C., Blair, M. W., and Cheng, X. (2015b). Transcriptome Sequencing of Mung Bean (*Vigna radiate* L.) Genes and the Identification of EST-SSR Markers. *PLoS One* 10, e0120273. doi:10.1371/JOURNAL.PONE.0120273

Cheng, Z. J., Zhao, X. Y., Shao, X. X., Wang, F., Zhou, C., Liu, Y. G., et al. (2014). Abscisic Acid Regulates Early Seed Development in Arabidopsis by ABI5-Mediated Transcription of Short Hypocotyl Under Blue1. *Plant Cell* 26, 1053–1068. doi:10.1105/TPC.113.121566

Choudhary, S., Sethy, N. K., Shokeen, B., and Bhatia, S. (2008). Development of Chickpea EST-SSR Markers and Analysis of Allelic Variation across Related Species. *Theor. Appl. Genet.* 118, 591–608. doi:10.1007/S00122-008-0923-Z

Christian, M., Steffens, B., Schenck, D., Burmester, S., Böttger, M., and Lüthen, H. (2006). How Does Auxin Enhance Cell Elongation? Roles of Auxin-Binding Proteins and Potassium Channels in Growth Control. *Plant Biol.* 8, 346–352. doi:10.1055/S-2006-923965

Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A Developmental Transcriptome Map for Allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* 0, 1446. doi:10.3389/FPLS.2016.01446

Daele, I. V., Gonzalez, N., and Ilse, I. V. P. (2012). A Comparative Study of Seed Yield Parameters in *Arabidopsis thaliana* Mutants and Transgenics. *Plant Biol. J.* 10, 488–500. doi:10.1111/j.1467-7652.2012.00687.x

Desbrosses, G. J., and Stougaard, J. (2011). Root Nodulation: A Paradigm for How Plant-Microbe Symbiosis Influences Host Developmental Pathways. *Cell Host Microbe* 10, 348–358. doi:10.1016/J.CHOM.2011.09.005

Fornara, F., Panigrahi, K. C. S., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J. A., et al. (2009). Arabidopsis DOF Transcription Factors Act Redundantly to Reduce CONSTANS Expression and are Essential for a Photoperiodic Flowering Response. *Dev. Cell* 17, 75–86. doi:10.1016/J.DEVCEL.2009.06.015

Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., et al. (2011). Gene Discovery and Tissue-specific Transcriptome Analysis in Chickpea with Massively Parallel Pyrosequencing and Web Resource Development. *Plant Physiol.* 156, 1661–1678. doi:10.1104/pp.111.178616

Geng, X., Dong, N., Wang, Y., Li, G., Wang, L., Guo, X., et al. (2018). Correction: RNA-Seq Transcriptome Analysis of the Immature Seeds of Two Brassica Napus Lines with Extremely Different Thousand-Seed Weight to Identify the Candidate Genes Related to Seed Weight. *PLoS ONE* 13, 1. doi:10.1371/journal.pone.0218914

Glover, J., Grelon, M., Craig, S., Chaudhury, A., and Dennis, E. (1998). Cloning and Characterization of MS5 from Arabidopsis : A Gene Critical in Male Meiosis. *Plant J.* 15, 345–356. doi:10.1046/J.1365-313X.1998.00216.X

González, A. M., Yuste-Lisbona, F. J., Weller, J., Vander Schoor, J. K., Lozano, R., and Santalla, M. (2021). Characterization of QTL and Environmental Interactions Controlling Flowering Time in Andean Common Bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 11, 599462. doi:10.3389/FPLS.2020.599462

Gu, D., Andreev, K., and Dupre, E. M. (2021). Major Trends in Population Growth Around the World. *China CDC Wkly* 3, 604–613. doi:10.46234/ccdcw2021.160

Guo, B., Wei, Y., Xu, R., Lin, S., Luan, H., Lv, C., et al. (2016). Genome-Wide Analysis of APETALA2/Ethylene-Responsive Factor (AP2/ERF) Gene Family in Barley (*Hordeum vulgare* L.). *PLoS One* 11, e0161322. doi:10.1371/JOURNAL.PONE.0161322

Gupta, P., Singh, R., Malhotra, S., Boora, K. S., and Singal, H. R. (2010). Characterization of Seed Storage Proteins in High Protein Genotypes of Cowpea [*Vigna unguiculata* (L.) Walp.]. *Physiol. Mol. Biol. Plants* 161, 53–58. doi:10.1007/S12298-010-0007-9

Hanai, L. R., De Campos, T., Camargo, L. E. A., Benchimol, L. L., De Souza, A. P., Melotto, M., et al. (2007). Development, Characterization, and Comparative Analysis of Polymorphism at Common Bean SSR Loci Isolated from Genic and Genomic Sources. *Genome* 50, 266–277. doi:10.1139/G07-007

Hansen, B. G., Halkier, B. A., and Kliebenstein, D. J. (2008). Identifying the Molecular Basis of QTLs: eQTLs Add a New Dimension. *Trends Plant Sci.* 13, 72–77. doi:10.1016/J.TPLANTS.2007.11.008

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of Photoperiodic Control Pathways Produces Short-Day Flowering in Rice. *Nature* 422, 719–722. doi:10.1038/nature01549

Hayashi, K. (2012). The Interaction and Integration of Auxin Signaling Components. *Plant Cell Physiol.* 53, 965–975. doi:10.1093/PCP/PCS035

Hentrich, M., Sánchez-Parra, B., Alonso, M.-M. P., Loba, V. C., Carrillo, L., Vicente-Carbajosa, J., et al. (2013). YUCCA8 and YUCCA9 Overexpression Reveals a Link between Auxin Signaling and Lignification through the Induction of Ethylene Biosynthesis, *Plant Signal. Behav.* 8, e26363. doi:10.4161/psb.26363

Herridge, R. P., Day, R. C., Baldwin, S., and Macknight, R. C. (2011). Rapid Analysis of Seed Size in Arabidopsis for Mutant and QTL Discovery. *Plant Methods* 71 (7), 1. doi:10.1186/1746-4811-7-3

Heyl, A., Riefler, M., Romanov, G. A., and Schmülling, T. (2012). Properties, Functions and Evolution of Cytokinin Receptors. *Eur. J. Cell Biol.* 91, 246–256. doi:10.1016/J.EJCB.2011.02.009

Heyl, A., and Schmülling, T. (2003). Cytokinin Signal Perception and Transduction. *Curr. Opin. Plant Biol.* 6, 480–488. doi:10.1016/S1369-5266(03)00087-6

Hiz, M. C., Canher, B., Niron, H., and Turet, M. (2014). Transcriptome Analysis of Salt Tolerant Common Bean (*Phaseolus vulgaris* L.) under Saline Conditions. *PLoS One* 9, e92598. doi:10.1371/JOURNAL.PONE.0092598

Hornett, E. A., and Wheat, C. W. (2012). Quantitative RNA-Seq Analysis in Non-Model Species: Assessing Transcriptome Assemblies as a Scaffold and the Utility of Evolutionary Divergent Genomic Reference Species. *BMC Genom.* 13, 361. doi:10.1186/1471-2164-13-361

Huang, J., Deng, J., Shi, T., Chen, Q., Liang, C., Meng, Z., et al. (2017). Global Transcriptome Analysis and Identification of Genes Involved in Nutrients Accumulation during Seed Development of rice Tartary Buckwheat (*Fagopyrum tararicum*). *Sci. Rep.* 71, 1–14. doi:10.1038/s41598-017-11929-z

Huang, R., Jiang, L., Zheng, J., Wang, T., Wang, H., Huang, Y., et al. (2013). Genetic Bases of rice Grain Shape: So many Genes, So Little Known. *Trends Plant Sci.* 18, 218–226. doi:10.1016/J.TPLANTS.2012.11.001

Hwang, I., Sheen, J., and Müller, B. (2012). Cytokinin Signaling Networks. *Annu. Rev. Plant Biol.* 63, 353. doi:10.1146/annurev-arplant-042811-105503

Jain, M., Kaur, N., Garg, R., Thakur, J. K., Tyagi, A. K., and Khurana, J. P. (20052005). Structure and Expression Analysis of Early Auxin-Responsive Aux/IAA Gene Family in Rice (*Oryza sativa*). *Funct. Integr. Genom.* 61 6, 47–59. doi:10.1007/S10142-005-0005-0

Jiang, H., Li, Y., Qin, H., Li, Y., Qi, H., Li, C., et al. (2018). Identification of Major QTLs Associated with First Pod Height and Candidate Gene Mining in Soybean. *Front. Plant Sci.* 0, 1280. doi:10.3389/FPLS.2018.01280

Jones, S. I., and Vodkin, L. O. (2013). Using RNA-Seq to Profile Soybean Seed Development from Fertilization to Maturity. *PLoS One* 8. doi:10.1371/journal.pone.0059270

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., and Dhawan, A. K. (2011). Microsatellite Markers: An Overview of the Recent Progress in Plants. *Euphytica* 177, 309–334. doi:10.1007/s10681-010-0286-9

Kalluri, U. C., DiFazio, S. P., Brunner, A. M., and Tuskan, G. A. (2007). Genome-Wide Analysis of Aux/IAA and ARF Gene Families in Populus trichocarpa. *BMC Plant Biol.* 71 (7), 1. doi:10.1186/1471-2229-7-59

Katoch, R. (2013). Nutritional Potential of Rice Bean (*Vigna umbellata*): An Underutilized Legume. *J. Food Sci.* 78, 8–16. doi:10.1111/j.1750-3841.2012.02989.x

Kaul, T., Eswaran, M., Thangaraj, A., Meyyazhagan, A., Nehra, M., Raman, N. M., et al. (2019). Rice Bean (*Vigna umbellata*) Draft Genome Sequence: Unravelling the Late Flowering and Unpalatability Related Genomic Resources for Efficient Domestication of This Underutilized Crop. *bioRxiv*, 816595. doi:10.1101/816595

Komiya, R., Yokoi, S., and Shimamoto, K. (2009). A Gene Network for Long-Day Flowering Activates RFT1 Encoding a mobile Flowering Signal in Rice. *Development* 136, 3443–3450. doi:10.1242/DEV.040170

Kumar, A., Patel, J. S., Meena, V. S., and Srivastava, R. (2019). Recent Advances of PGPR Based Approaches for Stress Tolerance in Plants for Sustainable Agriculture. *Biocatal. Agric. Biotechnol.* 20, 101271. doi:10.1016/J.BCAB.2019.101271

Leyser, O. (2003). Molecular Genetics of Auxin Signaling. *Annu. Rev. Plant Biol.* 53, 377. doi:10.1146/annurev.arplant.53.100301.135227

Li, H., Lv, Q., Deng, J., Huang, J., Cai, F., Liang, C., et al. (2019a). Transcriptome Analysis Reveals Key Seed-Development Genes in Common Buckwheat (*Fagopyrum esculentum*). *Int. J. Mol. Sci.* 20, 4303. doi:10.3390/ijms20174303

Li, N., Wang, Y., Lu, J., and Liu, C. (2019b). Genome-Wide Identification and Characterization of the ALOG Domain Genes in Rice. *Int. J. Genomics* 2019, 2146391. doi:10.1155/2019/2146391

Liu, N., Zhang, G., Xu, S., Mao, W., Hu, Q., and Gong, Y. (2015). Comparative Transcriptomic Analyses of Vegetable and Grain Pea (*Pisum sativum* L.) Seed Development. *Front. Plant Sci.* 0, 1039. doi:10.3389/FPLS.2015.01039

Liu, T., Zhu, S., Tang, Q., Chen, P., Yu, Y., and Tang, S. (2013). *De Novo* Assembly and Characterization of Transcriptome Using Illumina Paired-End Sequencing and Identification of CesA Gene in Ramie (*Boehmeria nivea* L. Gaud). *BMC Genomics* 14, 125. doi:10.1186/1471-2164-14-125

Lo, S., Muñoz-Amatriaín, M., Boukar, O., Herniter, I., Cisse, N., Guo, Y.-N., et al. (2018). Identification of QTL Controlling Domestication-Related Traits in Cowpea (*Vigna unguiculata* L. Walp). *Sci. Rep.* 81 8, 1–9. doi:10.1038/s41598-018-24349-4

Lo, S., Muñoz-Amatriaín, M., Hokin, S. A., Cisse, N., Roberts, P. A., Farmer, A. D., et al. (2019). A Genome-wide Association and Meta-Analysis Reveal Regions Associated with Seed Size in Cowpea [*Vigna unguiculata* (L.) Walp]. *Theor. Appl. Genet.* 132, 3079–3087. doi:10.1007/S00122-019-03407-Z

Lohar, D. P., Schaff, J. E., Laskey, J. G., Kieber, J. J., Bilyeu, K. D., and Bird, D. M. (2004). Cytokinins Play Opposite Roles in Lateral Root Formation, and Nematode and Rhizobial Symbioses. *Plant J.* 38, 203–214. doi:10.1111/J.1365-313X.2004.02038.X

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. doi:10.1111/TPJ.14349

Ma, M., Wang, Q., Li, Z., Cheng, H., Li, Z., Liu, X., et al. (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

Mahto, A., Mathew, I. E., and Agarwal, P. (2017). Decoding the Transcriptome of Rice Seed During Development. *Adv. Seed Biol.*, 25 doi:10.5772/INTECHOPEN.70574

Möller, B., and Weijers, D. (2009). Auxin Control of Embryo Patterning. *Cold Spring Harb. Perspect. Biol.* 1, a001545. doi:10.1101/CSHPERSPECT.A001545

Muñoz, M., and Calderini, D. F. (2015). Volume, Water Content, Epidermal Cell Area, and XTH5 Expression in Growing Grains of Wheat across Ploidy Levels. *F. Crop Res.* 173, 30–40. doi:10.1016/J.FCR.2014.12.010

Nakashima, K., and Yamaguchi-Shinozaki, K. (2013). ABA Signaling in Stress-Response and Seed Development. *Plant Cell Rep.* 327 (32), 959–970. doi:10.1007/S00299-013-1418-1

Nakayama, A., Park, S., Zheng-Jun, X., Nakajima, M., and Yamaguchi, I. (2002). Immunohistochemistry of Active Gibberellins and Gibberellin-Inducible α-Amylase in Developing Seeds of Morning Glory. *Plant Physiol.* 129, 1045–1053. doi:10.1104/PP.010921

Nayak, S. N., Zhu, H., Varghese, N., Datta, S., Choi, H. K., Horres, R., et al. (2010). Integration of Novel SSR and Gene-Based SNP Marker Loci in the Chickpea Genetic Map and Establishment of New Anchor Points with Medicago Truncatula Genome. *Theor. Appl. Genet.* 120, 1415–1441. doi:10.1007/s00122-010-1265-1

Nelson, S. K., Ariizumi, T., and Steber, C. M. (2017). Biology in the Dry Seed: Transcriptome Changes Associated with Dry Seed Dormancy and Dormancy Loss in the Arabidopsis GA-insensitive Sleepy1-2 Mutant. *Front. Plant Sci.* 8, 1–21. doi:10.3389/fpls.2017.02158

Ozga, J. A., Yu, J., and Reinecke, D. M. (2003). Pollination-, Development-, and Auxin-specific Regulation of Gibberellin 3β-Hydroxylase Gene Expression in Pea Fruit and Seeds. *Plant Physiol.* 131, 1137–1146. doi:10.1104/PP.102.015974

Parry, G., Calderon-Villalobos, L. I., Prigge, M., Peret, B., Dharmasiri, S., Itoh, H., et al. (2009). Complex Regulation of the TIR1/AFB Family of Auxin Receptors. *Proc. Natl. Acad. Sci.* 106, 22540–22545. doi:10.1073/PNAS.0911967106

Pattanayak, A., Roy, S., Sood, S., Iangrai, B., Banerjee, A., Gupta, S., et al. (2019). Rice Bean: A Lesser Known Pulse with Well-Recognized Potential. *Planta* 250, 873–890. doi:10.1007/s00425-019-03196-1

Pazhamala, L. T., Agarwal, G., Bajaj, P., Kumar, V., Kulshreshtha, A., Saxena, R. K., et al. (2016). Deciphering Transcriptional Programming during Pod and Seed Development Using RNA-Seq in Pigeonpea (*Cajanus cajan*). *PLoS One* 11, e0164959. doi:10.1371/JOURNAL.PONE.0164959

Peng, L., Qian, L., Wang, M., Liu, W., Song, X., Cheng, H., et al. (2021). Comparative Transcriptome Analysis during Seeds Development between Two Soybean Cultivars. *PeerJ* 9, 1–20. doi:10.7717/peerj.10772

Pradhan, S., Bandhiwal, N., Shah, N., Kant, C., Gaur, R., and Bhatia, S. (2014). Global Transcriptome Analysis of Developing Chickpea (*Cicer arietinum* L.) Seeds. *Front. Plant Sci.* 5, 1–14. doi:10.3389/fpls.2014.00698

Qi, Z., Zhang, Z., Wang, Z., Yu, J., Qin, H., Mao, X., et al. (2018). Meta-analysis and Transcriptome Profiling Reveal Hub Genes for Soybean Seed Storage Composition during Seed Development. *Plant Cell Environ.* 41, 2109–2127. doi:10.1111/pce.13175

Rai, M. K., Phulwaria, M., and Shekhawat, N. S. (2013). Transferability of Simple Sequence Repeat (SSR) Markers Developed in Guava (*Psidium guajava* L.) to Four Myrtaceae Species. *Mol. Biol. Rep.* 40, 5067–5071. doi:10.1007/s11033-013-2608-1

Raizada, A., and Jegadeesan, S. (2020). *Comparative Transcriptomic Analysis Revealed Complex 2 Molecular Mechanisms Underlying Pests, Pathogens Resistance 3 and Seed Development in Wild and Cultivated Blackgram*. bioRxiv

Rashotte, A. M., Poupart, J., Waddell, C. S., and Muday, G. K. (2003). Transport of the Two Natural Auxins, Indole-3-Butyric Acid and Indole-3-Acetic Acid, in Arabidopsis. *Plant Physiol.* 133, 761–772. doi:10.1104/PP.103.022582

Reinhardt, D., Mandel, T., and Kuhlemeier, C. (2000). Auxin Regulates the Initiation and Radial Position of Plant Lateral Organs. *Plant Cell* 12, 507–518. doi:10.1105/TPC.12.4.507

Riefler, M., Novak, O., Strnad, M., and Schmülling, T. (2006). Arabidopsis Cytokinin Receptor Mutants Reveal Functions in Shoot Growth, Leaf

Senescence, Seed Size, Germination, Root Development, and Cytokinin Metabolism. *Plant Cell* 18, 40–54. doi:10.1105/TPC.105.037796

Robert, C., Noriega, A., Tocino, Á., and Cervantes, E. (2008). Morphological Analysis of Seed Shape in *Arabidopsis thaliana* Reveals Altered Polarity in Mutants of the Ethylene Signaling Pathway. *J. Plant Physiol.* 165, 911–919. doi:10.1016/J.JPLPH.2007.10.005

Ruuska, S. A., Schwender, J., and Ohlrogge, J. B. (2004). The Capacity of Green Oilseeds to Utilize Photosynthesis to Drive Biosynthetic Processes. *Plant Physiol.* 136, 2700–2709. doi:10.1104/PP.104.047977

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.* 467 46, 707–713. doi:10.1038/ng.3008

Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2006). The Auxin Response Factor 2 Gene of Arabidopsis Links Auxin Signalling, Cell Division, and the Size of Seeds and Other Organs. *Development* 133, 251–261. doi:10.1242/DEV.02194

Shen, C., Bai, Y., Wang, S., Zhang, S., Wu, Y., Chen, M., et al. (2010). Expression Profile of PIN, AUX/LAX and PGP Auxin Transporter Gene Families in Sorghumbicolorunder Phytohormone and Abiotic Stress. *FEBS J.* 277, 2954–2969. doi:10.1111/J.1742-4658.2010.07706.X

Shlamovitz, N., Ziv, M., and Zamski, E. (19951995). Light, Dark and Growth Regulator Involvement in Groundnut (*Arachis hypogaea* L.) Pod Development. *Plant Growth Regul.* 16, 37–42. doi:10.1007/BF00040505

Singh, D. P., Jermakow, A. M., and Swain, S. M. (2002). Gibberellins Are Required for Seed Development and Pollen Tube Growth in Arabidopsis. *Plant Cell* 14, 3133–3147. doi:10.1105/TPC.003046

Sinha, P., Bajaj, P., Pazhamala, L. T., Nayak, S. N., Pandey, M. K., Chitikineni, A., et al. (2020). *Arachis hypogaea* Gene Expression Atlas for Fastigiata Subspecies of Cultivated Groundnut to Accelerate Functional and Translational Genomics Applications. *Plant Biotechnol. J.* 18, 2187–2200. doi:10.1111/PBI.13374

Souframanien, J., and Reddy, K. S. (2015). De Novo Assembly, Characterization of Immature Seed Transcriptome and Development of Genic-SSR Markers in Black Gram [*Vigna mungo* (L.) Hepper]. *PLoS One* 10, e0128748. doi:10.1371/JOURNAL.PONE.0128748

Suanum, W., Somta, P., Kongjaimun, A., Yimram, T., Kaga, A., Tomooka, N., et al. (2016). Co-Localization of QTLs for Pod Fiber Content and Pod Shattering in F 2 and Backcross Populations between Yardlong Bean and Wild Cowpea. *Mol. Breed.* 36, 1–11. doi:10.1007/S11032-016-0505-8

Sudheesh, S., Sawbridge, T. I., Cogan, N. O., Kennedy, P., Forster, J. W., and Kaur, S. (2015). *De Novo* Assembly and Characterisation of the Field Pea Transcriptome Using RNA-Seq. *BMC Genomics* 16, 611. doi:10.1186/S12864-015-1815-7

Takahashi, T., Gasch, A., Nishizawa, N., and Chua, N. H. (1995). The DIMINUTO Gene of Arabidopsis is Involved in Regulating Cell Elongation. *Genes Dev.* 9, 97–107. doi:10.1101/GAD.9.1.97

Takahashi, Y., Sakai, H., Yoshitsu, Y., Muto, C., Anai, T., Pandiyan, M., et al. (2019). Domesticating Vigna Stipulacea: A Potential Legume Crop with Broad Resistance to Biotic Stresses. *Front. Plant Sci.* 10, 1607. doi:10.3389/FPLS.2019.01607

Tamaki, S., Matsuo, S., Hann, L. W., Yokoi, S., and Shimamoto, K. (2007). Hd3a Protein is a Mobile Flowering Signal in Rice. *Science* 316, 1033–1036. doi:10.1126/science.1141753

Tang, H., Cuevas, H. E., Das, S., Sezen, U. U., Zhou, C., Guo, H., et al. (2013). Seed Shattering in a Wild Sorghum is Conferred by a Locus Unrelated to Domestication. *Proc. Natl. Acad. Sci.* 110, 15824–15829. doi:10.1073/PNAS.1305213110

Tarkowski, P., Tarkowská, D., Novák, O., Mihaljević, S., Magnus, V., Strnad, M., et al. (2006). Cytokinins in the Perianth, Carpels, and Developing Fruit of *Helleborus niger* L. *J. Exp. Bot.* 57, 2237–2247. doi:10.1093/JXB/ERJ190

Tian, J., Isemura, T., Kaga, A., Vaughan, D. A., Tomooka, N., and McIntyre, C. L. (2013). Genetic Diversity of the rice Bean (Vigna Umbellata) Genepool as Assessed by SSR Markers. *Genome* 56, 717–727. doi:10.1139/gen-2013-0118

Tian, X., Li, S., Liu, Y., and Liu, X. (2016a). Transcriptomic Profiling Reveals Metabolic and Regulatory Pathways in the Desiccation Tolerance of Mungbean (*Vigna radiata* [L.] R. Wilczek). *Front. Plant Sci.* 7, 1–12. doi:10.3389/fpls.2016.01921

Tian, Y., Zhang, M., Hu, X., Wang, L., Dai, J., Xu, Y., et al. (2016b). Over-Expression of CYP78A98, a Cytochrome P450 Gene from *Jatropha curcas* L., Increases Seed Size of Transgenic Tobacco. *Electron. J. Biotechnol.* 19, 15–22. doi:10.1016/j.ejbt.2015.11.001

Walton, L. J., Kurepin, L. V., Yeung, E. C., Shah, S., Emery, R. J. N., Reid, D. M., et al. (2012). Ethylene Involvement in Silique and Seed Development of Canola, *Brassica napus* L. *Plant Physiol. Biochem.* 58, 142–150. doi:10.1016/J.PLAPHY.2012.06.016

Wan, L., Li, B., Lei, Y., Yan, L., Ren, X., Chen, Y., et al. (2017). Mutant Transcriptome Sequencing Provides Insights into Pod Development in Peanut (*Arachis hypogaea* L.). *Front. Plant Sci.* 0, 1900. doi:10.3389/FPLS.2017.01900

Watcharatpong, P., Kaga, A., Chen, X., and Somta, P. (2020). Narrowing Down a Major QTL Region Conferring Pod Fiber Contents in Yardlong Bean (*Vigna unguiculata*), a Vegetable Cowpea. *Genes* 11, 363. doi:10.3390/GENES11040363

Werner, T., Holst, K., Pörs, Y., Guivarc'h, A., Mustroph, A., Chriqui, D., et al. (2008). Cytokinin Deficiency Causes Distinct Changes of Sink and Source Parameters in Tobacco Shoots and Roots. *J. Exp. Bot.* 59, 2659–2672. doi:10.1093/JXB/ERN134

Xiang, J., Tang, S., Zhi, H., Jia, G., Wang, H., and Diao, X. (2017). Loose Panicle1 Encoding a Novel WRKY Transcription Factor, Regulates Panicle Development, Stem Elongation, and Seed Size in Foxtail Millet [*Setaria italica* (L.) P. Beauv.]. *PLoS One* 12, e0178730. doi:10.1371/JOURNAL.PONE.0178730

Xu, W., and Huang, W. (20172017). Calcium-Dependent Protein Kinases in Phytohormone Signaling Pathways. *Int. J. Mol. Sci.* 18, 2436. doi:10.3390/IJMS18112436,

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: A Web Tool for Plotting GO Annotations. *Nucleic Acids Res.* 34, W293–W297. doi:10.1093/NAR/GKL031

Yi, F., Gu, W., Chen, J., Song, N., Gao, X., Zhang, X., et al. (2019). High Temporal-Resolution Transcriptome Landscape of Early maize Seed Development. *Plant Cell* 31, 974–992. doi:10.1105/tpc.18.00961

Zhang, D. F., Zhang, N., Zhong, T., Wang, C., Xu, M. L., and Ye, J. R. (2016). Identification and Characterization of the GH3 Gene Family in Maize. *J. Integr. Agric.* 15, 249–261. doi:10.1016/S2095-3119(15)61076-0

Zhong, R., Kays, S. J., Schroeder, B. P., and Ye, Z.-H. (2002). Mutation of a Chitinase-Like Gene Causes Ectopic Deposition of Lignin, Aberrant Cell Shapes, and Overproduction of Ethylene. *Plant Cell* 14, 165–179. doi:10.1105/TPC.010278

Zhou, Y., Lu, D., Li, C., Luo, J., Zhu, B.-F., Zhu, J., et al. (2012). Genetic Control of Seed Shattering in Rice by the APETALA2 Transcription Factor Shattering Abortion1. *Plant Cell* 24, 1034–1048. doi:10.1105/TPC.111.094383

Zhu, W., Chen, X., Li, H., Zhu, F., Hong, Y., Varshney, R. K., et al. (2014). Comparative Transcriptome Analysis of Aerial and Subterranean Pods Development Provides Insights into Seed Abortion in Peanut. *Plant Mol. Biol.* 85, 395–409. doi:10.1007/s11103-014-0193-x

Zhu, Z., Chen, H., Xie, K., Liu, C., Li, L., Liu, L., et al. (2020). Characterization of Drought-Responsive Transcriptome during Seed Germination in Adzuki Bean (*Vigna angularis* L.) by PacBio SMRT and Illumina Sequencing. *Front. Genet.* 11, 1–14. doi:10.3389/fgene.2020.00996

Ziv, M., and Kahana, O. R. A. (1988). "Photo Morphogenetic Response of the Embryo of Early-Stage Embryos Depended on the Presence of an Intact Suspensor.in" *Injury to the Sus- Pensor at the Globular and Early Heart Stage Caused Abortion of Embryos Cultured* In Vitro. bioRxiv. 57, 159–164.

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 © 2022 Verma, Mittal, Gayacharan, Wankhede, Parida, Chattopadhyay, Prasad, Mishra, Joshi, Singh, 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: 14 March 2022
doi: 10.3389/fgene.2022.750814

Identification of QTLs/ Candidate Genes for Seed Mineral Contents in Common Bean (*Phaseolus vulgaris* L.) Through Genotyping-by-Sequencing

*Muslima Nazir[1], Reetika Mahajan[1], Sheikh Mansoor[2], Sheezan Rasool[1], Rakeeb Ahmad Mir[3], Ravinder Singh[4], Vandana Thakral[5], Virender Kumar[5], Parvaze A. Sofi[6], Hamed A. El-Serehy[7], Daniel Ingo Hefft[8] and Sajad Majeed Zargar[1]**

[1]Proteomics Laboratory, Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, India, [2]Division of Biochemistry, FBSc, Sher-e Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu, India, [3]Department of Biotechnology, BGSB University, Rajouri, India, [4]School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu, India, [5]National Agr Food Biotechnology Institute (NABI), Mohali, India, [6]Division of Genetics and Plant Breeding, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, India, [7]Department of Zoology, College of Science, King Saud University, Riyadh, Saudi Arabia, [8]University Centre Reaseheath, Reaseheath College, Nantwich, United Kingdom

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Parvaiz Ahmad,
Government Degree College, Pulwama, India
Vijay Gahlaut,
Institute of Himalayan Bioresource Technology (CSIR), India

***Correspondence:**
Sajad Majeed Zargar
smzargar@skuastkashmir.ac.in

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

Received: *31 July 2021*
Accepted: *17 January 2022*
Published: *14 March 2022*

Citation:
Nazir M, Mahajan R, Mansoor S, Rasool S, Mir RA, Singh R, Thakral V, Kumar V, Sofi PA, El-Serehy HA, Hefft DI and Zargar SM (2022) Identification of QTLs/ Candidate Genes for Seed Mineral Contents in Common Bean (Phaseolus vulgaris L.) Through Genotyping-by-Sequencing. Front. Genet. 13:750814. doi: 10.3389/fgene.2022.750814

Throughout the ages, the common bean has been consumed by humanity as an important food staple crop and source of nutrition on a global scale. Since its domestication, a wide spectrum of phenotypic and genotypic investigations have been carried out to unravel the potential of this crop and to understand the process of nutrient accumulation along with other desirable characteristics. The common bean is one of the essential legume crops due to its high protein and micronutrient content. The balance in micronutrients is critical for the growth and development of plants as well as humans. Iron (Fe), Zinc (Zn), Copper (Cu), Manganese (Mn), Magnesium (Mg), Calcium (Ca), and Molybdenum (Mo) are some of the important micronutrients present in legumes. Thus, we aimed to investigate the quantitative trait loci's (QTLs)/single nucleotide polymorphisms (SNPs) to identify the candidate genes associated with micronutrients through genotyping by sequencing (GBS). In our investigation, through GBS we identified SNPs linked with traits and assessed seven micronutrients in 96 selected common bean genotypes for screening nutritionally rich genotypes. Among 96399 SNPs total identified through GBS, 113 SNPs showed significant phenotypic variance, ranging from 13.50 to 21.74%. SNPs associated with most of the seed micronutrients (Mg, Mn, Fe, Ca, Cu) were found on chr3 & chr11 (Mg, Mn, Mo, Ca, Zn). The findings from this study could be used for haplotype-based selection of nutritionally rich genotypes and for marker-assisted genetic enhancement of the common bean. Further, the identified SNPs for candidate genes/transporters associated with micronutrient content may pave the way for the enrichment of seeds by employing genomics-assisted breeding programs.

Keywords: common bean, single nucleotide polymorphism (SNP), genome wide association studies (GWAS), ionome, transporters, population structure

1 INTRODUCTION

Common bean (*Phaseolus vulgaris* L.), an important food legume, constitutes 50% of the grain legumes consumed as a human food source (Câmara et al., 2013; FAOSTAT, 2017). Common bean has huge genetic variation and based on domestication it is distinguished into two main gene pools, Andean and Mesoamerican. Cultivars in the mesoamerican gene pool have small to medium seed size and with "S" or "B" phaseolin patterns, while Andean cultivars have a large seed size with phaseolin patterns "T"/"C"/"H"/"A" (Bitocchi et al., 2012; Bellucci et al., 2014).

Being a great source of carbohydrates, dietary proteins, soluble and insoluble fibers, vitamins, and essential micronutrients such as minerals including Iron (Fe), Zinc (Zn), Copper (Cu), Manganese (Mn), Magnesium (Mg), Calcium (Ca) and Molybdenum (Mo), beans have often been considered a "poor man's meat" (Fennema owen, 2000; Hayat et al., 2014; Yeken, 2018). These micronutrients play a pivotal role in the proper growth and development of plants and animals. The deficiency of any essential micronutrients hinders the proper functioning of biological processes leading to several metabolic and physiological implications. In many low- and lower-middle income countries, especially those in Asia, Africa, and Latin America, micronutrients Fe and Zn are the main components of hidden hunger (Darnton-Hill et al., 2006). Similarly, Fe and Zn deficiency cause severe yield loss in crops and metabolic disturbances in humans (Zargar et al., 2015). In addition, Cu is one of the important trace elements that play a vital role in maintaining metabolic activities. In humans, Cu deficiency leads to anemia, cardiac dysfunction, myeloneuropathy, and myelopathy, whereas in plants, its deficiency leads to lignification dysfunctioning (Geir, 2013; Lehmann and Rillig, 2015; Papamargaritis et al., 2015).

Food crops are the major source of essential minerals (Hardiman et al., 1984; Ali et al., 2014). Biofortification of nutrient-rich plants is important in addressing malnutrition-related issues. Thus, given today's population explosion and food shortage issues, there is a need to introduce smart food crops into our diet. So far, research has been carried out using both conventional and modern breeding approaches to increase the nutrient content in cultivated plants. It is now important to identify genetic loci that regulate the uptake of essential minerals, as each genetic loci is an essential factor in the success of the biofortification breeding effort. Association mapping is one of the modern breeding approaches used to identify genetic loci that determine desired traits (Lewontin and Kojima, 1960). Association mapping has several advantages over the bi-parental QTL mapping approach (Yu and Buckler, 2006; Tian et al., 2011; Wen et al., 2014; Nadeem et al., 2021). For example, association mapping explores the allelic diversity that exists in the diverse germplasm, while QTL mapping can examine the allele variation present in only two parental lines. Undoubtedly, both approaches are indispensable and have their own advantages and disadvantages (Sonah et al., 2013). We believe an integrated approach involving both association mapping as well as biparental mapping can lead to a breakthrough in crop improvement.

Since the whole genome sequence of the common bean is available in the public domain, as such there is an excellent opportunity to perform Genome-wide association studies (GWAS) to identify QTLs followed by candidate genes that govern the uptake and accumulation of minerals (Schmutz et al., 2014). Genome-wide association study is one of the modern breeding approaches for mapping genes associated with different traits (Lu et al., 2015; Liu et al., 2016). The introduction of Next-generation sequencing (NGS) technologies has sped up the identification of SNPs and subsequent genotyping (Verma et al., 2015). Genotyping-by-sequencing (GBS) is a robust and cost-effective method wherein selective small genome fragments obtained by restriction digestion are sequenced by NGS platform to identify SNPs (Narum et al., 2013; Peterson et al., 2014; Schilling et al., 2014). In the last two decades, numerous studies have been performed using the GBS approach in diverse plant species, including wheat, canola, barley, and soybean, which are known to have a complex genome. In this regard, GWAS in common bean will allow us to estimate population structure and linkage disequilibrium (LD), connecting the variation in the genome with the phenotypic variations in the population. The LD-based analysis is organized based on population structure and genetic relatedness among populations. Several association studies with respect to micronutrient contents have been carried out in food crops like rice (Shao et al., 2011; Norton et al., 2014), pea (Diapari et al., 2015), chickpea (Diapari et al., 2014; Upadhyaya et al., 2016; Ozkuru et al., 2019), common bean (Nemli et al., 2016; Mahajan et al., 2017a; Erdogmus et al., 2020).

In the present study, we examined the germplasm of common beans of the Jammu and Kashmir region of northwestern Himalayas, India, for various micronutrients, followed by the identification of QTLs associated with the accumulation of nutrients. To identify genes/QTLs that regulate micronutrients in beans, GBS-based SNPs were discovered from natural populations of beans. Based on the preliminary studies, a set of 96 different bean genotypes was created, which were collected from other regions of the northwestern Himalayas (Zargar et al., 2014). In the present study, the ionome profile, which contains Ca, Cu, Mg, Mn, and Mo, of the bean core set was deciphered, and the QTLs contributing to their accumulation by GWAS have been identified. The investigation led us to inventory the micronutrients of the 96 different types of beans to determine candidate lines (nutrient-rich for different micronutrient levels in the seed) that can be used in breeding high-nutrient, high yielding bean varieties. In addition, the identification of QTLs can serve as critical genomic resources for improving the micronutrient profile in beans. These studies can improve the understanding of possible correlations for the accumulation of different elements. To understand the function of the respective candidate genes that regulate micronutrient uptake in beans, further studies need to be performed that may include knockout or overexpression of responsive candidate genes.

2 EXPERIMENTAL PROCEDURES

2.1 Plant Material

A total of 96 common bean germplasm lines, mostly landraces and a few released varieties (SFB1, SR1, SR2, Arka Anoop, VLR-125) were used as plant material in the current study. Germplasm was collected from different geographical regions of Jammu and Kashmir (**Supplementary Table S1**) and maintained at the research fields at Division of Genetics and Plant Breeding, Faculty of Agriculture, Wadura, Sher-e-Kashmir University of Agricultural Sciences and Technology-Kashmir (SKUAST-K), Sopore. Most of the released varieties used in this study were developed through a single plant selection and have been used as checks.

2.2 Field Experiment and Micronutrient Profiling

Field experiments were conducted during 2016 and 2017 at the experimental field of the College of Agriculture, Wadura, SKUAST-Kashmir, India (34.34 North; 74.4 East; Altitude: 1,590.00 m). Clay loam textured soil with pH (7.2), organic carbon (.65%), electrical conductivity (.18 dS/m), CEC (16 meq/kg), and an available concentration of different elements in the soil i.e. P (4.91 mg kg^{-1}), K (5.55 mg kg^{-1}), Zn (.68 mg kg^{-1}), Fe (5.1 mg kg^{-1}), Cu (.29 mg kg^{-1}), Mn (6.2 mg kg^{-1}) was used for plantation of germplasm seeds in the experimental sites. The experiment was laid out according to augmented block design, which includes more than one released variety that is taken as replicated treatments, and these varieties are repeated in each block. Five released varieties (SFB1, SR1, SR2, Arka Anoop, VLR-125) were included in each block as checks. All the standard agronomical practices recommended were followed to raise healthy and disease-free crop plants. Harvesting was done at the time of 90% pod maturity. Further, the seed material of each genotype was powdered to analyze seven essential micronutrients i.e. Cu, Mn, Mg, Ca, Fe, Zn, and Mo. The elemental profiling of these genotypes was determined using a portable X-ray fluorescence spectrometer (pXFR). The pXRF instrument was calibrated as explained in Reidinger et al. (2012). In pXRF, a synthetic methylcellulose matrix was used to spike the known quantity of standard elements. Based on the methylcellulose pellet with know standards, an elemental composition standard curve was developed and subsequently used for sample evaluation. Similarly, samples were cross-verified with Energy Dispersive X-ray Fluorescence (ED-XRF).

2.3 Genotyping-by-Sequencing of Common Bean Genotypes

Genomic DNA was extracted from 15 day old leaves by using the CTAB method (Doyle and Doyle, 1990), and the quality, as well as quantity of DNA, was checked on both gel electrophoresis (.8% Agarose) and nano-drop (mySPEC, Wilmington, USA). The extracted genomic DNA was purified for the preparation of multiplex GBS libraries *via* Illumina HiSeq 2000 (SciGenom Pvt. Ltd., India). *Ape*K1 (from *Aeropyrum pernix K1*) restriction enzyme was employed for restriction digestion of genomic DNA. After quality filtering, de-multiplexed high-quality sequences were mapped to the reference common bean genome (Phtyozome v12.1 database), and SNPs were mined from the coding and non- coding regions of common bean genes and chromosomes. Subsequently, SNPs mined were structurally annotated on the diverse coding DNA sequence (CDS) and non-coding (upstream/downstream regulatory regions and introns) sequence components of genes and intergenic regions of the common bean genome.

2.4 Statistical Analysis

2.4.1 Micronutrient Profiling Analysis

All the observations were recorded in replicates of three, and values were then averaged. One-way ANOVA was applied to evaluate the variance of seven micronutrients among the genotype and Pearson's pairwise correlation coefficient was calculated for all elements using the SPSS program (ver. 16).

2.4.2 Population Genetic Analysis

Population structure was estimated using a Bayesian Markov Chain Monte Carlo model (MCMC) implemented in STRUCTURE (Pritchard et al., 2000). Five runs were performed for each number of populations (K) set from 01 to 12. Burn-in time and MCMC replication number were both set to 100,000 for each run. The most likely *K value* was determined by the log probability of the data LnP(D) and delta K, based on the rate of change in LnP(D) between successive K values. These analyses were performed using Structure Harvester (Earl and vonHoldt, 2012). The neighbor-joining tree was built using Phylip and MEGA5 (Felsenstein, 1989; Tamura et al., 2011).

2.4.3 Genome-wide Association Analysis

All the analyses were performed using TASSEL3.0 and the Genomic Association and Prediction Integrated Tool (GAPIT) (Bradbury et al., 2007; Lipka et al., 2012). Mixed linear models (MLM) were used for the identification of SNPs associated with these traits. Van Raden method (K) was used to calculate the kinship matrix (Loiselle et al., 1995; Hyun et al., 2008). Covariate like P from principal component analysis and Q from STRUCTURE along with kinship matrix (K) were used for mixed linear models (MLM). The negative log(1/n) was used to establish a significance threshold (Wang et al., 2012; Yang et al., 2013).

2.4.4 LD Plots and Haplotype Blocks

The SNP matrix for all the samples was converted to HapMap format and TASSEL was used for the filtering out of SNP with major allele frequency (MAF) of less than .05. The LD plots for individual chromosomes were created in Tassel > Analysis > Diversity > Linkage Disequilibrium with LD Type as "Sliding Window" and LD Window size set to 50. The heterozygous sites were treated as missing and the R^2 accumulated was calculated for 100 intervals with a size of .01. The R^2 data obtained following LD analysis was used for plotting chromosome-wise LD information using MS Excel.

The PLINK v1.90b6.24 64-bit was used to determine the block size of SNP haplotypes. The individual chromosome Tassel files

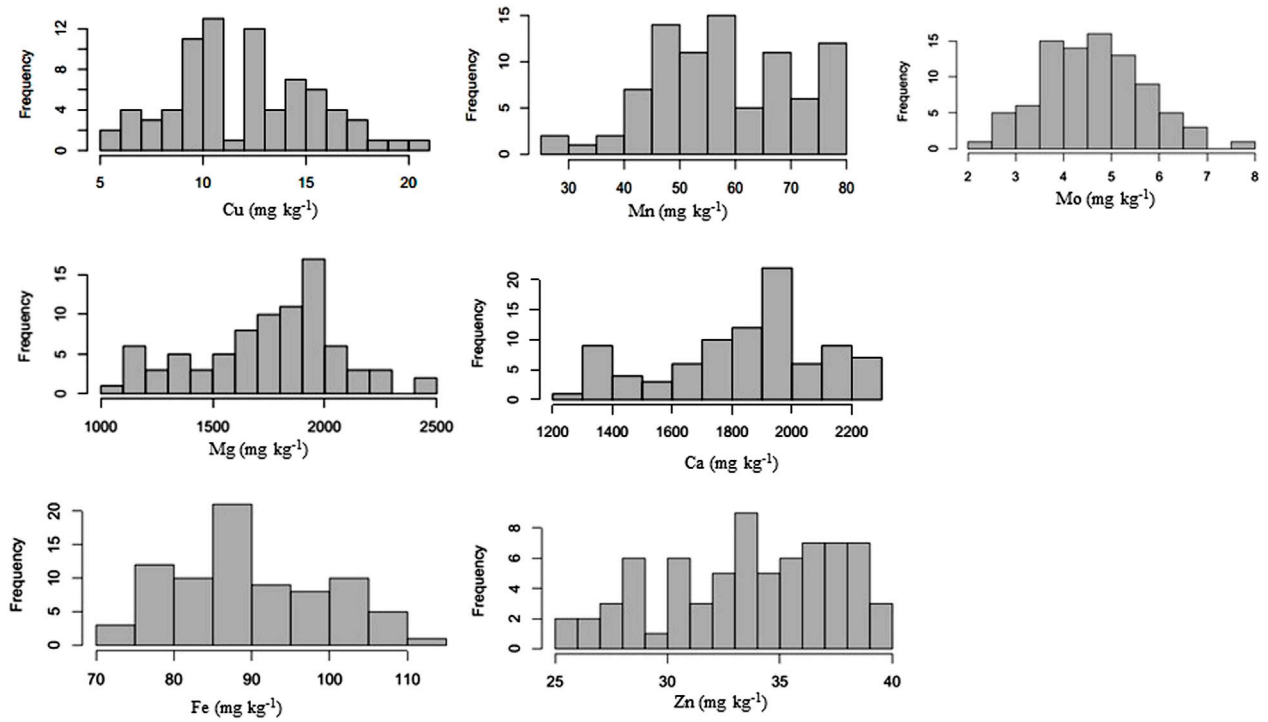


FIGURE 1 | Frequency distribution of various micronutrients in common bean.

were saved to Plink format (.map and .ped) using the option available in Tassel. The haplotype blocks were calculated using—blocks no-pheno-req-- blocks-max-kb 100--make-founders command in Plink. The Haplotype Block analysis revealed the size of Haplo Blocks, including the number of SNPs on each block. The data was plotted using MS Excel.

2.5 Candidate Gene Identification

For candidate genes identification, the reference genome of *P. vulgaris* (V2.1) was used. The candidate genes were identified in .1 Mb both flanking regions of significant SNPs using the BioMart tool (Smedley et al., 2009), and the information related to gene function, Pfam ID, Panther ID, KOG ID, gene ontology ID, and their description were downloaded.

3 RESULTS

3.1 Genetic Variation in Common Bean Seed Micronutrients

Micronutrient profiling of 96 genotypes was conducted to study the distribution and correlation among each other. Normal distribution was observed for Cu, Mn, Mg, Ca, Fe, Zn, and Mo (**Figure 1**; **Supplementary Tables S2, S3**). Seed micronutrient content showed continuous variation for Fe ranging from 67.35[1]–133.02 mg kg^{-1} (with average (av) of 94.21 mg kg^{-1}), 21.16–49.77 mg kg^{-1} (av 35.98 mg kg^{-1}) for Zn, 1,293.21–2,657.46 mg kg^{-1} (av 1893.56 mg kg^{-1}) for Ca, 2.02–28.02 mg kg^{-1} (av 13.11 mg kg^{-1}) for Cu, 1,057.55 mg kg^{-1} 208–2,492.26 mg kg^{-1} (av 1827.71 mg kg^{-1}) for Mg, 22.34–93.36 mg kg^{-1} (av 58.26 mg kg^{-1}) for Mn and 2.09–7.80 mg kg^{-1} (av 4.63 mg kg^{-1}) for Mo (**Table 1**). Very low coefficients of variation (CVs) were observed for all these micronutrients.

3.2 Correlation Analysis

Pearson's correlation analysis showed a highly significant positive correlation between Fe and Zn (r = .61**), Ca and Mg (r = .37**), Cu and Fe (r = .28**), and Cu and Zn (r = .26*) (**Table 2**). All the other micronutrients showed a non-significant correlation with each other. Correlation studies between the micronutrient contents of the seeds in the present study showed that Mn is the only micronutrient that correlates negatively with all other micronutrients such as Fe, Zn, Cu, Ca, Mg, and Mo. However, the Zn content was found positively correlated with Fe and Cu and negatively correlated with Ca, Mo, and Mg. Mo was found to correlate positively with Ca, Mg, Cu and negatively with Fe, Zn, and Mn. The Fe content was found to be positively correlated with Ca, Mg, and Cu, while it was negatively correlated with Mo and Mn.

3.3 Characterization and Distribution of SNPs in the Common Bean

A total of 96,399 SNPs were found among a whole set of 96 diverse genotypes. The highest number of SNPs (10,978) were observed on chr. 3, whereas the lowest number of SNPs (6524) were observed on chr. 6 (**Figure 2**).

TABLE 1 | Range, highest, lowest genotype and Coefficient of Variance (CV) of different micronutrient content of common bean seeds.

S.No	Micronutrient	Range (mg kg^{-1})	Genotype		Average (mg kg^{-1}±S.E)	CV
			Lowest	Highest		
1	Molybednum (Mo)	2.09–7.80	KD11	WB1680	4.63 ± 0.07	3.54
2	Zinc (Zn)	21.16–49.77	WB352	WB1190	35.98 ± 0.05	.30
3	Iron (Fe)	67.35–133.02	K13	WB1679	94.21 ± 0.04	.10
4	Calcium (Ca)	1,293.21–2,657.46	WB1643	KD7	1893.56 ± 0.03	.03
5	Magnesium (Mg)	1,057.55–2,492.26	WB371	K16	1827.71 ± 0.03	.004
6	Manganese (Mn)	22.34–93.36	R2	N15	58.26 ± 0.02	.27
7	Copper (Cu)	2.02–28.02	WB1136	UJ	13.11 ± 0.02	.07

TABLE 2 | Correlation among seed micronutrient content in common bean.

Correlation							
	Mo	Zn	Fe	Ca	Mg	Mn	Cu
Mo	1						
Zn	−.104	1					
Fe	−.032	.608[a]	1				
Ca	.106	−.076	.086	1			
Mg	.052	−.029	.031	.373[a]	1		
Mn	−.010	−.140	−.125	−.028	−.010	1	
Cu	.121	.258[b]	.283[a]	.125	.150	−.183	1

[a]*Correlation is significant at the .01 level.*
[b]*Correlation is significant at the .05 level.*

3.4 Genetic Diversity and Population Structure

In the present study, all paired genetic distances between the 96 bean lines were determined from SNP-based genotypic data. A neighboring tree based on these genetic distances showed that the genotypes were divided into five main and five subgroups (**Figure 3A**). The Dendrogram revealed that of the minor groups, one included only one genotype, WB6; another group included only two genotypes (N13, WB1137); while another two groups were found to have only three genotypes each (KD17, WB1664, KD5 in one group and WB877, KD11, WB864 in the second group) and yet another group comprised of four genotypes (WB1282, KD13, R9, WB1436). The released lines like SR2 and ARKAANOOP were grouped in one cluster with other local lines, whereas VLR-125 and SFB1 clustered in another group.

Principal component analysis (PCA) also showed diversity among the common bean genotypes (**Figure 3B**). The released bean lines and local lines have been pooled with no clear separation of local and released lines (**Figure 3B**). In addition, population structure analysis provides a robust analysis for understanding the genotypic origins of a particular crop. The population structure was scored for K values ranging from 1 to 12 across the panel using high-quality SNPs. The peak of delta K was found to be the highest at K = 5 and thus groups the 96 genotypes of the common bean into 5 populations (**Figure 3C**). Furthermore, this was in accordance with the neighbor-joining tree with little deviation (**Figure 3A**).

3.5 Genome-Wide Association Analysis

GWAS was performed for common bean seed micronutrient contents (Fe, Zn, Cu, Ca, Mn, Mg, and Mo). Out of 96,399 SNPs,

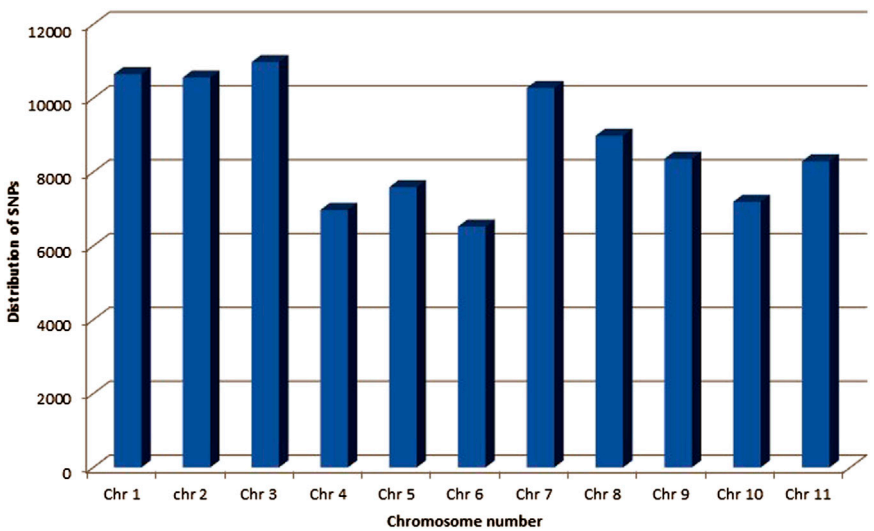


FIGURE 2 | Frequency distribution of SNPs identified using GWAS on different chromosomes of common bean.

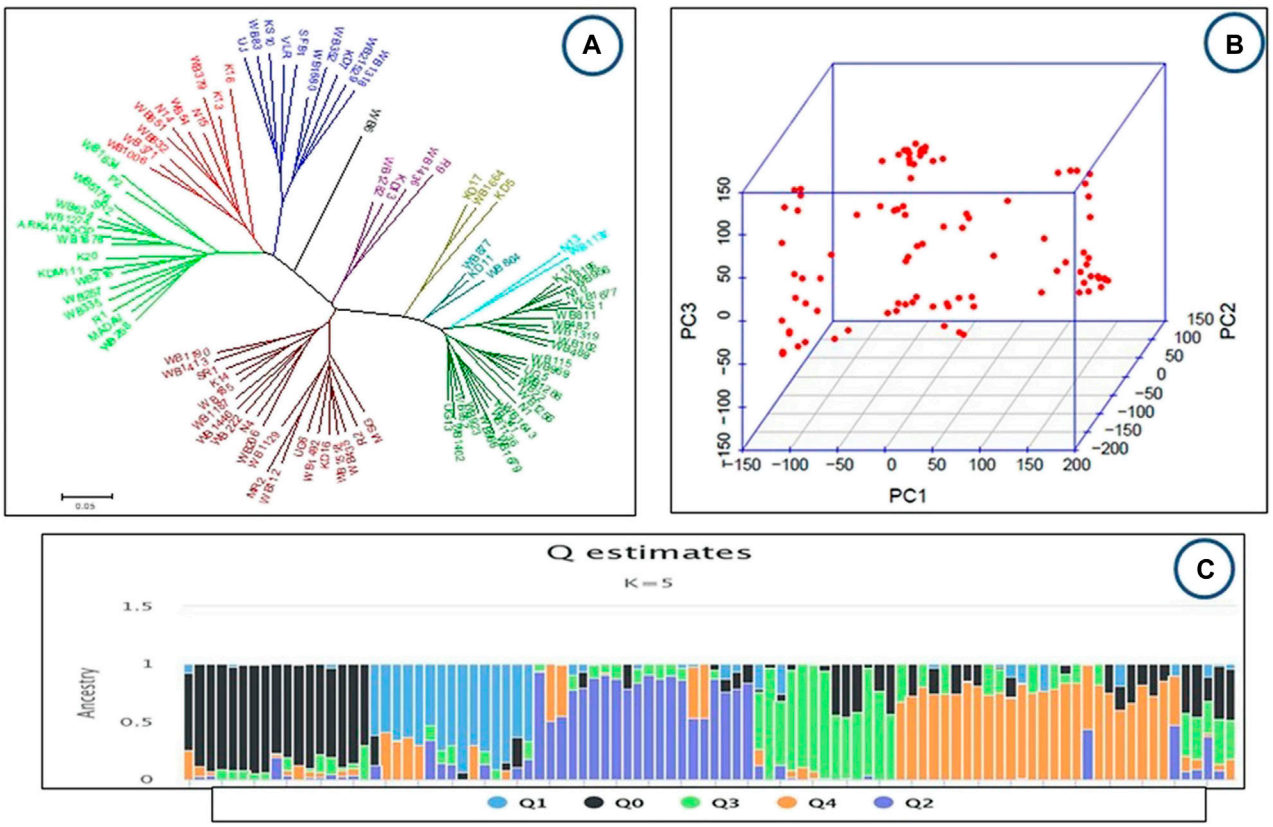


FIGURE 3 | (A–C): Genetic diversity and population structure of the studied common bean accessions **(A)** Phylogenetic trees constructed using the neighbor-joining method by Phiylip and MEGA 5 **(B)** PCA Scatter plots of the first two principal components (PCA analyses), each dot represents one accession **(C)**: Population structure K = 5, each accession is represented by a single vertical line and colors represent ancestries.

113 SNPs were found to be significantly associated with different seed micronutrient contents with 13.50–21.74% phenotypic variance (**Figures 4A–G**; **Table 3**). A total of 32 SNPs across all chromosomes except chr. 8 were found significantly associated with seed Mg content. The highest number of SNPs (7) associated with seed Mg content was found on chr.2, while only one associated SNP was found on chr.1 and chr.7. One SNP on chr.9 positioned at 31752041 (*p*-value = 1.61E-04) contributed 18.42% phenotypic variation. For seed Mo content, 29 SNPs were found significantly associated that are positioned on chr.4, chr.5, chr.6, chr.7, chr.10, and chr.11 with the highest number of SNPs (13) on chr.10 and the lowest (1) on chr.6 and chr.7. One SNP on chr.11 contributed to 21.42% phenotypic variation. About 15 SNPs were found associated with seed Ca content on each of chr.3, chr.4, chr.5, chr. 7, chr. 9, and chr. 11 and seed Cu content on chr.2, chr.3, chr.5, chr.6, and chr.8. For seed Ca content, SNP contributing to the highest phenotypic variation (17.10%) was found on chr. 3 whereas, SNP associated to seed Cu content located on chr.2 and chr.3 contributed 21.74% and 21.71% phenotypic variation. Around 13 SNPs were found associated with seed Mn content on chr.1, chr.3, chr.8, chr.10 and chr.11 whereas, only 5 SNPs were found associated with seed Zn content on chr.2, chr.10, and chr.11 and 4 SNPs were associated to seed Fe content on chr.1, chr.3, and chr.7.

3.6 LD Plot and Haplotype Blocks

Overall LD measured as R^2 was correlated in all chromosomes (**Figure 5**; **Supplementary Table S4**). The maximum number of SNPs (125,580) correlated with other chromosomes were found on chr.10, whereas, a minimum number of SNPs (50,003) were found on chr.1 at R^2 = .01. With the increase in R^2 value, decreasing pattern in the number of associated SNPs across the 11 chromosomes was observed. At R^2 = .93 the lowest number of SNPs correlation was found in all the chromosomes. The lowest (33) and highest (232) were found on chr. 6 and 1, respectively. Moreover, a total of 1879 SNPs on chr.1 were found to correlate with other chromosomes at R^2 = .96, whereas 421 SNPs were found on chr.10. No SNPs were found associated across 11 chromosomes at R^2 = .97–.99. Through haplotype analysis, a set of 7107 haploblocks representative of the 11 chromosomes, ranging from 1244 (chr. 1) to 503 (chr. 4) were identified. A total of 22,090 SNPs were distributed in these blocks, with an average of ~3 SNPs per block. Chr. 1 (17) and chr. 4 and 10 (7 each) had the highest and lowest number of SNPs within their haploblocks, respectively (**Figures 6A–K**; **Supplementary Table S5**).

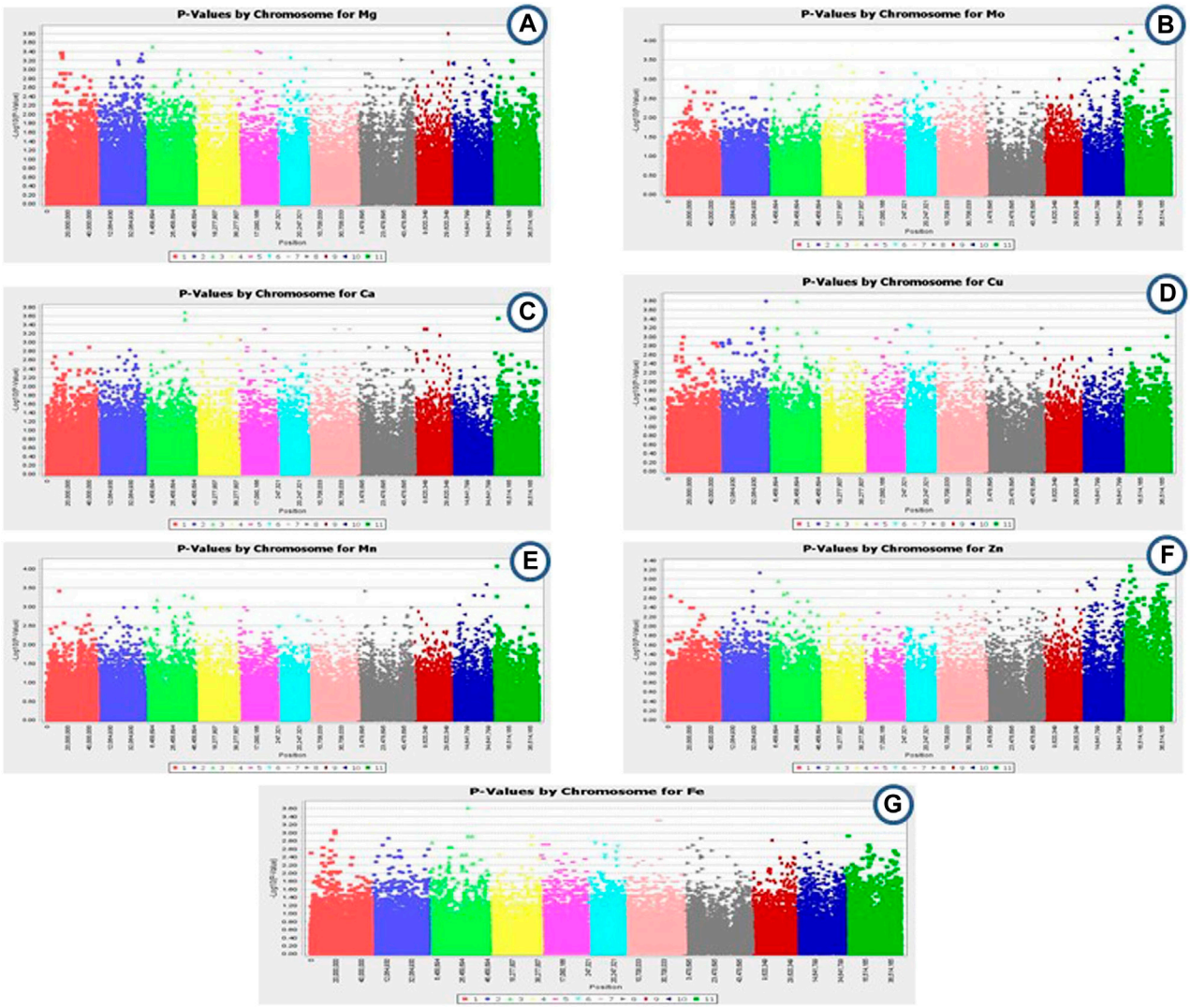


FIGURE 4 | (A–G): Summary of Genome-wide association result: Manhattan plots depicting association of 113 SNP markers with seed. **(A)** Mg, **(B)** Mo, **(C)** Ca, **(D)** Cu, **(E)** Mn, **(F)** Zn, **(G)**. Fe content in common bean.

3.7 Candidate Gene Analysis

A total of 840 genes were identified in the .1 Mb flanking region of significant SNPs related to different traits; however, five SNPs were not co-localized with any gene in the .1 Mb flanking region (**Supplementary Table S6**). Out of these 840 genes, 16 transporter genes were identified (**Table 4**), some being metal transporters. We have also identified genes for 24 metal-binding proteins (**Table 4**) like zinc finger, calcium, and iron-binding proteins.

4 DISCUSSION

4.1 Micronutrients Variation and Correlation Among Micronutrients

Micronutrients play an indispensable role in the growth and development of eukaryotic organisms. A deficiency in these essential micro-and macronutrients leads to abnormal growth in living systems. Humans get most of their micronutrients from plant and animal sources. Therefore, the biofortification of important food crops is necessary nowadays. In the present study, an initiative was taken to investigate different micronutrient concentrations in the germplasm of common beans. The micronutrient content of the seeds has been varied widely in common bean seeds (**Supplementary Table S2**). Previous studies have shown that mineral variation has been observed in almost all major legumes, including the common bean. The different mineral content in beans has been studied in different parts of the world including India (O Abidemi et al., 2012; Martinez Meyer et al., 2013; Zaccardelli et al., 2013; Kumar and Chopra, 2014; Mahajan R et al., 2015; Erdogmus et al., 2020). Earlier reports also suggested that seed mineral content showed huge variation in common bean germplasm.

TABLE 3 | Details of loci associated with accumulation of different elements.

Trait	Marker	Chr. No	Position	*p*-value	R_2 (%)
Cu	Pv36231	2	43570203	1.63E-04	21.74
	Pv42447	3	26899343	1.64E-04	21.71
	Pv63237	6	4804173	5.52E-04	17.94
	Pv63769	6	7107934	6.00E-04	17.68
	Pv38848	3	8664946	6.52E-04	17.43 17.43
	Pv33235	2	30814539	6.52E-04	17.43
	Pv35821	2	41211297	6.52E-04	17.43
	Pv87597	8	52162199	6.52E-04	17.43
	Pv60686	5	29919387	7.02E-04	17.21
	Pv35247	2	38960183	8.03E-04	16.81
	Pv35810	2	41160625	8.03E-04	16.81
	Pv46872	3	44944379	8.03E-04	16.81
	Pv67269	6	23404896	8.08E-04	16.78
	Pv42479	3	27006768	8.36E-04	16.68
	Pv35442	2	39678800	8.46E-04	16.65
Mn	Pv18707	11	5168818	8.51E-05	20.40
	Pv16743	10	32828781	2.60E-04	17.38
	Pv80480	8	6963598	3.88E-04	16.33
	Pv02970	1	12045066	3.88E-04	16.33
	Pv45092	3	37444587	4.99E-04	15.67
	Pv14912	10	21149317	5.01E-04	15.66
	Pv14921	10	21209873	5.01E-04	15.66
	Pv16067	10	29484108	5.18E-04	15.57
	Pv18722	11	5199874	5.39E-04	15.47
	Pv46926	3	45208418	5.69E-04	15.33
	Pv39599	3	11581681	6.73E-04	14.89
	Pv12158	10	6845256	8.94E-04	14.17
	Pv24110	11	33811979	9.88E-04	13.91
Mg	Pv71125	9	31752041	1.61E-04	18.43
	Pv38472	3	7168425	3.16E-04	16.64
	Pv58364	5	17263012	3.92E-04	16.07
	Pv52919	4	30632382	3.96E-04	16.05
	Pv58963	5	20353152	4.32E-04	15.83
	Pv03198	1	13513718	4.33E-04	15.82
	Pv03216	1	13577463	4.33E-04	15.82
	Pv03452	1	14989418	4.33E-04	15.82
	Pv36071	2	42540272	4.57E-04	15.69
	Pv03386	1	14603432	5.19E-04	15.36
	Pv03456	1	14999140	5.19E-04	15.36
	Pv03454	1	14992813	5.58E-04	15.17
	Pv64956	6	12382652	5.66E-04	15.13
	Pv35797	2	41116882	5.76E-04	15.09
	Pv73343	7	20643908	6.13E-04	14.93
	Pv85139	8	41708004	6.13E-04	14.93
	Pv16441	10	31481349	6.29E-04	14.86
	Pv21626	11	19327905	6.39E-04	14.82
	Pv30574	2	20002969	6.51E-04	14.78
	Pv35396	2	39493707	6.64E-04	14.73
	Pv35747	2	40920106	6.64E-04	14.73
	Pv36214	2	43486371	6.64E-04	14.73
	Pv21789	11	20127048	6.64E-04	14.73
	Pv10966	10	1524933	7.14E-04	14.54
	Pv71157	9	31933306	7.14E-04	14.54
	Pv10985	10	1595959	7.66E-04	14.37
	Pv71179	9	32052922	7.66E-04	14.37
	Pv17142	10	34496599	7.66E-04	14.37
	Pv30756	2	20684206	7.70E-04	14.35
	Pv14063	10	16507514	9.28E-04	13.88
	Pv68046	6	26644872	9.73E-04	13.76
	Pv43881	3	32599558	9.98E-04	13.69
Ca	Pv45223	3	38019351	2.07E-04	17.10
	Pv18893	11	6191264	2.90E-04	16.25

(Continued in next column)

TABLE 3 | (*Continued*) Details of loci associated with accumulation of different elements.

Trait	Marker	Chr. No	Position	*p*-value	R_2 (%)
	Pv45240	3	38092190	3.03E-04	16.13
	Pv45215	3	37988109	3.06E-04	16.11
	Pv65433	9	8733609	5.04E-04	14.87
	Pv66070	9	11355026	5.04E-04	14.87
	Pv66093	9	11451916	5.04E-04	14.87
	Pv66113	9	11504334	5.04E-04	14.87
	Pv59654	5	24452529	5.04E-04	14.87
	Pv73938	7	24798685	5.04E-04	14.87
	Pv77126	7	38956026	5.04E-04	14.87
	Pv69120	9	23846015	6.95E-04	14.08
	Pv51758	4	24279020	7.21E-04	13.99
	Pv54439	4	40294366	8.55E-04	13.58
	Pv54824	5	825380	8.80E-04	13.51
Fe	Pv43570	3	31289008	2.45E-04	18.82
	Pv74144	7	25770068	4.97E-04	16.81
	Pv74596	7	27864996	4.97E-04	16.81
	Pv04389	1	19220891	9.02E-04	15.15
Zn	Pv19080	11	7315158	5.31E-04	18.47
	Pv19085	11	7334690	6.62E-04	17.75
	Pv19088	11	7346014	6.62E-04	17.75
	Pv34942	2	37819199	7.36E-04	17.45
	Pv13503	10	12576659	9.54E-04	16.65
Mo	Pv19156	11	7595234	6.20E-05	21.42
	Pv16574	10	32099819	8.69E-05	20.49
	Pv16729	10	32768914	9.00E-05	20.40
	Pv19452	11	8815634	1.85E-04	18.43
	Pv21419	11	18293910	4.38E-04	16.13
	Pv51170	4	19580353	4.39E-04	16.12
	Pv16324	10	30756216	5.20E-04	15.68
	Pv20017	11	11345383	6.15E-04	15.24
	Pv16924	10	33461608	6.15E-04	15.24
	Pv16947	10	33552458	6.15E-04	15.24
	Pv16983	10	33671563	6.15E-04	15.24
	Pv58272	5	16746387	6.81E-04	14.97
	Pv58289	5	16870681	6.81E-04	14.97
	Pv53043	4	31371712	6.81E-04	14.97
	Pv20598	11	13856025	6.88E-04	14.95
	Pv64654	6	10916435	7.55E-04	14.71
	Pv19693	11	10155319	8.91E-04	14.28
	Pv19709	11	10212465	8.91E-04	14.28
	Pv16484	10	31648813	8.91E-04	14.28
	Pv16500	10	31724718	8.91E-04	14.28
	Pv16533	10	31876693	8.91E-04	14.28
	Pv16572	10	32090665	8.91E-04	14.28
	Pv14753	10	20089097	9.58E-04	14.10
	Pv19859	11	10700430	9.63E-04	14.08
	Pv19921	11	10908469	9.63E-04	14.08
	Pv78779	7	46163077	9.83E-04	14.03
	Pv17019	10	33837525	9.84E-04	14.03
	Pv20026	11	11415382	9.92E-04	14.00
	Pv16937	10	33511552	9.92E-04	14.00

Seed mineral concentrations such as Mg, Ca, Fe, Zn, and Cu of 60 common bean genotypes collected from the Western Himalayas varied from 1,220.5 to 2,737.5 ppm, 300–5,350 ppm, 80.5–180.6 ppm for Fe, 14.64–104.08 ppm, and .9–13.4 ppm, respectively (Jan et al., 2021). The average seed Ca concentration was recorded as 1.37 and 1.41 g kg^{-1}, whereas, the average Fe seed content was recorded as 79.57 and

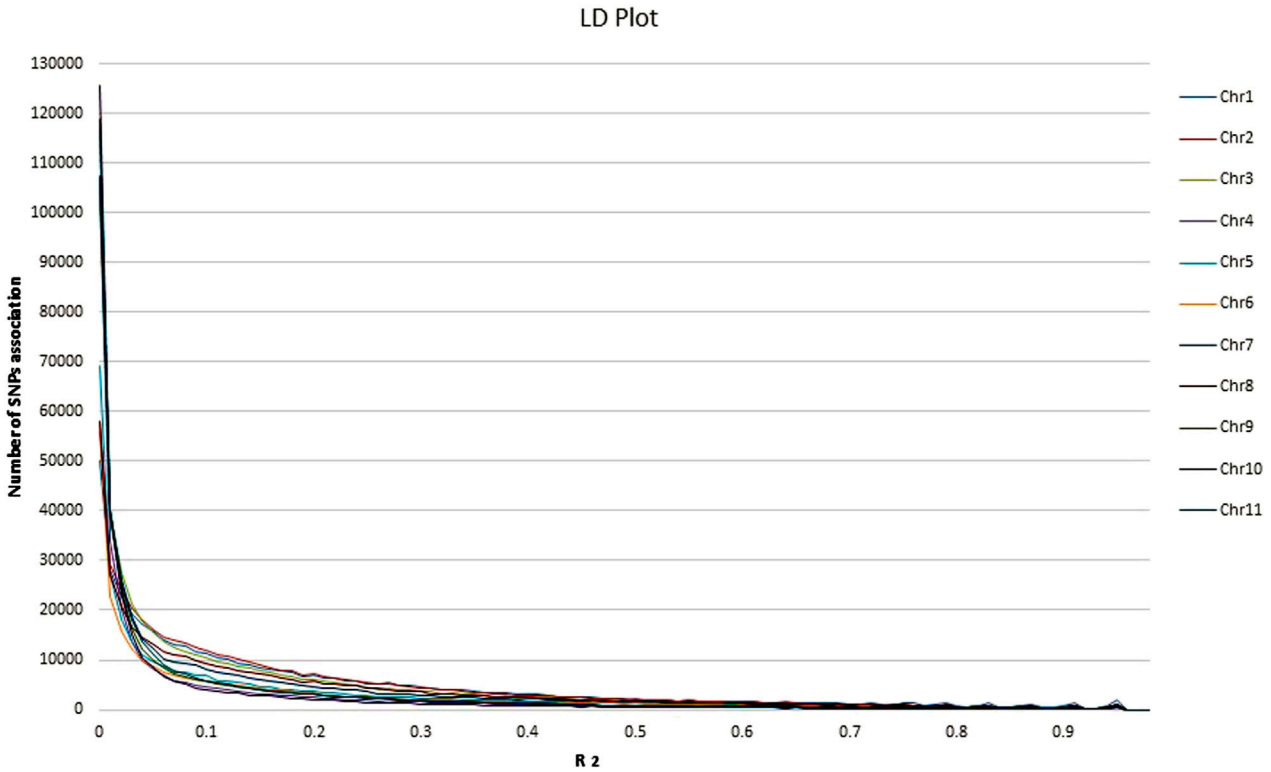


FIGURE 5 | Analysis of linkage disequilibrium (LD) plot across the 11 common bean chromosomes.

85.95 mg kg^{-1} from common bean seeds obtained through pedigree and single seed descent methods respectively (Ribeiro et al., 2014). Also, 88.14 mg $kg^{-1}$of Fe, 49.24 mg $kg^{-1}$of Zn, .25 g 100 g^{-1} of Mg, 11.30 mg $kg^{-1}$of Cu, and 22.71 mg $kg^{-1}$of Mn was found in common bean genotypes from Universidade Federal de Lavras (UFLA) (Silva et al., 2010) and 74.6 ppm Fe, 39.9 ppm Zn content in Ugandan common bean germplasm (Mukamuhirwa et al., 2012). The difference in micronutrient content in bean genotypes from different parts of the world may be due to the different number of samples taken for evaluation, environmental conditions such as climate and soil composition, and agricultural techniques. This diversity in the germplasm can help us identify potential candidate lines that can be used in the development of Multi-Parent Advanced Generation Intercross (MAGIC) or bi-parental mapping populations, the breeding of micronutrient-rich, high-yielding varieties of beans, can be used for investigating different levels of gene expression for different nutrients in common beans. The higher micronutrient bean lines could be used for biofortification programs.

Based on the Pearson's correlation analysis our study reveals a significant positive correlation between Fe and Zn; Ca and Mg; Cu, Fe, and Zn. In previous studies, a similar correlation pattern between Zn and Fe and other minerals was observed in bean genotypes (Beebe et al., 2000; House et al., 2002; Gelin et al., 2007; Pfeiffer and McClafferty, 2007). Correlation studies have been conducted on other agronomical traits in the common bean (Nadeem et al., 2020). The negative correlation between the trait suggests that these traits are interdependent. In the present study, Mn was found negatively correlated with all other micronutrients, which means with the increase in Mn concentration there will be a decrease in the concentration of other micronutrients in common bean seeds. In addition, a positive correlation indicated that increasing the concentration of one micronutrient would positively affect the concentration of other micronutrients.

4.2 GWAS for Micronutrients

Genetic diversity is an important parameter for studying variability in any crop and identifying superior alleles controlling qualitative and quantitative traits through association mapping (Nachimuthu et al., 2015). Molecular markers such as SSRs and SNPs have an important role in studying genetic diversity in most crops (Nachimuthu et al., 2015; Zargar et al., 2016). Insights into the genomic diversity and population structure of common bean germplasm can expedite the genetic gains in common bean-breeding programs (Blair et al., 2012; Blair et al., 2013). The diversity based on the dissimilarity coefficient divided selected germplasm into five main groups and subgroups. The results of the clustering showed that the local germplasm of the common bean of Jammu & Kashmir (J&K) is very diverse and could be used as advanced lines for the genetic enhancement of the common bean. Further, intermixing of released and local lines collected from

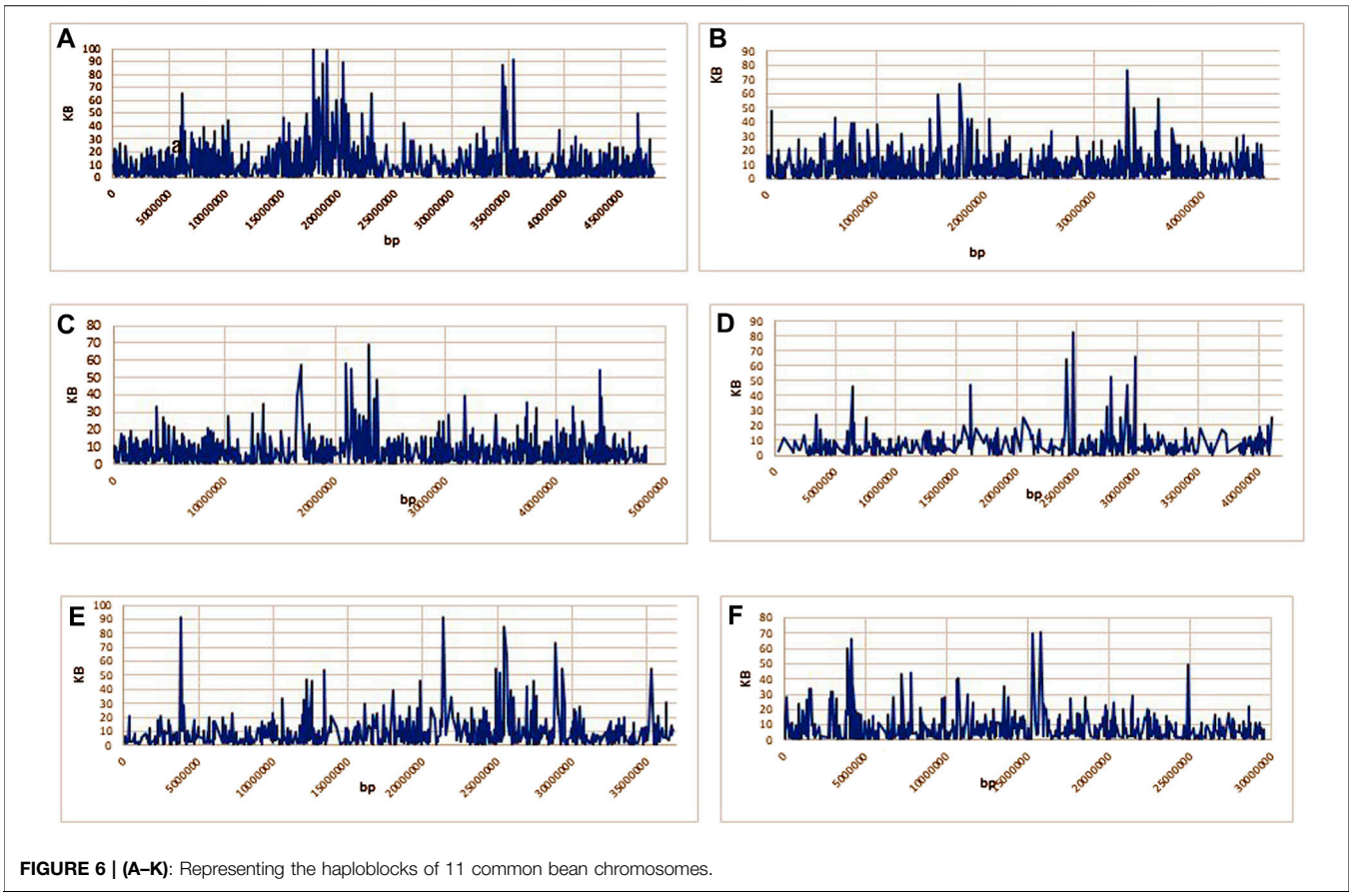


FIGURE 6 | (A–K): Representing the haploblocks of 11 common bean chromosomes.

different regions of J&K indicated that the selected germplasm is diverse. Some of the previously conducted diversity studies on local landraces of the common beans collected from the J&K region have also divided the germplasm into different groups (Zargar et al., 2016; Mahajan et al., 2017a; Mahajan et al., 2017b). In order to have knowledge about the sub-populations in a particular crop structure analysis was performed.

Based on the sharp peak for the delta-K value and the results of the PCA, 96 common bean genotypes were classified into five major groups. A similar pattern of population structure K = 5 was also found in common bean germplasm from Jammu and Kashmir (Mahajan et al., 2017a; Mahajan et al., 2017b). However, earlier studies also classified the common bean germplasm into K = 2 (Gupta et al., 2020; Nkhata et al., 2020; Mir et al., 2021); K = 3 (Blair et al., 2012; Blair et al., 2013; Dennis et al., 2014; Nemli et al., 2014). The difference in cluster and structure analysis could be due to different algorithms used by the two approaches. The cluster analysis is based on evolutionary dissimilarity, while the population structure is based on a Bayesian algorithm. The deviations in the results can be attributed to the different germplasm, the different marker system, and the different geographical locations.

For the identification of genes associated with different traits in a large population, GWAS offers much higher mapping resolutions (Mamo et al., 2014; Norton et al., 2014). To the best of our knowledge, there are only a few association studies on different traits in common bean germplasm collected from the Himalayan region (Mahajan et al., 2017b; Choudhary et al., 2018; Gupta et al., 2020; Mir et al., 2021). However, in earlier studies, genic and genomic SSRs were used for GWAS. The present study is the first report on the association of SNPs related to seed micronutrient content in common bean germplasm from the Himalayan region. In the present study, it was found that SNPs associated with most of the seed micronutrients (Mg, Mn, Fe, Ca, Cu) were found on chr.3 and chr.11 (Mg, Mn, Mo, Ca, Zn) whereas, chr.8 have SNPs associated with Cu, Mg and Mn and chr.9 have SNPs associated with only Mg and Ca. Earlier studies revealed that QTLs linked with Fe content were found on chr. 2, 5, 6, 7, 9, and 10, whereas Zn content was found on chr.1, 2, 3, 5, 7, 8, and 10 (Blair et al., 2011; Mahajan et al., 2017a). In a recent study, a single QTL on chr. 9 and chr. 8 was found to be associated with seed Ca content and seed Mg content, respectively, whereas two QTLs on chr. 6 were associated with seed Zn content (Gunjača et al., 2021). In our study, we found that all of the 11 chromosomes have SNPs associated with more than one seed micronutrient content, which indicates the preciseness of using high throughput genotypic data in the present study.

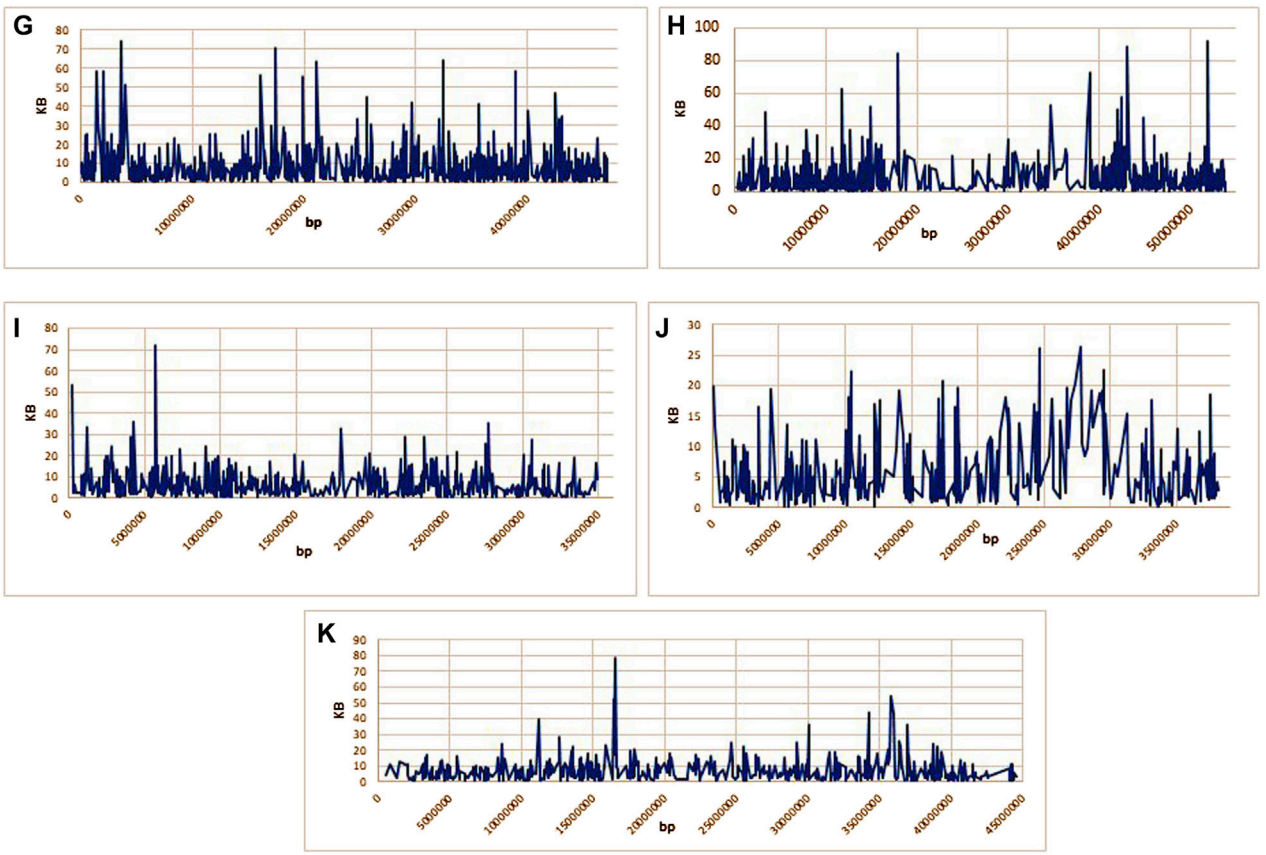


FIGURE 6 | (Continued)

4.3 LD Plot and Haploblock in Common Bean Germplasm

Knowledge about the LD properties in domesticated crops is important as it underlies all types of genetic mapping and may be used in the fine mapping of genes associated with complex traits in crop plants. It is said that in-depth LD in a crop is important for having more SNP-based associations to predict the average number of markers required for GWAS (Nordborg and Tavaré, 2002). Identification of QTLs inassociation mapping is based on Linkage Disequilibrium measurement. The population structure and genetic relatedness between the genotypes can lead to false-positive LD analysis. LD pattern in germplasm is affected by reducing genetic diversity by various factors like the type of selection, population admixture (Contreras-Soto et al., 2017). Our study is in accordance with the previous common bean LD estimation studies (Erdogmus et al., 2020; Gunjača et al., 2021). Hence, we can conclude that as the number of SNPs increases, there will be more R^2 and the higher the likelihood of association of markers with traits of interest. This also indicated that a significant association would be possible with LD block having a higher number of SNPs compared to those windows having lower SNPs. The present study suggests that GBS is an advanced approach to analyzing genetic diversity and population structure in the common bean. The haplotype-based analysis showed that more haplocks were found in the centromeric region than in the telomeric region. In Chr.10 (26,295 kb) higher haplotypes per kb were found than in the rest of the chromosomes, which suggests that the LD decay in Chr.10 is stronger than in other chromosomes.

5 CANDIDATE GENES

GWAS is often used for comparative genome analysis and helps explicitly in dissection as well as in understanding the complex quantitative feature analysis. The GWAS helps identify significant SNPs associated with a trait that is not the function of the region of interest, and sometimes these SNPs are present in non-coding or non-regulatory regions of chromosomes (Bararyenya et al., 2020). Thus, it is important to identify the candidate genes in the vicinity of the significant SNPs, and in our study, the significant SNP regions were examined for the identification of putative protein-coding genes using the *P. vulgaris* genome. The genes that are present in the .1 Mb flanking region of significant SNPs are given in **Supplementary Table S6**. Many previous studies also reported some markers linked with traits but no genes in the

TABLE 4 | Identified gene related to transporter/metal transporter and metal ion binding proteins.

QTL name	Gene ID	Chr name	Start	End	Strand	Description
QTL_Zn_2_37819199	Phvul.002G210700	Chr02	37776912	37782162	−1	ABC-2 TYPE TRANSPORTER
QTL_Zn_2_37819199	Phvul.002G210500	Chr02	37746541	37750458	1	ABC TRANSPORTER
QTL_Zn_2_37819199	Phvul.002G210600	Chr02	37769769	37772725	−1	ABC TRANSPORTER
QTL_Mn_1_12045066	Phvul.001G079200	Chr01	12009949	12013454	−1	CALCIUM-BINDING TRANSPORTER-LIKE PROTEIN
QTL_Mg_1_13513718	Phvul.001G087300	Chr01	13547691	13551468	1	ZINC FINGER PROTEIN CONSTANS-LIKE 14-RELATED
QTL_Mg_1_13577463	Phvul.001G087600	Chr01	13639243	13640935	−1	ZINC FINGER PROTEIN CONSTANS-LIKE 14-RELATED
QTL_Zn_2_37819199	Phvul.002G211500	Chr02	37863083	37865132	−1	NITRATE, FROMATE, IRON DEHYDROGENASE
QTL_Mg_2_39493707	Phvul.002G223300	Chr02	39466477	39468933	1	OXIDOREDUCTASE, 2OG-FE II OXYGENASE FAMILY PROTEIN
QTL_Cu_2_39678800	Phvul.002G224601	Chr02	39628554	39636892	1	ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN// SUBFAMILY NOT NAMED
QTL_Cu_2_39678800	Phvul.002G226100	Chr02	39760216	39762026	−1	SNF2, HELICASE AND ZINC-FINGER DOMAIN-CONTAINING PROTEIN-RELATED
QTL_Mg_2_43486371/ QTL_Cu_2_43570203	Phvul.002G264200	Chr02	43553358	43564776	−1	DOF DOMAIN, ZINC FINGER (ZF-DOF)
QTL_Mg_2_43486371	Phvul.002G262500	Chr02	43427768	43428934	−1	COPPER TRANSPORT PROTEIN ATOX1-RELATED
QTL_Mn_3_37444587	Phvul.003G157800	Chr03	37403436	37405196	−1	3.6.3.41 - HEME-TRANSPORTING ATPASE (1 OF 1)
QTL_Ca_3_38092190	Phvul.003G162600	Chr03	38180169	38182817	1	ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 39
QTL_Ca_3_38092190	Phvul.003G162500	Chr03	38160568	38162888	1	ZINC FINGER PROTEIN
QTL_Ca_3_38092190	Phvul.003G162400	Chr03	38151881	38153482	-1	ZINC FINGER PROTEIN
QTL_Cu_3_44944379	Phvul.003G221000	Chr03	44958664	44959509	-1	C2H2-TYPE ZINC FINGER
QTL_Mn_3_45208418	Phvul.003G223100	Chr03	45277144	45282157	1	OXIDOREDUCTASE, 2OG-FE II OXYGENASE FAMILY PROTEIN
QTL_Ca_5_825380	Phvul.005G008400	Chr05	739318	742827	−1	FOLATE-BIOPTERIN TRANSPORTER 8, CHLOROPLASTIC-RELATED
QTL_Cu_6_7107934	Phvul.006G015300	Chr06	7024910	7028768	1	CALCIUM-DEPENDENT PROTEIN KINASE 17-RELATED
QTL_Mg_6_26644872	Phvul.006G164200	Chr06	26717778	26719484	−1	PREDICTED TRANSPORTER (MAJOR FACILITATOR SUPERFAMILY)
QTL_Mg_6_26644872	Phvul.006G163300	Chr06	26634344	26644212	−1	ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 34-RELATED
QTL_Mg_7_20643908	Phvul.007G111200	Chr07	20693634	20695218	-1	CALCIUM BINDING PROTEIN
QTL_Ca_7_24798685	Phvul.007G149300	Chr07	24734652	2,4741700	-1	MEMBRANE MAGNESIUM TRANSPORTER (MMGT)
QTL_Fe_7_25770068	Phvul.007G152700	Chr07	2,5821756	25822875	1	CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE
QTL_Fe_7_27864996	Phvul.007G165166	Chr07	27868704	27875618	−1	MITOCHONDRIAL METAL TRANSPORTER 1-RELATED
QTL_Ca_7_38956026	Phvul.007G267600	Chr07	38871259	38872086	1	DOF DOMAIN, ZINC FINGER (ZF-DOF
QTL_Mn_8_6963598	Phvul.008G074100	Chr08	7022615	7031313	−1	C2 CALCIUM/LIPID-BINDING ENDONUCLEASE/ EXONUCLEASE/PHOSPHATASE-RELATED
QTL_Cu_8_52162199	Phvul.008G185000	Chr08	52129836	52132078	1	CALCIUM-ACTIVATED CHLORIDE CHANNEL REGULATOR
QTL_Ca_9_8733609	Phvul.009G043800	Chr09	8788738	8789490	−1	C2H2-TYPE ZINC FINGER (ZF-C2H2_6)
QTL_Ca_9_8733609	Phvul.009G042600	Chr09	8633736	8637094	1	PROBABLE ZINC-RIBBON DOMAIN (ZINC_RIBBON_12)
QTL_Ca_9_8733609	Phvul.009G043700	Chr09	8779518	8780270	−1	C2H2-TYPE ZINC FINGER (ZF-C2H2_6)
QTL_Mo_10_31876693	Phvul.010G071100	Chr10	31841020	31842480	−1	C2H2-TYPE ZINC FINGER (ZF-C2H2_6)
QTL_Mo_10_31876693	Phvul.010G071300	Chr10	31949277	31951593	1	C2H2-LIKE ZINC FINGER PROTEIN-RELATED
QTL_Mn_11_5168818/ QTL_Mn_11_5199874	Phvul.011G058100	Chr11	5183094	5185016	−1	AN1-TYPE ZINC FINGER PROTEIN
QTL_Mn_11_5168818/ QTL_Mn_11_5199874	Phvul.011G058500	Chr11	5228604	5232591	1	ZINC TRANSPORTER, ZIP FAMILY (TC.ZIP, ZUPT, ZRT3, ZIP2)
QTL_Ca_11_6191264	Phvul.011G068700	Chr11	6106515	6108448	−1	COPPER TRANSPORT PROTEIN ATOX1-RELATED
QTL_Mo_11_7595234	Phvul.011G081250	Chr11	7526302	7531699	−1	CALCIUM-ACTIVATED CHLORIDE CHANNEL REGULATOR
QTL_Mo_11_10155319/ QTL_Mo_11_10212465	Phvul.011G098800	Chr11	10254133	10256987	−1	ALUMINUM-ACTIVATED MALATE TRANSPORTER 1-RELATED
QTL_Mo_11_10700430	Phvul.011G100400	Chr11	10792608	10796881	1	CALCIUM-DEPENDENT PHOSPHOTRIESTERASE SUPERFAMILY PROTEIN-RELATED
QTL_Mo_11_11345383/ QTL_Mo_11_11415382	Phvul.011G102500	Chr11	11388457	11393199	−1	MULTI-COPPER OXIDASE

genomic region of markers (Bararyenya et al., 2020). The number of genes in SNPs revealed that chromosome 2, 6, 9, and 11 has more number of genes in 0.1 MB flanking region of SNPs has a high density of genes or the hot spot of QTLs.

6 CONCLUSION

Micronutrient deficiency is the leading cause of human health deterioration worldwide. An animal or plant-based diet alone

cannot provide humans with excessive amounts of essential micronutrients. Biofortification of food crops is therefore important in order to provide humans with essential micronutrients. Modern breeding approaches such as QTL mapping and association mapping are important to identify QTLs that are related to micronutrient levels in plants. In the present study, a significant variation in the micronutrient content of the seeds in the germplasm of common beans was found. The present data on genetic loci, particularly the key SNPs associated with seven elements, will be helpful in identifying candidate genes, understanding molecular mechanisms, and developing molecular markers for breeding applications. We firmly believe that the results of the current studies will help accelerate bean biofortification efforts to overcome nutritional deficiencies.

DATA AVAILABILITY STATEMENT

The data sets 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

MN did experimental work and wrote the manuscript; RM helped in data analysis and editing the manuscript; SM helped in writing the manuscript; SR helped in experimental work; RAM helped with writing the manuscript and analysis; RS helped in data analysis; VT and VK: molecular data analysis; PS helped in fieldwork; HE-S and DH help in editing the manuscript; SZ developed the concept and guided all the laboratory work and edited the manuscript.

ACKNOWLEDGMENTS

MN and SZ acknowledge the financial support of SERB, New Delhi (Project File No: PDF/2016/000486) for the award of the National Postdoc Fellowship (NPDF) to MN under the mentorship of SZ. SZ also acknowledges Dr. Swarup Parida (NIPGR, New Delhi, India) for his support in genotyping; Dr. Humaira Sonah and Dr. Rupesh Deshmukh (NABI, Mohali, India) for their support in data analysis and editing the manuscript. SZ also acknowledges the Hon'ble Vice-Chancellor SKUAST-K (Prof. Nazir A. Ganai) for his support. The authors would like to extend their sincere appreciation to the Researchers Supporting Project Number (RSP-2021/19), King Saud University, Riyadh, Saudi Arabia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.750814/full#supplementary-material

REFERENCES

Ali, M., Choudhury, T. R., Hossain, B., and Ali, M. P. (2014). Determination of Traces of Molybdenum and lead in Foods by X-ray Fluorescence Spectrometry. *SpringerPlus* 3. doi:10.1186/2193-1801-3-341

Bararyenya, A., Olukolu, B. A., Tukamuhabwa, P., Grüneberg, W. J., Ekaya, W., Low, J., et al. (2020). Genome-wide Association Study Identified Candidate Genes Controlling Continuous Storage Root Formation and Bulking in Hexaploid Sweetpotato. *BMC Plant Biol.* 20, 3. doi:10.1186/s12870-019-2217-9

Beebe, S., Gonzalez, A. V., and Rengifo, J. (2000). Research on Trace Minerals in the Common Bean. *Food Nutr. Bull.* 21, 387–391. doi:10.1177/156482650002100408

Bellucci, E., Bitocchi, E., Rau, D., Rodriguez, M., Biagetti, E., Giardini, A., et al. (2014). *Genomics of Origin, Domestica-Tion and Evolution of Phaseolus vulgaris*. doi:10.1007/978-94-007-7572-53

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., et al. (2012). Mesoamerican Origin of the Common Bean (Phaseolus vulgaris L.) Is Revealed by Sequence Data. *Proceedings of the National Academy of Sciences of the United States of America*, 109. doi:10.1073/pnas.1108973109PMID:22393017

Bjorklund, G. (2013). The Role of Zinc and Copper in Autism Spectrum Disorders. *Acta Neurobiol. Exp. Wars* 73, 225–236.

Blair, M. W., Astudillo, C., Rengifo, J., Beebe, S. E., and Graham, R. (2011). QTL Analyses for Seed Iron and Zinc Concentrations in an Intra-genepool Population of Andean Common Beans (Phaseolus vulgaris L.). *Theor. Appl. Genet.* 122, 511–521. doi:10.1007/s00122-010-1465-8

Blair, M. W., Brondani, R. V. P., Díaz, L. M., and Del Peloso, M. J. (2013). Diversity and Population Structure of Common Bean from Brazil. *Crop Sci.* 53, 1983–1993. doi:10.2135/cropsci2012.09.0516

Blair, M. W., Soler, A., and Cortés, A. J. (2012). Diversification and Population Structure in Common Beans (Phaseolus vulgaris L.). *PLoS ONE* 7, e49488. doi:10.1371/journal.pone.0049488

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

Câmara, C., Urrea, C., and Schlegel, V. (2013). Pinto Beans (Phaseolus vulgaris l.) as a Functional Food: Implications on Human Health. *Agriculture* 3, 90–111. doi:10.3390/agriculture3010090

Choudhary, N., Bawa, V., Paliwal, R., Singh, B., Bhat, M. A., Mir, J. I., et al. (2018). 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

Contreras-Soto, R. I., de Oliveira, M. B., Costenaro-da-Silva, D., Scapim, C. A., and Schuster, I. (2017). Population Structure, Genetic Relatedness and Linkage Disequilibrium Blocks in Cultivars of Tropical Soybean (Glycine max). *Euphytica* 213, 173. doi:10.1007/s10681-017-1966-5

Darnton-Hill, I., Bloem, M. W., and Chopra, M. (2006). Achieving the Millennium Development Goals through Mainstreaming Nutrition: Speaking with One Voice. *Public Health Nutr.* 9, 537–539. doi:10.1079/phn2006965

Dennis, O., Phinehas, T., James, K., Annet, N., Pamela, P., and Michael, U. (2014). The Genetic Diversity and Population Structure of Common Bean (Phaseolus vulgaris L) Germplasm in Uganda. *Afr. J. Biotechnol.* 13, 2935–2949. doi:10.5897/AJB2014.13916

Diapari, M., Sindhu, A., Bett, K., Deokar, A., Warkentin, T. D., Taran, B., et al. (2014). Genetic Diversity and Association Mapping of Iron and Zinc Concentrations in Chickpea (Cicer Arietinum L.). *Genome* 57. doi:10.1139/gen-2014-0108

Diapari, M., Sindhu, A., Warkentin, T. D., Bett, K., and Tar'an, B. (2015). Population Structure and Marker-Trait Association Studies of Iron, Zinc and Selenium Concentrations in Seed of Field Pea (Pisum Sativum L.). *Mol. Breed.* 35. doi:10.1007/s11032-015-0252-2

Doyle, J. J., and Doyle, L. H. (1990). Isolation of Plant DNA from Fresh Tissue. *Focus* 12, 13–15.

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. doi:10.1007/s12686-011-9548-7

Erdogmus, S., Ates, D., Nemli, S., Yagmur, B., Asciogul, T. K., Ozkuru, E., et al. (2020). Genome-wide Association Studies of Ca and Mn in the Seeds of the Common Bean (Phaseolus vulgaris L.). *Genomics* 112. doi:10.1016/j.ygeno.2020.03.030

Faostat (2017). FAOSTAT. Available at: http://www.fao.org/faostat/en/#data/QC2017 (Accessed March 21, 2021).

Felsenstein, J. (1989). {PHYLIP} – Phylogeny Inference Package ({V}ersion 3.2). *Cladistics* 5. doi:10.1016/0168-9525(89)90204-7

Fennema owen, R. (2000). in *Food Chemistry*. Editor R. Fennema owen. 3rd ed. (NEW YORKNEW YORK: Marcel Dekker).

Gelin, J. R., Forster, S., Grafton, K. F., McClean, P. E., and Rojas-Cifuentes, G. A. (2007). Analysis of Seed Zinc and Other Minerals in a Recombinant Inbred Population of Navy Bean (Phaseolus vulgaris L.). *Crop Sci.* 47. doi:10.2135/cropsci2006.08.0510

Gunjača, J., Carović-Stanko, K., Lazarević, B., Vidak, M., Petek, M., Liber, Z., et al. (2021). Genome-Wide Association Studies of Mineral Content in Common Bean. *Front. Plant Sci.* 12, 636484. doi:10.3389/fpls.2021.636484

Gupta, N., Zargar, S. M., Singh, R., Nazir, M., Mahajan, R., and Salgotra, R. K. (2020). Marker Association Study of Yield Attributing Traits in Common Bean (Phaseolus vulgaris L.). *Mol. Biol. Rep.* 47, 6769–6783. doi:10.1007/s11033-020-05735-6

Hardiman, R. T., Banin, A., and Jacoby, B. (1984). The Effect of Soil Type and Degree of Metal Contamination upon Uptake of Cd, Pb and Cu in bush Beans (Phaseolus vulgaris L.). *Plant Soil* 81, 3–15. doi:10.1007/BF02206889

Hayat, I., Ahmad, A., Masud, T., Ahmed, A., and Bashir, S. (2014). Nutritional and Health Perspectives of Beans (Phaseolus vulgaris L.): An Overview. *Crit. Rev. Food Sci. Nutr.* 54, 580–592. doi:10.1080/10408398.2011.596639

House, W. A., Welch, R. M., Beebe, S., and Cheng, Z. (2002). Potential for Increasing the Amounts of Bioavailable Zinc in Dry Beans (Phaseolus vulgaris L) through Plant Breeding. *J. Sci. Food Agric.* 82. doi:10.1002/jsfa.1146

Hyun, M. K., 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. doi:10.1534/genetics.107.080101

Jan, S., Rather, I. A., Sofi, P. A., Wani, M. A., Sheikh, F. A., Bhat, M. A., et al. (2021). Characterization of Common Bean (*PHASEOLUS VULGARIS* L.) Germplasm for Morphological and Seed Nutrient Traits from Western Himalayas. *Legume Sci.* 3. doi:10.1002/leg3.86

Kumar, V., and Chopra, A. K. (2014). Accumulation and Translocation of Metals in Soil and Different Parts of French Bean (Phaseolus Vulgaris L.) Amended with Sewage Sludge. *Bull. Environ. Contam. Toxicol.* 92. doi:10.1007/s00128-013-1142-0

Lehmann, A., and Rillig, M. C. (2015). Arbuscular Mycorrhizal Contribution to Copper, Manganese and Iron Nutrient Concentrations in Crops - A Meta-Analysis. *Soil Biol. Biochem.* 81. doi:10.1016/j.soilbio.2014.11.013

Lewontin, R. C., and Kojima, K. (1960). The Evolutionary Dynamics of Complex Polymorphisms. *Evolution* 14. doi:10.2307/2405995

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. doi:10.1093/bioinformatics/bts444

Liu, J., Wang, W., Mei, D., Wang, H., Fu, L., Liu, D., et al. (2016). Characterizing Variation of branch Angle and Genome-wide Association Mapping in Rapeseed (Brassica Napus L.). *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.00021

Loiselle, B. A., Sork, V. L., Nason, J., and Graham, C. (1995). Spatial Genetic Structure of a Tropical Understory Shrub, Psychotria Officinalis (Rubiaceae). *Am. J. Bot.* 82. doi:10.2307/244586910.1002/j.1537-2197.1995.tb12679.x

Lu, Q., Zhang, M., Niu, X., Wang, S., Xu, Q., Feng, Y., et al. (2015). Genetic Variation and Association Mapping for 12 Agronomic Traits in Indica rice. *BMC Genomics* 16. doi:10.1186/s12864-015-2245-2

Mahajan, R., Zargar, S. M., Aezum, A. M., Farhat, S., Gani, M., Hussain, I. S. A., et al. (2015). Evaluation of Iron, Zinc, and Protein Contents of Common Bean (Phaseolus vulgaris L.) Genotypes: A Collection from Jammu & Kashmir, India. *Legume Genomics Genet.* doi:10.5376/lgg.2015.06.0002

Mahajan, R., Zargar, S. M., Salgotra, R. K., Singh, R., Wani, A. A., Nazir, M., et al. (2017a). Linkage Disequilibrium Based Association Mapping of Micronutrients in Common Bean (Phaseolus vulgaris L.): a Collection of Jammu & Kashmir, India. *Biotech* 3, 7. doi:10.1007/s13205-017-0928-x

Mahajan, R., Zargar, S. M., Singh, R., Salgotra, R. K., Farhat, S., and Sonah, H. (2017b). Population Structure Analysis and Selection of Core Set Among Common Bean Genotypes from Jammu and Kashmir, India. *Appl. Biochem. Biotechnol.* 182. doi:10.1007/s12010-016-2307-1

Mamo, B. E., Barber, B. L., and Steffenson, B. J. (2014). Genome-wide Association Mapping of Zinc and Iron Concentration in Barley Landraces from Ethiopia and Eritrea. *J. Cereal Sci.* 60. doi:10.1016/j.jcs.2014.08.007

Martinez Meyer, M. R., Rojas, A., Santanen, A., and Stoddard, F. L. (2013). Content of Zinc, Iron and Their Absorption Inhibitors in Nicaraguan Common Beans (Phaseolus vulgaris L.). *Food Chem.* 136. doi:10.1016/j.foodchem.2012.07.105

Mir, R. R., Choudhary, N., Bawa, V., Jan, S., Singh, B., Bhat, M. A., et al. (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

Mukamuhirwa, F., Tusiime, G., Mukankusi, C., Gibson, P., and Edema, R. (2012). Potential Sources of High Iron and Zinc Content in Ugandan Bean Germplasm. *In: Third RUFORUM Biennial Meeting* 24-28 September 2012, Entebbe, Uganda. 153–157.

Nachimuthu, V. V., Raveendran, M., Duraialaguraja, S., Sivakami, R., Pandian, B. A., Ponniah, G., et al. (2015). Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers: An Initiative towards Association Mapping of Agronomic Traits in Oryza Sativa. *Rice* 8. doi:10.1186/s12284-015-0062-5

Nadeem, M. A., Karaköy, T., Yeken, M. Z., Habyarimana, E., Hatipoğlu, R., Çiftçi, V., et al. (2020). Phenotypic Characterization of 183 Turkish Common Bean Accessions for Agronomic, Trading, and Consumer-Preferred Plant Characteristics for Breeding Purposes. *Agronomy* 10, 272. doi:10.3390/agronomy10020272

Nadeem, M. A., Yeken, M. Z., Shahid, M. Q., Habyarimana, E., Yılmaz, H., Alsaleh, A., et al. (2021). Common Bean as a Potential Crop for Future Food Security: an Overview of Past, Current and Future Contributions in Genomics, Transcriptomics, Transgenics and Proteomics. *Biotechnol. Biotechnol. Equip.* 35, 758–786. doi:10.1080/13102818.2021.1920462

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., and Hohenlohe, P. A. (2013). Genotyping-by-sequencing in Ecological and Conservation Genomics. *Mol. Ecol.* 22. doi:10.1111/mec.12350

Nemli, S., Asciogul, T. K., Esiyok, D. D. A., and Tanyolac, M. B. (2016). Genome Wide Association Mapping (GWAS) of Nutrient Elements in Seed Using Genotyping by Sequencing (GWAS) in Common Bean (Phaseolus vulgaris L). in Proceeding of plant and animal genome conference XXIV.

Nemli, S., Asciogul, T. K., Kaya, H. B., Kahraman, A., Eşiyok, D., and Tanyolac, B. (2014). Association Mapping for Five Agronomic Traits in the Common Bean (*Phaseolus vulgaris* L.): Association Mapping for Five Agronomic Traits. *J. Sci. Food Agric.* 94, 3141–3151. doi:10.1002/jsfa.6664

Nkhata, W., Shimelis, H., Melis, R., Chirwa, R., Mzengeza, T., Mathew, I., et al. (2020). Population Structure and Genetic Diversity Analyses of Common Bean Germplasm Collections of East and Southern Africa Using Morphological Traits and High-Density SNP Markers. *PLOS ONE* 15, e0243238. doi:10.1371/journal.pone.0243238

Nordborg, M., and Tavaré, S. (2002). Linkage Disequilibrium: what History Has to Tell Us. *Trends Genet.* 18, 890. doi:10.1016/S0168-9525(02)02557-X

Norton, G. J., Douglas, A., Lahner, B., Yakubova, E., Guerinot, M. L., Pinson, S. R. M., et al. (2014). Genome Wide Association Mapping of Grain Arsenic, Copper, Molybdenum and Zinc in rice (Oryza Sativa L.) Grown at Four International Field Sites. *PLoS ONE* 9. doi:10.1371/journal.pone.0089685

O Abidemi, O., Shittu, S., and Rebbeca, A. O. (2012). Evaluation of Heavy Metals in Three Common Nigerian Cowpea (Vigna Unguiculata) Paste End Product ("Moinmoin") Using Different Packaging Materials-Indian Journals. *Int. J. Environ. Sci.*, 833–840.

Ozkuru, E., Ates, D., Nemli, S., Erdogmus, S., Karaca, N., Yilmaz, H., et al. (2019). Association Mapping of Loci Linked to Copper, Phosphorus, and Potassium

Concentrations in the Seeds of C. Arietinum and *C. reticulatum*. *Genomics* 111. doi:10.1016/j.ygeno.2018.12.010

Papamargaritis, D., Aasheim, E. T., Sampson, B., and le Roux, C. W. (2015). Copper, Selenium and Zinc Levels after Bariatric Surgery in Patients Recommended to Take Multivitamin- mineral Supplementation. *J. Trace Elem. Med. Biol.* 31, 167–172. doi:10.1016/j.jtemb.2014.09.005

Peterson, G. W., Dong, Y., Horbach, C., and Fu, Y. B. (2014). Genotyping-by-sequencing for Plant Genetic Diversity Analysis: A Lab Guide for SNP Genotyping. *Diversity* 6. doi:10.3390/d6040665

Pfeiffer, W. H., and McClafferty, B. (2007). "HarvestPlus: Breeding Crops for Better Nutrition," in *Crop Science*. doi:10.2135/cropsci2007.09.0020IPBS

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945959.

Ribeiro, N. D., Jost, E., Maziero, S. M., Storck, L., and Rosa, D. P. (2014). Selection of Common Bean Lines with High Grain Yield and High Grain Calcium and Iron Concentrations. *Rev. Ceres* 61. doi:10.1590/S0034-737X2014000100010

Reidinger, S., Ramsey, M. H., and Hartley, S. E. (2012). Rapid and Accurate Analyses of Silicon and Phosphorus in Plants Using a Portable X-ray Fluorescence Spectrometer. *New Phyt.* 195, 699706. doi:10.1111/j.1469-8137.2012.04179.x

Schilling, M. P., Wolf, P. G., Duffy, A. M., Rai, H. S., Rowe, C. A., Richardson, B. A., et al. (2014). Genotyping-by-sequencing for Populus Population Genomics: An Assessment of Genome Sampling Patterns and Filtering Approaches. *PLoS ONE* 9. doi:10.1371/journal.pone.0095292

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. doi:10.1038/ng.3008

Shao, Y., Jin, L., Zhang, G., Lu, Y., Shen, Y., and Bao, J. (2011). Association Mapping of Grain Color, Phenolic Content, Flavonoid Content and Antioxidant Capacity in Dehulled rice. *Theor. Appl. Genet.* 122. doi:10.1007/s00122-010-1505-4

Silva, C. A., De Fátima Barbosa Abreu, Â., Antonio, M., Ramalho, P., Duarte Correa, A., Gontijo, L., et al. (2010). Genetic Variability for Protein and Minerals Content in Common Bean Lines. *Phaseolus vulgaris, LEMBRAPA, Embrapa Scientific Information-Alice*, Brazil.

Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kaspryzyk, A. (2009). BioMart–Biological Queries Made Easy. *BMC Genomics* 10, 22. doi:10.1186/1471-2164-10-22

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An Improved Genotyping by Sequencing (GWAS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. *PLoS ONE* 8. doi:10.1371/journal.pone.0054603

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* 28. doi:10.1093/molbev/msr121

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., et al. (2011). Genome-wide Association Study of Leaf Architecture in the maize Nested Association Mapping Population. *Nat. Genet.* 43. doi:10.1038/ng.746

Upadhyaya, H. D., Bajaj, D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Genetic Dissection of Seed-Iron and Zinc Concentrations in Chickpea. *Sci. Rep.* 6. doi:10.1038/srep24050

Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., and Bhatia, S. (2015). High- Density Linkage Map Construction and Mapping of Seed Trait QTLs in Chickpea (Cicer Arietinum L.) Using Genotyping-By-Sequencing (GWAS). *Sci. Rep.* 5. doi:10.1038/srep17512

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., et al. (2012). Genome-wide Association Study (GWAS) of Resistance to Head Smut in maize. *Plant Sci.* 196. doi:10.1016/j.plantsci.2012.08.004

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., et al. (2014). Metabolome-based Genome-wide Association Study of maize Kernel Leads to Novel Biochemical Insights. *Nat. Commun.* 5. doi:10.1038/ncomms4438

Yang, Z., Li, Z., and Bickel, D. R. (2013). Empirical Bayes Estimation of Posterior Probabilities of Enrichment: A Comparative Study of Five Estimators of the Local False Discovery Rate. *BMC Bioinformatics* 14. doi:10.1186/1471-2105-14-87

Yeken, M. Z. Assessment of Mineral Content Variations for Biofortification of the Bean Seed FasulyeTohumunun Biyofortifikasyonuic in MineralÌc erikVasyasyonlarının Değerlendirilmesi.2018;4: 261–269. doi:10.24180/ijaws.455311

Yu, J., and Buckler, E. S. (2006). Genetic Association Mapping and Genome Organization of maize. *Curr. Opin. Biotechnol.* 17. doi:10.1016/j.copbio.2006.02.003

Zaccardelli, M., Pentangelo, A., and Tripodi, P. (2013). Characterization of bean (Phaseolus vulgaris L.) ecotype "Fagiolo occhio nero di Oliveto Citra" using agronomic, biochemical and molecular approaches. *Pak. J. Biol. Sci.* 16. doi:10.3923/pjbs.2013.901.910

Zargar, S. M., Farhat, S., Mahajan, R., Bhakhri, A., and Sharma, A. (2016). Unraveling the Efficiency of RAPD and SSR Markers in Diversity Analysis and Population Structure Estimation in Common Bean. *Saudi J. Biol. Sci.* 23. doi:10.1016/j.sjbs.2014.11.011

Zargar, S. M., Mahajan, R., Farhat, S., Nazir, M., Mir, R. A., Nazir, M., et al. (2015). Understanding the Role of Iron and Zinc in Animals and Crop Plants from Genomics Perspective. *Curr. Trends Biotechnol. Pharm.* 9.

Zargar, S. M., Sharma, A., Sadhu, A., Agrawal, G. K., and Rakwal, R. (2014). Exploring Genetic Diversity in Common Bean from Unexploited Regions of Jammu & Kashmir-India. *Mol. Plant Breed.* doi:10.5376/mpb.2014.05.0002

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 © 2022 Nazir, Mahajan, Mansoor, Rasool, Mir, Singh, Thakral, Kumar, Sofi, El-Serehy, Hefft and Zargar. 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 2022
doi: 10.3389/fgene.2022.845602

High-Density Genetic Variation Map Reveals Key Candidate Loci and Genes Associated With Important Agronomic Traits in Peanut

Huiling Zhao[1,2†], Ruizheng Tian[1†], Han Xia[1,2†], Changsheng Li[1], Guanghui Li[1], Aiqin Li[1], Xianying Zhang[1], Ximeng Zhou[1,2], Jing Ma[1,2], Huailing Huang[1,2], Kun Zhang[1,3], Mahendar Thudi[1,4], Changle Ma[2], Xingjun Wang[1,2] and Chuanzhi Zhao[1,2*]*

[1]Institute of Crop Germplasm Resources (Institute of Biotechnology), Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan, China, [2]College of Life Sciences, Shandong Normal University, Jinan, China, [3]College of Agricultural Science and Technology, Shandong Agriculture and Engineering University, Jinan, China, [4]Rajendra Prasad Central Agricultural University, Samsthipur, India

OPEN ACCESS

Edited by:
Ahmed Sallam, Assiut University, Egypt

Reviewed by:
Sunil S. Gangurde, University of Georgia, United States
Xiaohong Yang, China Agricultural University, China
Arun K. Pandey, China Jiliang University, China

***Correspondence:**
Xingjun Wang
xingjunw@hotmail.com
Chuanzhi Zhao
chuanzhiz@126.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: *30 December 2021*
Accepted: *22 February 2022*
Published: *25 March 2022*

Citation:
Zhao H, Tian R, Xia H, Li C, Li G, Li A, Zhang X, Zhou X, Ma J, Huang H, Zhang K, Thudi M, Ma C, Wang X and Zhao C (2022) High-Density Genetic Variation Map Reveals Key Candidate Loci and Genes Associated With Important Agronomic Traits in Peanut. Front. Genet. 13:845602. doi: 10.3389/fgene.2022.845602

Peanut is one of the most important cash crops with high quality oil, high protein content, and many other nutritional elements, and grown globally. Cultivated peanut (*Arachis hypogaea* L.) is allotetraploid with a narrow genetic base, and its genetics and molecular mechanisms controlling the agronomic traits are poorly understood. Here, we report a comprehensive genome variation map based on the genotyping of a panel of 178 peanut cultivars using *Axiom_Arachis2* SNP array, including 163 representative varieties of different provinces in China, and 15 cultivars from 9 other countries. According to principal component analysis (PCA) and phylogenetic analysis, the peanut varieties were divided into 7 groups, notable genetic divergences between the different areas were shaped by environment and domestication. Using genome-wide association study (GWAS) analysis, we identified several marker-trait associations (MTAs) and candidate genes potentially involved in regulating several agronomic traits of peanut, including one MTA related with hundred seed weight, one MTA related with total number of branches, and 14 MTAs related with pod shape. This study outlines the genetic basis of these peanut cultivars and provides 13,125 polymorphic SNP markers for further distinguishing and utility of these elite cultivars. In addition, the candidate loci and genes provide valuable information for further fine mapping of QTLs and improving the quality and yield of peanut using a genomic-assisted breeding method.

Keywords: peanut, GWAS, SNP array, molecular markers, agronomic traits

INTRODUCTION

Cultivated peanut (*Arachis hypogaea*. L) is one of the most important oil crops and cash crops in the world. In 2019, the total production of peanuts was approximately 48.8 million tons (http://www.fao.org). Peanuts are full of high-quality vegetable oil, protein, more than 30 essential vitamins, and many other nutrients, and is part of a balanced diet for human health. Peanuts are widely planted in India, China, United States, Argentina, Australia, and Brazil. In the past 20 years, the average yield of peanut increased from 1.95 t/ha in 1999 to 3.3 t/ha in 2019 in the above six countries

(http://www.fao.org). However, peanut is often grown on marginal soils with lesser inputs and usually intercropped with cereals in many countries such as China and India, the top two countries with the largest peanut harvest area. At present, almost all peanut varieties are developed through conventional breeding methods, marker-assisted selection (MAS) technology was only used in a few traits in the peanut breed program, such as oleic acid content, and root-knot nematode resistance (Chu et al., 2011; Shasidhar et al., 2020). The average yield of peanut is significantly lower than that of the staple food crop, rice, and corn. Peanut germplasm resources have a narrow genetic background. It has been difficult to get significant improvement of the yield and quality through traditional cross-breeding. In the future, MAS will be an important alternative approach for increasing the yield and improving the quality of peanut.

Compared to the conventional breeding approach, MAS technology can significantly accelerate breeding process and improve breeding efficiency by increasing the genetic gains per selection cycles (Collard and Mackill, 2007; Varshney et al., 2013). The utility of MAS is becoming more and more popular in crop breeding programs. For example, in wheat, hundreds of resistance (R) genes to powdery mildew, leaf rust, and stripe rust have been mapped (Pinto da Silva et al., 2018; Qureshi et al., 2018; Shah et al., 2018), and many of them have been successfully used to improve the resistance of wheat through MAS. For MAS technology, identification of quantitative trait loci (QTL) or genes, and development of the closely linked markers is necessary. With the availability of genome sequence information and high-throughput genotyping technologies, genome-wide association study (GWAS) has become a powerful way to identify the tightly linked markers and QTLs from the genome, superseding the traditional QTL mapping method from the structured populations derived from two parents (Pujar et al., 2020). GWAS has been successfully used in identifying the QTLs and the key genes related with the complex traits on peanut. (Gangurde et al., 2020). In peanut, the markers associated with oil, protein, oleic acid, and linoleic acid through a preliminary GWAS analysis with 120 simple sequence repeat (SSR) and transposable element (TE) markers have been reported (Zhang et al., 2020b). Recently, using genotyping-by-sequencing based SNP markers, 79 loci significantly associated for the six yield-related traits were also reported (Zhou et al., 2021).

In the last decade, advances in high throughput sequencing and bioinformatics technologies provided a good platform for peanut genome research including marker development and trait mapping as well as development of molecular breeding products (Zhao et al., 2017; Han et al., 2018; Luo et al., 2018; Bertioli et al., 2019; Zhuang et al., 2019; Ma et al., 2020; Pandey et al., 2020; Zhao et al., 2020). The availability of large-scale genomic resources was used for identifying a large number of genome-wide SNPs, and high-throughput genotyping platforms like 48 K SNP array (*Axiom_Arachis*2, version II) (Clevenger et al., 2018) and 58K SNP array (Pandey et al., 2017). The 48 K SNP array is also used for GWAS analysis for 96 peanut genotypes and revealed that current Korean genetic resources lacked variability compared to US mini-core genotypes (Nabi et al., 2021). Zhang et al. reported the identification of 36 QTLs related with the 13 nutrient elements and 46 QTLs related with leaf spots resistance using the SNP array based GWAS analysis for 120 mini-core germplasms (Zhang H. et al., 2019; Zhang et al., 2020a). All these studies have successfully discovered loci associated with the agronomic traits of peanut.

In this study, we analyzed 20 phenotypic traits of 178 peanut cultivars from diverse origins in China and 9 other countries. We analyzed the genetic divergences of these peanut cultivars and identified a number of significant genetic loci related to phenotypic traits, which will be helpful for further fine mapping and genomic-assisted breeding.

MATERIALS AND METHODS

Plant Materials

For genome-wide association study, a set of 178 peanut varieties (of which 119 varieties represented varieties from 13 provinces and regions including Shandong, Henan, Guangdong, and Fujian; 41 featured germplasm resources preserved in laboratories, and 18 were exotic, i.e., imported from resources abroad) were selected based on phenotypic characteristics, including plant height, number of total branches, seed and pod size, protein, etc. The botanical information for the selected Chinese varieties was derived from available monographs (Feng, 1987; Yu, 2008) and an online database (http://www.peanutdata.cn). The 178 peanut varieties consisted of *var. hypogaea*, *var. vulgaris*, *var. fastigiata*, *var. hirsuta,* and irregular type varieties. The detailed information of each sample was also listed on the **Supplementary File S1**.

Phenotyping for Agronomic Traits

The test materials used were planted in the Jiyang Agricultural Planting base in Jinan City in the summer for 3 consecutive years from 2018 to 2020. The field experiment is a completely random design and adopts conventional cultivation management (Wan, 2003). At harvest, there are three peanut plants randomly selected from each peanut variety and the main stem height, lateral branch angle, total number of branches, pod length, seed length, linoleic acid content, and other traits were measured. The phenotypic data obtained were analyzed using Excel data analysis tools for descriptive statistics and normal distribution test, and Origin software was used for drawing.

Genotyping of Peanut Cultivars

DNA was extracted from 15-day-old seedlings using Plant Genome Extraction Kit (Beijing, China), following the manufacturer's instructions (https://www.tiangen.com/). The DNA was visualized in agarose gel containing Super GelRed (US Everbright Inc., Suzhou, China), and then quality and concentration were determined using Nanodrop™ 2000 spectrophotometer (Thermo Scientific, Shanghai, China). The second-generation 48K SNP array of peanut was used to obtain genotyping data of 178 materials (Clevenger et al., 2018). SNPs with low call rates were removed with selection criteria of missing data rate (>10%) and minor allele frequency (<5%). Only

TABLE 1 | Phenotypic statistics of peanut major agronomic traits.

Traits	Abbreviation	Maximum	Minimum	Median	Mean	Standard deviation	Coefficient of variation (%)	Skewness	Kurtosis
Lateral branch length (cm)	LBL	99.0	44.20	66.00	67.73	12.72	19.08	0.40	–0.52
Main stem height (cm)	MSH	82.80	34.20	51.20	52.95	10.08	19.00	0.68	–0.01
Lateral branch angle (°)	LBA	90.00	30.00	50.00	53.32	13.82	25.91	0.33	–0.51
Total number of branches	TNB	16.40	3.60	8.80	8.72	2.19	25.16	0.29	0.59
Pod branching number	PBN	12.00	2.60	6.40	6.48	1.62	24.96	0.50	1.00
Pod number per plant	PNP	38.80	3.80	17.60	17.90	7.10	39.90	0.46	0.05
Hundred pod weight (gm)	HPW	329.20	57.20	194.40	190.05	51.04	26.86	–0.10	–0.20
Hundred seed weight (gm)	HSW	113.00	29.50	72.40	72.92	17.89	24.54	–0.11	–0.71
Pod length (mm)	PL	50.96	21.60	35.18	35.44	5.31	15.03	0.16	–0.29
Pod width (mm)	PW	20.02	10.01	14.97	14.79	2.05	13.80	–0.12	–0.51
Seed length (mm)	SL	23.23	7.63	17.16	17.30	6.36	14.10	–0.27	0.81
Seed width (mm)	SW	13.30	6.70	9.26	9.40	1.28	13.65	0.30	–0.18
Peel thickness	PT	2.69	0.52	1.25	1.25	0.37	30.70	0.99	2.07
Filled pods number	FPN	30.60	1.00	13.00	12.97	6.87	52.88	0.39	0.05
Oleic acid content (%)	OAC	82.02	28.51	42.15	42.15	12.36	27.78	1.41	1.63
Linoleic acid content (%)	LAC	49.15	4.66	38.81	36.69	10.27	27.99	–1.57	2.06
Behenic acid content (%)	BAC	3.08	2.37	2.66	2.64	0.18	6.70	0.33	–0.96
Palmitic acid content (%)	PAC	13.55	6.26	11.52	11.24	1.48	13.13	–1.40	1.97
Arachidic acid content (%)	AAC	1.75	0.12	1.23	1.21	0.27	22.47	–1.35	4.17
Stearic acid content (%)	SAC	3.07	0.05	2.00	1.93	0.56	28.39	–0.87	1.76

high-quality SNPs were selected for further analysis. Reference genome builds were acquired from arahy (https://peanutbase.org/peanut_genome).

Population Genetic Analysis

The phylogenetic tree was constructed based on the SNPs identified above by maximum likelihood (ML) method in IQ-tree v1.6.12 (Minh et al., 2020) (http://www.iqtree.org/), which was visualized with ITOL software (Letunic and Bork, 2021) (https://itol.embl.de/). The bootstrap values were calculated with 1000 replicates. The population structure of the 178 samples was first evaluated using PCA by the GCTA package and later using Admixture v1.3.0 (Alexander et al., 2009). We used the default parameters in Admixture to test the number of ancestral populations (K) with a cross-validation (CV) process, and the one with minimum CV error calculated was selected as best K value (http://software.genetics.ucla.edu/admixture/admixture-manual.pdf) (Alexander et al., 2009), which was visualized in R script next.

Genome-Wide Association Study

The TASSEL v5.2.1 software was used for the genome-wide association study (GWAS) analysis of 7 aforementioned yield and quality related traits with the high-quality SNPs (Bradbury et al., 2007). Both generalized linear model (GLM) and mixed linear model (MLM) were used to determine MTAs. In general, the GLM model focuses on the SNP effects, which only contains the fixed effects such as population structure and genotype, and the MLM model additionally adds random effects (kinship matrix) to correct for the cryptic relatedness. The Q-Q plots were used for selecting the best model of each trait. The Bonferroni-corrected *p*-value was used for mining the trait-related genome regions, and the markers that *p*-value of 0.05/13,125 (the total number of SNPs) or less were defined as significant. Based on the loci of MTAs, we used the online software - genome browser of peanut (https://www.peanutbase.org/gbrowse_peanut1.0) to screen the trait-related candidate genes among the trait-related regions.

RESULTS

Phenotypic Analysis of Peanut Varieties

Based on 3 years' evaluation data, we observed a large phenotypic variability among 178 peanut cultivars for all the traits studied. Phenotype identification and statistical analysis showed that the 178 peanut genotypes displayed wide ranges of phenotypic variation for most of the agronomic traits. For example, the length of the lateral branch, the height of the main stem, and the angle of the lateral branch are important factors for determining the peanut plant type, which varied from 44.2 to 106.6 cm, 82.8–34.2 cm, and 30–90°, respectively (**Table 1**). Besides, many traits related to yield and quality as well as significant variation included the weight and the length of seed and the number of branches with the pod, and the latter varied from average 2.6 to 12 (**Table 1**). Moreover, the testa color also had high variation in these peanut varieties, and the seeds with pink, red, black, white, and variegation are included. In addition, many quality traits are also varied, including the content of oleic acid and linoleic acid (**Table 1**). Most of these traits accord with normal distributions indicating these traits could be quantitative traits (**Figure 1**).

Genome-Wide Distribution of SNP Markers

The *Axiom_Arachis2* 48 K SNP array was used for genotyping the 178 peanut varieties (Nabi et al., 2021). A total of 34,712 SNPs

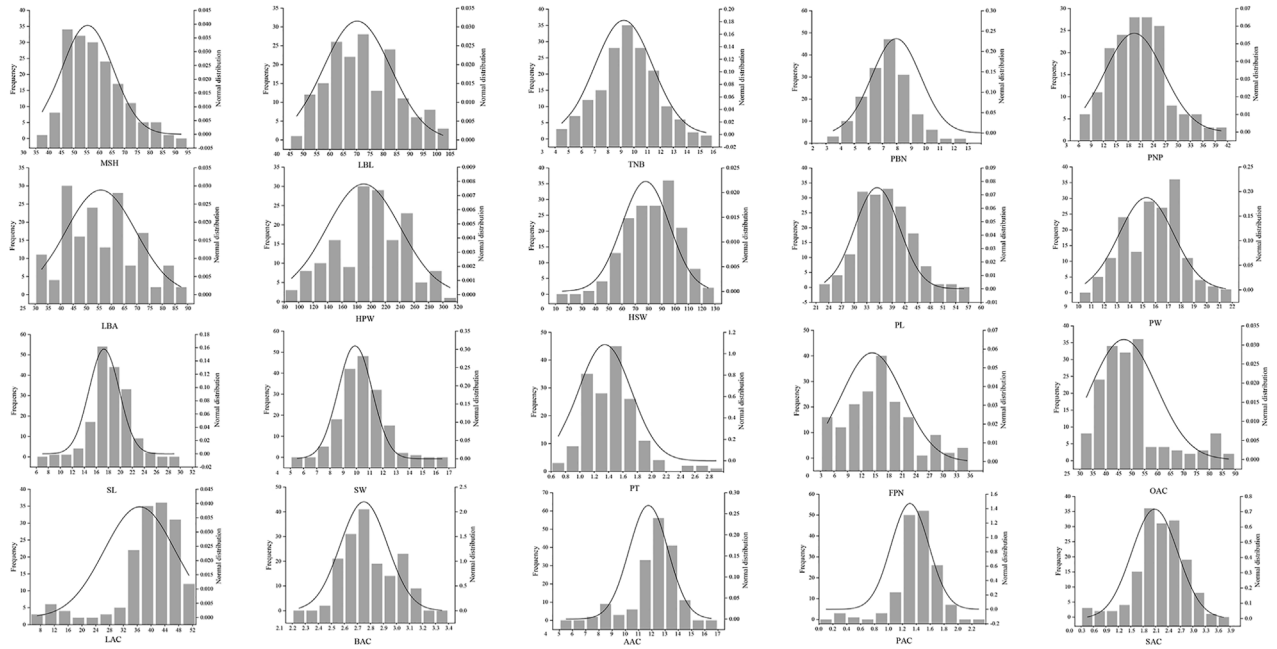


FIGURE 1 | Frequency distribution of 178 peanut cultivars for 20 traits. MSH, main stem height; LBL, lateral branch length; TNB, total number of branches; PBN, pod-bearing branches number; PNP, pod number per plant; LBA, lateral branch angle; HPW, hundred pod weight; HSW, hundred seed weight; PL, pod length; PW, pod width; SL, seed length; SW, seed width; PT, peel thickness; FPN, filled pods number; OAC, oleic acid content; LAC, linoleic acid content; BAC, behenic acid content; AAC, arachidic acid content; PAC, palmitic acid content; SAC, stearic acid content.

TABLE 2 | Distribution and density of SNPs on 20 chromosomes and scaffolds of peanut.

Chromosome	Length of chromosome (Mb)	Number of SNPs	Density of SNPs
Arahy.01	112.42	797	7.09
Arahy.02	102.98	460	4.47
Arahy.03	143.81	684	4.76
Arahy.04	128.8	532	4.13
Arahy.05	115.93	503	4.34
Arahy.06	115.5	681	5.90
Arahy.07	81.12	637	7.84
Arahy.08	51.9	340	6.55
Arahy.09	120.52	546	4.53
Arahy.10	117.09	524	4.48
Arahy.11	149.3	569	3.81
Arahy.12	120.58	451	3.74
Arahy.13	146.73	628	4.28
Arahy.14	143.24	800	5.59
Arahy.15	160.88	758	4.71
Arahy.16	154.81	677	4.37
Arahy.17	134.92	572	4.24
Arahy.18	135.15	645	4.77
Arahy.19	158.63	788	4.97
Arahy.20	143.98	768	5.33
Scaffold		765	
Total	2538.29	13125	4.69

were excluded based on filtering criterion: (1) SNPs with missing data rate (>10%) and (2) minor allele frequency (<5%). After filtration, 13,125 (27.43%) high quality SNPs were obtained (**Supplementary File S2**). On an average 4.69 SNPs/Mb were found distributed on 20 peanut chromosomes (Arahy.01 to Arahy.20) ranging from 3.74 SNPs/Mb to 7.84 SNPs/Mb (**Table 2**; **Figure 2A**). The maximum number of SNPs (800) were found on chromosome Arahy.14, followed by Arahy.01

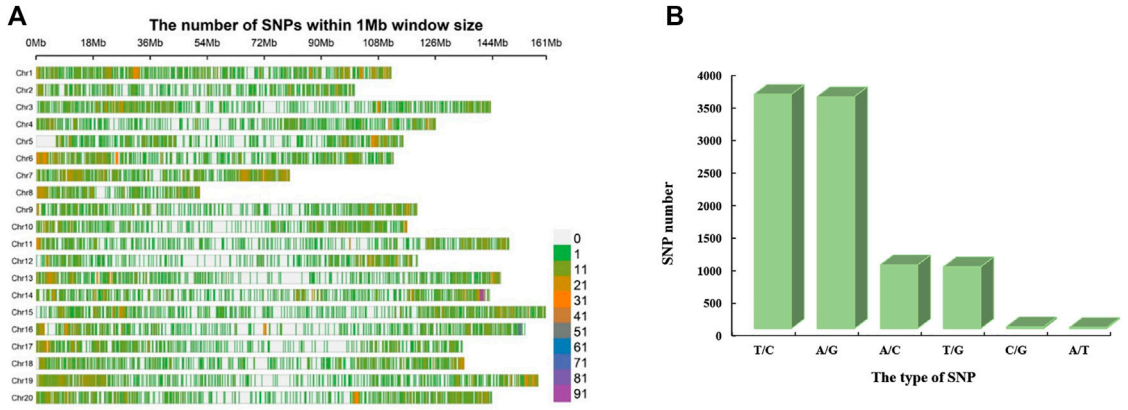


FIGURE 2 | Distribution and types of SNPs. **(A)** Distribution and density of SNPs in 20 peanut chromosomes. The horizontal axis shows the length of the chromosome (Mb), and the vertical axis represents 20 chromosomes. The shades of assorted color represent the SNP density on corresponding loci. **(B)** Frequency of several types of SNPs.

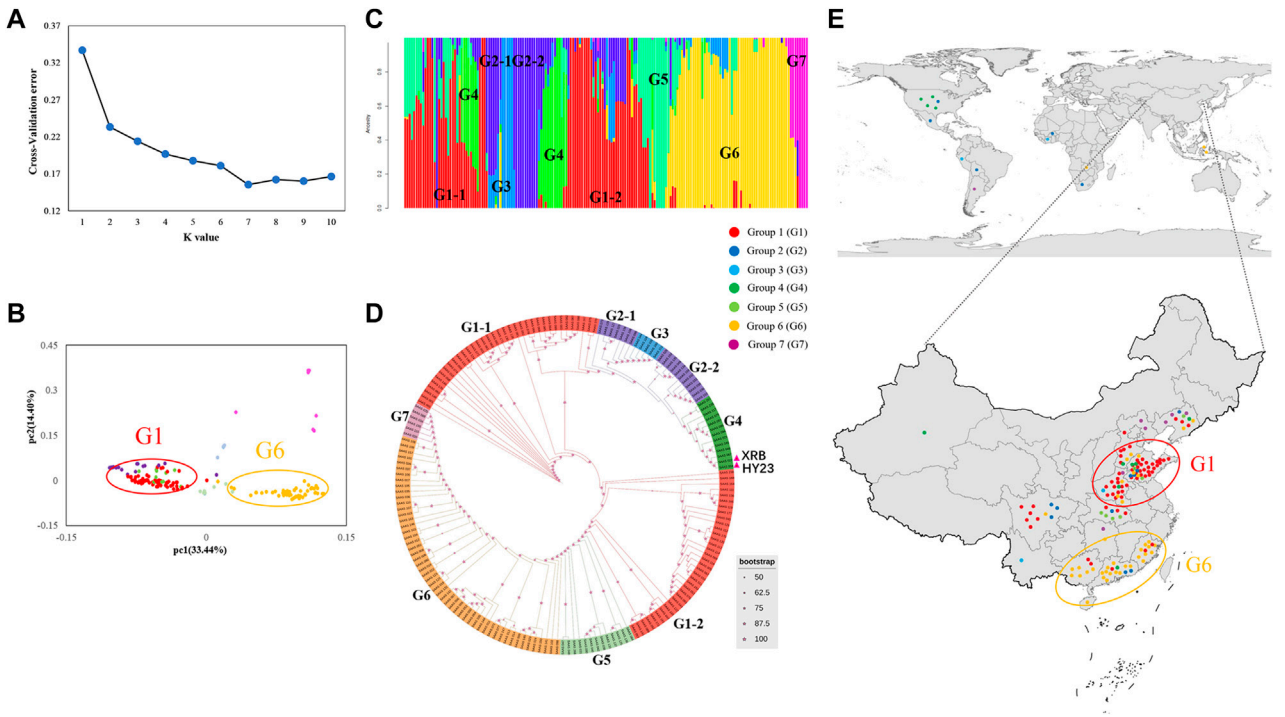


FIGURE 3 | Population structure and genetic diversity of the 178 peanut varieties. **(A)** Cross-validation value of each K ranging from 1 to 10. **(B)** The PCA analysis of the total accessions. Each dot represents one variety. **(C)** Population structure. Each variety is represented with a single vertical line, and the color represents ancestry. **(D)** Phylogenetic trees constructed by the maximum likelihood method. **(E)** Geographical distribution of total varieties.

(797) and Arahy.19 (788). As the smallest chromosome of the peanut genome, Arahy.08 contains only 340 SNPs. The density on chromosome Arahy.07 was the highest density (7.84 SNPs/Mb), while that on chromosome Arahy.12 was the lowest density (3.74 SNPs/Mb) (**Table 2**).

A total of six types of SNPs were observed including "T/C", "A/G", "A/C", "T/G", "C/G", and "A/T". We found that "T/C" is the most abundant type of SNPs, accounting for 39.43% of the total SNPs, followed by "A/G" which accounted for 39.05% of the total SNPs (**Figure 2B**). The "A/C" and "T/G" account for 10.71 and 10.37% of the total SNPs, respectively (**Figure 2B**). The "C/G" accounts for 0.29% of the total SNPs. The "A/G" is the least type of the SNPs, accounting for only 0.15% of the total SNPs (**Figure 2B**).

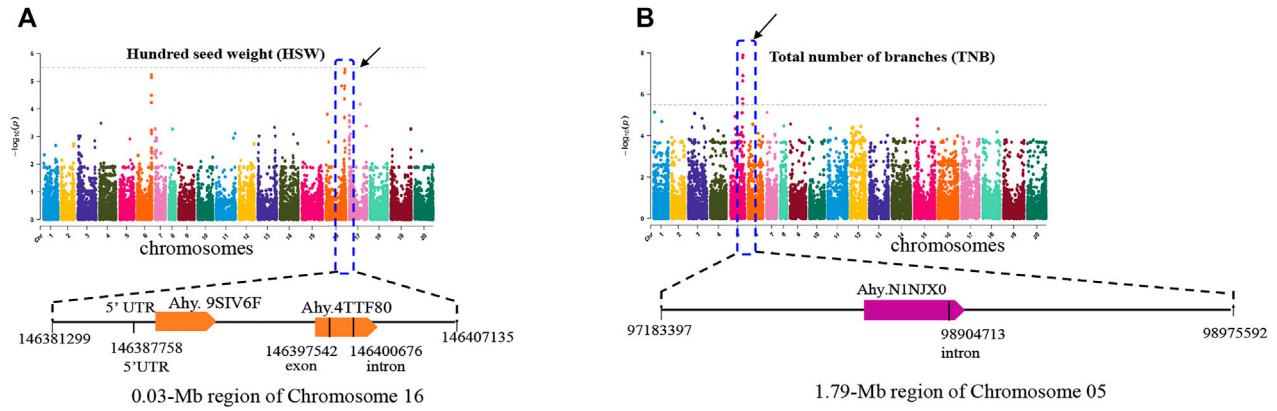


FIGURE 4 | GWAS signals for hundred seed weight **(A)** and total number of branches **(B)** of peanut. The significance level is log10 (0.05/13125) = 5.4 (the gray horizontal line). The characteristic analysis of functional genes in the screening intervals is shown below each Manhattan plot.

Population Structure Analysis

The population structure of the panel of peanut varieties was first investigated with the assessment of K value (**Figure 3A**), followed by validation via PCA (**Figure 3B**). The magnitude of CV error suggested that the best K (number of groups) was seven in the model-based group analysis. Based on their genotypes, the peanut panel could be divided into seven groups, group 1 (G1) to group 7 (G7), and the number of peanut varieties per group ranged from 5 to 72. The G1 containing 72 peanut varieties is the biggest group, followed by G6 which contains 53 peanut varieties (**Figures 3C,D**). Furthermore, the population structure and phylogenetic analysis results also suggested the presence of two subgroups of G1 (G1-1) and G2 (G2-1) (**Figures 3C,D**). The groups exhibited geographic distribution patterns, and the peanut varieties derived from the same planted areas of origin were usually in the same group. Most of the varieties originating from northern provinces including Shandong, Henan, and Hebei Provinces belong to G1, while the varieties originating from the southern provinces (Fujian, Guangdong, and Guangxi Provinces) were grouped into G6 (**Figure 3E**).

Genome-Wide Association Study

Based on the Q-Q plot analysis, GLM was selected as the best model for GWAS signals among five traits: total number of branches, oil patch, peel thickness, main stem height and testa color; while the MLM was used for the other two traits: hundred seed weight (HSW) and pod shape (**Supplementary Figure S1**).

Hundred seed weight and the total number of branches are important agronomic traits related to peanut yield. Under the threshold of -log10 > 5.4, MTA related to HSW was detected on chromosome 16 (**Figure 4A**). In addition, the associated SNPs were identified. The *SNP_Chr16:146387758* is located on the 5′UTR of the gene *Ahy. 9SIV6F* which encodes an unknown function protein. *SNP_Chr16:146400676* and *SNP_Chr16: 146397542* were all located in the gene region of *Ahy.4TTF80*, and the latter was in the exon of this gene. Function annotation showed that *Ahy.4TTF80* encodes an ABC-2 type transporter (**Figure 4A**). A previous study has shown that the ABC-2 type transporter protein was related with increasing size of plant seed and content of fat stored within the seed (Kim et al., 2013). For total number of branches, 1 MAT was detected in a 1.79-Mb region of chromosome 5 (97,904,713 to 98,975,592 bp), and 9 associated SNPs were enriched (**Figure 4B**). Among them, *SNP_Chr05:98904713* is in the intron of *Ahy.N1NJX0*, which is annotated as a calmodulin-binding transcription activator 2-like isoform X1. The other 8 SNPs were all located on intergenic regions (**Figure 4B**).

The oil patch is the spot in the seed coat of the peanut (**Figure 5A**). The presence of an oil patch will affect peanut quality; however, the genetic and molecular mechanism of an oil patch are unclear. GWAS analysis showed that 6 SNPs were associated with the oil patch, and all of them were in a 3.88-Mb region of chromosome 5 ranging from chr.05:111.93–115.81 (**Figure 5B**). *SNP_Chr05:111936057* is on the exon of *Arahy.7X9WBQ*, which encodes a peroxidase superfamily protein. Pod shape, peel thickness, and testa color are also the important appearance traits of peanut. Pod shape is one of the important characteristics for the classification of peanut. A total of 14 MTAs were detected for pod shape, and distributed in Chr2, Chr3, Chr5, Chr8, Chr10, Chr12, Chr13, Chr14, Chr15, Chr16, Chr17, Chr18, and Chr20 (**Figure 6A**). Among them, the most significant association loci were detected on Chr8 and Chr18 (**Figure 6A**). For peel thickness, one MAT was detected on a 3.72 -Mb region of chromosome 2 (Chr2:86.18–89.45 Mb) (**Figure 6B**). For the main stem height, only one SNP was identified on chromosome 6 (**Figure 6C**).

The seed coat (testa) is an important trait of peanut which is not only as an important protective barrier for peanut seed against the pathogen, but also important for health nutrition such as anthocyanins and procyanidins. Testa color is also a complex trait which is controlled by at least 12 genes (Branch, 2011). In this study, the color of the 178 peanut genotypes displayed significant variations, including pink, red, black, purple, white, and variegation (**Figure 7A**). GWAS analysis showed that the associated SNPs were detected in most of the chromosomes except to Chr.01 and Chr.07 (**Figure 7B**). Among them, the *SNP_AX-177640068* of chr.10 is only 222 kb to the gene *AhTc1*, one of the key gene controlling black testa identified

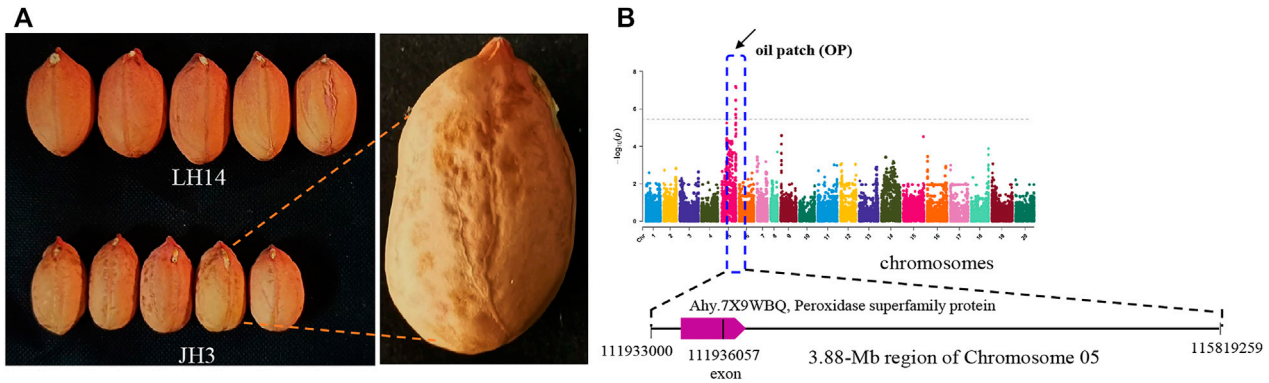


FIGURE 5 | GWAS signals for oil patch (spots) of peanut. **(A)** Peanut cultivars without an oil patch (LH14) and with an oil patch (JH3). **(B)** Manhattan plot. The characteristic analysis for one gene encoding peroxidase superfamily protein is shown below.

previously (Zhao et al., 2020). The *SNP_AX-176811136* in chr.03 is close to *AhRt1* locus contributing to red testa of peanut (Chen et al., 2021) (**Figure 7B**).

DISCUSSION

GWAS is a useful method for identifying linked loci and candidate genes by analyzing the association between the genotypes and the phenotypes of individuals in a population. Whole genome resequencing (WGS), high-density SNP array, and target genome resequencing (TGS) are the popular methods for acquiring the high throughput genotyping. WGS is with the maximum coverage across the genome, however, it is more expensive. TGS is a low-cost method that relies on the sequencing of target regions of the genome. Recently, several TGS methods have been developed and used in GWAS analysis, including genotyping by sequencing (GBS), restriction-site associated DNA sequencing (RAD-seq) and specific-locus amplified fragment sequencing (SLAF) (Zhang S. et al., 2019; Dodia et al., 2019; Wang et al., 2019; Yu et al., 2020; Jadhav et al., 2021). The SNP array is a low-cost and wide coverage method. The commercial high-density SNP array chips have developed in many crops, such as 660K of wheat (Hassan et al., 2021), 90K of rice (Yang et al., 2020), and 55K of maize (Li et al., 2021). In peanut, the available genome resource was used for identifying a large number of genome-wide SNPs, and large-scale 58 K SNP array (*Axiom_Arachis*) (Clevenger et al., 2017; Pandey et al., 2017) and 48K SNP array (*Axiom_Arachis2*) (Clevenger et al., 2018) have been developed. *Axiom_Arachis2* containing 47,837 SNPs is the second generation of peanut gene chip which has been successfully used for genetic diversity analysis and identification of QTLs related with the nutrient elements and leaf spots resistance of peanut (Zhang H. et al., 2019; Zhang et al., 2020a). However, the *Axiom_Arachis2* was developed prior to the release of a cultivated peanut genome. The positions of these SNPs were according to the genomes of wild type diploid peanut species, *A. duranensis* and *A. ipaensis*. In this study, we used *Axiom_Arachis2* to genotype 178 cultivars of peanut, and all of them are tetraploid cultivated peanut. Thus, we first remapped the probe sequences with the genome of cultivated peanut *Tifrunner* and updated the position information of the SNPs. In total, 45,608 SNPs were mapped in the 20 chromosomes of the peanut genomes and 2229 SNPs were mapped in the scaffolds. In the past few years, the SNP array has shown great potential for mapping the traits on peanut, and the updated position information of these SNPs will provide important references for future utility of the peanut SNP array (Pandey et al., 2021).

As an important index to evaluate seed size, HSW has been one of the hotspots in peanut genetics and QTL mapping. In this study, the HSW of the 178 peanut genotypes displayed wide ranges of variation, ranging from 29.5 to 113.0 g. Our results also showed that HSWs displayed variation within different groups. The G1 represented the varieties from Shandong, Henan, and Hebei Provinces. The average HSW of G1 is 85.1 g, which is significantly heavier than that of G6 (average HSW 66.0 g), in which most of the varieties come from southern provinces of China including Fujian, Guangdong, and Guangxi Provinces. Previous studies have reported many QTLs related with HSW, which is distributed in chr02, 03, 05, 07, 08, 12, 13, 14, 16, 17, and 18. Among them, QTLs in chr16 could be detected at least from four populations, including Fuchuandahuasheng × ICG 637, ZH16 × sd-H1, Zhonghua 16× J11, and Huayu 36 × 6–13, explaining up to 35.39% of the phenotypic variation (Huang et al., 2015; Wang et al., 2018; Mondal and Badigannavar, 2019; Zhang et al., 2019). In this study, the MTAs related with HSW were also detected in chromosome 16, and associated with the candidate gene *Ahy.4TTF80*, which encodes an ABC-2 type transporter. In tomato, the natural variation of the ABC transporter gene was associated with the seed size (Orsi and Tanksley, 2009). In rice, the ABC transporter gene, *OsABCG18* controls the shootward transport of cytokinin and is related with the grain yield of rice (Zhao et al., 2019). We found that there are two SNP substitutes in the *Ahy.4TTF80* gene, which provided an important clue for further fine mapped and revealed the key genes controlling the seed size of peanut. Besides, oil patch and pod shape are important agronomic traits of peanut. The oil patch and pod shape affect the appearance and commodity value. However,

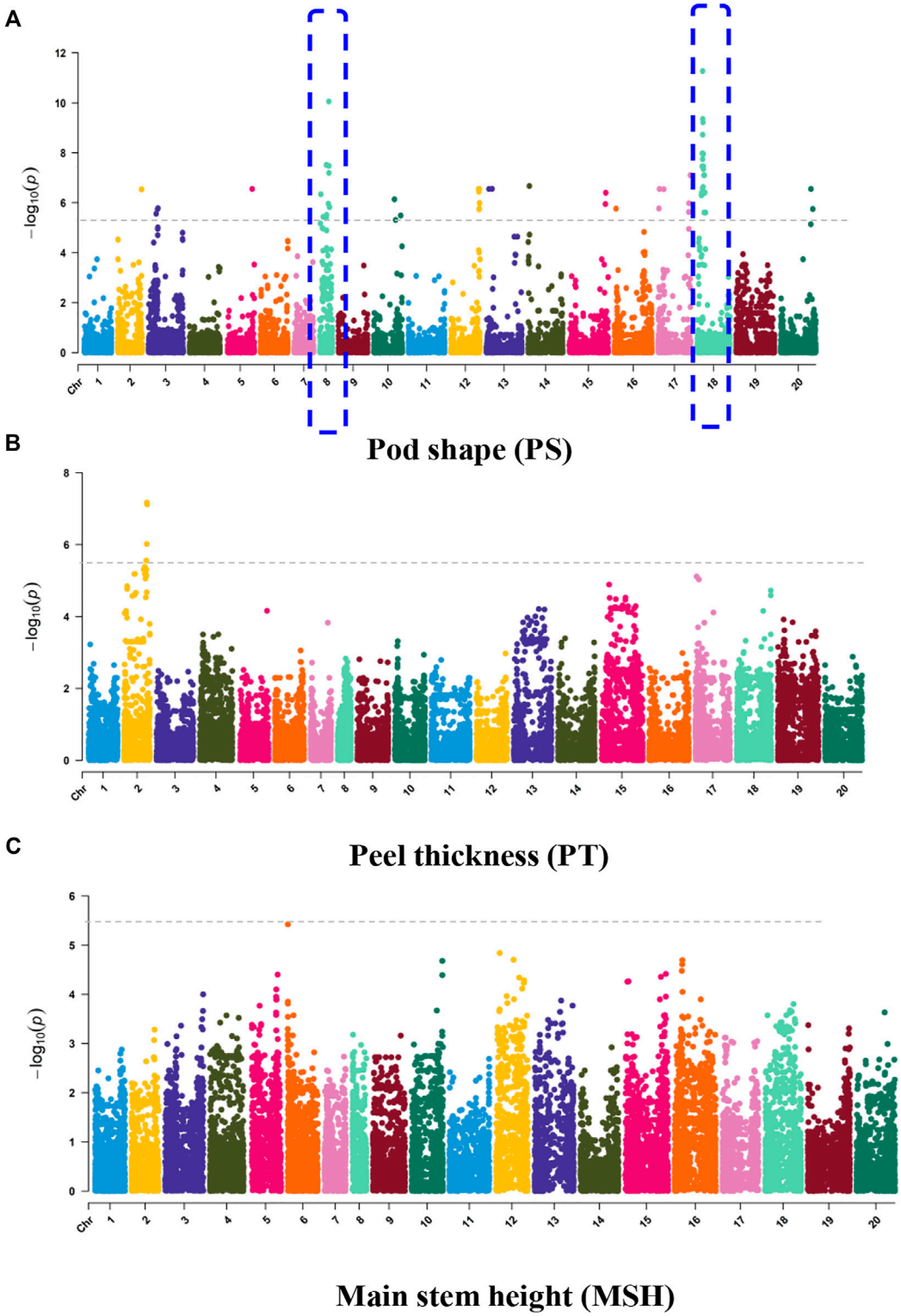


FIGURE 6 | GWAS signals for **(A)** pod shape, **(B)** peel thickness, and **(C)** main stem height, related to the appearance of peanut.

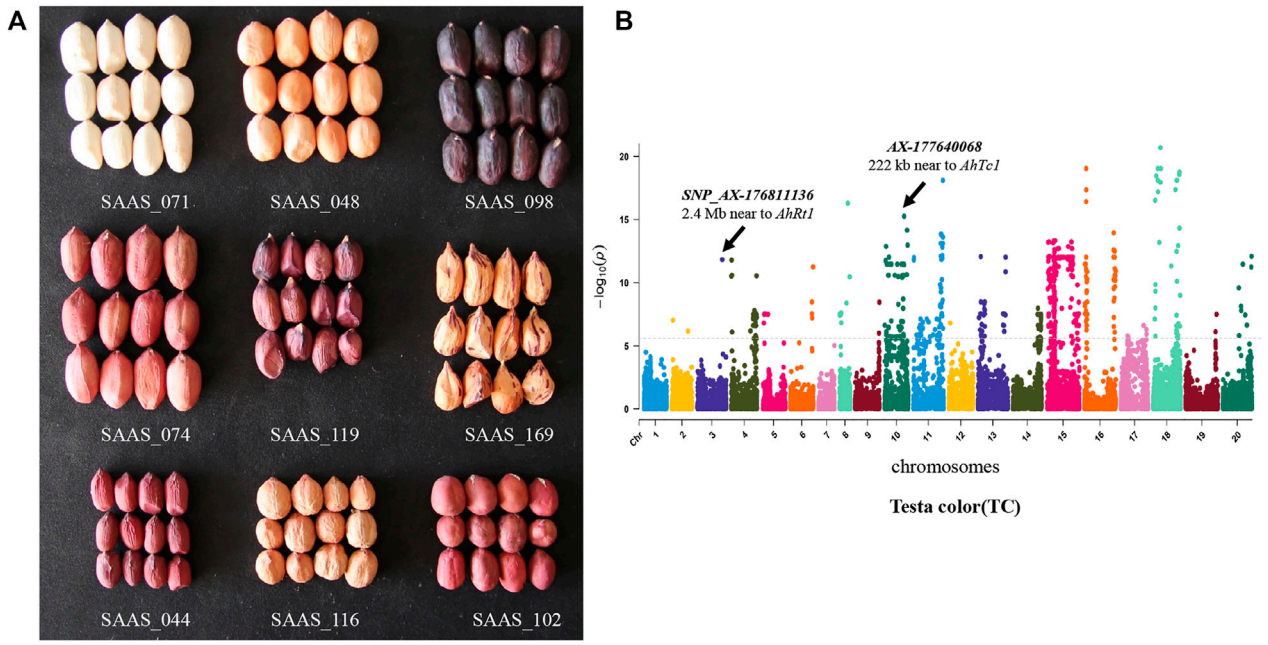


FIGURE 7 | Manhattan plots showing significant marker-trait associations for testa color of peanut. **(A)** Peanuts with different testa color, and its **(B)** GWAS signal. The characteristic analysis for one gene encoding lysosomal cystine transporter is shown below.

the physiology and genetics of them are rarely studied. In this study, we identified three MTAs related with the oil patch and pod shape. The details of those selected MTAs were shown in **Supplementary Figure S3**. These MTAs and candidate genes offer the opportunity to further study the molecular mechanism and improve these traits through the MAS approach.

Cultivated peanut is allotetraploid (AABB, 2n = 4 × = 40), derived from a hybridization event between *A. duranensis* (AA, 2n = 2 × = 20) and *A. ipaensis* (BB, 2n = 2 × = 20) about 3500 years ago (Kochert et al., 1996; Lavia et al., 2011). The molecular marker analysis demonstrated that cultivated peanut possesses a narrow genetic base (Halward et al., 1992), and some elite germplasm lines were overused in the peanut breeding program. For example, a previous study showed that more than 70% of peanut cultivars in China derived from two germplasms, Fuhuasheng and Shitouqi, directly or indirectly (Liao, 2004). In this study, a pedigree survey of these 120 peanut germplasms from 13 provinces and regions in China showed that 83 were from Fuhuasheng, accounting for 69.17%, and 42 were from Shitouqi, accounting for 35.00%. Thus, analysis of the genetic relationship between different germplasm resources is especially important for further designing hybrid combinations. In this study, we constructed the phylogenetic tree of these peanut germplasms through the 13,125 polymorphic SNPs. The results suggested that the geographical distribution is not exactly consistent with the genetic relationship among Chinese indigenous peanut breeds, which might be due to the exchange of germplasm resources across China (**Figure 3**). For example, Yueyou551 (SAAS_015) is classified into G1, however, it is a cultivar from the southern region of China. Pedigree analysis showed that the Yueyou551 derived from the cross-combination of Yueyou 22 and Yueyou 431, and the latter derived from the cross-combination of Shitouqi and Fuhuasheng. Fuhuasheng is a very typical elite peanut germplasm in the north of China. Besides, another two cultivars from the south, Tianfu3 (SAAS_128) and Guihua36 (SAAS_99), also have a close relationship with Fuhuasheng and is classified into G1. In addition, we found two peanut cultivars from Indonesia and one germplasm from Zambia (PI268586) are closely related to peanut varieties in south China, and classified into G6, which might be due to the exchange of germplasms between China and other countries. These results provide an important reference for further use of these germplasms.

CONCLUSION

In this study, we analyzed 20 phenotypic traits of 178 peanut germplasms and genotyped them using the 48 K *Axiom_Arachis2* SNP array. We analyzed the genetic diversity of these cultivars and identified a number of MTAs related to different traits. The candidate SNPs and candidate genes for these MTAs are helpful for further fine mapping and improving the quality and yield of peanut via a molecular breeding method.

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: NCBI's GEO database, accession number GSE197103.

AUTHOR CONTRIBUTIONS

XW, CZ, RT, and CM conceived and designed the experiments. HZ, XZ, JM, HH, and KZ performed the experiments. CZ, RT, HX, and HZ analyzed data. CZ, RT, HX, HZ, and MT wrote and revised the manuscript. All authors read and approved the manuscript.

FUNDING

This research is supported by Key Research and Development Project of Shandong Province (2020LZGC001), Shandong Province Natural Science Foundation (ZR2020MC105), National Natural Science Foundation of China (32072090, 31861143009, and 31801301), agricultural scientific and technological innovation project of Shandong Academy of Agricultural Sciences (CXGC2021B05, CXGC2021B24, and CXGC2021A30), High-level Foreign Experts Introduction Program (GL20200123001), and Taishan Scholar Project of Shandong Province (ts20190964).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.845602/full#supplementary-material

REFERENCES

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., et al. (2019). The Genome Sequence of Segmental Allotetraploid Peanut *Arachis H. Nat. Genet.* 51, 877–884. doi:10.1038/s41588-019-0405-z

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

Branch, W. D. (2011). First 100 Years - Inheritance of Testa Color in Peanut (Arachis HypogaeaL.). *Crop Sci.* 51, 1–4. doi:10.2135/cropsci2010.06.0349

Chen, H., Chen, X., Xu, R., Liu, W., Liu, N., Huang, L., et al. (2021). Fine-Mapping and Gene Candidate Analysis for *AhRt1*, a Major Dominant Locus Responsible for Testa Color in Cultivated Peanut. *Theor. Appl. Genet.* 134, 3721–3730. doi:10.1007/s00122-021-03924-w

Chu, Y., Wu, C. L., Holbrook, C. C., Tillman, B. L., Person, G., and Ozias-Akins, P. (2011). Marker-Assisted Selection to Pyramid Nematode Resistance and the High Oleic Trait in Peanut. *The plant genome* 4, 110–117. doi:10.3835/plantgenome2011.01.0001

Clevenger, J., Chu, Y., Chavarro, C., Agarwal, G., Bertioli, D. J., Leal-Bertioli, S. C. M., et al. (2017). Genome-Wide SNP Genotyping Resolves Signatures of Selection and Tetrasomic Recombination in Peanut. *Mol. Plant* 10, 309–322. doi:10.1016/j.molp.2016.11.015

Clevenger, J. P., Korani, W., Ozias-Akins, P., and Jackson, S. (2018). Haplotype-Based Genotyping in Polyploids. *Front. Plant Sci.* 9, 564. doi:10.3389/fpls.2018.00564

Collard, B. C. Y., and Mackill, D. J. (2007). Marker-Assisted Selection: An Approach for Precision Plant Breeding in the Twenty-First century. *Phil. Trans. R. Soc. B* 363, 557–572. doi:10.1098/rstb.2007.2170

Dodia, S. M., Joshi, B., Gangurde, S. S., Thirumalaisamy, P. P., Mishra, G. P., Narandrakumar, D., et al. (2019). Genotyping-by-Sequencing Based Genetic Mapping Reveals Large Number of Epistatic Interactions for Stem Rot Resistance in Groundnut. *Theor. Appl. Genet.* 132, 1001–1016. doi:10.1007/s00122-018-3255-7

Feng, H. S. (1987). *Chinese Peanut Varieties*. Beijing: Agriculture press.

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., et al. (2020). Nested-Association Mapping (NAM)-Based Genetic Dissection Uncovers Candidate Genes for Seed and Pod Weights in Peanut (*Arachis Hypogaea*). *Plant Biotechnol. J.* 18, 1457–1471. doi:10.1111/pbi.13311

Halward, T., Stalker, T., LaRue, E., and Kochert, G. (1992). Use of Single-Primer DNA Amplifications in Genetic Studies of Peanut (*Arachis hypogaea* L.). *Plant Mol. Biol.* 18, 315–325. doi:10.1007/BF00034958

Han, S., Yuan, M., Clevenger, J. P., Li, C., Hagan, A., Zhang, X., et al. (2018). A SNP-Based Linkage Map Revealed QTLs for Resistance to Early and Late Leaf Spot Diseases in Peanut (*Arachis Hypogaea* L.). *Front. Plant Sci.* 9, 1012. doi:10.3389/fpls.2018.01012

Hassan, M. A., Yang, M., Rasheed, A., Tian, X., Reynolds, M., Xia, X., et al. (2021). Quantifying Senescence in Bread Wheat Using Multispectral Imaging from an Unmanned Aerial Vehicle and QTL Mapping. *Plant Physiol.* 187, 2623–2636. doi:10.1093/plphys/kiab431

Huang, L., He, H., Chen, W., Ren, X., Chen, Y., Zhou, X., et al. (2015). Quantitative Trait Locus Analysis of Agronomic and Quality-Related Traits in Cultivated Peanut (*Arachis H* L.). *Theor. Appl. Genet.* 128, 1103–1115. doi:10.1007/s00122-015-2493-1

Jadhav, M. P., Gangurde, S. S., Hake, A. A., Yadawad, A., Mahadevaiah, S. S., Pattanashetti, S. K., et al. (2021). Genotyping-by-Sequencing Based Genetic Mapping Identified Major and Consistent Genomic Regions for Productivity and Quality Traits in Peanut. *Front. Plant Sci.* 12, 668020. doi:10.3389/fpls.2021.668020

Kim, S., Lee, Y., Nishida, I., Yamaoka, Y., and Ono, H. (2013). *Composition Containing Gene Encoding Abc Transporter Proteins for Increasing Size of Plant Seed and Content of Fat Stored within Seed*. US, US20130283481 A1.

Kochert, G., Stalker, H. T., Gimenes, M., Galgaro, L., Lopes, C. R., and Moore, K. (1996). RFLP and Cytogenetic Evidence on the Origin and Evolution of Allotetraploid Domesticated Peanut, *Arachis Hypogaea* (Leguminosae). *Am. J. Bot.* 83, 1282–1291. doi:10.1002/j.1537-2197.1996.tb13912.x

Lavia, G. I., Ortiz, A. M., Robledo, G., Fernández, A., and Seijo, G. (2011). Origin of Triploid *Arachis Pintoi* (Leguminosae) by Autopolyploidy Evidenced by FISH and Meiotic Behaviour. *Ann. Bot.* 108, 103–111. doi:10.1093/aob/mcr108

Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) V5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* 49, W293–W296. doi:10.1093/nar/gkab301

Li, S., Zhang, C., Yang, D., Lu, M., Qian, Y., Jin, F., et al. (2021). Detection of QTNs for Kernel Moisture Concentration and Kernel Dehydration Rate before Physiological Maturity in maize Using Multi-Locus GWAS. *Sci. Rep.* 11, 1764. doi:10.1038/s41598-020-80391-1

Liao, B. (2004). Present Situation and prospect of Peanut Genetic Improvement in China. *J. Agr. Sci. Tech-Iran.* 6, 60–62.

Luo, H., Pandey, M. K., Khan, A. W., Guo, J., Wu, B., Cai, Y., et al. (2018). Discovery of Genomic Regions and Candidate Genes Controlling Shelling Percentage Using QTL-seq Approach in Cultivated Peanut (*Arachis Hypogaea* L.). *Plant Biotechnol. J.* 17, 1248–1260. doi:10.1111/pbi.13050

Ma, J., Zhao, Y., Chen, H., Fu, C., Zhu, L., Zhou, X., et al. (2020). Genome-Wide Development of Polymorphic Microsatellite Markers and Their Application in Peanut Breeding Program. *Electron. J. Biotechnol.* 44, 25–32. doi:10.1016/j.ejbt.2020.01.004

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 2461. doi:10.1093/molbev/msaa131

Mondal, S., and Badigannavar, A. M. (2019). Identification of Major Consensus QTLs for Seed Size and Minor QTLs for Pod Traits in Cultivated Groundnut (*Arachis Hypogaea* L.). *3 Biotech.* 9, 347. doi:10.1007/s13205-019-1881-7

Nabi, R. B. S., Cho, K.-S., Tayade, R., Oh, K. W., Lee, M. H., Kim, J. I., et al. (2021). Genetic Diversity Analysis of Korean Peanut Germplasm Using 48 K SNPs '*Axiom_Arachis*' Array and its Application for Cultivar Differentiation. *Sci. Rep.* 11, 16630. doi:10.1038/s41598-021-96074-4

Orsi, C. H., and Tanksley, S. D. (2009). Natural Variation in an ABC Transporter Gene Associated with Seed Size Evolution in Tomato Species. *Plos Genet.* 5, e1000347. doi:10.1371/journal.pgen.1000347

Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., et al. (2017). Development and Evaluation of a High Density Genotyping 'Axiom_Arachis' Array with 58 K SNPs for Accelerating Genetics and Breeding in Groundnut. *Sci. Rep.* 7, 40577. doi:10.1038/srep40577

Pandey, M. K., Gangurde, S. S., Sharma, V., Pattanashetti, S. K., Naidu, G. K., Faye, I., et al. (2021). Improved Genetic Map Identified Major QTLs for Drought Tolerance- and Iron Deficiency Tolerance-Related Traits in Groundnut. *Genes* 12 (1), 37. doi:10.3390/genes12010037

Pandey, M. K., Pandey, A. K., Kumar, R., Nwosu, C. V., Guo, B., Wright, G. C., et al. (2020). Translational Genomics for Achieving Higher Genetic Gains in Groundnut. *Theor. Appl. Genet.* 133, 1679–1702. doi:10.1007/s00122-020-03592-2

Pinto da Silva, G. B., Zanella, C. M., Martinelli, J. A., Chaves, M. S., Hiebert, C. W., McCallum, B. D., et al. (2018). Quantitative Trait Loci Conferring Leaf Rust Resistance in Hexaploid Wheat. *Phytopathology* 108, 1344–1354. doi:10.1094/PHYTO-06-18-0208-RVW

Pujar, M., Gangaprasad, S., Govindaraj, M., Gangurde, S. S., Kanatti, A., and Kudapa, H. (2020). Genome-Wide Association Study Uncovers Genomic Regions Associated with Grain Iron, Zinc and Protein Content in Pearl Millet. *Sci. Rep.* 10, 19473. doi:10.1038/s41598-020-76230-y

Qureshi, N., Bariana, H., Kumran, V. V., Muruga, S., Forrest, K. L., Hayden, M. J., et al. (2018). A New Leaf Rust Resistance Gene Lr79 Mapped in Chromosome 3BL from the Durum Wheat Landrace Aus26582. *Theor. Appl. Genet.* 131, 1091–1098. doi:10.1007/s00122-018-3060-3

Shah, L., Rehman, S., Ali, A., Yahya, M., Riaz, M. W., Si, H., et al. (2018). Genes Responsible for Powdery Mildew Resistance and Improvement in Wheat Using Molecular Marker-Assisted Selection. *J. Plant Dis. Prot.* 125, 145–158. doi:10.1007/s41348-017-0132-6

Shasidhar, Y., Variath, M. T., Vishwakarma, M. K., Manohar, S. S., Gangurde, S. S., Sriswathi, M., et al. (2020). Improvement of Three Popular Indian Groundnut Varieties for Foliar Disease Resistance and High Oleic Acid Using SSR Markers and SNP Array in Marker-Assisted Backcrossing. *Crop J.* 8, 1–15. doi:10.1016/j.cj.2019.07.001

Varshney, R. K., Mohan, S. M., Gaur, P. M., Gangarao, N. V. P. R., Pandey, M. K., Bohra, A., et al. (2013). Achievements and Prospects of Genomics-Assisted Breeding in Three Legume Crops of the Semi-Arid Tropics. *Biotechnol. Adv.* 31, 1120–1134. doi:10.1016/j.biotechadv.2013.01.001

Wan, S. (2003). *Chinese Peanut Cultivation*. Shanghai: Shanghai Scientific and Technical Publishers.

Wang, J., Yan, C., Li, Y., Li, C., Zhao, X., Yuan, C., et al. (2019). GWAS Discovery of Candidate Genes for Yield-Related Traits in Peanut and Support from Earlier QTL Mapping Studies. *Genes* 10, 803. doi:10.3390/genes10100803

Wang, Z., Huai, D., Zhang, Z., Cheng, K., Kang, Y., Wan, L., et al. (2018). Development of a High-Density Genetic Map Based on Specific Length Amplified Fragment Sequencing and its Application in Quantitative Trait Loci Analysis for Yield-Related Traits in Cultivated Peanut. *Front. Plant Sci.* 9, 827. doi:10.3389/fpls.2018.00827

Yang, Y., Xu, X., Zhang, M., Xu, Q., Feng, Y., Yuan, X., et al. (2020). Genetic Basis Dissection for Eating and Cooking Qualities of *Japonica* rice in Northeast China. *Agronomy* 10, 423. doi:10.3390/agronomy10030423

Yu, B., Jiang, H., Pandey, M. K., Huang, L., Huai, D., Zhou, X., et al. (2020). Identification of Two Novel Peanut Genotypes Resistant to Aflatoxin Production and Their SNP Markers Associated with Resistance. *Toxins* 12, 156. doi:10.3390/toxins12030156

Yu, S. (2008). *Chinese Peanut Cultivar and Pedigree*. Shanghai: Shanghai Scientific Technology Press.

Zhang, H., Chu, Y., Dang, P., Tang, Y., Jiang, T., Clevenger, J. P., et al. (2020a). Identification of QTLs for Resistance to Leaf Spots in Cultivated Peanut (*Arachis Hypogaea* L.) through GWAS Analysis. *Theor. Appl. Genet.* 133, 2051–2061. doi:10.1007/s00122-020-03576-2

Zhang, H., Li Wang, M., Dang, P., Jiang, T., Zhao, S., Lamb, M., et al. (2020b). Identification of Potential QTLs and Genes Associated with Seed Composition Traits in Peanut (*Arachis Hypogaea* L.) Using GWAS and RNA-Seq Analysis. *Gene* 769, 145215. doi:10.1016/j.gene.2020.145215

Zhang, H., Wang, M. L., Schaefer, R., Dang, P., Jiang, T., and Chen, C. (2019). GWAS and Coexpression Network Reveal Ionomic Variation in Cultivated Peanut. *J. Agric. Food Chem.* 67, 12026–12036. doi:10.1021/acs.jafc.9b04939

Zhang, S., Hu, X., Miao, H., Chu, Y., Cui, F., Yang, W., et al. (2019). QTL Identification for Seed Weight and Size Based on a High-Density SLAF-Seq Genetic Map in Peanut (*Arachis Hypogaea* L.). *BMC Plant Biol.* 19, 537. doi:10.1186/s12870-019-2164-5

Zhao, C., Qiu, J., Agarwal, G., Wang, J., Ren, X., Xia, H., et al. (2017). Genome-Wide Discovery of Microsatellite Markers from Diploid Progenitor Species, *Arachis Duranensis* and *A. Ipaensis*, and Their Application in Cultivated Peanut (*A. Hypogaea*). *Front. Plant Sci.* 8, 1209. doi:10.3389/fpls.2017.01209

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, Y., Ma, J., Li, M., Deng, L., Li, G., Xia, H., et al. (2020). Whole-Genome Resequencing-Based QTL-Seq Identified *AhTc1* Gene Encoding a R2R3-MYB Transcription Factor Controlling Peanut Purple Testa Colour. *Plant Biotechnol. J.* 18, 96–105. doi:10.1111/pbi.13175

Zhou, X., Guo, J., Pandey, M. K., Varshney, R. K., Huang, L., Luo, H., et al. (2021). Dissection of the Genetic Basis of Yield-Related Traits in the Chinese Peanut Mini-Core Collection through Genome-Wide Association Studies. *Front. Plant Sci.* 12, 637284. doi:10.3389/fpls.2021.637284

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The Genome of Cultivated Peanut Provides Insight into Legume Karyotypes, Polyploid Evolution and Crop Domestication. *Nat. Genet.* 51, 865–876. doi:10.1038/s41588-019-0402-2

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.

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 © 2022 Zhao, Tian, Xia, Li, Li, Li, Zhang, Zhou, Ma, Huang, Zhang, Thudi, Ma, Wang and Zhao. 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 | Frontiers in Genetics

REVIEW
published: 07 April 2022
doi: 10.3389/fgene.2022.831656

Progress of Genomics-Driven Approaches for Sustaining Underutilized Legume Crops in the Post-Genomic Era

Uday Chand Jha[1], Harsh Nayyar[2], Swarup K Parida[3], Melike Bakır[4], Eric J. B. von Wettberg[5,6] and Kadambot H. M. Siddique[7]**

[1]ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, India, [2]Panjab University, Chandigarh, India, [3]National Institute of Plant Genome Research (NIPGR), New Delhi, India, [4]Department of Agricultural Biotechnology, Faculty of Agriculture, Erciyes University, Kayseri, Turkey, [5]Plant and Soil Science and Gund Institute for the Environment, The University of Vermont, Burlington, VT, United States, [6]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia, [7]The UWA Institute of Agriculture, The University of Western Australia, Perth, WA, Australia

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Leif Skot,
Aberystwyth University, United Kingdom
Sheikh Mansoor,
Sher-I-Kashmir Institute of Medical Sciences, India
Rafiul Amin Laskar,
Bahona College, India

***Correspondence:**
Uday Chand Jha
uday_gene@yahoo.co.in
Kadambot H. M. Siddique
kadambot.siddique@uwa.edu.au

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

Received: *08 December 2021*
Accepted: *24 February 2022*
Published: *07 April 2022*

Citation:
Jha UC, Nayyar H, Parida SK, Bakır M, von Wettberg EJ and Siddique KHM (2022) Progress of Genomics-Driven Approaches for Sustaining Underutilized Legume Crops in the Post-Genomic Era. Front. Genet. 13:831656. doi: 10.3389/fgene.2022.831656

Legume crops, belonging to the Fabaceae family, are of immense importance for sustaining global food security. Many legumes are profitable crops for smallholder farmers due to their unique ability to fix atmospheric nitrogen and their intrinsic ability to thrive on marginal land with minimum inputs and low cultivation costs. Recent progress in genomics shows promise for future genetic gains in major grain legumes. Still it remains limited in minor legumes/underutilized legumes, including adzuki bean, cluster bean, horse gram, lathyrus, red clover, urd bean, and winged bean. In the last decade, unprecedented progress in completing genome assemblies of various legume crops and resequencing efforts of large germplasm collections has helped to identify the underlying gene(s) for various traits of breeding importance for enhancing genetic gain and contributing to developing climate-resilient cultivars. This review discusses the progress of genomic resource development, including genome-wide molecular markers, key breakthroughs in genome sequencing, genetic linkage maps, and trait mapping for facilitating yield improvement in underutilized legumes. We focus on 1) the progress in genomic-assisted breeding, 2) the role of whole-genome resequencing, pangenomes for underpinning the novel genomic variants underlying trait gene(s), 3) how adaptive traits of wild underutilized legumes could be harnessed to develop climate-resilient cultivars, 4) the progress and status of functional genomics resources, deciphering the underlying trait candidate genes with putative function in underutilized legumes 5) and prospects of novel breeding technologies, such as speed breeding, genomic selection, and genome editing. We conclude the review by discussing the scope for genomic resources developed in underutilized legumes to enhance their production and play a critical role in achieving the "zero hunger" sustainable development goal by 2030 set by the United Nations.

Keywords: underutilized legumes, genomics, molecular marker, food security, transcriptomics

INTRODUCTION

Burgeoning pressure from the global human population, increasing food demands, and adverse effects of global climate change are serious concerns for global food and nutrition security (Godfray et al., 2010; Foley et al., 2011; Ebi and Loladze 2019). In addition, increasing outbreaks of plant diseases and pests, loss of arable land, and increasing environmental degradation due to excessive use of chemical fertilizers and pesticides have constrained crop yields (Godfray et al., 2010; Lesk et al., 2016). Of the various approaches for sustaining global food production without deteriorating soil and environmental health, crop diversification is needed to maintain sustainable agro-ecological systems and prevent biodiversity losses (Hufnagel J. et al., 2020; Tamburini et al., 2020). Legume crops remain the third most widely grown class of crops globally (Gepts et al., 2005), providing "one third of all dietary protein nitrogen" to the human population, enriching soil fertility by fixing atmospheric nitrogen in association with symbiotically active rhizobacteria in roots (Graham and Vance, 2003), and adding rotational value to subsequent crops (Yigezu et al., 2019; Marques et al., 2020). Likewise, legume fodder and forage mitigate the rising global demand for dietary protein by livestock and provide industrial raw materials (Das and Arora 1978; Elfaki and Abdelatti 2018). Most studies have focused on major grain legumes, such as soybean, common bean, and chickpea. However, some legume crops (**Supplementary Table S1**) with high nutrient contents are grown in limited areas on small scales in developing countries under low input conditions and marginal land (Cullis and Kunert 2017; Kamenya et al., 2021). Despite the enormous potential of these legumes, they are neglected and known as "underutilized" legumes (Cullis and Kunert 2017; Kamenya et al., 2021). Underutilized species are rarely grown outside of a narrow geographic area, are cultivated with low chemical inputs or mechanization, are not broadly used outside of traditional cuisines, and have not been the focus of major public and private breeding companies. In the last decade, major grain legume crops have witnessed unprecedented advances in genomic resource development, including the development of reference genome sequences due to rapid advances in genome sequencing technologies, especially, next-generation sequencing (NGS). However, underutilized legume crops are lagging behind in terms of developing genomic resources. Thus, in this review we analyze the present global status of these underutilized legumes in terms of area, production, major production and nutritional quality limitation and origin (**Supplementary Table S1**) and discuss the available genomic resources, including their molecular marker repertoire and genome assemblies. We review the progress in genetic linkage maps and identification of trait QTLs through bi-parental mapping and genome-wide association studies of various underutilized legumes, including the downstream application of genomic assisted breeding (GAB). The discovery of various trait candidate gene(s) with putative function through transcriptome sequencing are discussed with examples. We also brief how crop wild relatives (CWRs), whole-genome resequencing (WGRS), and pangenome sequences could underpin novel structural variants across the whole genome in these crops. Finally, we propose the prospects and scope of novel breeding schemes—genomic selection, genome editing, and speed breeding—for enhancing genetic gain to achieve "zero hunger" in 2030.

Why Genomics and Advanced Breeding Tools for Underutilized Legumes

Underutilized legumes generally require few inputs, are rich in protein, vitamins, and minerals, and can often withstand harsh environments, including drought, extreme temperature, and waterlogging. Furthermore, these legumes replenish soil nitrogen by fixing atmospheric nitrogen through root nodules, ameliorate soil properties, and sustain agro-ecosystem services (Bhartiya et al., 2015; Ditzler et al., 2021). In addition to their role in combating nutritional and economic security, underutilized legumes play critical roles in various human diseases as they are rich in bioactive compounds and nutraceutical and medicinal properties (Prasad and Singh 2015; Bazzano et al., 2001). However, despite these benefits, there are several constraints and challenges related to the production and productivity of these legumes due to biotic and abiotic stresses (**Supplementary Table S1**). Furthermore, the edible seeds of some underutilized legumes contain antinutritional elements, constraining their use (Campbell et al., 1994; Tate and Ennenking 2006; Kroc et al., 2017). Combining modern genomic and traditional breeding approaches could help develop new plant types, reduce yield losses from biotic and abiotic stresses, add value for consumer preferences, and eliminate antinutritional properties.

How Minor Legumes can Catch up With Genomics

One of the aspects of the advances in DNA sequencing technology over the past two decades has been the potential to democratize research. Before the advent of next generation sequencing, performing molecular genetic research outside of a handful of species, such as fruit flies and *Arabidopsis*, was cost-prohibitive. Exponential declines in the cost of sequencing have made research in nearly any species not only feasible, but practical. Consequently, crops like chickpeas, pigeonpea and cowpea, once considered minor crops, now have a rich array of genome resources (e.g., Jha 2018; Varshney et al., 2021). However, there are still a range of crop species that have received less attention, due to being grown over a limited geographic extent or market demand that is mostly restricted to a small region.

For those crop species that still trail behind others for genomic resources, there is hope that lessons learned in other species can be applied to others. In legumes, where there is substantial genome synteny across the entire family [e.g., (Ren et al., 2019)], the potential for comparative genomics to speed up research in understudied species is particularly high. With improving databases for mining genomic information from more widespread cultivated legumes [e.g., (Bauchet et al., 2019; Berendzen et al., 2021)], this task has become easier than in the past.

In a range of minor legume crops, one of the foci for improvement are "domestication syndrome" traits, such as

TABLE 1 | Genomic resources in underutilized legumes developed during the last decade.

Common name	Genome size	Mapping populations	SSRs/SNPs discovered
Adzuki bean	538 Mbp (Parida et al., 1990)	~6 (Yang et al., 2015a; Wang et al., 2021b)	7,947 EST-SSR (Chen et al., 2015a)
—	—	—	143,113 SSRs (Kang et al., 2015)
Bambara	550 Mb(Lonardi et al., 2019)	~2 (Ho et al. (2017)	1292 SSR (Chapman (2015); 3,343 SNP Uba et al. (2021)
Groundnut	—	—	—
Clusterbean	580.9 Mbp (Tyagi et al., 2019)	—	5,773 (Tanwar et al., 2017); 8,687 (Rawal et al., 2017)
—	—	—	18,792 (Thakur and Randhawa 2018); 25,280 (Kumar et al., 2020)
—	—	—	27,066 (Al-Qurainy et al., 2019); 1,859 genomic SSRs
—	—	—	(Tribhuvan et al., 2019); 3,594 SNPs (Tanwar et al., 2017)
—	—	—	5,999 SNPs and 249 InDels (Thakur and Randhawa 2018)
Common	1.8 Gb (Shirasawa et al., 2021b)	—	6,848 SSRs and 7,246 high quality SNPs (De la Rosa et al., 2020)
Vetch	—	—	—
Dolichos bean	—	~4 (Konduri et al., 2000; Yuan et al., 2011)	9,320 DArT seq based SNPs and 15,719 SilicoDart markers
or hyacinth bean	—	Ramtekey et al., 2019)	(Sserumaga et al., 2021)
—	—	—	2,529 SSRs (Chapman 2015)
Grasspea	8.2 Gb (Bennett and Leitch, 2012)	~2 (Santos et al., 2018)	651,827 SSRs and 288 SSRs (Yang et al., 2014);
—	—	—	3,204 EST-SSR (Hao et al., 2017); 146,406 SNPs (Hao et al., 2017)
Horse gram	400 Mb (Bhardwaj et al., 2013)	~1 (Shirasawa et al., 2021a)	6,195 SSRs(Bhardwaj et al., 2013)
—	—	—	3,942 SNPs (Mahesh et al., 2021)
Lima bean	~622 Mbp/1 C	~1 (Garcia et al., 2021)	10,497 SNPs(Garcia et al., 2021)
—	(Mercado-Ruaro, P. & Delgado-Salinas 1998)	—	—
Mothbean	—	~1 (Yundeang et al., 2019)	—
Mungbean	494–555 Mb (Liu et al., 2016b)	~19 (Wang et al., 2020)	13,134 EST-SSRs (Chen et al., 2015b)
—	—	—	and 200,808 SSRs in mungbean
—	—	—	(Kang et al., 2014); 775,831 high-confidence SNPs (Kang et al., 2014)
—	—	—	8,966 SNPs (Ha et al., 2021); 233,799 SNPs (Bangar et al., 2021)
Narrow-leafed lupin	924 Mbp (Kasprzak et al., 2006)	~9 (Zhou et al., 2018; Kozak et al., 2019)	4830 SNPs (Książkiewicz et al., 2017)
—	—	—	38,948 SNPs (Mousavi-Derazmahalleh et al., 2018)
—	—	—	1,723 058 SNPs (Wang P. et al., 2021)
Red clover	420 Mb(Sato et al., 2005)	~3 (Riday and Krohn 2010)	264,927 SNPs (Jones et al., 2020); 69,975 SNPs (Li et al., 2019)
—	—	—	6,749 SSR, 343,027 SNPs (Ištvánek et al., 2017)
Ricebean	414 Mbp (Kaul et al., 2019)	~2 (Somta et al., 2006; Isemura et al., 2010)	300 SSR (Chen et al., 2016); 261,458 SSRs (Wang et al., 2016)
White lupin	451 Mb (Hufnagel et al., 2020b)	1	2,659,837 SNPs (Hufnagel et al., 2020b)
—	—	—	3,527,872 SNPs (Hufnagel et al., 2021)
Yellow lupin	—	1 (Iqbal et al. (2020)	13,462 SNPs in yellow lupin (Iqbal et al., 2019)
—	—	—	3,942 SNPs (De Vega et al., 2015)
Urd bean	574 Mbp	3 (Somta et al., 2019)	166,014 SSRs (Jegadeesan et al., 2021)
—	Arumuganathan and Earle (1991)	—	1,621 genic-SSR and 1844 SNPs (Raizada and Souframanien 2019)
—	—	—	3,675 SNPs (Somta et al., 2019)
Winged bean	1.22 Gbp/C (Vatanparast et al., 2016)	—	1853 SSRs (Chapman 2015); 12,956 SSRs(Vatanparast et al., 2016)
—	—	—	5,190 SNPs (Vatanparast et al., 2016)
Zombi pea	—	2 (Dachapak et al., 2018; Amkul et al., 2019)	4,044,822 SNPs in zombipea (Amkul et al., 2019)

pod shattering, seed dormancy, seed size, and palatability. There is growing evidence that at least some of the loci controlling these traits are shared, such as for pod shattering (Ogutcen et al., 2018). With shared loci and extensive genomic synteny, either finding natural variation at these loci or using genetic modification become much easier.

Advances in Genomic Resource Development in Underutilized Legumes

In the last decade, rapid advances in genome sequencing technologies have enriched the genomic resources, including genome-wide distributed high-throughput molecular markers especially, simple sequence repeats (SSRs) and single

TABLE 2 | List of genome sequence assembly of underutilised legume crops.

Crop name	Genotype	Pubmed ID	Chromosome no.	Size of genome	No. of protein coding genes	Genome coverage	Sequencing platform used	References
Cyamopsis tetragonoloba (L.) Taub.	Vaviloskij 130	—	—	1.2 Gb	—	5×	Illumina and Oxford Nanopore	Grigoreva et al. (2019)
Cyamopsis tetragonoloba (L.) Taub.	RGC-936	—	—	550.31 Mbp	34680	366.73×	Illumina, 10× Chromium and Oxford Nanopore	Gaikwad et al. (2020)
L. purpureus	—	30535374	2n = 22	395.47 Mb	20,946	—	HiSeq 2000 platform (Illumina)	Chang et al. (2019)
Lupinus albus	AMIGA	31980615	2n = 50	451 Mb	38258	164×	PacBio Sequel platform	Bárbara Hufnagel et al. (2020)
Lupinus angustifolius	Tanjil	27557478	2n = 40	609 Mb	33 076	162.8×	Illumina	Hane et al. (2017)
Lupinus angustifolius	Tanjil	33249667	2n = 40	615.8 Mb	33907	156x	PacBio Sequel II platform	Wang P. et al. (2021)
Macrotyloma uniflorum	HPK-4	—	—	259.2 Mb	—	—	Illumina HiSeq 2000	Shirasawa et al. (2021a)
Macrotyloma uniflorum	PHG-9	—	2n = 40	279.1 Mb	24,521	—	Illumina HiSeq	Mahesh et al. (2021)
Medicago polymorpha	Huaiyang Jinhuacai	33642569	—	441.83 Mb	36,087	123.89×	Illumina, PacBio and Hi-C technologies	Cui et al. (2021)
M. ruthenica	—	—	—	904.13 Mb	50,162	—	PacBio, Illumina, 10×Genomics, and Hi-C	Wang T. et al. (2021)
M. ruthenica	—	33615703	—	903.56 Mb	50,268	—	Illumina, PacBio, and Hi-C	Mou Yin et al. (2021)
Narrow leafed lupin	Tanjil	23734219	—	538	57,807	27×	—	Yang et al. (2013a)
Narrow leafed lupin	Tanjil	—	—	521.2	—	25×	—	Kamphuis et al. (2015)
Phaseolus lunatus L.	G27455	33514713	—	512 Mbp	28,326	10×	Illumina HiSeq	Garcia et al. (2021)
Phaseolus acutifolis A. Gray	Frijol Bayo	—	2n = 22	684 Mb	27,538	101.28×	Illumina HiSeq platforms	Moghaddam et al. (2021)
Phaseolus acutifolis A. Gray	wild accession	—	2n = 22	676 Mb	27,095	—	Illumina HiSeq platforms	Moghaddam et al. (2021)
—	W6 15578	—	—	—	—	—	—	—
Red clover	Tatra	24500806	—	314.6	47,398	50×	Illumina HiSeq 2000	Isõtvánek et al. (2014)
Red clover	Milvus B	26617401	—	309 Mb	40,868	30×	Illumina HiSeq 2000	De Vega et al. (2015)
T. subterraneum L.	Daliak	27545089	—	471.8 Mb	42,706	—	Illumina MiSeq and HiSeq 2000	Hirakawa et al. (2016)
T. subterraneum L.	TSUd_r1.1	28111887	—	512 Mb	31,272	341×	Illumina HiSeq 2000	Kaur et al. (2017)
Vigna radiata	VC 1973A	25384727	2n = 22	579 Mb	22,427	—	Illumina HiSeq 2000, GS FLX +	Kang et al. (2014)
Vigna radiata	VC 1973A	34275211	2n = 22	475 Mb	30,958	—	PacBio RS II platform	Ha et al. (2021)
Vigna angularis	Jingnong 6	26460024	2n = 22	450 Mb	34,183	168×	HiSeq 2000	Kai Yang et al. (2015)
Vigna angularis var. *angularis*	IT213134	25626881	2n = 22	612 Mb	26,857	—	Illumina HiSeq 2000	Kang et al. (2015)
Vicia sativa	KSR5	—	2n = 14	1.5 Gb	31,146	146×	HiSeq2000	Shirasawa et al. (2021b)
Vigna mungo	Pant U-31	—	2n = 22	475 Mb	18655	—	Illumina and Nanopore sequencing	Jegadeesan et al. (2021)
Vigna mungo	Chai Nat 80	—	2n = 22	499 Mb	29,411	21.72×	Illumina HiSeq × Ten	Pootakham et al. (2021)
V. subterranea	—	30535374	—	535.05 Mb	31,707	—	HiSeq 2000 platform (Illumina)	Chang et al. (2019)
Vigna umbellata	VRB3	—	2n = 22	414 Mb	31276	30×	Illumina and PacBio platform	Kaul et al. (2019)

nucleotide polymorphisms (SNPs), transcriptomes, and whole-genome assemblies, of various underutilized legumes.

Molecular Marker Resources

Hybridization-based molecular markers, such as restriction fragment length polymorphisms (RFLP), and PCR-based molecular markers, such as RAPD, SSR markers, have been used to analyze, tag, and map trait gene(s) in various underutilized legumes (Bohra et al., 2014). However, the arrival of next-generation sequencing technology (NGS) based high-throughput (HTP) markers, especially SNPs, has replaced traditional PCR-based molecular markers for genotyping.

Second- and third-generation sequencing technologies have enabled the mining of massive numbers of SSRs and SNPs through whole-genome sequencing, WGRS, and transcriptome sequencing efforts in various crops, including underutilized legumes (Edwards and Batley 2010).

Likewise, the advent of NGS-based HTP genotyping platforms, such as Illumina's GoldenGate assay, Illumina's HiSeq 4000 platform, and Illumina's Infinium SNP array, enabled the discovery of copious SNPs across multiple genomes that facilitate a range of investigations, including the diversity of genebank collections (Sokolkova et al., 2020). Aiming at comprehensive mining of SSR markers for *Vigna* species including cowpea, mungbean and adzuki bean, microsatellite database *VigSatDB* has been developed (Jasrotia et al., 2019). A comprehensive list of molecular markers, mapping populations available in various underutilized legumes are in **Table 1**. Thus, these molecular markers will provide the foundation for implementing genomic assisted breeding for improving genetic gain in underutilized grain legumes.

De Novo Genome Sequencing of Underutilized Legumes

Adzuki bean (*Vigna angularis* var. *angularis*) (2n = 2x = 22) is an important grain legume of Asiatic origin (Kang et al., 2015). The draft genome sequence of adzuki bean was assembled on 11 pseudo-chromosomes, estimating 612 Mb or 75% of the estimated genome and high-confidence 26,857 protein-coding genes (Kang et al., 2015) (**Table 2**). Yang K. et al. (2015) assembled a draft genome assembly of "Jingnong 6" cultivar covering 450 Mb of the total genome.

Bambara groundnut (*Vigna subterranean*) (2n = 2x = 22) is an important legume crop, rich in protein (18–26%), carbohydrate (63%), and fat (6.5%) and having inherent drought tolerance capacity (Shegro et al., 2013). It originated from West Africa and is mainly grown in sub-Saharan areas, especially Nigeria (Olukolu et al., 2012). Chang et al. (2019) assembled the genome sequence of bambara groundnut, with a genome size of ~535.05 Mb with 31,707 protein-coding genes.

Mungbean (*Vigna radiata*, 2n = 2X = 22) is a warm-season legume crop, originated from India and mostly grown in South and Southeast Asian countries. Kang et al. (2014) first assembled the mungbean genome sequence, estimating 421 Mb or 80% of the total genome size and 22,427 protein-coding genes, with scaffold length 431 Mb and N50 length 35.4 Mb covering 314 Mb. Recently, a mungbean genome sequence was assembled with a total scaffold size of 475 Mb and N50 scaffold value of 5.2 Mb (Ha et al., 2021).

Urdbean (*Vigna mungo*, 2n = 2x = 22), native to Indian subcontinent, mostly grown in South and Southeast Asian countries (Kaewwongwal et al., 2015), is a rich source of dietary protein, vitamins, folate, and iron (Kakati et al., 2010). The genome assembly of Chai Nat 80 cultivar measured 499 Mb with an N50 length of 5.2 Mb (Pootakham et al., 2021). Subsequently, Jegadeesan et al. (2021) assembled a genome assembly of urdbean, measuring 475 Mb or 82% of the genome with scaffold N50 of 1.42 Mb and 42,115 genes with coding sequence.

Cluster bean (*Cyamopsis tetragonoloba*, 2n = 2x = 14), native to west Africa and India, an important commercial legume crop widely grown in India and parts of Africa, contains hetero-polysaccharide called guar gum or galactomannan used extensively in the cosmetic and pharmaceutical industries (Gillett 1958). Gaikwad et al. (2020) assembled the first genome sequence of RGC-936 cultivar, measuring 550.31 Mbp with N50 length of 78.27 Mbps and 34,680 protein-coding genes.

Dolichos bean (*Lablab purpureus*) (2n = 2x = 22) is a versatile legume crop of African origin, rich in seed protein and highly tolerant to various abiotic stresses (Maass et al., 2010). It is mostly cultivated in tropical and sub-tropical regions globally (Maass et al., 2010). The genome assembly of *Lablab purpureus* was constructed recently, with an estimated 395.47 Mb genome size and 20,946 protein-coding genes (Chang et al., 2019).

Grass pea (*Lathyrus sativus*) is a climate-resilient legume of Central Asia and Abyssinia origin, diploid (2n = 2x = 14), cool-season legume species (Kamphuis et al., 2015; Emmrich et al., 2020) primarily grown on the Indian subcontinent and in northern and eastern Africa, including Ethiopia (Kumar et al., 2011). The assembled genome size of EIv1 was measured at 8.12 Gbp with scaffold N50 value of 59.7 kbp and 33,819 high-confidence genes (Kamphuis et al., 2015; Emmrich et al., 2020).

Horsegram [*Macrotyloma uniflorum* (Lam.) Verdc.], native to tropical southern Asia, is a diploid legume (2n = 20, 22) grown in India, Africa, and Australia (Arora and Chandel 1972). The genome sequence of the HPK-4 genotype was assembled on ten pseudomolecules measuring 259.2 Mb or 89% of the total length of the assembled sequence (Shirasawa et al., 2021a). Another genome assembly of accession PHG-9, measuring 279.1 Mb with 24,521 annotated genes has recently been constructed (Mahesh et al., 2021).

Red clover (*Trifolium pratense* L.; Fabaceae, 2n = 2x = 14) is an important forage legume of European origin, with a genome size of 418 Mbp. Isõtvánek et al. (2014) completed a *de novo* assembly of the red clover genome, comprising ~314.6 Mbp.

Likewise, subterranean clover presumed to be originated from Southern Australia, belonging to *Trifolium* genus, is an annual diploid (2n = 2x = 16) pasture legume with 540 Mbps genome size (Kaur et al., 2017). Hirakawa et al. (2016) assembled the genome sequence of *T. subterraneum* L., measuring 471.8 Mb or 85.4% of the whole genome and containing 42,706 protein-coding genes. Subsequently, Kaur et al. (2017) assembled an advanced genome assembly of *T. subterraneum* L., estimating 512 Mb with 31,272 protein-coding genes.

Tepary bean (*Phaseolus acutifolius* A. Gray), native to the Sonoran Desert and a sister species of common bean, is gaining attention due to its inherent capacity for biotic and abiotic stress tolerance (Moghaddam et al., 2021) and important source traits for improving biotic and abiotic stress tolerance in common bean (Moghaddam et al., 2021). A reference genome assembly of cultivated landrace *Frijol Bayo*, possessing inherent heat tolerance, was constructed using Illumina X10 and HiSeq platforms and PACBIO with 101.28× sequence coverage, and

TABLE 3 | List of high density genetic maps developed in various underutilised legumes.

Crop	Mapping population	Type of population	Size of LG map	Number of marker/loci assigned	Marker density	References
Trifolium pratense L.	HR × R130, NS10 × HR, NS10×H17L	BC_1F_1	836.6 cM	1804 loci	0.46 cM	Isobe et al. (2009)
—	H17L × R130, 272 × WF1680	—	—	—	—	—
—	pC × pV	—	—	—	—	—
Vigna mungo	BC48 × TC2210	—	1,588.7 cM	3,675 SNPs	0.57 cM	Somta et al. (2019)
Vigna radiata	*Vigna radiata* × *V. umbellata*	RIL	1,291.7 cM	538 SNPs	2.40 cM	Mathivathana et al. (2019)
Vigna angularis	Ass001 × CWA108	F_2(150)	1,031.17 cM	1571 SNP	0.67 cM	Kai Yang et al. (2015)
Vigna angularis	*Vigna nipponensis*: Yesheng10 × Jihong9218	RIL,153	1,628.15 cM	2032 SLAF	0.80 cM	Liu et al. (2016a)
—		—	—	—	—	—
Yellow lupin	Wodjil cultivar × P28213	RIL(154)	—	2,450	—	Iqbal et al. (2020)
Yellow lupin	Wodjil cultivar × P28213	RIL(154)	2,261.3 cM	2,458	2.29 cM	Iqbal et al. (2019)
Lupinus angustifolius L	83A:476 × P27255	RIL(87)	2,399 cM	34,574 markers/3,508 loci	—	Zhou et al. (2018)
Lupinus angustifolius L	Emir × LAE-1	RIL(92)	3,042 cM	4602 markers	—	Kozak et al. (2019)
Lupinus angustifolius L	83A:476 × P27255	RIL(153)	2,500.8 cM	9,972 loci	0.85 cM	Hane et al. (2017)
Lupinus angustifolius L	Chittick × Geebung	RIL(185)	781.2 cM	2,315	—	Taylor et al. (2021)
Phaseolus lunatus L.	UC 92 ×UC Haskell	RIL	1064 cM	522 loci	2.18 cM	Garcia et al. (2021)
Horsegram	HPK-4 × HPKM-193	F_2	980 cM	1,263 SNPs	—	Shirasawa et al. (2021a)
Vigna vexillata	TVNu 240 × TVNu 1,623	F_2(198)	1,740.9 cM	6,529	0.27 cM	Amkul et al. (2019)
Vigna radiata	*Dahuaye* × *Jilyu 9–1*	RIL	1,060.2 cM	1,946 bin markers	0.54	Wang et al. (2020)
Vigna angularis	*Vigna angularis* ×	F_2(143)	1,365.0 cM	2,904	0.47 cM	Wang et al. (2021b)
—	*V. angularis* var. *nipponensis*	—	—	—	—	—
Vigna aconitifolia	TN67 × ICPMO056	F_2(188)	1,016.8	172	7.34 cM	Yundeng et al. (2019)
Horsegram	HPK4 × HPKM249	RIL(190)	1,423.4 cM	211	9.6 cM	Chahota et al. (2020)
Vigna radiata	VC 1973A × V2984	190, RIL	—	1,321	—	Kang et al. (2014)
Lathyrus	BGE008277× BGE023542	103, RIL	724.2 cM	307	2.4 cM	Santos et al. (2018)
Vigna vexillata	*V. vexillata* (JP235863) ×	F_2	704.8 cM	262	2.87 cM	Dachapak et al. (2018)
—	wild *V. vexillata* (AusTRCF66514)	—	—	—	—	—
Bamara ground nut	IITA686 × Ankpa4	263 F_2	1,395.2 cM	223 markers	—	Ho et al. (2017)
—	Tiga Nicuru × DipC	71 F_3	1,376.7 cM	293 markers	—	—

measured 512,626,114 bp with 27,538 high-confidence genes (Moghaddam et al., 2021).

White lupin (*Lupinus albus* L. 2n = 50) originated from Mediterranean region, contains high protein content (30–40% whole seed) (Bähr et al., 2014) and can use higher soil phosphorus than other legume crops due to its special "cluster root" structure (Lambers et al., 2013). However, improving yield stability and minimizing anti-nutritional alkaloids in white lupin seed through conventional breeding remains challenging. Hence, to elucidate the function of various trait gene(s) related to quality and quantitative importance, Bárbara Hufnagel et al. (2020) assembled a high-quality genome sequence of white lupin, scaling 451 Mb and 38,258 annotated protein-coding genes.

Likewise, narrow-leafed lupin (*Lupinus angustifolious*) is an important grain legume of Mediterranean origin, enriched with dietary protein (40–45%) and fiber (25–30%) (Lee et al., 2006). Hane et al. (2017) assembled the draft genome sequence of Tanjil cultivar, estimating 609 Mb and 33,076 protein-coding genes. Subsequently, Wang et al. (2021a) constructed an improved genome assembly of Tanjil, measuring 615.8 Mb with contig N50 = 5.65 Mb, using a long-read whole-genome sequencing approach.

Common vetch (*Vicia sativa*, 2n = 14) originated from Near Eastern centre of diversity, is a wild and partially domesticated legume crop with a genome size of 1.8 Gb (Shirasawa et al., 2021b). It is used as silage and hay for livestock feeding. The reference genome assembly has been assembled, spanning 1.5 Gb and 31,146 genes (Shirasawa et al., 2021b).

Quantitative Trait Mapping Through Bi-parental and Multi-Parental Schemes

As most of the traits with agricultural importance including biotic, abiotic stress tolerance and quality traits are governed by multiple gene(s)/quantitative trait loci (QTL). In order to map these traits various molecular breeding approaches are available to breeders, including family based bi-parental mapping approach, marker-assisted backcrossing. Subsequently, the availability of high-throughput molecular markers has accelerated the precise mapping of various trait QTLs through employing novel molecular breeding schemes including MutMap, multi-parental cross (MAGIC), genome-wide association mapping, genomic selection and QTL seq approach (Meuwissen et al., 2001; Cavanagh et al., 2008; Takagi et al., 2013; Takagi et al., 2015). In underutilized legumes several bi-parental mapping populations based on interspecific and intraspecific crosses have been developed aiming at constructing genetic linkage map and mapping/tagging targeted trait QTLs of agronomic importance (for details **Table 3**). However, mapping resolution of detected QTLs through bi-parental mapping approach remains low.

TABLE 4 | List of selected QTLs identified in various underutilised legume crops.

Crop	Trait	Mapping population	QTL	LG group	Type of marker	PV%	References
Bambara groundnut	Internode length	IITA686 × Ankpa4, F_2 263	One major QTL	LG2	DArTseq markers	33.4	Ho et al. (2017)
—	—	Tiga Nicuru × DipC, F3 71	—	—	—	—	—
Lupinus angustifolius	Gray leaf spot	83A:476 × P27255, F_8 RIL	One major QTL, *LOC109334326*	LG19	Microsatellite fragment	98	Zhou et al. (2021)
—	—	—	*LOC109334327*	—	Length polymorphism	—	—
Lupinus angustifolius	Vernalisation	Chittick × Geebung, F_2 and RIL	*efl, Trimethylguanosine*	LG14	SNP	81.95%	Taylor et al. (2021)
—	—	—	*Synthase1-like (LanTGS1)*	—	—	—	—
—	—	—	*Lup005529.1*	—	—	—	—
Lupinus albus	Anthracnose	Kiev×P27174 F8, RIL	*antr04_1,antr05_1,antr04_2, antr05_2*	ALB02, ALB04	SNP	14.6–25	Rychel-Bielska et al.(.2020)
—	—	—	*Lalb_Chr02g0142231*	—	—	—	—
—	—	—	*Lalb_Chr02g0141611*	—	—	—	—
—	—	—	*Lalb_Chr02g0141701*	—	—	—	—
—	—	—	*Lalb_Chr04g0264801*	—	—	—	—
Lupinus luteous	Domestication related traits	Wodjil×P28213, RIL(156)	Vernalisation responsiveness locus	YL-21, YL-06	SNP, presence	83%	Iqbal et al. (2020)
—	—	—	Alkaloid content, flower and seed	YL-03 and YL-38	Absence variation	—	—
—	—	—	Colour loci	—	Marker	—	—
Lupinus luteous	Anthracnose resistance	Alu*Prot*-CGNA × PI385149	Anthracnose resistance QTL	4, 10, 11, 13, 23	SNP	75–83%	Lichtin et al. (2020)
—	and early flowering	F_2 (188)	Days to flowering QTL	—	—	—	—
Macrotyloma	Drought and yield	HPK4 × HPKM249 (RIL,190)	*qDFW01, qDFW02, qDTM01*	LG1,4,6 7	SSR, RAPD, COS	7.3–55.3%	Chahota et al. (2020)
uniflorum	—	—	*qRL01, qNSPP01*	—	—	—	—
Vigna radiata	Drought	RIL	58 QTLs	—	SNP	6.40–20	Liu et al. (2017)
Vigna radiata	Plant height	VC 1973A × V2984, RIL, 187	*Height4-1, Height5-1*	LG4, 5	SNP	6.2–30	Ha et al. (2021)
—	Flower initiation	—	*Fl4-1, Fl9-1*	LG4,9	SNP	6.4–24	Ha et al. (2021)
—	No. of branches	—	*Branch3-1*	LG3	SNP	6.4	Ha et al. (2021)
—	No. of nodes	—	*Node4-1, Node11-1*	LG4, 11	SNP	6.3–20	Ha et al. (2021)
—	Synchronous maturity	—	*SPM4-1, SPM7-1*	LG4, 7	SNP	6.8–10.3	Ha et al. (2021)
Vigna aconitifolia	*C. chinensis*	TN67× IPCMO056, F2(188)	*qVacBrc2.1* and *qVacBrc5.1*	LG2 and 5	SSR	—	Somta et al. (2018)
Vigna mungo	*C. maculatus*resistance.	BC48 × TC2210, RIL(150)	*qCm_PDS2.1, qCm_AUDPS6.1*	LG2, 6 and 7	SNP	7.28–30%	Somta et al. (2019)
—	—	—	*qCm_AUDPS6.2, qCm_AUDPS7.1*	—	—	—	—
—	—	—	*qVmunBr6.1* and *qVmunBr6.2*	—	—	—	—
Vigna radiata	Indented Leaflet	*Dahuaye* × *Jilyu 9–1*	Indented Leaflet QTL	LG3 and LG10	SNP	39.70%	Wang et al. (2020)
—	—	—	—	—	—	and 45.4%	—
Vigna radiata	*C. chinensis*	—	*VrPGIP1 and VrPGIP2*	LG5	—	—	Zhang et al. (2021)
Vigna angularis	Flowering time	*Vigna nipponensis: Yesheng10* × *Jihong9218*	*Fld3.2* and *Fld3.3, Fld5.1 vs. Fld5., and Fld5.2 vs. Fld5.5*	LG03, LG05	SLAF	66–71%	Mao-Sen Liu et al. (2016)
Vigna angularis	*Seed size*	*Vigna angularis* ×	12 seed size related QTLs	LG2, 4,5,6 and 9	Indels	3–22%	Wang et al. (2021b)
—	—	*V. angularis* var. *nipponensis*	—	—	—	—	—
			50 QTLs related to	LG 1, 2, 4, 7, and 10	SSR	4.26–53.66%	

(Continued on following page)

TABLE 4 | (*Continued*) List of selected QTLs identified in various underutilised legume crops.

Crop	Trait	Mapping population	QTL	LG group	Type of marker	PV%	References
Vigna aconitifolia	Domestication related traits	TN67 × ICPMO056 F_2(188)					Yundeng et al. (2019)
—	—	—	Domestication related trait	—	—	—	—
Vigna vexillata	Domestication related traits	JP235863 × AusTRCF66514	37 QTLs related to	LGs 5, 6, 7, 8, 10 and 11	SSR, RAD-seq	5.9–52%	Dachapak et al. (2018)
—	—	F_2(139)	Domestication related trait				
Vigna vexillata	22 domestication-related traits	*V. vexillata* (JP235863) ×	37 QTLs	LG1,2,3,4,5,6,7,8,9	SSR, RAD-seq	upto 52%	Dachapak et al. (2018)
—	—	wild V. vexillata (AusTRCF66514)	—	—	—	—	—
—	—	F_2(139)	—	—	—	—	—
Vigna vexillata	*C.chinensis* resistance	TVNu 240 × TVNu 1,623	One major and three minor QTLs	—	SNP	—	Amkul et al. (2019)
—	*C.maculatus* resistance	F_2(198)	one major and	—	—	—	—
—	—	—	one minor QTLs for *C.maculatus*	—	—	—	—

Therefore, to increase the resolution of trait QTLs novel breeding scheme viz., genome-wide association study (GWAS), nested association mapping and MAGIC has been developed. We believe these approaches could be implemented in underutilized legumes to increase the resolution of trait QTLs.

Progress in High-Density Genetic Map Development for Trait Quantitative Trait Loci Discovery and Mapping

Initially, morphological-based markers, isozymes, RFLP, amplified fragment length polymorphisms (AFLP), randomly amplified polymorphic DNAs (RAPD), and SSR markers were used to construct preliminary genetic linkage maps in various underutilized legumes [for details, (Bohra et al., 2014)]. However, the increasing ease of developing high-throughput SNP markers derived by GBS, restriction site-associated DNA sequencing (RAD-seq), and whole genome resequencing has facilitated developing highly dense/saturated consensus linkage maps in various underutilized legumes.

Several genetic maps of mungbean based on SSR markers have been developed (Bohra et al., 2014). Later, a genetic map measuring 1,060.2 cM was developed from an intraspecific mapping population (Wang et al., 2020) and a denser genetic map with 1,291.7 cM and harboring 538 SNPs was developed from an interspecific mapping population derived from *Vigna radiata* × *V. umbellate* cross (Mathivathana et al., 2019) (**Table 4**).

A comprehensive genetic map of urd bean (*V. mungo*) covering 1,588.7 cM with 3,675 SNPs was developed (Somta et al., 2019). Based on a F_2 population, Kai Yang et al. (2015) developed an initial genetic map in adzuki bean measuring 1,031.17 cM. Wang et al. (2021b) presented a denser genetic map measuring 1,365.0 cM in adzuki bean (*V angularis*). In zombi pea (*V. vexillata*), a high-density linkage map spanning 1740.9 cM harboring 6,529 SNPs with an average distance of 0.27 cM between markers was developed from an F_2 mapping population of TVNu 240' × "TVNu 1,623" (Amkul et al., 2019).

Hane et al. (2017) presented a high-density linkage map of narrow-leafed lupin measuring 2,500.8 cM with 9,972 loci and Iqbal et al. (2019) developed a high-density linkage map of yellow lupin measuring 2,261.3 cM. Santos et al. (2018) developed a genetic map of lathryus covering 724.2 cM with 307 loci. Chahota et al. (2020) presented a genetic map for horse gram measuring 1,423.4 cM with 211 loci (**Table4**).

The above linkage maps can be used to identify various traits of biotic, abiotic stress tolerance, agronomic, and culinary importance in numerous underutilized legumes. The selected major trait QTLs identified in the last decade based on bi-parental mapping populations are listed in (**Table 4**). Biotic stress remains the most significant yield stress in underutilized grain legumes globally. The increased availability of genomic resources, especially molecular markers, has identified/tagged various disease-resistant QTLs/gene(s); for example, one major QTL *qCc_PDS6.1* against *Callosobruchus chinensis* (bean weevil) and another QTL *qCm_PDS6.1*against *Callosobruchus maculatus* (cowpea weevil) have been identified (Amkul et al., 2019). Likewise, four major QTLs (*antr04_1*, *antr05_1, antr04_2* and *antr05_2*) controlling anthracnose resistance explaining 14–25% (Rychel-Bielska et al., 2020) of the phenotypic variation in white lupin. Restriction site-associated DNA sequencing derived SNP markers were used as candidate markers for the R gene of phomopsis stem blight disease resistance in narrow-leafed lupin (Yang et al., 2013b). Recently, one major QTL with *LOC109334326, LOC109334327* underlying candidate genes was deciphered for gray leaf spot disease in narrow-leafed lupin (Zhou et al., 2021).

Like biotic stresses, abiotic stresses, particularly drought, causes significant yield losses in underutilized legumes (Liu et al., 2017; Chahota et al., 2020). Several QTLs contributing to drought tolerance have been discovered in mungbean (Liu et al., 2017), and horse gram (Chahota et al., 2020).

Low seed-alkaloid content (<0.02%) is a prime objective of quality improvement in lupin. In lupin the *iucundus* allele is a major gene regulating seed alkaloid content. Several mapping populations have been developed for identifying low alkaloid controlling QTLs and gene(s). Li et al. (2011) identified a microsatellite-anchored fragment length polymorphism-derived PCR marker (lucLi) linked to the low-alkaloid locus *iucundus* (0.9 cM). Likewise, Lin et al. (2009) developed a sequence-specific PCR marker (PauperM1) closely linked (1.4 cM) to the low-alkaloid locus *pauper* in white lupin (*Lupinus albus* L.). Moreover, of five SNP markers co-segregating the *pauper* locus in a set of 140 lupin accessions, the *LAGI01_35805_F1_R1* marker was highly linked with this gene and could be used in low seed alkaloid lupin breeding programs (Rychel and Książkiewicz, 2019). Subsequently, Kroc et al. (2019) developed a co-dominant derived cleaved amplified polymorphic sequence (dCAPS) marker (iuc_RAP2-7) from the *RAP2-7* candidate gene of alkaloid locus *iucundus* responsible for seed alkaloid content in narrow-leafed lupin, which could be used in marker-assisted breeding for low alkaloid content in lupin. Furthermore, fine mapping of this seed alkaloid controlling genomic region unveiled four candidate gene(s)—*LOC109339893, LOC109339862, LOC109339875* and *LOC109339876*—on LG7 in the interval of 20.70–20.89 Mb (Wang et al., 2021a).

Genome-Wide Association Study Approach for Trait Quantitative Trait Loci Identification With Increased Resolution

GWAS is gaining popularity for uncovering genotype–phenotype associations in various plant species, including underutilized legumes (Huang and Han 2014; Liu and Yan 2019), by establishing the genetic basis of the genotype–phenotype association for the trait of interest in a large panel of diverse accessions based on multiple crossing-over events over the recent demographic history of a taxa (Huang and Han 2014). Due to the unprecedented advances in NGS technology, an increasing repertoire of HTP markers in several underutilized legumes have helped to identify loci associated with aspects of complex trait architecture. GWAS has been assisted by the subsequent availability of genome-wide SNP markers for various traits, including phenological traits, quality/nutritional traits, biotic and abiotic stresses, and yield and yield-related traits, in many underutilized legumes (Plewiński et al., 2020). In narrow-leafed lupin, a GWAS incorporating massive analysis of cDNA ends (MACE) markers in 126 gentoypes uncovered significant MTAs related to flower initiation, maturity, plant height, and yield traits (Plewiński et al., 2020). The underlying candidate genes were *Lup019134, Lup015264, Lup021911,* and *Lup021909* for flower initiation, *Lup015264* and *Lup004734* for maturity, *Medtr1g030750* for plant height, and *Lup021835* and *Lup022535* for yield traits (Plewiński et al., 2020).

GWAS has been used increasingly for dissecting complex QTLs controlling various abiotic stresses in crop plants, including underutilized legumes. To elucidate the underlying genomic regions attributing macro- and micro-nutrients in mungbean seeds, Wu et al. (2020) identified 43 MTAs related to calcium, iron, manganese, phosphorus, sulfur, and zinc using inductively coupled plasma (ICP) spectroscopy and GBS-derived SNPs in a set of 95 global mungbean accessions. The explained phenotypic variation ranged from 1 to 38%. Further, Reddy et al. (2021) used a GBS-based GWAS study to dissect the molecular basis of phosphorus uptake efficiency and phosphorus utilization efficiency in 120 mungbean genotypes. The authors uncovered 116 SNPs in 61 protein-coding genes related to phosphorus uptake efficiency and phosphorus utilization efficiency traits. The significantly associated SNPs explained phenotypic variation ranging from 17 to 20% for total phosphorus utilization (under low phosphorus) and it ranged from 15 to 21% for phosphorus utilization efficiency. Six candidate genes—*VRADI01G04370*, *VRADI05G20860*, *VRADI06G12490, VRADI08G20910, VRADI08G00070* and *VRADI09G09030*—regulating phosphorus uptake efficiency and phosphorus utilization efficiency were deciphered (Reddy et al., 2021).

Recently, recruiting 5,041 SNPs in a minicore collection of 293 mungbean accessions identified four significant MTAs for maturation and hypocotyl color within the *Vradi02g04380* gene on chromosome 2 encoding zinc finger A20 and AN1 domain stress-associated protein (Sokolkova et al., 2020). Despite the popularity of GWAS for elucidating marker-trait associations, it has some drawbacks regarding population structure and low-frequency causal alleles causing false negative results (Korte and Farlow 2013). To minimize and overcome the population structure related problems, artificially designed populations such as MAGIC and nested association mapping, could be used [for details (Alseekh et al., 2021)].

Crop Wild Relatives and Their Genome Assembly for Exploring Novel Trait Genes in Underutilized Legumes

CWRs, including those of underutilized legumes, are a hidden reservoir of novel trait gene(s), offering scope for broadening genetic diversity in crop breeding programs (Warschefsky et al., 2014; Zhang and Batley, 2020). In the past, during domestication process, several genes associated with adaptive traits conferring abiotic stress tolerance were lost rendering modern cultivated crop plants to adapt poorly under stress condition (Warschefsky et al., 2014; Zhang and Batley, 2020). However, CWRs serve as reservoir of these biotic and abiotic stress adaptive genes. Thus, recapturing these genes from CWRs through introgression and novel breeding tools could facilitate in increasing the fitness of genepool (Burgarella et al., 2019). Several CWRs of underutilized legumes. e.g., *V. nakashimae*, are potential sources of bruchid resistance (Somta et al., 2006) and salinity tolerance (Yoshida et al., 2016) in adzuki bean. Likewise, harnessing bruchid resistance genes/genomic regions from *Vigna radiata* var. *sublobata* can improve bruchid resistance in mungbean (Schafleitner et al., 2016) (**Table 5**). In urd bean, *V. mungo* var. *silvestris* could be promising for transferring bruchid and mungbean yellow mosaic India virus resistance genes into high-yielding urd bean breeding lines (Souframanien and Gopalakrishna, 2006; Souframanien et al., 2010). Further, the

TABLE 5 | List of CWRs source of novel trait gene in various underutilized legumes.

Crop	Wild species	Importance	References
Adzuki bean	*V. nakashimae*	Bruchid resistance	Somta et al. (2006)
Adzuki bean	*V. angularis* var. *nipponensis*	Domestication- and fitness-related traits	Kaga et al. (2008)
Adzuki bean	JP205833 of *V.riukiuensis*	*Salinity tolerance*	Yoshida et al. (2016)
—	JP107879 of *V.nakashimae*	—	—
Grasspea	*L. articulatus* L. (*IG64782 and IG65197*	*Orobanche crenata*	Abdallah et al. (2021)
—	*IG116989*)	*O. foetida* Poir	—
—	*L. aphaca* L. and *L. ochrus*	—	—
Mungbean	JP 2118749	Bruchid resistance and domestication related traits	Isemura et al. (2012)
Mungbean	*Vigna radiata* var. *sublobata*	Bruchid resistance	Kaewwongwal et al. (2015)
Mungbean	V. radiata var. sublobata TC1966	Bruchid resistance	Schafleitner et al. (2016)
Mungbean	*Vigna umbellata*	Mungbean yellow mosaic virus	Sudha et al. (2015)
Wild vigna	*V. riukiuensis, V. trilobata, V. vexillata*	Salinity tolerance	Yoshida et al. (2020)
—	V. luteola, V. marina	—	—
Urd bean	*V. mungo* var. *silvestris*	Bruchid resistance	Souframanien and Gopalakrishna (2006)
—	—	Mungbean yellow mosaic	Souframanien et al. (2010)
—	—	India virus (MYMIV)	—

genomic sequences of wild underutilized legumes have been assembled to gain insight into the novel trait genes of CWRs. Whole-genome sequencing of *M. ruthenica* offered novel insights into many genes, including the FHY3/FAR1 gene family conferring higher drought tolerance in cultivated *M. sativa* (Wang et al., 2021c). Mou Yin et al. (2021) advocated evidence for multiple family genes and TF family genes, viz., C_2H_2, *CAMTA* and *NAC* attributing various abiotic stress tolerances through chromosome-scale genome sequencing of *M. ruthenica*. Novel SNP and InDel markers were recovered from genome sequencing of *V. radiata* var. *sublobata*; the wild relative accession TC1966 of mungbean could be useful for exploring biotic and abiotic stress tolerant genomic regions through comparative mapping of cultivated mung bean (Kang et al., 2014). Thus, these CWR genomic resources could be used to develop climate-resilient grain legume cultivars.

Implications of Genomic-Assisted Breeding in Underutilized Legumes

Current advances in genomic resource development in underutilized legumes have enabled breeders to develop improved cultivars. For example, tagging various traits in narrow-leafed lupin, such as LanFTc1 PCR-based INDEL markers for vernalization responsiveness locus *Ku/Julius* (Nelson et al., 2017; Plewiński et al., 2019; Taylor et al., 2019), InDel2, InDel10, and PhtjM7 for *PhtjR* (Yang et al., 2013b; Yang H. et al., 2015), Anseq3 and Anseq4 for *Lanr1* (Yang et al., 2012), and TP222136 and TP47110 markers for antr04_1/antr05_1 and TP338761 for antr04_2/antr05_2 (anthracnose resistance) (Rychel-Bielska et al., 2020), the iucLi co-dominant marker (Li et al., 2011) and RAP2-7 PCR-based dCAPS marker for major alkaloid content locus *iucundus* (Kroc et al., 2019) are available. Likewise, a diagnostic marker LAGI01_35805_F1_R1 linked to *pauper* locus controlling low alkaloid content in white lupin could be used for practicing MAS of white lupin lines with low-alkaloid content (Rychel and Książkiewicz 2019). Moreover, co-dominant markers linked to the *tardus* (Li et al., 2010) and *lentus* (Li et al., 2012) genes, attributed to low pod shattering, could be of interest for developing zero shattering narrow-leafed lupin using marker-assisted breeding.

Similarly, CEDG261 and DMB-SSR160 markers linked to bruchid resistance could be used in GAB in moth bean breeding programs (Somta et al., 2018). Downstream application of GAB in concert with other novel breeding approaches for enhancing genetic gain in various underutilized legumes is depicted in **Figure 1**.

Transcriptomics Resources as a Component of Functional Genomics for Gene Discovery With Function in Underutilized Legumes

The advent of NGS-based RNA-seq technology assessing global gene expression has offered a platform for the discovery of functional markers, including EST-SSRs and SNPs, capturing gene space and shedding light on a myriad of trait candidate genes and their plausible functions (O'Rourke et al., 2013; Yang et al., 2017; Glazińska et al., 2019). Previously, EST markers, microarrays, and cDNA libraries were the major functional genomic resources for investigating the function of various trait genes. For example, cDNA library sequencing identified 125,821 unique sequences (O'Rourke et al., 2013) in white lupin.

Subsequently, advances in transcriptome sequencing facilitated the discovery of many unigenes and differentially expressed genes for various traits of importance for details (see **Table 6**). Transcriptome studies have also shed light on the functional role of various underlying candidate gene(s) controlling seed biology, plant phenology, biotic and abiotic stress tolerance, yield traits, and nutritional quality traits, including alkaloid regulation in narrow-leafed lupin, β-N-oxalyl-L-α, β-diaminopropionic acid (β-ODAP) in grass pea and condensed tannin in winged bean (Kroc et al., 2019; Yang et al., 2017; Xu et al., 2018).

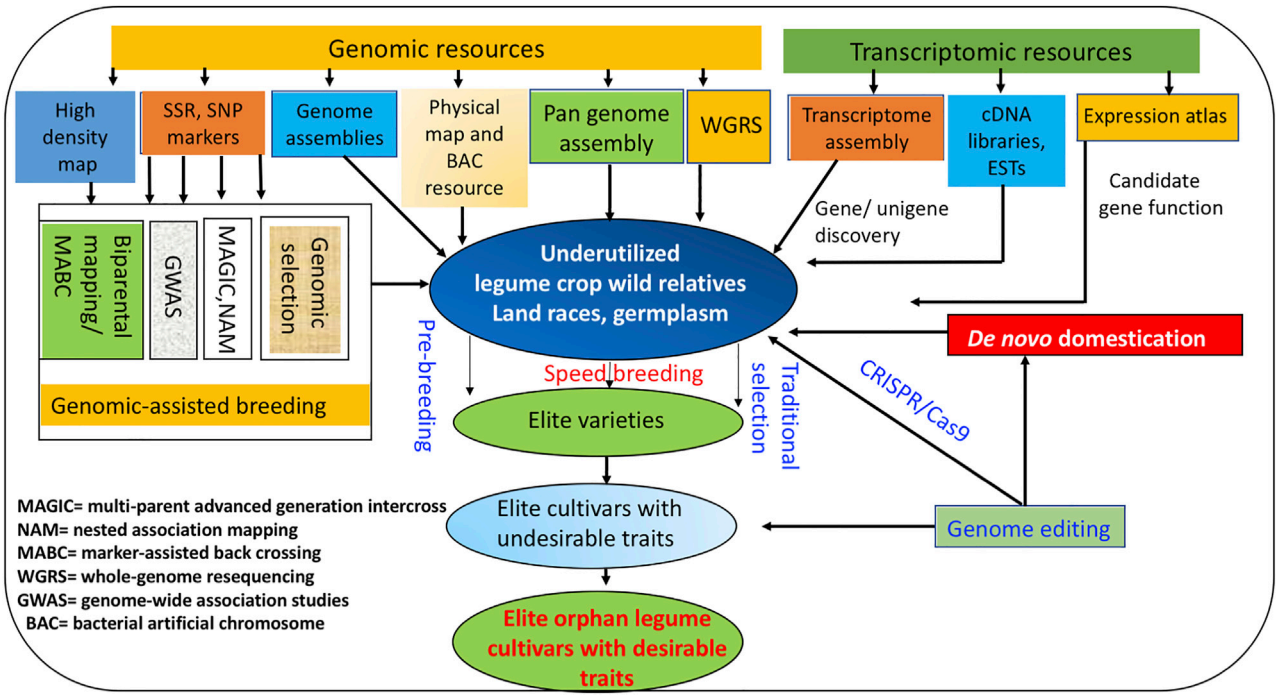


FIGURE 1 | "Omics" and emerging novel breeding approaches for improving genetic gain in underutilized legumes.

In association with small RNA sequencing, degradome sequencing and transcriptome sequencing helped unravel key molecular players, including various phytohormones and metabolic pathways involved in floral development and organ abscission of *L. luteus* (Glazinska et al., 2017; Glazińska et al., 2019). Moreover, participation of small RNA related to seed biology and the conglutin gene encoding seed storage protein was demonstrated in a transcriptome study in narrow-leafed lupin (DeBoer et al., 2019).

Transcriptome studies could improve our understanding of the regulatory mechanisms of the complex network of gene(s), pathogenesis-related genes, phytohormone signaling response, and non-coding RNAs mediating plant immune responses to attacking pathogens (Almeida et al., 2014; Dasgupta et al., 2021). To gain insight into the molecular mechanisms involved in conferring rust resistance in grasspea, an RNA-seq study in rust-responsive grasspea (resistant vs. susceptible) revealed the upregulation of salicylic acid and abscisic acid in the rust-resistant genotype and downregulation of jasmonate and ethylene pathways in the susceptible genotype (Almeida et al., 2014) (**Table 6**). Additionally, several pathogenesis-related genes and the mildew resistance locus O (MLO)-like resistance gene were discovered in this study.

An RNA-seq study offered insight into the participatory role of WRKY, NAC and MYB transcription factors, phytoene synthase, cytochrome P450, and JAZ and LOX genes attributing to mungbean yellow mosaic virus (MYMV) resistance (Dasgupta et al., 2021).

Likewise, transcriptome studies can decipher the complex molecular mechanisms and underlying possible candidate gene(s) networks during perceiving abiotic stress signaling and mediate various abiotic stress tolerances by activating antioxidant mechanisms and other cellular protective mechanisms, enabling plants to acclimate to abiotic stress (Bhardwaj et al., 2013; Butsayawarapat et al., 2019; De la Rosa et al., 2020).

A *de novo* transcriptome analysis of two contrasting horse gram genotypes for drought tolerance revealed the involvement of various TFs (NAC, MYB, and WRKY families) in conferring drought stress tolerance (Bhardwaj et al., 2013). De novo transcriptome sequencing of contrasting drought tolerant and sensitive genotypes of common vetch revealed a plethora of differentially expressed genes under water stress (De la Rosa et al., 2020). Most of the genes mediating drought tolerance are associated with cell wall modification, oxidative stress response and ABA response (De la Rosa et al., 2020). In zombi pea, a comparative transcriptome analysis revealed up-regulatory activity of glycolysis and fermentative genes in the waterlogging-sensitive genotype; in contrast, the waterlogging-tolerant genotype had enhanced activity of auxin-regulated lateral root initiation, aquaporin, and peroxidase genes (Butsayawarapat et al., 2019) (**Table 6**).

Deciphering the underlying genes and molecular function of quality parameters, including nutritional and industrially important parameters, using transcriptomic studies could improve these traits (Yang et al., 2017, Xu et al., 2018; Tyagi et al., 2018). Small RNA sequencing indicated the involvement of several miRNAs and their target genes coding for carbohydrate metabolism, kinase, and enzymes for regulating galactomannan biosynthesis in cluster bean (Tyagi et al., 2018) (**Table 6**). The authors also discovered two novel unigenes, mannosyltransferase/mannan synthase (ManS) and UDP-

TABLE 6 | List of various differentially expressed genes/candidate traits genes with putative function in underutilised legumes.

Crop	Trait	Candidate genes/Unigenes/DEG	Function	Platform used	References
Common vetch	Drought stress tolerance	2,646 transcripts are DEG	Redox homeostasis, cell wall modifications	Illumina HiSeq 2,500	De la Rosa et al. (2020)
Common vetch	Pod shattering	1,285 DEGs and 575 upregulated unigenes	Hydrolase activity	HiSeq 2000	Dong et al. (2017)
—	—	710 downregulated unigenes	Carbohydrate metabolic process	—	—
Guar	Root development	102,479 unigenes	Root development	Illumina HiSeq 2,500	Thakur and Randhawa (2018)
—	stress tolerance	—	Stress tolerance	—	—
Guar	—	11,308	Carbohydrate, protein, lipid, energy	Illumina HiSeq	Tanwar et al. (2017)
—	—	—	Nucleotide metabolism	—	—
Guar	Galactomannan	187 known and 171 novel miRNAs differentially expressed	Regulating galactomannan pathway	Illumina NextSeq 500	Tyagi et al. (2018)
—	Biosynthesis	—	—	—	—
Guar	—	38423 DEGs	Metabolic process, cellular process	Illumina	Rawal et al. (2017)
Guar	Various abiotic stress	61,508 putative genes	Biological process, cellular component and molecular function	Illumina HiSeq 2,500	Al-Qurainy et al. (2019)
Guar	Galactomannan	Cellulose synthase D1, GAUT-like gene	Galactomannan biosynthesis pathway	Illumina HiSeq. 4000	Chaudhury et al. (2019)
—	Biosynthesis pathway	—	—	—	—
Guar—	Galactomannan	5,147 DEGs	*LBD, BZIP, NAC,* and *C2H2, BHLH, MYB*	Illumina Hiseq X Ten	Sharma et al. (2021)
—	Biosynthesis pathway	—	—	—	—
Horsegram	Drought	21,887 unigenes	Calmodulin binding factor, heat shock protein	Illumina GAIIx	Bhardwaj et al. (2013)
—	—	—	DEAD-box ATP dependent RNA helicase	—	—
Lathyrus	Rust tolerance	134,914 contigs	Regulating phytohormone signalling	Illumina Hiseq2000	Almeida et al. (2014)
Lathyrus	*Ascochyta lathyri*	738 unitags	Cell wall metabolism	DeepSuperSAGE	Almeida et al. (2015)
Lathyrus	—	27,431 unigenes	—	Ilumina NextSeqTM 500	Hao et al. (2017)
Lathyrus	Rust	4520 and 3,498 contigs down regulated	Hormone metabolism, cell wall degradation	Illumina Hiseq2000	Santos et al. (2018)
—	—	—	Secondary metabolism, ROS production	—	—
Lathyrus	β-ODAP	213,258 unigenes	Carbohydrate and	Illumina-HiSeq 3,000	Xu et al. (2018)
—	—	—	sulfur assimilation/metabolism	—	—
—	—	—	nucleic acid metabolism like purine and pyrimidine		—
Lupinus albus	Phosphorus	2,128 sequences differentially	Cluster root development	Illumina GA-IIx	O'Rourke et al. (2015)
—	—	expressed in response to Pi deficiency	—	—	—
Lupinus angustifolius	—	10,240 transcripts	Peroxidase and anthocyanin biosynthesis	Illumina HiSeq 2000	Kamphuis et al. (2015)
—	—	—	Basal pathogen defences	—	—
Lupinus angustifolius	quinolizidine alkaloids	12 candidate genes, RAP2-7, AP2/ERF TF	Quinolizidine synthesis	Illumina HiSeq 1,500	Kroc et al. (2019)
L. angustifolius	quinolizidine alkaloids	33 genes related to lupin alkaloid biosynthesis	Copper amine oxidase	Illumina HiSeq 2,500	Yang et al. (2017)
Medicago ruthenica	Drought tolerance	3,905 genes and 50 miRNAs	gma-miR171j-5p and mtr-miR396a-5p down regulated	Illumina Hiseq4000	Shi et al. (2021)
Mungbean	MYMV	1881, 1,449, 1,583 and 1,140 genes as up-regulated	Defence related activity	Illumina HiSeq 2,500	Dasgupta et al. (2021)
—	—	1,423, 1,154, 1,396 and 1,152 genes as down-regulated		—	—
Mungbean	Osmotic response	13 *OSCA* genes	Contributes in salinity and drought tress tolerance		Mou Yin et al. (2021)
Mungbean	—	1,245	biological process, cellular component and molecular function	Illumina	Changyou Liu et al. (2016)

(Continued on following page)

TABLE 6 | (*Continued*) List of various differentially expressed genes/candidate traits genes with putative function in underutilised legumes.

Crop	Trait	Candidate genes/Unigenes/DEG	Function	Platform used	References
Psophocarpus tetragonolobus	—	5,053 transcript have predicted functions	biological process, cellular component	Illumina platform	Vatanparast et al. (2016)
—	—	—	—	—	—
Trifolium ambiguum	Rhizome development	betaine aldehyde dehydrogenase	rhizome growth and development	PacBio sequencing	Yin et al. (2020)
—	—	276 DEGs involved in hormone signalling and transduction		and Illumina sequencing	—
Trifolium pratense	Drought	45181 contigs	Role of proline, malate and pinitol Contributing to drought tolerance	Illumina HiSeq 2000 (Illumina, United States)	Yates et al. (2014)
Trifolium pratense	Seed setting	1,196 DEGs	These gene(s) involved in seed setting	Illumina sequencing platform (HISEQ 2000)	Kovi et al. (2017)
Trifolium pratense	Regrowth	Phytohormone related genes	Gibberellin-related genes regulate regrowth in association with other phytohormones	Illumina Hiseq2000	Herbert et al. (2021)
Trifolium pratense	Iso-flavonoid	143 iso-flavonoid synthesis genes	Role various genes and long non coding RNAs contributing to iso-flavonoid synthesis	Illumina HiSeq X Ten platform	Shi et al. (2021)
—	—	stem specific genes (*TpPAL, TpC4H*, and *Tp4CL*)	—	—	—
—	—	Root specific genes (*TpCHS*, *TpCHI1*, and *TpIFS*)	—	—	—
Trifolium repens	Flower pigmentation	6,282 DEGs, *CHS, F3′H, F3′5′H, UFGT, FLS, LAR, ANS,* and *DFR*)	Anthocyanin flavonoid biosynthetic pathway and flavonoid biosynthetic pathway	Illumina Hiseq ×10	Heshan Zhang et al. (2018)
urd bean	—	2,306 DEGs	Cytochrome c-type biogenesis protein	Illumina MiSeq	Raizada and Souframanien (2019)
—	—	—	DnaJ protein homolog 1	—	
—	—	—	Uncharac- terized protein LOC108329961	—	—
Urdbean	—	29564 transcript contigs	Purine metabolism, pyrimidine metabolism	Illumina	Souframanien and Reddy (2015)
Vigna angularis	—	65,950 unigenes	RING-H2 finger protein	—	Chen et al. (2015a)
—	—	—	A serine/threonine protein kinase	—	—
—	—	—	Lipase ROG1-like protein	—	—
Vigna angularis	—	324,219 and 280,056 transcripts	—	Illumina HiSeq 2000	Jo et al. (2016)
Vigna angularis	Drought	5,337 DEGs	Hormone signal transduction	Illumina HiSeqX	Zhu et al. (2020)
—	—	—	Transcript or translation processes	—	—
—	—	—	Ubiquitin proteasome system	—	—
Trifolium repens	Heat stress	*Upregulation of PIP1-1* and *PIP2-7* in leaves and the TIP2-1	Induction of aquaporin genes	qRT-PCR	Qi et al. (2021)
—	—		Causing heat stress tolerance	—	—
Zombipea	Water logging	982 and 1,133 DEGs	Induction of Cell wall modification	Illumina HiSeq 4000	Butsayawarapat et al. (2019)
—	—	—	Aquaporin, and peroxidase genes	—	—
—	—	—	Auxin Metabolism	—	—

D-glucose 4-epimerase (UGE), targeted by Ct-miR3130, Ct-miR3135, and Ct-miR3157 miRNAs. Likewise, an RNA-seq study revealed preferential expression of 2,535 and 2,724 genes in endosperm and 3,720 and 2,530 genes in the embryo involved in guar gum biosynthesis (Hu et al., 2019).

Transcriptome assembly through RNA-seq identified several candidate genes regulating quinolizidine alkaloids (QAs) biosynthesis, an anti-nutritional factor in narrow-leafed lupin (Kamphuis et al., 2015; Yang et al., 2017; Kroc et al., 2019). Short-read sequencing using Illumina HiSeq2500 in association with long-read sequencing using PacBio technology of high QA-containing genotypes identified 33 candidate genes associated with QA biosynthesis in narrow-leafed lupin (Yang et al., 2017). Furthermore, transcriptome profiling offered insight into the genes involved in the accumulation and degradation of β-N-oxalyl-L-α, β-diaminopropionic acid (β-ODAP), a neurotoxin found in grasspea (Xu et al., 2018). Similarly, RNA-seq analysis of high- and low-tannin-containing lines of winged bean, using Illumina Nextseq 500, revealed 1,235 differentially expressed contigs in these two lines. Several genes related to condensed tannin were elucidated, including *anthocyanidin 3-O-glucosyltransferase (A-3GOT), anthocyanidin synthase* (*ANS*), *chalcone synthase* (*CHS*) *phenylalanine ammonia-lyase (PAL)* (Singh et al., 2017).

Scope of Genomic Selection/Genomic Prediction for Increasing Genetic Gain in Underutilized Legumes

The decoding of various underutilized legume genome sequences and resequencing efforts have made SNP markers accessible, providing great opportunities to perform genomic selection (GS). This approach has been used for estimating the genomic breeding value of tested individuals without any prior phenotypic information by measuring the genome-wide marker effect based on various prediction models (Meuwissen et al., 2001). Thus, the benefits of GS could be harnessed for the selection of progenies with known genotypic scores with high "genetic merit" for improving genetic gain.

Assessing anthracnose resistance in white lupin using GS based on GBS-derived SNPs in the ridge regression BLUP model, Rychel-Bielska et al. (2020) reported a moderately high predictive ability (0.56). Application of GS is very limited in minor legumes; however, increasing repertoire of genome wide SNP markers will greatly assist in implementing GS for improving future genetic gain in these legumes.

Scope of Speed Breeding, an Innovative Approach for Enhancing Breeding Efficiency in Underutilized Legumes

Speed breeding could be used to increase breeding efficiency by shortening the breeding cycle and reducing plant space, cost, and labor resources, thereby increasing genetic gain (Watson et al., 2018; Hickey et al., 2019). Speed breeding protocols have been established by optimizing photoperiod, daylength, and temperature in various legume crops, including soybean (Fang et al., 2021), chickpea (Samineni et al., 2019), pigeonpea (Saxena et al., 2019), and pea (Mobini and Warkentin, 2016). However, this approach has not been implemented in any underutilized legumes. Thus, the establishment of a speed breeding protocol could open up new avenues for improving genetic gain in various underutilized legumes more quickly than traditional breeding methods.

Resequencing and Pangenome Assembly for Capturing Novel Structural Variations Across the Whole Genome

With the declining costs of genome assembly construction, whole genome resequencing is gaining popularity for uncovering genomic regions controlling traits of agronomic importance in a large set of global crop germplasm (Hufnagel B. et al., 2020).

The WGRS approach can elucidate the causal candidate gene(s)/genomic regions associated with traits of interest. Like other major grain legumes, WGRS has been used in underutilized legume crops (Yang H. et al., 2015; Hufnagel B. et al., 2020). The resequencing of nine lupin cultivars discovered 180,596–795,735 SNP markers and 243 candidate diagnostic markers linked to the *PhtjR* (phomopsis stem blight disease) gene (Yang H. et al., 2015). Of these candidate diagnostic markers, nine were validated in commercial cultivars, offering an opportunity to practice marker-assisted breeding for phomopsis stem blight disease resistance in narrow-leafed lupin.

Resequencing 11 modern cultivars, two landraces, and one wild relative of white lupin and comparing them with the reference genome sequence revealed the recent breeding history of white lupin (Hufnagel B. et al., 2020). Similarly, 38 narrow-leafed lupin accessions, including 19 wild and 19 cultivated types, with 19× coverage of the genome were resequenced to reveal the genomic signal for domestication and genes associated with the domestication process (Wang et al., 2021a). A selective sweep analysis in the same study identified 303 genomic regions under strong selection, with 8.2% of the genome under selection associated with domestication. Further, these selective sweeps harbored nine key domestication-related traits, including early flowering, reduced pod shattering, white flower, and low alkaloid (Wang et al., 2021a). WGRS efforts of three mungbean accessions using the Ion Torrent Personal Genome Machine™ (PGM™) platform identified 233,799 SNPs and 9,544 insertions and deletions in coding and non-coding regions, revealing great opportunity for future mung bean improvement using genomic-assisted breeding (Bangar et al., 2021).

Previously, molecular biologists and geneticists have relied mainly on the "single reference genome sequence" of a species for genetic analyses within and across species (Sherman and Salzberg, 2020; Della Coletta et al., 2021). However, the single reference genome sequence does not explain all of the genomic variation/structural variants available within and across species; "pangenomics" studies can capture all of the genomic information in a species. The pangenome refers to the entire non-redundant DNA sequences existing in a species, constituting the "core" genome common to all individuals in a species, with

"dispensable" genome the variable fraction or "accessory" genome (Tettelin et al., 2005; Sherman and Salzberg, 2020; Della Coletta et al., 2021; Lei et al., 2021). In the context, Hufnagel et al. (2021) constructed the pangenome of white lupin using a "map to pan" approach (Hu et al., 2017) by sequencing 39 accessions, which identified 32,068 core genes and 14,822 dispensable genes. They also identified 333 selection sweeps related to low alkaloid content and candidate genes (*LaDHDPS, LaHLT,* and *LaAT*) controlling alkaloid content. Pangenome analyses of other underutilized legumes could provide novel insights into genomic variation for future trait discovery.

Several legume genera have multiple domesticated species. For example, *Vigna* has 10 domesticated taxa, *Phaseolus* seven, and *Lupinus* four. Super pan-genomes across these genera might have immense power to provide insight into similarities in domestication syndromes, the genetic basis of traits influencing geographic distribution, and disease and pest resistance.

Hope and Progress of Genome Editing in Underutilized Grain Legumes

Despite the success of transferring gene(s) of interest into high-yielding cultivars, environmental biosafety and regulatory governing bodies have not allowed the widespread adoption of transgenic technology (Zhang Y. et al., 2018).

Genome editing tools, especially the CRISPR/Cas9 based technique, has revolutionized functional genomics and plant breeding, creating novel genetic variation in plants by editing targeted genes of interest with precision and efficiency (Chen and Gao, 2014). Examples of genome editing in various crops are increasing (Chen and Gao 2014; Zhang Y. et al., 2018); however, there has been limited success in legume species. Notable instances of CRISPR/Cas9 mediated genome editing have been reported in soybean (Cai et al., 2015; Sun et al., 2015; Han et al., 2019), cowpea (Ji et al., 2019) and *Medicago trancatula* (Michno et al., 2015). In case of cowpea, Ji et al. (2019) employed CRISPR/Cas9 based genome editing tool in the symbiosis receptor -like kinase target gene *VuSYMRK* that controls nodule symbiosis in cowpea. The edited plant exhibited complete inhibition in nodule formation and consequently, the mutant plants were unable to synthesise nodules in association with *Sinorhizobium* sp. strain NGR234. Furthermore, complete male and female sterile plants were generated by editing *SPO11-1* gene through CRISPR/Cas9 technology in cowpea (Juranić et al., 2020). In the context of underutilized legume, the CRISPR/Cas9 genome engineering technique was used to edit the isoflavone synthase gene contributing to rhizobial defense signaling in red clover (Dinkins et al., 2021). Furthermore, gene-editing technology in association with base editors and prime-editing could be harnessed for *de novo* domestication of CWRs of underutilized legumes and "reengineering of metabolism" to increase resilience and enhance nutritive value (Gasparini et al., 2021; Nasti and Voytas 2021).

Scope of *de Novo* Domestication of Underutilized Legumes

Crop wild relatives are the richest reservoir of genetic diversity for improving various biotic and abiotic stress resistance in crop plants and could therefore be used as new crops through "*de novo* domestication" or "redomestication" process (Fernie and Yan 2019; Von Wettberg et al., 2021). Domestication of new legume underutilized crops from their wild relatives could strengthen crop diversity, and thus be vital for sustainable agriculture (Zhang et al., 2018b). Among the various underutilized grain legume species, *Vigna stipulaceae* could be targeted for *de novo* domestication due to its inherent capacity for drought and salinity stress tolerance and reduced pod shattering (Takahashi et al., 2019). Likewise, being an "incompletely domesticated species" and having inherent stress tolerance ability against biotic and abiotic stress, hairy vetch (*Vicia villosa*) is an ideal legume crop for *de novo* domestication (Renzi et al., 2020).

Of the various approaches, mutagenesis and forward screening and CRISPR/Cas9 based gene editing are important techniques for introducing domestication-related traits in wild relatives for *de novo* domestication (Shapter et al., 2013; Li et al., 2018). Ethyl methanesulfonate mutagenesis and forward screening enabled the domestication of *Vigna stipulacea* Kuntze by selecting mutants with reduced pod shattering and reduced seed dormancy (Takahashi et al., 2019). Likewise, CRISPR/Cas9 genome editing technology could be used to eliminate g-glutamyl-b-cyano-alanine (GBCA) toxin from seeds of common vetch (*Vicia sativa*), providing a zero-toxin vetch variety for combating the rising global protein demand (Nguyen et al., 2020).

CONCLUSION AND FUTURE PERSPECTIVES

Given the rising demand for food, feed, and forage, there is an urgent need to develop sustainable food resources. Underutilized legumes are versatile crops with great potential for mitigating global food security challenges, but they are lagging behind major legumes in terms of genomic resource development. More genomic sequencing of CWRs, landraces, and improved breeding lines will provide novel insights into genomic variations for investigating evolution, domestication events, and the diversification of underutilized legumes. Increasing genomic resources will allow increased genome-assisted breeding of these legumes. Likewise, WGRS in association with GWAS and pangenome integration with GWAS could underpin the causal genes/haplotypes of complex traits of interest. Emerging genome editing techniques could play a critical role in minimizing toxins or negative parameters associated with various nutritional quality traits, such as editing GBCA encoding gene(s) in common vetch, BOAA encoding gene(s) in grasspea, and genes involved in producing QAs in white lupin. These technologies also have great potential

for introducing *de novo* domestication in CWRs by removing phenotypically undesired traits in various CWRs of underutilized legumes.

Moreover, genomic selection and speed breeding approaches could enhance genetic gain in underutilized legumes. The rich diversity in these underutilized legumes needs proper collection, conservation, and characterization (Kamenya et al., 2021). Furthermore, the establishment of sound varietal releases and seed distribution systems could play a central role in popularizing these climate-smart underutilized legumes among farmers (Bohra et al., 2020). Disseminating knowledge on the global demand and profitability of these legumes needs strengthening via extension services, especially in developing countries (Kamenya et al., 2021). Hence, collective genomics, novel breeding knowledge, and sound seed system approaches could improve underutilized legume productivity for securing global food security.

AUTHOR CONTRIBUTIONS

UJ and HN developed the conceptual structure. UJ and KS prepared the original draft. EvW, SP, and MB contributed specific sections. KS edited manuscript. All authors read and approved the final manuscript.

FUNDING

EvW is partially supported in work on underutilized legumes by USDA Hatch funding and by the Ministry of Science and Higher Education of the Russian Federation as part of World-class Research Center program: Advanced Digital Technologies (Contract No. 075-15-2020-934 dated 17 November 2020).

ACKNOWLEDGMENTS

UJ acknowledges support from Indian Council of Agricultural Research (ICAR), New Delhi.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.831656/full#supplementary-material

REFERENCES

Abdallah, F., Kumar, S., Amri, A., Mentag, R., Kehel, Z., Mejri, R. K., et al. (2021). Wild Lathyrus Species as a Great Source of Resistance for Introgression into Cultivated Grass Pea (*Lathyrus sativus* L.) against Broomrape Weeds (*Orobanche crenata* Forsk. and *Orobanche foetida* Poir.). *Crop Sci.* 61, 263–276. doi:10.1002/csc2.20399

Adegboyega, T. T., Abberton, M. T., AbdelGadir, A. H., Dianda, M., Maziya-Dixon, B., Oyatomi, O. A., et al. (2019). Nutrient and Antinutrient Composition of Winged Bean (*Psophocarpus tetragonolobus* (L.) DC.) Seeds and Tubers. *J. Food Qual.* 2019, 1–8. doi:10.1155/2019/3075208

Al-Qurainy, F., Alshameri, A., Gaafar, A. R., Khan, S., Nadeem, M., Alameri, A. A., et al. (2019). Comprehensive Stress-Based De Novo Transcriptome Assembly and Annotation of Guar (*Cyamopsis tetragonoloba* (L.) Taub.): An Important Industrial and Forage Crop. *Int. J. Genomics* 2019, 7295859. doi:10.1155/2019/7295859

Almeida, N. F., Krezdorn, N., Rotter, B. r., Winter, P., Rubiales, D., and Vaz Patto, M. C. (2015). Lathyrus Sativus Transcriptome Resistance Response to *Ascochyta lathyri* Investigated by deepSuperSAGE Analysis. *Front. Plant Sci.* 6, 178. doi:10.3389/fpls.2015.00178

Almeida, N. F., Leitão, S. T., Krezdorn, N., Rotter, B., Winter, P., Rubiales, D., et al. (2014). Allelic Diversity in the Transcriptomes of Contrasting Rust-Infected Genotypes of *Lathyrus sativus*, a Lasting Resource for Smart Breeding. *BMC Plant Biol.* 14 (1), 1–15. doi:10.1186/s12870-014-0376-2

Alseekh, S., Kostova, D., Bulut, M., and Fernie, A. R. (2021). Genome-wide Association Studies: Assessing Trait Characteristics in Model and Crop Plants. *Cell. Mol. Life Sci.* 78, 5743–5754. doi:10.1007/s00018-021-03868-w

Amarowicz, R., Estrella, I., Hernández, T., and Troszyńska, A. (2008). Antioxidant Activity of Extract of Adzuki Bean and its Fractions. *J. Food Lipids* 15 (1), 119–136. doi:10.1111/j.1745-4522.2007.00106.x

Amkul, K., Wang, L., Somta, P., Wang, S., and Cheng, X. (2019). Construction of a High Density Linkage Map and Genome Dissection of Bruchid Resistance in Zombi Pea (*Vigna vexillata* (L.) A. Rich). *Sci. Rep.* 9 (1), 1–10. doi:10.1038/s41598-019-48239-5

Arora, R. K., and Chandel, K. P. S. (1972). Botanical Source Area of Wild Herbage Legumes in India. *Trop. Grassland* 6, 213–221.

Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA Content of Some Important Plant Species. *Plant Mol. Biol. Rep.* 9, 208–218. doi:10.1007/bf02672069

Awurum, A. N., and Emechebe, A. M. (2001). The Effects of Leaf Spot Diseases and Staking on Yield and Yield Attributes of Winded Bean Psophocarpus Tetragonolobus (L. Dc.). *J. Appl. Chem. Agric. Res.* 7, 42–47. doi:10.4314/jacar.v7i1.41123

Bähr, M., Fechner, A., Hasenkopf, K., Mittermaier, S., and Jahreis, G. (2014). Chemical Composition of Dehulled Seeds of Selected Lupin Cultivars in Comparison to Pea and Soya Bean. *Lwt-food Sci. Technol.* 59, 587–590. doi:10.1016/j.lwt.2014.05.026

Bangar, P., Tyagi, N., Tiwari, B., Kumar, S., Barman, P., Kumari, R., et al. (2021). Identification and Characterization of SNPs in Released, Landrace and Wild Accessions of Mungbean (*Vigna radiata* (L.) Wilczek) Using Whole Genome Re-sequencing. *J. Crop Sci. Biotechnol.* 24, 153–165. doi:10.1007/s12892-020-00067-0

Barbetti, M. J. (2007). Resistance in Annual Medicago Spp. To Phoma Medicaginis and Leptosphaerulina Trifolii and its Relationship to Induced Production of a Phytoestrogen. *Plant Dis.* 91, 239–244. doi:10.1094/pdis-91-3-0239

Bassal, H., Merah, O., Ali, A. M., Hijazi, A., and El Omar, F. (2020). Psophocarpus Tetragonolobus: An Underused Species with Multiple Potential Uses. *Plants* 9, 1730. doi:10.3390/plants9121730

Bauchet, G. J., Bett, K. E., Cameron, C. T., Campbell, J. D., Cannon, E. K., Cannon, S. B., et al. (2019). The Future of Legume Genetic Data Resources: Challenges, Opportunities, and Priorities. *Legume Sci.* 1 (1), e16. doi:10.1002/leg3.16

Bazzano, L. A., He, J., Ogden, L. G., Loria, C., Vupputuri, S., Myers, L., et al. (2001). Legume Consumption and Risk of Coronary Heart Disease in US Men and Women. *Arch. Intern. Med.* 161, 2573–2578. doi:10.1001/archinte.161.21.2573

Bennett, M. D., and Leitch, I. J. (2012). Angiosperm DNA C-Values Database. Available online at: http://data.kew.org/cvalues/.

Berendzen, J., Brown, A. V., Cameron, C. T., Campbell, J. D., Cleary, A. M., Dash, S., et al. (2021). The Legume Information System and Associated Online Genomic Resources. *Legume Sci.* 2021, e74. doi:10.1002/leg3.74

Bhardwaj, J., Chauhan, R., Swarnkar, M. K., Chahota, R. K., Singh, A. K., Shankar, R., et al. (2013). Comprehensive Transcriptomic Study on Horse Gram (*Macrotyloma uniflorum*): De Novo Assembly, Functional Characterization

and Comparative Analysis in Relation to Drought Stress. *BMC Genomics* 14, 647–648. doi:10.1186/1471-2164-14-647

Bhartiya, A., Aditya, J. P., and Kant, L. (2015). Nutritional and Remedial Potential of an Underutilized Food Legume Horsegram [*Macrotyloma Uni Orum* (Lam.) Verdc.] : a Review. *J. Anim. Plant Sci* 25, 908–920.

Blink, M., and Jansen, P. C. M. (2006). "*Vigna Aconitifolia* (Jacq.) Maréchal," in *Plant Genetic Resources of Tropical Africa. 1. Cereals and Pulses*. Editors M. Blink and G. Belay (Wageningen, Netherlands, Leiden, Netherlands: Backhuys Publishers), 200–203. Prota Foundation.

Bohra, A., Chand Jha, U., Godwin, I. D., and Kumar Varshney, R. (2020). Genomic Interventions for Sustainable Agriculture. *Plant Biotechnol. J.* 18, 2388–2405. doi:10.1111/pbi.13472

Bohra, A., Jha, U. C., Kishor, P. B. K., Pandey, S., and Singh, N. P. (2014). 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

Boschin, G., and Arnoldi, A. (2011). Legumes Are Valuable Sources of Tocopherols. *Food Chem.* 127, 1199–1203. doi:10.1016/j.foodchem.2011.01.124

Boschin, G., D'Agostina, A., Annicchiarico, P., and Arnoldi, A. (2008). Effect of Genotype and Environment on Fatty Acid Composition of Lupinus Albus L. Seed. *Food Chem.* 108, 600–606. doi:10.1016/j.foodchem.2007.11.016

Bounejmate, M., Robson, A., and Beale, P. (1992). Annual Medicago Species in Morocco. II. Distribution in Relation to Soil and Climate. *Aust. J. Agric. Res.* 43, 751–763. doi:10.1071/ar9920751

Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., et al. (2019). Adaptive Introgression: an Untapped Evolutionary Mechanism for Crop Adaptation. *Front. Plant Sci.* 10, 4. doi:10.3389/fpls.2019.00004

Butsayawarapat, P., Juntawong, P., Khamsuk, O., and Somta, P. (2019). Comparative Transcriptome Analysis of Waterlogging-Sensitive and Tolerant Zombi Pea (*Vigna vexillata*) Reveals Energy Conservation and Root Plasticity Controlling Waterlogging Tolerance. *Plants* 8, 264. doi:10.3390/plants8080264

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., et al. (2015). CRISPR/Cas9-Mediated Genome Editing in Soybean Hairy Roots. *PLoS One* 10, e0136064. doi:10.1371/journal.pone.0136064

Campbell, C. G. (1997). *Grass Pea: Lathyrus Sativus L. Promoting the Conservation and Use of Underutilized and Neglected Crops*. International Plant Genetic Resources Institute, Vol. 18.

Campbell, C. G., Mehra, R. B., Agrawal, S. K., Chen, Y. Z., Abdel Moneim, A. M., Khawaja, H. I. T., et al. (1994). "Current Status and Future Strategy in Breeding Grasspea (*Lathyrus sativus*)," in *Expanding the Production and Use of Cool Season Food Legumes* (Dordrecht: Springer), 617–630. doi:10.1007/978-94-011-0798-3_37

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From Mutations to MAGIC: Resources for Gene Discovery, Validation and Delivery in Crop Plants. *Curr. Opin. Plant Biol.* 11, 215–221. doi:10.1016/j.pbi.2008.01.002

Cerda-Hurtado, I. M., Mayek-Pérez, N., Hernández-Delgado, S., Muruaga-Martínez, J. S., Reyes-Lara, M. A., Reyes-Valdés, M. H., et al. (2018). Climatic Adaptation and Ecological Descriptors of Wild Beans from Mexico. *Ecol. Evol.* 8, 6492–6504.

Chahota, R. K., Sharma, V., Rana, M., Sharma, R., Choudhary, S., Sharma, T. R., et al. (2020). Construction of a Framework Linkage Map and Genetic Dissection of Drought-and Yield-Related QTLs in Horsegram (*Macrotyloma uniflorum*). *Euphytica* 216, 1. doi:10.1007/s10681-020-02583-0

Chandel, K. P. S., Lester, R. N., and Starling, R. J. (1984). The Wild Ancestors of Urid and Mung Beans (*Vigna mungo* (L.) Hepper and *V. radiata* (L.) Wilczek). *Bot. J Linn. Soc.* 89, 85–96. doi:10.1111/j.1095-8339.1984.tb01002.x

Chang, Y., Liu, H., Liu, M., Liao, X., Sahu, S. K., Fu, Y., et al. (2019). The Draft Genomes of Five Agriculturally Important African Orphan Crops. *GigaScience* 8, giy152. doi:10.1093/gigascience/giy152

Chapman, M. A. (2015). Transcriptome Sequencing and Marker Development for Four Underutilized Legumes. *Appl. Plant Sci.* 3, 1400111. doi:10.3732/apps.1400111

Chaudhury, A., Kaila, T., and Gaikwad, K. (2019). Elucidation of Galactomannan Biosynthesis Pathway Genes through Transcriptome Sequencing of Seeds Collected at Different Developmental Stages of Commercially Important Indian Varieties of Cluster Bean (*Cyamopsis tetragonoloba* L.). *Sci. Rep.* 9 (1), 1–17. doi:10.1038/s41598-019-48072-w

Chen, K., and Gao, C. (2014). Targeted Genome Modification Technologies and Their Applications in Crop Improvements. *Plant Cel. Rep.* 33, 575–583. doi:10.1007/s00299-013-1539-6

Chen, H., Chen, X., Tian, J., Yang, Y., Liu, Z., Hao, X., et al. (2016). Development of Gene-Based SSR Markers in rice Bean (*Vigna umbellata* L.) Based on Transcriptome Data. *PloS One* 11, e0151040. doi:10.1371/journal.pone.0151040

Chen, H., Liu, L., Wang, L., Wang, S., Somta, P., and Cheng, X. (2015a). Development and Validation of EST-SSR Markers from the Transcriptome of Adzuki Bean (*Vigna angularis*). *PloS One* 10, e0131939. doi:10.1371/journal.pone.0131939

Chen, H., Wang, L., Wang, S., Liu, C., Blair, M. W., and Cheng, X. (2015b). Transcriptome Sequencing of Mung Bean (*Vigna radiate* L.) Genes and the Identification of EST-SSR Markers. *PLoS ONE* 10, e0120273. doi:10.1371/journal.pone.0120273

Cheng, A., Raai, M. N., Zain, N. A. M., Massawe, F., Singh, A., and Wan-Mohtar, W. A. A. Q. I. (2019). In Search of Alternative Proteins: Unlocking the Potential of Underutilized Tropical Legumes. *Food Sec.* 11 (6), 1205–1215. doi:10.1007/s12571-019-00977-0

Cowling, W. A., Hamblin, J., Wood, P. M., and Gladstones, J. S. (1987). Resistance to Phomopsis Stem Blight in Lupinus Angustifolius L. 1. *Crop Sci.* 27, 648–652. doi:10.2135/cropsci1987.0011183x002700040007x

Cui, J., Lu, Z., Wang, T., Chen, G., Mostafa, S., Ren, H., et al. (2021). The Genome of Medicago Polymorpha Provides Insights into its Edibility and Nutritional Value as a Vegetable and Forage Legume. *Hort. Res.* 8, 1–2. doi:10.1038/s41438-021-00483-5

Cullis, C., and Kunert, K. J. (2017). Unlocking the Potential of Underutilised Legumes. *J. Expt. Bot.* 68, 1895–1903. doi:10.1093/jxb/erw437

Dachapak, S., Tomooka, N., Somta, P., Naito, K., Kaga, A., and Srinives, P. (2018). QTL Analysis of Domestication Syndrome in Zombi Pea (*Vigna vexillata*), an Underutilized Legume Crop. *PLoS ONE* 13, e0200116. doi:10.1371/journal.pone.0200116

Das, B., and Arora, S. K. (1978). 'Guar Seed–Its Chemistry and Industrial Utilization of gum.' Guar– its Improvement and Management. *Forage Res.* 4, 79–101.

Dasgupta, U., Mishra, G. P., Dikshit, H. K., Mishra, D. C., Bosamia, T., Roy, A., et al. (2021). Comparative RNA-Seq Analysis Unfolds a Complex Regulatory Network Imparting Yellow Mosaic Disease Resistance in Mungbean [*Vigna radiata* (L.) R. Wilczek]. *PLoS ONE* 16, e0244593. doi:10.1371/journal.pone.0244593

De la Rosa, L., López-Román, M. I., González, J. M., Zambrana, E., Marcos-Prado, T., and Ramírez-Parra, E. (2021). Common Vetch Valuable Germplasm for Resilient Agriculture: Genetic Characterization and Spanish Core Collection Development. *Front. Plant Sci.* 12, 617873. doi:10.3389/fpls.2021.617873

De la Rosa, L., Zambrana, E., and Ramirez-Parra, E. (2020). Molecular Bases for Drought Tolerance in Common Vetch: Designing New Molecular Breeding Tools. *BMC Plant Biol.* 20, 1–8. doi:10.1186/s12870-020-2267-z

De Vega, J. J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J. L., Ergon, Å., et al. (2015). Red clover (*Trifolium pratense* L.) Draft Genome Provides a Platform for Trait Improvement. *Sci. Reps.* 5 (1), 1–0. doi:10.1038/srep17394

DeBoer, K., Melser, S., Sperschneider, J., Kamphuis, L. G., Garg, G., Gao, L. L., et al. (2019). Identification and Profiling of Narrow-Leafed Lupin (*Lupinus angustifolius*) microRNAs during Seed Development. *BMC Genome* 20, 135. doi:10.1186/s12864-019-5521-8

Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., and Hirsch, C. N. (2021). How the Pan-Genome is Changing Crop Genomics and Improvement. *Genome Biol.* 22, 1–9. doi:10.1186/s13059-020-02224-8

Dinkins, R. D., Hancock, J., Coe, B. L., May, J. B., Goodman, J. P., Bass, W. T., et al. (2021). Isoflavone Levels, Nodulation and Gene Expression Profiles of a CRISPR/Cas9 Deletion Mutant in the Isoflavone Synthase Gene of Red clover. *Plant Cell Reps* 40, 517–528. doi:10.1007/s00299-020-02647-4

Dong, R., Dong, D., Luo, D., Zhou, Q., Chai, X., Zhang, J., et al. (2017). Transcriptome Analyses Reveal Candidate Pod Shattering-Associated Genes Involved in the Pod Ventral Sutures of Common Vetch (*Vicia sativa* L.). *Front. Plant Sci.* 8, 649. doi:10.3389/fpls.2017.00649

Drummond, C. S., Eastwood, R. J., Miotto, S. T., and Hughes, C. E. (2012). Multiple continental Radiations and Correlates of Diversification in Lupinus (Leguminosae): Testing for Key Innovation with Incomplete Taxon Sampling. *Syst. Biol.* 61, 443–460. doi:10.1093/sysbio/syr126

Ditzler, L., van Apeldoorn, D. F., Pellegrini, F., Antichi, D., Bàrberi, P., and Rossing, W. A. H. (2021). Current Research on the Ecosystem Service Potential of Legume Inclusive Cropping Systems in Europe. A Review. *Agron. Sustain. Dev.* 41, 26. doi:10.1007/s13593-021-00678-z

Ebi, K. L., and Loladze, I. (2019). Elevated Atmospheric CO2 Concentrations and Climate Change Will Affect Our Food's Quality and Quantity. *Lancet Planet. Health* 3, e283–e284. doi:10.1016/s2542-5196(19)30108-1

Edwards, D., and Batley, J. (2010). Plant Genome Sequencing: Applications for Crop Improvement. *Plant Biotechnol. J.* 8, 2–9. doi:10.1111/j.1467-7652.2009.00459.x

Elfaki, M. O., and Abdelatti, K. A. (2018). Rumen Content as Animal Feed: a Review. *J. Vet. Med. Anim. Prod.* 20 (2), 7.

Emmrich, P. M., Sarkar, A., Njaci, I., Kaithakottil, G. G., Ellis, N., Moore, C., et al. (2020). A Draft Genome of Grass Pea (*Lathyrus sativus)*, a Resilient Diploid Legume. *bioRxiv*. doi:10.1101/2020.04.24.058164

Etheridge, J. V., and Bateman, G. L. (1999). Fungicidal Control of Foliar Diseases of White Lupin (Lupinus Albus). *Crop Prot.* 18, 349–354. doi:10.1016/s0261-2194(99)00031-9

Fang, Y., Wang, L., Sapey, E., Fu, S., Wu, T., Zeng, H., et al. (2021). Speed-Breeding System in Soybean: Integrating Off-Site Generation Advancement, Fresh Seeding, and Marker-Assisted Selection. *Front. Plant Sci.* 12, 717077. doi:10.3389/fpls.2021.717077

FAO (2013). FAOStat 2013 Food and Agriculture Organization of the United Nations Statistics Division.

FAO (2018). Definitions and Standards Used in FAOSTAT. Available online: http://www.fao.org/faostat/en/#data/QC (Accessed on August 18, 2018).

FAO (2017). FAOStat 2017 Food and Agriculture Organization of the United Nations Statistics Division.

Feng, X., Orellana, G. E., Green, J. C., Melzer, M. J., Hu, J. S., and Karasev, A. V. (2019). A New Strain of Bean Common Mosaic Virus from lima Bean (Phaseolus Lunatus): Biological and Molecular Characterization. *Plant Dis.* 103, 1220–1227. doi:10.1094/pdis-08-18-1307-re

Fernie, A. R., and Yan, J. (2019). De Novo domestication: an Alternative Route toward New Crops for the Future. *Mol. Plant* 12, 615–631. doi:10.1016/j.molp.2019.03.016

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. doi:10.1038/nature10452

Fontanari, G. G., Batistuti, J. P., da Cruz, R. J., Saldiva, P. H., and Arêas, J. A. (2012). Cholesterol-lowering Effect of Whole Lupin (Lupinus Albus) Seed and its Protein Isolate. *Food Chem.* 132, 1521–1526. doi:10.1016/j.foodchem.2011.11.145

Francis, C. M., Enneking, D., and Abd El Moneim, A. M. (2000). "When and Where Will Vetches Have an Impact as Grain Legumes?," in *Linking Research and Marketing Opportunities for Pulses in the 21st Century 2000* (Dordrecht: Springer), 375–384. doi:10.1007/978-94-011-4385-1_34

Frick, K. M., Kamphuis, L. G., Siddique, K. H., Singh, K. B., and Foley, R. C. (2017). Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement. *Front. Plant Sci.* 8, 87. doi:10.3389/fpls.2017.00087

Fuller, D. Q. (2007). Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. *Ann. Bot. (Lond.)* 2007 (100), 903–924. doi:10.1093/aob/mcm048

Gaikwad, K., Ramakrishna, G., Srivastava, H., Saxena, S., Kaila, T., Tyagi, A., et al. (2020). Chromosome Scale Reference Genome of Cluster Bean (*Cyamopsis tetragonoloba* (L.) Taub.). *bioRxiv*. 2020 Jan 1. doi:10.1101/2020.05.16.098434

Garcia, T., Duitama, J., Zullo, S. S., Gil, J., Ariani, A., Dohle, S., et al. (2021). Comprehensive Genomic Resources Related to Domestication and Crop Improvement Traits in Lima Bean. *Nat. Commun.* 12, 1–7. doi:10.1038/s41467-021-20921-1

Gasparini, K., dos Reis Moreira, J., Peres, L. E., and Zsögön, A. (2021). De Novo domestication of Wild Species to Create Crops with Increased Resilience and Nutritional Value. *Curr. Opin. Plant Biol.* 60, 102006. doi:10.1016/j.pbi.2021.102006

Gautam, N. K., Kumar, K., and Prasad, M. (2016). Leaf Crinkle Disease in Urdbean (*Vigna mungo* L. Hepper): An Overview on Causal Agent, Vector and Host. *Protoplasma* 253, 729–746. doi:10.1007/s00709-015-0933-z

Gepts, P., Beavis, W. D., Brummer, E. C., Shoemaker, R. C., Stalker, H. T., Weeden, N. F., et al. (2005). Legumes as a Model Plant Family. Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. *Plant Physiol.* 137 (4), 1228–1235. doi:10.1104/pp.105.060871

Gillett, J. B. (1958). *Indigofera (Microcharis) in Tropical Africa with the Related Genera Cyamopsis and Rhynchotropis*, 1–166. H.M.S.O Kew Bull.

Gladstones, J. S. (1998). in *Lupins as Crop Plants: Biology, Production and Utilization*. Editors J. S. Gladstones, C. Atkins, and J. Hamblin (Oxon, New York: CAB International), 1–39.

Glazińska, P., Kulasek, M., Glinkowski, W., Wojciechowski, W., and Kosiński, J. (2019). Integrated Analysis of Small RNA, Transcriptome and Degradome Sequencing Provides New Insights into floral Development and Abscission in Yellow Lupine (*Lupinus luteus* L.). *Intl. J. Mol. Sci.* 20, 5122. doi:10.3390/ijms20205122

Glazinska, P., Wojciechowski, W., Kulasek, M., Glinkowski, W., Marciniak, K., Klajn, N., et al. (2017). De Novo transcriptome Profiling of Flowers, Flower Pedicels and Pods of Lupinus Luteus (Yellow Lupine) Reveals Complex Expression Changes during Organ Abscission. *Front. Plant Sci.* 8, 641. doi:10.3389/fpls.2017.00641

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science* 327, 812–818. doi:10.1126/science.1185383

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

Gresta, F. A., Wink, M. I., Prins, U., Abberton, M. I., Capraro, J. E., Scarafoni, A. L., et al. (2017). "Lupins in European Cropping Systems," in *Legumes in Cropping Systems*, 88–108. doi:10.1079/9781780644981.0088

Grigoreva, E., Ulianich, P., Ben, C., Gentzbittel, L., and Potokina, E. (2019). First Insights into the Guar (*Cyamopsis tetragonoloba* (L.) Taub.) Genome of the 'Vavilovskij 130'Accession, Using Second and Third-Generation Sequencing Technologies. *Russ. J. Genet.* 55, 1406–1416. doi:10.1134/s102279541911005x

Ha, J., Satyawan, D., Jeong, H., Lee, E., Cho, K. H., Kim, M. Y., et al. (2021). A Near-complete Genome Sequence of Mungbean (*Vigna radiata* L.) Provides Key Insights into the Modern Breeding Program. *Plant Genome* 18, e20121. doi:10.1002/tpg2.20121

Han, J., Guo, B., Guo, Y., Zhang, B., Wang, X., and Qiu, L. J. (2019). Creation of Early Flowering Germplasm of Soybean by CRISPR/Cas9 Technology. *Front. Plant Sci.* 10, 1446. doi:10.3389/fpls.2019.01446

Hane, J. K., Ming, Y., Kamphuis, L. G., Nelson, M. N., Garg, G., Atkins, C. A., et al. (2017). A Comprehensive Draft Genome Sequence for Lupin (*Lupinus angustifolius*), an Emerging Health Food: Insights into Plant–Microbe Interactions and Legume Evolution. *Plant Biotechnol. J.* 15, 318–330. doi:10.1111/pbi.12615

Hao, X., Yang, T., Liu, R., Hu, J., Yao, Y., Burlyaeva, M., et al. (2017). An RNA Sequencing Transcriptome Analysis of Grasspea (*Lathyrus sativus* L.) and Development of SSR and KASP Markers. *Front. Plant Sci.* 8, 1873. doi:10.3389/fpls.2017.01873

Hepper, F. N. (1963). The Bambara Groundnut (*Voandzeia subterranea*) and Kersting's Groundnut (Kerstingiella geocarpa) Wild in West Africa. *Kew Bull.* 16, 395–407. doi:10.2307/4114681

Herbert, D. B., Gross, T., Rupp, O., and Becker, A. (2021). Transcriptome Analysis Reveals Major Transcriptional Changes during Regrowth after Mowing of Red clover (*Trifolium pratense*). *BMC Plant Biol.* 21, 1–20. doi:10.1186/s12870-021-02867-0

Hickey, L. T., Hafeez, N. A., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding Crops to Feed 10 Billion. *Nat. Biotechnol.* 37, 744–754. doi:10.1038/s41587-019-0152-9

Hirakawa, H., Kaur, P., Shirasawa, K., Nichols, P., Nagano, S., Appels, R., et al. (2016). Draft Genome Sequence of Subterranean Clover, a Reference for Genus Trifolium. *Sci. Reps.* 22, 1–9. doi:10.1038/srep30358

Hillocks, R. J., Bennett, C., and Mponda, O. M. (2012). BambaraNut:A Review of Utilisation, Market Potential and Crop Improvement. *Afr. Crop Sci. J.* 20, 1–16.

Ho, W. K., Chai, H. H., Kendabie, P., Ahmad, N. S., Jani, J., Massawe, F., et al. (2017). Integrating Genetic Maps in Bambara Groundnut [*Vigna subterranea* (L) Verdc.] and Their Syntenic Relationships Among Closely Related Legumes. *BMC Genome* 18, 1–9. doi:10.1186/s12864-016-3393-8

Hu, H., Wang, H., Zhang, Y., Kan, B., Ding, Y., and Huang, J. (2019). Characterization of Genes in Guar Gum Biosynthesis Based on Quantitative RNA-Sequencing in Guar Bean (*Cyamopsis tetragonoloba*). *Sci. Reps.* 9, 1–0. doi:10.1038/s41598-019-47518-5

Hu, Z., Sun, C., Lu, K. C., Chu, X., Zhao, Y., Lu, J., et al. (2017). EUPAN Enables Pan-Genome Studies of a Large Number of Eukaryotic Genomes. *Bioinformatics* 33, 2408–2409. doi:10.1093/bioinformatics/btx170

Huang, X., and Han, B. (2014). Natural Variations and Genome-wide Association Studies in Crop Plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi:10.1146/annurev-arplant-050213-035715

Hufnagel, B., Marques, A., Soriano, A., Marquès, L., Divol, F., Doumas, P., et al. (2020). High-quality Genome Sequence of white Lupin Provides Insight into Soil Exploration and Seed Quality. *Nat. Commun.* 11, 1–2. doi:10.1038/s41467-019-14197-9

Hufnagel, B., Soriano, A., Taylor, J., Divol, F., Kroc, M., Sanders, H., et al. (2021). Pangenome of white Lupin Provides Insights into the Diversity of the Species. *bioRxiv*. 2021 Jan 1. doi:10.1111/pbi.13678

Hufnagel, J., Reckling, M., and Ewert, F. (2020). Diverse Approaches to Crop Diversification in Agricultural Research. A Review. *Agron. Sust Dev.* 40, 1–17. doi:10.1007/s13593-020-00617-4

Huyghe, C. (1997). White Lupin (*Lupinus albus* L.). *Field Crops Res.* 53, 147–160. doi:10.1016/s0378-4290(97)00028-2

Hymowitz, T., and Boyd, J. (1977). Origin, Ethnobotany and Agricultural Potential of the Winged Bean: Psophocarpus Tetragonolobus. *Econ. Bot.* 31, 180–188. doi:10.1007/bf02866589

Iizuka, N. (1990). Studies on Virus Diseases of Adzuki Bean (Vigna Angularis Wight) in Japan. *Bull. Tohoku Natl. Agric. Exp. Station* 82, 77–113.

Iqbal, M. M., Erskine, W., Berger, J. D., and Nelson, M. N. (2020). Phenotypic Characterisation and Linkage Mapping of Domestication Syndrome Traits in Yellow Lupin (*Lupinus luteus* L.). *Theor. Appl. Genet.* 133, 2975–2987. doi:10.1007/s00122-020-03650-9

Iqbal, M. M., Huynh, M., Udall, J. A., Kilian, A., Adhikari, K. N., Berger, J. D., et al. (2019). The First Genetic Map for Yellow Lupin Enables Genetic Dissection of Adaptation Traits in an Underutilized Grain Legume Crop. *BMC Genet.* 20, 68. doi:10.1186/s12863-019-0767-3

Isemura, T., Kaga, A., Tomooka, N., Shimizu, T., and Vaughan, D. A. (2010). Construction of a Genetic Linkage Map and Genetic Analysis of the Domestication Related Traits in rice Bean (*Vigna umbellata*). *Ann. Bot.* 106, 927–944. doi:10.1093/aob/mcq188

Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construction of a Genetic Linkage Map and Genetic Analysis of Domestication Related Traits in Mungbean (*Vigna radiata*). *PLoS One* 7, e41304. doi:10.1371/journal.pone.0041304

Isobe, S., Kölliker, R., Hisano, H., Sasamoto, S., Wada, T., Klimenko, I., et al. (2009). Construction of a Consensus Linkage Map for Red Clover (*Trifolium pratense* L.). *BMC Plant Biol.* 9 (1), 1–11.

Ištvánek, J., Dluhošová, J., Dluhoš, P., Pátková, L., Nedělník, J., and Řepková, J. (2017). Gene Classification and Mining of Molecular Markers Useful in Red clover (*Trifolium pratense*) Breeding. *Front. Plant Sci.* 8, 367. doi:10.3389/fpls.2017.00367

Ištvánek, J., Jaroš, M., Křenek, A.,, and Řepková, J. (2014). Genome Assembly and Annotation for Red clover (Trifolium Pratense; Fabaceae). *Am. J. Bot.* 101, 327–337.

Jasrotia, R. S., Yadav, P. K., Iquebal, M. A., Bhatt, S. B., Arora, V., Angadi, U. B., et al. (2019). VigSatDB: Genome-wide Microsatellite DNA Marker Database of Three Species of Vigna for Germplasm Characterization and Improvement. *Database*. 2019 Jan 1. doi:10.1093/database/baz055

Jegadeesan, S., Raizada, A., Dhanasekar, P., and Suprasanna, P. (2021). Draft Genome Sequence of the Pulse Crop Blackgram [*Vigna mungo* (L.) Hepper] Reveals Potential R-Genes. *Sci. Reps.* 11, 1–0. doi:10.1038/s41598-021-90683-9

Jha, U. C. (2018). Current Advances in Chickpea Genomics: Applications and Future Perspectives. *Plant Cell Reps* 37, 947–965. doi:10.1007/s00299-018-2305-6

Ji, J., Zhang, C., Sun, Z., Wang, L., Duanmu, D., and Fan, Q. (2019). Genome Editing in Cowpea Vigna Unguiculata Using CRISPR-Cas9. *Int. J. Mol. Sci.* 20 (10), 2471. doi:10.3390/ijms20102471

Jo, Y., Lian, S., Cho, J. K., Choi, H., Kim, S. M., Kim, S. L., et al. (2016). De Novo transcriptome Assembly of Two Vigna Angularis Varieties Collected from Korea. *Genomics Data* 8, 119–120. doi:10.1016/j.gdata.2016.05.004

Jones, C., De Vega, J., Lloyd, D., Hegarty, M., Ayling, S., Powell, W., et al. (2020). Population Structure and Genetic Diversity in Red clover (*Trifolium pratense* L.) Germplasm. *Sci. Reps.* 10, 1–2. doi:10.1038/s41598-020-64989-z

Jones, D. B., Gersdorff, C. E., Johns, C. O., and Finks, A. J. (1922). The Proteins of the Lima Bean, Phaseolus Lunatus. *J. Biol. Chem.* 53, 231–240. doi:10.1016/s0021-9258(18)85778-7

Juranić, M., Nagahatenna, D. S., Salinas-Gamboa, R., Hand, M. L., Sánchez-León, N., Leong, W. H., et al. (2020). A Detached Leaf Assay for Testing Transient Gene Expression and Gene Editing in Cowpea (*Vigna unguiculata* [L.] Walp.). *Plant Methods* 16 (1), 1–7. doi:10.1186/s13007-020-00630-4

Kadam, S. S., Salunkhe, D. K., and Luh, B. S. (1984). Winged Bean in Human Nutrition. *Crit. Rev. Food Sci. Nutr.* 21, 1–40. doi:10.1080/10408398409527395

Kaewwongwal, A., Kongjaimun, A., Somta, P., Chankaew, S., Yimram, T., and Srinives, P. (2015). Genetic Diversity of the Black Gram [*Vigna mungo* (L.) Hepper] Gene Pool as Revealed by SSR Markers. *Breed. Sci.* 65, 127–137. doi:10.1270/jsbbs.65.127

Kaga, A., Isemura, T., Tomooka, N., and Vaughan, D. A. (2008). The Genetics of Domestication of the Azuki Bean (*Vigna angularis*). *Genetics* 178, 1013–1036. doi:10.1534/genetics.107.078451

Kakati, P., Deka, S., Kotoki, D., and Saikia, S. (2010). Effect of Traditional Methods of Processing on the Nutrient Contents and Some Antinu- Tritional Factors in Newly Developed Cultivars of green Gram [*Vigna radiata* (L.) Wilezek] and Black Gram [Vigna Mungo (L.) Hepper] of Assam. India. *Int. Food Res. J.* 17, 377–384.

Kamenya, S. N., Mikwa, E. O., Song, B., and Odeny, D. A. (2021). Genetics and Breeding for Climate Change in Underutilised Crops. *Theor. Appl. Genet.* 23, 1–29.

Kamphuis, L. G., Hane, J. K., Nelson, M. N., Gao, L., Atkins, C. A., and Singh, K. B. (2015). Transcriptome Sequencing of Different Narrow-leafed Lupin Tissue Types Provides a Comprehensive Uni-gene Assembly and Extensive Gene-based Molecular Markers. *Plant Biotechnol. J.* 13, 14–25. doi:10.1111/pbi.12229

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome Sequence of Mungbean and Insights into Evolution within Vigna Species. *Nat. Commun.* 5, 1–9. doi:10.1038/ncomms6443

Kang, Y. J., Satyawan, D., Shim, S., Lee, T., Lee, J., Hwang, W. J., et al. (2015). Draft Genome Sequence of Adzuki Bean, Vigna Angularis. *Sci. Reps.* 5, 1–8. doi:10.1038/srep08069

Kasprzak, A., Šafář, J., Janda, J., Doležel, J., Wolko, B., and Naganowska, B. (2006). The Bacterial Artificial Chromosome (BAC) Library of the Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Cell. Mol. Biol. Lett.* 11, 396–407. doi:10.2478/s11658-006-0033-3

Katoch, R. (2020). "Common Diseases and Insect–Pests of Rice Bean," in *Ricebean* (Singapore: Springer), 347–353. doi:10.1007/978-981-15-5293-9_18

Kaul, T., Eswaran, M., Thangaraj, A., Meyyazhagan, A., Nehra, M., Raman, N. M., et al. (2019). Rice Bean (*Vigna umbellata*) Draft Genome Sequence: Unravelling the Late Flowering and Unpalatability Related Genomic Resources for Efficient Domestication of This Underutilized Crop. *bioRxiv*. 2019 Jan 1. doi:10.1101/816595

Kaur, P., Bayer, P. E., Milec, Z., Vrána, J., Yuan, Y., Appels, R., et al. (2017). An Advanced Reference Genome of Trifolium Subterraneum L. Reveals Genes Related to Agronomic Performance. *Plant Biotechnol. J.* 15, 1034–1046. doi:10.1111/pbi.12697

Keatinge, J. D., Easdown, W. J., Yang, R. Y., Chadha, M. L., and Shanmugasundaram, S. (2011). Overcoming Chronic Malnutrition in a Future Warming World: the Key Importance of Mungbean and Vegetable Soybean. *Euphytica* 180, 129–141. doi:10.1007/s10681-011-0401-6

Kim, T. S., Raveendar, S., Suresh, S., Lee, G. A., Lee, J. R., Cho, J. H., et al. (2015). Transcriptome Analysis of Two Vicia Sativa Subspecies: Mining Molecular Markers to Enhance Genomic Resources for Vetch Improvement. *Genes* 6 (4), 1164–1182. doi:10.3390/genes6041164

Kislev, M. E. (1989). Origins of the Cultivation of Lathyrus Sativus and L. Cicera (Fabaceae). *Econ. Bot.* 43, 262–270. doi:10.1007/bf02859868

Kitano-Okada, T., Ito, A., Koide, A., Nakamura, Y., Han, K. H., Shimada, K., et al. (2012). Anti-obesity Role of Adzuki Bean Extract Containing Polyphenols: *In Vivo* and *In Vitro* Effects. *J. Sci. Food Agric.* 92 (13), 2644–2651. doi:10.1002/jsfa.5680

Konduri, V., Godwin, I. D., and Liu, C. J. (2000). Genetic Mapping of the *Lablab Purpureus* Genome Suggests the Presence of "Cuckoo" Gene(s) in This Species. *Theor. Appl. Genet.* 100, 866–871. doi:10.1007/s001220051363

Korte, A., and Farlow, A. (2013). The Advantages and Limitations of Trait Analysis with GWAS: a Review. *Plant Methods* 9, 29. doi:10.1186/1746-4811-9-29

Kovi, M. R., Amdahl, H., Alsheikh, M., and Rognli, O. A. (2017). De novo and Reference Transcriptome Assembly of Transcripts Expressed During Flowering Provide Insight into Seed Setting in Tetraploid Red Clover. *Sci. Rep.* 7 (1), 1–11.

Kozak, B., Galek, R., Zalewski, D., and Sawicka-Sienkiewicz, E. (2019). Preliminary Genetic Map of a New Recombinant Inbred Line Population for Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Agronomy* 9, 653. doi:10.3390/agronomy9100653

Kroc, M., Rybiński, W., Wilczura, P., Kamel, K., Kaczmarek, Z., Barzyk, P., et al. (2017). Quantitative and Qualitative Analysis of Alkaloids Composition in the Seeds of a white Lupin (*Lupinus albus* L.) Collection. *Genet. Resour. Crop Evol.* 64, 1853–1860. doi:10.1007/s10722-016-0473-1

Kroc, M., Czepiel, K., Wilczura, P., Mokrzycka, M., and Święcicki, W. (2019). Development and Validation of a Gene-Targeted dCAPS Marker for Marker-Assisted Selection of Low-Alkaloid Content in Seeds of Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Genes* 10, 428. doi:10.3390/genes10060428

Książkiewicz, M., Nazzicari, N., Nelson, M. N., Renshaw, D., Rychel, S., Ferrari, B., et al. (2017). A High-Density Consensus Linkage Map of white Lupin Highlights Synteny with Narrow-Leafed Lupin and Provides Markers Tagging Key Agronomic Traits. *Sci. Reps.* 7, 1–5.

Kumar, S., Bejiga, G., Ahmed, S., Nakkoul, H., and Sarker, A. (2011). Genetic Improvement of Grass Pea for Low Neurotoxin (β-ODAP) Content. *Food Chem. Toxicol.* 49, 589–600. doi:10.1016/j.fct.2010.06.051

Kumar, S., Palve, A. S., Patel, S. K., Selvanayagam, S., Sharma, R., and Rathore, A. (2020). Development of Genomic Microsatellite Markers in Cluster Bean Using Next-Generation DNA Sequencing and Their Utility in Diversity Analysis. *Curr. Plant Biol.* 21, 100134. doi:10.1016/j.cpb.2019.100134

Lambers, H., Clements, J. C., and Nelson, M. N. (2013). How a Phosphorus-acquisition Strategy Based on Carboxylate Exudation powers the success and Agronomic Potential of Lupines (Lupinus, Fabaceae). *Am. J. Bot.* 100, 263–288. doi:10.3732/ajb.1200474

Lee, G. A. (2013). Archaeological Perspectives on the Origins of Azuki (Vigna Angularis). *Holocene* 23, 453–459. doi:10.1177/0959683612460788

Lee, Y. P., Mori, T. A., Sipsas, S., Barden, A., Puddey, I. B., Burke, V., et al. (2006). Lupin-enriched Bread Increases Satiety and Reduces Energy Intake Acutely. *Amer J. Clinic. Nutr.* 84, 975–980. doi:10.1093/ajcn/84.5.975

Lei, L., Goltsman, E., Goodstein, D., Wu, G. A., Rokhsar, D. S., and Vogel, J. P. (2021). Plant Pan-Genomics Comes of Age. *Annu. Rev. Plant Biol.* 72, 411–435. doi:10.1146/annurev-arplant-080720-105454

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of Extreme Weather Disasters on Global Crop Production. *Nature* 529, 84–87. doi:10.1038/nature16467

Li, T., Yang, X., Yu, Y., Si, X., Zhai, X., Zhang, H., et al. (2018). Domestication of Wild Tomato Is Accelerated by Genome Editing. *Nat. Biotechnol.* 36, 1160–1163. doi:10.1038/nbt.4273

Li, W., Riday, H., Riehle, C., Edwards, A., and Dinkins, R. (2019). Identification of Single Nucleotide Polymorphism in Red clover (*Trifolium pratense* L.) Using Targeted Genomic Amplicon Sequencing and RNA-Seq. *Front. Plant Sci.* 10, 1257. doi:10.3389/fpls.2019.01257

Li, X., Yan, G., Renshaw, D., and Yang, H. (2010). Development of a Co-dominant DNA Marker Tightly Linked to Gene Tardus Conferring Reduced Pod Shattering in Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Euphytica* 176, 49–58. doi:10.1007/s10681-010-0212-1

Li, X., Yang, H., and Yan, G. (2012). Development of a Co-dominant DNA Marker Linked to the Gene Lentus Conferring Reduced Pod Shattering for Marker-Assisted Selection in Narrow-Leafed Lupin (*Lupinus angustifolius*) Breeding. *Plant Breed* 131, 540–544. doi:10.1111/j.1439-0523.2012.01978.x

Li, X., Yang, H., Buirchell, B., and Yan, G. (2011). Development of a DNA Marker Tightly Linked to Low-Alkaloid Gene Iucundus in Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Crop Pasture Sci.* 62, 218–224. doi:10.1071/cp10352

Lichtin, N., Salvo-Garrido, H., Till, B., Caligari, P. D., Rupayan, A., Westermeyer, F., et al. (2020). Genetic and Comparative Mapping of Lupinus Luteus L. Highlight Syntenic Regions with Major Orthologous Genes Controlling Anthracnose Resistance and Flowering Time. *Sci. Reps* 10, 1–5. doi:10.1038/s41598-020-76197-w

Lin, R., Renshaw, D., Luckett, D., Clements, J., Yan, G., Adhikari, K., et al. (2009). Development of a Sequence-specific PCR Marker Linked to the Gene "Pauper" Conferring Low-Alkaloids in white Lupin (*Lupinus albus* L.) for Marker Assisted Selection. *Mol. Breed.* 23, 153–161. doi:10.1007/s11032-008-9222-2

Liu, C., Fan, B., Cao, Z., Su, Q., Wang, Y., Zhang, Z., et al. (2016). Development of a High-Density Genetic Linkage Map and Identification of Flowering Time QTLs in Adzuki Bean (*Vigna angularis*). *Sci. Reps.* 6, 1–0. doi:10.1038/srep39523

Liu, C., Fan, B., Cao, Z., Su, Q., Wang, Y. A., Zhang, Z., et al. (2016). A Deep Sequencing Analysis of Transcriptomes and the Development of EST-SSR Markers in Mungbean (*Vigna radiata*). *J. Genet.* 95, 527–535. doi:10.1007/s12041-016-0663-9

Liu, C., Wu, J., Wang, L., Fan, B., Cao, Z., Su, Q., et al. (2017). Quantitative Trait Locus Mapping under Irrigated and Drought Treatments Based on a Novel Genetic Linkage Map in Mungbean (*Vigna radiata* L.). *Theor. Appl. Genet.* 130, 2375–2393. doi:10.1007/s00122-017-2965-6

Liu, H. J., and Yan, J. (2019). Crop Genome-wide Association Study: a Harvest of Biological Relevance. *Plant J.* 97, 8–18. doi:10.1111/tpj.14139

Liu, L., Bestel, S., Shi, J., Song, Y., and Chen, X. (2013). Paleolithic Human Exploitation of Plant Foods during the Last Glacial Maximum in North China. *Proc. Natl. Acad. Sci. USA* 110, 5380–5385. doi:10.1073/pnas.1217864110

Liu, M. S., Kuo, T. C., Ko, C. Y., Wu, D. C., Li, K. Y., Lin, W. J., et al. (2016). Genomic and Transcriptomic Comparison of Nucleotide Variations for Insights into Bruchid Resistance of Mungbean (*Vigna radiata* [L.] R. Wilczek). *BMC Plant Biol.* 16, 1–6. doi:10.1186/s12870-016-0736-1

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. doi:10.1111/tpj.14349

Maass, B. L., Knox, M. R., Venkatesha, S. C., Angessa, T. T., Ramme, S., and Pengelly, B. C. (2010). *Lablab Purpureus*–a Crop Lost for Africa? *Trop. Plant Biol.* 3 (3), 123–135. doi:10.1007/s12042-010-9046-1

Mahesh, H. B., Prasannakumar, M. K., Manasa, K. G., Perumal, S., Khedikar, Y., Kagale, S., et al. (2021). Genome, Transcriptome, and Germplasm Sequencing Uncovers Functional Variation in the Warm-Season Grain Legume Horsegram Macrotyloma Uni Orum (Lam.) Verdc. *Front. Plant Sci.* 12, 758119. doi:10.3389/fpls.2021.758119

Marques, E., Kur, A., Bueno, E., and von Wettberg, E. (2020). Defining and Improving the Rotational and Intercropping Value of a Crop Using a Plant–Soil Feedbacks Approach. *Crop Sci.* 60 (5), 2195–2203. doi:10.1002/csc2.20200

Mathivathana, M. K., Murukarthick, J., Karthikeyan, A., Jang, W., Dhasarathan, M., Jagadeeshselvam, N., et al. (2019). Detection of QTLs Associated with Mungbean Yellow Mosaic Virus (MYMV) Resistance Using the Interspecific Cross of *Vigna radiata* × *Vigna umbellata*. *J. Appl. Genet.* 60, 255–268. doi:10.1007/s13353-019-00506-x

Mayes, S., Ho, W. K., Chai, H. H., Gao, X., Kundy, A. C., Mateva, K. I., et al. (2019). Bambara Groundnut: An Exemplar Underutilised Legume for Resilience under Climate Change. *Planta* 250, 803–820. doi:10.1007/s00425-019-03191-6

McGuire, W. S. (1985). Subterranean clover. *Clover Sci. Technol.* 25, 515–534.

Mercado-Ruaro, P., and Delgado-Salinas, A. (1998). Karyotypic Studies on Species of *Phaseolus* (Fabaceae: Phaseolinae). *Am. J. Bot.* 85, 1–9. doi:10.2307/2446547

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. doi:10.1093/genetics/157.4.1819

Michno, J. M., Wang, X., Liu, J., Curtin, S. J., Kono, T. J., and Stupar, R. M. (2015). CRISPR/Cas Mutagenesis of Soybean and Medicago Truncatula Using a New Web-Tool and a Modified Cas9 Enzyme. *GM crops & Food* 6, 243–252. doi:10.1080/21645698.2015.1106063

Mishra, S. K., Macedo, M. L. R., Panda, S. K., and Panigrahi, J. (2018). Bruchid Pest Management in Pulses: Past Practices, Present Status and Use of Modern Breeding Tools for Development of Resistant Varieties. *Ann. Appl. Biol.* 172, 4–19. doi:10.1111/aab.12401

Mobini, S. H., and Warkentin, T. D. (2016). A Simple and Eficient Method of *In Vivo* Rapid Generation Technology in Pea (*Pisum sativum* L.) *In Vitro*. *Cell Dev-Plant* 52, 530–536. doi:10.1007/s11627-016-9772-7

Moghaddam, S. M., Oladzad, A., Koh, C., Ramsay, L., Hart, J. P., Mamidi, S., et al. (2021). The Tepary Bean Genome Provides Insight into Evolution and Domestication under Heat Stress. *Nat. Commun.* 12 (1), 1–4. doi:10.1038/s41467-021-22858-x

Mousavi-Derazmahalleh, M., Bayer, P. E., Nevado, B., Hurgobin, B., Filatov, D., Kilian, A., et al. (2018). Exploring the Genetic and Adaptive Diversity of a Pan-Mediterranean Crop Wild Relative: Narrow-Leafed Lupin. *Theor. Appl. Genet.* 131, 887–901. doi:10.1007/s00122-017-3045-7

Nair, R. M., Schafleitner, R., Kenyon, L., Srinivasan, R., Easdown, W., Ebert, A. W., et al. (2012). Genetic Improvement of Mungbean. *SABRAO J. Breed. Genet.* 44, 177–190.

Nair, R. M., Pandey, A. K., War, A. R., Hanumantharao, B., Shwe, T., Alam, A. K. M. M., et al. (2019). Biotic and Abiotic Constraints in Mungbean Production—Progress in Genetic Improvement. *Front. Plant Sci.* 10, 1340. doi:10.3389/fpls.2019.01340

Nasti, R. A., and Voytas, D. F. (2021). Attaining the Promise of Plant Gene Editing at Scale. *Proc. Nat. Acad. Sci.* 118 (22), e2004846117. doi:10.1073/pnas.2004846117

Nelson, M. N., Książkiewicz, M., Rychel, S., Besharat, N., Taylor, C. M., Wyrwa, K., et al. (2017). The Loss of Vernalization Requirement in Narrow-leafed Lupin Is Associated with a Deletion in the Promoter and De-repressed Expression of a Flowering Locus T (FT) Homologue. *New Phytol.* 213, 220–232. doi:10.1111/nph.14094

Nguyen, V., Riley, S., Nagel, S., Fisk, I., and Searle, I. R. (2020). Common Vetch: A Drought Tolerant, High Protein Neglected Leguminous Crop with Potential as a Sustainable Food Source. *Front. Plant Sci.* 11, 818. doi:10.3389/fpls.2020.00818

Nichols, P. G., Revell, C. K., Humphries, A. W., Howie, J. H., Hall, E. J., Sandral, G. A., et al. (2012). Temperate Pasture Legumes in Australia—their History, Current Use, and Future Prospects. *Crop Pasture Sci.* 63, 691–725. doi:10.1071/cp12194

O'Rourke, J. A., Fu, F., Bucciarelli, B., Yang, S. S., Samac, D. A., Lamb, J. F., et al. (2015). The Medicago Sativa Gene index 1.2: a Web-Accessible Gene Expression Atlas for Investigating Expression Differences between Medicago Sativa Subspecies. *BMC Genom* 16, 1–7. doi:10.1186/s12864-015-1718-7

O'Rourke, J. A., Yang, S. S., Miller, S. S., Bucciarelli, B., Liu, J., Rydeen, A., et al. (2013). An RNA-Seq Transcriptome Analysis of Orthophosphate-Deficient white Lupin Reveals Novel Insights into Phosphorus Acclimation in Plants. *Plant Physiol.* 161, 705–724. doi:10.1104/pp.112.209254

Ogutcen, E., Pandey, A., Khan, M. K., Marques, E., Penmetsa, R. V., Kahraman, A., et al. (2018). Pod Shattering: a Homologous Series of Variation Underlying Domestication and an Avenue for Crop Improvement. *Agronomy* 8, 137. doi:10.3390/agronomy8080137

Olukolu, B. A., Mayes, S., Stadler, F., Ng, N. Q., Fawole, I., Dominique, D., et al. (2012). Genetic Diversity in Bambara Groundnut (*Vigna subterranea* (L.) Verdc.) as Revealed by Phenotypic Descriptors and DArT Marker Analysis. *Genet. Resour. Crop Evol.* 59, 347–358. doi:10.1007/s10722-011-9686-5

Parida, A., Raina, S. N., and Narayan, R. K. J. (1990). Quantitative DNA Variation between and within Chromosome Complements of Vigna Species (Fabaceae). *Genetica* 82, 125–133. doi:10.1007/bf00124642

Parimala, K., Meenakumari, K. V. S., Sudhakar, R., and Durga, K. K. (2011). Screening of Horsegram Genotypes against Yellow Mosaic Virus and Powdery Mildew Diseases. Indian. *J. Plant Prot.* 39, 160–161.

Patel, D. P., Dabas, B. S., Sapra, R. S., and Mandal, S. (1995). *Evaluation of Horsegram (Macrotyloma uniflorum) (Lam.) Germplasm*. New Delhi, India: National Bureau of Plant Genetic Resources Publication.

Plewiński, P., Ćwiek-Kupczyńska, H., Rudy, E., Bielski, W., Rychel-Bielska, S., Stawiński, S., et al. (2020). Innovative Transcriptome-based Genotyping Highlights Environmentally Responsive Genes for Phenology, Growth and Yield in a Non-model Grain Legume. *Plant Cel. Environ.* 43, 2680–2698. doi:10.1111/pce.13880

Plewiński, P., Książkiewicz, M., Rychel-Bielska, S., Rudy, E., and Wolko, B. (2019). Candidate Domestication-Related Genes Revealed by Expression Quantitative Trait Loci Mapping of Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Int. J. Mol. Sci.* 20 (22), 5670. doi:10.3390/ijms20225670

Pootakham, W., Nawae, W., Naktang, C., Sonthirod, C., Yoocha, T., Kongkachana, W., et al. (2021). A Chromosome-scale Assembly of the Black Gram (*Vigna mungo*) Genome. *Mol. Ecol. Resour.* 21, 238–250. doi:10.1111/1755-0998.13243

Porch, T. G., Beaver, J. S., Debouck, D. G., Jackson, S. A., Kelly, J. D., and Dempewolf, H. (2013). Use of Wild Relatives and Closely Related Species to Adapt Common Bean to Climate Change. *Agronomy* 3, 433–461. doi:10.3390/agronomy3020433

Prasad, S. K., and Singh, M. K. (2015). Horse Gram-An Underutilized Nutraceutical Pulse Crop: a Review. *J. Food Sci. Technol.* 52, 2489–2499. doi:10.1007/s13197-014-1312-z

Purseglove, J. W. (1974). *Tropical Crops: Dicotyledons*. London: Longman.

Qi, H., Kang, D., Zeng, W., Jawad Hassan, M., Peng, Y., Zhang, X., et al. (2021). Alterations of Endogenous Hormones, Antioxidant Metabolism, and Aquaporin Gene Expression in Relation to γ-Aminobutyric Acid-Regulated Thermotolerance in White Clover. *Antioxidants* 10, 1099. doi:10.3390/antiox10071099

Raizada, A., and Souframanien, J. (2019). Transcriptome Sequencing, De Novo Assembly, Characterisation of Wild Accession of Blackgram (*Vigna mungo Var. silvestris*) as a Rich Resource for Development of Molecular Markers and Validation of SNPs by High Resolution Melting (HRM) Analysis. *BMC Plant Biol.* 191-6. doi:10.1186/s12870-019-1954-0

Ramtekey, V., Bhuriya, A., Ayer, D., Parekh, V., Modha, K., Kale, B., et al. (2019). Molecular Tagging of Photoperiod Responsive Flowering in Indian Bean [*Lablab purpureus* (L.) Sweet]. *Indian J. Genet.* 79, 269. doi:10.31742/ijgpb.79s.1.17

Rawal, H. C., Kumar, S., Amitha Mithra, S. V., Solanke, A. U., Nigam, D., Saxena, S., et al. (2017). High Quality Unigenes and Microsatellites Markers from Tissue Specific Transcriptome and Development of a Database in Clusterbean (*Cymopsis tetragonoloba* (L.) Taub.). *Genes* 8, 313. doi:10.3390/genes8110313

Reddy, V. R., Das, S., Dikshit, H. K., Mishra, G. P., Aski, M. S., Singh, A., et al. (2021). Genetic Dissection of Phosphorous Uptake and Utilization Efficiency Traits Using GWAS in Mungbean. *Agronomy* 11, 1401. doi:10.3390/agronomy11071401

Ren, L., Huang, W., and Cannon, S. B. (2019). Reconstruction of Ancestral Genome Reveals Chromosome Evolution History for Selected Legume Species. *New Phytol.* 223, 2090–2103. doi:10.1111/nph.15770

Ren, Y. Z., Yue, Y. L., Jin, G. X., and Du, Q. (2014). First Report of Bacterial Blight of Guar Caused by Xanthomonas Axonopodis Pv. Cyamopsidis in China. *Plant Dis.* 98, 840. doi:10.1094/pdis-09-13-0983-pdn

Renzi, J. P., Chantre, G. R., Smýkal, P., Presotto, A. D., Zubiaga, L., Garayalde, A. F., et al. (2020). Diversity of Naturalized Hairy Vetch (Vicia Villosa Roth) Populations in Central Argentina as a Source of Potential Adaptive Traits for Breeding. *Front. Plant Sci.* 11, 189. doi:10.3389/fpls.2020.00189

Riday, H., and Krohn, A. L. (2010). Genetic Map-Based Location of the Red clover (*Trifolium pratense* L.) Gametophytic Self-Incompatibility Locus. *Theor. Appl. Genet.* 121, 761–767. doi:10.1007/s00122-010-1347-0

Robotham, O., and Chapman, M. (2017). Population Genetic Analysis of Hyacinth Bean (*Lablab purpureus* (L.) Sweet, Leguminosae) Indicates an East African Origin and Variation in Drought Tolerance. *Genet. Resour. Crop Evol.* 64, 139–148. doi:10.1007/s10722-015-0339-y

Rubatzky, V. E., and Yamaguchi, M. (1997). World Vegetables Principles, Production, and Nutritive Values. *Fruits* 5 (51), 381.

Rychel-Bielska, S., Nazzicari, N., Plewiński, P., Bielski, W., Annicchiarico, P., and Książkiewicz, M. (2020). Development of PCR-Based Markers and Whole-Genome Selection Model for Anthracnose Resistance in white Lupin (*Lupinus albus* L.). *J. Appl. Genet.* 61, 531–545. doi:10.1007/s13353-020-00585-1

Rychel, S., and Książkiewicz, M. (2019). Development of Gene-Based Molecular Markers Tagging Low Alkaloid Pauper Locus in white Lupin (*Lupinus albus* L.). *J. Appl. Genet.* 60, 269–281. doi:10.1007/s13353-019-00508-9

Samineni, S., Sen, M., Sajja, S. B., and Gaur, P. M. (2019). Rapid Generation advance (RGA) in Chickpea to Produce up to Seven Generations Per Year and Enable Speed Breeding. *Crop J.* 8, 164–169. doi:10.1016/j.cj.2019.08.003

Sammour, R. H. (2014). Genetic Diversity in *Lathyrus sativus* L. Germplasm. *Res. Rev. Biosci.* 8, 325–336.

Sankar, A. U., Anitha, K., Sivaraj, N., Kumari, K. M., Sunil, N., and Chakrabarty, S. K. (2015). Screening of Horsegram Germplasm Collected from Andhra Pradesh against Anthracnose. *Legume Res.* 38, 753–757. doi:10.18805/lr.v38i6.6719

Santos, C., Almeida, N. F., Alves, M. L., Horres, R., Krezdorn, N., Leitão, S. T., et al. (2018). First Genetic Linkage Map of Lathyrus Cicera Based on RNA Sequencing-Derived Markers: Key Tool for Genetic Mapping of Disease Resistance. *Hort. Res.* 5, 1–4. doi:10.1038/s41438-018-0047-9

Sato, S., Isobe, S., Asamizu, E., Ohmido, N., Kataoka, R., Nakamura, Y., et al. (2005). Comprehensive Structural Analysis of the Genome of Red clover (*Trifolium pratense* L.). *DNA Res.* 12, 301–364. doi:10.1093/dnares/dsi018

Saxena, K. B., Saxena, R. K., Hickey, L. T., and Varshney, R. K. (2019). Can a Speed Breeding Approach Accelerate Genetic Gain in Pigeonpea? *Euphytica* 215, 202. doi:10.1007/s10681-019-2520-4

Schafleitner, R., Huang, S. M., Chu, S. H., Yen, J. Y., Lin, C. Y., Yan, M. R., et al. (2016). Identification of Single Nucleotide Polymorphism Markers Associated

with Resistance to Bruchids (Callosobruchus spp.) in Wild Mungbean (*Vigna radiata* Var. *sublobata*) and Cultivated *V. radiata* through Genotyping by Sequencing and Quantitative Trait Locus Analysis. *BMC Plant Biol.* 16, 1–5. doi:10.1186/s12870-016-0847-8

Sengooba, T. N., Spence, N. J., Walkey, D. G. A., Allen, D. J., and Femi Lana, A. (1997). The Occurrence of Bean Common Mosaic Necrosis Virus in Wild and Forage Legumes in Uganda. *Plant Pathol.* 46, 95–103. doi:10.1046/j.1365-3059.1997.d01-12.x

Shapter, F. M., Cross, M., Ablett, G., Malory, S., Chivers, I. H., King, G. J., et al. (2013). High-throughput Sequencing and Mutagenesis to Accelerate the Domestication of Microlaena Stipoides as a New Food Crop. *PLoS One* 8, e82641. doi:10.1371/journal.pone.0082641

Sharma, S., Tyagi, A., Srivastava, H., Ramakrishna, G., Sharma, P., Sevanthi, A. M., et al. (2021). Exploring the Edible Gum (Galactomannan) Biosynthesis and its Regulation during Pod Developmental Stages in Clusterbean Using Comparative Transcriptomic Approach. *Sci. Reps.* 11, 1–2. doi:10.1038/s41598-021-83507-3

Shegro, A., Van Rensburg, W. J., and Adebola, P. O. (2013). Assessment of Genetic Variability in Bambara Groundnut (*Vigna subterrenea* L. Verdc.) Using Morphological Quantitative Traits. *Academia J. Agric. Res.* 1 (3), 045–051.

Sherman, R. M., and Salzberg, S. L. (2020). Pan-genomics in the Human Genome Era. *Nat. Rev. Genet.* 21, 243–254. doi:10.1038/s41576-020-0210-7

Shi, R., Jiao, W., Yun, L., Zhang, Z., Zhang, X., Wang, Q., et al. (2021). Utilization of Transcriptome, Small RNA, and Degradome Sequencing to Provide Insights into Drought Stress and Rewatering Treatment in Medicago Ruthenica. *Front. Plant Sci.* 12, 675903. doi:10.3389/fpls.2021.675903

Shirasawa, K., Kosugi, S., Sasaki, K., Ghelfi, A., Okazaki, K., Toyoda, A., et al. (2021b). Genome Features of Common Vetch (*Vicia sativa*) in Natural Habitats. *bioRxiv*. doi:10.1101/2021.03.09.434686

Shirasawa, K., Chahota, R., Hirakawa, H., Nagano, S., Nagasaki, H., Sharma, T. R., et al. (2021a). A Chromosome-Scale Draft Genome Sequence of Horsegram (*Macrotyloma uniflorum*). *bioRxiv*. 2021a Jan 1. doi:10.1101/2021.01.18.427074

Singh, T., Ramakrishnan, S., Mahanta, S. K., Tyagi, V. C., and Roy, A. K. (2018). *Tropical Forage Legumes in India: Status and Scope for Sustaining Livestock Production*. London, UK: Intech Open, 123–143.

Singh, V., Goel, R., Pande, V., Asif, M. H., and Mohanty, C. S. (2017). De Novo sequencing and Comparative Analysis of Leaf Transcriptomes of Diverse Condensed Tannin-Containing Lines of Underutilized *Psophocarpus tetragonolobus* (L.) DC. *Sci. Reps.* 7, 1–3. doi:10.1038/srep44733

Sokolkova, A., Burlyaeva, M., Valiannikova, T., Vishnyakova, M., Schafleitner, R., Lee, C. R., et al. (2020). Genome-wide Association Study in Accessions of the Mini-Core Collection of Mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan). *BMC Plant Biol.* 20, 1–9. doi:10.1186/s12870-020-02579-x

Somta, P., Chen, J., Yundaeng, C., Yuan, X., Yimram, T., Tomooka, N., et al. (2019). Development of an SNP-Based High-Density Linkage Map and QTL Analysis for Bruchid (*Callosobruchus maculatus* F.) Resistance in Black Gram (*Vigna mungo* (L.) Hepper). *Sci. Reps* 9 (1), 1–9. doi:10.1038/s41598-019-40669-5

Somta, P., Jomsangawong, A., Yundaeng, C., Yuan, X., Chen, J., Tomooka, N., et al. (2018). Genetic Dissection of Azuki Bean Weevil (*Callosobruchus chinensis* L.) Resistance in Moth Bean (*Vigna aconitifolia* [Jaqc.] Maréchal). *Genes* 9, 555. doi:10.3390/genes9110555

Somta, P., Kaga, A., Tomooka, N., Kashiwaba, K., Isemura, T., Chaitieng, B., et al. (2006). Development of an Interspecific Vigna Linkage Map between *Vigna umbellata* (Thunb.) Ohwi & Ohashi and *V. nakashimae* (Ohwi) Ohwi & Ohashi and its Use in Analysis of Bruchid Resistance and Comparative Genomics. *Plant Breed* 125, 77–84. doi:10.1111/j.1439-0523.2006.01123.x

Souframanien, J., and Gopalakrishna, T. (2006). ISSR and SCAR Markers Linked to the Mungbean Yellow Mosaic Virus (MYMV) Resistance Gene in Blackgram [*Vigna mungo* (L.) Hepper]. *Plant Breed* 125, 619–622. doi:10.1111/j.1439-0523.2006.01260.x

Souframanien, J., Gupta, S. K., and Gopalakrishna, T. (2010). Identification of Quantitative Trait Loci for Bruchid (*Callosobruchus maculatus*) Resistance in Black Gram [*Vigna mungo* (L.) Hepper]. *Euphytica* 176, 349–356. doi:10.1007/s10681-010-0210-3

Souframanien, J., and Reddy, K. S. (2015). De Novo Assembly, Characterization of Immature Seed Transcriptome and Development of Genic-SSR Markers in Black Gram [*Vigna mungo* (L.) Hepper]. *PLoS One* 10, e0128748. doi:10.1371/journal.pone.0128748

Soylu, S., Soylu, E. M., Kara, M., Kurt, Ş., and Choi, Y. J. (2020). First Report of Downy Mildew Disease Caused by Peronospora Viciae on Common Vetch (Vicia Sativa) in Turkey. *Plant Dis.* 104, 1876. doi:10.1094/pdis-12-19-2568-pdn

Sserumaga, J. P., Kayondo, S. I., Kigozi, A., Kiggundu, M., Namazzi, C., Walusimbi, K., et al. (2021). Genome-wide Diversity and Structure Variation Among Lablab [*Lablab purpureus* (L.) Sweet] Accessions and Their Implication in a Forage Breeding Program. *Genet. Resour. Crop Evol.* 19, 1–4.

Sudha, M., Karthikeyan, A., Shobhana, V. G., Nagarajan, P., Raveendran, M., Senthil, N., et al. (2015). Search for Vigna Species Conferring Resistance to Mungbean Yellow Mosaic Virus in Mungbean. *Plant Genet. Resour.* 13, 162. doi:10.1017/s1479262114000859

Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., et al. (2015). Targeted Mutagenesis in Soybean Using the CRISPR-Cas9 System. *Sci. Reps.* 5, 10342. doi:10.1038/srep10342

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

Takagi, H., Tamiru, M., Abe, A., Yoshida, K., Uemura, A., Yaegashi, H., et al. (2015). MutMap Accelerates Breeding of a Salt-Tolerant Rice Cultivar. *Nat. Biotechnol.* 33, 445–449. doi:10.1038/nbt.3188

Takahashi, Y., Sakai, H., Yoshitsu, Y., Muto, C., Anai, T., Pandiyan, M., et al. (2019). Domesticating Vigna Stipulacea: A Potential Legume Crop with Broad Resistance to Biotic Stresses. *Front. Plant Sci.* 10, 1607. doi:10.3389/fpls.2019.01607

Talhinhas, P., Baroncelli, R., and Le Floch, G. (2016). Anthracnose of Lupins Caused by Colletotrichum Lupini: A Recent Disease and a Successful Worldwide Pathogen. *J. Plant Pathol.* 5, 10342. doi:10.1038/srep10342

Tamburini, G., Bommarco, R., Wanger, T. C., Kremen, C., van der Heijden, M. G., Liebman, M., et al. (2020). Agricultural Diversification Promotes Multiple Ecosystem Services Without Compromising Yield. *Sci. Adv.* 6, eaba1715. doi:10.1126/sciadv.aba1715

Tan, L. X., Azam-Ali, S., Goh, E. V., Mustafa, M. A., Chai, H. H., Kuan, Ho. W., et al. (2020). Bambara Groundnut: An Underutilized Leguminous Crop for Global Food Security and Nutrition. *Front. Nutr.* 7, 276. doi:10.3389/fnut.2020.601496

Tanwar, U. K., Pruthi, V., and Randhawa, G. S. (2017). RNA-seq of Guar (*Cyamopsis tetragonoloba*, L. Taub.) Leaves: De Novo Transcriptome Assembly, Functional Annotation and Development of Genomic Resources. *Front. Plant Sci.* 8, 91. doi:10.3389/fpls.2017.00091

Tate, M., and Enneking, D. (2006). Common Vetch (Vicia Sativa Ssp. Sativa): Feed or Future Food? *Grain Legumes* 47, 16–17.

Taylor, C. M., Kamphuis, L. G., Zhang, W., Garg, G., Berger, J. D., Mousavi-Derazmahalleh, M., et al. (2019). INDEL Variation in the Regulatory Region of the Major Flowering Time Gene LanFTc1 Is Associated with Vernalization Response and Flowering Time in Narrow-leafed Lupin (*Lupinus angustifolius* L.). *Plant Cel Environ* 42, 174–187. doi:10.1111/pce.13320

Taylor, C. M., Garg, G., Berger, J. D., Ribalta, F. M., Croser, J. S., Singh, K. B., et al. (2021). A Trimethylguanosine Synthase1-like (TGS1) Homologue Is Implicated in Vernalisation and Flowering Time Control. *Theor. Appl. Genet.* 134, 3411–3426. doi:10.1007/s00122-021-03910-2

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. Nat. Acad. Sci.* 102, 13950–13955. doi:10.1073/pnas.0506758102

Thakur, O., and Randhawa, G. S. (2018). Identification and Characterization of SSR, SNP and InDel Molecular Markers from RNA-Seq Data of Guar (*Cyamopsis tetragonoloba*, L. Taub.) Roots. *BMC Genome* 19, 951. doi:10.1186/s12864-018-5205-9

Thirumaran, A. S., and Kanchana, S. (2000). "Role of Pulses in Human Diets," in *Pulses Production Strategies in Tamil Nadu* (Coimbatore, India: Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University), 129. Available at: http://www.tnau.ac.in/tech/pulsesbook. pdf (Accessed July 13, 2012).

Tian, J., Isemura, T., Kaga, A., Vaughan, D. A., and Tomooka, N. (2013). Genetic Diversity of the rice Bean (*Vigna umbellata*) Genepool as Assessed by SSR Markers. *Genome* 56, 717–727. doi:10.1139/gen-2013-0118

Tomooka, N., Vaughan, D. A., Moss, H., and Maxted, N. (2002). *The Asian* Vigna*: Genus* Vigna *Subgenus* Ceratotropis *Genetic Resources*. New York: Kluwer Academic Publishers.

Tribhuvan, K. U., Amitha Mithra, S. V., Sharma, P., Das, A., Kumar, K., Tyagi, A., et al. (2019). Identification of Genomic SSRs in Cluster Bean (*Cyamopsis tetragonoloba*) and Demonstration of Their Utility in Genetic Diversity Analysis. *Ind. Crops Prod.* 133, 221–231. doi:10.1016/j.indcrop.2019.03.028

Tripathy, S., and Das, M. K. (2013). Guar Gum: Present Status and Applications. *J. Pharm. Sci. Innov.* 4, 24–28. doi:10.7897/2277-4572.02447

Tyagi, A., Nigam, D., Sv, A. M., Solanke, A. U., Singh, N. K., Sharma, T. R., et al. (2018). Genome-wide Discovery of Tissue-Specific miRNAs in Clusterbean (*Cyamopsis tetragonoloba*) Indicates Their Association with Galactomannan Biosynthesis. *Plant Biotechnol. J.* 16, 1241–1257. doi:10.1111/pbi.12866

Tyagi, A., Sharma, P., Saxena, S., Sharma, R., Mithra, S. A., Solanke, A. U., et al. (2019). The Genome Size of Clusterbean (*Cyamopsis tetragonoloba*) Is Significantly Smaller Compared to its Wild Relatives as Estimated by Flow Cytometry. *Gene* 707, 205–211. doi:10.1016/j.gene.2019.02.090

Uba, C. U., Oselebe, H. O., Tesfaye, A. A., and Abtew, W. G. (2021). Genetic Diversity and Population Structure Analysis of Bambara Groundnut (*Vigna subterrenea* L) Landraces Using DArT SNP Markers. *PLoS One* 16, e0253600. doi:10.1371/journal.pone.0253600

Vaijayanthi, P. V., Ramesh, S., Gowda, M. B., and Mohan, R. A. (2015). Genetic Variability for Morpho-Metric Traits in Dolichos Bean (*Lablab purpureus* L. Sweet) Germplasm. *J. Food Legumes* 28 (1), 5–10.

van Oers, C. C. C. M. (1989). "Vigna Aconitifolia (Jacq.) Maréchal," in *Plant Resources of South-East Asia No 1. Pulses*. Editors L. J. G. van der Maesen and S. Somaatmadja (Wageningen, Netherlands: Pudoc), 66–67.

Varshney, R. K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., et al. (2021). A Chickpea Genetic Variation Map Based on the Sequencing of 3,366 Genomes. *Nature* 599 (7886), 622–627. doi:10.1038/s41586-021-04066-1

Vatanparast, M., Shetty, P., Chopra, R., Doyle, J. J., Sathyanarayana, N., and Egan, A. N. (2016). Transcriptome Sequencing and Marker Development in Winged Bean (Psophocarpus Tetragonolobus; Leguminosae). *Sci. Reps.* 6, 1–4. doi:10.1038/srep29070

Vaughan, D. A., Tomooka, N., and Kaga, A. (2005). "Azuki Bean [*Vigna Angularis* (L.) Ohwi and Ohashi]," in *Genetic Resources, Chromosome Engineering and Crop Improvement*. Editors R. J. Singh and P. P. Jauhar (Boca Roca, FL: CRC Press), 1, 347–359. Grain Legumes. doi:10.1201/9780203489284-15

Vaz Patto, M. C., Skiba, B., Pang, E. C. K., Ochatt, S. J., Lambein, F., and Rubiales, D. (2006). Lathyrus Improvement for Resistance against Biotic and Abiotic Stresses: From Classical Breeding to Marker Assisted Selection. *Euphytica* 147, 133–147. doi:10.1007/s10681-006-3607-2

Vleugels, T., Baert, J., and Van Bockstaele, E. (2013). Evaluation of a Diverse Red clover Collection for Clover Rot Resistance (*Sclerotinia trifoliorum*). *Commun. Agricult. Appl. Biol. Sci.* 78, 519–522.

Von Wettberg, E., Davis, T. M., and Smýkal, P. (2020). Wild Plants as Source of New Crops. *Front. Plant Sci.* 11, 1426. doi:10.3389/fpls.2020.591554

Wang, J., Li, J., Liu, Z., Yuan, X., Wang, S., Chen, H., et al. (2020). Construction of a High-Density Genetic Map and its Application for QTL Mapping of Leaflet Shapes in Mung Bean (*Vigna radiata* L.). *Front. Genet.* 11. doi:10.3389/fgene.2020.01032

Wang, L., Do Kim, K., Gao, D., Chen, H., Wang, S., Lee, S., et al. (2016). Analysis of Simple Sequence Repeats in rice Bean (*Vigna umbellata*) Using an SSR-Enriched Library. *Crop J.* 4, 40–47. doi:10.1016/j.cj.2015.09.004

Wang, L. X., Jie, W. A., Luo, G. L., Yuan, X. X., Dan, G. O., Hu, L. L., et al. (2021b). Construction of a High-Density Adzuki Bean Genetic Map and Evaluation of its Utility Based on a QTL Analysis of Seed Size. *J. Integr. Agric.* 20, 1753–1761. doi:10.1016/s2095-3119(20)63343-3

Wang, M. L., Morris, J. B., Barkley, N. A., Dean, R. E., Jenkins, T. M., and Pederson, G. A. (2007). Evaluation of Genetic Diversity of the USDALablab Purpureus Germplasm Collection Using Simple Sequence Repeat Markers. *J. Hortic. Sci. Biotech.* 82, 571–578. doi:10.1080/14620316.2007.11512275

Wang, P., Zhou, G., Jian, J., Yang, H., Renshaw, D., Aubert, M. K., et al. (2021a). Whole-Genome Assembly and Resequencing Reveal Genomic Imprint and Key Genes of Rapid Domestication in Narrow-Leafed Lupin. *Plant J.* 105, 1192–1210. doi:10.1111/tpj.15100

Wang, T., Ren, L., Li, C., Zhang, D., Zhang, X., Zhou, G., et al. (2021c). The Genome of a Wild Medicago Species Provides Insights into the Tolerant Mechanisms of Legume Forage to Environmental Stress. *BMC Plant Biol.* 19, 1–7. doi:10.1186/s12915-021-01033-0

Warschefsky, E., Penmetsa, R. V., Cook, D. R., and Von Wettberg, E. J. (2014). Back to the Wilds: Tapping Evolutionary Adaptations for Resilient Crops through Systematic Hybridization with Crop Wild Relatives. *Amer. J. Bot.* 101, 1791–1800. doi:10.3732/ajb.1400116

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M.-D., et al. (2018). Speed Breeding Is a Powerful Tool to Accelerate Crop Research and Breeding. *Nat. Plants* 4, 23–29. doi:10.1038/s41477-017-0083-8

Wei, Z. W., Hailong, R., Zinian, W., Guozhi, L., Xiang, C., and Zhihong, Q. (2015). A New Medicago Polymorpha Cultivar 'Huaiyang Jinhuacai. *Acta Hortic. Sin* 42, 2335–2336.

Wu, X., Islam, A. S., Limpot, N., Mackasmiel, L., Mierzwa, J., Cortés, A. J., et al. (2020). Genome-wide SNP Identification and Association Mapping for Seed Mineral Concentration in Mung Bean (*Vigna radiata* L.). *Front. Genet.* 11, 656. http://www.fao.org/faostat. doi:10.3389/fgene.2020.00656

Xu, Q., Liu, F., Qu, R., Gillman, J. D., Bi, C., Hu, X., et al. (2018). Transcriptomic Profiling of Lathyrus Sativus L. Metabolism of β-ODAP, a Neuroexcitatory Amino Acid Associated with Neurodegenerative Lower Limb Paralysis. *Plant Mol. Biol. Rep.* 36, 832–843. doi:10.1007/s11105-018-1123-x

Yadav, S. S., Bejiga, G., Brink, M., and Belay, G. (2006). *'Lathyrus Sativus L.' PROTA4U*. Wageningen, Netherlands: PROTA (Plant Resources of Tropical Africa/Ressources végétales de l'Afrique tropicale). http://www.prota4u.org/search.asp.

Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., et al. (2015). Genome Sequencing of Adzuki Bean (*Vigna angularis*) Provides Insight into High Starch and Low Fat Accumulation and Domestication. *Proc. Nat. Acad. Sci.* 112 (43), 13213–13218. doi:10.1073/pnas.1420949112

Yang, H., Jian, J., Li, X., Renshaw, D., Clements, J., Sweetingham, M. W., et al. (2015). Application of Whole Genome Re-sequencing Data in the Development of Diagnostic DNA Markers Tightly Linked to a Disease-Resistance Locus for Marker-Assisted Selection in Lupin (*Lupinus angustifolius*). *BMC Genome* 16, 1–7. doi:10.1186/s12864-015-1878-5

Yang, H., Tao, Y., Zheng, Z., Li, C., Sweetingham, M. W., and Howieson, J. G. (2012). Application of Next-Generation Sequencing for Rapid Marker Development in Molecular Plant Breeding: a Case Study on Anthracnose Disease Resistance in *Lupinus angustifolius* L. *BMC Genom.* 13, 1–2. doi:10.1186/1471-2164-13-318

Yang, H., Tao, Y., Zheng, Z., Shao, D., Li, Z., Sweetingham, M. W., et al. (2013b). Rapid Development of Molecular Markers by Next-Generation Sequencing Linked to a Gene Conferring Phomopsis Stem Blight Disease Resistance for Marker-Assisted Selection in Lupin (*Lupinus angustifolius* L.) Breeding. *Theor. Appl. Genet.* 126, 511–522. doi:10.1007/s00122-012-1997-1

Yang, H., Tao, Y., Zheng, Z., Zhang, Q., Zhou, G., Sweetingham, M. W., et al. (2013a). Draft Genome Sequence, and a Sequence-Defined Genetic Linkage Map of the Legume Crop Species *Lupinus angustifolius* L. *PLoS One* 8, e64799. doi:10.1371/journal.pone.0064799

Yang, T., Jiang, J., Burlyaeva, M., Hu, J., Coyne, C. J., Kumar, S., et al. (2014). Large-Scale Microsatellite Development in Grasspea (*Lathyrus sativus* L.), an Underutilized Legume of the Arid Areas. *BMC Plant Biol.* 14, 1–2. doi:10.1186/1471-2229-14-65

Yang, T., Nagy, I., Mancinotti, D., Otterbach, S. L., Andersen, T. B., Motawia, M. S., et al. (2017). Transcript Profiling of a Bitter Variety of Narrow-Leafed Lupin to Discover Alkaloid Biosynthetic Genes. *J. Expt Bot.* 68, 5527–5537. doi:10.1093/jxb/erx362

Yao, L. M., Wang, B., Cheng, L. J., and Wu, T. L . (2013). Identification of Key Drought Stress-Related Genes in the Hyacinth Bean. *PLoS ONE* 8, e58108. doi:10.1371/journal.pone.0058108

Yaqoob, M., Khan, N. U., Mansoor, M., Gul, S., Khalil, I. A., and Afridi, K. (2015). Moth Bean Germplasm Screening against Yellow Mosaic Virus, II. Developmentof Moth Bean High-Yielding Seed and Fodder Cultivars. *Turk. J. Agric. For.* 39, 212–226. doi:10.3906/tar-1406-3

Yates, S. A., Swain, M. T., Hegarty, M. J., Chernukin, I., Lowe, M., Allison, G. G., et al. (2014). De novo Assembly of Red Clover Transcriptome Based on RNA-

Seq Data Provides Insight into Drought Response, Gene Discovery and Marker Identification. *BMC Genom.* 15 (1), 1–15.

Yigezu, Y. A., El-Shater, T., Boughlala, M., Bishaw, Z., Niane, A. A., Maalouf, F., et al. (2019). Legume-based Rotations Have clear Economic Advantages over Cereal Monocropping in Dry Areas. *Agron. Sust. Dev.* 39, 1–14. doi:10.1007/s13593-019-0602-2

Yin, L., Zhang, M., Wu, R., Chen, X., Liu, F., and Xing, B. (2021). Genome-wide Analysis of OSCA Gene Family Members in Vigna Radiata and Their Involvement in the Osmotic Response. *BMC Plant Biol.* 21, 1–2. doi:10.1186/s12870-021-03184-2

Yin, M., Zhang, S., Du, X., Mateo, R. G., Guo, W., Li, A., et al. (2021). Genomic Analysis of Medicago Ruthenica Provides Insights into its Tolerance to Abiotic Stress and Demographic History. *Mol. Ecol. Resour.* 21 (5), 1641–1657. doi:10.1111/1755-0998.13363

Yin, X., Yi, K., Zhao, Y., Hu, Y., Li, X., He, T., et al. (2020). Revealing the Full-Length Transcriptome of Caucasian clover Rhizome Development. *BMC Plant Biol.* 20, 1–5. doi:10.1186/s12870-020-02637-4

Yoshida, Y., Marubodee, R., Ogiso-Tanaka, E., Iseki, K., Isemura, T., Takahashi, Y., et al. (2016). Salt tolerance in wild relatives of adzuki bean, Vigna angularis (Willd.) Ohwi et Ohashi. *Genet. Resour. Crop Evol.* 63, 627–637. doi:10.1007/s10722-015-0272-0

Yoshida, J., Tomooka, N., Yee Khaing, T., Shantha, P. S., Naito, H., Matsuda, Y., et al. (2020). Unique Responses of Three Highly Salt-Tolerant Wild Vigna Species against Salt Stress. *Plant Prod. Sci.* 23, 114–128. doi:10.1080/1343943x.2019.1698968

You, M. P., Nichols, P. G., Katusiime, R., and Barbetti, M. J. (2021). Novel Disease Host Resistances in the World Core Collection of Trifolium Subterraneum. *Plant Dis.*. PDIS-09. doi:10.1094/pdis-09-20-1985-re

Yuan, J., Wang, B., and Wu, T. L. (2011). Quantitative Trait Loci (QTL) Mapping for Inflorescence Length Traits in *Lablab purpureus* (L.) Sweet. *Afr. J. Biotechnol.* 10, 3558–3566.

Yundaeng, C., Somta, P., Amkul, K., Kongjaimun, A., Kaga, A., and Tomooka, N. (2019). Construction of Genetic Linkage Map and Genome Dissection of Domestication-Related Traits of Moth Bean (*Vigna aconitifolia*), a Legume Crop of Arid Areas. *Mol. Genet. Genome* 294, 621–635. doi:10.1007/s00438-019-01536-0

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, H., Li, Y., and Zhu, J. K. (2018). Developing Naturally Stress-Resistant Crops for as Sustainable Agriculture. *Nat. Plants* 4, 989–996. doi:10.1038/s41477-018-0309-4

Zhang, H., Tian, H., Chen, M., Xiong, J., Cai, H., and Liu, Y. (2018). Transcriptome Analysis Reveals Potential Genes Involved in Flower Pigmentation in a Red-Flowered Mutant of white clover (*Trifolium repens* L.). *Genomics* 110, 191–200. doi:10.1016/j.ygeno.2017.09.011

Zhang, Q., Yan, Q., Yuan, X., Lin, Y., Chen, J., Wu, R., et al. (2021). Two Polygalacturonase-Inhibiting Proteins (VrPGIP) of *Vigna Radiata* Confer Resistance to Bruchids (*Callosobruchus* Spp). *J. Plant Physiol.* 258, 153376. doi:10.1016/j.jplph.2021.153376

Zhang, Y., Massel, K., Godwin, I. D., and Gao, C. (2018). Applications and Potential of Genome Editing in Crop Improvement. *Genome Biol.* 19, 1. doi:10.1186/s13059-018-1586-y

Zhou, G., Jian, J., Wang, P., Li, C., Tao, Y., Li, X., et al. (2018). Construction of an Ultra-high-density Consensus Genetic Map, and Enhancement of the Physical Map from Genome Sequencing in Lupinus Angustifolius. *Theor. Appl. Genet.* 131, 209–223. doi:10.1007/s00122-017-2997-y

Zhou, G., Yang, H., Renshaw, D., Zou, M., Thomas, G., and Li, C. (2021). Identification of Gray Leaf Spot Disease Candidate Gene in Narrow-Leafed Lupin (*Lupinus angustifolius* L.). *Front. Genet.* 2021, 1375. doi:10.3389/fgene.2021.695791

Zhu, Z., Chen, H., Xie, K., Liu, C., Li, L., Liu, L., et al. (2020). Characterization of Drought-Responsive Transcriptome during Seed Germination in Adzuki Bean (*Vigna angularis* L.) by PacBio SMRT and Illumina Sequencing. *Front. Genet.* 11, 996. doi:10.3389/fgene.2020.00996

Zongo, E., Néya, B. J., Traoré, V. S. E., Palanga, E., Zabré, J., Barro, N., et al. (2018). Impact of Cowpea Mottle Virus on the Growth and Yield of Bambara Groundnut (*Vigna subterranean* (L.) Verdc.). *Amer. J. Plant Sci.* 9, 2053. doi:10.4236/ajps.2018.910149

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 © 2022 Jha, Nayyar, Parida, Bakır, von Wettberg and Siddique. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fgene.2022.847647

Introgressing cry1Ac for Pod Borer Resistance in Chickpea Through Marker-Assisted Backcross Breeding

*Ajinder Kaur[1], Urvashi Sharma[1], Sarvjeet Singh[2], Ravinder Singh[2], Yogesh Vikal[1], Satnam Singh[3], Palvi Malik[1], Khushpreet Kaur[3], Inderjit Singh[2], Shayla Bindra[2], Bidyut Kumar Sarmah[4] and Jagdeep Singh Sandhu[1]**

[1]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, [2]Pulses Section, Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India, [3]Punjab Agricultural University, Regional Research Station, Faridkot, India, [4]Department of Biotechnology-Assam Agricultural University Centre, Assam Agricultural University, Jorhat, India

OPEN ACCESS

Edited by:
Nasya Borisova Tomlekova,
Maritsa Vegetable Crops Research Institute (MVCRI), Bulgaria

Reviewed by:
Salej Sood,
Indian Council of Agricultural Research (ICAR), India
Mohar Singh Thakur,
Indian Council of Agricultural Research (ICAR), India

***Correspondence:**
Jagdeep Singh Sandhu
js_sandhu@pau.edu

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

Received: 03 January 2022
Accepted: 09 March 2022
Published: 12 April 2022

Citation:
Kaur A, Sharma U, Singh S, Singh R, Vikal Y, Singh S, Malik P, Kaur K, Singh I, Bindra S, Sarmah BK and Sandhu JS (2022) Introgressing cry1Ac for Pod Borer Resistance in Chickpea Through Marker-Assisted Backcross Breeding. Front. Genet. 13:847647. doi: 10.3389/fgene.2022.847647

The gram pod borer *Helicoverpa armigera* is a major constraint to chickpea (*Cicer arietinum* L.) production worldwide, reducing crop yield by up to 90%. The constraint is difficult to overcome as chickpea germplasm including wild species either lacks pod borer resistance or if possessing resistance is cross-incompatible. This study describes conversion of elite but pod borer-susceptible commercial chickpea cultivars into resistant cultivars through introgression of cry1Ac using marker-assisted backcross breeding. The chickpea cultivars (PBG7 and L552) were crossed with pod borer-resistant transgenic lines (BS 100B and BS 100E) carrying cry1Ac that led to the development of BC_1F_1, BC_1F_2, BC_1F_3, BC_2F_1, BC_2F_2, and BC_2F_3 populations from three cross combinations. The foreground selection revealed that 35.38% BC_1F_1 and 8.4% BC_1F_2 plants obtained from Cross A (PBG7 × BS 100B), 50% BC_1F_1 and 76.5% BC_1F_2 plants from Cross B (L552 × BS 100E), and 12.05% BC_2F_2 and 82.81% (average) BC_2F_3 plants derived from Cross C (PBG7 × BS 100E) carried the cry1Ac gene. The bioassay of backcross populations for toxicity to *H. armigera* displayed up to 100% larval mortality. BC_1F_1 and BC_1F_2 populations derived from Cross B and BC_2F_3 population from Cross C segregated in the Mendelian ratio for cry1Ac confirmed inheritance of a single copy of transgene, whereas BC_1F_1 and BC_1F_2 populations obtained from Cross A and BC_2F_2 population from Cross C exhibited distorted segregation ratios. BC_1F_1 plants of Cross A and Cross B accumulated Cry1Ac protein ranging from 11.03 to 11.71 $\mu g g^{-1}$ in leaf tissue. Cry1Ac-positive BC_2F_2 plants from Cross C demonstrated high recurrent parent genome recovery (91.3%) through background selection using SSR markers and phenome recovery of 90.94%, amongst these 30% plants, were homozygous for transgene. The performance of BC_2F_3 progenies derived from homozygous plants was similar to that of the recurrent parent for main agronomic traits, such as number of pods and seed yield per plant. These progenies are a valuable source for *H. armigera* resistance in chickpea breeding programs.

Keywords: Cicer arietinum, Mendelian inheritance, transgene introgression, Helicoverpa armigera resistance, marker-assisted backcross breeding

INTRODUCTION

Chickpea (*Cicer arietinum* L., *2n* = 16), belonging to the family Leguminoseae, is an economical source of protein (18–22%), minerals, fiber, β-carotene, and unsaturated fatty acids (Jukanti et al., 2012). The crop is grown in nearly 57 countries with India, Australia, Myanmar, Ethiopia, Turkey, and Russia as the major producers (Merga and Haji 2019). The crop production is severely affected by various biotic and abiotic stresses leading up to 90% yield losses (Kumar et al., 2018). Among biotic stresses, gram pod borer *Helicoverpa armigera* (Hübner) [Lepidoptera: Noctuidae] causes significant crop damage annually (90%) estimated at US $330 million worldwide (Rao et al., 2013; Patil et al., 2017). *H. armigera* is difficult to control as it has migratory behavior, numerous generations per year, adaptability to different environmental conditions, high fecundity, and insecticidal resistance (Fitt 1989). Furthermore, the biopesticides used to control the insect have high production costs coupled with poor product quality control systems (Cherry et al., 2000; Jenkins and Grzywacz 2000). The development of pod borer-resistant chickpea cultivars through conventional breeding is hampered due to the narrow crop genetic base and crossability barriers between cultivated chickpea and wild *Cicer* species (Mallikarjuna et al., 2007).

The pod borer larvae have been effectively controlled through specific insecticidal crystal proteins of *Bacillus thuringiensis,* and Cry1Ac is the most effective toxin against *H. armigera* (Chakrabarti et al., 1998). Cry1Ac protein acts by targeting the insect midgut in which the prevalence of high pH solubilizes the protein; the activated protein forms a pore complex in the insect epithelial membrane causing lysis and eventually larval death (Bravo et al., 2008). The pod borer attack has been countered efficiently by transgenic chickpea plants carrying cry1Ac, cry1Ab, cry2Aa, and cry1Aa3 (Kar et al., 1997; Sanyal et al., 2005; Acharjee et al., 2010; Mehrotra et al., 2011; Khatodia et al., 2014). The introgression of cry genes from transgenic plants into elite cultivars/lines through marker-assisted backcross breeding leads to precise trait transfer, for e.g., enhanced resistance against striped stem borer in rice by introgression of cry1Ab (Wang et al., 2012), improved resistance against corn borer with cry1A.105 and cry2ab2 in maize inbred lines (Venkatesh et al., 2015), increased tolerance to fruit/shoot borer in eggplant following cry1Ac transfer (Ripalda et al., 2012), and improved insect resistance in cotton *via* cryIA introgression (Guo et al., 2005), etc. Marker-assisted backcross breeding, an effective molecular breeding technique, enables the transfer of desirable genes from an agronomically inferior donor into an elite recipient in a few generations, without linkage drag and in a smaller population size (Hospital and Charcosset 1997).

The introgression of cry genes from transgenic chickpea lines to commercial chickpea cultivars for imparting resistance against *Helicoverpa* following marker-assisted backcross breeding is not reported so far. In the present study, an attempt was made to convert two elite but pod borer-susceptible chickpea cultivars, namely, PBG7 and L552, into resistant cultivars by introgressing cry1Ac from pod borer-resistant transgenic lines, namely, BS 100B and BS 100E through marker-assisted backcross breeding. PBG7 is a high-yielding cultivar of desi chickpea, whereas L552 is a bold-seeded high-yielding cultivar of kabuli chickpea; both cultivars are recommended for commercial cultivation in the North Indian state, Punjab, and possess good cooking quality (Sandhu et al., 2012; Singh et al., 2015). The backcross populations were analyzed for the presence of transgene, evaluated for Cry1Ac concentration, and bioassayed for toxicity to *H. armigera*. The highlighting feature of this study was the introgression of cry1Ac in BC_1F_1 populations and its subsequent transmission to BC_1F_2, BC_1F_3, BC_2F_2, and BC_2F_3 that displayed up to 100% *H. armigera* larval mortality, and agronomic performance of selected BC_2F_2 and BC_2F_3 plants was similar to that of the recurrent parent.

MATERIALS AND METHODS

Plant Material

T_5 seeds (15 in number) for each of two transgenic chickpea lines, namely, BS 100B and BS 100E expressing cry1Ac gene under the control of the *Arabidopsis* Rubisco small subunit gene promoter and tobacco SSU terminator (**Supplementary Figure S1**), were procured from the Department of Biotechnology-Assam Agricultural University Centre, Assam Agricultural University, Jorhat, Assam, India, during 2013. The transgenic lines carrying cry1Ac at a single locus were used as donor (male) parents in chickpea backcrossing program; the lines are reported to accumulate a high level of Cry1Ac protein (˃ 50 $\mu g\, g^{-1}$ leaf tissue) that causes 80–100% neonatal *H. armigera* larval mortality (Hazarika et al., 2019). The high-yielding commercial cultivars PBG7 (desi) and L552 (kabuli) were used as recipient (female) parents. F_1 plants of PBG7 × BS 100B (designated as Cross A), L552 × BS 100E (Cross B), and PBG7 × BS 100E (Cross C) were backcrossed with their respective recipient parents to obtain BC_1F_1 seeds that were sown to generate BC_1F_1 populations. F_1 plants, BC_1F_1, BC_1F_2, BC_1F_3, BC_2F_1, BC_2F_2, and BC_2F_3 populations were raised under contained conditions (**Supplementary Figure S2**; **Supplementary Table S1**) in a net house (30-mesh screen) at Experimental Farms, Department of Plant Breeding & Genetics, Punjab Agricultural University (PAU), Ludhiana. The populations were grown in plots comprising 25 rows with 2 m length and row-to-row distance of 40 cm following normal agronomic practices during the month of October. The schematic overview of marker assisted-backcross breeding of commercial chickpea cultivars with transgenic lines is shown in **Figure 1**.

DNA Extraction

Genomic DNA was extracted from tender twigs of 20-day-old BC_1F_1, BC_1F_2, BC_2F_2, and BC_2F_3 populations, transgenic donor parents BS 100B and BS 100E, and non-transgenic recipient parents PBG7 and L552 according to the miniprep method. For quantification, the extracted DNA was electrophoresed on 0.8% (w/v) agarose gel using PowerPacHC (Bio-Rad, United States) at 50 V for 2 h; ethidium bromide-stained gel

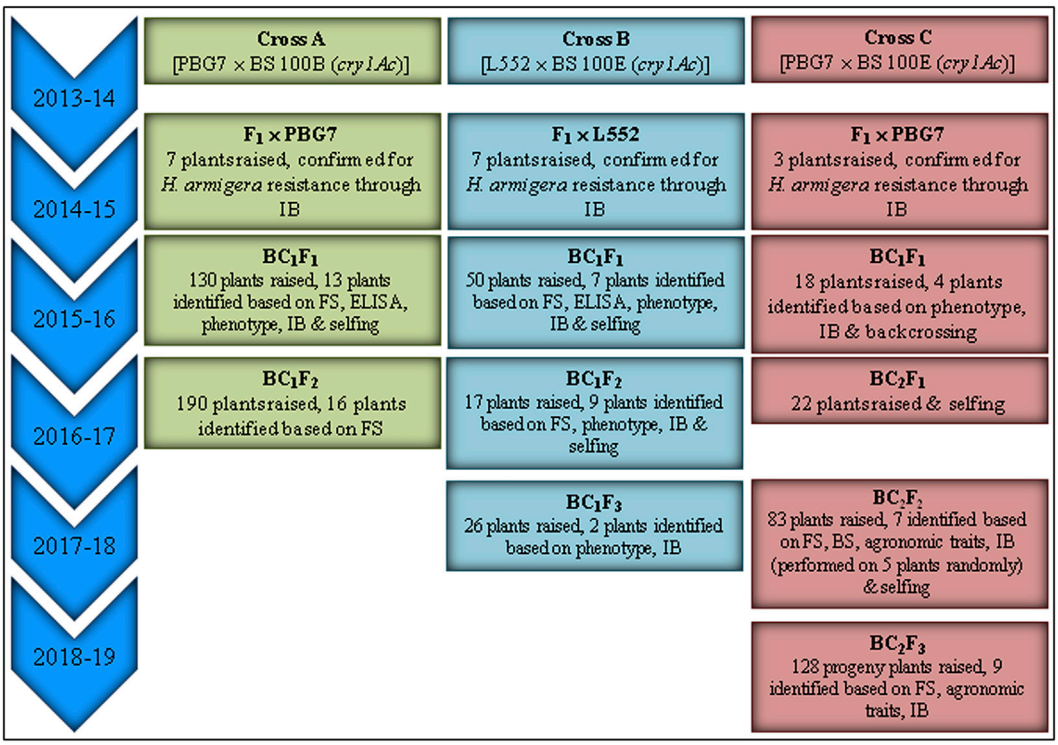


FIGURE 1 | Schematic overview of marker-assisted backcross breeding of commercial chickpea cultivars × transgenic lines.

was visualized under UV light and photographed on a 110 V AlphaImager HP imaging system (ProteinSimple, United Kingdom).

Foreground Selection for cry1Ac-Positive Plants

The foreground selection of backcross populations was carried out through PCR using cry1Ac-specific (Accession Number M11068, Hazarika et al., 2019), internal forward 5-TATCTT TGGTCCATCTCAATGGG-3 and reverse 5-GTGTCCAGA CCAGTAATACTC-3 primers to amplify 757 bp transgene. PCR mixture (20 µl) contained 50 ng genomic DNA (2 µl), 10 µM of each primer (0.6 µl), 1 mM dNTPs (4 µl), 25 mM $MgCl_2$ (1.5 µl), 5 × Green GoTaq Flexi buffer (4 µl), 5 units GoTaq DNA polymerase (1 µl) [Promega, United States] and nuclease-free water (6.3 µl). The reaction mixtures were placed in a GeneAmp PCR System 9700 (Thermo Fisher Scientific, United States) programmed for an initial denaturation at 94 C for 4 min, followed by 35 cycles of denaturation at 94 C for 50 s, annealing at 58 C for 1 min, extension at 72 C for 1 min, and concluded by a final extension at 72 C for 7 min and held at 4 C prior to storage. The amplicons were resolved on 1.5% (w/v) agarose gel, visualized, and photographed. The statistical significance for cry1Ac segregation data was determined by Chi-square analysis using the formula: $\chi^2 = (O\text{-}E)^2/E$, where O is the observed value and E is the expected value.

Enzyme-Linked Immunosorbent Assay

Cry1Ac expression in BC_1F_1 plants and transgenic donor and non-transgenic recipient parents was quantified through ELISA using Cry1Ac QuantiPlate kit (EnviroLogix, United States) according to the manufacturer's instructions. A total of two leaflets (10 mg) of each plant were homogenized in an Eppendorf grinding tube for 20–30 s by adding 500 µl of 1 × extraction buffer. Each leaf tissue sample (50 µl) was diluted in a 1:11 ratio by adding 550 µl of 1 × extraction buffer; thereafter, 100 µl each of diluted sample, negative control, and positive calibrator was dispensed in the ELISA plate, followed by parafilm masking and incubation at an ambient temperature for 15 min. The assay was performed in triplicate. Cry1Ac-enzyme conjugate (100 µl) was added to each well, and the plate was again covered with parafilm and incubated for 1 h. After incubation, the parafilm mask was removed and well contents were agitated vigorously to decant the wells. The vacant wells were flooded with washing buffer and agitated to decant; the washing step was performed thrice. Then substrate (100 µl) was added to each well and mixed thoroughly, followed by plate covering with parafilm and incubation for 20 min. The reaction was terminated by adding 100 µl of stop solution to each well. The ELISA plate was read in a 96-well ELISA plate reader Infinite 200 Pro (Tecan, Switzerland) at 450 and 600 nm. The optical density (OD) values of samples and positive calibrators were analyzed using a Microsoft Excel sheet to generate a linear scale graph of

the mean OD of each calibrator against its Cry1Ac concentration (**Supplementary Table S2**). The amount of Cry1Ac protein in each leaf tissue sample (μg g^{-1}) was determined using the formula {(OD of sample - mean OD of negative control) - 0.425/0.127} × dilution factor 1 (38.46) × dilution factor 2 (11)/1,000 (**Supplementary Table S2**). The data were analyzed for mean ± standard deviation using Microsoft Excel 2007 software at default settings.

Background Selection for Recurrent Parent Genome Recovery

The background selection of cry1Ac-positive BC_2F_2 plants was carried out using Simple Sequence Repeat (SSR) markers. As a preliminary step, polymorphism analysis was undertaken on parents PBG7 and BS 100E using 210 markers belonging to the following series: CGMM, CaM, GA, GAA, TA, TAA, TS, TR, NCPGR, H, and CaSTMS 11 (**Supplementary Table S3**). The amplified products were resolved on 6% (w/v) PAGE, and marker data were scored based on differential separation of amplicon(s). BC_2F_2 plants possessing maximum recurrent parent genome were identified with reproducible polymorphic SSR markers (**Supplementary Table S3**). The percent recurrent parent genome recovery in a BC_2F_2 plant was calculated as the sum of the number of alleles corresponding to recurrent parent detected by polymorphic markers divided by the total number of alleles detected by polymorphic and cry1Ac-specific markers.

Assessment for Agronomic Traits

The agronomic performance of BC_2F_2 population was assessed for plants analyzed for recurrent parent genome recovery, and of BC_2F_3 population was based on three progeny plants (from each BC_2F_2 plant) having phenotype similar to the recurrent parent. The data were recorded on days to 50% flowering, number of branches per plant, days to maturity, plant height, number of pods per plant, number of seeds per plant, 100-seed weight, biological yield, seed yield per plant, and harvest index, compared with the recurrent parent and analyzed for percent phenome recovery in BC_2F_2 plants and for mean ± standard deviation in BC_2F_3 plants.

Bioassay for Determining Toxicity to *H. armigera*

Four-month-old morphologically healthy plants were analyzed for toxicity to *H. armigera* using two approaches, i.e., detached leaf bioassay and whole plant bioassay given by Sharma et al. (2005a), Sharma et al. (2005b) with modifications. Detached leaf bioassay: The terminal twigs having fully expanded leaflets were plucked from F_1, backcross population (BC_1F_1, BC_1F_2 and BC_2F_2), transgenic donor parent and non-transgenic recipient parent plants, and placed on 3% (w/v) agar (HiMedia, India) medium slants in sterile 500 ml plastic cups and used for bioassay. *H. armigera* larvae (3rd to 4th instar) collected in February from chickpea fields were reared individually in bioassay cups and maintained initially on non-transformed tender chickpea twigs, followed by growth on a semi-synthetic diet (Armes et al., 1992) until pupation. The pupae were kept on moist sponges covered with filter paper (Whatman, United States) in plastic containers till the emergence of adults that were paired in oviposition chambers i.e., cell pots wrapped in black paper on all sides and covered with muslin cloth on top. The adults were fed on 5% (v/v) honey solution by hanging honey-soaked cotton swab inside each oviposition chamber. Subsequently, egg laying occurred on the muslin cloth that was shifted to bioassay cup containing semi-synthetic diet for egg hatching, thereafter, neonates were used for bioassay of plant twigs. Ten neonate larvae were released in each bioassay cup and incubated in a growth chamber (Saveer Biotech Limited, India) maintained at 25 ± 2 C, 14 h light: 10 h dark period and >65 ± 5 percent relative humidity. The bioassay was replicated thrice and performed in the Pulses Entomology Laboratory, Department of Plant Breeding & Genetics, PAU, Ludhiana. The assayed plants were visually scored for the damage caused by neonate larvae after 96 h of release on a scale of 1–9 (1 = < 10% leaf area damaged, and 9 = > 80% leaf area damaged) given by Sharma et al. (2005a) for detached leaf assay. The larval mortality rate was compared among plants of backcross populations and donor and recipient parents to monitor the relationship between percent larval mortality and Cry1Ac protein concentration. The data were analyzed for mean ± standard deviation.

Whole plant bioassay: The assay was carried out under net house conditions on plants grown in plots with row to row distance of 40 cm and plant to plant spacing of 10 cm, according to the method given by Sharma et al. (2005b) for screening chickpea against *H. armigera* under greenhouse conditions with modifications. The healthy plants at the flowering stage from BC_1F_3 and BC_2F_3 populations and donor and recipient parents were covered with cages sized 25 × 25 × 75 cm^3. The cages made of galvanized iron wire (2 mm in diameter) were supported by four vertical bars and covered with a muslin cloth bag. The experiment was performed in triplicate by caging three plants of each population, parent individually and releasing 10 *H. armigera* neonatal larvae on each plant, and terminated after 120 h when significant leaf area was damaged in recipient parents. The plants were scored for leaf-feeding visually on a 1-9 scale (where 1 = < 10%, 2 = 11–20%, 3 = 21–30%, 4 = 31–40%, 5 = 41–50%, 6 = 51–60%, 7 = 61–70%, 8 = 71–80% and 9 = > 80% leaf area and/or pods damaged). The number of surviving larvae was recorded and individually placed in 25 ml plastic cups to express the data as percent larval mortality that were analyzed using Microsoft Excel 2007 software. No insecticide was applied in the experiment.

Statistical Analysis

Data on Cry1Ac protein concentration, leaf-feeding, and larval mortality in backcross populations are presented as mean ± SD of three replicates. Statistical significance for the segregation data was determined using Chi-square analysis; calculated Chi-square value > table value was considered statistically significant at 5 percent level of significance.

TABLE 1 | Bioassay on F_1 plants and backcross populations raised by crossing commercial chickpea cultivars with cry1Ac transgenic lines for toxicity to *H. armigera* neonatal larvae.

S. No	Plant number/ parent	Cry1Ac protein concentration (µg G-1 leaf tissue)			Average protein concentration (µg G-1 leaf tissue)	Leaf feeding score[a]			Average leaf feeding score	Leaf feeding damage (%)	Larval mortality (%)			Average larval mortality (%)
		R_1	R_2	R_3		R_1	R_2	R_3			R_1	R_2	R_3	
F_1 plants (derived from Cross A) through detached leaf bioassay														
1	1	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
2	2	NA	NA	NA	NA	1	1	2	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
3	3	NA	NA	NA	NA	1	1	2	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
4	4	NA	NA	NA	NA	1	2	1	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
5	5	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
6	6	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
7	7	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
	BS 100B	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
	PBG7	NA	NA	NA	NA	6	6	6	6 ± 0.0	51–60	30	30	30	30 ± 0.0
F1 plants (derived from Cross B) through detached leaf bioassay														
1	1	NA	NA	NA	NA	2	1	1	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
2	2	NA	NA	NA	NA	2	1	1	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
3	3	NA	NA	NA	NA	1	2	1	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
4	4	NA	NA	NA	NA	1	1	2	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
5	5	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
6	6	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
7	7	NA	NA	NA	NA	1	2	2	1.67 ± 0.58	< 10–20	100	100	100	100 ± 0.0
	BS 100E	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
	L552	NA	NA	NA	NA	7	7	7	7 ± 0.0	61–70	30	20	20	23.33 ± 5.77
F1 plants (derived from Cross C) through detached leaf bioassay														
1	1	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
2	2	NA	NA	NA	NA	2	2	2	2 ± 0.0	11–20	100	100	100	100 ± 0.0
3	3	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
	BS 100E	NA	NA	NA	NA	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
	PBG7	NA	NA	NA	NA	6	6	6	6 ± 0.0	51–60	30	30	30	30 ± 0.0
BC_1F_1 population (derived from Cross A) through detached leaf bioassay														
1	4	11.68	11.71	11.71	11.70 ± 0.02	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
2	6	11.71	11.71	11.71	11.71 ± 0.0	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
3	7	11.27	11.28	11.27	11.27 ± 0.01	2	2	2	2 ± 0.0	11–20	100	100	100	100 ± 0.0
4	16	11.68	11.68	11.68	11.68 ± 0.0	2	2	2	2 ± 0.0	11–20	100	100	100	100 ± 0.0
5	17	11.71	11.71	11.71	11.71 ± 0.0	2	2	2	2 ± 0.0	11–20	100	100	100	100 ± 0.0
6	18	11.71	11.71	11.71	11.71 ± 0.0	2	2	2	2 ± 0.0	11–20	100	100	100	100 ± 0.0
7	21	11.57	11.61	11.61	11.60 ± 0.02	1	1	2	1.33 ± 0.58	< 10–20	100	100	100	100 ± 0.0
8	22	11.27	11.38	11.31	11.32 ± 0.06	2	2	2	2 ± 0.0	11–20	100	100	100	100 ± 0.0
9	24	11.64	11.61	11.64	11.63 ± 0.02	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
10	77	11.64	11.64	11.64	11.64 ± 0.0	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
11	81	11.34	11.34	11.34	11.34 ± 0.0	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
12	89	11.64	11.64	11.64	11.64 ± 0.0	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0
13	90	11.04	11.01	11.04	11.03 ± 0.02	1	1	1	1 ± 0.0	< 10	100	100	100	100 ± 0.0

(Continued on following page)

TABLE 1 | (*Continued*) Bioassay on F_1 plants and backcross populations raised by crossing commercial chickpea cultivars with cry1Ac transgenic lines for toxicity to *H. armigera* neonatal larvae.

S. No	Plant number/ parent	Cry1Ac protein concentration (µg G-1 leaf tissue)			Average protein concentration (µg G-1 leaf tissue)	Leaf feeding score[a]			Average leaf feeding score	Leaf feeding damage (%)	Larval mortality (%)			Average larval mortality (%)
		R_1	R_2	R_3		R_1	R_2	R_3			R_1	R_2	R_3	
	BS 100B	11.38	11.34	11.34	11.35 ± 0.02	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	PBG7	0	0	0	0 ± 0.0	6	6	6	6 ± 0.0	51–60	30	30	30	30 ± 0.0
BC_1F_1 population (derived from Cross B) through detached leaf bioassay														
1	1	11.61	11.61	11.61	11.61 ± 0.0	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
2	2	11.57	11.57	11.57	11.57 ± 0.0	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
3	4	11.51	11.51	11.51	11.51 ± 0.0	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
4	25	11.51	11.51	11.51	11.51 ± 0.0	1	2	1	1.33 ± 0.58	˂ 10–20	100	80	100	93.33 ± 11.55
5	34	11.68	11.68	11.68	11.68 ± 0.0	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
6	36	11.61	11.61	11.61	11.61 ± 0.0	1	2	1	1.33 ± 0.58	˂ 10–20	100	80	100	93.33 ± 11.55
7	41	11.68	11.68	11.68	11.68 ± 0.0	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
8	42	11.51	11.51	11.51	11.51 ± 0.0	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
9	45	11.54	11.51	11.54	11.53 ± 0.02	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	BS 100E	11.68	11.61	11.64	11.64 ± 0.04	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	L552	0	0	0	0 ± 0.0	7	7	7	7 ± 0.0	61–70	30	20	20	23.33 ± 5.77
BC1F1 population (derived from Cross C) through detached leaf bioassay														
1	1	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
2	2	NA	NA	NA	NA	1	2	1	1.33 ± 0.58	˂ 10–20	100	80	100	93.33 ± 11.55
3	3	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
4	4	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
5	8	NA	NA	NA	NA	1	2	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	BS 100E	NA	NA	NA	NA	0	0	0	0 ± 0.0	0	100	100	100	100 ± 0.0
	PBG7	NA	NA	NA	NA	6	6	6	6 ± 0.0	51–60	30	30	30	30 ± 0.0
BC_1F2 population (derived from Cross B) through detached leaf bioassay														
1	1	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
2	3	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
3	9	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
4	10	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
5	12	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
6	14	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
7	15	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
8	16	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
9	17	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	BS 100E	NA	NA	NA	NA	0	0	0	0 ± 0.0	0	100	100	100	100 ± 0.0
	L552	NA	NA	NA	NA	7	7	7	7 ± 0.0	61–70	30	20	20	23.33 ± 5.77
BC_1F_3 population (derived from Cross B) through whole plant bioassay														
1	1	NA	NA	NA	NA	2	3	2	2.33 ± 0.58[€]	11–30[b]	70	40	60	56.67 ± 15.27
2	3	NA	NA	NA	NA	2	2	2	2 ± 0.0	11–20	60	80	60	66.67 ± 11.55
3	9	NA	NA	NA	NA	1	2	1	1.33 ± 0.58	˂ 10–20	100	80	100	93.33 ± 11.55
4	10	NA	NA	NA	NA	5	6	5	5.33 ± 0.58	41–60	20	10	20	16.67 ± 5.77

(Continued on following page)

TABLE 1 | (*Continued*) Bioassay on F_1 plants and backcross populations raised by crossing commercial chickpea cultivars with cry1Ac transgenic lines for toxicity to *H. armigera* neonatal larvae.

S. No	Plant number/ parent	Cry1Ac protein concentration (µg G-1 leaf tissue)			Average protein concentration (µg G-1 leaf tissue)	Leaf feeding score[a]			Average leaf feeding score	Leaf feeding damage (%)	Larval mortality (%)			Average larval mortality (%)
		R_1	R_2	R_3		R_1	R_2	R_3			R_1	R_2	R_3	
5	12	NA	NA	NA	NA	3	2	2	2.33 ± 0.58	11–30	60	80	80	73.33 ± 11.55
6	14	NA	NA	NA	NA	2	1	1	1.33 ± 0.58	˂ 10–20	80	100	100	93.33 ± 11.55
	BS 100E	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	L552	NA	NA	NA	NA	7	7	7	7 ± 0.0	61–70	30	20	20	23.33 ± 5.77
BC_2F_2 population (derived from Cross C) through detached leaf bioassay														
1	2	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
2	8	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
3	20	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
4	33	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
5	39	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	BS 100E	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	PBG7	NA	NA	NA	NA	6	6	6	6 ± 0.0	51–60	0	0	0	0 ± 0.0
BC_2F_3 population (derived from Cross C) through whole plant bioassay														
1	2	NA	NA	NA	NA	3	4	3	3.33 ± 0.58[€]	21–40[b]	40	60	60	53.33 ± 11.55
2	8	NA	NA	NA	NA	3	4	3	3.33 ± 0.58	21–40	40	60	60	53.33 ± 11.55
3	20	NA	NA	NA	NA	2	3	2	2.33 ± 0.58	11–30	80	60	80	73.33 ± 11.55
4	26	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
5	33	NA	NA	NA	NA	3	2	2	2.33 ± 0.58	11–30	60	80	80	73.33 ± 11.55
6	39	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
7	44	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	BS 100E	NA	NA	NA	NA	1	1	1	1 ± 0.0	˂ 10	100	100	100	100 ± 0.0
	PBG7	NA	NA	NA	NA	6	6	6	6 ± 0.0	51–60	0	0	0	0 ± 0.0

Data on Cry1Ac protein concentration, leaf feeding, and larval mortality are presented as mean ± SD of three replicates. Leaf feeding score and larval mortality in detached leaf bioassay were recorded after 96 h of incubation whereas in whole plant bioassay after 120 h.

[a]Leaf feeding score: The plants were scored visually for the extent of damage caused on a 1–9 scale, where 1 = ˂ 10% and 9 = ˃ 80% leaf area damaged in detached leaf bioassay (Sharma et al., 2005a), and in whole plant bioassay, 1 = ˂ 10%, 2 = 11–20%, 3 = 21–30%, 4 = 31–40%, 5 = 41–50%, 6 = 51–60%, 7 = 61–70%, 8 = 71–80%, and 9 = ˃ 80% leaf area and/or pods damaged (Sharma et al., 2005b). NA, not analyzed.

[b]Average leaf and/or pod feeding score.

[c]Percent leaf and/or pod feeding damage; Cross A: PBG7 × BS 100B; Cross B: L552 × BS 100E; Cross C: PBG7 × BS 100E.

TABLE 2 | Segregation analyses of backcross populations developed by crossing commercial chickpea cultivars with cry1Ac transgenic lines.

S. No	Backcross population	Observed number		Observed ratio	Expected number		Expected Ratio	Calculated χ^2 value	*p*-value
		cry1Ac-positive plants	cry1Ac-negative plants		cry1Ac-positive plants	cry1Ac-negative plants			
1	BC_1F_1 (derived from Cross A)	46	84	1:1.8	65	65	1:1	11.10[a]	0.000863
2	BC_1F_1 (derived from Cross B)	25	25	1:1	25	25	1:1	0.0	1
3	BC_1F_2 (derived from Cross A)	16	174	1:10.9	142.50	47.50	3:1	449.18[a]	0.00001
4	BC_1F_2 (derived from Cross B)	13	4	3.3:1	12.75	4.25	3:1	0.02	0.8875
5	BC_2F_2 (derived from Cross C)	10	73	1:7.3	62.25	20.75	3:1	175.43[a]	0.00001
6	BC_2F_3 (derived from Cross C) progeny of: BC_2F_2 plant no. 1	18	3	6:1	15.75	5.25	3:1	1.28	0.257,899
	BC_2F_2 plant no. 2	7	3	2.3:1	7.50	2.50	3:1	0.13	0.718
	BC_2F_2 plant no. 8	5	4	1.3:1	6.75	2.25	3:1	1.81	0.1785
	BC_2F_2 plant no. 9	12	4	3:1	12	4	3:1	0.0	1
	BC_2F_2 plant no. 12	8	5	1.6:1	9.75	3.25	3:1	1.25	0.26355
	BC_2F_2 plant no. 20	4	1	4:1	3.75	1.25	3:1	0.07	0.791
	BC_2F_2 plant no. 26	16	0	16:0[§]	-	-	-	-	-
	BC_2F_2 plant no. 33	6	2	3:1	6	2	3:1	0.0	1
	BC_2F_2 plant no. 39	10	0	10:0[§]	-	-	-	-	-
	BC_2F_2 plant no. 44	20	0	20:0[§]	-	-	-	-	-

Data expressed as the number of cry1Ac-positive and -negative plants based on PCR.

[a]*indicates significance at 0.05 level (χ^2 table value = 3.84, 1 df)*

[b]*implies homozygous nature of respective BC_2F_2 plants; Cross A: PBG7 × BS 100B; Cross B: L552 × BS 100E; Cross C: PBG7 × BS 100E.*

RESULTS

Analysis on F_1 Plants for Determining Toxicity to *H. armigera*

F_1 plants developed from Cross A, Cross B, Cross C, and transgenic donor parent and non-transgenic recipient parents were analyzed for toxicity to *H. armigera*. F_1 plants obtained from Cross A (seven in number), Cross B (seven), Cross C (three), and donor parent displayed 100% *H. armigera* neonatal larval mortality and negligible (<10–20%) leaf-feeding damage, whereas recipient parents exhibited 23.33–30% larval mortality on an average with significant (51 to 70%) leaf-feeding damage (**Table 1**). F_1 plants toxic to *H. armigera* were backcrossed to generate BC_1F_1 populations.

Analysis on BC_1F_1 Populations for Foreground Selection and Determining Toxicity to *H. armigera*

The foreground selection of two BC_1F_1 populations derived from Cross A and Cross B, comprising 130 and 50 plants, respectively, was carried out through PCR using cry1Ac-specific primers. An amplicon corresponding to cry1Ac was detected in 46 (35.38%) BC_1F_1 plants obtained from Cross A (**Supplementary Figure S3**; **Supplementary Table S1**). The transgene segregation in a ratio of 1:1.8 deviated significantly from the 1:1 ratio expected if transgene was inserted at a single locus (**Table 2**). BC_1F_1 raised from Cross B segregated for the transgene in an expected Mendelian ratio of 1:1, as 25 (50.0%) plants were found to be cry1Ac positive (**Supplementary Figure S4**).

The recombinant protein concentration was estimated in 13 healthy BC_1F_1 plants derived from Cross A showing amplification of cry1Ac, nine Cross B plants along with transgenic donor parents BS 100B and BS 100E, and non-transgenic recipient parents PBG7 and L552 through ELISA (**Supplementary Table S2**). The average Cry1Ac protein concentration in both populations (11.03 to 11.71 $\mu g\ g^{-1}$ leaf tissue) was at par with donor parents (11.35 to 11.64 $\mu g\ g^{-1}$), whereas recipient parents did not exhibit any Cry1Ac concentration (**Table 1**). The BC_1F_1 plants (13 obtained from Cross A, nine from Cross B, and five from Cross C) had a phenotype similar to the recurrent parent, their bioassay for toxicity to *H. armigera* revealed that 13, 7, 4 plants from respective crosses, donor parents showed 100% *H. armigera* mortality and minor (<10–20%) leaf-feeding damage; in contrast, recipient parents exhibited 23.33–30% larval mortality with significant (51 to 70%) leaf-feeding damage (**Table 1**; **Supplementary Figure S5**). BC_1F_1 plants displaying toxicity to *H. armigera* were advanced for raising BC_1F_2 populations.

Analysis on BC_1F_2 Populations for Foreground Selection and Determining Toxicity to *H. armigera*

The foreground selection was carried out on two BC_1F_2 populations: the first comprising of 190 plants derived from Cross A, and the second consisting of 17 plants obtained from Cross B; cry1Ac

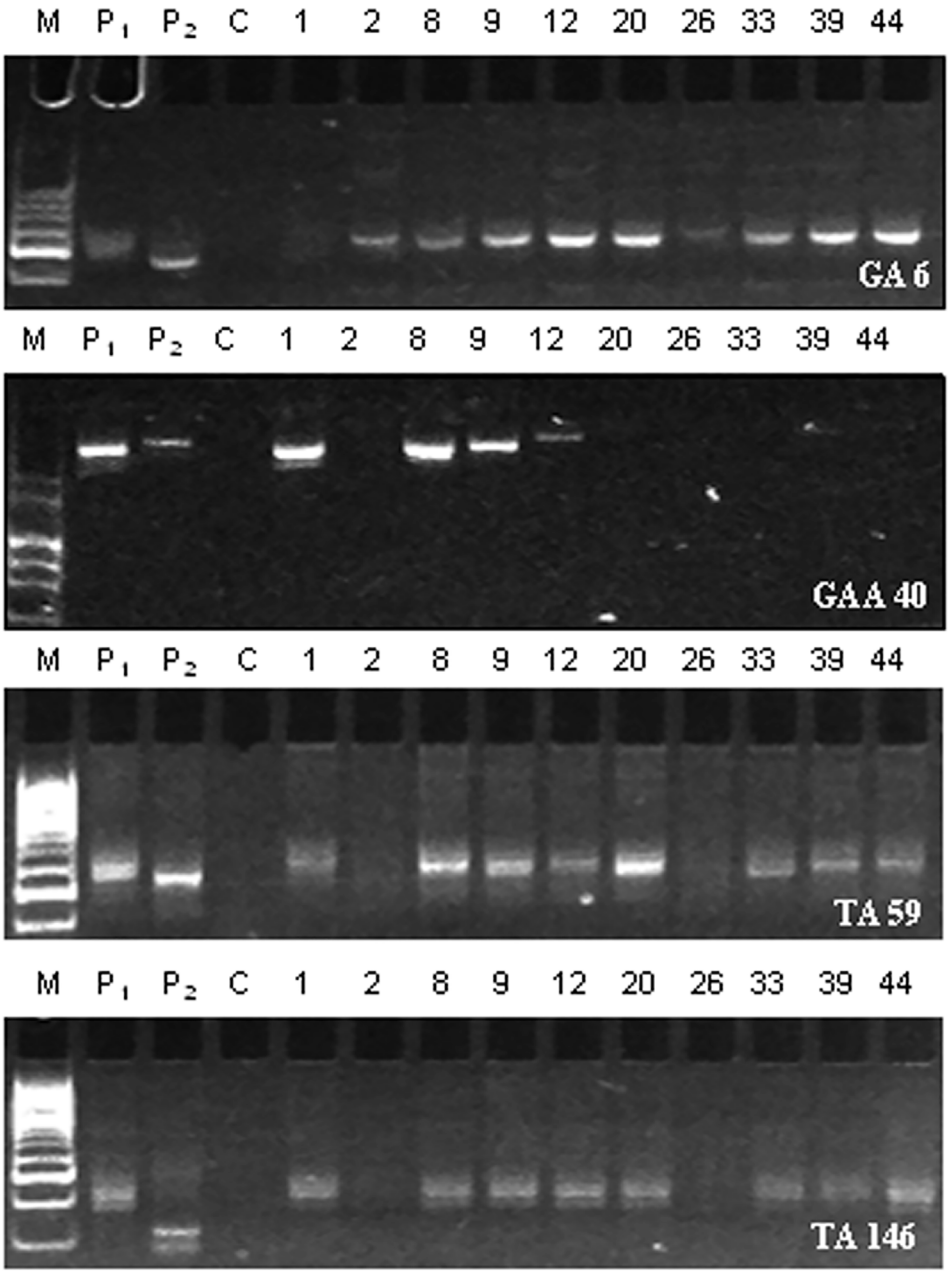


FIGURE 2 | SSR amplification profiles of BC_2F_2 plants using polymorphic markers, namely, GA 6, GAA 40, TA 59, and TA 146. P_1 indicates non-transgenic recipient parent PBG7; P_2 represents transgenic donor parent BS 100E; C refers to control PCR reaction without template DNA; the numbers 1, 2, 8, 9, 12, 20, 26, 33, 39, and 44 denote BC_2F_2 plants; and M represents 50 bp DNA ladder (Cat. No. DM1100, SMOBIO Technology, Inc., Taiwan).

amplification was detected in 16 (8.4%) and 13 (76.5%) plants (**Supplementary Figures S6, S7**; **Supplementary Table S1**), exhibiting non-Mendelian (1:10.9) and Mendelian (3.3:1) segregation ratios in the two populations, respectively (**Table 2**).

The insect bioassay was performed on nine BC_1F_2 plants raised from Cross B displaying phenotypic growth similar to the recurrent parent, along with transgenic donor and non-transgenic recipient parents. In the BC_1F_2 plants, donor parent displayed 100% *H. armigera* mortality and negligible (<10%) leaf-feeding damage; however, recipient parent showed 23.33% larval mortality with significant (61 to 70%) leaf-feeding damage (**Table 1**). BC_1F_2 plants showing toxicity to *H. armigera* were used to raise BC_1F_3 population.

Analysis on BC_1F_3 Population for Determining Toxicity to *H. armigera*

Six out of 26 BC_1F_3 plants having comparable phenotype to the recurrent parent developed from Cross B, transgenic donor

TABLE 3 | Agronomic traits of BC_2F_2 and BC_2F_3 populations derived from Cross C (PBG7 × BS 100E).

S. No	Plant number/ parent	Agronomic trait										Recurrent parent phenome recovery (%)
		Days to 50% flowering	Number of branches per plant	Days to maturity	Plant height (cm)	Number of pods per plant	Number of seeds per plant	100-seed weight (g)	Biological yield (g)	Seed yield per plant (g)	Harvest index[a] (%)	
BC_2F_2												
1	1	84 (93.33)	13 (76.47)	149 (96.13)	51.3 (86.51)	44 (73.33)	80 (70.80)	15.6 (93.97)	40.22 (77.97)	13.21 (76.36)	32.84 (97.91)	84.28
2	2	89 (98.89)	15 (88.23)	153 (98.71)	55.5 (93.59)	55 (91.67)	102 (90.26)	15.5 (93.37)	47.16 (91.43)	15.45 (89.31)	32.76 (97.67)	93.31
3	8	87 (96.67)	16 (94.12)	152 (98.06)	57.5 (96.96)	46 (76.67)	85 (75.22)	16.2 (97.59)	44.36 (86.00)	14.12 (81.62)	31.83 (94.90)	89.78
4	9	88 (97.78)	12 (70.59)	154 (99.35)	58.1 (97.98)	43 (71.67)	78 (69.03)	15.8 (95.18)	38.34 (74.33)	12.43 (71.85)	32.42 (96.66)	84.44
5	12	85 (94.44)	12 (70.59)	151 (97.42)	50.2 (84.65)	49 (81.67)	94 (83.18)	15.6 (93.97)	43.78 (84.88)	14.32 (82.77)	32.71 (97.52)	87.11
6	20	89 (98.89)	15 (88.23)	153 (98.71)	58.3 (98.31)	53 (88.33)	98 (86.72)	15.1 (90.96)	49.56 (96.08)	15.20 (87.86)	30.67 (91.44)	92.55
7	26	89 (98.89)	14 (82.35)	154 (99.35)	57.6 (97.13)	59 (98.33)	112 (99.11)	15.2 (91.57)	51.00 (98.87)	16.86 (97.46)	33.06 (97.46)	96.05
8	33	86 (95.55)	16 (94.12)	150 (96.77)	51.4 (86.68)	54 (90.00)	99 (87.61)	15.3 (92.17)	47.62 (92.32)	15.23 (88.03)	31.98 (95.35)	91.86
9	39	87 (96.67)	15 (88.23)	151 (97.42)	55.6 (93.76)	57 (95.00)	108 (95.57)	15.3 (92.17)	50.50 (97.91)	16.38 (94.68)	32.43 (96.69)	94.81
10	44	86 (95.55)	15 (88.23)	151 (97.42)	56.2 (94.77)	58 (96.67)	108 (95.57)	15.3 (92.17)	51.02 (98.91)	16.60 (95.95)	32.54 (97.02)	95.23
	BS 100E	82	8	145	48.4	16	28	13.3	23.30	4.13	17.72	-
	PBG7	90	17	155	59.3	60	113	16.6	51.58	17.30	33.54	-
Average recurrent parent phenome recovery = 90.94%												
BC_2F_3												
1	2	88.33 ± 0.58	14.00 ± 1.00	150.67 ± 0.58	55.70 ± 1.58	50.67 ± 2.52	97.67 ± 3.51	15.36 ± 0.21	44.34 ± 2.78	13.99 ± 1.12	31.59 ± 2.59	
2	8	86.00 ± 1.00	16.00 ± 1.00	151.67 ± 1.15	55.80 ± 2.66	44.33 ± 2.08	83.33 ± 5.51	15.83 ± 0.15	44.07 ± 3.09	14.28 ± 0.16	32.50 ± 2.04	
3	20	87.33 ± 1.15	15.67 ± 0.58	154.30 ± 0.58	57.47 ± 2.20	53.67 ± 1.53	99.00 ± 1.00	15.37 ± 0.49	52.08 ± 2.08	15.64 ± 0.51	30.03 ± 0.26	
4	26	89.00 ± 1.00	13.67 ± 0.58	152.00 ± 0.27	58.13 ± 2.37	54.00 ± 3.00	111.00 ± 4.58	15.13 ± 0.35	52.13 ± 3.56	15.33 ± 0.98	31.35 ± 0.27	
5	33	86.67 ± 1.53	14.00 ± 1.00	152.67 ± 1.15	55.30 ± 4.05	44.00 ± 3.00	79.33 ± 4.16	15.60 ± 0.26	39.42 ± 2.61	12.88 ± 0.22	32.77 ± 1.86	
6	39	86.67 ± 1.53	15.33 ± 1.15	152.33 ± 1.53	53.76 ± 3.16	55.00 ± 4.00	103.67 ± 7.64	15.23 ± 0.66	48.36 ± 1.61	15.74 ± 0.62	32.55 ± 0.66	
7	44	86.33 ± 1.53	14.00 ± 1.00	150.67 ± 0.58	57.03 ± 2.05	53.67 ± 3.78	102.67 ± 7.02	15.23 ± 0.55	45.20 ± 1.16	14.84 ± 0.53	32.83 ± 0.56	
	BS 100E	82.33 ± 1.15	8.33 ± 1.15	147.00 ± 1.00	46.50 ± 2.52	14.67 ± 2.08	25.33 ± 3.51	13.40 ± 0.43	22.71 ± 1.63	4.29 ± 0.22	18.94 ± 0.64	
	PBG7	88.33 ± 0.58	16.33 ± 0.58	154.00 ± 1.00	57.13 ± 2.43	54.67 ± 3.05	107.00 ± 7.00	15.93 ± 0.25	49.54 ± 1.91	15.90 ± 0.56	32.10 ± 0.22	

Data on BC_2F_2 population are presented for the plants analyzed for recurrent parent genome recovery; Data on BC_2F_3 population are based on three plants phenotypically similar to PBG7 and presented as mean ± SD; figures in parentheses are recurrent parent recovery percentages for agronomic traits calculated as plant trait value/value of PBG7 for that trait × 100.
[a]Harvest index = seed yield per plant/biological yield × 100; recurrent parent phenome recovery percentage was calculated as the sum of recurrent parent recovery percentages for different traits/10.

parent, and non-transgenic recipient parent were analyzed for toxicity to *H. armigera*. The plants revealed 16.67–93.33% larval mortality and variable (˂ 10–60%) leaf and pod feeding damage; donor parent exhibited 100% *H. armigera* mortality with negligible (<10%) leaf and pod feeding damage, whereas recipient parent showed 23.33% larval mortality and significant (61 to 70%) damage to leaves and pods (**Table 1**). Two BC_1F_3 plants were observed to display 93.33% insect mortality.

Analysis on BC_2F_2 Population for Foreground and Background Selection and Determining Toxicity to *H. armigera*.

The foreground selection of BC_2F_2 population derived from Cross C and comprising of 83 plants led to the identification of 10 (12.05%) plants showing amplification of cry1Ac (**Supplementary Table S1**). The population deviated significantly for transgene segregation (1:7.3) from the Mendelian ratio (3:1) for a single insertion site (**Table 2**).

The donor and recipient parents were assessed for polymorphism using 210 SSR markers leading to the identification of 25 (11.9%) polymorphic markers (**Supplementary Figure S8**; **Supplementary Table S3**). The background selection using reproducible polymorphic markers on cry1Ac-positive BC_2F_2 plants demonstrated amplification pattern in ten BC_2F_2 plants (designated as 1, 2, 8, 9, 12, 20, 26, 33, 39, and 44) to be similar to recurrent parent "PBG7" profile (**Figure 2**), and the average recurrent parent genome recovery in these plants after two backcrosses was calculated to be 91.3% (**Supplementary Table S4**). The comparison of agronomic traits in BC_2F_2 plants with PBG7 revealed an average recurrent parent phenome recovery of 90.94% in BC_2F_2 plants (**Table 3**; **Supplementary Table S5**).

The randomly selected BC_2F_2 plants (designated as 2, 8, 20, 33, and 39) were bioassayed for toxicity to *H. armigera*. The results revealed that the selected plants and transgenic donor parent exhibited 100% larval mortality and negligible (<10%) leaf-feeding damage, whereas non-transgenic recipient parent was vulnerable to *H. armigera* with no larval mortality and significant (51 to 60%) leaf-feeding damage (**Table 1**). Subsequently, seeds of all ten BC_2F_2 plants were sown to obtain BC_2F_3 population.

Analysis on BC_2F_3 Population for Foreground Selection, Agronomic Traits, and Determining Toxicity to *H. armigera*.

BC_2F_3 population obtained from Cross C, consisting of 128 plants was subjected to foreground selection for identifying BC_2F_2 plants homozygous for cry1Ac through recognition of BC_2F_3 plants carrying cry1Ac gene. The results revealed that on an average, 82.81% BC_2F_3 plants carried cry1Ac, and three (30%) BC_2F_2

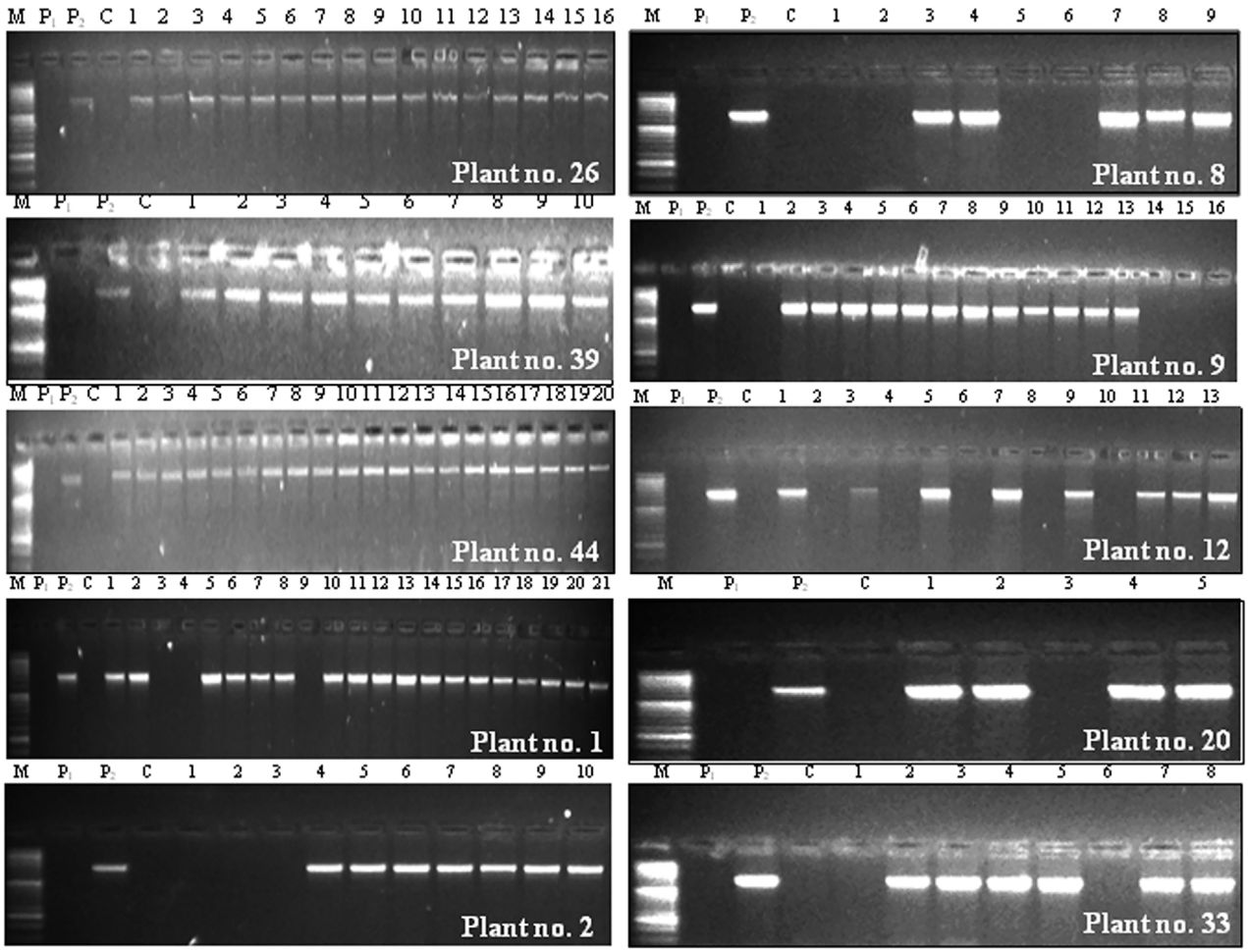


FIGURE 3 | Foreground selection of BC_2F_3 population derived from Cross C (PBG7 × BS 100E) through PCR using cry1Ac-specific primers. P_1 indicates non-transgenic recipient parent PBG7; P_2 represents transgenic donor parent BS 100E; C refers to control PCR reaction without template DNA; the numbers on top of each gel represent BC_2F_3 progenies of a specific plant, and its identity is mentioned in the right bottom corner of each gel; the plants designated as 26, 39, and 44 were homozygous for cry1Ac, and those designated as 1, 2, 8, 9, 12, 20, and 33 were hemizygous for the transgene; M represents 50 bp DNA ladder (Cat. No. DM1100).

plants designated as 26, 39, and 44 were homozygous for the transgene as all progeny plants (16 of plant no. 26, 10 of plant no. 39, and 20 of plant no. 44) contained the transgene (**Table 2**: **Figure 3**). On the contrary, the remaining seven BC_2F_2 plants designated as 1, 2, 8, 9, 12, 20, and 33 were hemizygous for cry1Ac with their BC_2F_3 progeny plants segregating in a ratio of 6:1, 2.3:1, 1.3:1, 3:1, 1.6:1, 4:1, and 3:1, respectively for transgenes that were found to fit in Mendelian 3:1 ratio expected for a selfed population (**Table 2**).

BC_2F_3 progeny plants belonging to seven BC_2F_2 plants, namely 2, 8, 20, 26, 33, 39, and 44, were assessed for agronomic performance. The results showed that the mean number of pods and seed yield of BC_2F_3 progeny plants derived from BC_2F_2 plant no. 20 and homozygous BC_2F_2 plants, namely 26, 39, and 44 were 53.67 ± 1.53, 54.00 ± 3.00, 55.00 ± 4.00, 53.67 ± 3.78, and 15.64 ± 0.51 g, 15.33 ± 0.98 g, 15.74 ± 0.62 g, 14.84 ± 0.53 g, respectively were statistically similar to mean number of pods (54.67 ± 3.05) and seed yield (15.90 ± 0.56 g) of recurrent parent (PBG 7) (**Table 3**; **Supplementary Table S5**).

The bioassay of BC_2F_3 progeny plants revealed 53.33–100% *H. armigera* larval mortality and variable (<10–40%) leaf and pod feeding damage; amongst these, the progeny of homozygous BC_2F_2 plants displayed 100% mortality with negligible (<10%) leaf and pod feeding damage (**Table 1**; **Figure 4A,B**). The larval mortality in transgenic donor parent and BC_2F_3 progeny plants was similar, whereas the non-transgenic recipient parent displayed no larval mortality and significant (51 to 60%) leaf, pod feeding damage (**Figure 4C**).

DISCUSSION

The elite, commercial chickpea cultivars susceptible to pod borer were converted into resistant by introgressing cry1Ac from transgenic lines through marker-assisted backcross

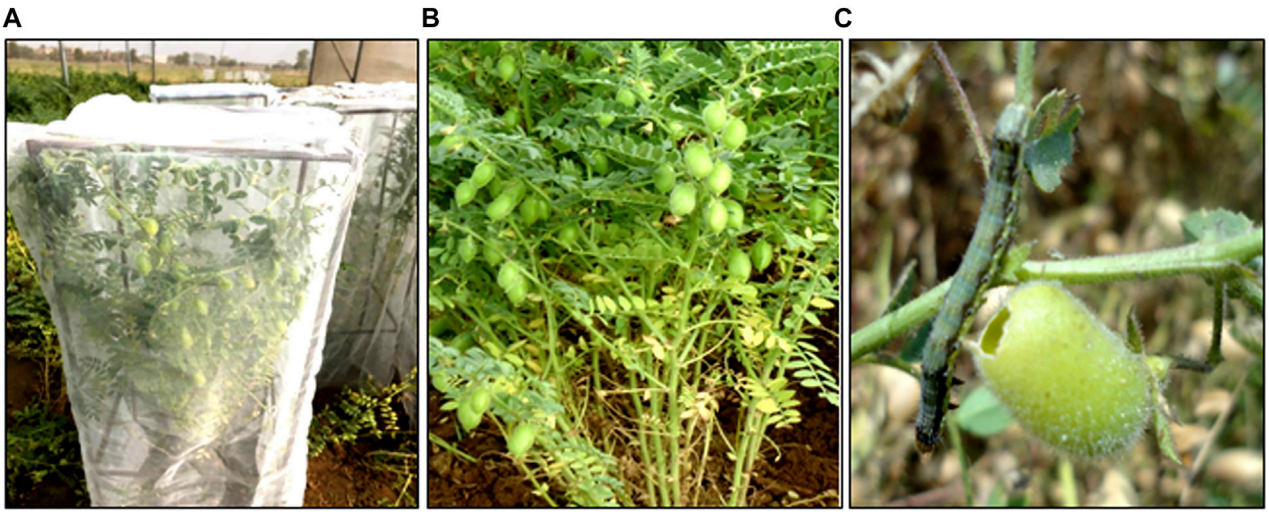


FIGURE 4 | Bioassay on BC_2F_3 plants obtained from Cross C (PBG7 × BS 100E) expressing cry1Ac for toxicity to *H. armigera* through whole plant screening. **(A)** Caged plants displaying healthy leaves and pods. **(B)** Closer view of plant showing healthy leaves and pods. **(C)** Non-transgenic recipient parent PBG7 exhibiting damaged pod and surviving larva.

breeding. F_1 plants and their backcross populations i.e., BC_1F_1, BC_1F_2, BC_1F_3, BC_2F_2, and BC_2F_3, exhibited up to 100% *H. armigera* neonatal larval mortality with agronomic performance similar to that of the recurrent parent. The high larval mortality was a result of Cry1Ac protein accumulation up to 11.00 $\mu g\, g^{-1}$ in backcross populations; Bt protein concentration as low as 0.9–3.1 $\mu g\, g^{-1}$ is reported to be highly insecticidal to corn earworm, *Helicoverpa* spp. in backcross populations of *Brassica napus* lines × wild *B. rapa* (Halfhill et al., 2001; Zhu et al., 2004). Zhang et al. (2012) and Chen et al. (2018) demonstrated that an even lower Cry1Ac concentration (0.5 to 1.2 $\mu g\, g^{-1}$) in the artificial diet of *H. armigera* larvae induced distinct histopathological changes in goblet cells of larval midgut epithelial lining, such as breakage of microvilli, endoplasmic reticulum, disorganization of mitochondria and chromatin, 2–36 h after Cry1Ac ingestion that eventually caused mortality.

BC_1F_1 and BC_1F_2 populations derived from Cross B segregated in Mendelian ratios of 1:1 and 3:1, respectively, for cry1Ac under contained field conditions; similarly, BC_2F_3 progenies of hemizygous BC_2F_2 plants raised from Cross C also segregated in Mendelian ratio of 3:1, pointing toward stable inheritance of cry1Ac as a single dominant gene in plants of different backcross populations. The typical 3:1 segregation ratio in selfed population and 1:1 in backcross population (Peng et al., 1992; Datta et al., 1998) often results from the insertion of one copy of the foreign gene in the host genome. The introgression of cry1Ab transgene following marker-assisted breeding has been reported in BC_2F_2 and BC_1 generations of cotton and rice, respectively (Agbios-Agriculture & Biotechnology Strategies (Canada), Inc., 2007; Kiani et al., 2009). The recurrent parent genome recovery in BC_2F_2 plants was higher (91.3%) in this study as compared to 87.5% genetic similarity to the recurrent parent obtained after two backcrosses through conventional breeding (Venkatesh et al., 2015). A recurrent parent genome recovery of 95.9% in BC_2F_2 rice plants was reported using polymorphic SSR markers by Chukwu et al. (2020). The marker-assisted backcross breeding is a dynamic approach for conveniently recognizing plants that have recovered over 98% of the recurrent parent genome in two to three backcross generations depending upon the availability of polymorphic markers (Stojsin 2010). The similarity of BC_2F_2 plants for agronomic traits with the recurrent parent in our study pointed toward the recurrent parent genome recovery, suggesting that in a situation where a limited number of polymorphic markers is available, the phenotypic characterization for agronomic traits is important. Joseph et al. (2004) reported that phenotypic selection coupled with fewer polymorphic markers between the parental lines maximizes recurrent parent genome recovery. We observed that the agronomic performance of BC_2F_3 progeny plants (derived from homozygous BC_2F_2 plants 26, 39 and 44) for main traits i.e., the number of pods and seed yield was statistically similar to the recurrent parent. Likewise, marker-assisted breeding between β-carotene-rich inbred lines UMI1200β^+, UMI1230β^+ × HKI163 in maize resulted in the development of improved BC_2F_3 lines exhibiting agronomic traits e.g., cob weight and single plant yield similar to the recurrent parents (Chandran et al., 2019).

The distorted segregation ratios were detected in BC_1F_1 and BC_1F_2 populations developed from Cross A. The distorted ratios generally arise due to transgene inactivation (Matzke and Matzke 1995), low viability/fertilization ability of transgenic pollen (Zhang et al., 1996), reduced germination (Sachs et al., 1998), genetic background (Scott et al., 1998; Wu et al., 2002), recessive lethal (Scott et al., 1998) etc. In the

present study, segregation distortion in BC_1F_1 and BC_1F_2 populations might be a result of reduced germination and not due to 1) transgene inactivation: as cry1Ac amplicon was observed in 46 plants from a total of 130 BC_1F_1 plants, and 16 out of 190 BC_1F_2 plants 2) low viability/fertility of transgenic pollen: as the pollen from recipient parent PBG7 (and not from the transgenic line) was used to pollinate F_1 plants to obtain BC_1F_1 plants; and further both BC_1F_1 and BC_1F_2 populations had resulted from a cross between desi PBG7 and desi BS 100B. Wu et al. (2002) observed that crosses between *japonica* and *japonica* rice had no significant effect on segregation ratios of cry1Ab, whereas *japonica* × *indica* resulted in distorted gene segregation in F_2 population, 3) genetic background: as both PBG7 and BS 100B are desi chickpeas, or 4) recessive lethal: as BC_1F_1 plants were hemizygous in nature for cry1Ac. Our assumption of reduced germination responsible for distorted segregation ratios draws support from observations by Sachs et al. (1998) on non-Mendelian segregation of cry1Ac in F_2 populations derived from MON 249 × CAMD-E due to failure of a large number of F_2 seeds inheriting cry1Ac to germinate. They further suggested that reduced germination associated with the inheritance of cry1Ac in MON 249 plants was a result of direct insertion effect leading to silencing of one or more native genes. In our case, we conclude that reduced germination was possibly associated with the inheritance of cry1Ac gene present in BS 100B plants. cry1Ac integration in the genomes of homozygous BC_2F_2 plants and BC_2F_3 progeny plants is probably at the same position as backcross populations obtained from a single transformation event are reported to carry transgene at a constant position in the genomes (Bakó et al., 2013).

The genetic background of BC_1F_1 populations raised from Cross A and Cross B did not affect the transgene expression as Cry1Ac protein concentration in the two BC_1F_1 populations was similar to each other and transgenic donor line. This observation is consistent with reports on hybrids of *Bt* maize (Fearing et al., 1997; Bakó et al., 2013) and eggplant (Ripalda et al., 2012) producing a similar amount of Cry protein in backcross populations irrespective of genetic background. However, this might not always be true as Sachs et al. (1998) observed that cryIA gene expression in cotton lines was influenced by the background genotype. The backcross populations of chickpea F_1 plants displayed a high degree of resistance to pod borer as compared to PBG7 and L552, implying stable expression of Cry1Ac throughout different generations.

In conclusion, cry1Ac was introgressed from transgenic chickpea lines into commercial cultivars through marker-assisted backcross breeding for imparting pod borer resistance; consequently, the backcross populations exhibited up to 100% *H. armigera* larval mortality. The BC_2F_2 plants homozygous for cry1Ac with high recurrent parent phenome recovery were identified; their BC_2F_3 progeny plants displaying agronomic performance similar to the recurrent parent are a valuable source of *H. armigera* resistance and can be used in chickpea 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.

AUTHOR CONTRIBUTIONS

AK analyzed and interpreted data on foreground selection, background selection, insect bioassay, agronomic traits, prepared the original draft, and edited the manuscript. US carried out foreground selection, background selection, and data compilation. SrS conceptualized the project, arranged funding, attempted crosses, monitored backcrosses and backcross populations, and edited the manuscript. RS reared *H. armigera* neonatal larvae, carried out insect bioassay, and collected data. YV planned and coordinated background selection. StS and KK carried out ELISA and interpreted the results. PM helped in analyzing segregation data. IS and SB attempted backcrosses, raised backcross populations, and collected agronomic data. BS provided seed of transgenic lines. JS planned and coordinated foreground selection and interpreted the data, prepared the original draft, and edited the manuscript. All the authors approved the final version of the manuscript.

FUNDING

This work was supported by the Department of Biotechnology, Government of India, with grant number BT/ISCB/Chickpea-PAU/2016 dated 18.5.2017 on "Introgression of Bt-genes (cry1Ac) into elite chickpea lines through backcross breeding and evaluation of introgressed lines against *Helicoverpa armigera*."

ACKNOWLEDGMENTS

All authors thankfully acknowledge infrastructural support provided by the Pulses Section, Department of Plant Breeding & Genetics, School of Agricultural Biotechnology, PAU, Ludhiana, and Regional Research Station, Faridkot, for carrying out the research work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.847647/full#supplementary-material

REFERENCES

Acharjee, S., Sarmah, B. K., Kumar, P. A., Olsen, K., Mahon, R., Moar, W. J., et al. (2010). Transgenic Chickpeas (*Cicer Arietinum* L.) Expressing a Sequence-Modified *cry2Aa* Gene. *Plant Sci.* 178, 333–339. doi:10.1016/j.plantsci.2010.02.001

Agbios-Agriculture & Biotechnology Strategies (Canada), Inc. (2007). Pre-market Environmental Risk Assessment of Transgenic Plants: A Case-Study Approach Utilizing Mon 15985 Cotton, 1–81.

Armes, N. J., Bond, G. S., and Cooters, R. J. (1992). *The Laboratory Culture and Development of Helicoverpa Armigera*. Chatham, United Kingdom: Natural Resources Institute.

Bakó, A., Gell, G., Zámbó, Á., Spitkó, T., Pók, I., Pintér, J., et al. (2013). Monitoring Transgene Expression Levels in Different Genotypes of Field Grown maize (*Zea mays* L.). *South Afr. J. Bot.* 84, 6–10. doi:10.1016/j.sajb.2012.09.005

Bravo, A., Gill, S. S., and Soberón, M. (2008). Mode of Action of *Bacillus Thuringiensis* Cry and Cyt Toxins and Their Potential for Insect Control. *Toxicon* 49, 423–435. doi:10.1016/j.toxicon.2006.11.022

Chakrabarti, S. K., Mandaokar, A., Kumar, P. A., and Sharma, R. P. (1998). Efficacy of Lepidopteran Specific δ-Endotoxins ofBacillus thuringiensisagainstHelicoverpa Armigera. *J. Invertebr. Pathol.* 72, 336–337. doi:10.1006/jipa.1998.4786

Chandran, S., Pukalenthy, B., Adhimoolam, K., Manickam, D., Sampathrajan, V., Chocklingam, V., et al. (2019). Marker-Assisted Selection to Pyramid the Opaque-2 (O2) and β-Carotene (crtRB1) Genes in Maize. *Front. Genet.* 10, 859. doi:10.3389/fgene.2019.00859

Chen, W., Liu, C., Lu, G., Cheng, H., Shen, Z., and Wu, K. (2018). Effects of Vip3AcAa+Cry1Ac Cotton on Midgut Tissue in *Helicoverpa Armigera* (Lepidoptera: Noctuidae). *J. Insect Sci.* 18, 1–6. doi:10.1093/jisesa/iey075

Cherry, A. J., Rabindra, R. J., Parnell, M. A., Geetha, N., Kennedy, J. S., and Grzywacz, D. (2000). Field Evaluation of Helicoverpa Armigera Nucleopolyhedrovirus Formulations for Control of the Chickpea Pod-Borer, H. Armigera (Hubn.), on Chickpea (Cicer Arietinum Var. Shoba) in Southern India. *Crop Prot.* 19, 51–60. doi:10.1016/s0261-2194(99)00089-7

Chukwu, S. C., Rafii, M. Y., Ramlee, S. I., Ismail, S. I., Oladosu, Y., Muhammad, I. i., et al. (2020). Recovery of Recurrent Parent Genome in a Marker-Assisted Backcrossing against rice Blast and Blight Infections Using Functional Markers and SSRs. *Plants* 9, 1411. doi:10.3390/plants9111411

Datta, K., Vasquez, A., Tu, J., Torrizo, L., Alam, M. F., Oliva, N., et al. (1998). Constitutive and Tissue-specific Differential Expression of the cryIA(b) Gene in Transgenic rice Plants Conferring Resistance to rice Insect Pest. *Theor. Appl. Genet.* 97, 20–30. doi:10.1007/s001220050862

Fearing, P. L., Brown, D., Vlachos, D., Meghji, M., and Privalle, L. (1997). Quantitative Analysis of CryIA(b) Protein Expression in Bt maize Plants, Tissues, and Silage and Stability of Expression over Successive Generations. *Mol. Breed.* 3, 169–176. doi:10.1023/a:1009611613475

Fitt, G. P. (1989). The Ecology of *Heliothis* Species in Relation to Agroecosystems. *Annu. Rev. Entomol.* 34, 17–53. doi:10.1146/annurev.en.34.010189.000313

Guo, W., Zhang, T., Zhu, X., and Pan, J. (2005). Modified Backcross Pyramiding Breeding with Molecular Marker-Assisted Selection and its Applications in Cotton. *Zuo Wu Xue Bao* 31, 963–970.

Halfhill, M. D., Richards, H. A., Mabon, S. A., and Stewart, C. N. (2001). Expression of GFP and Bt Transgenes in *Brassica Napus* and Hybridization with *Brassica Rapa*. *Theor. Appl. Genet.* 103, 659–667. doi:10.1007/s001220100613

Hazarika, N., Acharjee, S., Boruah, R. R., Babar, K., Parimi, S., Char, B., et al. (2019). Enhanced Expression of Arabidopsis Rubisco Small Subunit Gene Promoter Regulated Cry1Ac Gene in Chickpea Conferred Complete Resistance to Helicoverpa Armigera. *J. Plant Biochem. Biotechnol.*, 1–11. doi:10.1007/s13562-019-00531-1

Hospital, F., and Charcosset, A. (1997). Marker-assisted Introgression of Quantitative Trait Loci. *Genetics* 147, 1469–1485. doi:10.1093/genetics/147.3.1469

Jenkins, N. E., and Grzywacz, D. (2000). Quality Control of Fungal and Viral Biocontrol Agents - Assurance of Product Performance. *Biocontrol Sci. Tech.* 10, 753–777. doi:10.1080/09583150020011717

Joseph, M., Gopalakrishnan, S., Sharma, R. K., Singh, V. P., Singh, A. K., K. Singh, N., et al. (2004). Combining Bacterial Blight Resistance and Basmati Quality Characteristics by Phenotypic and Molecular Marker-Assisted Selection in rice. *Mol. Breed.* 13, 377–387. doi:10.1023/b:molb.0000034093.63593.4c

Jukanti, A. K., Gaur, P. M., Gowda, C. L. L., and Chibbar, R. N. (2012). Nutritional Quality and Health Benefits of Chickpea (Cicer arietinumL.): a Review. *Br. J. Nutr.* 108, S11–S26. doi:10.1017/s0007114512000797

Kar, S., Basu, D., Das, S., Ramkrishnan, N. A., Mukherjee, P., Nayak, P., et al. (1997). Expression of *cryIA(c)* Gene of *Bacillus Thuringiensis* in Transgenic Chickpea Plants Inhibits Development of Pod-Borer (*Heliothis Armigera*) Larvae. *Transgenic Res.* 6, 177–185. doi:10.1023/a:1018433922766

Khatodia, S., Kharb, P., Batra, P., Kumar, P. A., and Chowdhury, V. K. (2014). Molecular Characterization of Bt Chickpea (*Cicer Arietinum* L.) Plants Carrying *cry1Aa3* Gene. *Int. J. Curr. Microbiol. App. Sci.* 3, 632–642.

Kiani, G., Nematzadeh, G. A., Ghareyazie, B., and Sattari, M. (2009). Genetic Analysis of *cry1Ab* Gene in Segregating Populations of rice. *Afr. J. Biotechnol.* 8, 3703–3707. doi:10.5897/AJB09.483

Kumar, M., Yusuf, M. A., Nigam, M., and Kumar, M. (2018). An Update on Genetic Modification of Chickpea for Increased Yield and Stress Tolerance. *Mol. Biotechnol.* 60, 651–663. doi:10.1007/s12033-018-0096-1

Mallikarjuna, N., Sharma, H. C., and Upadhyaya, H. D. (2007). Exploitation of Wild Relatives of Pigeonpea and Chickpea for Resistance to *Helicoverpa Armigera*. *J. SAT Agric. Res.* 3, 4.

Matzke, M. A., and Matzke, A. (1995). How and Why Do Plants Inactivate Homologous (Trans)genes? *Plant Physiol.* 107, 679–685. doi:10.1104/pp.107.3.679

Mehrotra, M., Singh, A. K., Sanyal, I., Altosaar, I., and Amla, D. V. (2011). Pyramiding of Modified cry1Ab and cry1Ac Genes of Bacillus Thuringiensis in Transgenic Chickpea (Cicer Arietinum L.) for Improved Resistance to Pod Borer Insect Helicoverpa Armigera. *Euphytica* 182, 87. doi:10.1007/s10681-011-0501-3

Merga, B., and Haji, J. (2019). Economic Importance of Chickpea: Production, Value, and World Trade. *Cogent Food Agric.* 5, 1615718. doi:10.1080/23311932.2019.1615718

Patil, S. B., Goyal, A., Chitgupekar, S. S., Kumar, S., and El-Bouhssini, M. (2017). Sustainable Management of Chickpea Pod Borer. A Review. *Agron. Sustain. Dev.* 37, 20. doi:10.1007/s13593-017-0428-8

Peng, J., Kononowicz, H., and Hodges, T. K. (1992). Transgenic Indica rice Plants. *Theoret. Appl. Genet.* 83-83, 855–863. doi:10.1007/bf00226708

Rao, G. V. R., Rameshwar Rao, V., and Ghaffar, M. A. (2013). *Handbook On Chickpea And Pigeonpea Insect Pests Identification And Management, Information Bulletin No. 57. International Crops Research Institute for the Semi-arid Tropics*. Patancheru, Andhra Pradesh.

Ripalda, R. R., Hautea, D. M., Narciso, J. O., and Canama, A. O. (2012). Inheritance and Expression of *cry1Ac* Gene in Bt Eggplant Backcross Populations in the Philippines. *Philipp. J. Crop Sci.* 37, 1–9.

Sachs, E. S., Benedict, J. H., Stelly, D. M., Taylor, J. F., Altman, D. W., Berberich, S. A., et al. (1998). Expression and Segregation of Genes Encoding CrylA Insecticidal Proteins in Cotton. *Crop Sci.* 38, 1–11. doi:10.2135/cropsci1998.0011183x003800010001x

Sandhu, J. S., Gupta, S. K., Singh, I., and Bains, T. S. (2012). L 552: A New Bold Seeded Variety of *Kabuli* Gram. *J. Res. Punjab Agric. Univ.* 49, 292.

Sanyal, I., Singh, A. K., Kaushik, M., and Amla, D. V. (2005). *Agrobacterium*-mediated Transformation of Chickpea (*Cicer Arietinum* L.) with *Bacillus Thuringiensis cry1Ac* Gene for Resistance against Pod Borer Insect *Helicoverpa Armigera*. *Plant Sci.* 168, 1135–1146. doi:10.1016/j.plantsci.2004.12.015

Scott, A., Woodfield, D., and White, D. W. R. (1998). Allelic Composition and Genetic Background Effects on Transgene Expression and Inheritance in white clover. *Mol. Breed.* 4, 479–490. doi:10.1023/a:1009601321343

Sharma, H. C., Pampapathy, G., Dhillon, M. K., and Ridsdill-Smith, J. T. (2005a). Detached Leaf Assay to Screen for Host Plant Resistance to *Helicoverpa Armigera*. *J. Econ. Entomol.* 98, 568–576. doi:10.1093/jee/98.2.568

Sharma, H. C., Pampapathy, G., and Kumar, R. (2005b). Standardization of Cage Techniques to Screen Chickpeas for Resistance to *Helicoverpa Armigera* (Lepidoptera: Noctuidae) in Greenhouse and Field Conditions. *J. Econ. Entomol.* 98, 210–216. doi:10.1093/jee/98.1.210

Singh, S., Singh, I., Sandhu, J. S., Gupta, S. K., Bains, T. S., Rathore, P., et al. (2015). PBG7: A New High Yielding Variety of *Desi* Gram (*Cicer Arietinum* L.). *Agric. Res. J.* 52, 212–213. doi:10.5958/2395-146x.2015.00024.1

Stojsin, D. (2010). "Transgenes: Plant Breeding," in *Encyclopedia of Biotechnology in Agriculture and Food*. Editors D. R. Heldman, D. G. Hoover, and M. B. Wheeler. 1st ed. (Boca Raton: CRC Press). doi:10.1081/E-EBAF

Venkatesh, T. V., Cook, K., Liu, B., Perez, T., Willse, A., Tichich, R., et al. (2015). Compositional Differences between Near-Isogenic GM and Conventional maize Hybrids Are Associated with Backcross Breeding Practices in Conventional Breeding. *Plant Biotechnol. J.* 13, 200–210. doi:10.1111/pbi.12248

Wang, Z., Yu, C., and Jiang, L. (2012). Segregation and Expression of Transgenes in the Progenies of *Bt* Transgenic rice Crossed to Conventional rice Varieties. *Afr. J. Biotechnol.* 11, 7812–7818. doi:10.5897/ajb12.119

Wu, G., Cui, H., Ye, G., Xia, Y., Sardana, R., Cheng, X., et al. (2002). Inheritance and Expression of the *cry1Ab* Gene in Bt (*Bacillus Thuringiensis*) Transgenic rice. *Theor. Appl. Genet.* 104, 727–734. doi:10.1007/s001220100689

Zhang, S., Warkentin, D., Sun, B., Zhong, H., and Sticklen, M. (1996). Variation in the Inheritance of Expression Among Subclones for Unselected (*uidA*) and Selected (*Bar*) Transgenes in maize (*Zea mays* L.). *Theor. Appl. Genet.* 92, 752–761. doi:10.1007/bf00226098

Zhang, Y., Liang, G., Zhang, L., and Wei, J. (2012). Pathological Changes in Midgut Tissues of Larvae of the Cotton Bollworm, *Helicoverpa Armigera* (Lepidoptera: Noctuidae), after Feeding Vip3Aa Protein. *Acta Entomol. Sinica* 55, 869–876.

Zhu, B., Lawrence, J. R., Warwick, S. I., Mason, P., Braun, L., Halfhill, M. D., et al. (2004). Stable *Bacillus Thuringiensis* (Bt) Toxin Content in Interspecific F_1 and Backcross Populations of Wild *Brassica Rapa* after Bt Gene Transfer. *Mol. Ecol.* 13, 237–241. doi:10.1046/j.1365-294x.2004.02018.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 reviewer MST declared a past co-authorship with the author(s) SS to the handling editor.

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 © 2022 Kaur, Sharma, Singh, Singh, Vikal, Singh, Malik, Kaur, Singh, Bindra, Sarmah and Sandhu. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fgene.2022.847612

The bZIP Transcription Factor Family in Adzuki Bean (*Vigna Angularis*): Genome-Wide Identification, Evolution, and Expression Under Abiotic Stress During the Bud Stage

*Zhengong Yin, Xianxin Meng, Yifan Guo, Shuhong Wei, Yongcai Lai and Qiang Wang**

Crop Resources Institute of Heilongjiang Academy of Agricultural Sciences Harbin, Heilongjiang, China

Adzuki bean (*Vigna angularis*) is an important dietary legume crop that was first cultivated and domesticated in Asia. Currently, little is known concerning the evolution and expression patterns of the basic leucine zipper (bZIP) family transcription factors in the adzuki bean. Through the PFAM search, 72 bZIP members of adzuki bean (VabZIP) were identified from the reference genome. Most of them were located on 11 chromosomes and seven on an unknown chromosome. A comprehensive analysis, including evolutionary, motifs, gene structure, *cis*-elements, and collinearity was performed to identify VabZIP members. The subcellular localization results showed VabZIPs might locate on the nuclear. Quantitative real-time PCR (qRT-PCR) analysis of the relative expression of VabZIPs in different tissues at the bud stage revealed that VabZIPs had a tissue-specific expression pattern, and its expression was influenced by abiotic stress. These characteristics of *VabZIPs* provide insights for future research aimed at developing interventions to improve abiotic stress resistance.

Keywords: adzuki bean, bZIP members, analysis, relative expression, bud stage, abiotic stress

OPEN ACCESS

Edited by:
Nasya Borisova Tomlekova, Maritsa Vegetable Crops Research Institute (MVCRI), Bulgaria

Reviewed by:
Eleonora Cominelli, Italian National Research Council, Italy
Aamir Raina, Aligarh Muslim University, India

****Correspondence:***
Qiang Wang
shi.yongdou@163.com

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

Received: *03 January 2022*
Accepted: *05 April 2022*
Published: *25 April 2022*

Citation:
Yin Z, Meng X, Guo Y, Wei S, Lai Y and Wang Q (2022) The bZIP Transcription Factor Family in Adzuki Bean (Vigna Angularis): Genome-Wide Identification, Evolution, and Expression Under Abiotic Stress During the Bud Stage. Front. Genet. 13:847612. doi: 10.3389/fgene.2022.847612

INTRODUCTION

Transcription factors (TFs), which constitute approximately 8% of the protein-encoding regulators in eukaryotic genomes, are critical transcriptional regulatory factors (Pruneda-Paz et al., 2014). Therefore, functional characterization of transcription factors (TFs) is critical for understanding transcriptional regulatory networks and biological processes (Liu et al., 2014). The basic leucine zipper (bZIP) family is one of the largest and most diverse TF families (Pérez-Rodríguez et al., 2010). The bZIP domain is highly conserved and contains two structural features located on a contiguous α-helix i.e., the leucine zipper composed of several heptad repeats of Leu or other bulky hydrophobic amino acids for dimerization specificity, and the N-x7-R/K-x9 domain for specific binding (Jakoby et al., 2002; Lee et al., 2006; Nijhawan et al., 2008). Apart from the bZIP domain, several domains of the bZIP family have been found to function as transcriptional activators (Liao et al., 2008). To bind DNA, half of the basic region in the N-terminal binds double-stranded DNA, and half of the Leu zipper in the C-terminal undergoes dimerization, leading to the formation of a superimposed coiled structure (Ellenberger et al., 1992).

Members of the bZIP transcription factor family are involved in the regulation of growth and developmental processes such as seed germination, embryogenesis, flower and vascular

development, hormonal control, and senescence (Jakoby et al., 2002; Schütze et al., 2008; Toh et al., 2012; Sornaraj et al., 2016). Overexpression of *OsbZIP23*, a member of bZIP in rice (*Oryza sativa*), rescued the pre-harvest budding phenotype and the decrease in expression of genes associated with ABA signaling in transgenic plants (Song et al., 2020). *CAREB1*, an important trans-acting factor of bZIP members, was found to regulate somatic embryogenesis in carrot (*Daucus carota*) (Guan et al., 2009). Eleven *TabZIP* genes in wheat (*Triticum aestivum*) were highly expressed in anthers, suggesting that they were involved in flower development (Li D. et al., 2015). In *Arabidopsis*, a bZIP transcription factor that control monopteros (MP) output and modulate vascular gene expression (Smit et al., 2020). The bZIP Transcription factor PERIANTHIA interacts with a variety of developmental pathways, including light and plant hormones, both of which participate in meristem formation (Maier et al., 2011). Furthermore, the bZIP members regulate response to abiotic/biotic stresses such as drought, salt, hypoxia, cold, pests, and diseases (Uno et al., 2000; Shimizu et al., 2005; Zander et al., 2012; Alves et al., 2013; E et al., 2014; Amorim et al., 2017). *GmbZIP44* and *GmbZIP62*, the bZIP genes of soybean (*Glycine max*), conferred tolerance to salt and freezing stress in transgenic *Arabidopsis* plants (Liao et al., 2008). Overexpression of *CabZIP25*, a member of bZIP in pepper (*Capsicum annuum*), enhanced salt tolerance in transgenic *Arabidopsis* and promoted salt sensitivity by decreasing virus induced gene silencing (VIGS) expression in pepper (Gai et al., 2020). The study by Hsieh et al. (2010) showed that *SlAREB*, a member of bZIP in tomato (*Solanum lycopersicum*), regulated stress-responsive genes and improved water logging deficit and salt stress response. Elsewhere, it was reported that *AREB1*, an *Arabidopsis* bZIP transcription factor, conferred tolerance to water deficit (including drought and flooding stresses) in modified soybeans overexpressing *AREB1* (Fuhrmann-Aoyagi et al., 2021). PPI_1, a bZIP in pepper, regulated expression of genes involved in defense mechanisms (Lee et al., 2002).

The application of genome sequencing has led to identification of bZIP family members (Jakoby et al., 2002), in *Arabidopsis* (Jakoby et al., 2002), rice (E et al., 2014), *Carthamus tinctorius* (Li H. et al., 2020), Chinese jujube (*Ziziphus jujuba*) (Zhang Q. et al., 2020), Olive (*Olea europaea*) (Rong et al., 2020), common bean (*Phaseolus vulgaris*) (Zhang et al., 2021), and potato (*Solanum tuberosum*) (Herath and Verchot, 2020). However, few studies have investigated bZIP family members in adzuki bean (*Vigna angularis*). Adzuki bean (*Vigna angularis*) is an important dietary legume crop that was first cultivated in China (Han et al., 2005). Its grains have high protein content, a low-fat content, and high iron content. They contain several bioactive substances such as triterpenoids, flavonoids, and saponins. It was traditionally used as an iron supplement, to remove damp and swelling (Amarowicz et al., 2008; Yang et al., 2015). Being a sensitive species, adzuki beans are particularly vulnerable to environmental stressors such as cold, drought, salt, and heavy metals (Srivastava et al., 2018; Li W.-Y. et al., 2020). In this study, bZIP members in the adzuki bean were identified, and characterized in terms of phylogeny and evolutionary expansion in different tissues under different stress conditions such as drought (D), cold (C), salt (NaCl) and heavy metal ($CdCl_2$). The findings will provide new insights about bZIP members which can be applied in resistance breeding.

MATERIALS AND METHODS

Identification of bZIP Members in *Vigna Angularis*

The basic information for the reference genome (including genes, cDNAs, and proteins) in adzuki bean (Vigan1.1) was obtained from the Esembl plant's database (https://plants.ensembl.org/Vigna angularis/Info/Index). The bZIP domain information was obtained from the PFAM database (http://pfam.xfam.org/), with PF00170 as the search key. The bZIP members in adzuki bean (*Vigna angularis*) were identified using the HMMER software (Finn et al., 2015) and screened using a database that included the ExPASy Proteomics Server (http://www.expasy.org) (Hoogland et al., 2008) and P3DB (http://www.p3db.org) (Yao and Xu, 2017). After deduplication, the remaining bZIP members were considered to be members of the bZIP family in adzuki bean, and were named VabZIP. VabZIPs were named according to their location in the reference genome in the Esembl database, which was determined using the TBtools software (Chen et al., 2020).

Analysis of VabZIP Members

Protein sequences of the VabZIP members were aligned using MEGA X (Kumar et al., 2018) while bootstrap values (1,000 replicates) were used for the maximum likelihood analysis. MEGA predicted the optimal model. Ten motifs from VabZIP members were identified using the MEME tool (Bailey et al., 2009), with an E-value of less than $1e^{-20}$ for motifs containing 10–50 amino acids. Gene structures for VabZIP members were analyzed using GSDS (Hu et al., 2015) and Gene-wise (Simmons et al., 2019), in which the coordinates corresponded to DNA and protein sequences. *Cis*-acting elements of VabZIP members were identified and their functions predicted by the plantCARE software (Lescot et al., 2002). Gene duplication events for VabZIP members were evaluated by MCScanX (Wang et al., 2013) and circus (Krzywinski et al., 2009) software. Subcellular locations for VabZIP members were predicted by the CELLO database (http://cello.life.nctu.edu.tw/) (Yu et al., 2006). Expression data for orthologous genes of VabZIP members in *Arabidopsis* and soybean (*Glycine max*) were obtained from the phytozome database (https://phytozome-next.jgi.doe.gov/).

Plant Materials, Stress Concentrations, and qRT-PCR Analysis

Plant materials for this study were "Longxiaodou 5", which was provided by the Institute of Crop Resources, Heilongjiang

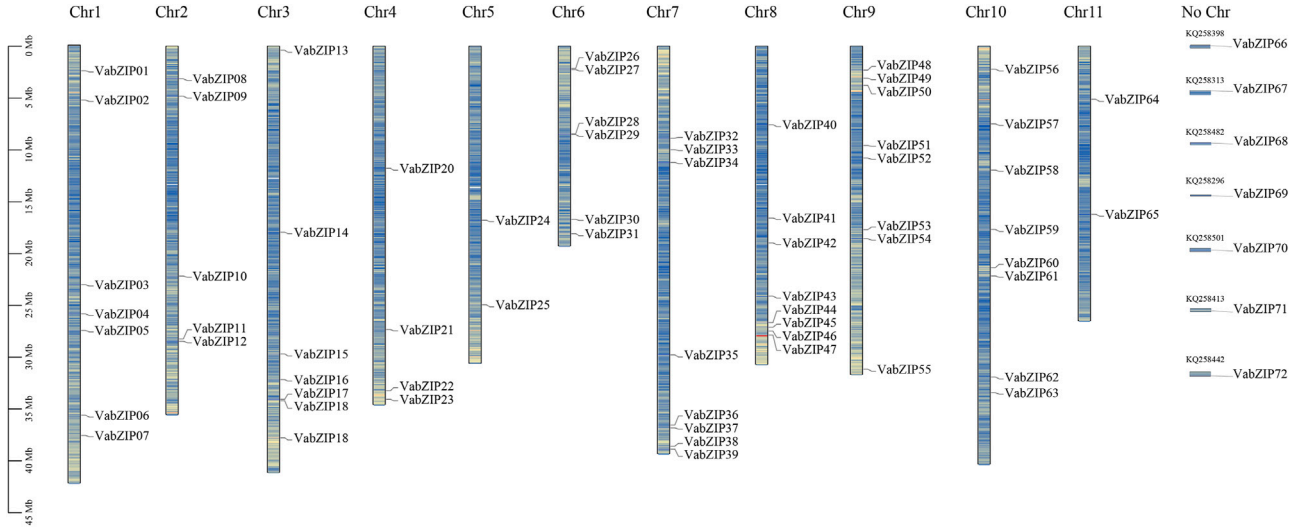


FIGURE 1 | Locations of VabZIP members. The 11 pillars correspond to the 11 chromosomes, whereas No Chr depict members lacking a chromosome. The blue lines indicate gene density on chromosomes.

Academy of Agricultural Sciences (Harbin, Heilongjiang, China). For the seedlings to bud, they were incubated at 26 ˚C without light (Qi et al., 2021).

During treatment, the following stressors were prioritized: drought, salt, cold, and heavy metals. Salt stress concentration was 70 mmol/L (Zhang Y. et al., 2020) while heavy metal stress concentration was 0.5 mg/L $CdCl_2$ (Zhao et al., 2020). A temperature of 4 ˚C was used to induce cold stress (Wang et al., 2020) while 15% PEG was the concentrate drought stress (Ahmad et al., 2020). The stresses were separately induced on the third day, with water treatment used as the control (CK). The hypocotyl, radicle, cotyledon, and germs were collected as samples for tissue-specific analysis expressions. The radicles were collected in response to these abiotic stress treatments. The RNA Easy Fast Kit (DP452, Tiangen, Beijing) was used for sample RNA extraction, which was used for cDNA synthesis using HiScript SuperMix (R223-01, Vazyme, Nanjing). The *VabZIPs* primers were designed using the Primer premier5.0 software (PREMIER Biosoft, San Francisco, United States) while *Va-actin* was used as the reference control gene (Li W.-Y. et al., 2020). qRT-PCR analyses for expressions of three biological replicates of each VabZIP member were performed using the Light Cycler system (Roche 480II, Roche, Switzerland) and *TransStart*® Top Green qPCR SuperMix (AQ131-04, TransGen Biotech, Beijing). Relative expressions were calculated as described by Livak and Schmittgen (2001).

Subcellular Localization

The coding sequence (CDS) of VabZIP members (VabZIP17 and VabZIP56) was PCR amplified from the cDNAs, which without stop codon. The primers used for cloning the VabZIP17 and VabZIP56 was shown in **Supplementary Table S2**. Then, the sequence was cloned into the vector, which had the green fluorescent protein (GFP) tag and a CaMV35S promoter. The VabZIP17-GFP and VabZIP56-GFP construct were transformed into Agrobacterium competent cells and transiently expressed in the leaves of Nicotiana benthamiana with the empty vector was used as a control. After injection for 2 days, the leaves were observed under a confocal laser microscope (TCS-SP8 Leica, Wetzlar, Germany) to find fluorescence signals (A1Si, Nikon, Japan).

RESULTS

Identification of bZIP Members in *Vigna angularis*

Following a HMMER-search of the bZIP domain, 72 members of the bZIP family were identified from the reference genome in the Esembl database (*Vigna angularis*), which had no duplications. These members were located on all *Vigna angularis* chromosomes. Eight of the members were located on chromosomes 7, 8, 9, and 10 while chromosome 5 had the fewest members (2). Seven members were located on an unknown chromosome, which may be positioned on the scaffold. The bZIP members were named based on their location (VabZIP1-VabZIP72) (**Figure 1**). Information on the VabZIP members is presented in **Supplementary Table S1**. Protein lengths of VabZIPs ranged from 80 to 773, with VabZIP56 having the longest protein (773) and CDS (2,322). Isoelectric points of VabZIP members ranged from 4.76–11.56, while their molecular weights ranged from 9,438.77 Kilodalton (Kd) to 84105.11 (Kd) (**Supplementary Table S3**).

Evolutionary Analysis of the VabZIP Members

Evolutionary history of VabZIP members was determined using the Maximum Likelihoodphy (ML) analysis, with the lg + g model predicted by MEGA X software used as the model

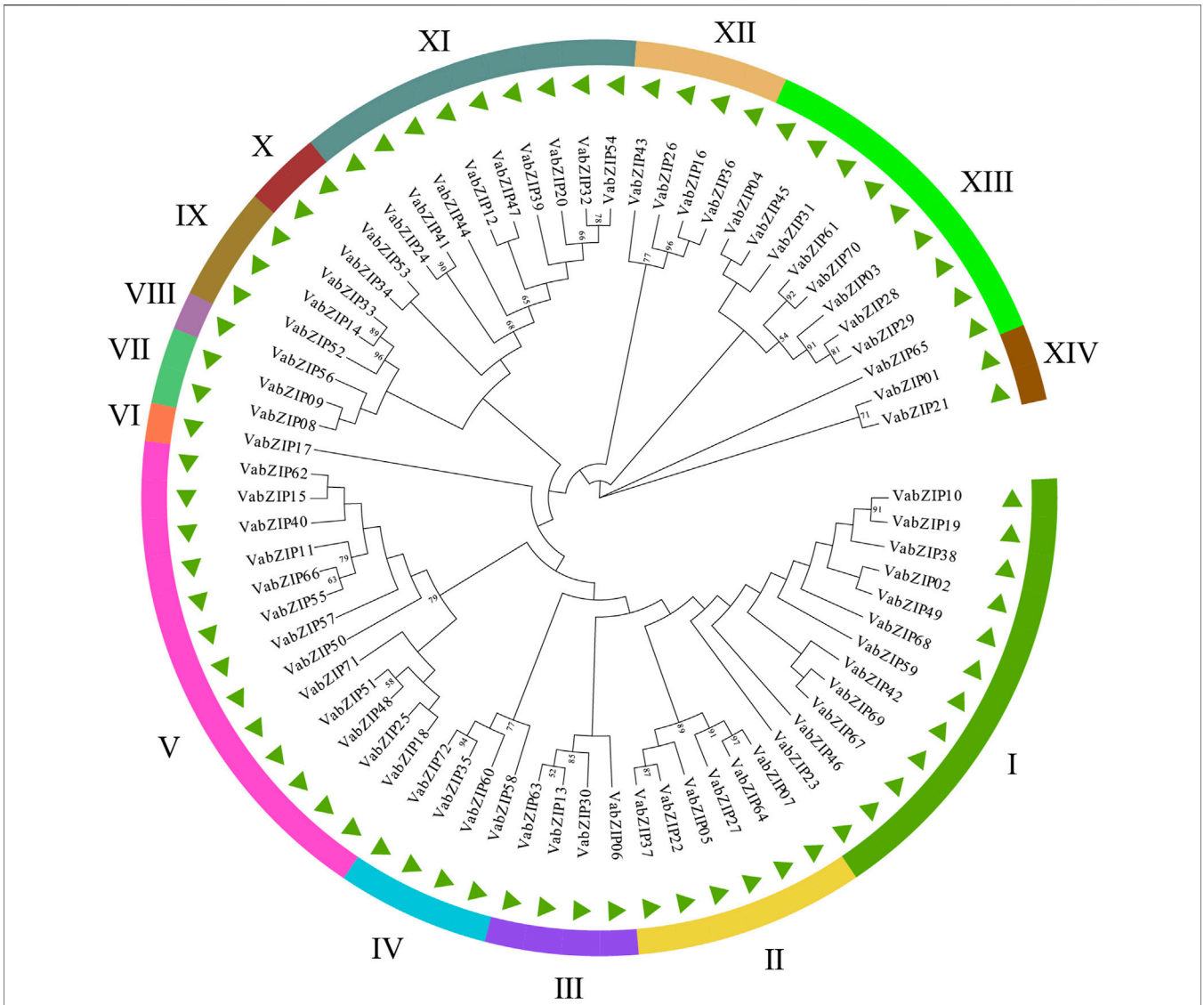


FIGURE 2 | Evolutionary analysis of VabZIPs. Different sub-families are painted in different colors.

for analysis. Findings from MEGA X analysis are presented in **Figure 2**. These 72 members were divided into 14 subfamilies, with sub-family V having the most VabZIP members (13), while sub-families VI and VIII each had only one VabZIP member, making them the least-membered sub-families. There were 4, 2, and 9 VabZIP members in subfamily III, IV, and XII, subfamily VII, X, and XIV, and subfamily XI and XIII, respectively.

Motifs and Structure in VabZIPs

Analysis of VabZIPs using the MEME software revealed ten motifs (**Figures 3A,B** and **Supplementary Figure S1**). Apart from subfamilies VII and VIII, motif-1 was found in most VabZIPs sub-families while motif-3 and motif-5 were only found in sub-family X. Motif 9 was found only in subfamily XII. The VabZIP members in each subfamily had similar motifs. Gene structures for VabZIP members were assessed by GSDS, which revealed exon and intron structures (**Figure 3C**). The bZIP structure was located above all members in these VabZIPs, and sub-families I to III members had shorter introns than members of the other subfamilies.

Evolution of bZIP Members in Two Species

The 72 VabZIPs were compared to sequences encoded by bZIP members from *Arabidopsis* to determine their evolutionary history, motif, and gene structure (**Figure 4** and **Supplementary Figure S2**). MEGA X predicted the lg + g model as the best model, and 14 subfamilies were defined based on the results in these two species, consistent with the evolutionary of VabZIPs (**Figure 2**). Each subfamily had bZIP members of these two species. Each member of the subfamily had comparable motifs and gene structures.

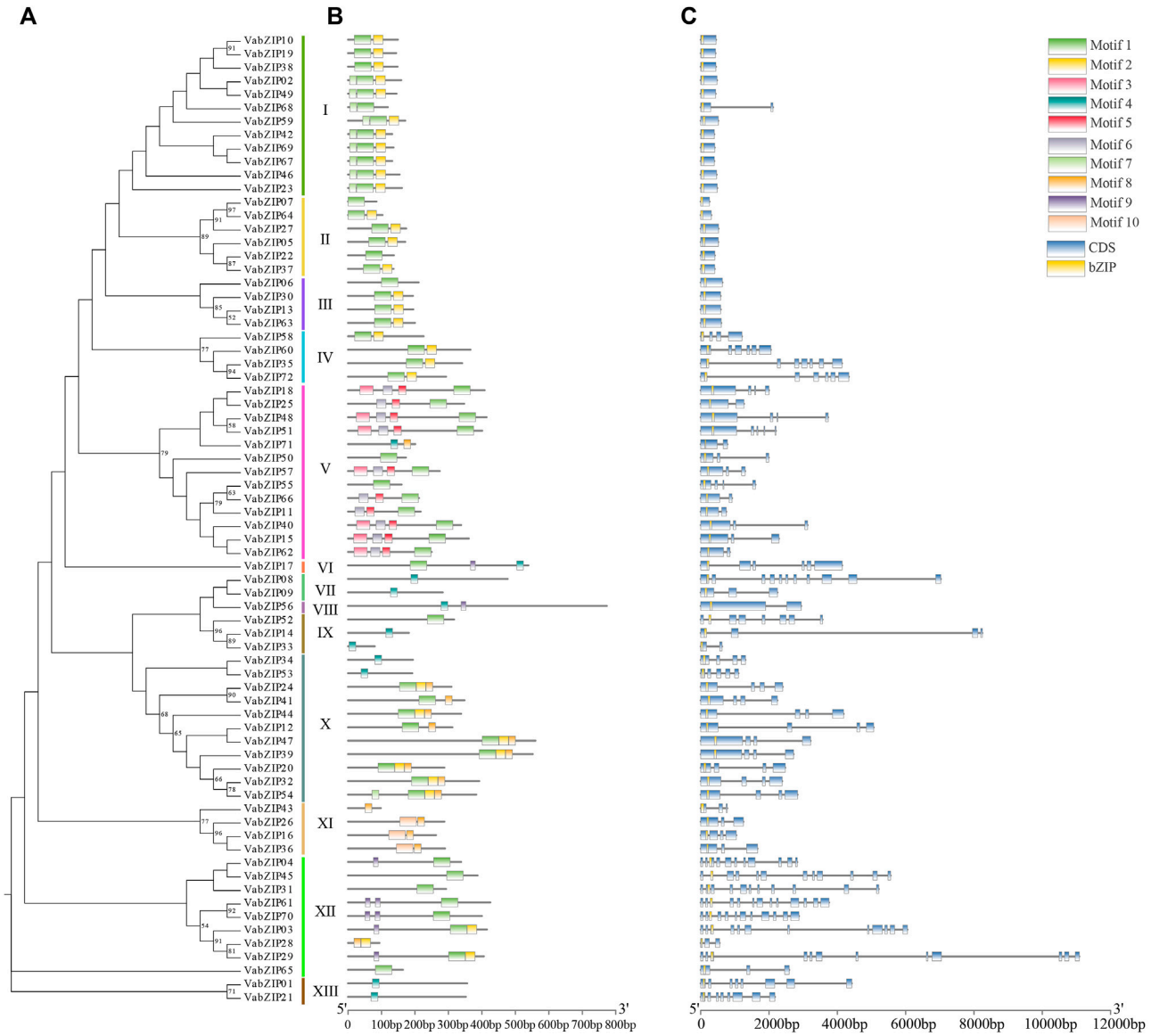


FIGURE 3 | Motifs and gene structure of VabZIPs. **(A)** Evolutionary analysis of VabZIPs. Different colors represent different subfamilies **(B)** The motifs of VabZIPs **(C)** Gene structure of VabZIPs. Blue squares represent CDS while yellow squares represent the bZIP structure.

Cis-Elements of *VabZIPs*

Cis-elements of *VabZIP* members were analyzed using the plantCARE software. PlantCARE predicted the functions of the ten *cis*-elements (**Supplementary Table S4**), which were divided into three categories: hormone responsiveness (red), environmental stress (blue), and germination (yellow). Hormone responsiveness elements, including TATC-box, P-box, and GARE-motif were involved in gibberellin responsiveness, while ABRE was the *cis*-acting element involved in abscisic acid responsiveness. Environmental stress elements, such as LTR were involved in low-temperature responsiveness, while MBS was involved in drought-inducibility. The RY-element and NON-box elements had seed-specific regulation function. These findings indicate that *VabZIP* family members are involved in hormone regulation, stress resistance, and seed germination (**Figure 5**).

Collinearity Analysis of *VabZIPs*

There was collinearity between ten pairs of *VabZIPs* with *VABZIP05* and *VABZIP22* being the most collinear with other *VabZIP* members (three pairs). *VaBZIP27* and *VaBZIP37* had two pairs of collinearity (**Figure 6A**). Twenty five collinearity pairs were identified between *VabZIP* members and *Arabidopsis*, with *VabZIP13*, *VabZIP23*, *VaBZIP26*, *VaBZIP46,* and *VaBZIP47* having two collinear members in *Arabidopsis*, implying that the 21 *VabZIP* members may have the same function as collinear *Arabidopsis genes* (**Figure 6B**).

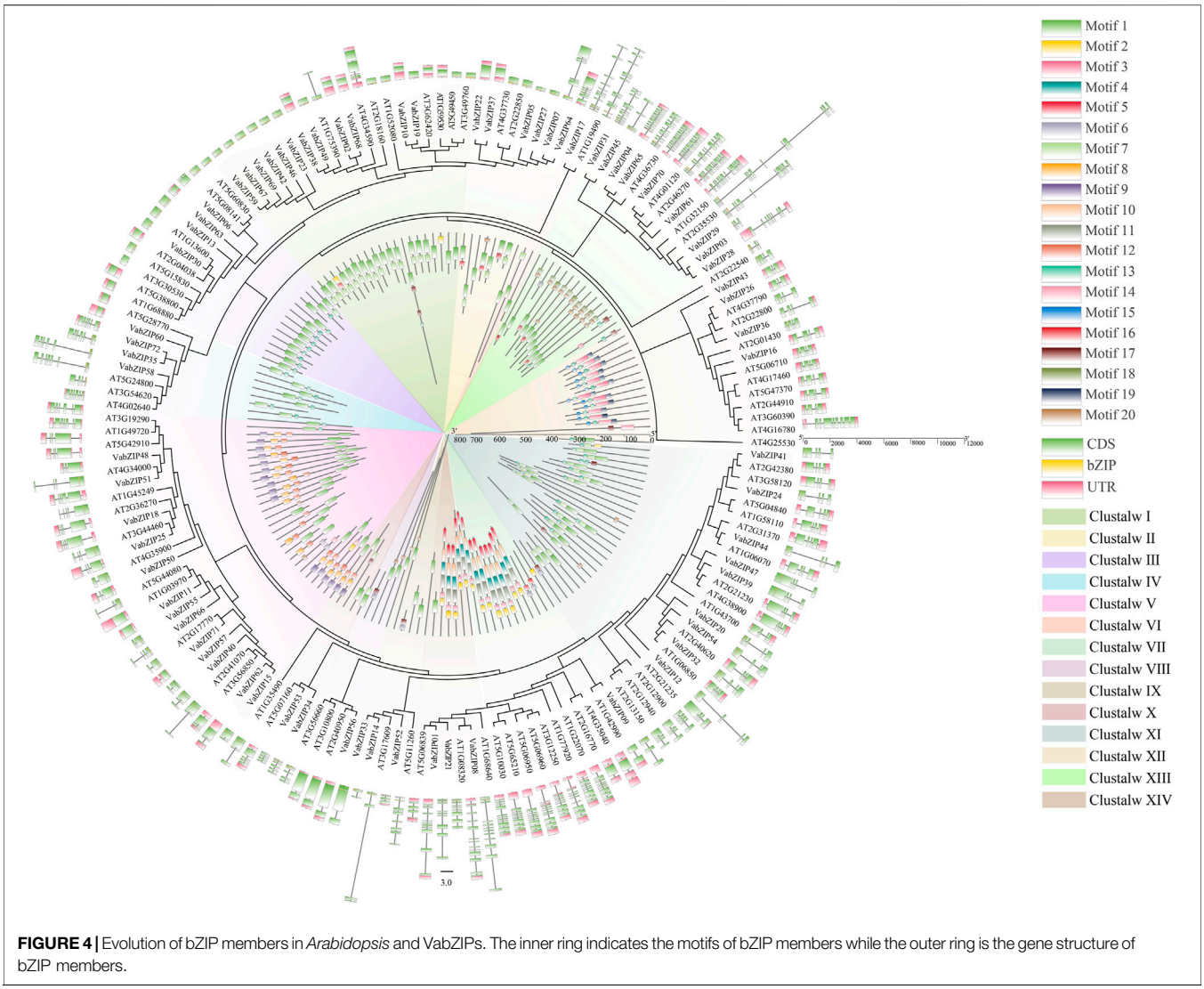


FIGURE 4 | Evolution of bZIP members in *Arabidopsis* and VabZIPs. The inner ring indicates the motifs of bZIP members while the outer ring is the gene structure of bZIP members.

Subcellular Location Analysis of VabZIPs

Subcellular locations for VabZIPs were analyzed using the CELLO database, with locations predicted by Molecular bioinformatics center. Almost all VabZIP members were predicted to be expressed in the nucleus, with only VabZIP11 predicted to be located on chloroplasts or in the cytoplasm (**Supplementary Table S5**).

In order to analyze the subcellular localization of VabZIP members, two members in different subfamilies were selected (VabZIP17 and VabZIP56) for subcellular location analysis. The result showed that the control (GFP) was distributed on the membrane and nuclear, while the VabZIP17-GFP and VabZIP56-GFP fusion proteins were only found on the nuclear, which indicated VabZIP17 and VabZIP56 were located on the nuclear (**Figure 7**).

Tissue-specific Expression Analysis at the Bud Stage

During the budding stage, the germ, cotyledon, hypocotyl, and radicle were used to investigate the expressions of bZIP members in different tissues in the adzuki bean. The twenty *VabZIP* members were selected randomly for qRT-PCR analysis which these twenty *VabZIP* members covered all of the sub-families. The *VabZIP* members were found to be expressed in a tissue-specific manner. *VabZIP06* was abundantly expressed in the germ and cotyledon, while *VabZIP11* and *VabZIP26* were highly expressed in the germ. The radicle was highly enriched with *VabZIP17*, *VabZIP30*, *VabZIP35*, and *VabZIP47*, compared to the other tissues (**Figure 8**).

Expressions of *VabZIPs* Under Abiotic Stress at the Budding Stage

Also, these twenty *VabZIP* members were selected randomly for qRT-PCR analysis to determine variations in expressions in response to abiotic stress. Expressions of the nine *VabZIP* members varied in response to various stresses (**Figure 9**). Expressions of some *VabZIP* members (such as *VabZIP06*, *VabZIP11*, *VabZIP21*, *VabZIP47* and *VabZIP51*) were up-

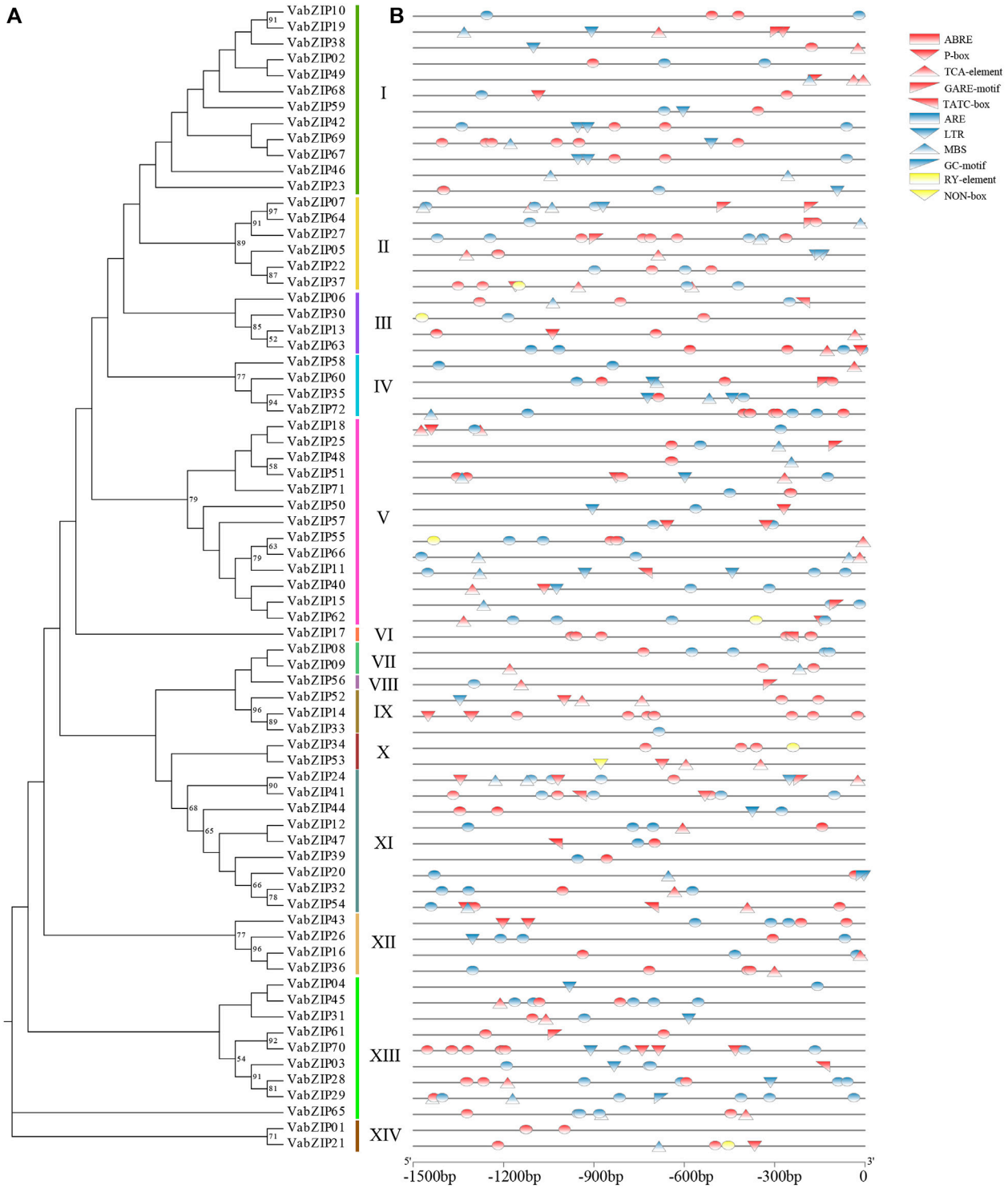


FIGURE 5 | *Cis*-elements of *VabZIPs*. **(A)** Evolutionary analysis of *VabZIPs*. **(B)** *Cis*-elements of *VabZIPs*. The red models represent the hormone-related elements. The blue models represent the stress-related elements. The yellow models represent the elements in germination.

regulated in response to drought, cold, salt, and heavy metal stress, whereas others *VabZIP* members (such as *VabZIP24*, *VabZIP34*, *VabZIP35* and *VabZIP56*) were down-regulated. Differences in expressions of *VabZIPs* in response to various types of abiotic stress were significant, such that while *VabZIP26* and *VabZIP15* did not exhibit marked changes in

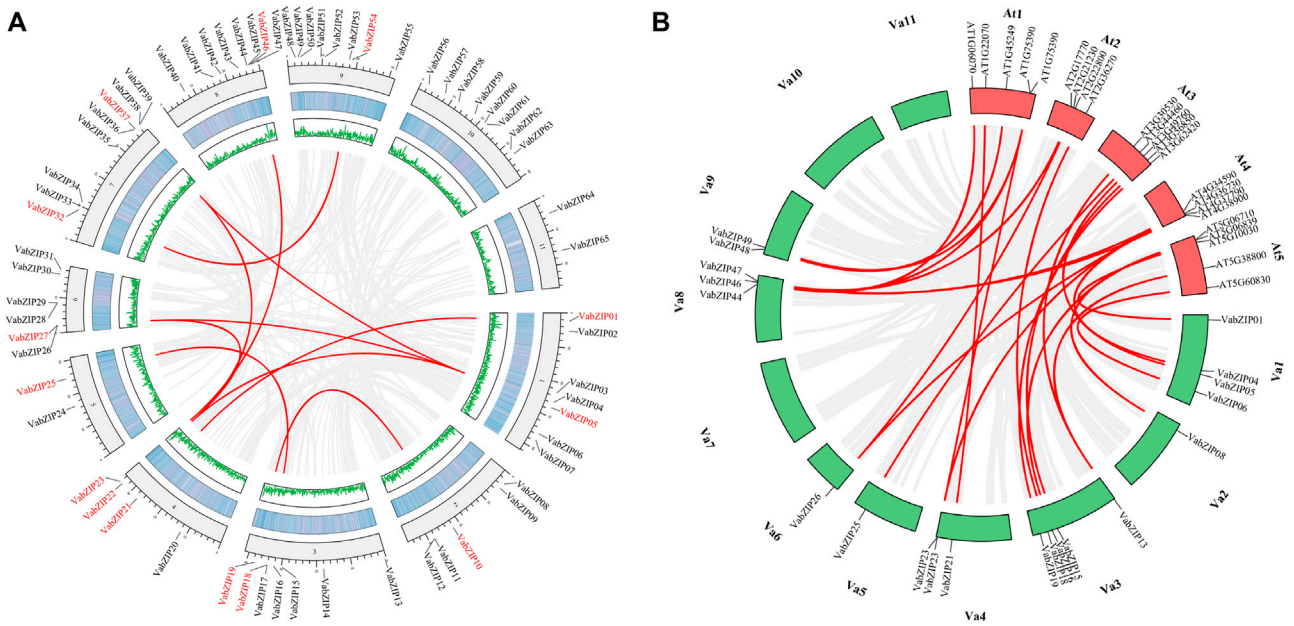


FIGURE 6 | Collinearity analysis of *VabZIPs*. **(A)** Collinearity of *PvHsf* members. The red *PvHsfs* indicate collinearity while the black ones have no collinearity. The middle two rings indicate gene density. The gray background line indicates a collinear background while the green lines indicate a collinear relationship. **(B)** Collinearity of *VabZIPs* with *Arabidopsis*. Red boxes are the chromosomes of *Arabidopsis* while the green boxes are the chromosomes of the adzuki bean. The gray lines indicate the collinearity background while the red lines indicate collinearity between *VabZIPs* and *Arabidopsis* members.

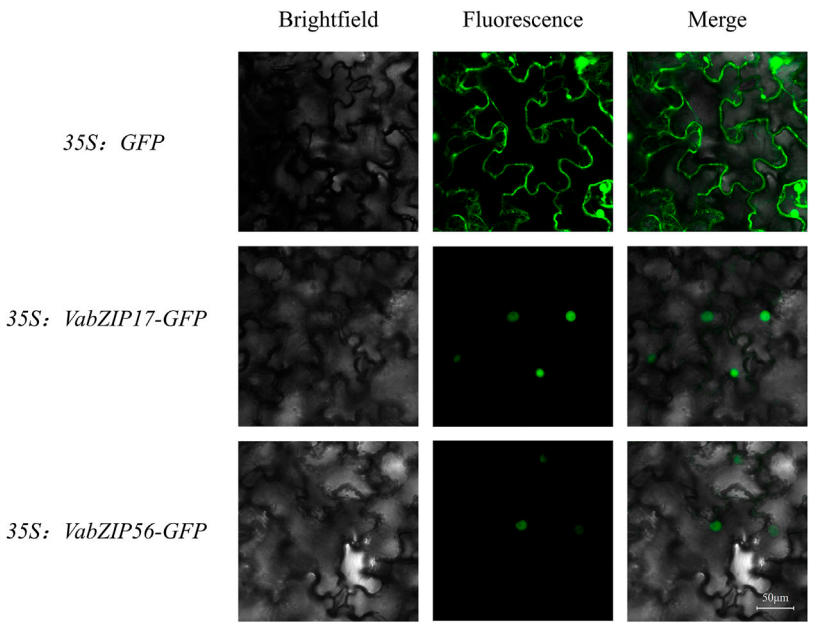


FIGURE 7 | The subcellular localization of VabZIP17 and VabZIP56.

response to heavy metal stress, they did change significantly in response to drought and cold stress, indicating that these two members may respond to other abiotic stressors other than heavy metals.

DISCUSSION

The bZIP members are present in various species, and the number of members vary from one species to another. There

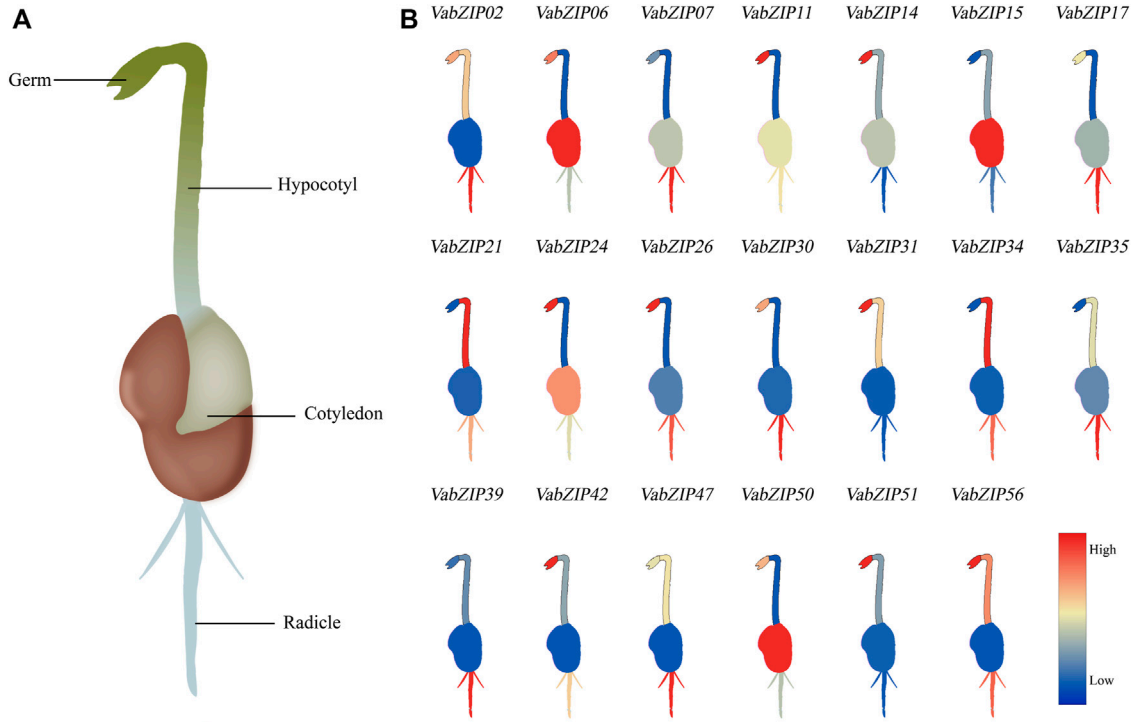


FIGURE 8 | Tissue-specific expression analysis of *VabZIPs* at the bud stage. **(A)** Schematic illustration of tissues at the bud stage of adzuki bean. **(B)** Expressions of *VabZIPs* in different tissues. The change in color from red to blue indicates a high to low expression.

are 78 members in *Arabidopsis* (Dröge-Laser et al., 2018), 80 members in potato (*Solanum tuberosum*) (Herath and Verchot, 2020), 160 members in soybean (Glycine max) (Zhang et al., 2018), 89 members in rice (*Oryza sativa* L.) (Nijhawan et al., 2008) and 69 members in tomato (*Solanum lycopersicum*) (Li X. et al., 2015). In this study, VabZIP members were identified from the reference genome of the adzuki bean, which contained 72 members. The number of VabZIP members was found to exceed the number of bZIP members in tomato, while the number of bZIP members was less than in *Arabidopsis*, potato, soybean, and rice. These findings suggest that the number of bZIP members may be related to the size of the reference genome and that after differentiation from their early ancestors, the adzuki bean may have experienced fewer genomic replication events, when compared to other species.

Evolutionary analysis revealed that VabZIP members could be divided into 14 subfamilies in the unrooted Maximum Likelihoodphy (ML) tree, and the result of evolutionary combined with bZIP members in *Arabidopsis* and adzuki bean was also revealed that bZIP members had 14 sub-families, which was similar to the results in wheat (*Triticum aestivum*) (Li D. et al., 2015) and Chinese jujube (*Ziziphus jujuba* Mill.) (Zhang Q. et al., 2020). With regards to motif and gene structures of VabZIP members, motif constitutions differed in different sub-families. Within the same sub-family, the motifs were similar and the motif of the coded bZIP (motif-1) was highly conserved (Zhou et al., 2017). Motif-3 and motif-5 were only found in sub-family X while motif 9 was only found in subfamily XII, which was similarly found in tartaty buckwheat (*Fagopyrum tataricum*) (Liu et al., 2019). VabZIP members from the same subfamily exhibited a similar gene structure, whereas VabZIPs from sub-families I, II and, III had no more than two introns, suggesting a relationship between the low number of introns and stress responses in the three sub-families (Zhao et al., 2016). The combined results of the motif and gene structure for two species of bZIP members revealed similar results.

Cis-elements in promoter regions of *VabZIP* members regulate the functions of *VabZIPs* that contain related *cis*-elements (Lescot et al., 2002). Similarly, bZIP members have *cis*-elements that are comparable across species: ABRE, TATC-box, TCA-element, and P-box are hormone-related *cis*-elements in *VabZIP* and bZIP members in common bean and potato (Wang et al., 2021; Zhang et al., 2021), implying that *VabZIPs* may regulate hormones involved in plant growth. Stress-related *cis*-elements, such as MBS and LTR, were found in *VabZIPs* and sesame bZIP transcription factor members (Wang et al., 2018), leading to the hypothesis that *VabZIPs* are involved in abiotic stress responses. Moreover, since they contain the RY element, which is similar to that found in the common bean, *VabZIP* members may have had an effect at the bud stage (Zhang et al., 2021).

Collinear analysis allows the transfer of functional information from a well-studied taxon to a less-studied taxon (Ghiurcuta and Moret, 2014). In this study, 25 *VabZIPs* pairs exhibited collinearity with an *Arabidopsis* member, which was found to be involved in plant growth regulation, abiotic stress responses,

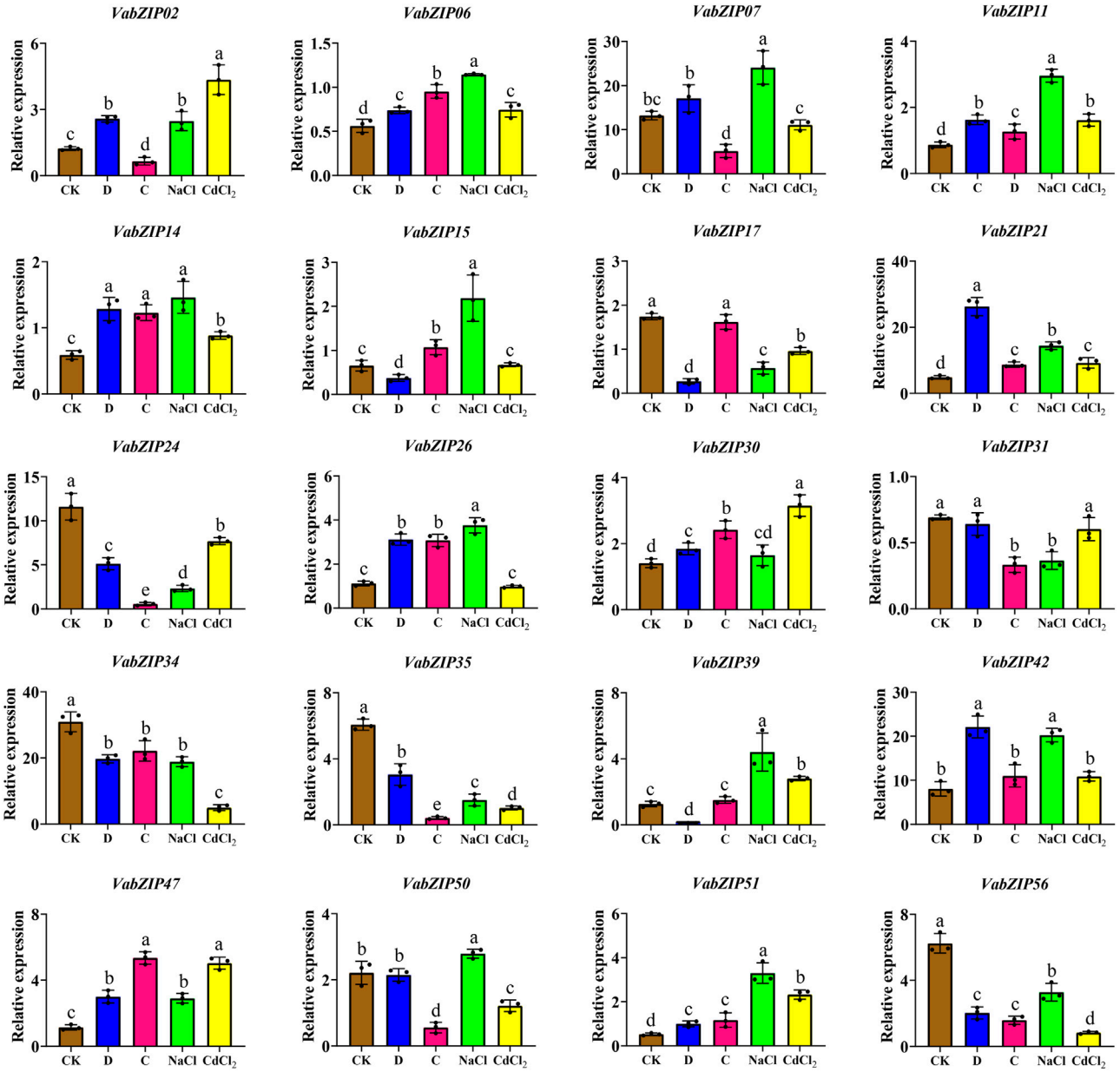


FIGURE 9 | Relative expressions of *VabZIPs* in radicles under different abiotic stress levels at the bud stage. Brown squares denote CK treatment, whereas the blue, pink, green, and yellow squares denote drought, cold, salt, and heavy metal stress, respectively.

responses to hormones, and germination in plants. *AT4G37790*, *AT1G45249,* and *AT2G36270* were the collinearity genes for *VABZIP26*, *VaBZIP48,* and *VaBZIP18* respectively, which play a function in salt stress responses (Lopez-Molina et al., 2001; Nakashima et al., 2006; Liu et al., 2016). Collinearity members for *Vigna angularis* such as *VaBZIP23*, *VaBZIP46*, *VaBZIP18,* and *VaBZIP47* in *Arabidopsis*, have a role in the bud stage of the plant: *AT1G75390*, the collinearity member for *VabZIP23* and *VabZIP46*, positively regulates plant seed germination rate. Its knock-out was associated with significantly slower germination rate (Iglesias-Fernández et al., 2013). *AT2G36270* had collinearity with *VabZIP18*, which increases proteins for preventing seed germination (Piskurewicz et al., 2008); *AT4G38900*, the collinearity gene for *VabZIP47* was expressed in meristematic tissues and negatively modulates *Arabidopsis* growth (Lozano-Sotomayor et al., 2016). In *Arabidopsis*, *VaBZIPs* member collinearity genes, such as *AT1G22070* (*VaBZIP21*) were shown to be involved in the salicylic acid (SA) signaling pathway (Zhou et al., 2000), whereas *AT1G45249* (*VaBZIP48*) and *AT2G36270* (*VaBZIP18*) were found to be involved in abscisic acid (ABA) responses (Fujita et al., 2005). The collinear analysis results indicate that *VaBZIP* members are

involved in responses to hormones, coping with environmental pressures, and regulating the bud stage.

The bZIP members have previously been reported to exhibit tissue-specific expressions, including in Olive (*Olea europaea*) (Rong et al., 2020), radish (*Raphanus sativus*) (Fan et al., 2019), and poplar (Zhao et al., 2021). Expressions of *VabZIPs* at the bud stage revealed tissue-specificity, with the radicle having higher expressions than other tissues, indicating that the radicle could be used as a target tissue for *VabZIPs*' research. Gene expression changes under abiotic stress conditions might lead to abiotic stress responses in plants, and differentially expressed genes under abiotic stress can be used as candidate genes for further research on abiotic stress responses (Qi et al., 2021). In this study, expressions of selected *VabZIPs* under abiotic stress indicated that *VabZIPs* are involved in abiotic stress responses. Expressions of *VabZIP06*, *VabZIP11*, *VabZIP21*, *VabZIP47* and *VabZIP51* were markedly up-regulates under drought, cold, salt and heavy metal stress, implying that these bZIP members are involved in abiotic stress responses. Moreover, some bZIP members are involved in abiotic stress responses in other plants, such as *StbZIP25* in potato (Wang et al., 2021), *TabZIP96* in wheat (Liang et al., 2022) and *CabZIP25* in pepper (Gai et al., 2020). Expressions of bZIP members under abiotic stress revealed that some *bZIPs* can be used in plant breeding for abiotic stress resistance, such as in watermelon (Yang et al., 2019), sesame (Wang et al., 2018), and apple (Zhao et al., 2016). These results indicate that *bZIPs* might be useful in molecular breeding under abiotic stress and the *VabZIPs* that were differentially expressed under stress can be used for further research, particularly in stress-resistance breeding.

CONCLUSION

In this study, 72 VabZIP members were identified and divided into 14 subfamilies. The members of each sub-family had motifs and gene structures that were comparable. These VabZIP members exhibited hormonal responsiveness, environmental stress, and germination *cis*-elements, indicating that the *VabZIPs* might be involved in plant hormone and abiotic stress regulation. The *VabZIPs,* whose expressions were tissue specific, might be involved in abiotic stress responses. And VabZIP17 and VabZIP56 were located on the nuclear in subcellular localization ananlysis. Furthermore, expressions of *VabZIPs* under stress conditions such as drought, cold, salt, and heavy metal stress at the bud stage revealed that some *VabZIPs* (such as *VabZIP06*, *VabZIP11*, *VabZIP21*, *VabZIP47* and *VabZIP51*) might regulate abiotic stresses responses in the adzuki bean. This study provides valuable information and insights into the development of *VabZIPs* and establishes a foundation for the use of related characteristics of *VabZIPs* in adzuki beans.

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

ZY and QW conceived the study and designed and managed the experiments. XM and YG provided the plant lines. SW and YL performed the trials, collected the data and completed the statistical analysis of the data, ZY and QW wrote the paper. All authors contributed to writing the paper.

FUNDING

This work was supported by the Key Research and Development Program of Heilongjiang Province (GA21B009-03), The Research Funds of Heilongjiang provincial science and Technology Department (CZKYF-2-C004), the China Agriculture Research System (CARS-08-G05).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.847612/full#supplementary-material

REFERENCES

Ahmad, M. A., Javed, R., Adeel, M., Rizwan, M., and Yang, Y. (2020). PEG 6000-Stimulated Drought Stress Improves the Attributes of *In Vitro* Growth, Steviol Glycosides Production, and Antioxidant Activities in Stevia Rebaudiana Bertoni. *Plants* 9, 1552. doi:10.3390/plants9111552

Alves, M., Dadalto, S., Gonçalves, A., De Souza, G., Barros, V., and Fietto, L. (2013). Plant bZIP Transcription Factors Responsive to Pathogens: A Review. *Int. J. Mol. Sci.* 14, 7815–7828. doi:10.3390/ijms14047815

Amarowicz, R., Estrella, I., Hernández, T., and Troszyńska, A. (2008). Antioxidant Activity of Extract of Adzuki Bean and its Fractions. *J. Food Lipids* 15 (1), 119–136. doi:10.1111/j.1745-4522.2007.00106.x

Amorim, L., Santos, R., Neto, J., Guida-Santos, M., Crovella, S., and Benko-Iseppon, A. (2017). Transcription Factors Involved in Plant Resistance to Pathogens. *Curr. Protein Pept. Sci.* 18, 335–351. doi:10.2174/1389203717666160619185308

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., 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

Dröge-Laser, W., Snoek, B. L., Snel, B., and Weiste, C. (2018). The Arabidopsis bZIP Transcription Factor Family - an Update. *Curr. Opin. Plant Biol.* 45, 36–49. doi:10.1016/j.pbi.2018.05.001

E, Z. G., Zhang, Y. P., Zhou, J. H., and Wang, L. (2014). Mini Review Roles of the bZIP Gene Family in Rice. *Genet. Mol. Res.* 13, 3025–3036. doi:10.4238/2014.April.16.11

Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992). The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted α Helices: Crystal Structure of the Protein-DNA Complex. *Cell* 71, 1223–1237. doi:10.1016/s0092-8674(05)80070-4

Fan, L., Xu, L., Wang, Y., Tang, M., and Liu, L. (2019). Genome- and Transcriptome-Wide Characterization of bZIP Gene Family Identifies Potential Members Involved in Abiotic Stress Response and Anthocyanin Biosynthesis in Radish (Raphanus Sativus L.). *Int. J. Mol. Sci.* 20, 6334. doi:10.3390/ijms20246334

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER Web Server: 2015 Update. *Nucleic Acids Res.* 43, W30–W38. doi:10.1093/nar/gkv397

Fuhrmann-Aoyagi, M. B., de Fátima Ruas, C., Barbosa, E. G. G., Braga, P., Moraes, L. A. C., de Oliveira, A. C. B., et al. (2021). Constitutive Expression of Arabidopsis bZIP Transcription Factor AREB1 Activates Cross-Signaling Responses in Soybean under Drought and Flooding Stresses. *J. Plant Physiol.* 257, 153338. doi:10.1016/j.jplph.2020.153338

Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M. M., Seki, M., et al. (2005). AREB1 Is a Transcription Activator of Novel ABRE-Dependent ABA Signaling that Enhances Drought Stress Tolerance in Arabidopsis. *The Plant cell* 17, 3470–3488. doi:10.1105/tpc.105.035659

Gai, W.-X., Ma, X., Qiao, Y.-M., Shi, B.-H., Ul Haq, S., Li, Q.-H., et al. (2020). Characterization of the bZIP Transcription Factor Family in Pepper (Capsicum Annuum L.): CabZIP25 Positively Modulates the Salt Tolerance. *Front. Plant Sci.* 11, 139. doi:10.3389/fpls.2020.00139

Ghiurcuta, C. G., and Moret, B. M. E. (2014). Evaluating Synteny for Improved Comparative Studies. *Bioinformatics* 30, i9–i18. doi:10.1093/bioinformatics/btu259

Guan, Y., Ren, H., Xie, H., Ma, Z., and Chen, F. (2009). Identification and Characterization of bZIP-Type Transcription Factors Involved in Carrot (Daucus CarotaL.) Somatic Embryogenesis. *Plant J.* 60, 207–217. doi:10.1111/j.1365-313X.2009.03948.x

Han, O. K., Kaga, A., Isemura, T., Wang, X. W., Tomooka, N., and Vaughan, D. A. (2005). A Genetic Linkage Map for Azuki Bean [Vigna Angularis (Willd.) Ohwi & Ohashi]. TAG. Theoretical and Applied Genetics. *Theor. Appl. Genet.* 111, 1278–1287. doi:10.1007/s00122-005-0046-8

Herath, V., and Verchot, J. (2020). Insight into the bZIP Gene Family in Solanum Tuberosum: Genome and Transcriptome Analysis to Understand the Roles of Gene Diversification in Spatiotemporal Gene Expression and Function. *Int. J. Mol. Sci.* 22, 253. doi:10.3390/ijms22010253

Hoogland, C., Mostaguir, K., Appel, R. D., and Lisacek, F. (2008). The World-2DPAGE Constellation to Promote and Publish Gel-Based Proteomics Data through the ExPASy Server. *J. Proteomics* 71, 245–248. doi:10.1016/j.jprot.2008.02.005

Hsieh, T.-H., Li, C.-W., Su, R.-C., Cheng, C.-P., Sanjaya, S. T., Tsai, Y.-C., et al. (2010). A Tomato bZIP Transcription Factor, SlAREB, Is Involved in Water Deficit and Salt Stress Response. *Planta* 231, 1459–1473. doi:10.1007/s00425-010-1147-4

Hu, B., Jin, J., Guo, A.-Y., 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

Iglesias-Fernández, R., Barrero-Sicilia, C., Carrillo-Barral, N., Oñate-Sánchez, L., and Carbonero, P. (2013). *Arabidopsis T* bZIP44: A Transcription Factor Affecting Seed Germination and Expression of the Mannanase-Encoding Gene AtMAN7. *Plant J.* 74, 767–780. doi:10.1111/tpj.12162

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). bZIP Transcription Factors in Arabidopsis. *Trends Plant Sci.* 7, 106–111. doi:10.1016/s1360-1385(01)02223-3

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., 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

Lee, S. C., Choi, H. W., Hwang, I. S., Choi, D. S., and Hwang, B. K. (2006). Functional Roles of the Pepper Pathogen-Induced bZIP Transcription Factor, CAbZIP1, in Enhanced Resistance to Pathogen Infection and Environmental Stresses. *Planta* 224, 1209–1225. doi:10.1007/s00425-006-0302-4

Lee, S. J., Lee, M. Y., Yi, S. Y., Oh, S. K., Choi, S. H., Her, N. H., et al. (2002). PPI1: A Novel Pathogen-Induced Basic Region-Leucine Zipper (bZIP) Transcription Factor from Pepper. *Mol. Plant-Microbe Interactions* 15, 540–548. doi:10.1094/mpmi.2002.15.6.540

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, D., Fu, F., Zhang, H., and Song, F. (2015a). Genome-Wide Systematic Characterization of the bZIP Transcriptional Factor Family in Tomato (Solanum lycopersicum L.). *BMC Genomics* 16, 771. doi:10.1186/s12864-015-1990-6

Li, H., Li, L., ShangGuan, G., Jia, C., Deng, S., Noman, M., et al. (2020a). Genome-Wide Identification and Expression Analysis of bZIP Gene Family in Carthamus Tinctorius L. *Sci. Rep.* 10, 15521. doi:10.1038/s41598-020-72390-z

Li, W.-Y., Wang, C., Shi, H.-H., Wang, B., Wang, J.-X., Liu, Y.-S., et al. (2020b). Genome-Wide Analysis of Ethylene-Response Factor Family in Adzuki Bean and Functional Determination of VaERF3 under Saline-Alkaline Stress. *Plant Physiol. Biochem.* 147, 215–222. doi:10.1016/j.plaphy.2019.12.019

Li, X., Gao, S., Tang, Y., Li, L., Zhang, F., Feng, B., et al. (2015b). Genome-Wide Identification and Evolutionary Analyses of bZIP Transcription Factors in Wheat and its Relatives and Expression Profiles of Anther Development Related TabZIP Genes. *BMC Genomics* 16, 976. doi:10.1186/s12864-015-2196-7

Liang, Y., Xia, J., Jiang, Y., Bao, Y., Chen, H., Wang, D., et al. (2022). Genome-Wide Identification and Analysis of bZIP Gene Family and Resistance of TaABI5 (TabZIP96) under Freezing Stress in Wheat (*Triticum A*). *Int. J. Mol. Sci.* 23, 2351. doi:10.3390/ijms23042351

Liao, Y., Zou, H.-F., Wei, W., Hao, Y.-J., Tian, A.-G., Huang, J., et al. (2008). Soybean GmbZIP44, GmbZIP62 and GmbZIP78 Genes Function as Negative Regulator of ABA Signaling and Confer Salt and Freezing Tolerance in Transgenic Arabidopsis. *Planta* 228, 225–240. doi:10.1007/s00425-008-0731-3

Liu, J., Chen, N., Chen, F., Cai, B., Dal Santo, S., Tornielli, G. B., et al. (2014). Genome-Wide Analysis and Expression Profile of the bZIP Transcription Factor Gene Family in Grapevine (Vitis Vinifera). *BMC Genomics* 15, 281. doi:10.1186/1471-2164-15-281

Liu, M., Wen, Y., Sun, W., Ma, Z., Huang, L., Wu, Q., et al. (2019). Genome-Wide Identification, Phylogeny, Evolutionary Expansion and Expression Analyses of bZIP Transcription Factor Family in Tartary Buckwheat. *BMC Genomics* 20, 483. doi:10.1186/s12864-019-5882-z

Liu, T., Longhurst, A. D., Talavera-Rauh, F., Hokin, S. A., and Barton, M. K. (2016). The Arabidopsis Transcription Factor ABIG1 Relays ABA Signaled Growth Inhibition and Drought Induced Senescence. *eLife* 5, e13768. doi:10.7554/eLife.13768

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

Lopez-Molina, L., Mongrand, S., and Chua, N.-H. (2001). A Postgermination Developmental Arrest Checkpoint Is Mediated by Abscisic Acid and Requires the ABI5 Transcription Factor in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4782–4787. doi:10.1073/pnas.081594298

Lozano-Sotomayor, P., Chávez Montes, R. A., Silvestre-Vañó, M., Herrera-Ubaldo, H., Greco, R., Pablo-Villa, J., et al. (2016). Altered Expression of the bZIP Transcription Factor DRINK ME Affects Growth and Reproductive Development in Arabidopsis Thaliana. *Plant J.* 88, 437–451. doi:10.1111/tpj.13264

Maier, A. T., Stehling-Sun, S., Offenburger, S.-L., and Lohmann, J. U. (2011). The bZIP Transcription Factor PERIANTHIA: A Multifunctional Hub for Meristem Control. *Front. Plant Sci.* 2, 79. doi:10.3389/fpls.2011.00079

Nakashima, K., Fujita, Y., Katsura, K., Maruyama, K., Narusaka, Y., Seki, M., et al. (2006). Transcriptional Regulation of ABI3- and ABA-Responsive Genes Including RD29B and RD29A in Seeds, Germinating Embryos, and Seedlings of Arabidopsis. *Plant Mol. Biol.* 60, 51–68. doi:10.1007/s11103-005-2418-5

Nijhawan, A., Jain, M., Tyagi, A. K., and Khurana, J. P. (2008). Genomic Survey and Gene Expression Analysis of the Basic Leucine Zipper Transcription Factor Family in rice. *Plant Physiol.* 146, 323–324. doi:10.1104/pp.107.112821

Pérez-Rodríguez, P., Riaño-Pachón, D. M., Corrêa, L. G. G., Rensing, S. A., Kersten, B., and Mueller-Roeber, B. (2010). PlnTFDB: Updated Content and New Features of the Plant Transcription Factor Database. *Nucleic Acids Res.* 38, D822–D827. doi:10.1093/nar/gkp805

Piskurewicz, U., Jikumaru, Y., Kinoshita, N., Nambara, E., Kamiya, Y., and Lopez-Molina, L. (2008). The Gibberellic Acid Signaling Repressor RGL2

Inhibits Arabidopsis Seed Germination by Stimulating Abscisic Acid Synthesis and ABI5 Activity. *The Plant Cell* 20, 2729–2745. doi:10.1105/tpc.108.061515

Pruneda-Paz, J. L., Breton, G., Nagel, D. H., Kang, S. E., Bonaldi, K., Doherty, C. J., et al. (2014). A Genome-Scale Resource for the Functional Characterization of Arabidopsis Transcription Factors. *Cel Rep.* 8, 622–632. doi:10.1016/j.celrep.2014.06.033

Qi, Z., Li, M., Xia, C. Y., ZhangZhang, W. J., Yin, Z. G., Zhang, Y. L., et al. (2021). Transcriptome-Based Analysis of Salt-Related Genes during the Sprout Stage of Common Bean (Phaseolus Vulgaris) under Salt Stress Conditions. *Biotechnol. Biotechnol. Equip.* 35, 1086–1098. doi:10.1080/13102818.2021.1954091

Rong, S., Wu, Z., Cheng, Z., Zhang, S., Liu, H., and Huang, Q. (2020). Genome-Wide Identification, Evolutionary Patterns, and Expression Analysis of bZIP Gene Family in Olive (Olea Europaea L.). *Genes* 11, 510. doi:10.3390/genes11050510

Schütze, K., Harter, K., and Chaban, C. (2008). Post-Translational Regulation of Plant bZIP Factors. *Trends Plant Sci.* 13, 247–255. doi:10.1016/j.tplants.2008.03.002

Shimizu, H., Sato, K., Berberich, T., Miyazaki, A., Ozaki, R., Imai, R., et al. (2005). LIP19, a Basic Region Leucine Zipper Protein, Is a Fos-Like Molecular Switch in the Cold Signaling of rice Plants. *Plant Cel Physiol.* 46, 1623–1634. doi:10.1093/pcp/pci178

Simmons, M. P., Sloan, D. B., Springer, M. S., and Gatesy, J. (2019). Gene-Wise Resampling Outperforms Site-Wise Resampling in Phylogenetic Coalescence Analyses. *Mol. Phylogenet. Evol.* 131, 80–92. doi:10.1016/j.ympev.2018.10.001

Smit, M. E., Llavata-Peris, C. I., Roosjen, M., van Beijnum, H., Novikova, D., Levitsky, V., et al. (2020). Specification and Regulation of Vascular Tissue Identity in the Arabidopsis Embryo. *Development* 147 (8), dev186130. doi:10.1242/dev.186130

Song, S., Wang, G., Wu, H., Fan, X., Liang, L., Zhao, H., et al. (2020). OsMFT2 Is Involved in the Regulation of ABA Signaling-mediated Seed Germination through Interacting with OsbZIP23/66/72 in Rice. *Plant J.* 103, 532–546. doi:10.1111/tpj.14748

Sornaraj, P., Luang, S., Lopato, S., and Hrmova, M. (2016). Basic Leucine Zipper (bZIP) Transcription Factors Involved in Abiotic Stresses: A Molecular Model of a Wheat bZIP Factor and Implications of its Structure in Function. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1860, 46–56. doi:10.1016/j.bbagen.2015.10.014

Srivastava, R., Kumar, S., Kobayashi, Y., Kusunoki, K., Tripathi, P., Kobayashi, Y., et al. (2018). Comparative Genome-Wide Analysis of WRKY Transcription Factors in Two Asian Legume Crops: Adzuki Bean and Mung Bean. *Sci. Rep.* 8, 16971. doi:10.1038/s41598-018-34920-8

Toh, S., McCourt, P., and Tsuchiya, Y. (2012). HY5is Involved in Strigolactone-Dependent Seed Germination in Arabidopsis. *Plant Signaling Behav.* 7, 556–558. doi:10.4161/psb.19839

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2000). Arabidopsis Basic Leucine Zipper Transcription Factors Involved in an Abscisic Acid-Dependent Signal Transduction Pathway under Drought and High-Salinity Conditions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11632–11637. doi:10.1073/pnas.190309197

Wang, F., Chen, X., Dong, S., Jiang, X., Wang, L., Yu, J., et al. (2020). Crosstalk of PIF4 and DELLA Modulates CBF Transcript and Hormone Homeostasis in Cold Response in Tomato. *Plant Biotechnol. J.* 18, 1041–1055. doi:10.1111/pbi.13272

Wang, Q., Guo, C., Li, Z., Sun, J., Wang, D., Xu, L., et al. (2021). Identification and Analysis of bZIP Family Genes in Potato and Their Potential Roles in Stress Responses. *Front. Plant Sci.* 12, 637343. doi:10.3389/fpls.2021.637343

Wang, Y., Li, J., and Paterson, A. H. (2013). MCScanX-Transposed: Detecting Transposed Gene Duplications Based on Multiple Colinearity Scans. *Bioinformatics* 29, 1458–1460. doi:10.1093/bioinformatics/btt150

Wang, Y., Zhang, Y., Zhou, R., Dossa, K., Yu, J., Li, D., et al. (2018). Identification and Characterization of the bZIP Transcription Factor Family and its Expression in Response to Abiotic Stresses in Sesame. *PLoS One* 13, e0200850. doi:10.1371/journal.pone.0200850

Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., et al. (2015). Genome Sequencing of Adzuki Bean (Vigna Angularis) Provides Insight into High Starch and Low Fat Accumulation and Domestication. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13213–13218. doi:10.1073/pnas.1420949112

Yang, Y., Li, J., Li, H., Yang, Y., Guang, Y., and Zhou, Y. (2019). The bZIP Gene Family in Watermelon: Genome-Wide Identification and Expression Analysis under Cold Stress and Root-Knot Nematode Infection. *PeerJ* 7, e7878. doi:10.7717/peerj.7878

Yao, Q., and Xu, D. (2017). Bioinformatics Analysis of Protein Phosphorylation in Plant Systems Biology Using P3DB. *Methods Mol. Biol. (Clifton, N.J.)* 1558, 127–138. doi:10.1007/978-1-4939-6783-4_6

Yu, C.-S., Chen, Y.-C., Lu, C.-H., and Hwang, J.-K. (2006). Prediction of Protein Subcellular Localization. *Proteins* 64, 643–651. doi:10.1002/prot.21018

Zander, M., Chen, S., Imkampe, J., Thurow, C., and Gatz, C. (2012). Repression of the *Arabidopsis T* Jasmonic Acid/Ethylene-Induced Defense Pathway by TGA-Interacting Glutaredoxins Depends on Their C-Terminal ALWL Motif. *Mol. Plant* 5, 831–840. doi:10.1093/mp/ssr113

Zhang, M., Liu, Y., Shi, H., Guo, M., Chai, M., He, Q., et al. (2018). Evolutionary and Expression Analyses of Soybean Basic Leucine Zipper Transcription Factor Family. *BMC Genomics* 19, 159. doi:10.1186/s12864-018-4511-6

Zhang, Q., Zhang, W.-J., Yin, Z.-G., Li, W.-J., Xia, C.-Y., Sun, H.-Y., et al. (2021). Genome-Wide Identification Reveals the Potential Functions of the bZIP Gene Family in Common Bean (Phaseolus vulgaris) in Response to Salt Stress During the Sprouting Stage. *J. Plant Growth Regul.* doi:10.1007/s00344-021-10497-x

Zhang, Q., Zhang, W.-J., Yin, Z.-G., Li, W.-J., Zhao, H.-H., Zhang, S., et al. (2020a). Genome- and Transcriptome-Wide Identification of C3Hs in Common Bean (Phaseolus vulgaris L.) and Structural and Expression-Based Analyses of Their Functions During the Sprout Stage under Salt-Stress Conditions. *Front. Genet.* 11, 564607. doi:10.3389/fgene.2020.564607

Zhang, Y., Gao, W., Li, H., Wang, Y., Li, D., Xue, C., et al. (2020b). Genome-Wide Analysis of the bZIP Gene Family in Chinese Jujube (Ziziphus Jujuba Mill.). *BMC genomics* 21, 483. doi:10.1186/s12864-020-06890-7

Zhao, J., Guo, R., Guo, C., Hou, H., Wang, X., and Gao, H. (2016). Evolutionary and Expression Analyses of the Apple Basic Leucine Zipper Transcription Factor Family. *Front. Plant Sci.* 7, 376. doi:10.3389/fpls.2016.00376

Zhao, K., Chen, S., Yao, W., Cheng, Z., Zhou, B., and Jiang, T. (2021). Genome-Wide Analysis and Expression Profile of the bZIP Gene Family in poplar. *BMC Plant Biol.* 21, 122. doi:10.1186/s12870-021-02879-w

Zhao, Q., Wang, H., Du, Y., Rogers, H. J., Wu, Z., Jia, S., et al. (2020). MSH2 and MSH6 in Mismatch Repair System Account for Soybean (Glycine max (L.) Merr.) Tolerance to Cadmium Toxicity by Determining DNA Damage Response. *J. Agric. Food Chem.* 68, 1974–1985. doi:10.1021/acs.jafc.9b06599

Zhou, J.-M., Trifa, Y., Silva, H., Pontier, D., Lam, E., Shah, J., et al. (2000). NPR1 Differentially Interacts with Members of the TGA/OBF Family of Transcription Factors that Bind an Element of the PR-1 Gene Required for Induction by Salicylic Acid. *Mol. Plant-Microbe Interactions* 13, 191–202. doi:10.1094/mpmi.2000.13.2.191

Zhou, Y., Xu, D., Jia, L., Huang, X., Ma, G., Wang, S., et al. (2017). Genome-Wide Identification and Structural Analysis of bZIP Transcription Factor Genes in Brassica Napus. *Genes* 8, 288. doi:10.3390/genes8100288

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 © 2022 Yin, Meng, Guo, Wei, Lai and Wang. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 02 May 2022
doi: 10.3389/fgene.2022.848663

Magnesium- a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm

Faheem Shehzad Baloch[1], Muhammad Azhar Nadeem[1], Ferit Sönmez[2], Ephrem Habyarimana[3], Zemran Mustafa[1], Tolga Karaköy[1], Gönül Cömertpay[4], Ahmad Alsaleh[5], Vahdettin Çiftçi[6], Sangmi Sun[7], Gyuhwa Chung[7] and Yong Suk Chung[8]**

[1]Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas, Turkey, [2]Department of Seed Science and Technology, Faculty of Agriculture, Bolu Abant Izzet Baysal University, Bolu, Turkey, [3]International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India, [4]Eastern Mediterranean Agricultural Research Institute, Adana, Turkey, [5]Molecular Genetic Laboratory, Science and Technology Application and Research Center, Institute for Hemp Research, Yozgat Bozok University, Yozgat, Turkey, [6]Department of Field Crops, Faculty of Agriculture, Bolu Abant Izzet Baysal University, Bolu, Turkey, [7]Department of Biotechnology, Chonnam National University, Chonnam, South Korea, [8]Department of Plant Resources and Environment, Jeju National University, Jeju, South Korea

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Ali Raza,
Fujian Agriculture and Forestry University, China
Galal Bakr Anis,
China National Rice Research Institute (CAAS), China
Shoaib Ur Rehman,
Muhammad Nawaz Shareef University of Agriculture, Pakistan

***Correspondence:**
Faheem Shehzad Baloch
balochfaheem13@gmail.com
Yong Suk Chung
yschung@jejunu.ac.kr

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

Received: 04 January 2022
Accepted: 21 February 2022
Published: 02 May 2022

Citation:
Baloch FS, Nadeem MA, Sönmez F, Habyarimana E, Mustafa Z, Karaköy T, Cömertpay G, Alsaleh A, Çiftçi V, Sun S, Chung G and Chung YS (2022) Magnesium- a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm. Front. Genet. 13:848663. doi: 10.3389/fgene.2022.848663

Magnesium (Mg) is the fourth most abundant element in the human body and plays the role of cofactor for more than 300 enzymatic reactions. In plants, Mg is involved in various key physiological and biochemical processes like growth, development, photophosphorylation, chlorophyll formation, protein synthesis, and resistance to biotic and abiotic stresses. Keeping in view the importance of this element, the present investigation aimed to explore the Mg contents diversity in the seeds of Turkish common bean germplasm and to identify the genomic regions associated with this element. A total of 183 common bean accessions collected from 19 provinces of Turkey were used as plant material. Field experiments were conducted according to an augmented block design during 2018 in two provinces of Turkey, and six commercial cultivars were used as a control group. Analysis of variance depicted that Mg concentration among common bean accessions was statistically significant ($p < 0.05$) within each environment, however genotype × environment interaction was non-significant. A moderate level (0.60) of heritability was found in this study. Overall mean Mg contents for both environments varied from 0.33 for Nigde-Dermasyon to 1.52 mg kg^{-1} for Nigde-Derinkuyu landraces, while gross mean Mg contents were 0.92 mg kg^{-1}. At the province level, landraces from Bolu were rich while the landraces from Bitlis were poor in seed Mg contents respectively. The cluster constellation plot divided the studied germplasm into two populations on the basis of their Mg contents. Marker-trait association was performed using a mixed linear model (Q + K) with a total of 7,900 DArTseq markers. A total of six markers present on various chromosomes (two at Pv01, and one marker at each chromosome i.e., Pv03, Pv07, Pv08, Pv11) showed statistically significant association for seed Mg contents. Among these identified markers, the DArT-3367607 marker present on chromosome Pv03 contributed to maximum phenotypic variation (7.5%). Additionally, this marker was found

within a narrow region of previously reported markers. We are confident that the results of this study will contribute significantly to start common bean breeding activities using marker assisted selection regarding improved Mg contents.

Keywords: phaseolus vulgaris, food legume, mg contents, DArTseq, GWAS

1 INTRODUCTION

Incensement of mineral contents in staple food to provide the recommended daily intake is crucial to fight nutrient deficiencies in the diets of humans as more than half of the world population receives insufficient essential mineral elements (Frossard et al., 2000; White and Broadley, 2009). Magnesium is the fourth most abundant mineral in human body and serves as a cofactor for more than 300 enzymes. It has a crucial role in protein synthesis, muscle contraction nerve transmission, blood pressure regulation, glucose metabolism, and signal transduction (Gröber et al., 2015). Magnesium deficiency is linked with insulin resistance, diabetes, cardiovascular diseases, stroke, and obesity (Cakmak, 2013; Bertinato et al., 2015; Gröber et al., 2015). The adult human body contains around 25 g Mg (Elin, 1987) and the estimated average daily requirement (EAR) of Mg is 265 and 350 mg for adult females and males, respectively (Rude, 1998). In plants, Mg plays many important roles in metabolism and its deficiency causes reduction in growth and yield. Since up to 35% of Mg is found in chloroplasts, chlorosis and yellowing in leaves is a typical symptom of its deficiency (Farhat et al., 2016). It is also utilized as an adaptation against aluminium toxicity where it is released from the roots to chelate aluminium ions in the soil (Cakmak and Yazici, 2010).

The Fabaceae or Leguminosae is one of the most important family in the kingdom Plantae and individuals (legumes) of this family are a great source of high quality protein, minerals, and vitamins. Legumes are multi-benefit crops as they contribute significantly to atmospheric nitrogen fixation, increase the high quality organic matter content in soil, and improve water retention. These benefits have increased the importance of legumes for sustainable agriculture under a changing climate (Stagnari et al., 2017). Among various legumes, common bean is considered as a "grain of hope" due to its nutritional potential (Nadeem et al., 2021a). Currently, this crop is cultivated all over the world and a total of 34 mha area was under common bean cultivation globally during 2020, resulting in a production of 27.5 mtones (FAO, 2022). Earlier studies confirmed that common bean was originated and domesticated in Mesoamerica and arrived in Europe through Columbian exchange (Rodiño et al., 2006; Gioia et al., 2013; Nadeem et al., 2021a), and then to Turkey through Ottoman traders. Turkey is considered a hotspot for agricultural biodiversity for most of the crops we use in our kitchen today (Nadeem et al., 2018; Nadeem et al., 2020a). Since common bean is not native to Turkey, however, it has a unique place in Turkish agriculture and hundreds of common bean landraces have emerged over time in different parts of Turkey due to variations in the agricultural practices, topography, and taste preference of local people (Nadeem et al., 2018; Nadeem et al., 2020a).

In Turkey, common bean is one of the important sources of protein, minerals, and calories after cereals as Turkish people use common bean at least once in a week either as unripe pods as vegetables, dry seeds, or in the form of salads (Nadeem et al., 2020a). Annual common bean production in Turkey in dry or fresh form was around 279,518 tons, making Turkey 3rd largest producer of the common bean in the world (Yeken et al., 2019) and 1st largest producer in the Mediterranean region (Yeken et al., 2019). An increase in common bean production has been recorded in Turkey in the last decade. This increase in common bean production can possibly be due to a good number of breeding activities carried out in Turkey. To date, a good number of common bean cultivars (200 fresh and 39 dry) have been registered in Turkey (Variety registration and seed certification center; www.tarimorman.gov.tr). As common bean has a critical place in Turkish diets and around the world, so breeding common bean cultivars with high Mg contents is crucial for fighting mineral malnutrition.

Breeding methodologies developed rapidly in the last few decades due to increasing demand of the safe foods for an increasing population of the world. The first step for breeding the common bean cultivars for high Mg contents is to evaluate the natural and ancestral germplasm, particularly from its area of diversity (Baloch et al., 2014). Various agencies of the world engaged with biodiversity have put emphasis on the collection and characterization of the germplasm, as they harbour novel alleles for the traits of interest (Baloch and Nadeem, 2022; Nadeem et al., 2020a,b). Characterization of common bean germplasm is crucial to explore the variations in order to select the elite genotypes having high Mg contents and to identify the genetic regions controlling the Mg contents in common bean seed. There is plentiful variability of seed Mg concentration in common bean. Variation of Mg contents in the seeds of common bean germplasm from various parts of the world is well documented in the earlier studies (Barampama and Simard, 1993; Sangronis and Machado, 2007; Wang et al., 2010; Akond et al., 2011; Brigide et al., 2014; Yeken et al., 2019). In our previous study, we characterized 80 Turkish common bean accessions to explore mineral elements diversity and identified Mg contents in a range of 0.63–0.94 (mg kg^{-1}). Magnesium contents harbored by the common bean germplasm could be utilized in common bean breeding to increase Mg concentrations in edible parts as common bean is frequently used in the human diet in all parts of the world (Yeken et al., 2019).

During germplasm characterization for yield and mineral traits, environment and genotype interaction should be considered one of the most important factors, as the same plant can be affected hugely from its surroundings (Shrestha et al., 2012; Misra et al., 2020). To effectively breed crops with advanced phenotypic performances, knowledge about its

adaptations and its reaction in different growing conditions and environments should be elucidated (Falconer et al., 1996). Both environmental and genetic factors affect the accumulation of Mg concentration in dry and fresh common bean seeds (Moraghan et al., 2006). Therefore, breeding common bean cultivars require the characterization of the germplasm under various environmental conditions.

Genome-wide association studies (GWAS) are a powerful genomic tool for the identification of linked markers using variation harbored by natural germplasm. GWAS, a strong structural genomics technique to screen large number of accessions using next generation sequencing (NGS) based markers covering the whole genome of common bean, has been used to identify the linked marker for various traits of agricultural and nutritional importance with high resolution (Nadeem et al., 2020b; Nadeem et al., 2021b). However very few studies evaluated the germplasm for Mg contents variations in common bean and identified QTLs/linked-markers. Delfini et al. (2021) used a Brazilian germplasm, Gunjaca et al. (2021) used a Croatian germplasm, Blair et al. (2016) and Casañas et al. (2013) used RIL populations for the identification of QTLs/linked-markers associated with Mg contents. Despite the importance of the Mg for human health and crop production, little research work is documented for breeding the common bean for Mg concentration. Therefore, Mg is referred to as "A Forgotten Element". However, under a changing climate and a rapidly increasing world population scenario, Mg deficiency is becoming a critical limiting factor for common bean production and indirectly for human health.

In the present research, mini core collection of 183 common bean genetic resources collected from 19 provinces of Turkey was established to identify the chromosomal regions associated with seed Mg contents. This study also aimed to check whether the markers identified in our study fall within the same genetic region or whether new QTLs for seed Mg contents are available in the Turkish common bean germplasm.

2 MATERIALS AND METHODS

2.1 Plant Material

During this study, a total of 177 common bean landraces collected from 19 provinces of Turkey and six commercial cultivars were used as plant material. The studied germplasm was collected from a farmer's field and the core collection was established at Bolu Abant Izzet Baysal University. Detailed information about this plant material can be found in **Supplementary Table S1**. This material was sown at the research and implementation area of Bolu Abant Izzet Baysal University and single plant selection was performed for two consecutive years during 2014 and 2015. In later years, seed multiplication was performed to get a high enough quantity of seeds for each accession for further genetics and breeding studies.

2.2 Field Experimentation

Field experiments were conducted at two geographical locations; Bolu and Sivas during 2018 according to Augmented Block design. A total of six commercial cultivars (Akman, Goynuk, Karacaşehir, Onceler, Goksun, and Akdag) were used as control groups. These cultivars were repeated in each block to standardize the mean of all accession. Sowing was performed on 12th and 17th april 2018 in Sivas and Bolu locations, respectively. All accessions were sown in a single row of 3 m length with 50 cm row to row and 10 cm plant to plant distance. All standard agronomic practices were followed during this study. Detailed information about the field experiments and applied agricultural practices can be found in our previous work (Nadeem et al., 2020a).

2.3 Phenotypic Analysis for Mg Contents of Common Bean Germplasm

Harvesting was performed at 90% pod maturity and seed samples were taken from each accession in three replicates. Seed Mg contents were investigated according to the methodology described by Yeken et al. (2019). Firstly, seeds were grinded and a fine powder was obtained. A total of 0.2 g seed sample from each accession was digested with 5 ml of concentrated nitric acid (65%) and 2 ml of hydrogen peroxide (35%) in a closed microwave digestion system (ETHOS EASY, Milestone, Italy) (Bremner, 1965; Seco-Gesto et al., 2007). After the completion of the digestion process, solutions were transferred to flasks and a final volume of 20.0 ml was maintained with ultra-pure water. This prepared solution was used for the investigation of seed Mg contents with Atomic Absorption Spectrophotometer (Shimadzu AA-7000). Seed Mg contents were repeated three times for each sample and recorded as mg kg^{-1}.

2.4 Genotyping of Common Bean Germplasm

DNA was extracted from the single selected plants according to CTAB protocol of Doyle and Doyle (1990) with a specific protocol suggested by Diversity Arrays Technology (available at "http://www.diversityarrays.com/orderinstructions/plant-dnaextraction-protocol-for-dart/)" \o "http://www.diversityarrays.com/orderinstructions/plant-dnaextraction-protocol-for-dart/)" \h www.diversityarrays.com/orderinstructions/plant-dnaextraction-protocol-for-dart/). Quality and quantity of DNA was calculated on the agarose gel (0.8%). DNA was diluted to a final concentration of 50 ng/ul and DNA samples were sent to diversity array technology (http://www.diversityarrays.com/) for DArTseq analysis based genotyping by sequencing technology (GBS). The detailed information about GBS analysis for DArTseq markers could be traced from our previously published study (Nadeem et al., 2018).

2.5 Statistical Analysis

2.5.1 Phenotypic Analysis

In this study, the sowing of the germplasm was performed in eight blocks, while six commercial cultivars were repeated eight times as a control. Repetition of commercial cultivars was used in the standardization of data and for the calculation of adjusted means. Analysis of variance (ANOVA) was conducted to get an

TABLE 1 | Summary of analysis of variance in Turkish common bean germplasm.

Variables	Df	Sum Sq	Mean Sq	F value	Pr (>F)
Genotype	182	1,089.683	5.827182	1.958887	2.13E-08
GxE	183	238.4785	1.268,503	0.426424	1
Residuals	366	1,118.503	2.974741	NA	NA
Heritability			0.597895		

*NA: Not available.

idea about the effect of genotype and genotype × environment interaction for seed Mg contents in studied germplasm. Analysis of variance (ANOVA) was calculated utilizing these evaluated adjusted means. Firstly, ANOVA was calculated within environments and later ANOVA was performed across the environments using agricolae: an R package (De Mendiburu and Simon 2015). Mean, maximum, and minimum Mg contents for the studied environments were investigated through XLSTAT statistical software (www.xlstat.com). Frequency distribution and provinces based Mg contents were calculated through XLSTAT statistical software. The most stable common bean accessions for Mg contents were investigated through the online software "STABILITYSOFT" (Pour-Aboughadareh et al., 2019). The constellation plot for common bean accessions was constructed through JMP 14.1.0 statistical software (2018, SAS Institute Inc., Cary, NC, United States).

2.5.2. Marker-Trait Investigation for Seeds Mg Contents

The Q-matrix and Kinship are basic requirements while performing bioinformatics analysis for GWAS studies as both are used to correct the population and family structure during the association analysis. Population structure of studied germplasm was performed previously and has been reported in our published study (Nadeem et al., 2018). Therefore, the required Q-matrix for GWAS analysis was evaluated from a previous study (Nadeem et al., 2018). Marker trait association was performed using mixed linear model approach (MLM, Q + K). Tassel 5.2.50 (https://tassel.bitbucket.io) program was used to investigate the kinship matrix according to the methodology of Bradbury et al. (2007). False discovery rate (FDR) and Bonferroni (p = 0.01) thresholds were used in the present study to investigate the significantly associated markers. A Manhattan plot was developed to visualize the statistically significant markers through R 3.4.1 statistical software (http://www.r-project.org/) by using qq-man R Package (Turner, 2014). A physical map was constructed for the identified linked DArTseq markers through R 3.4.1 statistical software to confirm whether they were present or not at same chromosomal region.

3 RESULTS

3.1 Phenotypic Diversity

The results showed that there was a plentiful and continuous diversity for seed Mg concentrations among common bean accessions (**Supplementary Table S2**). The ANOVA results depicted that Mg concentration among common bean accessions was statistically significant within each environment (data now shown), however genotype × environment interaction was non-significant (**Table 1**). Heritability analysis showed a moderate level (0.60) of heritability (**Table 1**).

During 2018 at Bolu, Mg contents ranged from 0.32–1.49 mg kg^{-1} for Nigde-Dermasyon and Nigde- Derinkuyu landraces respectively, while mean Mg contents were 0.90 mg kg^{-1} (**Table 2**). Similarly, at Sivas, seed Mg contents ranged from 0.34–1.55 mg kg^{-1} for Nigde-Dermasyon and Nigde-Derinkuyu landraces respectively, while mean Mg contents were 0.95 mg kg^{-1}. By taking the mean of both locations, overall Mg contents among accessions varied from 0.33–1.52 mg kg^{-1} for the above reported landraces, respectively. Frequency distribution of the Mg contents among common bean accessions is shown in **Figure 1**, which clearly depicted that most of the landraces with high Mg concentration in the common bean seeds had a value above 0.90 mg kg^{-1}.

Mg contents variations were also observed at the geographical provinces from where the studied germplasm was collected. By taking the mean of two locations, we observed that maximum Mg contents were reflected by landraces from Bolu (1.13 mg kg^{-1}), while landraces from Bitlis province were found poor (0.76 mg kg^{-1}) in seed Mg contents (**Figure 2**). We performed the stability analysis using the mean of two environments and succeeded in identifying eight landraces with high Mg contents in common bean seeds (**Table 3**). The constellation plot separated the studied germplasm into two main populations A and B (**Figure 3**). Population A clustered accessions with poor Mg contents compared to population B. Population A was further subdivided into two subpopulations A1 and A2. Population B was further subdivided into B1 and B2, while subpopulation B2 clustered accessions rich in Mg contents.

3.2 Genomic Regions and Putative Genes Associated With Mg Contents

The mean data of two environments (Bolu and Sivas) were used for the identification of chromosomal regions associated with seed Mg contents and a total of 6 DArTseq markers were found statistically significant for Mg contents in common bean seeds (**Table 4** and **Figure 4**). A total of two markers (DArT-3365938 and DArT-3367358) were present on chromosome Pv01, while rest of four markers (one marker for each chromosome) were present on chromosome Pv03, Pv07, Pv08 and Pv11. Among these identified markers, DArT-3367607 marker contributed in maximum (7.5%) variations. A total of five putative candidate genes were also identified from sequences reflecting homology to six identified DArTseq markers. Vigun01g245600 putative gene was predicted as a putative gene for DArT-3365938 and DArT-3367358 markers. The constructed physical map of identified markers revealed a very narrow region for 0.00000101 Mbp for DArT-3367358 and DArT-3365938 markers (**Figure 5**).

4 DISCUSSION

Despite the well-known role of Mg for human health and plant metabolism, little or negligible research has been conducted on

TABLE 2 | Minimum, maximum and mean Mg (mg kg^{-1}) contents in Turkish common bean germplasm under multi-year/environments.

Environment	Minimum	Maximum	Mean	Std. deviation
Bolu2018	0.320 (Nigde-Dermasyon)	1.490 (Nigde-Derinkuyu)	0.90	0.172
Sivas2018	0.340 (Nigde-Dermasyon)	1.550 (Nigde-Derinkuyu)	0.95	0.209
Mean	0.330 (Nigde-Dermasyon)	1.520 (Nigde-Derinkuyu)	0.92	0.186

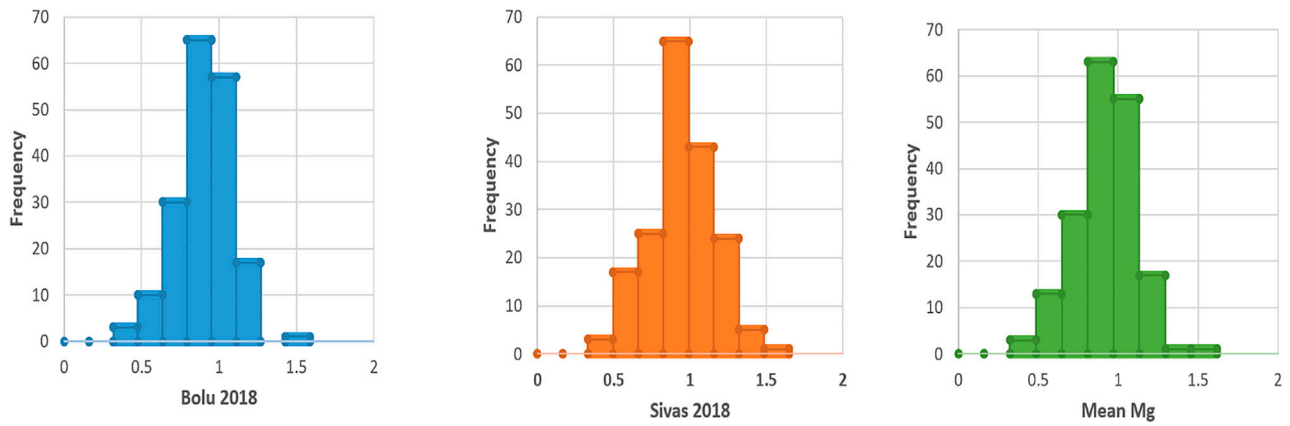


FIGURE 1 | Frequency distribution of Mg contents in Turkish common bean seeds.

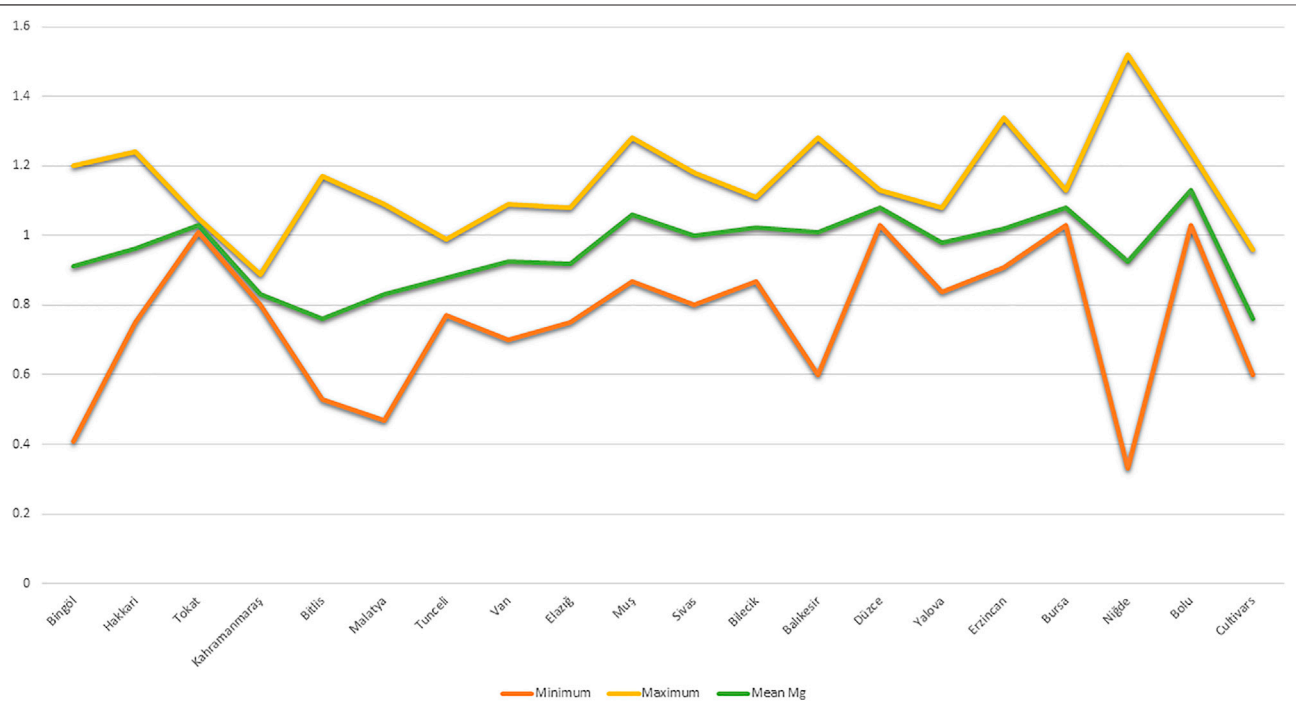


FIGURE 2 | Variation of seed Mg contents in Turkish common bean germplasm on the basis of their collection provinces.

this important element. As discussed earlier, Mg is considered a "Forgotten element" and there is an utmost need to characterize the germplasm of different crops in order to identify the accessions with high Mg contents in the edible parts of crops to eradicate the malnutrition of this important mineral element in the human population. Common bean is one of the most

TABLE 3 | The most stable common bean accessions for Mg contents.

Landraces	Mean Mg contents	s^2d_i	b_i	CVi
Hakkari-28	0.94	0	0	0
Bitlis-35	1.05	0	0	0
Bitlis-69	1.17	0	0	0
Bursa-1	1.13	2.31E-06	0.199184	0.509427
Duzce-9	1.13	2.45E-06	0.197323	0.512441
Hakkari-69	1.05	2.31E-06	0.199184	0.548117
Van-33	1.05	2.31E-06	0.199184	0.548117
Hakkari-65	1.04	2.45E-06	−0.19732	0.553371

CVi, Coefficient of variance, s^2d_i = Deviation from regression, b_i = Regression coefficient 691.

important legume crops used frequently in the diets of the human populations of the developing and least developed countries, and therefore crops like common bean are considered as "poor man's meat". Moreover, common bean is referred to as a "grain of hope" as its seeds are rich sources of various mineral elements crucial for human health (Nadeem et al., 2021a). The present work was done with an objective of characterizing the mini core collection of common bean germplasms from Turkey, an important area of diversity for common bean, to find the promising accessions with high Mg contents in the common bean seeds and also to unlock the chromosomal region associated with this mineral element and to discuss whether the genomic regions identified here are novel QTLs or whether they fall into the same genetic region reported earlier.

Analysis of variance (ANOVA) revealed significant differences ($p < 0.05$) for Mg contents in both environments, however the genotype x environment was non-significant (**Table 1**). Heritability analysis showed a moderate level of heritability. These results were found in line with the very recent study of Delfini et al. (2021). They also observed significant effects within the environment, while there was no Genotype x environment (G x E) interactions. Moreover they also found a low range of Heritability (0.18–0.47). During this study, plentiful variation (0.330–1.520 mg kg^{-1}) was seen for seed Mg contents (**Table 2**).

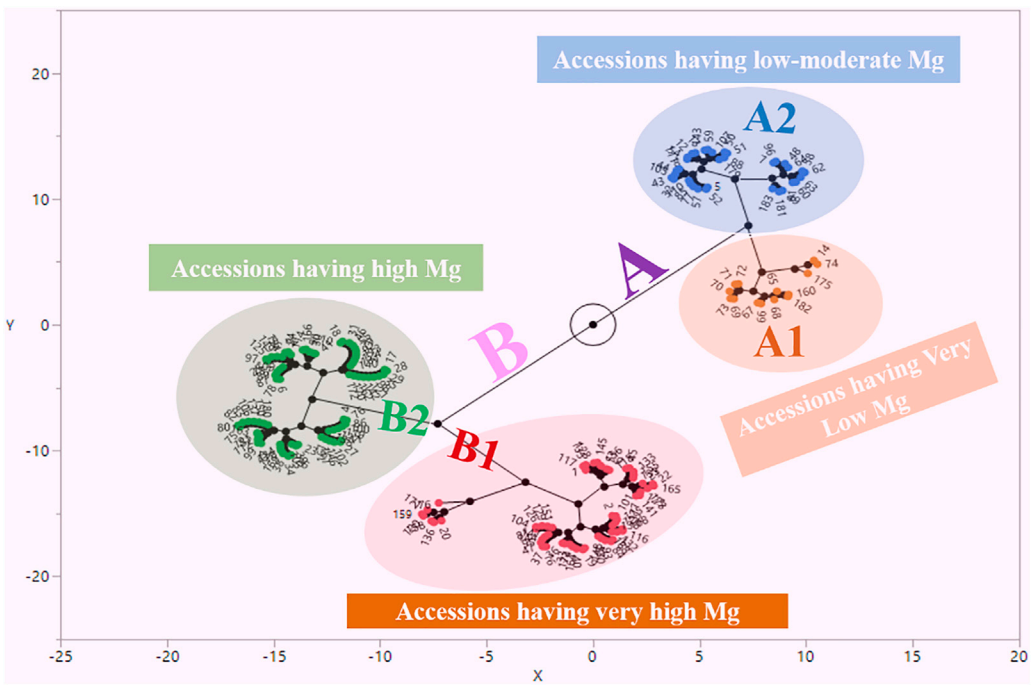


FIGURE 3 | Constellation plot for magnesium content in Turkish common bean germplasm.

TABLE 4 | Chromosomal regions associated with seed Mg contents in Turkish common bean germplasm.

Marker	Chromosome	Position (bp)	*p*-value	R2%	Putative gene (bp)
DArT-3367607	3	521,185	4.29E-04	7.5	Phvul.003G001300
DArT-3365938	1	51,520,369	4.62E-04	7.4	Vigun01g245600
DArT-3367358	1	51,520,303	5.31E-04	7.3	Vigun01g245600
DArT-22345,410	11	6,622,226	5.61E-04	6.9	Phvul.011G071900
DArT-3375642	7	45,783,153	8.19E-04	6.5	Phvul.005G079500
DArT-16652019	8	52172087	8.65E-04	6.7	Phvul.008G185200

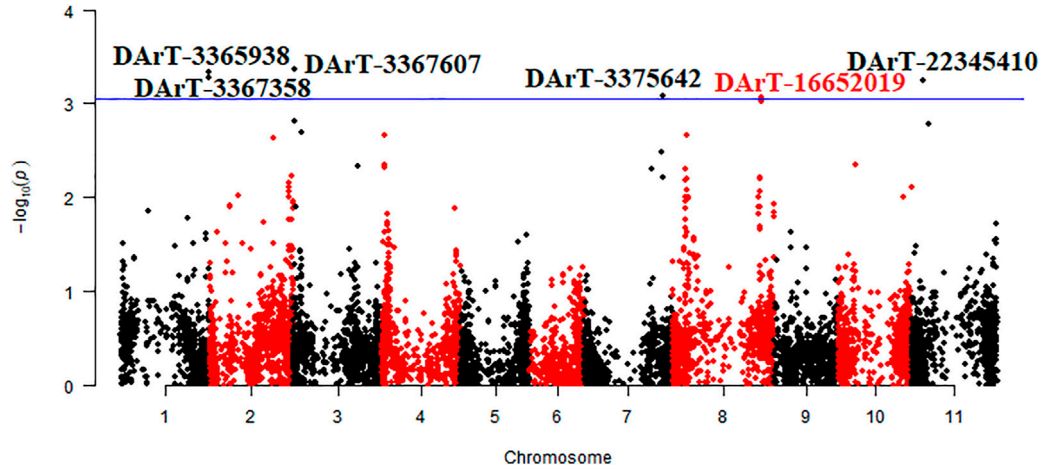


FIGURE 4 | Manhattan plot of marker trait association for magnesium content in Turkish common bean germplasm.

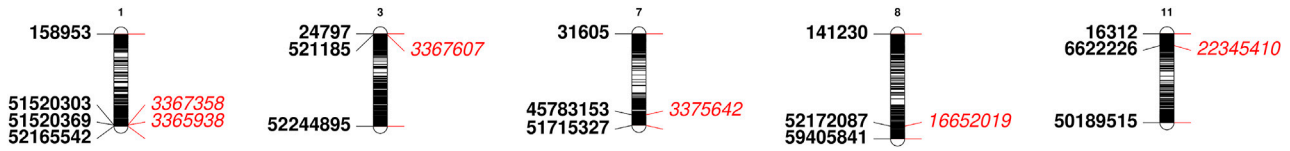


FIGURE 5 | Physical map of identified DArTseq markers having association with seed Mg contents in Turkish common bean germplasm.

Moreover, frequency distribution clearly depicted that most of the accessions have high (above than 0.9 mg kg-1) Mg contents (**Figure 1**). The mean and range of Mg contents in the studied germplasm (**Table 2**) were found higher than in our previous study (Yeken et al., 2019). This could be because the germplasm used in the present study was different from our previous studies and the size of the germplasm in the earlier study was small compared with the germplasm used in the current study. The amount of the Mg contents in common bean accessions used in this study was comparable or slightly lower than the previous studies. Very recently, Gunjača et al. (2021) aimed to explore marker-trait association in common bean germplasm for mineral contents and reported Mg contents in a range of 0.13–0.24% in dry weight (DW). Palčić et al. (2018) also reported Mg content in a range of 0.17–0.2% DW, with the average of 0.18%. Delfini et al., 2021 analyzed 178 Mesoamerican accessions in three different conditions and found the Mg content to range between 164—290 mg/100 g. Augustin et al. (1981) used samples from nine classes of common bean germplasm from United States of America, and found that raw common beans contain 16–230 mg magnesium whereas cooked common bean have slightly reduced Mg content of 130–220 mg per 100 g dry weight. Ray et al. (2014) used 10 common bean cultivars grown in Saskatchewan, Canada in six locations and found the Mg content in a range of 184.5–238.3 mg/100 g. In another study, seven common bean genotypes from Manitoba and Saskatchewan, Canada were evaluated and Mg contents varied from 143.0–199.5 mg/100 g (Wang et al., 2010). In their assay Akond et al. (2011) used 29 common bean genotypes from CIAT (International Center for Tropical Agriculture), United States, India and Brazil and found the Mg content 0.647–1.105 mg/100 g. Brigide et al. (2014), in their study, used four biofortified and one control common bean variety. The Mg content in 100 g seeds was measured 11.2–17.3 mg in the raw treatment and 16.1–17.1 mg in the macerated/cooked treatment. Existence of good variation in the Turkish common bean germplasm for Mg contents could be successfully used for breeding the common bean with a higher Mg composition of seeds.

The world is facing the disaster of climate change, and therefore selection of the stable genotypes is one of the most important criteria for effective breeding programs. The environment has always had a magnificent effect on the genotypes, therefore genotypes with maximum stability are identified with the least environmental effect. In this study, eight common bean accessions reflecting the highest stability for seed Mg contents were evaluated and can be used for future common bean breeding programs (**Table 3**). These stable accessions were evaluated according to coefficient of variance, deviation from regression, and regression coefficient as described in our previous study (Nadeem et al., 2021b). According to Francis and Kannenberg (1978), accessions with a low coefficient of variance have minimal environmental variance and can be used as the most stable accessions. All of the identified and stable accessions reflected a good range of mean Mg contents (0.94–1.17 mg kg^{-1}). Therefore, these identified accessions can serve as a source of genetic resource for the biofortification of common bean regarding Mg contents.

The germplasm used in this research was collected from various geographical provinces with the different topography, climate, and agricultural practices of Turkey. Therefore, we also analyzed the germplasm according to their geographical provinces (**Figure 2**). Accessions from the Bolu province showed the highest Mg contents while the Bitlis province reflected the lowest Mg contents. Bolu is located in the west Black sea region of the Turkey, while the Bitlis province is located in the East Anatolian region of Turkey. During our previous study regarding morpho-agronomic characterization of Turkish common bean germplasm, we found that accessions from the Bolu province have higher 100 seeds weight compared to accessions from Bitlis. It is clearly understandable by making a comparison of seed Mg contents with 100 seeds weight at the province level from our previous study (Nadeem et al., 2020a), most of the provinces with higher 100 seeds weight reflected higher Mg contents. Keeping in view these findings, we can postulate that accessions with higher seed weight may have higher Mg contents (Nadeem and Baloch's personal perception).

To see the pattern of variation, clustering analysis was performed to observe the grouping of studied germplasm. The studied germplasm was divided into two populations i.e., A and B on the basis of their Mg contents (**Figure 3**). Most of the accessions with comparatively low Mg contents were clustered in group A. Population A was further subdivided into A1 and A2, while the A1 subpopulation clustered those accessions with very low Mg contents. Mean minimum Mg contents (0.330 mg kg^{-1}) were observed for Nigde-Dermasyon landraces that were also present in sub-population A1. Sub-population A2 clustered accessions had low to moderate Mg contents (0.60–0.90 mg kg^{-1}). Population B clustered accessions had high Mg contents. Sub-population B1 clustered accessions had very high Mg contents compared to B2. Most of the accession present in B1 sub-population reflected Mg contents above 1 mg kg^{-1}. Nigde-Derinkuyu landrace reflected maximum mean Mg contents during this study and it was present in B1 group. Accessions present in the B2 group reflected higher Mg contents compared to population A and lower than sub-population B1. As accessions present in B1 sub-population were rich in Mg contents. It is recommended to utilize the accessions of this sub-population for the breeding perspective of common bean.

4.1 Marker-Trait Association

A total of six DArTseq markers showed significant association with seed Mg contents in Turkish common bean germplasm (**Table 4** and **Figure 4**). DArT-3367607 was the only marker present at Pv03 and contributed in highest phenotypic variations (7.5%). A total of two DArTseq markers (DArT- 3365938 and DArT-3367358) showed their distribution on chromosome Pv01 and reflected 7.4 and 7.3% variations respectively. Similarly, one statistically significant marker on each of chromosome 333 Pv07 (DArT-3375642), Pv08 (DArT-16652019) and Pv011 (DArT-22345,410) was observed with phenotypic variation of 6.5, 6.7 and 6.9 respectively (**Table 4**). Very recently, Delfini et al. (2021) performed GWAS analysis to identify quantitative trait nucleotides (QTNs) for mineral contents in common bean diversity panel from Brazil. They reported distribution of QTNs for Mg contents on various chromosomes of common bean. They reported S03_552367 as a QTN on chromosome Pv03 at the position of 552,367 bp. Similarly, DArT-3367607 marker identified in this study was present on the chromosome Pv03 at the position of 521185 bp. Both markers were present within a very narrow region of 0.031182 Mbp. Similarly, Gunjaca et al. (2021) identified only one marker on chromosome Pv08 for Mg contents in Croatian common bean germplasm and phenotypic variation explained by this marker was low when compared with this study. Their marker chromosomal position was 50,916,423 bp, while our identified marker on the same chromosome was at 52172087 bp. Both markers were present in a region of 1.255 Mbp. Blair et al. (2016) also reported the distribution of QTLs for Mg contents on Pv07, Pv08, and Pv10. They stated that the identified QTL (P9DB1D) present on Pv07 chromosome was in the region of the *Phs* locus which has been found to be a very important region with multiple genes that influenced Fe and Zn concentration (Blair et al., 2009). As Blair et al. (2016) also found QTL present on Pv08, they stated that this QTL near the marker Bng96 aligned with a previous QTL for Fe contents (Blair et al., 2012). Casañas et al. (2013) reported a QTL ($Mg7^{xc}$) for Mg contents in common bean on Pv07 and reported *P* gene as a closest marker to this QTL. They also reported that this *P* gene has association with calcium, ashes, dietary fiber, and uronic acid contents in common bean. Some markers identified in this study were found in the same chromosomal regions reported by Delfini et al. (2021) and Gunjaca et al. (2021). Therefore, it could be further studied for validation through candidate gene association mapping. Most of the markers found in this study could be associated with novel/new QTLs that could be present in Turkish germplasm and can be used for marker-assisted breeding of common bean. Additionally, the physical map disclosed that two markers i.e. DArT-3367358 and DArT-3365938 with association for seed Mg contents in common bean were present on the chromosomes Pv01 at 51.52 Mbp and 52.16 Mbp respectively (**Figure 5**). Both markers were present in a very narrow chromosomal region with a distance of 0.00000101 Mbp. Therefore, this region containing both markers with association for seed Mg contents should be considered for future common bean breeding.

During the present investigation, sequences of the identified markers were used to BLAST-search against the common bean genome in the legume information system (LIS: https://legumeinfo.org/) and putative genes were investigated. Phvul.003G001300 was identified as a putative gene for DArT-3367607 marker. This gene encodes for Pentatricopeptide repeat (PPR) superfamily protein. This family is characterized by tandem 30–40 amino acid sequence motifs and considered one of the largest protein families in land plants (Xing et al., 2018). Zhang et al. (2020) stated that this family is involved in the post-transcriptional processing of RNA in chloroplasts and mitochondria, which is very important for plant development and evolutionary adaption. Previous studies confirmed that mitochondria have the capability of accumulation of Mg and ultimately act as an important intracellular Mgstore (Kubota et al., 2005; Shindo et al., 2011). Vigun01g245600

present on chromosome 01 of *Vigna unguiculata* was found to be a putative gene for DArT-3365938 and DArT-3367358. This gene encodes for Ankyrin repeat family protein, which is considered to be one of the largest protein families. This protein family is involved in various processes like plant growth and development, hormone response, and contributes significantly to resistance to abiotic and biotic stresses (Lopez-Ortiz et al., 2020). Zhao et al. (2020) revealed the role of this protein family in salt and drought tolerance in Arabidopsis and Soybean. Verbruggen and Hermans (2013) clearly explored the role of Mg in various physiological and biochemical processes in plants. Phvul.011G071900 resulted as a putative gene for DArT-22345410 that encodes for DOF zinc finger protein. DOF is a family of transcription factors that contributes significantly to various fundamental processes like seed germination, seed maturation phytohormone production and response to light (Kang et al., 2016). Zhang et al. (2020) clearly explored the role of Mg for yield and seed germination traits in wax gourd crop. They concluded that seeds derived from Mg-sufficient plants were more vigorous and have earlier emergence, better seedling establishment, and better development compared to the seeds collected from Mg deficient plants. Phvul.005G079500 was found to be a putative gene for DArT-3375642 and this gene encodes for zinc finger (Ran-binding) family protein. Zinc figure proteins comprise one of the largest transcription factor families and play a significant role in various abiotic stress resistances (Han et al., 2020). Phvul. 008G185200 was found as a putative gene for DArT-16652019 marker and this gene encodes for mate efflux family protein (MATE). Members of this family are present abundantly in plants and contribute to growth and developmental processes (Chen et al., 2015). Transporters of this family are directly or indirectly involved in detoxification of toxic compounds, heavy metals resistance, disease resistance, and response to hormone regulation (Wu et al., 2014; Santos et al., 2017). Previous studies explored the role of this protein family against aluminum toxicity (Liu et al., 2016; Wang et al., 2017). Bose et al. (2011) comprehensively explored the role of Mg in reducing aluminum toxicity in plants.

5 CONCLUSION

The present investigation provided a deep insight into the existence of the wide range of Mg contents diversity in Turkish common bean germplasm. genotype × environment interaction showed non-significant effects, while a moderate level of heritability was observed for the studied trait. Accessions from Nigde province showed maximum range of variation in seed Mg contents. Some stable bean accessions were also identified which can be explored in the crossing program as parents for developing bean varieties with stable Mg contents under various environmental conditions. The present investigation reported six DArTseq markers with association for seed Mg contents. Identified markers with association for Mg contents were found within a narrow region in which previous markers for Mg contents have been reported by earlier studies. Keeping this in view, identified markers in this study should be validated along with previously reported markers. After validating these markers, they can be effectively used in marker assisted selection for breeding bean with higher Mg contents. We are confident that the information presented in this study will be helpful for common bean breeding regarding Mg contents.

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 authors.

AUTHOR CONTRIBUTIONS

FSB and MAN designed the experiments. MAN, TK, FS and FSB performed the experiments. MAN, TK, and FS performed elemental analysis. EH, MAN, and GCö, AA performed the bioinformatics analysis. FSB, VÇ, SS, GCh and YSC provided resources. FSB, MAN, and ZM wrote the manuscript. All authors read and approved the final manuscript for the publication.

FUNDING

The authors express their gratitude to Scientific and Technological Research Council of Turkey (TÜBİTAK) to FB for providing research grants for this study under the project no: TOVAG- 2015O630 and Basic Science Research Program supported this research through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2019R1A6A1A11052070).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.848663/full#supplementary-material

REFERENCES

Akond, A. S. M. G. M., Crawford, H., Berthold, J., Talukder, Z. I., and Hossain, K. (2011). Minerals (Zn, Fe, Ca and Mg) and Antinutrient (Phytic Acid) Constituents in Common Bean. *Am. J. Food Technology* 6, 235–243. doi:10.3923/AJFT.2011.235.243

Augustin, J., Beck, C. B., Kalbfleish, G., Kagel, L. C., and Matthews, R. H. (1981). Variation in the Vitamin and mineral Content of Raw and Cooked Commercial Phaseolus vulgaris Classes. *J. Food Sci.* 42646, 1701–1706. doi:10.1111/j.1365-2621.1981.tb04467.x

Baloch, F. S., Karaköy, T., Demïrbaş, A., Toklu, F., Özkan, H., and Hatïpoğlu, R. (2014). Variation of Some Seed mineral Contents in Open Pollinated Faba Bean

(Vicia faba L.) Landraces from Turkey. *Turk. J. Agric. For.* 38, 591–602. doi:10.3906/tar-1311-31

Baloch, F. S., and Nadeem, M. A. (2022). Unlocking the Genomic Regions Associated with Seed Protein Contents in Turkish Common Bean Germplasm through Genome-wide Association Study. *Turk J. 432 Agric. For.* 46. doi:10.3906/tar-2104-63

Barampama, Z., and Simard, R. E. (1993). Nutrient Composition, Protein Quality and Antinutritional Factors of Some Varieties of Dry Beans (*Phaseolus vulgaris*) Grown in Burundi. *Food Chem.* 47, 159–167. doi:10.1016/0308-8146(93)90238-b

Bertinato, J., Xiao, C. W., Ratnayake, W. M. N., Fernandez, L., Lavergne, C., Wood, C., et al. (2015). Lower Serum Magnesium Concentration Is Associated with Diabetes, Insulin Resistance, and Obesity in South Asian and white Canadian Women but Not Men. *Food Nutr. Res.* 59, 25974. doi:10.3402/fnr.v59.25974

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., Herrera, A. L., Sandoval, T. A., Caldas, G. V., Filleppi, M., and Sparvoli, F. (2012). Inheritance of Seed Phytate and Phosphorus Levels in Common Bean (*Phaseolus vulgaris* L.) and Association with Newly-Mapped Candidate Genes. *Mol. Breed.* 30, 1265–1277. doi:10.1007/s11032-012-9713-z

Blair, M. W., Wu, X., Bhandari, D., and Astudillo, C. (2016). Genetic Dissection of ICP-Detected Nutrient Accumulation in the Whole Seed of Common Bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 7, 219. doi:10.3389/FPLS.2016.00219

Bose, J., Babourina, O., and Rengel, Z. (2011). Role of Magnesium in Alleviation of Aluminium Toxicity in Plants. *J. Exp. Bot.* 62, 2251–2264. doi:10.1093/jxb/erq456

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

Bremner, J. M. (1965). "Total Nitrogen," in *Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties*. Editor A. G. Norman (Amer. Soc. of Agron. Inc. Pub. Agron.), 1149–1178. doi:10.2134/agronmonogr9.2.c32

Brigide, P., Canniatt-Brazaca, S. G., and Silva, M. O. (2014). Nutritional Characteristics of Biofortified Common Beans. *Food Sci. Technol. (Campinas)* 34, 493–500. doi:10.1590/1678-457x.6245

Cakmak, I. (2013). Magnesium in Crop Production, Food Quality and Human Health. *Plant Soil* 368, 4621–4624. doi:10.1007/s11104-013-1781-2

Cakmak, I., and Yazici, A. M. (2010). Magnesium: a Forgotten Element in Crop Production. *Bet. Crop* 94, 23–25.

Casañas, F., Pérez-Vega, E., Almirall, A., Plans, M., Sabaté, J., and Ferreira, J. J. (2013). Mapping of QTL Associated with Seed Chemical Content in a RIL Population of Common Bean (*Phaseolus vulgaris* L.). *Euphytica* 192, 279–288. doi:10.1007/s10681-013-0880-8

Chen, L., Liu, Y., Liu, H., Kang, L., Geng, J., Gai, Y., et al. (2015). Identification and Expression Analysis of MATE Genes Involved in Flavonoid Transport in Blueberry Plants. *PLoS ONE* 10, e0118578. doi:10.1371/journal.pone.0118578

De Mendiburu, F., and Simon, R. (2015). Agricolae-Ten years of an open source statistical tool for experiments in breeding, agriculture and biology (No. e1748). *Peer J. Pre. Prints.* 3, e1404v1. doi:10.7287/peerj.preprints.2015.1404v1

Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Zeffa, D. M., Nogueira, A. F., Ribeiro, L. A. B., et al. (2021). Genome-wide Association Study for Grain mineral Content in a Brazilian Common Bean Diversity Panel. *Theor. Appl. Genet.* 134, 2795–2811. doi:10.1007/s00122-021-03859-2

Doyle, J. J., and Doyle, J. L. (1990). Isolation of Plant DNA from Fresh Tissue. *Focus* 12, 139000007291403–139000007291415.

Elin, R. J. (1987). Status of the Mononuclear Blood Cell Magnesium Assay. *J. Am. Coll. Nutr.* 6, 105–107. doi:10.1080/07315724.1987.10720166

Falconer, D. S., Mackay, T. F., and Hardy-Weinberg, E. (19961996). *Introduction to Quantitative Genetics*. 4th ed. Harlow, UK: AddisonWesley Longman.

FAO (Food and Agriculture Organization) (2022). Food and Agriculture Organization. Available online: http://www.faostat.fao.org/site/339/default.aspx> (Accessed on January 20, 2022).

Farhat, N., Elkhouni, A., Zorrig, W., Smaoui, A., Abdelly, C., and Rabhi, M. (2016). Effects of Magnesium Deficiency on Photosynthesis and Carbohydrate Partitioning. *Acta Physiol. Plant* 38, 145. doi:10.1007/s11738-016-2165-z

Francis, T. R., and Kannenberg, L. W. (1978). Yield Stability Studies in Short-Season maize. I. A Descriptive Method for Grouping Genotypes. *Can. J. Plant Sci.* 58, 1029–1034. doi:10.4141/cjps78-157

Frossard, E., Bucher, M., Machler, F., Mozafar, A., and Hurrell, R. (2000). Potential for Increasing the Content and Bioavailability of Fe, Zn and Ca in Plants for Human Nutrition. *J. Sci. Food Agric.* 80, 861–879. doi:10.1002/(sici)1097-0010(20000515)80:7<861::aid-jsfa601>3.0.co;2-p

Gioia, T., Logozzo, G., Attene, G., Bellucci, E., Benedettelli, S., Negri, V., et al. (2013). Evidence for Introduction Bottleneck and Extensive Inter-gene Pool (Mesoamerica X Andes) Hybridization in the European Common Bean (Phaseolus vulgaris L.) Germplasm. *PLoS ONE* 8, e75974. doi:10.1371/journal.pone.0075974

Gröber, U., Schmidt, J., and Kisters, K. (2015). Magnesium in Prevention and Therapy. *Nutrients* 7, 503 8199–8226. doi:10.3390/nu7095388

Gunjača, J., Carović-Stanko, K., Lazarević, B., Vidak, M., Petek, M., Liber, Z., et al. (2021). Genome-Wide Association Studies of mineral Content in Common Bean. *Front. Plant Sci.* 12, 305. doi:10.3389/fpls.2021.636484

Han, G., Lu, C., Guo, J., Qiao, Z., Sui, N., Qiu, N., et al. (2020). C2H2 Zinc finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. *Front. Plant Sci.* 11, 115. doi:10.3389/fpls.2020.00115

Kang, W. H., Kim, S., Lee, H. A., Choi, D., and Yeom, S. I. (2016). Genome-wide Analysis of Dof Transcription Factors Reveals Functional Characteristics during Development and Response to Biotic Stresses in Pepper. *Sci. Rep.* 6, 33332–33412. doi:10.1038/srep33332

Kubota, T., Shindo, Y., Tokuno, K., Komatsu, H., Ogawa, H., Kudo, S., et al. (2005). Mitochondria Are Intracellular Magnesium Stores: Investigation by Simultaneous Fluorescent Imagings in PC12 Cells. *Biochim. Biophys. Acta (BBA) - Mol. Cel Res.* 1744, 19–28. doi:10.1016/j.bbamcr.2004.10.013

Liu, J., Li, Y., Wang, W., Gai, J., and Li, Y. (2016). Genome-wide Analysis of MATE Transporters and Expression Patterns of a Subgroup of MATE Genes in Response to Aluminum Toxicity in Soybean. *BMC Genomics* 17, 1–15. doi:10.1186/s12864-016-2559-8

Lopez-Ortiz, C., Peña-Garcia, Y., Natarajan, P., Bhandari, M., Abburi, V., Dutta, S. K., et al. (2020). The Ankyrin Repeat Gene Family in Capsicum Spp: Genome-wide Survey, Characterization and Gene Expression Profile. *Sci. Rep.* 10, 1–16. doi:10.1038/s41598-020-61057-4

Misra, G., Joshi-Saha, A., Salaskar, D., Reddy, K. S., Dixit, G. P., Srivastava, A. K., et al. (2020). Baseline Status and Effect of Genotype, Environment and Genotype × Environment Interactions on Iron and Zinc Content in Indian Chickpeas (Cicer Arietinum L.). *Euphytica* 216, 1–16. doi:10.1007/s10681-020-02673-z

Moraghan, J. T., Etchevers, J. D., and Padilla, J. (2006). Contrasting Accumulations of Calcium and Magnesium in Seed coats and Embryos of Common Bean and Soybean. *Food Chem.* 95, 554–561. doi:10.1016/j.foodchem.2004.10.060

Nadeem, M. A., Gündoğdu, M., Ercişli, S., Karaköy, T., Saracoğlu, O., Habyarimana, E., et al. (2020b). Uncovering Phenotypic Diversity and DArTseq Marker Loci Associated with Antioxidant Activity in Common Bean. *Genes* 11, 36. doi:10.3390/genes11010036

Nadeem, M. A., Habyarimana, E., Çiftçi, V., Nawaz, M. A., Karaköy, T., Comertpay, G., et al. (2018). Characterization of Genetic Diversity in Turkish Common Bean Gene Pool Using Phenotypic and Whole-Genome DArTseq-Generated silicoDArT Marker Information. *PLoS ONE* 13, e0205363. doi:10.1371/journal.pone.0205363

Nadeem, M. A., Habyarimana, E., Karaköy, T., and Baloch, F. S. (2021b). Genetic Dissection of Days to Flowering via Genome-wide Association Studies in Turkish Common Bean Germplasm. *Physiol. Mol. Biol. Plants* 27, 1609–1622. doi:10.1007/s12298-021-01029-8

Nadeem, M. A., Karaköy, T., Yeken, M. Z., Habyarimana, E., Hatipoğlu, R., Çiftçi, V., et al. (2020a). Phenotypic Characterization of 183 Turkish Common Bean Accessions for Agronomic, Trading, and Consumer-Preferred Plant Characteristics for Breeding Purposes. *Agronomy* 10, 272. doi:10.3390/agronomy10020272

Nadeem, M. A., Yeken, M. Z., Shahid, M. Q., Habyarimana, E., Yılmaz, H., Alsaleh, A., et al. (2021a). Common Bean as a Potential Crop for Future Food Security: an Overview of Past, Current and Future Contributions in Genomics, Transcriptomics, Transgenics and Proteomics. *Biotechnol. Biotechnological Equipment* 35 (35), 758–786. doi:10.1080/13102818.2021.1920462

Palčić, I., Karažija, T., Petek, M., Lazarević, B., Herak Ćustić, M., Gunjača, J., et al. (2018). Relationship between Origin and Nutrient Content of Croatian Common Bean Landraces. *Jcea* 19, 490–502. doi:10.5513/JCEA01/19.3.2103

Pour-Aboughadareh, A., Yousefian, M., Moradkhani, H., Poczai, P., and Siddique, K. H. M. (2019). STABILITYSOFT: A New Online Program to Calculate Parametric and Non-parametric Stability Statistics for Crop Traits. *Appl. Plant Sci.* 7, e01211. doi:10.1002/aps3.1211

Rathore, S. S., Epstein, A. J., Volpp, K. G. M., and Krumholz, H. M. (2004). Hospital Coronary Artery Bypass Graft Surgery Volume and Patient Mortality, 1998-2000. *Ann. Surg.* 239, 110–117. doi:10.1097/01.sla.0000103066.22732.b8

Ray, H., Bett, K., Tar'an, B., Vandenberg, A., Thavarajah, D., and Warkentin, T. (2014). Mineral Micronutrient Content of Cultivars of Field Pea, Chickpea, Common Bean, and Lentil Grown in Saskatchewan, Canada. *Crop Sci.* 54, 1698–1708. doi:10.2135/cropsci2013.08.0568

Rodiño, A. P., Santalla, M., González, A. M., De Ron, A. M., and Singh, S. P. (2006). Novel Genetic Variation in Common Bean from the Iberian Peninsula. *Crop Sci.* 46, 2540–2546. doi:10.2135/cropsci2006.02.0104

Rude, R. K. (1998). Magnesium Deficiency: A Cause of Heterogenous Disease in Humans. *J. Bone Miner. Res.* 13, 749–758. doi:10.1359/jbmr.1998.13.4.749

Sangronis, E., and Machado, C. J. (2007). Influence of Germination on the Nutritional Quality of Phaseolus vulgaris and Cajanus Cajan. *LWT - Food Sci. Technology* 40, 116–120. doi:10.1016/j.lwt.2005.08.003

Santos, A. L. D., Chaves-Silva, S., Yang, L., Maia, L. G. S., Chalfun-Júnior, A., Sinharoy, S., et al. (2017). Global Analysis of the MATE Gene Family of Metabolite Transporters in Tomato. *BMC Plant Biol.* 17, 1–13. doi:10.1186/s12870-017-1115-2

Seco-Gesto, E. M., Moreda-Piñeiro, A., Bermejo-Barrera, A., and Bermejo-Barrera, P. (2007). Multi-element Determination in Raft Mussels by Fast Microwave-Assisted Acid Leaching and Inductively Coupled Plasma-Optical Emission Spectrometry. *Talanta* 72, 1178–1185. doi:10.1016/J.talanta.2007.01.009

Shindo, Y., Fujii, T., Komatsu, H., Citterio, D., Hotta, K., Suzuki, K., et al. (2011). Newly Developed Mg2+-Selective Fluorescent Probe Enables Visualization of Mg2+ Dynamics in Mitochondria. *PLoS ONE* 6, e23684. doi:10.1371/journal.pone.0023684

Shrestha, S., Asch, F., Dusserre, J., Ramanantsoanirina, A., and Brueck, H. (2012). Climate Effects on Yield Components as Affected by Genotypic Responses to Variable Environmental Conditions in upland rice Systems at Different Altitudes. *Field Crops Res.* 134, 216–228. doi:10.1016/j.fcr.2012.06.011

Stagnari, F., Maggio, A., Galieni, A., and Pisante, M. (2017). Multiple Benefits of Legumes for Agriculture Sustainability: an Overview. *Chem. Bio. Technol.* 4, 1–13. doi:10.1186/s40538-016-0085-1

Turner, S. D. (2014). Qqman: an R Package for Visualizing GWAS Results Using QQ and manhattan 587 Plots. *Biorxiv*, 005165. doi:10.1101/Biorxiv.2014.005165

Verbruggen, N., and Hermans, C. (2013). Physiological and Molecular Responses to Magnesium Nutritional Imbalance in Plants. *Plant Soil* 368, 87–99. doi:10.1007/s11104-013-1589-0

Wang, N., Hatcher, D. W., Tyler, R. T., Toews, R., and Gawalko, E. J. (2010). Effect of Cooking on the Composition of Beans (*Phaseolus vulgaris* L.) and Chickpeas (*Cicer Arietinum* L.). *Food Res. Int.* 43, 589–594. doi:10.1016/j.foodres.2009.07.012

Wang, Y., Li, R., Li, D., Jia, X., Zhou, D., Li, J., et al. (2017). NIP1;2 Is a Plasma Membrane-Localized Transporter Mediating Aluminum Uptake, Translocation, and Tolerance in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5047–5052. doi:10.1073/pnas.1618557114

White, P. J., and Broadley, M. R. (2009). Biofortification of Crops with Seven mineral Elements Often Lacking in Human Diets - Iron, Zinc, Copper, Calcium, Magnesium, Selenium and Iodine. *New Phytol.* 182, 49–84. doi:10.1111/j.1469-8137.2008.02738.x

Wu, X., Li, R., Shi, J., Wang, J., Sun, Q., Zhang, H., et al. (2014). *Brassica oleracea* MATE Encodes a Citrate Transporter and Enhances Aluminum Tolerance in *Arabidopsis thaliana*. *Plant Cel Physiol* 55, 1426–1436. doi:10.1093/pcp/pcu067

Xing, H., Fu, X., Yang, C., Tang, X., Guo, L., Li, C., et al. (2018). Genome-wide Investigation of Pentatricopeptide Repeat Gene Family in poplar and Their Expression Analysis in Response to Biotic and Abiotic Stresses. *Sci. Rep.* 8, 1–9. doi:10.1038/s41598-018-21269-1

Yeken, M. Z., Nadeem, M. A., Karaköy, T., Baloch, F. S., and Çiftçi, V. (2019). Determination of Turkish Common Bean Germplasm for Morpho-Agronomic and mineral Variations for Breeding Perspectives in Turkey. *Kahramanmaraş Sütçü İmam Üniversitesi Tarım Ve Doğa Dergisiksu J. Agric. Nat.* 22, 38–50. doi:10.18016/ksutarimdoga.2019.vi563740

Zhang, Q., Xu, Y., Huang, J., Zhang, K., Xiao, H., Qin, X., et al. (2020). The rice Pentatricopeptide Repeat Protein PPR756 Is Involved in Pollen Development by Affecting Multiple RNA Editing in Mitochondria. *Front. Plant Sci.* 11, 749. doi:10.3389/fpls.2020.00749

Zhao, J. Y., Lu, Z. W., Sun, Y., Fang, Z. W., Chen, J., Zhou, Y. B., et al. (2020). The Ankyrin-Repeat Gene GmANK114 Confers Drought and Salt Tolerance in Arabidopsis and Soybean. *Front. Plant Sci.* 11, 584167. doi:10.3389/fpls.2020.584167

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 © 2022 Baloch, Nadeem, Sönmez, Habyarimana, Mustafa, Karaköy, Cömertpay, Alsaleh, Çiftçi, Sun, Chung and Chung. 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 | Frontiers in Genetics

ERRATUM
published: 15 July 2022
doi: 10.3389/fgene.2022.964612

Erratum: Magnesium-a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm

*Frontiers Production Office**

Frontiers Media SA, Lausanne, Switzerland

Keywords: ***Phaseolus vulgaris*, food legume, Mg contents, DArTseq, GWAS**

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Frontiers Production Office
production.office@frontiersin.org

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

Received: *08 June 2022*
Accepted: *08 June 2022*
Published: *15 July 2022*

Citation:
Frontiers Production Office (2022) Erratum: Magnesium-a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm. Front. Genet. 13:964612. doi: 10.3389/fgene.2022.964612

An Erratum on

Magnesium-a Forgotten Element: Phenotypic Variation and Genome Wide Association Study in Turkish Common Bean Germplasm
by Baloch, F. S., Nadeem, M. A., Sönmez, F., Habyarimana, E., Mustafa, Z., Karaköy, T., Cömertpa, G., Alsaleh, A., Çiftçi, V., Sun, S., Chung, G., and Chung, Y. S. (2022). Front. Genet. 13:848663. doi: 10.3389/fgene.2022.848663

Due to a production error, an author name was incorrectly spelled as "Gönül Cömertpa". The correct spelling is "Gönül Cömertpay".

The publisher apologizes for this mistake. The original version of this article has been updated.

Copyright © 2022 Frontiers Production Office. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fgene.2022.889795

GmARP is Related to the Type III Effector NopAA to Promote Nodulation in Soybean (*Glycine max*)

Jinhui Wang[1†], Chao Ma[1†], Shengnan Ma[1†], Haiyang Zheng[1], Haojie Feng[1], Yue Wang[1], Jiangxu Wang[2], Chunyan Liu[1], Dawei Xin[1], Qingshan Chen[1]* and Mingliang Yang[1]**

[1]College of Agriculture, Key Laboratory of Soybean Biology in Chinese Ministry of Education, Northeast Agricultural University, Harbin, China, [2]Heilongjiang Academy of Agricultural Sciences, Harbin, China

Type III effectors secreted by rhizobia regulate nodulation in the host plant and are important modulators of symbiosis between rhizobia and soybean (*Glycine max*), although the underlying mechanisms are poorly understood. Here, we studied the type III effector NopAA in *Sinorhizobium fredii* HH103, confirming its secretion into the extracellular environment under the action of genistein. The enzyme activity of NopAA was investigated *in vitro*, using xyloglucan and β-glucan as substrates. NopAA functions were investigated by the generation of a NopAA mutant and the effects of NopAA deficiency on symbiosis were analyzed. Soybean genes associated with NopAA were identified in a recombinant inbred line (RIL) population and their functions were verified. NopAA was confirmed to be a type III effector with glycosyl hydrolase activity, and its mutant did not promote nodulation. Quantitative trait locus (QTL) analysis identified 10 QTLs with one, *Glyma.19g074200* (*GmARP*), found to be associated with NopAA and to positively regulate the establishment of symbiosis. All these results support the hypothesis that type III effectors interact with host proteins to regulate the establishment of symbiosis and suggest the possibility of manipulating the symbiotic soybean–rhizobia interaction to promote efficient nitrogen fixation.

Keywords: soybean, symbiosis, nodulation, glycosyl hydrolase, QTL

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Javaid Akhter Bhat,
Nanjing Agricultural University, China
Yongqing Jiao,
Henan Agricultural University, China
Hongwei Jiang,
Jilin Academy of Agricultural Sciences (CAAS), China

***Correspondence:**
Dawei Xin
xdawei@163.com
Qingshan Chen
qshchen@126.com
Mingliang Yang
yml5419@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: *04 March 2022*
Accepted: *05 May 2022*
Published: *27 May 2022*

Citation:
Wang J, Ma C, Ma S, Zheng H, Feng H, Wang Y, Wang J, Liu C, Xin D, Chen Q and Yang M (2022) GmARP is Related to the Type III Effector NopAA to Promote Nodulation in Soybean (Glycine max). Front. Genet. 13:889795. doi: 10.3389/fgene.2022.889795

INTRODUCTION

Soybean is an important food crop, could provide rich plant protein and oil for human. With the increasing demand for protein and oil, more and more attention has been paid to soybean agricultural production (Li et al., 2019). In agricultural production, the yield of soybean is increased by applying a large amount of industrial nitrogen fertilizer (Wang et al., 2020a). However, excess application of nitrogen fertilizers has had adverse effects on the environment, leading also to economic and health concerns (Jordan et al., 2011; Chen et al., 2017; Guo and Wang, 2021). The use of symbiotic nodule-forming soybean maintains soil fertility and sustainable crop production, leading to reduced requirements for chemical N fertilizers in agricultural systems (Goyal et al., 2021). Many native *Rhizobium* species have been isolated and inoculated into soybean agricultural production to reduce the amount of applied N fertilizer (Chen et al., 2021). Rhizobia inoculations benefit the production of soybean as they promote plant growth by producing hormones, antibiotics, vitamins, and siderophores (Roy et al., 2020); in addition, rhizobia can enhance disease resistance (Plett et al., 2021; Zboralski et al., 2022).

The establishment of symbiosis system is a complex process, which requires signal exchange between soybean and rhizobia (Roy et al., 2020). In rhizobia, six secretion systems, namely, type I to VI systems, have been identified and classified according to their function (Freiberg et al., 1997; Krehenbrink and Downie, 2008; Pukatzki et al., 2009; Tampakaki, 2014; van Ulsen et al., 2014). Each secretion system plays a different role in symbiotic nitrogen fixation, with the type III secretion system (T3SS) being the most widely studied (Kambara et al., 2009; Miwa and Okazaki, 2017). The T3SS secretes nodulation outer proteins (Nops) into host cells which play key roles in the establishment of symbiosis. Several Nops have been identified in *Sinorhizobium fredii* HH103, including NopD, NopL, NopM, NopP, and NopT (Ausmees et al., 2004; Bartsev et al., 2004; Dai et al., 2008). NopD and NopM have been identified by LC-MALDI and LC-ESI (Rodrigues et al., 2007), and protein domain prediction has shown that the C-terminal region of NopD contains a domain homologous to the ubiquitin-like protease Ulp1, thus suggesting that NopD may combine with SUMO (small ubiquitin-like modifier)-conjugated proteins, leading to the removal of the SUMO conjugate in HH103-infected soybean (Xiang et al., 2020; Ratu et al., 2021). NopM is an E3 ubiquitin ligase and may target the MAPK pathway in host plants during nodule formation (Xin et al., 2012). NopL and NopP are substrates for soybean kinases (Skorpil et al., 2005; Zhang et al., 2011a; Ge et al., 2016). The NopL mutant was found to inhibit nodule formation, while the NopP mutant increased nodule numbers (Wang et al., 2018; Zhang et al., 2018; Zhang et al., 2021a). These reports indicate that the various type III effectors have different roles in symbiosis. The analysis and identification of type III effector proteins in soybean could provide a good basis for understanding the functions of these proteins in symbiosis. The release of the soybean (*Glycine max*) reference genome (Lam et al., 2010; Schmutz et al., 2010) has allowed the identification of many novel genes based on quantitative trait loci (QTLs) (Qi et al., 2014a; Lu et al., 2017). The Rj2 protein was cloned and found to recognize T3SS proteins (Yang et al., 2010). *GmNNL1*, an *R* gene that interacts with the rhizobial effector NopP, was identified by GWAS (Zhang et al., 2021a). NopD promotes HH103 infection of soybean by direct or indirect regulation of *GmPP2C*, and different haplotypes of *GmPP2C* were found to vary their expression patterns in response to NopD (Wang et al., 2020b). *GmRPK* was found to be related to NopL and to promote HH103 infection (Zhang et al., 2018). Although many proteins related to type III effectors have been identified in soybean, much work is still needed to uncover the mechanisms by which these proteins function.

In this study, we aimed to identify the genes associated with NopAA in regulating nodulation. We identified NopAA as a glycosyl hydrolase that hydrolyzed xyloglucan and β-glucan into glucose. Nodule tests in 13 soybean natural varieties were used to study the effects of NopAA mutation, indicating that NopAA positively influenced symbiosis. We identified 10 QTLs related to NopAA using the Dongnong594 and Charleston germplasms. Further haplotypic, genetic, and transgenic analyses showed that *Glyma.19g074200* (*GmARP*) responded to NopAA to regulate nodulation. These findings thus provide useful information on the role of NopAA in nodulation, and the identification and study of soybean host genes can provide support for the effective utilization of symbiotic nitrogen fixation in soybean.

MATERIALS AND METHODS

Strains, Vectors, and Primers

S. fredii HH103 and *Escherichia coli* DH5α and BL21 (DE3.1) were used in this study (**Supplementary Table S2**). HH103 was cultured in or on TY medium supplemented with rifampicin at 28°C, and all *E. coli* strains were cultured in or on LB media supplemented with appropriate antibiotics at 37°C. The plasmids used for the construction of mutants and the primers used for PCT and qRT-PCR are listed in **Supplementary Table S1**. Antibiotics were used at the following concentrations: 50 μg/ml rifampicin; 50 μg/ml kanamycin; 34 μg/ml chloramphenicol; 50 μg/ml spectinomycin; 50 μg/ml streptomycin.

Construction of the NopAA Mutant

The NopAA mutant was constructed by triparental mating as previously described (Wang et al., 2019) with the kanamycin-resistant DNA fragment inserted into the coding sequence of NopAA by homologous recombination. The mutant was confirmed by PCR and Southern blotting (Chen and Kuo, 1993).

qRT-PCR and Analysis of Nodulation Outer Proteins

HH103, the NopAA mutant, and the TtsI mutant were grown in YM medium (Yeast Mannitol Broth) in the presence or absence of genistein, and extracellular proteins were extracted as previously described (Jim é nez-Guerrero et al., 2015). Extracted RNA was subjected to qRT-PCR analysis by a Roche LightCycler LC480 (Roche, Switzerland), and the relative expression levels of *NopAA* were normalized using *16S rRNA*. All sample collections were analyzed using three independent biological replicates as well as three technical replicates.

Enzyme Activity Assays

The NopAA genes were then independently subcloned into pET28b as *Bam*H I-*Sal* I fragments to create pET28b-NopAA. This vector was used for active protein purifications as described (Xin et al., 2012). The His-NopAA protein was purified from BL21 (DE3) *E. coli* after induction with 0.1 mM isopropylthio-β-galactoside using Ni-NTA resin beads (Qiagen, Hilden, Germany). The purified protein was identified by western blotting with an anti-His antibody and its enzyme activity was measured using pure xyloglucan and beta-glucan (Megazyme, Bray, Ireland) as substrates. Activity assays were performed by glucose as the standard and a iMark Microplate Absorbance Reader (Bio-Rad, United States) as previously described (Grishutin et al., 2004).

Agrobacterium Infiltration Assays

Agrobacterium-mediated transformation of tobacco leaves was performed as described (Wang et al., 2011). *A. tumefaciens* EHA105 containing pSoy1-*NopAA*, the empty vector pSoy1,

and pSoy1 *35S-NopM* were transformed into leaves of 30-day-old *Nicotiana benthamiana*. The OD_{600} of the *A. tumefaciens* culture was adjusted to 0.2 with buffer (Ma et al., 2015). Infiltrated leaves were harvested each day for analysis of electrical conductivity by a conductivity meter (Lei-ci, China) (Wang et al., 2020a).

Nodulation Tests

The RILs population were planted in Xiangyang Farm, Harbin, Heilongjiang Province (Harbin, latitude 45°450″N, longitude 126°380″E) in 2016, and nodulation tests were performed in 2017. 100 soybean germplasms used in the study came from different ecoregions were planted in Xiangyang Farm in 2019, and nodulation tests were performed in 2020. Sterilized soybean seeds (15 seeds from each variety) were grown under greenhouse conditions in sterilized vermiculite containing a low-nitrogen nutrient solution (Zhang et al., 2018). All plants were grown in a greenhouse at 25°C with a photoperiod of 16 h. At the Vc stage (unifoliolate leaves expand), all the plants were inoculated with HH103 and the NopAA mutant. Nodulation in the germplasms was evaluated by nodule number (NN) and nodule dry weight (NDW) (28 days post-inoculation, dpi). Three different biological replicates and 15 plants for each replicate were conducted and the *t*-test was used to detect the statistical significance of differences in NN and NDW.

Observation of Rhizobial Infection Events

DN594 and Charleston were inoculated with GFP-conjugated HH103 and NopAA mutant strains. One-centimeter samples of infected lateral roots were collected at 1 dpi, comprising ten lateral roots per plant, and infection in the entire plant was evaluated using confocal fluorescence microscopy (Zeiss LSM700, Germany). The number of infection events was then multiplied by the length of the roots. Infection events included foci, infection threads in epidermal cells (ITs), and infection threads extending into the cortical cells (rITs) (Liu et al., 2020). Ten independent lateral root segments from each plant were used for each biological replicate and three biological replicates were used for each condition.

Quantitative Trait Locus Mapping and Identification of Genes in Quantitative Trait Locus Regions

Charleston and DN594 were used to construct the RIL population, using a genetic map that had been constructed in a previous study (Qi et al., 2014b). We used WinQTL Cartographer and composite interval mapping methods to map QTLs related to nodulation in the RIL population. QTLs for NN and NDW were defined by LOD scores greater than 3.0 (Brensha et al., 2012; Wang et al., 2020b). The reference genome of Williams 82 was used to identify genes within the QTL regions, and candidate genes were annotated according to the gene annotation information of Williams 82 (Xin et al., 2016).

Analysis of the Expression Patterns of Candidate Genes Related to NopAA

qRT-PCR was performed to analyze the expression patterns of candidate genes in Charleston and DN594 inoculated with HH103, the NopAA mutant, and the control. Roots were harvested at several time points and rapidly submerged in liquid nitrogen, after which the total RNA was extracted by the TRIzol method. cDNA was synthesized using the PrimeScript™ RT Reagent Kit (Takara Biotech Co., Beijing, China) and qRT-PCR was performed using TB Green® Premix Ex Taq™ II (Takara Biotech Co.). Relative expression levels of each gene were normalized using *GmUNK1* (*Glyma.12g020500*) (Wang et al., 2020a).

Hairy-Root Transformation of Soybean

The *A. rhizogenes* strain K599 containing pSoy10-*GmARP-GFP*, pSoy10-*GFP*, pB7GWIWG2-*GmARP-DsRed,* and pB7GWIWG2-*DsRed* were used for hairy-root transformation as described previously (Kereszt et al., 2007). GFP and DsRED were used as markers to detect positive hairy roots using LUYOR-3415RG. Hairy roots were inoculated with HH103, the NopAA mutant, and the control to identify the influence of gene overexpression and interference on nodulation. All nodules were harvested at 28 dpi and transcript levels of *GmARP* were measured using qRT-PCR. Three different biological replicates and 15 plants for each replicate were conducted and the *t*-test was used to detect the statistical significance of differences in NN and NDW.

Haplotype Analysis

Haplotype analysis of *GmARP* was performed in 100 soybean natural varieties. The genomic sequence of *GmARP,* including the coding sequence and the 3000-bp promoter sequence, was obtained from the resequencing of the genomes of the 100 varieties, and the genomic sequences were subjected to local BLAST analysis to obtain SNP information (Wang et al., 2020a). Haploview 4.2 software (Cambridge, MA, United States) was used for analysis using the Haps Format module, and GraphPad Prism 8 software was used to analyze the correlations between the *GmARP* haplotypes and nodule traits in the 100 natural varieties (Li et al., 2019).

RESULTS

NopAA is Induced by Genistein and Secreted by Rhizobia

A phylogenetic analysis was performed to study the evolutionary relationships of NopAA in different rhizobial species. This showed that the NopAA tree was separated into three branches (blue, green, and pink regions in **Figure 1A**). The blue region contained only four *Sinorhizobium* species, the green region contained only three *Bradyrhizobium* species, and the pink region contained only two *Mesorhizobium* species (**Figure 1A**). The phylogenetic analysis showed that there is obvious diversity in NopAA among different rhizobial genera. Previous RNA-seq results showed that NopAA might be induced by genistein; thus, to further study whether NopAA was a type III effector, qRT-PCR and nodulation analyses were performed. The qRT-PCR results showed that genistein increased NopAA expression significantly in HH103 compared with rhizobia not

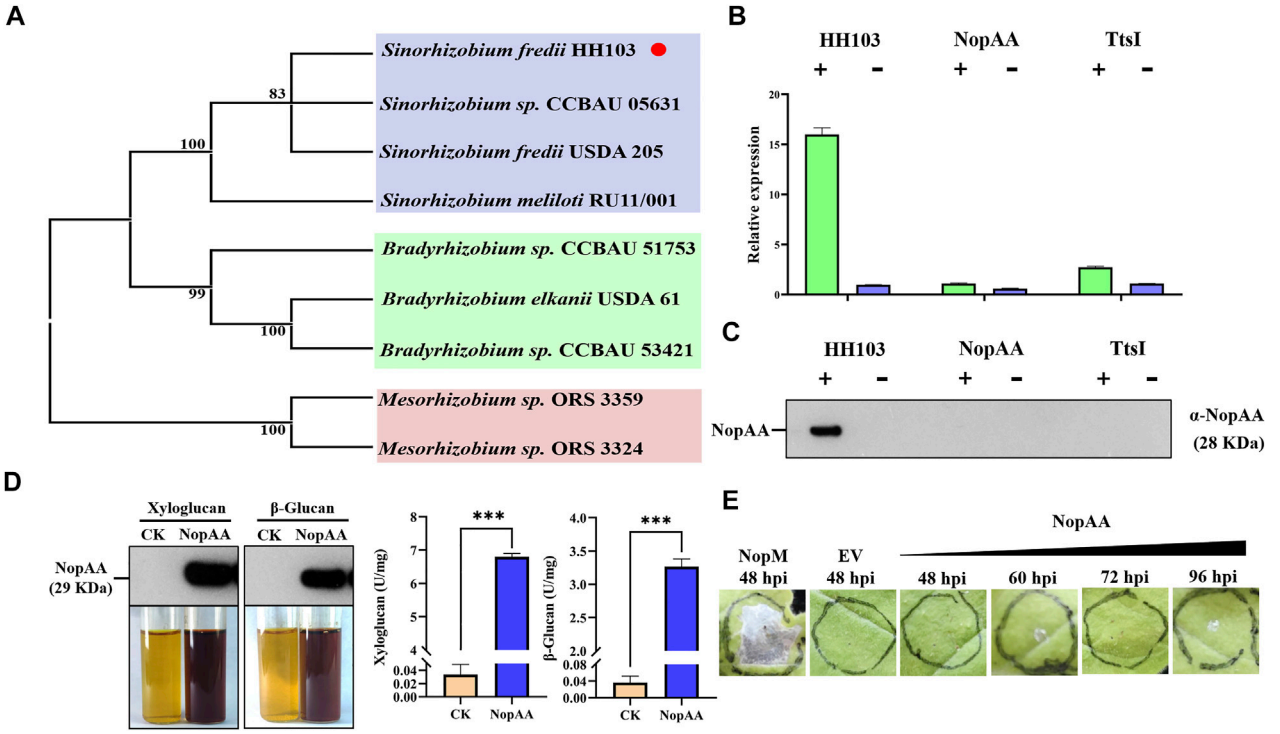


FIGURE 1 | NopAA is a rhizobial type III effector with glycoside hydrolase activity. **(A)** Phylogenetic tree analysis of NopAA proteins from different rhizobia. **(B)** Relative expression of NopAA in rhizobia in the presence and absence of genistein. **(C)** Analysis of NopAA, a nodulation-promoting outer protein secreted by rhizobia, by immunoassay in the presence and absence of genistein. In **(B)** and **(C)**, "+" represents induction by the addition of genistein, and "–" represents the absence of genistein. **(D)** Analysis of glycoside hydrolase activity of NopAA. Pure xyloglucan and β-glucan were used as substrates for measuring enzyme activity, and NopAA protein in the reaction system was detected by immunoblotting. Data are presented as the average of three different biological replicates. **(E)** *N. benthamiana* leaves 48 h–96 h after inoculation with EH105 carrying *NopAA* and positive control *NopM*.

treated with genistein but did not induce NopAA expression in the NopAA mutant or the TtsI mutant regardless of the presence or absence of genistein (**Figure 1B**). After extraction of the nodulation outer proteins, NopAA was detected in the supernatants after genistein treatment but not in the supernatants of the NopAA mutant or TtsI mutant outer proteins induced by genistein (**Figure 1C**). These results indicated that NopAA could be induced by genistein and TtsI and could be secreted into the extracellular milieu of the rhizobia.

NopAA has Glycosyl Hydrolase Activity and Does not Induce Cell Death in Tobacco

To determine whether NopAA had glycosyl hydrolase activity, 6 × His-tagged NopAA was expressed and purified by Ni-affinity chromatography. The molecular weight of the His-NopAA fusion protein was ~28 kDa (**Figure 1D**). Xyloglucan and β-glucan were used as substrates to analyze hydrolase activity, showing that NopAA hydrolyzed both xyloglucan and β-glucan directly to sugars (**Figure 1D**). The *Phytophthora sojae* apoplastic effector PsXEG1 and NopAA belong to the glycosyl hydrolase family, which is known to induce cell death in tobacco leaves. To study the effect of NopAA on tobacco leaves, we infiltrated NopAA into tobacco leaves using *Agrobacterium tumefaciens* EHA105. In contrast to PsXEG1, NopAA did not trigger cell death (**Figure 1E**). These results indicated that NopAA had glycosyl hydrolase activity and did not induce cell death in tobacco.

NopAA Mutant Does not Promote Nodulation in Several Soybean Germplasms

In this study, we selected 13 soybean germplasms to study the effect of NopAA on nodulation, including five cultivated, six landrace soybean, and two wild soybean germplasms (**Supplementary Figure S1**). Nodulation tests showed that while the NopAA mutant significantly reduced the nodule numbers in 11 of the germplasms, the nodule numbers did not differ significantly in Dongnong594 (DN594) and Qingdou compared with the wild strain HH103. Inoculation with the NopAA mutant led to lower nodule dry weights in 12 of the 13 germplasms except for DN594, compared with the wild strain HH103. Nodulation tests of DN594 and Qingdou showed that the NopAA mutant did not influence nodule numbers, nor did it alter the change in the dry weight of DN594. These results indicated that NopAA did not negatively influence symbiosis in different types of soybean germplasms, and that differences in the genetic backgrounds of the germplasms might be the explanation for the variations in

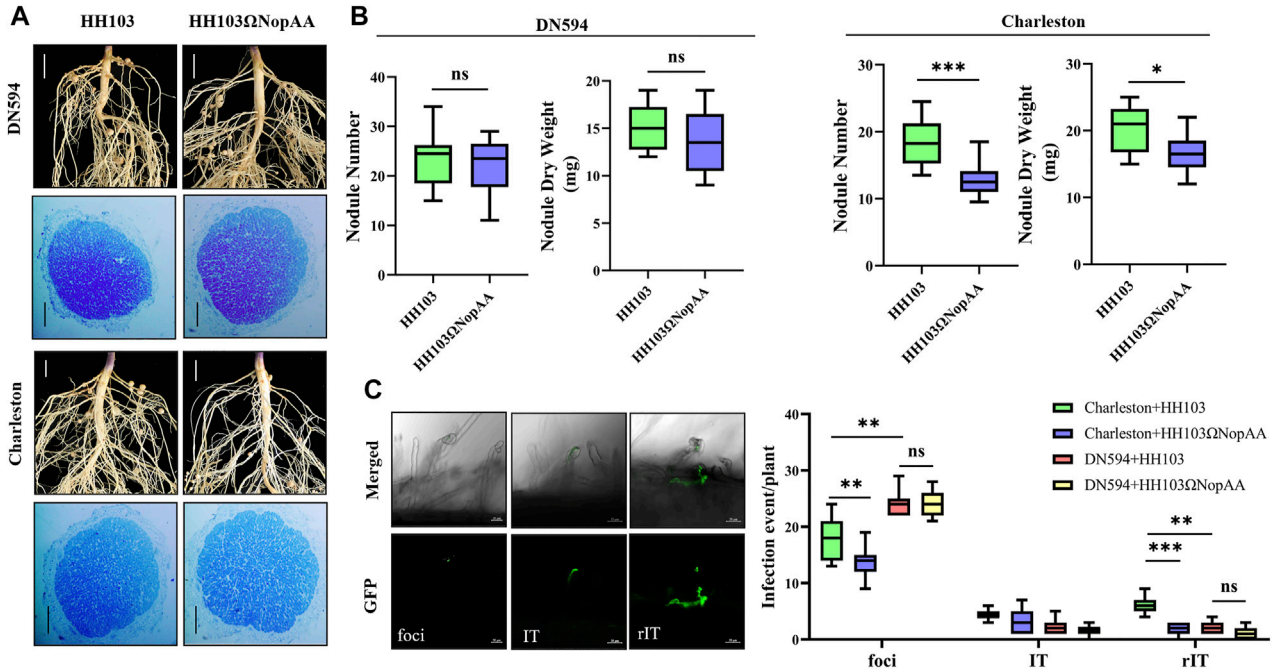


FIGURE 2 | Differences in infection events between DN594 and Charleston after inoculation with HH103 or the NopAA mutant. **(A)** Phenotypes of DN594 and Charleston inoculated with HH103 and HH103ΩNopAA; root phenotype scale bars represent 0.5 cm; toluidine blue staining of nodules scale bars represent 50 μm. **(B)** Boxplots of NN and NDW. **(C)** Boxplots of the total number of infection events per plant of Charleston and DN594 (1 dpi). Data are presented as the average of three different biological replicates and ten root segments for each replicate, and all results were analyzed for significance using *t*-tests. Foci, infection foci; IT, infection thread in an epidermal cell; rIT, infection thread extending into a cortex cell. dpi, day post-inoculation.

nodule traits associated with NopAA. This suggested that these genetic differences could be used to identify genes or loci associated with symbiosis.

The Influence of NopAA on Dongnong594 and Charleston

Based on the results of the nodule tests on various soybean germplasms, DN594 and Charleston were selected for further investigation of the role of NopAA in symbiosis. Compared with the wild strain HH103, the NopAA mutants caused a significant reduction in both the nodule number (NN) and nodule dry weight (NDW) in Charleston but did not affect DN594 (**Figures 2A,B**). In the mature nodule, the infection zone and infection cells did not differ significantly between the two germplasms after inoculation with HH103, and inoculation with the NopAA mutant did not change this (**Figure 2A**). Because of the differences in the nodule traits of DN594 and Charleston, we further investigated the infection events after inoculation with GFP-labeled HH103 and the NopAA mutant. The results showed there were more infection events in curled root hairs in Charleston compared with DN594 3 days after inoculation with GFP-labeled HH103 compared with the wild strain, while the GFP -labeled NopAA mutant induced fewer infection events in Charleston but had no significant effects on DN594 (**Figure 2C**). This implied that the type III effector NopAA might influence nodulation through involvement in infection events and that the genetic differences between DN594 and Charleston might result in different symbiotic phenotypes, with some of the genetic differences being associated with NopAA.

Quantitative Trait Locus for Nodule Number and Nodule Dry Weight Related to NopAA in Soybean Recombinant Inbred Line Populations

The soybean RILs were derived from the cross and continuous self-cross of DN594 and Charleston and were used to identify genes related to NopAA. Nodulation tests were performed on 150 RILs after inoculation with the wild strain and the NopAA mutant, with the results showing that the NopAA mutant could significantly reduce both the NN and NDW in whole RILs compared with HH103, further suggesting that NopAA plays a positive role in symbiosis. WinQTL Cartographer was used to identify the QTLs underlying nodule-related traits using a composite interval mapping method. Two main-effect QTLs underlying NDW and five QTLs underlying NN were identified with the background inoculated with *S. fredii* HH103 (**Table. 1**). Two QTLs, *QNDW3-1* and *QNDW7-1* were associated with NDW and explained 4.40% and 2.42% of the phenotypic variation, respectively. Five QTLs, *QNN9-1, QNN13-1, QNN13-2, QNN16-1,* and *QNN17-1.* were associated with NN, explaining 1.84%, 0.99%, 0.39%, 5.89%, and 6.13% of

TABLE 1 | Nodule traits in RILs inoculated with HH103 and HH103ΩNopAA.

Strain	RILs (*n* = 150)				Parents (average)	
	Trait	Average	Standard deviation	Coefficient of variation	Charleston	DN594
HH103	NN	11.4	5.6	48.95	21.0 ± 4.0	31.6 ± 5.3
	NDW	14.0	17.6	125.71	12.3 ± 2.8	19.5 ± 2.1
NopAA mutant	NN	8.1*	5.9	62.50	14.3 ± 4.2**	29.9 ± 6.2
	NDW	6.2*	3.9	185.85	5.5 ± 1.7**	18.8 ± 3.4

*Note: NN, nodule number; NDW, nodule dry weight. * indicates significant differences with different inoculations,* $p \leq 0.05$, *** indicates* $p \leq 0.01$.

TABLE 2 | Main QTLs identified in the RIL.

Strain	Trait	QTL	Chr.	Start position	End position	LOD	R^2(%)	Putative causal genes or QTLs identified in previous study
HH103	NDW	*QNDW6-1*	6	13,842,716	14,216,591	3.0	4.40	Nodule size 1–3 Hwang et al. (2013)
		QNDW7-1	7	5, 985, 880	6, 014, 687	2.7	2.42	
	NN	*QNN9-1*	9	36, 047, 415	40, 663, 169	3.3	1.84	
		QNN11-1	11	30,891,632	33,719,971	2.8	0.99	Nodule weight per plant, dry 1–9 Nicol á s et al. (2006)
		QNN13-1	13	8, 930, 546	9, 544, 895	2.4	0.39	
		QNN16-1	16	8, 511, 616	9, 216, 892	3.7	5.89	
		QNN17-1	17	32, 105, 983	32, 555, 254	3.5	6.13	
NopAA mutant	NDW	*QNDW19-1*	19	19, 068, 552	19, 533, 854	3.3	0.15	
		QNDW19-2	19	25, 922, 562	27, 078, 098	3.9	2.59	
	NN	*QNN19-1*	19	25, 922, 286	27, 078, 098	3.6	4.19	

Note: Chr., chromosome; LOD, log-of-odds; QTL, quantitative trait loci; LOD score cutoff of major QTLs was determined by permutation tests (1,000 times; $p < 0.05$).

the phenotypic variation, respectively. A total of 33 candidate genes were found in these seven loci (**Supplementary Table S3**). In the background inoculated with the *S. fredii* NopAA mutant, two QTLs underlying NDW were located on chromosome 19, while only one QTL underlying NN was identified on chromosome 19 (**Table 2**). The QTLs *QNDW19-1* and *QNDW19-2* explained 0.15% and 2.59% of the phenotypic variation, respectively QTLs *QNDW19-2* and *QNN19-1* were in the same region, with very small R^2 values. Twenty genes were identified in these loci (**Supplementary Table S3**). As no comparable region was identified in the backgrounds of *S. fredii* HH103 and the *S. fredii* NopAA mutant and as *QNDW19-2* and *QNN19-1* were in the same region, we considered that this QTL region might contain the key candidate genes for interaction with NopAA. Gene functional annotation showed that gene *Glyma.19G074000* encoded a nodulin protein, while *Glyma.19G073900* encoded a DNA lyase, and *Glyma.19G074200* was co-expressed with genes in the root-specific co-expression subnetwork. *Glyma.19G073800* and *Glyma.19G074100* had no specific functional annotations.

qRT-PCR Verification of Candidate Genes

The candidate genes were verified by qRT-PCR to confirm their association with NopAA, using the roots of both DN594 and Charleston, as well as non-inoculated roots (**Figure 3**). Among these genes, *Glyma.19G074200* was confirmed to be upregulated with an almost two-fold change after inoculation with *S. fredii* HH103 but not with the *S. fredii* NopAA mutant 6 h after inoculation. This supported the association between *Glyma.19G074200* (named *GmARP*) and NopAA. The remaining genes were not significantly affected by the wild-type and mutant rhizobial inoculations.

The Effects of *GmARP* Silencing and Overexpression on Nodulation

To further investigate the role of *GmARP* in soybean nodulation, the *Agrobacterium rhizogenes* strain K599 carrying pB7GWIWG2(II)-*Dsred-GmARP* (for RNA interference), pSoy10-*GmARP-GFP* (for overexpression), and the corresponding empty vectors were used for the transformation of soybean transgenic hairy roots. Silencing and overexpression were confirmed by qRT-PCR after transformation (**Supplementary Figure S2**). Reduced NN and NDW values were seen after *GmARP* silencing compared with the controls inoculated with HH103 and the NopAA mutant, respectively (**Figure 4A**). In the silenced hairy roots, the NopAA mutant did not induce significantly different NN or NDW compared with HH103 (**Figure 4B**). In the overexpressed hairy roots, *GmARP* overexpression resulted in increased NN and NDW values compared with the controls inoculated with HH103 and the NopAA mutant, respectively (**Figure 4B**). We found that overexpression of *GmARP* led to elevated NN and NDW in comparison with the control hairy roots inoculated with the NopAA mutant. However, both the NN and NDW after inoculation with the NopAA mutant were still lower than HH103 in the overexpressed hairy roots, suggesting that overexpression could partially compensate for the loss of

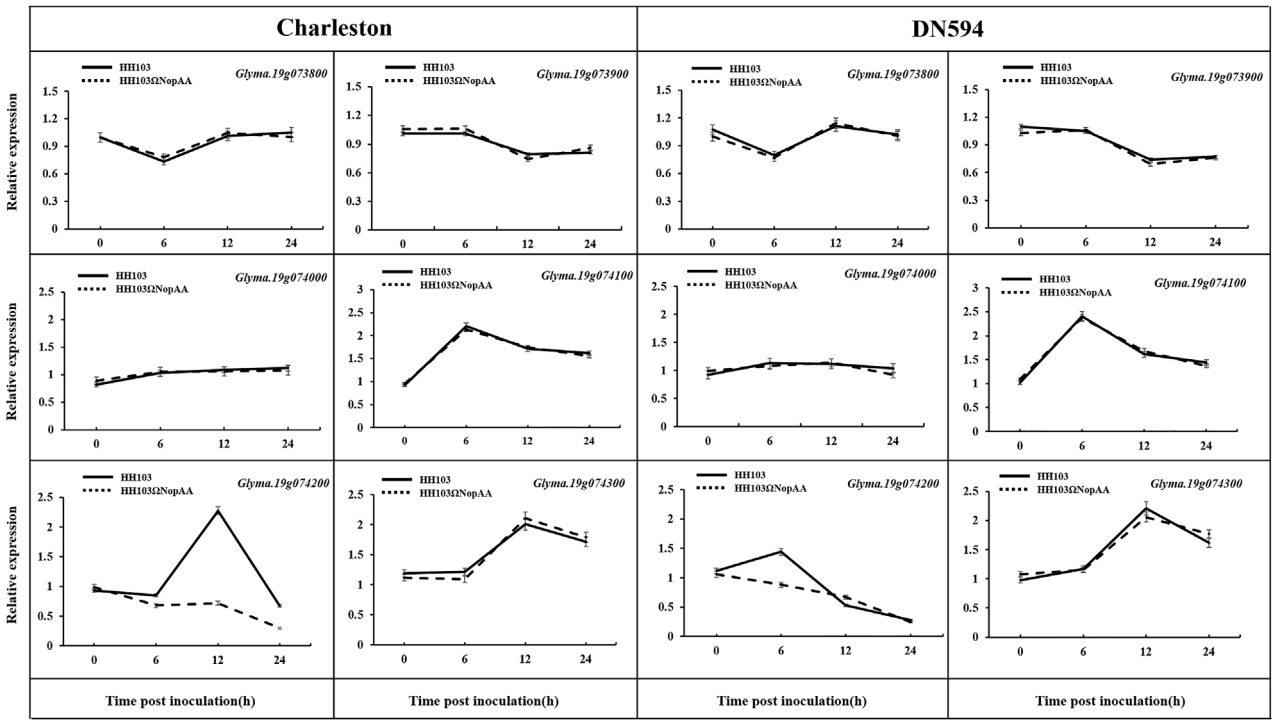


FIGURE 3 | Analysis of the relative expression levels of candidate genes in the QTL interval by qRT-PCR. Relative expression of genes was calculated by the $2^{-\Delta\Delta Ct}$ method, and the relative expression levels of each gene were normalized using *GmUNK1* (*Glyma.12g020500*). Data are means ± SE of three replicates.

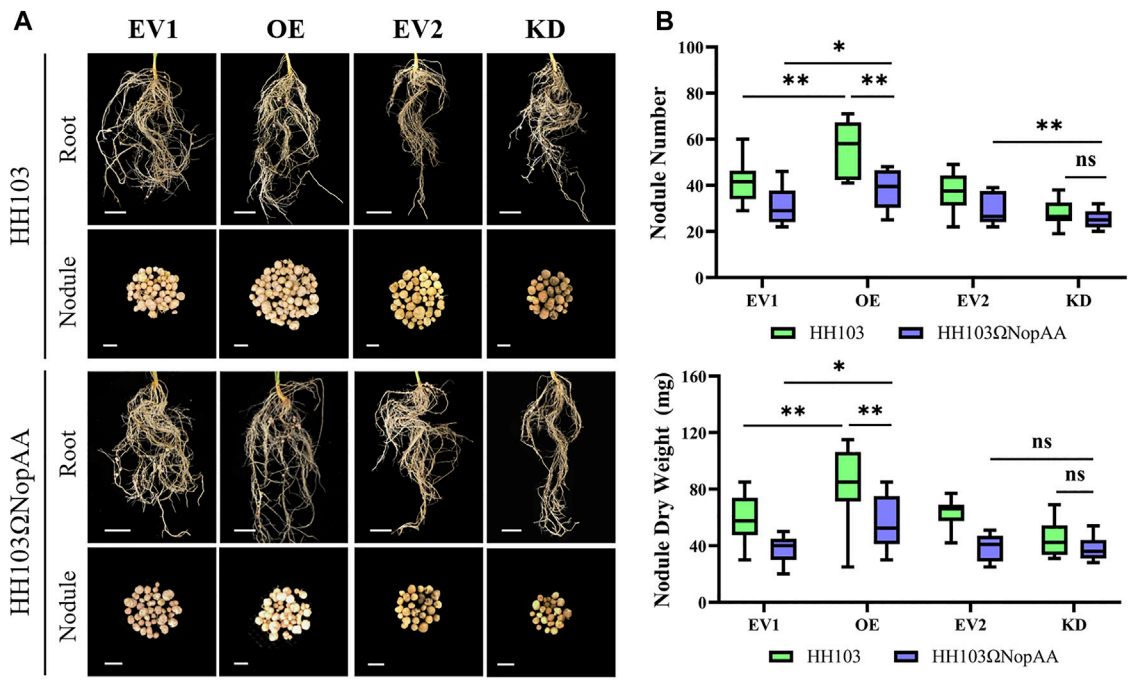
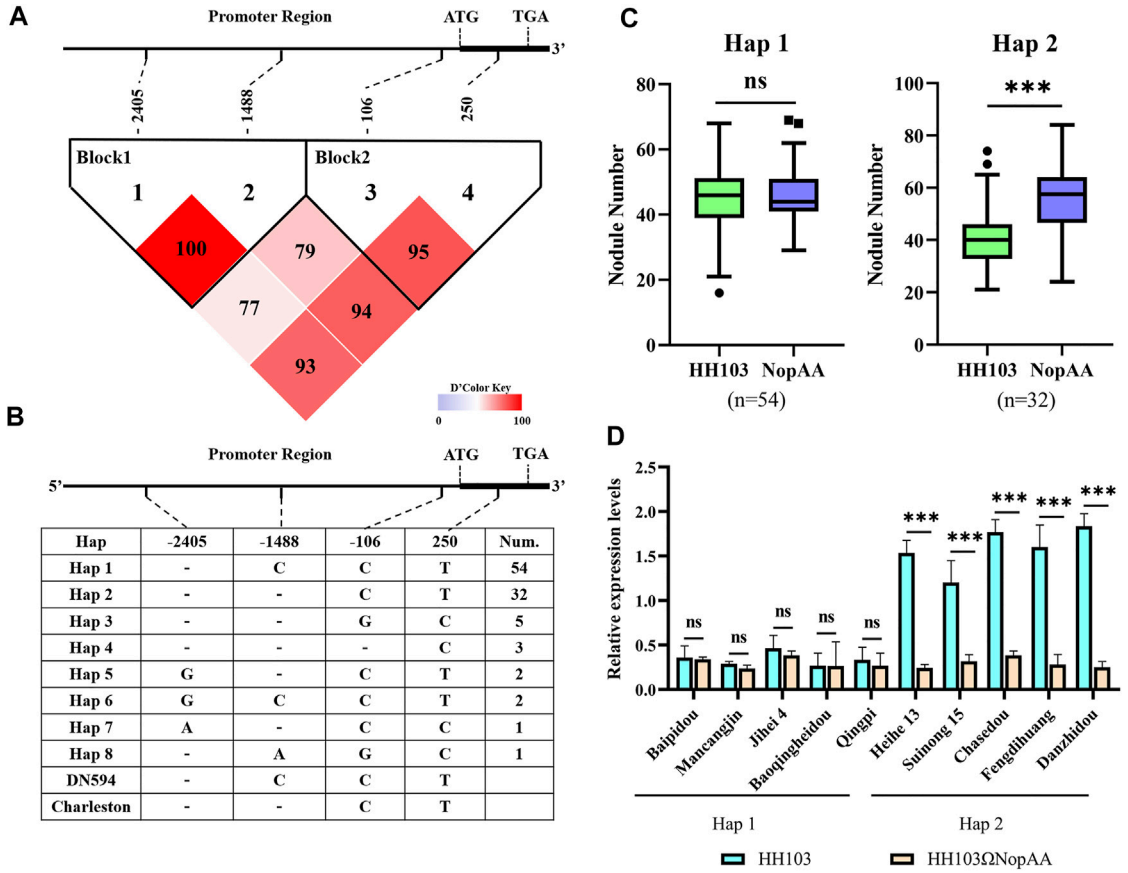


FIGURE 4 | Nodule phenotypes of EV1, OE, and KD. **(A)** Nodular phenotypes of hairy roots transformed with EV1, OE, EV2, and KD after inoculation with HH103 and HH103ΩNopAA (28 dpi). EV1, Empty vector for gene overexpression; OE, Overexpression of *GmARP*; EV2, Empty vector for gene silencing; KD, Gene silencing of *GmARP* by RNAi. Root: scale bars represent 1 cm; nodule: scale bars represent 2 mm. **(B)** Boxplots of nodule phenotypes. Data are presented as the average of three different biological replicates and 15 plants for each replicate and significance was determined by *t*-tests.



Hap	-2405	-1488	-106	250	Num.
Hap 1	-	C	C	T	54
Hap 2	-	-	C	T	32
Hap 3	-	-	G	C	5
Hap 4	-	-	-	C	3
Hap 5	G	-	C	T	2
Hap 6	G	C	C	T	2
Hap 7	A	-	C	C	1
Hap 8	-	A	G	C	1
DN594	-	C	C	T	
Charleston	-	-	C	T	

FIGURE 5 | Haplotype analysis of *GmARP*. **(A,B)** Haplotype analysis of *GmARP* from 100 soybean resources. **(C)** NN of Hap1 and Hap2 inoculated with HH103 and HH103ΩNopAA, data are presented as the average of three different biological replicates and 15 plants for each replicate. **(D)** Gene expression of *GmARP* of Hap1 and Hap2 inoculated with HH103 and HH103ΩNopAA (12 hpi). Significance was measured by *t*-tests. NN, nodule number.

NopAA in HH103 during symbiotic formation. These results indicated that *GmARP* may be involved in signal transduction after recognition of NopAA by the soybean, leading to the regulation of nodulation.

Haplotype Analysis Suggests That *GmARP* Regulates Nodule Number and Nodule Dry Weight in Soybean

Based on the observed differences in the recognition of NopAA, haplotype (Hap) analysis of *GmARP* was performed in 100 soybean natural varieties using Dnasp5.0 software. Using the results of the resequencing of soybean natural varieties, eight Haps were identified. Two Haps containing more than 10 accessions were considered the dominant Haps, and only one SNP and one indel were observed in the promoter regions of Hap1 and Hap2 (**Figures 5A,B**). What is even more remarkable is that there was no difference in the NN and NDW values in the Hap1 accessions inoculated with HH103 compared with the NopAA mutant; however, in the Hap2 accessions, the NopAA mutant induced both lower NN and NDW than HH103 (**Figure 5C**). To identify whether the difference in the promoter region caused the expression difference between Hap1 and Hap2, qRT-PCR was used to analyze the expression of *GmARP* at 12 h post-inoculation with HH103 and the NopAA mutant in some varieties of Hap1 and Hap2. The expression patterns suggested that the relative expression level of *GmARP* did not differ between several Hap1 accessions (Heihe13 and Suinong15) regardless of the presence of NopAA in HH103 (**Figure 5D**). However, in the Hap2 accessions, HH103 induced significant expression of *GmARP* compared with the NopAA mutant. The haplotype analysis further confirmed that *GmARP* was associated with the type III effector NopAA to regulate nodulation in soybean.

DISCUSSION

In this study, we confirmed that NopAA was a type III effector and could hydrolyze xyloglucan and β-glucan directly into glucose. Nodulation tests indicated that NopAA did not have a negative effect on nodulation, and the soybean gene *GmARP* was found to be associated with NopAA in the soybean RIL population. There are few studies on NopAA function and the mechanisms involved, and few address the question of its role in symbiosis. Our study of NopAA

provides an essential foundation for understanding the role of NopAA and its associated signaling pathways in symbiotic nitrogen fixation.

The type III effectors of rhizobia have obvious characteristics, that is, they are induced and regulated by flavonoids and TtsI, and the most important is that they can be secreted into the extracellular of rhizobia (Teulet et al., 2022). Similar to the reported type III effectors, such as NopD, NopL and NopM (Zhang et al., 2011b; Xin et al., 2012; Wang et al., 2020a), NopAA can also be induced by flavonoids and its expression can be regulated by TtsI, these confirmed NopAA is a type III effector. In the previous study, several type III effectors can act as enzymes and have biochemical activities (Teulet et al., 2022), and elucidating the biochemical functions of type III effectors, particularly those that are conserved among different species, has greatly enhanced our understanding of the mechanisms underlying bacterial pathogenesis (Chen et al., 2021; Zboralski et al., 2022). By analysis of NopAA amino acid sequence, we found that NopAA belonged to the Glycoside hydrolase 12 (GH12) family and had the potential to hydrolyze polysaccharides (Zhang et al., 2021b). Nodulation requires the assistance of cell-wall degrading enzymes, such as hemicellulases, pectinases, polygalacturonases, glucanases, cellulases, xyloglucanases, and pectinases, which allow root penetration by beneficial microorganisms, such as *Rhizobium, Frankia*, and the arbuscular mycorrhizal (Mateos et al., 1992; Zhang et al., 2021b; Tsyganova et al., 2021). GH12 is a cell wall-degrading enzyme that, along with the bacterial effector Cif and the *Phytophthora sojae* apoplastic effector PsXEG1, have been identified as playing pivotal roles in pathogen survival and systemic infection processes (Ma et al., 2015; Xia et al., 2020). Identifying the host protein interactors could elucidate the mechanisms underlying host-microbe interactions. NopAA was identified as a glycoside hydrolase belonging to the GH12 family. In *Phytophthora sojae*, PsXEG1 is also a glycoside hydrolase 12 family member with glycoside hydrolase activity and hydrolases the same substrates as NopAA. Because PsXEG1 is an effector of *P. sojae*, the replication of a *P. sojae* virus could be promoted using a PsXLP1 decoy that was similar to PsXEG1 but without its enzymatic activity. Additionally, PsXLP1 can protect PsXEG1 from GmGIP1 binding *in vitro* and *in planta* (Ma et al., 2015; Ma et al., 2017). We did not find any proteins similar to NopAA in the *S. fredii* HH103 genome, suggesting that HH103 does not infect soybeans using the same decoy pattern as *P. sojae* by NopAA. The investigation of the effect of NopAA on tobacco leaves did not show any evidence of cell death. Rhizobia may be involved in soybean immune signaling pathways during infection and the establishment of the symbiotic relationship. However, the establishment of an effective symbiosis does not require as strong or prolonged an immune response as occurs against pathogens. These results differ significantly from the actions of PsXEG1 in tobacco leaves. Thus, we propose that the role of NopAA may be different from that of PsXEG1. In this study, xyloglucan and β-glucan were found to be hydrolyzed to glucose; both xyloglucan and β-glucan are important components of cellulose and hemicellulose in plant cell walls, and their hydrolysis could promote the entry of rhizobia into host cells. These results are consistent with the observations of infection sites as fewer infection sites were observed with the NopAA mutant, indicating that NopAA could degrade the cell walls of soybean root hair cells, thus promoting infection and the formation of the infection thread.

A total of 10 QTLs were identified in this study, two of which were related to NDW, and five were related to NN with HH103, with two QTLs related to NDW and only one QTL related to NN seen with the NopAA mutant. No overlapping QTL loci were found when comparing inoculated HH103 and NopAA mutant, because NopAA mutation obviously changed the NN and NDW of DN594, Charleston and RIL populations, so no overlapping QTL could be identified, this result was similar to previous studies of other effectors, such NopL and NopT (Liu et al., 2021; Ni et al., 2022). Of the 10 identified QTLs, two were found to overlap with the reported symbiotic loci *Ns1-3* (Hwang et al., 2013) and *Nw1-9* (Nicol á s et al., 2006), respectively. These results indicated the accuracy of the QTL mapping and genetic analyses in the DN594 and Charleston RIL populations. The NopAA mutant was derived from the wild strain HH103 and was found to induce fewer NNs and lower NDW, as well as to generate fewer infection sites. Analysis of the expression patterns of several symbiotic marker genes showed that the NopAA mutant could also change the expression pattern of genes, such as PR1. It is possible that the interaction network might change without the involvement of NopAA. These results suggested that the soybean response to the NopAA mutant differed significantly from its response to the wild strain, leading to the observed QTL differences between HH103 and the NopAA mutant.

In the overlapped QTLs, *Glyma.19G074200* was found to be associated with NopAA using gene annotation and qRT-PCR. *Glyma.19G074200* is a root-specific co-expression gene belonging to the uncharacterized protein At3g27210 protein family. In *Arabidopsis*, the At3g27210 protein interacts with AtRKL1 (Tarutani et al., 2004) and participates in the response to pathogen infection. Soybean RNA-seq results confirmed that *Glyma.19G074200* was involved in symbiosis during *Bradyrhizobium japonicum* infection, although the function of the gene was not determined (Libault et al., 2010). In soybean, the function of this gene has not been studied, in this study, overexpression of *GmARP* promoted nodule formation after inoculation with either the wild strain or the NopAA mutant but, after *GmARP* silencing, no significant differences in either NN or NDW between plants inoculated with HH103 or the NopAA mutant were observed. These results suggest that *GmARP* is a positive regulator of nodule formation and that it mediates NopAA signaling in plants, the mechanism of the synergistic regulation of *GmARP* and NopAA in symbiosis remains to be further studied. We demonstrated that NopAA had glycosyl hydrolase activity and may promote rhizobia infection by hydrolyzing the host cell wall, and that host recognition of NopAA activated the expression of symbiotic genes. When HH103 was used for infection, NopAA induced *GmARP* expression to complete the establishment of symbiosis. We have two hypotheses concerning the mechanism of this activated expression: the first is that NopAA induces the expression of *GmARP* by activating its downstream signals while the second is that the products resulting from NopAA hydrolysis of cell wall polysaccharides induce signaling changes in the host cells, which further activate the expression of *GmARP*. However, these hypotheses require further experimental verification.

CONCLUSION

We demonstrated that NopAA is a type III effector with glycosyl hydrolase activity. Investigation of the effects of the NopAA mutant on nodulation showed that the mutant did not promote nodulation. Using QTL mapping and analysis of soybean genes associated with NopAA, we found that the soybean protein GmARP positively regulated nodulation through its association with NopAA. This study provides an essential reference for analyzing the function of NopAA and provides support for efficient utilization of symbiotic nitrogen fixation in soybean agricultural production.

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 authors.

AUTHOR CONTRIBUTIONS

JHW, QC, MY, and DX conceived the study and designed and managed the experiments. CM and SM performed trials and collected data. JXW, HZ, HF, and YW completed statistical analyses of phenotypic data. CL provided plant lines. JXW, QC, MY, DX, and CM all contributed to writing the paper.

FUNDING

This work was supported by the China Postdoctoral Science Foundation (Grant No. 2020M681072), the National Natural Science Foundation of China (Grant No. 32072014, U20A2027).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.889795/full#supplementary-material

Supplementary Figure S1 | Identification of nodulation ability of soybean germplasm resources. **(A)** Boxplot of NNs of soybean varieties inoculated with HH103 or HH103ΩNopAA (28 dpi). **(B)** Boxplot of NDWs of soybean varieties inoculated with HH103 or HH103ΩNopAA (28 dpi). Data are presented as the average of three different biological replicates and 15 plants for each replicate, results were analyzed for significance using the *t*-test. NDW, nodule dry weight; NN, nodule number.

Supplementary Figure S2 | Relative expression levels of *GmARP* in transformed hairy roots. Data are means ± SE.

REFERENCES

Ausmees, N., Kobayashi, H., Deakin, W. J., Marie, C., Krishnan, H. B., Broughton, W. J., et al. (2004). Characterization of NopP, a Type III Secreted Effector of *Rhizobium Sp*. Strain NGR234. *J. Bacteriol.* 186 (14), 4774–4780. doi:10.1128/JB.186.14.4774-4780.2004

Bartsev, A. V., Deakin, W. J., Boukli, N. M., McAlvin, C. B., Stacey, G., Malnoë, P., et al. (2004). NopL, an Effector Protein of *Rhizobium Sp*. NGR234, Thwarts Activation of Plant Defense Reactions. *Plant Physiol.* 134 (2), 871–879. doi:10.1104/pp.103.031740

Brensha, W., Kantartzi, S. K., Meksem, K., Grier, I. V. R. L., Barakat, A., Lightfoot, D. A., et al. (2012). Genetic Analysis of Root and Shoot Traits in the 'Essex'by 'Forrest'recombinant Inbred Line (RIL) Population of Soybean [*Glycine Max* (L.) Merr.]. *Plant Genet. Genomics, Biotechnol.* 1 (1), 1–9. doi:10.1007/s00425-018-2952-4

Chen, J., Qian, H., and Wu, H. (2017). Nitrogen Contamination in Groundwater in an Agricultural Region along the New Silk Road, Northwest China: Distribution and Factors Controlling its Fate. *Environ. Sci. Pollut. Res.* 24 (15), 13154–13167. doi:10.1007/s11356-017-8881-0

Chen, W.-p., and Kuo, T.-t. (1993). A Simple and Rapid Method for the Preparation of Gram-Negative Bacterial Genomic DNA. *Nucl. Acids Res.* 21 (9), 2260. doi:10.1093/nar/21.9.2260

Chen, W. F., Wang, E. T., Ji, Z. J., and Zhang, J. J. (2021). Recent Development and New Insight of Diversification and Symbiosis Specificity of Legume Rhizobia: Mechanism and Application. *J. Appl. Microbiol.* 131 (2), 553–563. doi:10.1111/jam.14960

Dai, W.-J., Zeng, Y., Xie, Z.-P., and Staehelin, C. (2008). Symbiosis-promoting and Deleterious Effects of NopT, a Novel Type 3 Effector of *Rhizobium Sp*. Strain NGR234. *J. Bacteriol.* 190 (14), 5101–5110. doi:10.1128/JB.00306-08

Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A., and Perret, X. (1997). Molecular Basis of Symbiosis between Rhizobium and Legumes. *Nature* 387 (6631), 394–401. doi:10.1038/387394a0

Ge, Y.-Y., Xiang, Q.-W., Wagner, C., Zhang, D., Xie, Z.-P., and Staehelin, C. (2016). The Type 3 Effector NopL of Sinorhizobiumsp. Strain NGR234 Is a Mitogen-Activated Protein Kinase Substrate. *Exbotj* 67 (8), 2483–2494. doi:10.1093/jxb/erw065

Goyal, R. K., Mattoo, A. K., and Schmidt, M. A. (2021). Rhizobial-Host Interactions and Symbiotic Nitrogen Fixation in Legume Crops toward Agriculture Sustainability. *Front. Microbiol.* 12, 669404. doi:10.3389/fmicb.2021.669404

Grishutin, S., Gusakov, A., Markov, A., Ustinov, B., Semenova, M., and Sinitsyn, A. (2004). Specific Xyloglucanases as a New Class of Polysaccharide-Degrading Enzymes. *Biochimica Biophysica Acta (BBA) - General Subj.* 1674 (3), 268–281. doi:10.1016/j.bbagen.2004.07.001

Guo, Y., and Wang, J. (2021). Spatiotemporal Changes of Chemical Fertilizer Application and its Environmental Risks in China from 2000 to 2019. *Ijerph* 18 (22), 11911. doi:10.3390/ijerph182211911

Hwang, S., Ray, J. D., Cregan, P. B., King, C. A., Davies, M. K., and Purcell, L. C. (2013). Genetics and Mapping of Quantitative Traits for Nodule Number, Weight, and Size in Soybean (*Glycine Max* L.[Merr.]). *Euphytica* 195 (3), 419–434. doi:10.1007/s10681-013-1005-0

Jiménez-Guerrero, I., Pérez-Montaño, F., Medina, C., Ollero, F. J., and López-Baena, F. J. (2015). NopC Is a Rhizobium-specific Type 3 Secretion System Effector Secreted by *Sinorhizobium* (*Ensifer) Fredii* HH103. *PLoS One* 10 (11), e0142866. doi:10.1371/journal.pone.0142866

Jordan, L. K., Mandelman, J. W., and Kajiura, S. M. (2011). Behavioral Responses to Weak Electric Fields and a Lanthanide Metal in Two Shark Species. *J. Exp. Mar. Biol. Ecol.* 409 (1), 345–350. doi:10.1016/j.jembe.2011.09.016

Kambara, K., Ardissone, S., Kobayashi, H., Saad, M. M., Schumpp, O., Broughton, W. J., et al. (2009). Rhizobia Utilize Pathogen-like Effector Proteins during Symbiosis. *Mol. Microbiol.* 71 (1), 92–106. doi:10.1111/j.1365-2958.2008.06507.x

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C. D., Nontachaiyapoom, S., Kinkema, M., et al. (2007). Agrobacterium Rhizogenes-Mediated Transformation of Soybean to Study Root Biology. *Nat. Protoc.* 2 (4), 948–952. doi:10.1038/nprot.2007.141

Krehenbrink, M., and Downie, J. A. (2008). Identification of Protein Secretion Systems and Novel Secreted Proteins in Rhizobium Leguminosarum Bv. Viciae. *BMC Genomics* 9 (1), 55–19. doi:10.1186/1471-2164-9-55

Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.-L., et al. (2010). Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns

of Genetic Diversity and Selection. *Nat. Genet.* 42 (12), 1053–1059. doi:10.1038/ng.715

Li, R., Jiang, H., Zhang, Z., Zhao, Y., Xie, J., Wang, Q., et al. (2019). Combined Linkage Mapping and BSA to Identify QTL and Candidate Genes for Plant Height and the Number of Nodes on the Main Stem in Soybean. *Ijms* 21 (1), 42. doi:10.3390/ijms21010042

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R. J., Bilgin, D. D., et al. (2010). Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and its Alteration in Response to Bradyrhizobium Japonicum Infection. *Plant physiol.* 152 (2), 541–552. doi:10.1104/pp.109.148379

Liu, J., Liu, M. X., Qiu, L. P., and Xie, F. (2020). SPIKE1 Activates the GTPase ROP6 to Guide the Polarized Growth of Infection Threads in Lotus Japonicus. *Plant Cell* 32 (12), 3774–3791. doi:10.1105/tpc.20.00109

Liu, Y., Sun, M., Chen, Q., Xin, D., and Sun, X. (2021). Mapping Quantitative Trait Loci Related to Nodule Number in Soybean (*Glycine Max* (L.) Merr.) in Response to the *Sinorhizobium* (*Ensifer) Fredii* HH103 NopT Type III Effector. *J. Plant Interact.* 16 (1), 126–135. doi:10.1080/17429145.2021.1908635

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., et al. (2017). Natural Variation at the Soybean *J* Locus Improves Adaptation to the Tropics and Enhances Yield. *Nat. Genet.* 49 (5), 773–779. doi:10.1038/ng.3819

Ma, Z., Song, T., Zhu, L., Ye, W., Wang, Y., Shao, Y., et al. (2015). A *Phytophthora Sojae* Glycoside Hydrolase 12 Protein Is a Major Virulence Factor during Soybean Infection and Is Recognized as a PAMP. *Plant Cell* 27 (7), 2057–2072. doi:10.1105/tpc.15.00390

Ma, Z., Zhu, L., Song, T., Wang, Y., Zhang, Q., Xia, Y., et al. (2017). A Paralogous Decoy Protects *Phytophthora Sojae* Apoplastic Effector PsXEG1 from a Host Inhibitor. *Science* 355 (6326), 710–714. doi:10.1126/science.aai7919

Mateos, P. F., Jimenez-Zurdo, J. I., Chen, J., Squartini, A. S., Haack, S. K., Martinez-Molina, E., et al. (1992). Cell-associated Pectinolytic and Cellulolytic Enzymes in *Rhizobium Leguminosarum* Biovar Trifolii. *Appl. Environ. Microbiol.* 58 (6), 1816–1822. doi:10.1128/aem.58.6.1816-1822.1992

Miwa, H., and Okazaki, S. (2017). How Effectors Promote Beneficial Interactions. *Curr. Opin. plant Biol.* 38, 148–154. doi:10.1016/j.pbi.2017.05.011

Ni, H., Peng, Y., Wang, J., Wang, J., Yuan, Y., Fu, T., et al. (2022). Mapping of Quantitative Trait Loci Underlying Nodule Traits in Soybean (Glycine Max (L.) Merr.) and Identification of Genes Whose Expression Is Affected by the Sinorhizobium Fredii HH103 Effector Proteins NopL and NopT. *Agronomy* 12 (4), 946. doi:10.3390/agronomy12040946

Nicolás, M. F., Hungria, M., and Arias, C. A. A. (2006). Identification of Quantitative Trait Loci Controlling Nodulation and Shoot Mass in Progenies from Two Brazilian Soybean Cultivars. *Field Crops Res.* 95 (2), 355–366. doi:10.1016/j.fcr.2005.04.012

Plett, K. L., Bithell, S. L., Dando, A., and Plett, J. M. (2021). Chickpea Shows Genotype-specific Nodulation Responses across Soil Nitrogen Environment and Root Disease Resistance Categories. *BMC Plant Biol.* 21 (1), 310. doi:10.1186/s12870-021-03102-6

Pukatzki, S., McAuley, S. B., and Miyata, S. T. (2009). The Type VI Secretion System: Translocation of Effectors and Effector-Domains. *Curr. Opin. Microbiol.* 12 (1), 11–17. doi:10.1016/j.mib.2008.11.010

Qi, X., Li, M.-W., Xie, M., Liu, X., Ni, M., Shao, G., et al. (2014). Identification of a Novel Salt Tolerance Gene in Wild Soybean by Whole-Genome Sequencing. *Nat. Commun.* 5 (1), 1–11. doi:10.1038/ncomms5340

Qi, Z., Huang, L., Zhu, R., Xin, D., Liu, C., Han, X., et al. (2014). A High-Density Genetic Map for Soybean Based on Specific Length Amplified Fragment Sequencing. *PloS one* 9 (8), e104871. doi:10.1371/journal.pone.0104871

Ratu, S. T. N., Hirata, A., Kalaw, C. O., Yasuda, M., Tabuchi, M., and Okazaki, S. (2021). Multiple Domains in the Rhizobial Type III Effector Bel2-5 Determine Symbiotic Efficiency with Soybean. *Front. Plant Sci.* 12, 1046. doi:10.3389/fpls.2021.689064

Rodrigues, J. A., López-Baena, F. J., Ollero, F. J., Vinardell, J. M., Espuny, M. d. R., Bellogín, R. A., et al. (2007). NopM and NopD Are Rhizobial Nodulation Outer Proteins: Identification Using LC-MALDI and LC-ESI with a Monolithic Capillary Column. *J. Proteome Res.* 6 (3), 1029–1037. doi:10.1021/pr060519f

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., et al. (2020). Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *Plant Cell* 32 (1), 15–41. doi:10.1105/tpc.19.00279

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome Sequence of the Palaeopolyploid Soybean. *nature* 463 (7278), 178–183. doi:10.1038/nature08670

Skorpil, P., Saad, M. M., Boukli, N. M., Kobayashi, H., Ares-Orpel, F., Broughton, W. J., et al. (2005). NopP, a Phosphorylated Effector ofRhizobiumsp. Strain NGR234, Is a Major Determinant of Nodulation of the Tropical legumesFlemingia congestaandTephrosia Vogelii. *Mol. Microbiol.* 57 (5), 1304–1317. doi:10.1111/j.1365-2958.2005.04768.x

Tampakaki, A. P. (2014). Commonalities and Differences of T3SSs in Rhizobia and Plant Pathogenic Bacteria. *Front. Plant Sci.* 5, 114. doi:10.3389/fpls.2014.00114

Tarutani, Y., Sasaki, A., Yasuda, M., Nakashita, H., Yoshida, S., Yamaguchi, I., et al. (2004). Identification of Three Clones Which Commonly Interact with the Kinase Domains of Highly Homologous Two Receptor-like Kinases, RLK902 and RKL1. *Biosci. Biotechnol. Biochem.* 68 (12), 2581–2587. doi:10.1271/bbb.68.2581

Teulet, A., Camuel, A., Perret, X., and Giraud, E. (2022). The Versatile Roles of Type III Secretion Systems in Rhizobia-Legume Symbioses. *Annu. Rev. Microbiol.* 76. doi:10.1146/annurev-micro-041020-032624

Tsyganova, A. V., Brewin, N. J., and Tsyganov, V. E. (2021). Structure and Development of the Legume-Rhizobial Symbiotic Interface in Infection Threads. *Cells* 10 (5), 1050. doi:10.3390/cells10051050

van Ulsen, P., Rahman, S. u., Jong, W. S. P., Daleke-Schermerhorn, M. H., and Luirink, J. (2014). Type V Secretion: from Biogenesis to Biotechnology. *Biochimica Biophysica Acta (BBA) - Mol. Cell Res.* 1843 (8), 1592–1611. doi:10.1016/j.bbamcr.2013.11.006

Wang, J., Wang, J., Liu, C., Ma, C., Li, C., Zhang, Y., et al. (2018). Identification of Soybean Genes Whose Expression Is Affected by the *Ensifer Fredii* HH103 Effector Protein NopP. *Ijms* 19 (11), 3438. doi:10.3390/ijms19113438

Wang, J., Wang, J., Ma, C., Zhou, Z., Yang, D., Zheng, J., et al. (2020). QTL Mapping and Data Mining to Identify Genes Associated with the *Sinorhizobium Fredii* HH103 T3SS Effector NopD in Soybean. *Front. Plant Sci.* 11, 453. doi:10.3389/fpls.2020.00453

Wang, J., Wang, J., Ma, C., Zhou, Z., Yang, D., Zheng, J., et al. (2020). QTL Mapping and Data Mining to Identify Genes Associated with the *Sinorhizobium Fredii* HH103 T3SS Effector NopD in Soybean. *Front. Plant Sci.* 11, 453. doi:10.3389/fpls.2020.00453

Wang, J., Wang, J., Tian, B., Li, Q., Zhu, J., Liu, X., et al. (2019). QTL Mapping and Gene Mining to Identify Genes on Soybean (Glycine Max) Associated with NopB of Sinorhizobium Fredii HH103. *Plant Breed.* 138 (6), 677–685. doi:10.1111/pbr.12714

Wang, Q., Han, C., Ferreira, A. O., Yu, X., Ye, W., Tripathy, S., et al. (2011). Transcriptional Programming and Functional Interactions within the Phytophthora Sojae RXLR Effector Repertoire. *Plant Cell* 23 (6), 2064–2086. doi:10.1105/tpc.111.086082

Xia, Y., Ma, Z., Qiu, M., Guo, B., Zhang, Q., Jiang, H., et al. (2020). N -Glycosylation Shields Phytophthora Sojae Apoplastic Effector PsXEG1 from a Specific Host Aspartic Protease. *Proc. Natl. Acad. Sci. U.S.A.* 117 (44), 27685–27693. doi:10.1073/pnas.2012149117

Xiang, Q.-W., Bai, J., Cai, J., Huang, Q.-Y., Wang, Y., Liang, Y., et al. (2020). NopD of *Bradyrhizobium Sp*. XS1150 Possesses SUMO Protease Activity. *Front. Microbiol.* 11, 386. doi:10.3389/fmicb.2020.00386

Xin, D.-W., Liao, S., Xie, Z.-P., Hann, D. R., Steinle, L., Boller, T., et al. (2012). Functional Analysis of NopM, a Novel E3 Ubiquitin Ligase (NEL) Domain Effector of *Rhizobium Sp*. Strain NGR234. *PLoS Pathog.* 8 (5), e1002707. doi:10.1371/journal.ppat.1002707

Xin, D., Qi, Z., Jiang, H., Hu, Z., Zhu, R., Hu, J., et al. (2016). QTL Location and Epistatic Effect Analysis of 100-seed Weight Using Wild Soybean (*Glycine Soja* Sieb. & Zucc.) Chromosome Segment Substitution Lines. *PloS one* 11 (3), e0149380. doi:10.1371/journal.pone.0149380

Yang, S., Tang, F., Gao, M., Krishnan, H. B., and Zhu, H. (2010). R Gene-Controlled Host Specificity in the Legume-Rhizobia Symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 107 (43), 18735–18740. doi:10.1073/pnas.1011957107

Zboralski, A., Biessy, A., and Filion, M. (2022). Bridging the Gap: Type III Secretion Systems in Plant-Beneficial Bacteria. *Microorganisms* 10 (1), 187. doi:10.3390/microorganisms10010187

Zhang, B., Wang, M., Sun, Y., Zhao, P., Liu, C., Qing, K., et al. (2021). *Glycine Max NNL1* Restricts Symbiotic Compatibility with Widely Distributed Bradyrhizobia via Root Hair Infection. *Nat. Plants* 7 (1), 73–86. doi:10.1038/s41477-020-00832-7

Zhang, L., Chen, X.-J., Lu, H.-B., Xie, Z.-P., and Staehelin, C. (2011). Functional Analysis of the Type 3 Effector Nodulation Outer Protein L (NopL) from Rhizobium Sp. NGR234. *J. Biol. Chem.* 286 (37), 32178–32187. doi:10.1074/jbc.M111.265942

Zhang, L., Chen, X.-J., Lu, H.-B., Xie, Z.-P., and Staehelin, C. (2011). Functional Analysis of the Type 3 Effector Nodulation Outer Protein L (NopL) from *Rhizobium Sp*. NGR234: Symbiotic Effects, Phosphorylation, and Interference with Mitogen-Activated Protein Kinase Signaling. *J. Biol. Chem.* 286 (37), 32178–32187. doi:10.1074/jbc.M111.265942

Zhang, L., Yan, J., Fu, Z., Shi, W., Ninkuu, V., Li, G., et al. (2021). FoEG1, a Secreted Glycoside Hydrolase Family 12 Protein from Fusarium Oxysporum , Triggers Cell Death and Modulates Plant Immunity. *Mol. Plant Pathol.* 22 (5), 522–538. doi:10.1111/mpp.13041

Zhang, Y., Liu, X., Chen, L., Fu, Y., Li, C., Qi, Z., et al. (2018). Mining for Genes Encoding Proteins Associated with NopL of *Sinorhizobium Fredii* HH103 Using Quantitative Trait Loci in Soybean (*Glycine Max* Merr.) Recombinant Inbred Lines. *Plant Soil* 431 (1), 245–255. doi:10.1007/s11104-018-3745-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.

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 © 2022 Wang, Ma, Ma, Zheng, Feng, Wang, Wang, Liu, Xin, Chen and Yang. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 31 May 2022
doi: 10.3389/fgene.2022.853114

Genome-Wide Association Study and Genomic Prediction for Bacterial Wilt Resistance in Common Bean (*Phaseolus vulgaris*) Core Collection

Bazgha Zia[1], Ainong Shi[1], Dotun Olaoye[1], Haizheng Xiong[1], Waltram Ravelombola[2], Paul Gepts[3], Howard F. Schwartz[4], Mark A. Brick[5], Kristen Otto[4], Barry Ogg[5] and Senyu Chen[6]**

[1]Department of Horticulture, University of Arkansas, Fayetteville, AR, United States, [2]Organic & Specialty Crop Breeding, Texas A&M AgriLife Research, Vernon, TX, United States, [3]Department of Plant Sciences/MS1, University of California, Davis, Davis, CA, United States, [4]Department of Agricultural Biology, Colorado State University, Fort Collins, CO, United States, [5]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, United States, [6]Department of Plant Pathology, University of Minnesota, Minneapolis, MN, United States

OPEN ACCESS

Edited by:
Madhav P. Nepal, South Dakota State University, United States

Reviewed by:
Satinder Kaur, Punjab Agricultural University, India
Tika Adhikari, North Carolina State University, United States

***Correspondence:**
Ainong Shi
ashi@uark.edu
Senyu Chen
chenx099@umn.edu

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

Received: *12 January 2022*
Accepted: *14 April 2022*
Published: *31 May 2022*

Citation:
Zia B, Shi A, Olaoye D, Xiong H, Ravelombola W, Gepts P, Schwartz HF, Brick MA, Otto K, Ogg B and Chen S (2022) Genome-Wide Association Study and Genomic Prediction for Bacterial Wilt Resistance in Common Bean (Phaseolus vulgaris) Core Collection. Front. Genet. 13:853114. doi: 10.3389/fgene.2022.853114

Common bean (*Phaseolus vulgaris*) is one of the major legume crops cultivated worldwide. Bacterial wilt (BW) of common bean (*Curtobacterium flaccumfaciens pv. flaccumfaciens*), being a seed-borne disease, has been a challenge in common bean producing regions. A genome-wide association study (GWAS) was conducted to identify SNP markers associated with BW resistance in the USDA common bean core collection. A total of 168 accessions were evaluated for resistance against three different isolates of BW. Our study identified a total of 14 single nucleotide polymorphism (SNP) markers associated with the resistance to BW isolates 528, 557, and 597 using mixed linear models (MLMs) in BLINK, FarmCPU, GAPIT, and TASSEL 5. These SNPs were located on chromosomes *Phaseolus vulgaris* [Pv]02, Pv04, Pv08, and Pv09 for isolate 528; Pv07, Pv10, and Pv11 for isolate 557; and Pv04, Pv08, and Pv10 for isolate 597. The genomic prediction accuracy was assessed by utilizing seven GP models with 1) all the 4,568 SNPs and 2) the 14 SNP markers. The overall prediction accuracy (PA) ranged from 0.30 to 0.56 for resistance against the three BW isolates. A total of 14 candidate genes were discovered for BW resistance located on chromosomes Pv02, Pv04, Pv07, Pv08, and Pv09. This study revealed vital information for developing genetic resistance against the BW pathogen in common bean. Accordingly, the identified SNP markers and candidate genes can be utilized in common bean molecular breeding programs to develop novel resistant cultivars.

Keywords: common bean, bacterial wilt, genome-wide association study, genomic prediction, single nucleotide polymorphism, *Phaseolus vulgaris*, *Curtobacterium flaccumfaciens* pv. flaccumfaciens

INTRODUCTION

Common bean (*Phaseolus vulgaris L.*) is an important legume crop known for its edible seeds and pods worldwide (Allen 2013). It is an important source of protein for humans and livestock. Among legume crops, common bean is considered an outstanding source of nutrition and value in comparison to lentils (Ganesan and Xu 2017), fava beans (Juncus 1998), and chickpeas (Allen 2013). It is called the perfect food due to its content in protein (Guzmán-Maldonado et al., 2000),

fiber (Hughes and Swanson 1989), and carbohydrates (Celmeli et al., 2018). It is mostly consumed as dry bean and green bean or snap bean in different parts of the world. On average, nearly 1.5 to 1.7 million acres of common bean is produced annually in the United States of America USDA-NASS Dry Beans (2022).

Common bean production has been affected by several seed-borne diseases (Sendi et al., 2020). Bacterial wilt (will be abbreviated as BW) of common bean caused by *Curtobacterium flaccumfaciens* pv. flaccumfaciens (Cff), affects production of common bean in a major way due to its seed-borne nature and is caused by various isolates of Cff (González et al., 2005). The pathogen is known to primarily cause disease in legume crops such as cowpea (*Vigna unguiculata*), common bean (*Phaseolus vulgaris*), mung bean (*Vigna radiata*), pea (*Pisum sativum*), and soybean (*Glycine max*) (Osdaghi et al., 2020). BW was first reported in South Dakota in 1926 (Hedges, 1926) and was later discovered in Mexico (Yerkes and Crispin, 1956), Canada (Hsieh et al., 2003), and in several other parts of the world. Due to its virulent nature and economic impact on legume crops, it is considered a high-risk pathogen that is subjected to quarantine regulations in Europe (CABI and Eppo, 1996).

The disease is transmitted *via* infected seeds (Hsieh et al., 2006). The *Cff* pathogen has five different isolates based on different color variants, that is, orange, yellow, purple, red, and pink (González et al., 2005). The infected seeds specifically turn to the color variant 3: making them yellow, orange, or purple as the infection proceeds. The disease symptoms include chlorotic areas on leaves with necrosis leading to a yellow halo progressing to irreversible plant wilt (Harveson et al., 2011). The leaf wilting is accompanied by hindrance of the normal water movement within the plant vascular system (Huang et al., 2009). The symptoms are worsened to tearing and shredding of leaves under unfavorable weather conditions. Young seedlings and plants are more susceptible to disease and prone to early mortality than mature plants (Hsieh et al., 2006; Osdaghi et al., 2020). The disease occurrence is primarily attributed to seed discoloration as a common symptom of BW in common bean (Hsieh et al., 2006). Mature seeds of infected plants are discolored and show yellow, orange, or purple seed coats (Harveson et al., 2011).

BW causes economic losses due to a substantial decrease in crop yield and marketability of the grain produced due to the visual appearance, size, shape, and color of the infected seeds (Huang et al., 2009). Crop rotation coupled with the use of pathogen-free seeds has been used to control the disease (Harveson et al., 2011). However, a cost-effective and reliable measure for disease management is to explore genetic resources to develop resistant cultivars (Assefa et al., 2019). Limited research has been conducted for BW management in common bean (Assefa et al., 2019). Early studies, based on a segregating population resulting from a cross between a resistant and susceptible genotype, identified the susceptibility to BW to be governed by two complimentary dominant genes. However the inheritance pattern for resistance was not clearly determined (Coyne et al., 1965). A more recent study identified a genotype showing some degree of resistance through inoculation tests, but it required substantial level of backcrossing to be acceptable for open cultivation in farmer's fields (Urrea and Harveson 2014).

More recently, resistant cultivars such as the great northern bean "Resolute," pinto bean "Agrinto," pink bean "Early Rose" (Mündel et al., 2005), and an advanced black bean line L02F132 have been identified in Canada, which are resistant to three isolates of *Cff* (Mündel et al., 2005). The bean breeding program at Alberta has also evaluated the identified resistant lines for resistance against different diseases of common bean (Zienkiewicz., 2016). Limited research has been conducted in the United States, resulting in the development of a tolerant variety, namely, great northern cv. "Emerson" (Coyne et al., 1971) as the first cultivar, tolerant to three isolates of the BW pathogen, which was derived by pedigree selection between a BW resistant genotype and the great northern bean type. However, under hot dry field conditions, the symptoms of BW were again observed at early stages of plant growth for this tolerant cultivar (Coyne et al., 1971). Moreover, the resurgence of BW, specifically in Nebraska, suggests there is a need to conduct comprehensive studies to identify genetic resistance to this pathogen (Huang et al., 2009).

Evaluating the existing bean germplasm for the identification of resistance to BW is vital and a cost-efficient method of disease management. BW resistant bean cultivars can be a useful resource in worldwide common bean breeding programs. The identified new sources of resistance to *Cff* will enable breeders to develop reliably resistant cultivars for the future. The source of genetic resistance identified in common bean commercial cultivars can also be transferred to susceptible, elite cultivars through conventional breeding to enhance sources of resistance.

Molecular breeding in plants has played a vital role for crop improvement by expediting crop breeding through the use of molecular tools (Mammadov et al., 2012; Id, Id, and Mayer 2018). Major genes and alleles have been tagged to facilitate marker-assisted selection (MAS) (Heffner et al., 2009; Assefa et al., 2019; Larkin et al., 2019). Recently, genomic selection (GS) has emerged as a valuable tool for crop improvement through predictive breeding (Visscher et al., 2012; Chung et al., 2017; Keller et al., 2020). GS employs the use of genomic estimated breeding value (GEBV) to select individuals based on their performance and has been successfully employed in the breeding programs for crops such as soybean (Jarquin et al., 2016; Qin et al., 2019; Ravelombola et al., 2019; Wen et al., 2019), maize (Liu et al., 2019; Cui et al., 2020), rice (Spindel et al., 2015; Wang et al., 2020), and wheat (Larkin et al., 2019) targeting disease resistance (Heffner et al., 2009; Vallejo et al., 2017; Carpenter et al., 2018) and other important agronomic traits (Chung et al., 2017). Similarly, GS and GWAS have been deployed to study environmental stresses affecting important agronomic traits in common bean (López-Hernández and Cortés, 2019; Keller et al., 2020; Delfini et al., 2021). However, no published GWAS studies have been reported in common bean that specifically address resistance to BW.

Historically, SNP genetic maps have been constructed in common bean using 6K SNP BeadChips (Santos et al., 2003). The availability of several genome assemblies of common bean (e.g., Schmutz et al., 2014; Vlasova et al., 2016; Rendón-Anaya

et al., 2017) has helped breeders conduct SNP studies for different traits allowing identification of candidate genes for important agronomic traits such as drought tolerance (Villordo-Pineda et al., 2015; Valdisser et al., 2020).

Hence, GWAS and GS serve as valuable tools for genetic improvement of important traits in crop species (Chung et al., 2017). The reduced cost of genotyping and improved methods of statistical analysis have increased the availability of valuable genetic information in large populations for complex traits (Visscher et al., 2012). Accordingly, this study primarily focuses on the evaluation of BW resistance in a publicly available USDA common bean core collection (Kuzay et al., 2020; Shi et al., 2021), using association mapping to identify SNP markers associated with BW resistance and conduct GS with the associated SNPs followed by candidate gene discovery.

MATERIALS AND METHODS

Plant Materials

A subset of 168 accessions from the USDA common bean core collection was used in this study. Around 50% (85 accessions) of the total accessions were collected from Mexico. The remaining 83 accessions were collected from Guatemala (20), Colombia (18), Costa Rica (10), Nicaragua (10), Ecuador (9), El Salvador (5), Honduras (4), and Peru (7) (**Supplementary Table S1**).

Bacterial Wilt Isolates

Three isolates of BW were used to study resistance to the *Cff* pathogen in this study. The yellow (BW_528), orange (BW_557), and purple (BW_597) isolates of the BW pathogen were previously recovered and maintained from infected common beans in Nebraska (Harveson et al., 2011) or Colorado. The purple isolate was obtained from the collection of R. Harveson at the University of Nebraska (Harveson et al., 2011), and the other two isolates were obtained from the collection of H. Schwartz (Schwartz et al., 2009) at Colorado State University.

Phenotyping for Bacterial Wilt Resistance

The 168 common bean accessions were tested for the three isolates of BW. The experiment was conducted using the cotyledonary node inoculation method (Hsieh et al., 2003) by Howard F. Schwartz at Colorado State University and Mark A. Brick, Kristen Otto, and Barry Ogg at Colorado State University (Schwartz et al., 2010). The data set with disease scores is already published and available for public at the USDA GRIN website (https://www.ars-grin.gov/Pages/Collections).

In brief, 7–8 seeds were sown at a depth of 2.5 cm using 15 cm plastic pots with a standard potting mix. The seedlings were thinned to five plants upon emergence. The 7- to 10-day-old seedlings were inoculated with the respective isolate using a sterile needle. The inoculated seedlings were then incubated at a daily temperature of 28°C/22°C for 16 h per day and 8 h per night photoperiod in a greenhouse. A total of 9–12 plants per accession were used for evaluation for each isolate. In addition, ten plants for the resistant and susceptible controls were included for each BW isolate. The symptoms were evaluated 4 weeks post inoculations. Data were recorded as average severity for the replicated plants for each isolate. A 2-month cycle was used to evaluate the germplasm for each isolate individually.

A standard rating scale from 1 to 4 was used to evaluate the plants, with 1 as highly resistant demonstrating no wilt or discoloration, 2 being moderately resistant with wilt or discoloration at one of the unifoliolate leaves, 3 showing wilt or discoloration on both unifoliolate leaves but asymptomatic on the 1st trifoliolate leaf, and 4 as highly susceptible with wilt or discoloration on the 1st trifoliolate leaf (Schwartz et al., 2009).

Genotyping

The common bean core set was genotyped (Kuzay et al., 2020) using BARCBean6K_3 Infinium BeadChips (Song et al., 2013). A total of 4,568 SNPs were obtained from the BARCBean6K_3 Infinium BeadChips (https://datadryad.org/stash/dataset/doi:10.25338/B8KP45) for genotyping. SNP filtering was conducted with removal of SNPs; data missing rate >20%, heterogeneous >10%, and MAF (minor allele frequency) <5%.

Phenotypic Data Analysis and Estimation of Plant Distribution for Bacterial Wilt Isolates

The phenotypic data for the three BW isolates resistance was analyzed using ANOVA and GLM functions in JMP Genomics 7 (Cary 2008). The mean (X), variance (V), standard deviation (SD), and standard error (SE) were estimated using the "Tabulate" function in JMP Genomics 7 followed by the "Distribution" function to graphically present the phenotypic data for each of the BW isolates.

Estimation of Population Structure and Genetic Diversity

The principal component analysis (PCA) and genetic diversity were analyzed using GAPIT 3 (genomic association and prediction integrated tool version 3) by setting PCA = 2 to 10 and NJ tree = 2 to 10, and phylogenetic trees were drawn using the neighbor-joining (NJ) method (Lipka et al., 2012; Wang et al., 2021; https://github.com/jiabowang/GAPIT3).

Association Analysis

The phenotypic and genotypic data obtained for the 168 common bean core collection was subjected to genome-wide association mapping using the mixed linear model (MLM) methods in TASSEL 5 (Bradbury et al., 2007). The compressed mixed linear modeling (cMLM) in GAPIT (Lipka et al., 2012), FarmCPU (Liu et al., 2016), and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) (Huang et al., 2019) were performed using the GAPIT 3 tool (Lipka et al., 2012; Wang et al., 2021; https://zzlab.net/GAPIT/index.html; https://github.com/jiabowang/GAPIT3). A threshold LOD [log10(p)] value >3.0 was used to select significant SNP markers associated with resistance to the BW_528, BW_557, and BW_597 isolates. Squared correlation coefficient (R^2) was used to calculate the linkage disequilibrium (LD) between the markers.

Candidate Gene Prediction

The SNP regions were subjected to candidate gene discovery analysis for the identification of candidate genes spanning the 50 kb (50 kb on each side of SNP) regions around the significant SNPs. The Andean whole-genome reference sequence *Pvulgaris* 442_v2.1 available on the Phytozome website (https://phytozome.jgi.doe.gov/pz/portal.html) was used to retrieve the candidate genes from the reference annotation of the common bean genome.

Genomic Prediction for Genomic Selection of Bacterial Wilt Resistance

In addition to the identification of SNPs associated with BW resistance, the effect of these SNPs markers was also evaluated by using seven genomic prediction (GP) models. The ridge regression (RR); best linear unbiased prediction (BLUP) rrBLUP analysis; Bayesian models: Bayes A, Bayes B, Bayes ridge regression (BRR), and Bayes LASSO (BL); random forest (RF); and support vector machines (SVM) were deployed to assess GP (**Table 3**).

GP were carried out using the unbiased prediction in the rrBLUP (Wang J. et al., 2018) package to predict for GS utilizing the GEBV (Vallejo et al., 2017) with the R software version 3.5.0 (https://www.r-project.org). The Pearson correlation coefficient (r) was used to estimate the prediction accuracy (PA) using the GEBVs and observed values (Waldmann 2019) for the resistance to each of the three BW isolates. In addition, Bayesian models: Bayes A, Bayes B (Barili et al., 2018), Bayes ridge regression (BRR), and Bayes LASSO (BL) (Legarra et al., 2011); random forest (RF) (Ogutu et al., 2011); and support vector machines (SVM) (Maenhout et al., 2007) were deployed to assess the GP. Each combination of GP was run hundred times to estimate the GP statistical parameters, including variance (V), mean (X), standard deviation (SD), standard error (SE), and (r) values. Two approaches were used in combination with the seven prediction models 1) using all the 4,568 SNPs and 2) using the 14 selected SNP markers. The distribution plots were drawn using the R package ggplot2 and Microsoft Excel 2016.

RESULTS

Phenotypic Data Analysis and Plant Distribution for Bacterial Wilt Isolates

The common bean core collection assessed for resistance to the three BW isolates showed a distribution of accessions from BW score 1 to 4 (where 1 is highest resistance and 4 is highest susceptible) skewed to resistance (**Supplementary Table S1**; **Figure 1**) to each of the three tested BW isolates, suggesting that BW resistant common bean accessions existed. Among the 168 common bean accessions inoculated with isolate BW_528, 31 accessions were resistant with a score of 1, while 9 accessions were susceptible with a score of 4. The distribution accessions for resistance to isolate BW_528 had a mean value (x) of 2.29, variance (V) of 1.027, standard deviation (SD) of 1.013, standard error (SE) of 0.078, and a coefficient of variance (CV) of 44.2% (**Supplementary Table S2**).

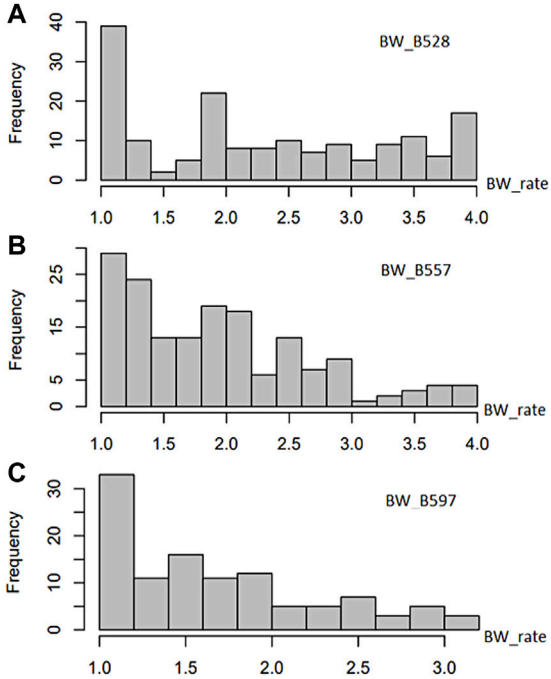


FIGURE 1 | Distribution of bacterial wilt (BW) disease scale (0–4 rate) in 168 USDA common bean germplasm accessions.

The distribution of resistance scores after inoculation of 165 accessions with the BW_557 isolate was skewed toward the left side with score 1 of BW resistance (**Figure 1**). Seven accessions were rated as highly resistant to BW_557 with a score of 1, while three accessions were scored as 4. The ANOVA analysis indicated a mean of 1.94 and an SD of 0.756 (**Supplementary Table S2**). Similarly, the graph for BW_597 was also skewed toward the left side. Among the 111 tested common bean accessions, nine accessions were rated 1 for resistance and one common bean accession imparted the highest susceptibility with a score of 3.17 (**Supplementary Table S1**; **Figure 1**). Overall, the distribution had a mean of 1.70, a median of 1.58, and an SD of 0.6 (**Supplementary Table S2**).

Based on the phenotypic analysis, PI203958 may be a good candidate for resistance to all three BW isolates. PI310611 was resistant to two BW isolates (BW_528 and BW_557) (**Supplementary Table S1**). Similarly, PI207336, PI313429, PI313531, PI325685, and PI451889 were potential candidates for resistance to the BW_528 and BW_597 isolates (**Supplementary Table S1**).

Genetic Diversity and Population Structure

The three subpopulations (Q1, Q2, and Q3) were well-differentiated with red, green, and blue colors (**Figure 2**; **Supplementary Figure S1**) in 168 common bean accessions based on 4,568 high-quality SNPs analyzed by GAPIT 3. From

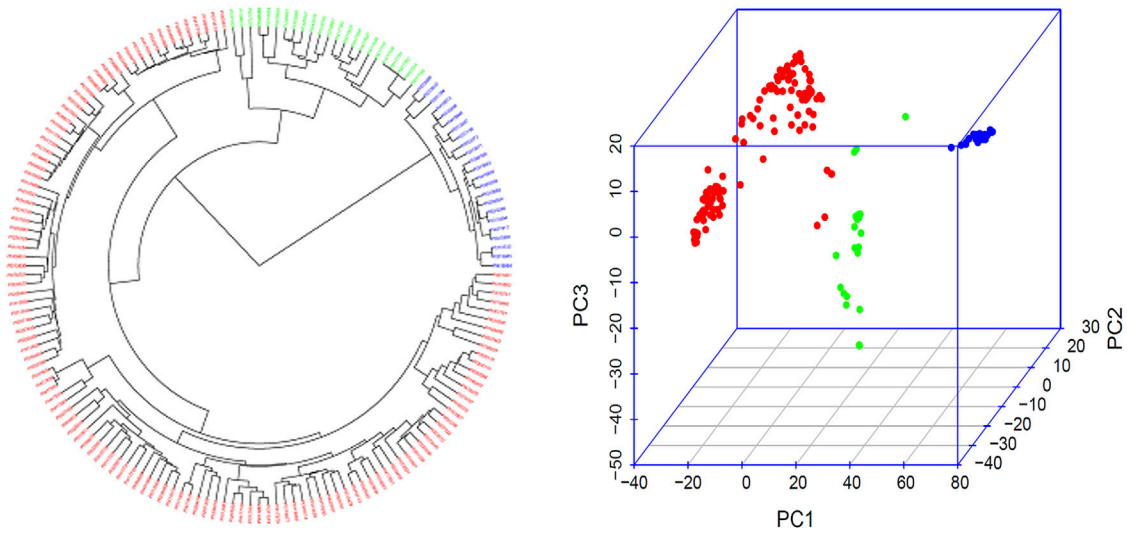


FIGURE 2 | Population genetic diversity analysis in the association panel consisted of 168 USDA common bean germplasm accessions. Phylogenetic trees drawn by using the neighbor-joining (NJ) method in three subpopulation (left) and 3D graphical plot of the principal component analysis (PCA) (right) drawn by using GAPIT 3. A large phylogenetic tree of the three subpopulation for each of the 168 common bean accessions is shown in **Supplementary Table S1**.

a total of 121 genotypes 72% of the total population was accountable for the cluster 1 (Q1); 13% of the total genotypes comprises 22 genotypes that made up the second cluster (Q2), and the remaining 22 genotypes makes up 13% of the total population and the third cluster (Q3) (**Figure 2**; **Supplementary Figure S1**).

Genetic Diversity of Bacterial Wilt Resistant Lines

Our analysis identified a total of 21 R-lines (**Table 1**). The identified R-lines for all the three BW type were assigned to Q1 and Q2 clusters. Among the 21 R lines, 17 R-lines originated from Mexico, two from Colombia, one from Costa Rica, and one from Guatemala (**Table 1**). Here again, PI203958 from the Q1 subpopulation proved to be a good candidate with resistance to all the three BW isolates (**Table 1**). Furthermore, the phylogenetic tree also depicted a similar trend (**Figure 3**).

GWAS and SNP Marker Identification

Collectively, 14 SNPs were associated with resistance to BW_528, BW_557, and BW_597, respectively, based on the four MLM models in TASSEL 5, FarmCPU, GAPIT, and BLINK using the 4,568 SNPs (**Table 2**). The identified SNPs were associated with only single isolate, respectively, and not a SNP marker was simultaneously associated with all three isolates with an LOD value >3.0 for one or more of the four MLM models for resistance to all three isolates (**Table 2**). A total of 4,568 SNPs were used to conduct LD analysis. The LD decay started at around 137 kb (**Supplementary Figure S2**).

The *t*-tests for the 14 SNP markers are listed in **Table 2**, showing their allelic association with the phenotypes in each of the three BW isolates. Except ss715648247, ss715639596, and ss715647928, 11 markers had a LOD value >1.6, showing significant differences between two alleles of the 11 SNPs at a p-value at the 0.05 level (**Table 2**). Highly significant difference with a LOD >4.0 were observed at the three SNPs, ss715647803, ss715640165, and ss715648541, for BW_528 resistance and at the four SNPs, ss715649344, ss715647896, ss715641991, and ss715649486, for BW_597 resistance (**Table 2**), suggesting that the presence of beneficial alleles associated with BW resistance.

GWAS for Bacterial Wilt_528 Isolate Resistance

The GWAS panel for BW_528 was subjected to four MLM analyses in TASSEL 5; a QQ-plot distribution was obtained for the observed vs. expected LOD values. Based on MLM, the distribution of QQ-plot between the observed vs. expected LOD value showed divergence from the expected distribution (**Supplementary Figure S3**). A similar trend was observed for the MLM QQ-plot with GAPIT, FarmCPU, and BLINK (**Supplementary Figure S3**). The QQ-plots obtained from GAPIT, BLINK, and FarmCPU showed the beginning of divergence between the observed vs. expected values starting at LOD >2 (**Supplementary Figure S3**).

These findings indicate the presence of SNPs at LOD scores greater than two to be associated with resistance to the BW_528 isolates (**Supplementary Figure S3**).

The TASSEL analysis showed a Manhattan plot for the MLM model with only one significant SNP (ss715648247 on chromosome Pv (04) had LOD close to 3 (actual value of 2.97) and other four SNPs with LOD >2.0 on Pv02, Pv04, Pv08, and Pv09, respectively, indicating a weak association for resistance to the BW_528 (**Supplementary Figure S3**). On the

TABLE 1 | List of 21 common bean accessions with resistance to three bacterial wilt (BW) isolates, B528, B557, and B597.

Campaign plant ID	Plant name	Country	Cluster	B528	B557	B597
PI207182	G918	Colombia	Q1	1	1.25	1
PI207322	Hidalgo 48-A	Colombia	Q1	1	1.08	1.08
PI207336	Jalisco 31-1	Costa Rica	Q1	1	1.09	1
PI309857	Col. No. 20670, lot #13	Guatemala	Q1	1	1.33	1.08
PI310726	Xucu mama	Mexico	Q1	1.17	1.36	1
PI310778	G2031	Mexico	Q1	1	1.08	1.25
PI311843	Frijol de gato	Mexico	Q1	1.08	1.17	1.25
PI451889	—	Mexico	Q1	1	1.17	1
PI201329	No. 3194	Mexico	Q1	1	1.08	—
PI203958	Negro	Mexico	Q1	1	1	1
PI309701	Frijol rosita	Mexico	Q1	1.08	1.08	1.17
PI310611	Frijol de bara	Mexico	Q1	1	1	1.17
PI312018	Frijol negro bolito	Mexico	Q1	1	1.08	1.17
PI313429	Morado claro	Mexico	Q1	1	1.17	1
PI313501	Parraleno colorado	Mexico	Q1	1	1.25	1.08
PI313512	Amarillo	Mexico	Q1	1.25	1	1.17
PI313531	Apetito	Mexico	Q1	1	1.17	1
PI317350	Frijol de raton	Mexico	Q1	1	1.08	1.09
PI325614	G16396	Mexico	Q1	1	1.17	1.17
PI325685	G12879	Mexico	Q2	1	—	1
PI325687	Frijol del raton	Mexico	Q2	1	1.08	1.18

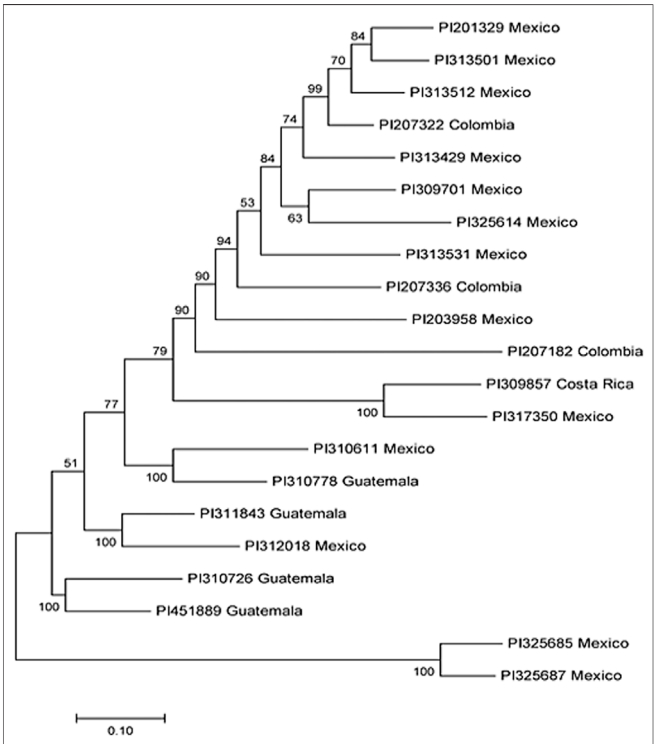


FIGURE 3 | Phylogenetic tree among 21 common bean accessions of bacterial wilt resistance drawn using Mega 7. In the tree, the taxon name consists of the accession ID and the accession original country.

other hand, the cMLM model in GAPIT showed a Manhattan plot with significant SNPs (LOD values >3) for resistance on Pv02 and Pv08 (**Supplementary Figure S3**). The Manhattan plot from BLINK also showed a similar trend with associated SNPs (LOD values >3) located on chromosomes Pv04, Pv08, and Pv09 (**Supplementary Figure S3**). Likewise, the Manhattan plot obtained from FarmCPU also showed similar results as BLINK with the associated SNPs located on chromosomes Pv02, Pv04, Pv08, and Pv09 (**Table 2**; **Supplementary Figure S3**).

The combined results from all the four models of MLM in GAPIT, MLM in FarmCPU, BLINK, and TASSEL showed a total of five SNPs associated with resistance to the BW_528 isolate (**Table 2**). The two SNPs, ss715640165 and ss715648247, were positioned at 12,907,955 and 38,819,373 bp, respectively, on Pv0 4 (**Table 2**), while another SNP, ss715647803, located on Pv02 was positioned at 3,915,879 bp. The SNP markers, ss715639596 and ss715648541, were positioned at 31,079,880 bp on Pv09 and 12,268,429 bp on Pv08, respectively (**Table 2**).

GWAS for Bacterial Wilt_557 Isolate Resistance

Based on MLM in TASSEL 5, the distribution of QQ-plot between the observed vs. expected LOD values showed divergence from the expected distribution. A similar trend was observed for the MLM QQ-plot obtained from the cMLM analysis from GAPIT and the MLM analysis from FarmCPU, and BLINK. The QQ-plots obtained from GAPIT, BLINK, and FarmCPU showed a larger divergence between the observed vs. expected values at LOD >2.5 (**Supplementary Figure S4**). This indicates the presence of SNPs at LOD >2.5 to be associated with resistance to the BW isolate 557 (**Supplementary Figure S4**).

The TASSEL analysis showed the Manhattan plot for the MLM model with SNPs associated with BW_550 resistance, being indicated as dots with LOD value greater than 3 to be located on chromosome Pv11 (**Supplementary Figure S4**). On the other hand, the MLM model resulted in a Manhattan plot with significant SNPs (LOD values >3) for resistance on Pv11 and

TABLE 2 | List of the selected SNP markers associated with resistance to three bacterial wilt (BW) isolates B528, B557, and B598 in common bean core collection obtained from four MLM models in Tassel 5, BLINK, GAPIT, and FarmCPU and a *t*-test.

SNP	Chr	Pos	LOD[-log(P-value)]					R^2%	Beneficial allele (resistant)	Unbeneficial allele (susceptible)	MAF %	BW strain
			Tassel_MLM	Blink	Gapit	FarmCPU	t-test	Tassel_MLM				
ss715647803	2	3915879	2.77	2.44	3.20	2.44	4.72	7.86	G	T	7.2	BW_528
ss715640165	4	12907955	2.17	3.24	2.72	3.24	10.93	6.12	C	T	7.8	
ss715648247	4	38819373	2.97	3.47	2.97	3.47	0.93	6.54	T	C	10.1	
ss715648541	8	12268429	2.74	4.26	3.12	4.26	6.99	7.81	T	G	37.2	
ss715639596	9	31079880	2.47	3.12	2.15	3.12	0.85	5.44	G	A	19.9	
ss715647928	7	11939824	2.37	3.03	2.77	3.03	1.02	5.34	G	A	3.0	BW_557
ss715648425	7	14455236	2.82	3.04	2.82	3.04	3.29	4.93	T	C	4.8	
ss715642582	7	14750979	2.82	3.04	2.82	3.04	3.29	4.93	G	T	4.8	
ss715648754	10	3784843	1.79	3.15	2.95	3.15	1.61	4.49	G	T	26.2	
ss715646271	11	2884160	3.01	3.59	3.18	3.59	2.26	6.88	T	C	3.6	
ss715649344	4	43584074	3.07	2.91	2.99	2.91	6.49	10.27	T	G	46.1	BW_597
ss715647896	8	42837392	3.04	2.84	3.09	2.84	17.05	10.16	G	A	6.5	
ss715641991	8	45046851	3.06	2.58	2.89	2.57	16.96	10.25	A	G	6.6	
ss715649486	10	8067409	2.39	3.16	3.04	3.15	4.07	10.12	T	C	36.4	

Pv10 (**Supplementary Figure S4**). The Manhattan plot from BLINK also showed a similar trend with associated SNPs (LOD value >3) on Pv07, Pv10, and Pv11 (**Supplementary Figure S4**). The Manhattan plot obtained from FarmCPU showed similar results with the associated SNPs on Pv7, Pv10, and Pv11. However, no SNP was found to be associated with a LOD higher than 5.5 (**Supplementary Figure S4**).

The combined results from all the four models of MLM in GAPIT, FarmCPU, BLINK, and TASSEL showed a total of five SNPs associated with resistance to the BW_557 isolate. Three SNPs, ss715647928, ss715648425, and ss715642582, were located on chromosome Pv07 with the latter two located closely together on positions 14,455,236 and 14,750,979 bp, respectively (**Table 2**). The other two SNPs were located at position 3,784,843 bp on Pv10 and position 2,884,160 on Pv11, respectively (**Table 2**).

GWAS for Bacterial Wilt_597 Isolate Resistance

The MLM QQ-plot distribution between the observed vs. expected LOD, obtained using the MLM in TASSEL 5, showed divergence from the expected distribution. A similar trend was observed for the MLM QQ-plot obtained from the cMLM analysis in GAPIT, and the MLM analysis in FarmCPU and BLINK (**Supplementary Figure S5**). The QQ-plots obtained from GAPIT, BLINK, and FarmCPU showed the beginning of divergence between the observed vs. expected values at LOD >2 (**Supplementary Figure S5**). This indicates the presence of SNPs at LOD >2 associated with resistance to the BW_597 isolate (**Supplementary Figure S5**).

The TASSEL analysis showed the Manhattan plot for the MLM model with only one SNP associated with resistance to the BW_597 isolate (LOD >3) located on chromosomes Pv04 and Pv08 (**Supplementary Figure S5**). The MLM model showed the Manhattan plot indicating associated SNPs (LOD values >3) for resistance on Pv04, Pv08, and Pv10 (**Supplementary Figure S5**). The Manhattan plot obtained from BLINK showed associated SNPs with a LOD >3 located on Pv02 and Pv10 (**Supplementary Figure S5**). The Manhattan plot obtained from FarmCPU showed associated SNPs located on Pv04, Pv08, and Pv10 (**Supplementary Figure S5**).

The combined results from all the four models of MLM in GAPIT, FarmCPU, BLINK, and TASSEL showed a total of four SNPs associated with resistance in common bean for the BW_597 isolate (**Table 2**). The two SNPs, ss715647896 and ss715641991, were closely positioned at 42,837,392 and 45,046,851 on chromosome Pv08, respectively (**Table 2**), while other SNPs, ss715649344 and ss715649486, were located at position 43,584,074 on Pv04 and position 8,067,409 on Pv10, respectively (**Table 2**).

Candidate Genes for Bacterial Wilt Resistance

The candidate gene discovery was carried out for 50 kb genomic regions upstream and downstream of the identified significant SNPs for each isolate. A total of 14 gene models were discovered 50 kb upstream and downstream of the identified SNP region on chromosomes Pv02, Pv04, Pv07, Pv08, and Pv09 (**Supplementary Table S3**). A total of six genes (*Phvul.002G041600*, *Phvul.004G076900*, *Phvul.004G119800*, *Phvul.008G107000*, *Phvul.009G210400*, and *Phvul.002G041000*) were identified as candidate genes for resistance to the BW_528 isolate. Out of these six genes, five genes (*Phvul.002G041600*, *Phvul.004G076900*, *Phvul.004G119800*, *Phvul.008G107000* and *Phvul.009G210400*) were located within the 50 kb region of the associated SNP (ss715647803, ss715640165, ss715648247, ss715648541, and ss715639596) region, while one gene model (*Phvul.002G041000*) included the SNP itself (**Supplementary Table S3**). These genes encoded the NAC domain–containing protein 87, XB3 ortholog 3, and duplicated homeodomain-like superfamily protein in *Arabidopsis thaliana*, while other

TABLE 3 | Genomic prediction of seven models for resistance to three bacterial wilt (BW) isolates in two SNP sets: 1) all 4,568 SNPs and 2) 14 SNP markers.

GP model	BW_528		BW_557		BW_597		Average (model)	SNP set
	rŸ100	SE	rŸ100	SE	rŸ100	SE		
rrBLUP	0.56	0.012	0.45	0.012	0.45	0.016	0.49	all 4568 SNPs
BL	0.58	0.010	0.46	0.013	0.43	0.017	0.49	
BA	0.55	0.013	0.46	0.013	0.47	0.016	0.49	
BB	0.55	0.010	0.44	0.013	0.45	0.016	0.48	
BRR	0.56	0.011	0.45	0.013	0.43	0.016	0.48	
SVM	0.54	0.013	0.47	0.013	0.44	0.018	0.48	
RF	0.51	0.012	0.37	0.015	0.41	0.016	0.43	
Average	0.55	—	0.44	—	0.44	—	0.48	
rrBLUP	0.40	0.012	0.30	0.016	0.43	0.015	0.38	14 SNP markers
BL	0.51	0.011	0.41	0.012	0.51	0.015	0.48	
BA	0.53	0.011	0.41	0.009	0.52	0.013	0.49	
BB	0.53	0.010	0.39	0.013	0.52	0.015	0.48	
BRR	0.52	0.011	0.44	0.010	0.51	0.015	0.49	
SVM	0.49	0.013	0.39	0.012	0.44	0.015	0.44	
RF	0.48	0.012	0.43	0.011	0.49	0.016	0.47	
Average	0.49	—	0.40	—	0.49	—	0.46	

identified gene models were unclassified (**Supplementary Table S3**).

Similarly, a total of four gene models, *Phvul.007G104800*, *Phvul.007G112100*, *Phvul.007G112900*, and *Phvul.007G104500*, were identified as candidate genes for resistance to the BW_557 isolate. Among these four gene models, three (*Phvul.007G104800*, *Phvul.007G112100*, and *Phvul.007G112900*) were located within the 50 kb distance of the identified SNPs (ss715647928, ss715648425, and ss715642582, respectively), while one gene model, *Phvul.007G104500* include the identified SNP (ss715647928). The *Phvul.007G112100* encoded a disease-associated Leucine-rich repeat protein kinase family protein positioned at 14,476,334 to 14,479,522 bp on chromosome Pv07 (**Supplementary Table S3**), while the other two gene models, *Phvul.007G104800* and *Phvul.007G104500*, encoded a cytochrome P450, family 77, subfamily A, polypeptide 4, and chlorophyllase, respectively.

A total of four gene models, *Phvul.004G154500*, *Phvul.008G166000*, *Phvul.004G154100*, and *Phvul.008G172000,* were presumed candidate genes for BW_597 resistance (**Supplementary Table S3**). These genes encoded a protein kinase superfamily protein, flavonol synthase 1, nuclease-related domain (NERD), and thioesterase superfamily protein, respectively. *Phvul.004G154500* and *Phvul.004G154100* were located on chromosome Pv04 and included the ss715649344 SNP. The other two genes (*Phvul.008G166000* and *Phvul.008G172000*) were located on chromosome Pv08, within the 50 kb region of ss715647896 and ss71564199 SNPs, respectively (**Supplementary Table S3**).

Genomic Prediction for Resistance to Bacterial Wilt Isolates

The use of all the seven models with aforementioned two approaches predicted the overall GA between the observed values and GEBV for the BW_528 isolate to fluctuate between 0.51 and 0.58 when 1) all the 4,568 SNPs were used and between 0.40 and 0.53 when 2) using the 14 selected SNPs. Similarly, the seven models resulted in a range of 0.37–0.46 when 1) all 4,568 SNPs were used for BW_557 in comparison to the reduced range of 0.30–0.44 when 2) using the selected 14 SNPs. A slightly higher range of average "r" value from 0.41 to 0.47 and 0.43 to 0.52 were observed for BW_597 when 1) all the 4,568 SNPs and 2) the 14 selected SNPs were used, respectively, in combination with the seven GP models (**Table 3**; **Supplementary Figures S6, S7**). The results were also verified to be similar through cross-validation across the seven GP models (**Supplementary Figures S6, S7**).

The general trend of PA was higher when a greater number of SNPs (4,568 SNPs) was utilized in combination with the seven models in comparison to the use of 14 selected SNPs for BW_528 and BW_557. However, observing the PA for individual models, the RF model indicated slightly higher PA when 14 SNPs set was used for BW_557 resistance. Conversely, the PA followed a general trend of higher range for the average values of (r) and for each of the individual models when using a lower number of SNPs (14 selected SNPs) (**Table 3**).

DISCUSSION

Genetic Diversity and Population Structure for the Common Bean Germplasm

The population structure and genetic diversity analyses in this study indicated the presence of three subpopulations (Q1, Q2, and Q3) among the tested germplasm as examined by the tool GAPIT 3 (**Figure 2**; **Supplementary Table S1**). Historically, Andean and Mesoamerican pools are reported as two centers for common bean origin (Bitocchi et al., 2013; Gepts et al., 1986; Kwak and Gepts 2009; Mamidi et al., 2013). Our study also confirmed the existence of two gene pools by consistent appearance of accessions from Mexico, Guatemala, Nicaragua, El Salvador, Costa Rica, Honduras, Colombia, Ecuador, and Peru in our two subpopulation clusters, Q1 and G2

(**Supplementary Table S1**, **Figure 2**). Hence, we can conclude that our tested germplasm is composed of diverse accessions and belong to the original two gene pools.

Genome Wide Association Study and SNP Marker Identification for Bacterial Wilt Resistance

The current study was focused on identifying SNP markers associated with resistance to the three isolates of BW in common beans. The phenotypic and genotypic data from the 168 accessions of the common bean core collection was subjected to the four MLM models in GAPIT, BLINK, FarmCPU, and TASSEL 5 to carry out GWAS analysis for each of the BW isolates. A total of 14 SNP markers were associated with resistance to the three different BW isolates, including five SNP markers for BW isolate 528, five SNP markers for BW isolate 557, and four SNP markers for BW isolate 597 (**Table 2**). These SNP markers were scattered on chromosomes Pv02, Pv04, Pv07, Pv08, Pv09, Pv10, and Pv11.

Genetic studies in BW studies have been primarily limited to other crops, such as extensive use of SSR (simple sequence repeat), AFLP (amplified fragment length polymorphism), and SCAR (sequence characterized amplified region) markers to map QTLs for BW resistance (Thoquet et al., 1996; Ashrafi et al., 2009; Wang et al., 2013) and generate a high-density genetic maps of inbred lines for BW resistance in tomato. The major QTLs were located on chromosomes 6 and 12 (Shin et al., 2020). Similarly, BW resistance has been explored using GWAS in peanut (*Arachis hypogaea L.*) with identification of four QTLs on chromosome 4 (Wang L. et al., 2018).

However, in common bean, studies have been limited to the use of genetic analysis for other diseases such as SCN. Jain et al. (2019) identified SCN resistance factors in common bean on chromosomes Pv04, Pv07, Pv09, and Pv11 based on the *Pvulgaris* v1.0_218 reference genome sequence (from Andean accession G19833) for various races (Jain et al., 2019). Likewise, our study also revealed that the resistance for BW for the three isolates is scattered on multiple chromosomes with the identification of SNPs on chromosomes Pv02, Pv04, Pv07, Pv08, Pv09, Pv010, and Pv011 (**Table 2**; **Supplementary Figures S3–S5**). Thus, our study is the first to report specifically the identification of SNPs associated with resistance to the BW isolates in common bean.

Similarly, a recent study was conducted to phenotype 467 accessions consisting of the NPGS core collection, 8 local, and 31 experimental lines from University of Nebraska for the orange BW isolate (Urrea and Harveson 2014). The results led to the identification of only one cultivar resistant to the tested BW isolate (Urrea and Harveson 2014). Likewise, our study has successfully led to the identification of potential SNPs for resistance to the BW isolates were identified, which can now be employed in marker-assisted selection to develop resistant cultivars. The identified SNPs for each BW isolate can be pyramided to develop a single cultivar with enhanced resistance to multiple isolates of BW.

Candidate Genes

Our results indicate the presence of a putative *chlorophyllase* encoded as the gene model *Phvul.007G104500* in the SNP ss715647928 region (**Supplementary Table S3**). The ortholog of the *chlorophyllase* gene in Arabidopsis encoded by *AtCLH1* is found to be induced following tissue damage by a bacterial necrotrophic pathogen (Kariola et al., 2005). The downregulation of *AtCLH1* is linked to enhanced susceptibility to the necrotrophic pathogen, which showed its role as modulator of defense to various pathogens (Kariola et al., 2005). Our findings also suggest that the identified *chlorophyllase* gene can be a good candidate for resistance to BW_528. However, the *Phvul.007G104800* gene model encoding the cytochrome P450, family 77, subfamily A, and polypeptide 4 protein was found to be located 50 kb upstream and downstream of SNP ss715647928 associated with resistance to the BW_557 isolate (**Supplementary Table S3**). The two genes near the ss715647928 SNP are suitable candidates for BW_528 resistance.

Similarly, the upstream and downstream regions of SNPs ss715648425 and ss715642582 also comprised gene models *Phvul.007G112100* putatively encoding the leucine-rich repeat protein kinase family protein and an unclassified gene *Phvul.007G112900*, respectively (**Supplementary Table S3**).

Interestingly, the LRR domains have been explored as vital modulators of immunity in plant–pathogen interaction responses (Marone et al., 2013; Song et al., 2019). Song et al. (2019) reported a total of 348 NBS-LRR proteins and studied the loss of function characteristic of an LRR domain resulting in increased susceptibility to the BW pathogen in peanut (*Arachis hypogea*) (Song et al., 2019). Thus, the identified gene *Phvul.007G112100* for BW_557 isolate in our study can also be explored further as a source of resistance to the *Cff* pathogen in common bean.

Moreover, our study identified two additional genes, *Phvul.002G041000* and *Phvul.004G154100*, on chromosome Pv02 for BW_528 resistance associated with SNPs ss715647803 and ss715649344. The *Phvul.002G041000* gene encodes a nuclease-related domain (NERD), associated with BW resistance to BW isolate 597 in common bean. Other reported genes (**Supplementary Table S3**) include the Arabidopsis NAC domain–containing protein 87, XB3 ortholog 3 in Arabidopsis, and duplicated homeodomain-like superfamily protein near the SNP regions for BW_528. The NAC genes play a vital role in plant immune responses by acting as regulators modulating the hypersensitive response and receptors of pathogen effectors in host plants. The identified SNP with the candidate gene *Phvul.004G076900* encoding the Arabidopsis *NAC* gene can be studied further to develop a deeper understanding of the respective gene as a potential modulator of immunity for the BW pathogen in common bean.

However, the genes associated with resistance to BW_597 encoded putative protein kinase superfamily protein cytochrome P450, flavonol synthase 1, and thioesterase superfamily protein. These genes are associated with the mechanism of wilting in several plants. Reportedly, the cytochrome P450 is a major component of the underlying resistance molecular mechanism for verticillium wilt in cotton. The *flavanol synthase 1* gene has been reported to be a constituent of the flavonoid pathway, which

are important regulators of biotic and abiotic stresses as an integral component of hormone signaling pathways, such as in Arabidopsis (Owens et al., 2008). However, its role as a modulator of plant–pathogen interaction is not clear yet. On the other hand, the thioesterase superfamily protein has been studied as an enhancer of drought tolerance in tobacco (Zhang et al., 2012), which makes it a suitable candidate as the modulator of abiotic stress tolerance. However, our study reported a SNP encoding the thioesterase superfamily protein. Based on our results, more studies should be conducted to ascertain the putative role of the respective genes for the common bean–*Cff* interaction.

So far, no other studies have been reported for candidate gene discovery for BW resistance in common bean. Our study, on the other hand, has successfully identified the presence of putative candidate genes associated with the BW resistance in common bean.

Genomic Prediction for Genomic Selection

In this study, GP was conducted using two approaches: 1) by using all 4,568 SNPs and 2) by using the 14 selected SNPs, in combination with seven GP models 1) rrBLUP, 2) Bayes A, 3) Bayes B, 4) BL, 4) BRR, 5) RF, 6) SVM, and 7) BL for each of the three BW isolates.

The average (r) calculated for all the SNPs and the 14 SNPs (**Table 3**) indicated an overall lower value when the selected 14 SNPs were used (**Table 3**; **Supplementary Figures S6, S7**). However, exploring PA further using the entire seven models predicted slightly different trends for BW_557 and BW_597 in each of the two approaches. The BW_528 followed the similar trend of average (r) as obtained for the general PA. The BW_557 depicted slightly higher values of PA for the 14 SNPs set with the RF model (**Table 3**; **Supplementary Figures S6, S7**), deviating from the general trend previously obtained. In addition, the PA has been reported to be low with use of less number of SNPs (Ali et al., 2020). Using a SNP set of 2000 or more reportedly shows an r value of 0.85 in comparison to the r value of 0.80 when less SNPs (1000 SNPS) were used for a population for soybean accessions (Zhang et al., 2016). Likewise, our study also showed a similar trend for BW_528 and BW_557, suggesting that the use of a higher number of SNPs is more reliable for GP. On the other hand, BW_597 had a very different trend with a higher PA for the 14 selected SNPs set with the five models (Bayes A, Bayes B, BL, BRR, and RF) and with lower PA for the respective SNPs set for the rrBLUP and similar PA for the SVM model (**Table 3**; **Supplementary Figures S6, S7**). The different trends suggest that it might be more beneficial to perform GP using the 14 SNPs selected from the GWAS analysis for BW_597 rather than deploying the generic 4,568 SNPs, to estimate GP more accurately with the seven models for BW_597. Evidently, Qin et al. (2019) reported that the r values deviate from a higher range of 0.64–0.74 when GWAS-selected SNPs were deployed to carry out GP rather than using the randomly selected SNPs (Qin et al., 2019). Qin et al. (2019) also reported the average correlation coefficient (r) among the 15 amino acids to range from 0.18 to 0.61 when all the 23,279 SNPs were used for GP and 0.45 to 0.68 upon using 231 SNP markers, using the rrBLUP model (Qin et al., 2019). Accordingly, the trend for BW_597 in our study is justified and supports the use of 14 GWAS-derived SNPs in combination with the seven GP models to be more beneficial for genomic prediction rather than use of all the generic 4,568 SNPs (**Table 3**; **Supplementary Figures S6, S7**). The results confirmed the accuracy of using the predicted models accordingly, with similar results.

Moreover, GS, based on the estimation of PA through use of Pearson's correlation coefficient (r) between the observed values and GEBV, has been employed to assess 481 common bean elite lines for resistance to environmental stress with reported prediction abilities between 0.6 and 0.8 for various traits (Keller et al., 2020). Several other studies have been reported for various biotic and abiotic stresses in common beans with genomic prediction (Barili et al., 2018; Jain et al., 2019; Shi et al., 2021), but no study has been reported to date for BW resistance in common bean utilizing genomic prediction. Significantly, here we report the use of genomic prediction for BW resistance in common bean, deploying seven prediction models under two SNP scenarios. The changes in the values of genomic prediction accuracies in our study indicate that the PA was affected by the SNP population size for our set of tested germplasm. Thus, our results indicate that the GS prediction can be effectively used in combination with MAS to breed for BW resistance in common bean.

CONCLUSION

Our study successfully tested 168 common bean accessions from the USDA NPGS based on public phenotypic resistance data for genome-wide association study (GWAS) and genomic prediction (GP). A total of 14 SNPs, on chromosome Pv02, Pv04, Pv07, Pv08, Pv09, Pv010, and Pv011 with 14 candidate genes, and 21 lines with potential resistance to the BW_528, BW_557, and BW_597 types were identified as a result of this study. The different SNPs and candidate genes identified for each isolate can be pyramided to enhance resistance to multiple isolates of BW. Moreover, the identified SNPs and candidate genes can be further explored and employed using genome editing and breeding techniques to develop common bean cultivars with enhanced resistance to the three BW isolates.

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 authors.

AUTHOR CONTRIBUTIONS

SC, PG, and AS are the principal investigators (PI) for the study on common bean. BZ wrote the draft of the manuscript. HS, MB,

KO, and BO performed phenotyping of BW resistance in common bean accessions. AS performed data analysis. PG performed genotypic data collection using the SNP chip. DO, WR, and HX assisted in the study. All authors have edited, reviewed, and approved the manuscript.

FUNDING

This research was supported by the USDA Germplasm Evaluation grant with Agreement Number/FAIN: 58-2090-8-053 and Project No. 2090-21000-032-11S; Accession No. 43498 and USDA NIFA Hatch project ARK0VG2018 and ARK02440.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.853114/full#supplementary-material

Data Sheet 1 | Supplementary Figures

Data Sheet 2 | Supplementary Tables

REFERENCES

Ali, M., Zhang, Y., Rasheed, A., Wang, J., and Zhang, L. (2020). Genomic Prediction for Grain Yield and Yield-Related Traits in Chinese winter Wheat. *Int. J. Mol. Sci.* 21 (4), 1342. doi:10.3390/ijms21041342

Allen, L. H. (2013). "Legumes," in *Caballero BBT-E of HN*. Third E (Waltham: Academic Press), 74–79. doi:10.1016/B978-0-12-375083-9.00170-7

Ashrafi, H., Kinkade, M., and Foolad, M. R. (2009). A New Genetic Linkage Map of Tomato Based on aSolanum Lycopersicum ×S. pimpinellifoliumRIL Population Displaying Locations of Candidate Pathogen Response Genes. *Genome* 52 (11), 935–956. doi:10.1139/G09-065

Assefa, T., Assibi Mahama, 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 (2). doi:10.1007/s11032-018-0920-0

Barili, L. D., Vale, N. M. d., Silva, F. F. e., Carneiro, J. E. d. S., Oliveira, H. R. d., Vianello, R. P., et al. (2018). Genome Prediction Accuracy of Common Bean via Bayesian Models. *Cienc. Rural* 48 (8), 1–6. doi:10.1590/0103-8478cr20170497

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 (1), 300–313. doi:10.1111/j.1469-8137.2012.04377.x

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 (19), 2633–2635. doi:10.1093/bioinformatics/btm308

CABI and EPPO (1996). *Data Sheets on Quarantine Pests- Ralstonia Solanacearum*, 138, 1–8.

Carpenter, M. A., Goulden, D. S., Woods, C. J., Thomson, S. J., Kenel, F., Frew, T. J., et al. (2018). Genomic Selection for Ascochyta Blight Resistance in Pea. *Front. Plant Sci.* 9, 1878. doi:10.3389/fpls.2018.01878

Cary., J. M. P. (2008). *Genomics* 7. 978-1-59994–920-8.

Celmeli, T., Sari, H., Canci, H., Sari, D., Adak, A., Eker, T., et al. (2018). The Nutritional Content of Common Bean (Phaseolus Vulgaris l.) Landraces in Comparison to Modern Varieties. *Agronomy* 8 (9), 166. doi:10.3390/agronomy8090166

Chung, Y. S., Choi, S. C., Jun, T.-H., and Kim, C. (2017). Genotyping-by-sequencing: a Promising Tool for Plant Genetics Research and Breeding. *Hortic. Environ. Biotechnol.* 58 (5), 425–431. doi:10.1007/s13580-017-0297-8

Coyne, D. P., Schuster, M. L., and 'Emerson' (1971). The New Large-Seeded Great Northern Dry Bean Variety Tolerant to Bacterial Wilt Disease. *Publ. Univ. Nebr. Agric. Exp. Station* 516, 11.

Coyne, D. P., Schuster, M. L., and Young, J. O. (1965). A Genetic Study of Bacterial Wilt (*Corynebacterium Flaccumfaciens Var. Aurantiacum*) Tolerance in *Phaseolus vulgaris* Crosses in Field Beans. *Proc. Am. Soc. Hortic. Sci.* 87, 279–285.

Cui, Z., Dong, H., Zhang, A., Ruan, Y., He, Y., and Zhang, Z. (2020). Assessment of the Potential for Genomic Selection to Improve Husk Traits in maize. *G3: Genes, Genomes, Genet.* 10 (10), 3741–3749. doi:10.1534/g3.120.401600

Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Ruas, P. M., Sant'Ana, G. C., Gepts, P., et al. (2021). Population Structure, Genetic Diversity and Genomic Selection Signatures Among a Brazilian Common Bean Germplasm. *Sci. Rep.* 11 (1), 1–12. doi:10.1038/s41598-021-82437-4

Ganesan, K., and Xu, B. (2017). Polyphenol-rich Lentils and Their Health Promoting Effects. *Ijms* 18 (11), 2390. doi:10.3390/ijms18112390

González, A. J., Tello, J. C., and Rodicio, M. R. (2005). Bacterial Wilt of Beans (*Phaseolus vulgaris*) Caused by *Curtobacterium Flaccumfaciens* in Southeastern Spain. *Plant Dis.* 89 (12), 1361. doi:10.1094/PD-89-1361C

Guzmán-Maldonado, S. H., Acosta-Gallegos, J., and Paredes-López, O. (2000). Protein and mineral Content of a Novel Collection of Wild and Weedy Common Bean (Phaseolus vulgaris L). *J. Sci. Food Agric.* 80 (13), 1874–1881. doi:10.1002/1097-0010

Harveson, R. M., Urrea, C. A., and Schwartz, H. F. (2011). Bacterial Wilt of Dry Beans in Western Nebraska. *NebGuide*. 9. 10.1094/PDIS-03-15-0299-FE

Hedges, F. (1926). Bacterial Wilt of Beans (*Bacterium Flaccumfaciens Hedges*), Including Comparisons with *Bacterium Phaseoli*. *Phytopathology* 16 (1), 1–22.

Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic Selection for Crop Improvement. *Crop Sci.* 49 (1), 1–12. doi:10.2135/cropsci2008.08.0512

Hsieh, T., Huang, H., Mündel, H., Scott, R., Canada, A., and Box, P. O. (2003). A Rapid Indoor Technique for Screening Common Bean (Phaseolus vulgaris L.) for Resistance to Bacterial Wilt [Curtobacterium Flaccumfaciens Pv. Flaccumfaciens (Hedges) Collins and Jones]. *Revista Mexicana de Fitopatología.* 21 (3), 364–369.

Hsieh, T.-F., Huang, H. C., and Erickson, R. S. (2006). Bacterial Wilt of Common Bean: Effect of Seedborne Inoculum on Disease Incidence and Seedling Vigour. *Seed Sci. Technol.* 34 (1), 57–67. doi:10.15258/sst.2006.34.1.07

Huang, H. C., Erickson, R. S., Balasubramanian, P. M., Hsieh, T. F., and Conner, R. L. (2009). Resurgence of Bacterial Wilt of Common Bean in North America. *Can. J. Plant Pathol.* 31 (3), 290–300. doi:10.1080/07060660909507603

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a Package for the Next Level of Genome-wide Association Studies with Both Individuals and Markers in the Millions. *GigaScience* 8 (2). doi:10.1093/gigascience/giy154

Hughes, J. S., and Swanson, B. G. (1989). Soluble and Insoluble Dietary Fiber in Cooked Common Bean (*Phaseolus Vulgaris*) Seeds. *Food Microstructure* 8 (1), 15–21.

Jain, S., Poromarto, S., Osorno, J. M., McClean, P. E., and Nelson, B. D. (2019). Genome Wide Association Study Discovers Genomic Regions Involved in Resistance to Soybean Cyst Nematode (*Heterodera glycines*) in Common Bean. *PLoS ONE* 14 (2), e0212140–17. doi:10.1371/journal.pone.0212140

Jarquin, D., Specht, J., and Lorenz, A. (2016). Prospects of Genomic Prediction in the USDA Soybean Germplasm Collection: Historical Data Creates Robust Models for Enhancing Selection of Accessions. *G3: Genes, Genomes, Genet.* 6 (8), 2329–2341. doi:10.1534/g3.116.031443

Juncus, L. (1998). *Plant Guide*. (Figure 3):5–8.

Kariola, T., Brader, G., Li, J., and Palva, E. T. (2005). Chlorophyllase 1, a Damage Control Enzyme, Affects the Balance between Defense Pathways in Plants. *Plant Cell* 17 (1), 282–294. doi:10.1105/tpc.104.025817

Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H. F., et al. (2020). Genomic Prediction of Agronomic Traits in Common Bean (*Phaseolus vulgaris* L.) under Environmental Stress. *Front. Plant Sci.* 11, 1–15. doi:10.3389/fpls.2020.01001

Kuzay, S., Hamilton-Conaty, P., Palkovic, A., and Gepts, P. (2020). Is the USDA Core Collection of Common Bean Representative of Genetic Diversity of the

Species, as Assessed by SNP Diversity? *Crop Sci.* 60, 1398–1414. doi:10.1002/csc2.20032

Larkin, D. L., Lozada, D. N., and Mason, R. E. (2019). Genomic Selection-Considerations for Successful Implementation in Wheat Breeding Programs. *Agronomy* 9 (9), 479. doi:10.3390/agronomy9090479

Legarra, A., Robert-granié, C., Croiseau, P., Guillaume, F., and Fritz, S. (2011). Improved Lasso for Genomic Selection. *Genet. Res.* 93 (1), 77–87. doi:10.1017/S0016672310000534

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 (Oxford, England)* 28 (18), 2397–2399. doi:10.1093/bioinformatics/bts444

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 (2), e1005767. doi:10.1371/journal.pgen.1005767

Liu, X., Wang, H., Hu, X., Li, K., Liu, Z., Wu, Y., et al. (2019). Improving Genomic Selection with Quantitative Trait Loci and Nonadditive Effects Revealed by Empirical Evidence in maize. *Front. Plant Sci.* 10 (10), 1129. doi:10.3389/fpls.2019.01129

López-Hernández, F., and Cortés, A. J. (2019). Last-Generation Genome-Environment Associations Reveal the Genetic Basis of Heat Tolerance in Common Bean (Phaseolus vulgaris L). *Front. Genet.* 10, 954. doi:10.3389/fgene.2019.00954

Maenhout, S., De Baets, B., Haesaert, G., and Van Bockstaele, E. (2007). Support Vector Machine Regression for the Prediction of maize Hybrid Performance. *Theor. Appl. Genet.* 115 (7), 1003–1013. doi:10.1007/s00122-007-0627-9

Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., et al. (2013). Demographic Factors Shaped Diversity in the Two Gene Pools of Wild Common Bean Phaseolus vulgaris L. *Heredity* 110 (3), 267–276. doi:10.1038/hdy.2012.82

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP Markers and Their Impact on Plant Breeding. *Int. J. Plant Genomics* 2012, 1–11. doi:10.1155/2012/728398

Marone, D., Russo, M., Laidò, G., De Leonardis, A., and Mastrangelo, A. (2013). Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. *Ijms* 14 (4), 7302–7326. doi:10.3390/ijms14047302

Messerer, M., Lang, D., and Mayer, K. (2018). Analysis of Stress Resistance Using Next Generation Techniques. *Agronomy* 8 (8), 130. doi:10.3390/agronomy8080130

Mündel, H.-H., Kiehn, F. A., Huang, H. C., Conner, R. L., and Saindon, G. (2005). Resolute Great Northern Bean. *Can. J. Plant Sci.* 85 (2), 393–395. doi:10.4141/P04-109

Ogutu, J. O., Piepho, H.-P., and Schulz-Streeck, T. (2011). A Comparison of Random Forests, Boosting and Support Vector Machines for Genomic Selection. *BMC Proc.* 5 (Suppl. 3), S11. doi:10.1186/1753-6561-5-S3-S11

Osdaghi, E., Young, A. J., and Harveson, R. M. (2020). Bacterial Wilt of Dry Beans Caused by Curtobacterium Flaccumfaciens Pv. Flaccumfaciens : A New Threat from an Old Enemy. *Mol. Plant Pathol.* 21 (5), 605–621. doi:10.1111/mpp.12926

Owens, D. K., Alerding, A. B., Crosby, K. C., Bandara, A. B., Westwood, J. H., and Winkel, B. S. J. (2008). Functional Analysis of a Predicted Flavonol Synthase Gene Family in Arabidopsis. *Plant Physiol.* 147 (3), 1046–1061. doi:10.1104/pp.108.117457

Qin, J., Shi, A., Song, Q., Li, S., Wang, F., Cao, Y., et al. (2019). Genome Wide Association Study and Genomic Selection of Amino Acid Concentrations in Soybean Seeds. *Front. Plant Sci.* 10, 1445. doi:10.3389/fpls.2019.01445

Ravelombola, W., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2019). Genome-Wide Association Study and Genomic Selection for Soybean Chlorophyll Content Associated With Soybean Cyst Nematode. *BMC Genom.* 20, 904. doi:10.1186/s12864-019-6275-z

Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Alvarez, S., Vlasova, A., Capella-Gutiérrez, S., Ordaz-Ortiz, J. J., et al. (2017) Genomic History of the Origin and Domestications of Common Bean in the Americas Unveils its Closest Sister Species. *Genome Biol.* 18, 60. doi:10.1186/s13059-017-1190-6

Rice, B., and Lipka, A. E. (2019). Evaluation of RR-BLUP Genomic Selection Models that Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum. *Plant Genome* 12 (1), 180052. doi:10.3835/plantgenome2018.07.0052

Santos, A. S., Bressan-Smith, R. E., Pereira, M. G., Rodrigues, R., and Ferreira, C. F. (2003). Genetic Linkage Map of *Phaseolus vulgaris* and Identification of QTLs Responsible for Resistance to *Xanthomonas Axonopodis Pv. Phaseoli*. *Fitopatol. Bras.* 28 (1), 5–10. doi:10.1590/s0100-41582003000100001

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 (7), 707–713. doi:10.1038/ng.3008

Schwartz, H. F., Otto, K., Brick, M. A., and Ogg, J. B. (2009). Resistance to Bacterial Wilt in the Phaseolus Core Collection. *Ann. Rept. Bean Improv. Coop.* 52, 68–69.

Schwartz, H. F., Brick, M. A., Otto, K., and Ogg, J. B. (2010). Germplasm Evaluation for Resistance to Bacterial Wilt in Common Bean, 2008-2009. *APS Plant Dis. Manage. Rep.* 4V125, 2.

Sendi, Y., Pfeiffer, T., Koch, E., Mhadhbi, H., and Mrabet, M. (2020). Potential of Common Bean (Phaseolus vulgaris L.) Root Microbiome in the Biocontrol of Root Rot Disease and Traits of Performance. *J. Plant Dis. Prot.* 127 (4), 453–462. doi:10.1007/s41348-020-00338-6

Shi, A., Gepts, P., Song, Q., Xiong, H., Michaels, T.E., and Chen, S. (2021). Genome-Wide Association Study and Genomic Selection for Soybean Cyst Nematode Resistance in USDA Common Bean (*Phaseolus vulgaris*) Core Collection. *Front. Plant Sci.* doi:10.3389/fpls.2021.624156/full

Shin, I. S., Hsu, J.-C., Huang, S.-M., Chen, J.-R., Wang, J.-F., Hanson, P., et al. (2020). Construction of a Single Nucleotide Polymorphism Marker Based QTL Map and Validation of Resistance Loci to Bacterial Wilt Caused by *Ralstonia Solanacearum* Species Complex in Tomato. *Euphytica* 216 (3), 1–20. doi:10.1007/s10681-020-2576-1

Song, H., Guo, Z., Hu, X., Qian, L., Miao, F., Zhang, X., et al. (2019). Evolutionary Balance between LRR Domain Loss and Young NBS-LRR Genes Production Governs Disease Resistance in *Arachis hypogaea* Cv. Tifrunner. *BMC Genomics* 20 (1), 844. doi:10.1186/s12864-019-6212-1

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PloS one* 8 (1), e54985. doi:10.1371/journal.pone.0054985

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic Selection and Association Mapping in Rice (*Oryza Sativa*): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *Plos Genet.* 11 (2), e1004982. doi:10.1371/journal.pgen.1004982

Thoquet, P., Sperisen, O. C., and Rogowsky, P. (1996). Polygenic Resistance of Tomato Plants to Bacterial Wilt in the French West Indies. *Mpmi* 9 (9), 837–842. doi:10.1094/mpmi-9-0837

Urrea, C. A., and Harveson, R. M. (2014). Identification of Sources of Bacterial Wilt Resistance in Common Bean (*Phaseolus vulgaris*). *Plant Dis.* 98 (7), 973–976. doi:10.1094/PDIS-04-13-0391-RE

USDA-NASS Dry Beans (2022). *Planted Acreage by Year, US and Major States*. Available at: https://www.nass.usda.gov/Charts_and_Maps/Dry_Beans,_Dry_Peas,_and_Lentils/dbinten.php (Accessed March 15, 2022).

Valdisser, P. A. M. R., Müller, B. S. F., de Almeida Filho, J. E., Morais Júnior, O. P., Guimarães, C. M., Borba, T. C. O., et al. (2020). Genome-Wide Association Studies Detect Multiple QTLs for Productivity in Mesoamerican Diversity Panel of Common Bean under Drought Stress. *Front. Plant Sci.* 11, 1563. doi:10.3389/fpls.2020.574674

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017). Genomic Selection Models Double the Accuracy of Predicted Breeding Values for Bacterial Cold Water Disease Resistance Compared to a Traditional Pedigree-Based Model in Rainbow trout Aquaculture. *Genet. Sel Evol.* 49 (1), 17. doi:10.1186/s12711-017-0293-6

Villordo-Pineda, E., González-Chavira, M. M., Giraldo-Carbajo, P., Acosta-Gallegos, J. A., and Caballero-Pérez, J. (2015). Identification of Novel Drought-Tolerant-Associated SNPs in Common Bean (*Phaseolus vulgaris*). *Front. Plant Sci.* 6, 546. doi:10.3389/fpls.2015.00546

Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five Years of GWAS Discovery. *Am. J. Hum. Genet.* 90 (1), 7–24. doi:10.1016/j.ajhg.2011.11.029

Vlasova, A., Capella-Gutierrez, S., Rendon-Anaya, M., Hernandez-Onate, M., Minoche, A. E., et al. (2016). Genome and Transcriptome Analysis of the Mesoamerican Common Bean and the Role of Gene Duplications in Establishing Tissue and Temporal Specialization of Genes. *Genome Biol.* 17, 32.

Waldmann, P. (2019). On the Use of the pearson Correlation Coefficient for Model Evaluation in Genome-wide Prediction. *Front. Genet.* 10 (SEP), 1–4. doi:10.3389/fgene.2019.00899

Wang, J.-F., Ho, F.-I., Truong, H. T. H., Huang, S.-M., Balatero, C. H., Dittapongpitch, V., et al. (2013). Identification of Major QTLs Associated with Stable Resistance of Tomato Cultivar 'Hawaii 7996' to Ralstonia Solanacearum. *Euphytica* 190 (2), 241–252. doi:10.1007/s10681-012-0830-x

Wang, J., Zhou, Z., Zhang, Z., Li, H., Liu, D., Zhang, Q., et al. (2018a). Expanding the BLUP Alphabet for Genomic Prediction Adaptable to the Genetic Architectures of Complex Traits. *Heredity* 121 (6), 648–662. doi:10.1038/s41437-018-0075-0

Wang, L., Zhou, X., Ren, X., Huang, L., Luo, H., Chen, Y., et al. (2018b). A Major and Stable QTL for Bacterial Wilt Resistance on Chromosome B02 Identified Using a High-Density SNP-Based Genetic Linkage Map in Cultivated Peanut Yuanza 9102 Derived Population. *Front. Genet.* 9, 1–13. doi:10.3389/fgene.2018.00652

Wang, S., Xu, Y., Qu, H., Cui, Y., Li, R., Chater, J. M., et al. (2020). Boosting Predictabilities of Agronomic Traits in rice Using Bivariate Genomic Selection. *Brief. Bioinform.* 22. doi:10.1093/bib/bbaa103

Wen, L., Chang, H.-X., Brown, P. J., Domier, L. L., and Hartman, G. L. (2019). Genome-wide Association and Genomic Prediction Identifies Soybean Cyst Nematode Resistance in Common Bean Including a Syntenic Region to Soybean Rhg1 Locus. *Hortic. Res.* 6 (1). doi:10.1038/s41438-018-0085-3

Yerkes, W. D., and Crispin, A. M. (1956). Bean Diseases of Importance in Mexico in 1955. *Plant Dis. Reporter* 40, 222–223.

Zhang, J., Song, Q., Cregan, P. B., and Jiang, G.-L. (2016). Genome-wide Association Study, Genomic Prediction and Marker-Assisted Selection for Seed Weight in Soybean (Glycine max). *Theor. Appl. Genet.* 129 (1), 117–130. doi:10.1007/s00122-015-2614-x

Zhang, L.-H., Jia, B., Zhuo, R.-Y., Liu, J.-L., Pan, H.-Y., Baldwin, T. C., et al. (2012). An Acyl-Acyl Carrier Protein Thioesterase Gene Isolated from Wintersweet (Chimonanthus Praecox), CpFATB, Enhances Drought Tolerance in Transgenic Tobacco (Nicotiana Tobaccum). *Plant Mol. Biol. Rep.* 30 (2), 433–442. doi:10.1007/s11105-011-0359-5

Zienkiewicz, M. (2016). *Plant Breeding 101*. Alberta Seed Guide. Available at: https://www.seed.ab.ca/plant-breeding-101/.

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 © 2022 Zia, Shi, Olaoye, Xiong, Ravelombola, Gepts, Schwartz, Brick, Otto, Ogg and Chen. 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 | Frontiers in Genetics

REVIEW
published: 28 June 2022
doi: 10.3389/fgene.2022.859437

Gene-Editing Technologies and Applications in Legumes: Progress, Evolution, and Future Prospects

*Mehmet Cengiz Baloglu[1], Yasemin Celik Altunoglu[1], Pinar Baloglu[2], Ali Burak Yildiz[3], Nil Türkölmez[3] and Yelda Özden Çiftçi[3,4]**

[1]Department of Genetics and Bioengineering, Faculty of Engineering and Architecture, Kastamonu University, Kastamonu, Turkey, [2]Research and Application Center, Kastamonu University, Kastamonu, Turkey, [3]Department of Molecular Biology and Genetics, Gebze Technical University, Kocaeli, Turkey, [4]Smart Agriculture Research and Application Center, Gebze Technical University, Kocaeli, Turkey

Legumes are rich in protein and phytochemicals and have provided a healthy diet for human beings for thousands of years. In recognition of the important role they play in human nutrition and agricultural production, the researchers have made great efforts to gain new genetic traits in legumes such as yield, stress tolerance, and nutritional quality. In recent years, the significant increase in genomic resources for legume plants has prepared the groundwork for applying cutting-edge breeding technologies, such as transgenic technologies, genome editing, and genomic selection for crop improvement. In addition to the different genome editing technologies including the CRISPR/Cas9-based genome editing system, this review article discusses the recent advances in plant-specific gene-editing methods, as well as problems and potential benefits associated with the improvement of legume crops with important agronomic properties. The genome editing technologies have been effectively used in different legume plants including model legumes like alfalfa and lotus, as well as crops like soybean, cowpea, and chickpea. We also discussed gene-editing methods used in legumes and the improvements of agronomic traits in model and recalcitrant legumes. Despite the immense opportunities genome editing can offer to the breeding of legumes, governmental regulatory restrictions present a major concern. In this context, the comparison of the regulatory framework of genome editing strategies in the European Union and the United States of America was also discussed. Gene-editing technologies have opened up new possibilities for the improvement of significant agronomic traits in legume breeding.

Keywords: genome-editing methods, CRISPR/Cas9, TALEN, ZFN, legumes

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Muntazir Mushtaq,
National Bureau of Plant Genetic Resources (ICAR), India
Narayanan Manikanda Boopathi,
Tamil Nadu Agricultural University, India

****Correspondence:***
Yelda Özden Çiftçi
ozden@gtu.edu.tr

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

Received: *21 January 2022*
Accepted: *13 May 2022*
Published: *28 June 2022*

Citation:
Baloglu MC, Celik Altunoglu Y, Baloglu P, Yildiz AB, Türkölmez N and Özden Çiftçi Y (2022) Gene-Editing Technologies and Applications in Legumes: Progress, Evolution, and Future Prospects. Front. Genet. 13:859437. doi: 10.3389/fgene.2022.859437

1 INTRODUCTION

Legumes are members of the angiosperm group and contain 19,500 species in 751 genera (Lewis et al., 2005). In addition to their nutritional value, the legume family contains many crops that contain essential amino acids and plant-based proteins. Additionally, legumes play an essential role in cultivating sustainable agriculture by symbiotically fixing nitrogen and releasing high-quality organic matter into the soil. Although legumes provide health benefits as well as ecological significance, their cultivation is affected by lower crop yields due to stress factors. The current

focus is on accelerating genetic gains related to yield, stress tolerance, and nutritional quality. Most genetic improvement of legumes has been accomplished over the past half-century through pedigree and performance-based selection. To achieve faster genetic gains in legumes, novel genomics techniques and high-throughput phenomics are widely used and resulted in improved legume varieties that possess important agronomic traits (Varshney et al., 2018). To increase yield potential and reliability, different approaches have been used such as genomic selection (or marker-assisted selection) and precision breeding (gene editing) (Bhowmik et al., 2021). With increasing access to information on genes and haplotypes that contribute to agronomically significant traits, genome editing has allowed the modification of multiple SNPs without affecting the original characteristics of a popular cultivar. Genetic barriers such as ploidy differences prevent many legume crop species from exchanging genetic material naturally; so, the enormous genetic diversity they hold in their wild relatives, however, remains unused (Varshney et al., 2018). The availability of the complete genome sequence of organisms makes a significant contribution to the advancement of new-generation genome-editing studies. Compared to other family members of legumes, there are more new-generation genome-editing studies in *Lotus japonicus* (Sato et al., 2008), *Glycine max* (Schmutz et al., 2010), and *Medicago truncatula* (Young et al., 2011), all of which have been fully sequenced, which supports the importance of having a complete genome. Although the availability of the whole-genome sequences of other legume species includes common bean (Schmutz et al., 2014), mung bean (Kang et al., 2014), lentil (Bett et al., 2014; Bett et al., 2016), and pea (Kreplak et al., 2019), genome-editing trials for those plants have not been conducted.

This article presents the mechanisms of new-generation genome-editing technologies including TALEN (transcription activator-like effector nucleases), ZFN (zinc finger nuclease), and CRISPR/Cas9 [the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 nuclease (Cas9)] systems, and detailed application examples of those technologies in legume family members. Moreover, the regulatory framework and future of genome-editing technologies in those crops have been extensively mentioned. This current study offers the comprehensive coverage of genome-editing studies in plants in the legume family, making it a collective resource.

2 GENE-EDITING TECHNOLOGIES

2.1 CRISPR/Cas9

In recent years, CRISPR/Cas9 technology has become popular for genome editing in various organisms as well as in plants. It has also increased the scope of agricultural research by allowing the creation of novel plant varieties with the deletion of harmful features or the addition of prominent characters. CRISPR is a rapidly developing technique that can be used for a variety of genetic manipulations, including generating knockouts, making precise modifications, creating multiplex genome engineering, or activating and repressing genes (Arora and Narula, 2017). In the CRISPR/Cas9 system, there are two main elements: Cas9 protein and guide RNA (gRNA). Cas9 is an RNA-dependent DNA endonuclease that binds and makes a complex with gRNA (**Figure 1A**). The gRNA consists of 20 nucleotides that are complementary to the target DNA sequence and serves as a recruiting signal for Cas9. To recognize the target DNA sequence, CRISPR/Cas9 employs RNA–DNA interactions instead of DNA–protein interactions, which is the major difference between it and the other genome-editing technologies. Two different DNA-binding domains are needed for each target site for ZFNs and TALENs that employ DNA–protein interactions to target their specific sequences. This method is quite troublesome. When it comes to CRISPR/Cas9, which uses DNA–RNA interactions, only an 18–20 base pair needs to be designed. It is essential that Cas9 and gRNA attach to a specific protospacer adjacent motif (PAM), a sequence that is found at the 3′ end of target sequences (**Figure 1B**). The sequence 5′-NGG-3′ is the PAM for Cas9 from *Streptococcus pyogenes* that is the most frequently selected and utilized for genome editing. Through the introduction of double-stranded breaks (DSBs) in the DNA target, Cas9 induces DNA repair. The repair mechanism is achieved through non-homologous end joining (NHEJ) to make gene knockouts and homology-directed repair (HDR) to generate gene modification and gene insertion (Huang and Puchta, 2019) (**Figure 1C**). A frameshift mutation occurs when NHEJ randomly inserts or deletes a DNA strand within a coding region, resulting in a gene knockout. It is not necessary to use a homologous repair template for NHEJ. Therefore, the NHEJ repair mechanism is by far the most popular and optimized method of repairing DNA damage in all plants including legumes. The HDR technique, on the other hand, can exactly insert predetermined sequences coming from a donor DNA template. Due to the low editing efficiency of HDR, its application in plants has been restricted (Atkins and Voytas, 2020). With the CRISPR technology, potential applications include the analysis of gene expression, gain-of-function, and loss-of-function. Applications for CRISPR in legume farming have been performed, and this genome-editing technology can be used to produce high-quality, sustainable agricultural products, including legumes.

2.1.1 Design of gRNAs

The gRNA, one of the main components of CRISPR/Cas9, plays a vital role in determining the specificity and efficiency of gene editing. Because of big gene families in plants, a high amount of sequence similarities and repetitive motifs occur and cause off-target gRNA binding, which is the main problem. As a result, gRNA binds unintended targets and unpredictable effects can be observed. To avoid these concerns, there are several requirements and preferences to take into account when selecting target motifs (Kumlehn et al., 2018). The most important limitation for choosing target DNA sequences arises from the availability of PAM that is triplet uniquely attached by the Cas endonuclease. In the case of Cas9 from *S. pyogenes*, PAM is composed of a flexible nucleotide followed by two guanosines, and this is known as NGG. To perform site-directed mutagenesis, the selection of two or three targets that are positioned 20 nucleotides upstream of a

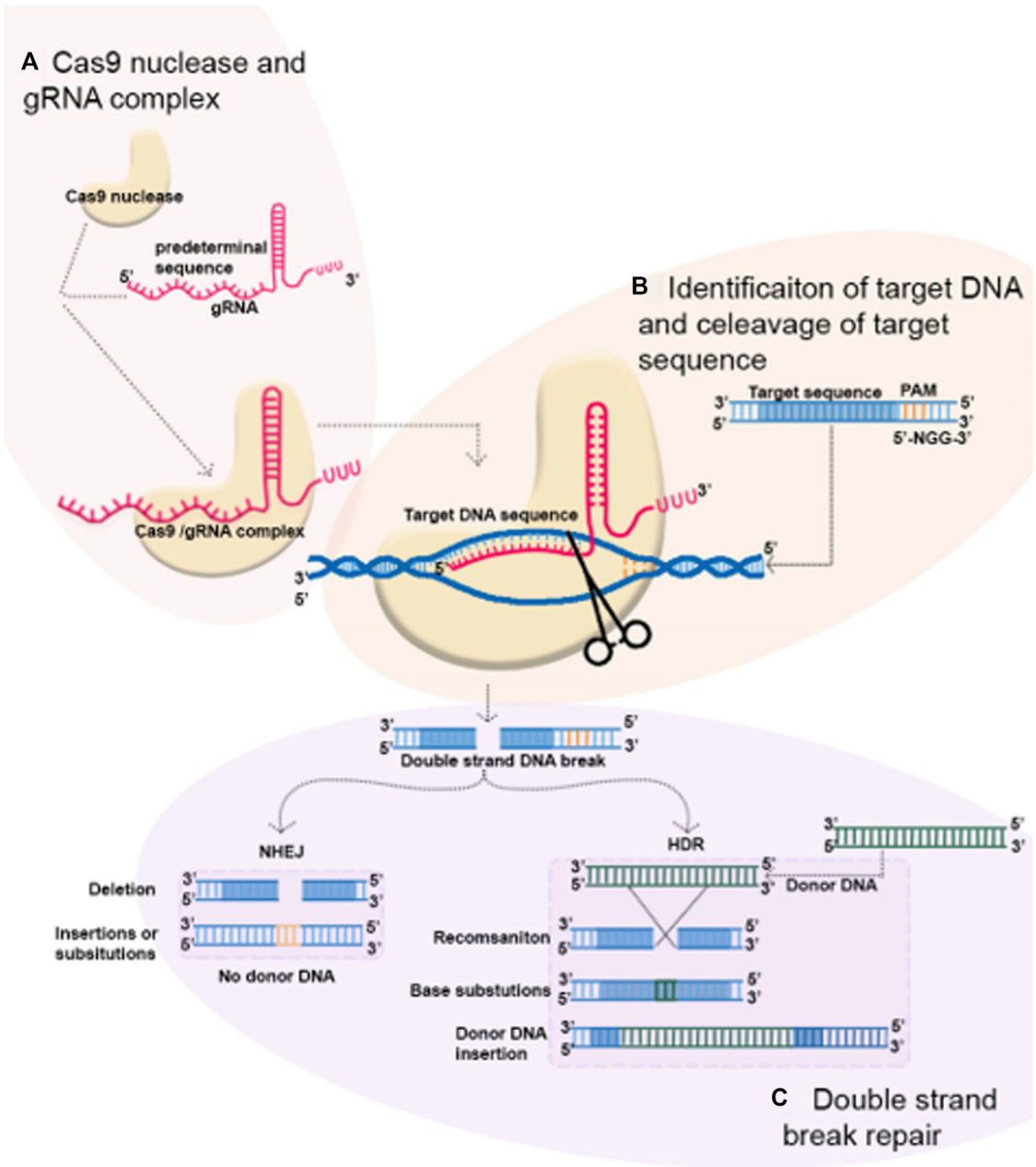


FIGURE 1 | Mechanism of CRISPR-Cas9 genome editing. **(A)** CRISPR-Cas9 system is composed of the Cas9 protein and gRNA and Cas9/gRNA complex occur. **(B)** Cas9/gRNA complex cleave targets DNA in a binary complex, causing a double-stranded DNA break. **(C)** DNA breaks are repaired by non-homologous end joining (NHEJ) and homology-directed repair (HDR). In the process, short insertion deletions, nucleotide substitutions, or gene insertion may occur.

double G within the coding sequence of the target gene is adequate. The activity of Cas cleavage mainly depends on the secondary structure formation in the gRNA as there are requirements for the 5-terminal part of the gRNA to pair with the target DNA and the Cas endonuclease. Two-dimensional (2D) structures form inside the gRNA 3′ terminal scaffold and play a very important role in the function of gRNA (Ma et al., 2015). There are numerous online tools such as RNAfold (rna.tbi.univie.ac.at/cgibin/RNAWebSuite/RNAfold.cgi) and MFOLD (Zuker, 2003) available for the prediction of the secondary structure of gRNAs *in silico*. There are also some software programs designing gRNAs that create or demolish restriction enzymes following editing, allowing users to perform a quick screening of editing events. Some of them are listed below: For CRISPR-Cas nucleases, CRISPOR (Concordet and Haeussler, 2018), CRISPR-P (Liu et al., 2017), RGEN Cas designer (Park et al., 2015), and CHOPCHOP (Labun et al., 2019). For base editors, RGEN BE-Designer (Hwang et al., 2018)

and PnB designer (Siegner et al., 2021). For prime editors, PrimeDesign (Hsu et al., 2021), pegFinder (Chow et al., 2021), and PlantPegDesigner (Lin et al., 2021). Despite the widespread use of software to design gRNAs, experienced users design gRNAs manually to suit specific purposes, such as detecting edits easily by restriction fragment length polymorphism analysis (Hassan et al., 2021).

2.1.2 Selection of the Best Cas Protein

There has been an enormous increase in the number and varieties of CRISPR/Cas genome editing technology over the past 5 years. As of 2015, Cas proteins were categorized into five types and 16 subtypes, under two major Cas classes that differ profoundly based on the elements of their effector modules that process and interfere with gRNAs (Makarova et al., 2015). As genome engineering technologies have improved, type VI RNA-targeting and numerous types V CRISPR/Cas subtypes were developed to extend the class two system capabilities (Stella et al., 2017). Later, different versions of types IV, I, and V systems were identified to reside in mobile genetic elements because they lack the ability to cleave targeted DNA (Klompe et al., 2019). A recent study uncovered new proteins in the type II system that serves functions other than adaptive immunity (Niewoehner et al., 2017). Cas1, Cas2, Cas3, Cas5, Cas6, and Cas7 in type 1, Cas1, Cas2, Cas9, and Csn2/Cas4 in type 2, Cas1, Cas2, Cas6, and Cas10 in type 3, Cas5/Csf3, Cas7/Csf2, and Cas8/Csf1 in type 4, Cas12/Cpf1/C2c1 in Type 5, and Cas13/C2c2/CasRx in Type 6 are the samples of different CRISPR proteins. DNA nuclease, ribonuclease, RNA cleavage, and crRNA processing are their main functions (Makarova et al., 2020). Despite a large number of Cas proteins, only a small portion has been used in genome editing in plants. PAM restrictions, codon-optimization of Cas proteins, off-side effects, and temperature sensitivity are the main troubles to the selection of appropriate Cas proteins. Genome editing is mediated by various classes of CRISPR/Cas systems. Depending on their mode of genome editing, they can be divided into four categories: 1) point mutations, 2) deletions, 3) insertions, or 4) a combination of them. All these mutation modes were performed in different legume species, and the details of strategies are discussed in **Section 3**.

For genome editing of crops, the type II CRISPR/Cas9 system has been the most chosen. Because of the simple design of the CRISPR/Cas9 system, the availability of only one single gRNA and defined PAM sequences, Cas9 proteins are the most selected and studied proteins in genome-editing research in plants. Practically, the CRISPR/Cas9 system is designed for the replacement of the tracrRNA and crRNA molecules of the bacteria with the guide RNA (gRNA) (Cong et al., 2013; Mali et al., 2013). It became possible to simplify the system by containing just a gRNA to guide the Cas9 protein to the target and resulting in cleavage of the target region on DNA. Genome editing is accomplished by delivering both of these components to the nucleus. As with Cas9, Cas12 works on similar principles and in the same manner; however, Cas12 effectors prefer T*rich PAMs instead of G-rich PAMs of Cas9. This allows it to target specific genomic regions with greater effectiveness (Zetsche et al., 2015; Kim et al., 2017). The Cas12 systems do not require tracrRNA for maturation and interference. In contrast to Cas9 gRNAs, a single molecule of RNA engineered to a length of 44 nucleotides is utilized in Cas12 systems. Rather than blunt ends created by Cas9 effectors, Cas12 effectors cause double-strand breaks with staggered ends. This makes them ideal for targeting specific genes. Cas12a/Cpf1 systems that are isolated from *Francisella tularensis novicida* (FnCas12a), *Acidaminococcus* sp. BV3L6 (AsCas12a), and *Lachnospiraceae bacterium* (LbCas12a) are frequently selected for genome editing in different plant species with high success rates (Zhan et al., 2021). Cas13a has also been utilized for plant genome-editing studies along with Cas9 and Cas12. The system has a non-specific RNase activity and can exhibit the cleavage of ssRNA similarly to types II and V CRISPR systems (Abudayyeh et al., 2017). For this reason, it was suggested that RNA interference studies be replaced with this method. According to the literature, Cas9 protein is the most preferred and used for the development of genome-edited plants (Kiryushkin et al., 2022). The expression of Cas9 under the cauliflower mosaic virus (CaMV) 35S promoter (p35S) was observed in a total of 78 plant-related genome-editing studies (Amack and Antunes, 2020). To date, the CRISPR/Cas 9 platform has been mainly used in legume crops for the improvement of agronomic traits (see **Section 3**). It has been optimized for routine use in legume crops. Therefore, it is suggested that the Cas9 proteins can be more suitable for legumes.

2.1.3 Vector Design

In recent years, CRISPR/Cas9—a genome-editing tool that has achieved worldwide fame—has been successfully used to edit the genomes of many monocot and dicot plants. To edit these genomes effectively, CRISPR/Cas9 components must be delivered to plants using an effective vector system that contains codon-optimized Cas9 gene and promoters for Cas9 and sgRNA. In addition, suitable target sites, efficient regeneration, and transformation methods must be specially optimized for the legume plants. It is necessary to deliver and express single guide RNA (sgRNA) and cas9 protein in the target cell for CRISPR editing (Arora and Narula, 2017). The expression of sgRNA is usually controlled by tissue-specific RNA polymerase III promoters such as AtU6 and TaU6. These promoters cause the production of specific small RNAs in various legume species. Like sgRNA, Cas9 has positioned downstream of RNA polymerase II promoters that guide the transcription of longer RNAs. For targeting nuclear DNA, Cas9 is mostly tagged with a nuclear localization sequence (NLS). The selection of suitable expression and Cas9 systems are critical factors for vector construction. Furthermore, restriction sites for the insertion of gRNA play a significant role. The website known as Addgene (http://www.addgene.org/crispr/plant/) provides information on the different types of plasmids for plant genome-editing studies. These plasmids in Addgene are empty backbones and usually possess three main components: sgRNA cassette, Cas9 endonuclease gene, and selection marker. RNA polymerase III promoters such as U3 or U6 have been obtained from monocot and dicot plants, and sgRNA has been directly expressed in plant cells. There are some genome-editing plasmids containing U3 or U6 promoters obtained from rice (Ma et al., 2015), maize (Qi et al.,

2018), wheat (Xing et al., 2014), and Arabidopsis (Tsutsui and Higashiyama, 2016) and are commercially available in Addgene. In dicots and some monocots, a codon-optimized Cas9 under the control of the CaMV 35S promoter has been used. The maize ubiquitin promoter has been an alternative option to obtain homozygous, hemizygous, or biallelic mutations in the T_0 generation that are passed on to subsequent generations (Shan et al., 2013; Zhang H. et al., 2014; Zhan et al., 2021). As the primary constituent of a CRISPR plasmid, the Cas endonuclease affects the rate of mutation during genome editing. SpCas9 (Cas9 from *S. pyogenes*) is the most preferred type of endonuclease used by researchers in plant genome-editing studies. For improvement of the performance of Cas9 endonuclease in the plant cell, different strategies have been developed. For example, codons of Cas9 have been optimized (Ma et al., 2015), the expression of Cas9 has been strengthened through strong promoters (Yue et al., 2020), and translational enhancers and nuclear localization signals have been added to the CRISPR cassette (Xie et al., 2015). For the efficient selection of genome-edited plants, different selection markers have been utilized. They are known as acetolactate synthase, phosphomannose isomerase, neomycin phosphotransferase, and hygromycin phosphotransferase (Hpt) (Yue et al., 2020). The *Hpt* gene, which confers tolerance to herbicide hygromycin, is the most frequently utilized marker selection gene in CRISPR-based plant breeding.

2.1.4 Advanced Designs

Although HDR provides precise nucleotide substitutions in some plants including sugar cane (Oz et al., 2021), tomato (Vu et al., 2020), and maize (Svitashev et al., 2016), its application in plants including recalcitrant legume species such as lentils, soybean, chickpea, bean, and pea has limited because of low editing efficiency (Huang and Puchta, 2019; Atkins and Voytas, 2020). Considering these limitations in precise genome-editing technologies and the lengthy breeding process for legume species, agronomically significant properties may be achieved in much less time with new alternative CRISPR-based genome-editing tools for legumes. They are called deaminase-mediated base editing and reverse transcriptase-mediated prime editing, which are more efficient than HDR in plants. These new technologies do not require DSB formation and donor DNA. A:T > G:C and C:G > T:A transitions can be introduced directly into targeted sites by using adenosine deaminase (adenine base editor, ABE) and cytidine deaminase (cytosine base editor, CBE), respectively (Zhu et al., 2020). Next-generation sequencing technologies have led to the development of genome assemblies for a number of legume crops even if they are fragmented (Garg et al., 2021). These genomic information are used to modify key regions of genes in order to increase yield and quality, as well as other agronomic traits. Different studies discussed in this review indicated examples of the application of the classical CRISPR-based genome editing. All details related to the modification of legume genomes are displayed in **Section 3** as a case study for each legume species. These studies may promote the fact that the modification of the complex nature of the legume genomes may also be defeated by these new base-editing and prime-editing CRISPR-based technologies. Utilization of the base-editing technology enables precise editing with high efficiency through both CBE and ABE systems. Various plant species including rice and tomato (Shimatani et al., 2017), rice, wheat, and maize (Zong et al., 2017), and wheat, rice, and potato (Zong et al., 2018) have been developed using different cytidine deaminase base-editing features. In recent studies, ABE systems have also been used in rice, wheat, *Arabidopsis thaliana,* and rapeseed (Kang et al., 2018; Li et al., 2018). In reviewing the literature, no data were found on the application of CBE and ABE systems for legume plants. Thus, various deaminase-mediated base-editing versions developed for model plant species may be useful for increasing the efficiency of the base editing in various legumes. Despite CBE and ABE systems having the ability to induct precise base transitions, their tools for base substitutions are restricted. Another highly promising technology known as "prime editor or prime editing" solves this problem by allowing the precise insertions of up to 44 bp, deletions of up to 80 bp, and combinations of these edits (Anzalone et al., 2019). The system has also been improved for plants and is able to perform multiple base substitutions, insertions, and deletions simultaneously in rice and wheat (Lin et al., 2020; Xu R. et al., 2020; Xu W. et al., 2020). Although different strategies and modifications such as the utilization of reverse transcriptase with different catalytic activities, the usage of ribozymes to obtain precise pegRNAs (prime-editing guide RNA), increasing culture temperature to raise catalytic activities, and modifications of the scaffold into pegRNA to augment the binding potential of Cas9 have been performed, there have been still limitations for the editing capacity of prime editor in plants (Zhu et al., 2020). Currently, the plant prime-editing technology has only been demonstrated in rice and wheat. Its performance still needs to be examined in a variety of plant species. Thus, plant prime-editing technology can be considered as an untouched deep blue cove for genome editing in legumes as well as in other significant agronomic plant species.

2.2 Transcription Activator-Like Effector Nucleases

The origin of transcription activator-like effector nucleases (TALENs) is quite extraordinary. Like CRISPR, TALEN is derived from a bacterium, and interestingly, this bacterium, called Xanthomonas, is quite pathogenic and responsible for serious diseases in various crops. During the infection, with the type III secretion system, *Xanthomonas* translocates transcription activator-like (TAL) proteins into the host cell cytoplasm. To enhance bacterial colonization during the infection, TALs act as host's-transcription factors and cause plant developmental changes that are beneficial for the disease. TALs mainly consist of three structures: the central domain of tandem repeats, transcriptional activation domain, and nuclear localization signals (Boch et al., 2009). Highly conserved repeat domains are mostly 33–35 amino acid lengths and are responsible for DNA binding. Specific target DNA to bind is decided by hypervariable residues that can be found at the 12th and 13th positions of the repeat domain. The pair of this position is named

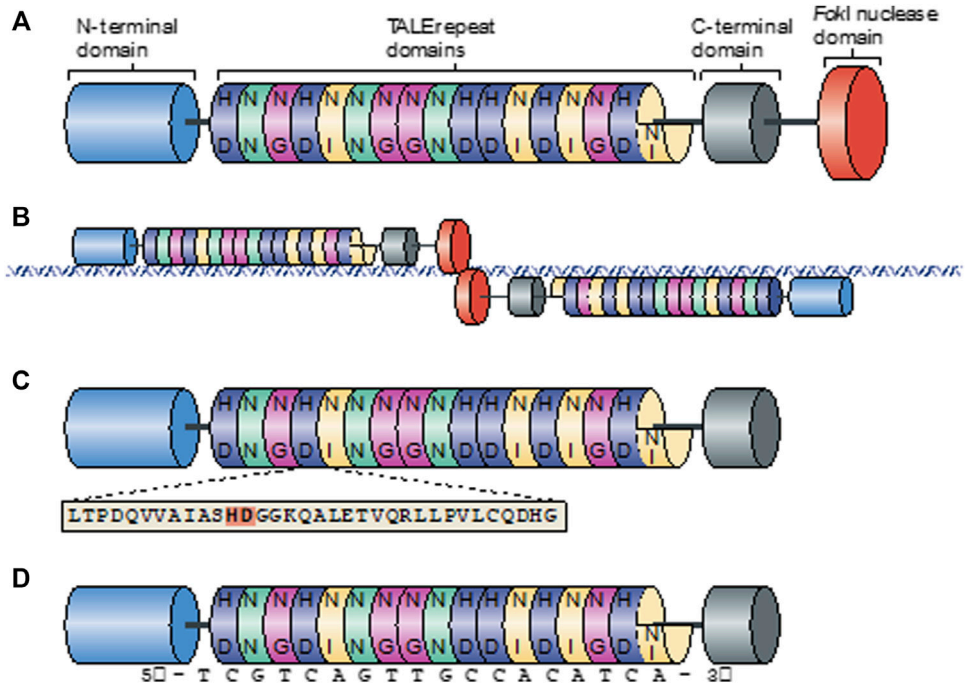


FIGURE 2 | Overview of the TALEN structure (Joung and Sander, 2013). **(A)** Schematic view of the TALEN structure. Colored discs with two letters inside represent the RVD. **(B)** Schematic view of TALEN pair binding to target site. Cleave of FokI occurs on the spacer domain on the target site. **(C)** Schematic diagram of the TALEN binding domain with an amino acid sequence. Two amino acids that represent RVD are shown in bold. **(D)** Amino acid sequence of a TALEN binding domain with a nucleotide representation of each RVD.

repeat-variable di-residue (RVD), and each RVD is associated with one of the four bases. On the other hand, in TALENs, *FokI* takes the place of the activator domain to become a target-specific genome-editing tool (**Figure 2**). To obtain double-strand breaks *via* FokI, TALENs are used in pairs. Each pair binds to the opposite strands of the target sequence and is separated with a spacer domain (Zhang X. et al., 2014). Before this groundbreaking development, *FokI* nuclease was placed along with the C-terminal activator domain on TALEN. Further studies have demonstrated that the truncation of large C-terminal sequences that are used to attend the *FokI* domain greatly increased the efficiency of TALEN (Bedell et al., 2012; Joung and Sander, 2013). These truncations were required due to the low efficiency of the custom TALEN application. Moreover, Bedell et al. (2012) aimed to increase the efficiency of TALEN by using differently truncated scaffolds. Their specifically designed GoldyTALEN scaffold, which is shorter due to the truncation of 215 amino acids from the pTAL scaffold, increased the success rate by up to sixfold, and some approaches showed 100% efficiency in zebrafish. All these results showed that design and construction have crucial roles and direct effects on the efficiency of the genome-editing process.

2.2.1 Vector Construction

Since the construction of TALEN cassettes depends on many factors, researchers have invested considerable time and effort in this step to simplify the construction and increase the efficiency of TALENs. One of the most common molecular methods for the plasmid construction of TALENs is the golden gate assembly. This method enables the simultaneous assembly of multiple DNA fragments to a single plasmid. Since type IIS restriction enzymes take part in this method, multiple insertions to a plasmid could be done "scarlessly" as this type of restriction enzymes cuts DNA outside of recognition sites. Cermak et al. (2011) showed that with two steps it is possible to construct a vector with an array of 12–31 RVDs. The first step is for the assembly of RVDs into arrays, up to 10 RVDs each array, and the second step is for inserting these repeat arrays into the plasmid backbone. On the other hand, the assembly of 10 RVDs together is quite challenging. To eliminate this challenge during the cloning, Hegazy and Youns (2016) modified this protocol. Instead of 10 RVDs, they constructed five RVD length arrays. Even though this modification increased the duration of the construction, it also increased the rate of the successfully constructed plasmids from an average of 4.4%–30%.

2.3 ZFNs

The early 1990s were a critical turning point in the genome-editing perspective of view. With a better understanding of DNA repair systems, one of the first precise genome-editing techniques was developed. The so-called ZFN technique was developed by merging an engineered zinc finger domain with a nuclease domain (ZFN). Similar to TALEN, ZFs are combined with FokI for DNA cleavage and they are used as pairs to obtain double-strand breaks. On the other hand, the zinc finger domain consists of up to six proteins and is responsible for binding to the DNA target point. Engineered Cys_2–His_2 residues in the structures of these ZF proteins are stabilized by Zn^{+2} ions, and

each one of these proteins interacts with three base pairs. For the construction, mainly three methods have been followed: 1) the modular assembly has been developed through the creation of identified ZF domain pool. During the application of ZFNs, the researchers can pick ZFs according to their target point from this pool and design their ZFN pair (Cathomen and Joung, 2008). 2) In context-sensitive selection strategies, the researchers have focused on developing new ZFN combinations of ZFs from customized libraries (Hurt et al., 2003; Durai et al., 2005). 3) The combination of two dual ZFs "(2 + 2 strategy)" from pre-existing libraries and optimization of these four ZFs according to the target loci with an algorithm. However, this method can only be used by researchers who have collaborated with Sangamo (Scott, 2005; Urnov et al., 2005).

Several advantages of CRISPR over ZFNs and TALENs have been pointed out in the literature, and the simplicity of the design has been highlighted as the main reason that made CRISPR the most widely utilized genome-editing tool. ZFN requires tremendous time and expertise during construction. Limited pairs have been identified so far, and optimization of zinc fingers is challenging. Especially protein engineering and the combination of new zinc fingers make this process impossible to perform in most of the laboratories. Although TALEN is much easier to construct when it is compared with ZFN, it is still far behind CRISPR. To construct a TALEN pair that targets a gene with 20-base pair length requires the design of 20 RVDs and the assembly of RVDs into a plasmid. These two steps make the process very challenging and again impossible to perform for researchers. Less than this effort, we can construct a CRISPR plasmid that targets 10 different genes in an organism. Even though it raises some questions, especially ethically inquiring questions, CRISPR became the most popular genome-editing tool due to its simplicity, efficiency, and multiplexed targeting potential. However, there are still some cases that make TALENs preferable. Due to the possibility of targeting longer DNA sequences, TALENs reduce the possibility of off-targets during the application. Researchers thus may choose TALENs over CRISPR to eliminate off-target mutations.

ZFNs, TALENs, and CRISPR-associated Cas9 endonucleases are the three major generations of genome-editing tools that have been mainly used for over a decade in plants including different legume species. The advantages and disadvantages of these methods have been discussed in many research articles (Gaj et al., 2013; Malzahn et al., 2017; Ahmar et al., 2020). Plant scientists face a major challenge in improving legume production and quality amidst changing climates and extreme environmental conditions. A promising option for achieving this goal is genome editing, and CRISPR/Cas technology is the most popular option because it is easy to use and convenient. However, no routine method has been proposed for the most efficient genome editing for legumes. The main reasons for this situation can be summarized as follows. The plantlet regeneration and genetic transformation of various legume species became a bottleneck for them. It is, therefore, necessary to develop an optimized protocol for the transformation and regeneration of legumes that is reproducible and reliable across species. In addition, there are some difficulties in CRISPR technology such as the on-target efficiency, off-target capacity, sgRNA design, and selection of proper Cas proteins. The next section summarizes all genetic modifications, including genetic transformation and genome editing, that have been done with legumes so far. At the end of the review, some suggestions are also presented to eliminate these problems.

3 APPLICATIONS IN LEGUMES

3.1 *M. truncatula* (Alfalfa)

M. truncatula is a model organism for legume crops because of its relatively easy transformation, short life cycle, self-fertility, diploidy nature, and small genome. Because of these features, it is widely studied in molecular and physiological research on legume crops. Michno et al. (2015) reported that they mutated soybean glutamine synthase (*GS1*) and chalcone-flavanone isomerase (*CHI20*) genes in *G. max* and β-glucuronidase (*GUS*) gene in *M. truncatula* by hairy root transformation (**Table 1**). Differently, Bottero et al. (2021) produced two transgenic alfalfa events (named as 3-1 and 5-1) by using CRISPR/Cas9 with the pBI121 binary vector containing the *GUS* gene and determined an average of 55% *GUS* inactivation. In the literature, many researchers prefer to study *phytoene desaturase* (*PDS*) genes because of its easy phenotypic observation to evaluate the success of an efficient CRISPR/Cas9 gene-editing tool. In 2017, Meng et al. (2017) developed an efficient CRISPR/Cas9 system for targeted *MtPDS* gene mutations in *M. truncatula,* and they observed that 32 of 309 T_0 transgenic plants exhibited the albino phenotype. Sequencing analysis of randomly selected 16 transgenic plants from this 32 showed that all these albino plants carry mutations at the targeted site of the *MtPDS* gene. In addition, Wolabu et al. (2020) showed that UBQ10 promoter-driven Cas9 provides high mutation efficiency (95% in Arabidopsis and 70% in *M. truncatula*). Zhang et al. (2020) also targeted *MtPDS* genes using the CRISPR/Cas9 system, and all regenerated seedlings derived from the homozygous/biallelic *MtPDS* mutant showed albino phenotypes.

M. truncatula forms indeterminate nodules, which are also found in pea (*Pisum sativum* L.), lentil (*Lens culinaris* Medik.), faba bean (*Vicia faba* L.), and chickpea (*Cicer arietinum* L.), that make it a good candidate plant to study nodulation in lentils (Bhowmik et al., 2021). In 2017, Curtin and colleagues used CRISPR/Cas9 nucleases, hairpin RNA interference constructs, and Tnt1 (the transposable element of *Nicotiana tabacum* cell type 1) retrotransposons together to evaluate the function of 10 candidate genes that exist in six clusters of strongly associated single nucleotide polymorphisms in *M. truncatula*. They found three candidate genes, ubiquitin conjugate24-like (*PHO-like*), Penetration3-like (*PEN3-like*), and partner of NOB1-like (*PNO1-like*), having statistically significant influences on nodule production.

In 2019, Trujillo et al. (2019) identified nodule-specific polycystin-1, lipoxygenase, and alpha toxin nodule-specific (PLAT) domain proteins (NPDs) and examined the *NPD* function with its knockout lines *via* CRISPR/Cas9. They created different combinations of *NPD* gene inactivations and

TABLE 1 | Gene-editing technology in different legume crops.

Legume	Technique	Target (gene and function)	Result	References
Medicago truncatula	CRISPR/Cas 9	*MtSUP* (regulates the floral organ number)	MtSUP was found to be orthologous of *AtSUP*	Rodas et al. (2021)
	CRISPR/Cas 9	*CYP93E2* and *CYP72A61* (soyasapogenol B biosynthesis)	51 *CYP93E2* mutant plant lines	Confalonieri et al. (2021)
	CRISPR/Cas 9	*MtPDS* (coding phytoene dehydrogenase/chromoplastic protein)	70% mutation efficiency	Wolabu et al. (2020)
		MtPDS (coding phytoene dehydrogenase/chromoplastic protein)	Homozygous and biallelic mutants	Zhang et al. (2020)
	CRISPR/Cas 9	*NPD* genes (nodulation)	Smaller nodules, earlier onset of nodule senescence, ineffective nodules	Trujillo et al. (2019)
	CRISPR/Cas 9	*MtHen1* (*Hua enhancer1* gene)	Efficient mutation	Curtin et al. (2018)
	CRISPR/Cas 9	*FMO1-like*, *RFP1-like*, *ERDJ2*, *MEL1*, *PEN3-like*, *ACRE1*, *HLZ1-like*, *PHO2-like*, *PNO1-like*, *FBL1-like* (root and nodules)	Statistically significant effects on nodule production	Curtin et al. (2017)
	CRISPR/Cas 9	*MtPDS* (coding phytoene dehydrogenase/chromoplastic protein)	Albino plants	Meng et al. (2017)
	CRISPR/Cas 9	*GmGS1*, *GmCHI20*, *MtGUS*	Mutated genes	Michno et al. (2015)
Lotus japonicus	CRISPR/Cas 9	*Lbs* genes (nodule senescence)	Early nodule senescence	Wang et al. (2019)
	CRISPR/Cas 9	*CYP716A51* (triterpenoid production)	Non-production of triterpenoids	Suzuki et al. (2019)
	CRISPR/Cas 9	*LjCZF1* and *LjCZF2* (root nodule symbiosis)	Decrease in nodule formation	Cai, K., et al. (2018)
	CRISPR/Cas 9	*SNF* (symbiotic nitrogen fixation) genes	CRISPR/Cas9 system can effectively induce mutations in SNF related genes	Wang et al. (2016)
Glycine max (Soybean)	CRISPR/Cas 9	*GmPRR37* (photoperiodic flowering)	Changes in flowering time	Wang, L., et al. (2020)
	CRISPR/Cas 9	*GmLox1*, *GmLox2*, *GmLox3* (encoding lipoxygenases)	Loss of lipoxygenase activity	Wang, J., et al. (2020)
	CRISPR/Cas 9	*GmAGO7a* and *GmAGO7b* (controlling leaf pattern)	Inherited mutation until T_2 lines	Zheng et al. (2020)
Glycine max (Soybean)	CRISPR/Cas 9	*GmFT2a* and *GmFT5a* (flowering time)	ft2a, ft5a, and ft2aft5a mutants	Cai et al. (2020)
	CRISPR/Cas 9	Pooled platform-102 candidate genes	Multiplex mutations	Bai et al. (2020)
	CRISPR/Cas 9	*KAS1* (conversion of sucrose to oil)	Deletion and insertion mutations	Virdi et al. (2020)
	CRISPR/Cas 9	*GmFAD2-1A*, *GmFAD2-2A* (biosynthesis of peakoil)	Increased oleic acid content	Wu et al. (2020)
	CRISPR/Cas 9	*GmSPL9* (plant architecture)	Altered plant architecture	Bao et al. (2019)
	CRISPR/Cas 9	*FAD2-2* (seed content improvement)	Increased oleic acid content	Al Amin et al. (2019)
	CRISPR/Cas 9	*Glyma03g36470* (eukaryotic translation initiation factor)	Insertion and deletion mutations	Di et al. (2019)
	CRISPR/Cas 9	*Glyma14g04180* (late-embryogenesis abundant protein)	Insertion and deletion mutations	Di et al. (2019)
	CRISPR/Cas 9	*Glyma06g136900* (uncharacterized protein)	Insertion and deletion mutations	Di et al. (2019)
	CRISPR/Cas 9	*GmFAD2-1A*, *GmFAD2-1B* (biosynthesis of peakoil)	Increased oleic acid content	Do et al. (2019)
	CRISPR/Cas 9	*Conglycinin* (*7S*) and *glycinin* (*11S*) (storage proteins)	Mutations in three of nine genes	Li et al. (2019)
	CRISPR/Cas 9	*E1* (flowering time)	Early flowering	Han et al. (2019)
	CRISPR/Cas 9	*GmDrb2a* and *GmDrb2b* (double-stranded RNA-binding2)	Biallelic double mutant	Curtin et al. (2018)
	TALEN	*Glycine max Dicer-like2* (Dicer-like protein)	*GmDicer-like2* mutant plants	Curtin et al. (2018)
	CRISPR/Cas 9	*GmFT2a* and *GmFT5a* (flowering time)	Deletion mutations	Cai, Y., et al. (2018)
	CRISPR/Cas 9 TALEN	*GmPDS11* and *GmPDS18* (coding phytoene dehydrogenase/chromoplastic protein)	Albino and dwarf buds	Du et al. (2016)
	CRISPR/Cas 9	*DD20* and *DD43* (two genomic sites on chromosome 4)	Mutations	Li et al. (2015)
	CRISPR/Cas 9	*Glyma06g14180* (uncharacterized protein)	Mutations	Sun et al. (2015)
	CRISPR/Cas 9	*Glyma08g02290* (potassium transporter)	Mutations	Sun et al. (2015)
	CRISPR/Cas 9	*Glyma12g37050* (ethylene receptor)	Mutations	Sun et al. (2015)
	CRISPR/Cas 9	*GmFEI2* (LRR receptor-like serine/threonine-protein kinase FEI 2)	Mutations	Cai et al. (2015)
	CRISPR/Cas 9	*GmSHR* (short root protein)	Mutations	Cai et al. (2015)
	TALEN	*FAD2-1A/B* (seed content improvement)	Increased oleic acid content, reduced linolenic acid content	Haun et al. (2014)
	TALEN	*FAD3A* (seed content improvement)	Reduced linolenic acid content	Demorest et al. (2016)

(Continued on following page)

TABLE 1 | (*Continued*) Gene-editing technology in different legume crops.

Legume	Technique	Target (gene and function)	Result	References
	TALEN	*GmDcl2b* (generating heritable mutations)	Combinatorial mutant plants	Curtin et al. (2017)
	ZFN	*RDR6a and RDR6b* (process optimization)	2 bp differences on target genes	Curtin et al. (2011)
	ZFN	*DCL4a/b* (Dicer-like protein)	Defective miRNA precursor transcript processing	Curtin et al., 2011, 2015
	ZFN	*DCLb* (Dicer-like protein)	Increased lateral root growth	Curtin et al. (2011)
Cowpea (*Vigna unguiculata*)	CRISPR/Cas 9	*VuSPO11-1* (cowpea meiosis gene)	Mutations	Che et al. (2021)
	CRISPR/Cas 9	*SPO11-1*, *REC8* and *OSD1* (meiosis genes)	Male and female sterilities	Juranic et al. (2020)
	CRISPR/Cas 9	*SNF* (symbiotic nitrogen fixation) genes	Blocked nodule formation	Ji et al. (2019)
Chickpea (*Cicer arietinum*)	CRISPR/Cas 9	*4CL* (*4-coumarate ligase*) *RVE7* (*Reveille 7*) (drought tolerance)	High efficiency in editing	Badhan et al. (2021)
Peanut (*Arachis hypogaea*)	CRISPR/Cas 9	*AhNFR1* and *AhNFR5* (nodulation)	Successfully edited genes	Shu et al. (2020)
	CRISPR/Cas 9	*AhFAD2* (seed content improvement)	G448A, 441_442insA, G451T mutations	Yuan et al. (2019)
	TALEN	*AhFAD2* (seed content improvement)	Increase in the oleic acid content	Wen et al. (2018)

observed that mutant lines showed an earlier onset of nodule senescence and smaller or ineffective nodules in comparison to the wild-type control.

CRISPR/Cas9 gene-editing tool has also been used in studies focused on flowering or secondary metabolite production. Rodas et al. (2021) edited the *M. truncatula SUPERMAN* (*MtSUP*) gene with CRISPR/Cas 9 and determined the impairment of *MtSUP* function with observing defects in floral development and inflorescence architecture in *mtsup* mutant allel carrying plants. Confalonieri et al. (2021) also used the CRISPR/Cas9 gene-editing tool to knock out the two cytochrome P_{450} genes (*CYP93E2* and *CYP72A61*) that are responsible for soyasapogenol B production in *Medicago* spp. Their results showed that 51 putative *CYP93E2* mutant plant lines with an 84% editing efficiency did not produce soyasapogenols in the leaves, stems, and roots with diverting the metabolic pathway toward the production of valuable hemolytic sapogenins.

3.2 *L. japonicus* (Lotus)

L. japonicus is also a model organism for legume crops with similar features to *M. truncatula,* but conversely to it, *L. japonicus* organizes determinate nodules, like soybean [*G. max* (L.) Merr.] and cowpea (*Vigna unguiculata* L.). Wang and colleagues (2016) studied *L. japonicus* and proved that SNF (symbiotic nitrogen fixation)-related gene mutations can be performed by CRISPR/Cas9 with hairy root transformation. In 2018, Cai K. et al. (2018) edited cytokinin receptor Lotus histidine kinaz I-interacting protein (*LjCZF1*) to reveal the mechanism of cytokinin signaling regulation of rhizobia-legume symbiosis. They determined that the knock-out mutant lines had a significantly reduced number of infection threads and nodules, supporting that LjCZF1 is a positive regulator of symbiotic nodulation. Later, Wang et al. (2019) used CRISPR/Cas9 technology to understand the role of leghemoglobin (Lbs) in *L. japonicus* and they observed that the lack of Lbs resulted in an early nodule senescence. In another study, the gene loss-of-function analysis of *CYP716A51* (which shows triterpenoid C-28 oxidation activity) was performed and the results showed that cyp716a51-mutant *L. japonicus* hairy roots did not produce C28 oxidized triterpenoids.

3.3 *G. max* (Soybean)

There is an ever-increasing need for soybean products since soybean has an important economic value with its rich protein and oil for animal and human nutrition. For this reason, genetic development with gene-editing tools must be accelerated in order to meet this increasing need and cope with the changing environmental conditions (Bao et al., 2020). The initial optimization approach was done by Curtin et al. (2011) by targeting the green fluorescent protein (GFP) coding region in soybean with a ZFN array that was developed *via* context-sensitive selection strategies. This approach resulted in up to 71 base pair deletions on the target. With the optimization of the process, they targeted two different RNA-dependent polymerases in soybean. The most interesting outcome of this study was that two independent ZFN pairs were designed, both recognizing their specific targets and causing two base pair differences in both genes. Another study by the same group (Curtin et al., 2015) was focused on disturbing miRNA maturation and miRNA gene expression regulation in 2015. For this purpose, they designed two different ZFN pairs to target *Dicer-like 1a* (*DCL1a*) and *Dicer-like 1b* (*DCL1b*) genes in soybean. While single mutants of *DCL* genes did not give any remarkable result, double *DCL* mutants expressed remarkable morphological outcomes; additionally defective miRNA precursor transcript processing efficiency and deregulated miRNA target gene expression were observed. In addition, Curtin et al. (2018) also used TALENs within the *G. max Dicer-like2* gene. They revealed multiple transgene insertion events by whole-genome sequencing and generated a suite of combinatorial mutant plants.

In the case of CRISPR-Cas, Cai et al. (2015) targeted different sites of two endogenous soybean genes (*GmFEI2* and *GmSHR*). For this, they designed an sgRNA that targeted a transgene (bar) and six sgRNAs resulted in targeted DNA mutations in hairy roots. Li and co-workers (2015) successfully applied CRISPR/

Cas9 and mutated two genomic sites DD20 and DD43 on chromosome 4 with 59% and 76%, respectively, success rates. Moreover, Sun and colleagues (2015) also constructed two vectors using the Arabidopsis U6-26 and soybean U6-10 promoters and targeted *Glyma06g14180*, *Glyma08g02290,* and *Glyma12g37050* in protoplast efficiently. In addition, *Glyma06g14180* and *Glyma08g02290* biallelic mutations were also observed in transgenic hairy roots. Later, Du et al. (2016) presented a comparative analysis of CRISPR/Cas9 and TAL-effector nuclease (TALEN) gene-editing technologies for two soybean *GmPDS11* and *GmPDS18* genes and they observed albino and dwarf buds (PDS knock-out) with the transformation of cotyledon nodes. The mutation efficiency of TALENs was slightly higher than the Cas9/sgRNA system using the AtU6-26 promoter but much lower when using the soybean GmU6-16g-1 promoter in hairy roots. According to the results, they declared that both gene-editing technologies can achieve gene targeting in soybean. In addition to this study, Curtin and friends (2018) also performed CRISPR/Cas9 and TALENs at the same time in *G. max* and *M. truncatula* and created a bi-allelic double mutant for the two soybean paralogous *double-stranded RNA-binding2* (*GmDrb2a* and *GmDrb2b*) genes and a mutation of the *M. truncatula Hua enhancer1* (*MtHen1*) gene.

Soybean flowering time is important due to its effect on increasing breeding speed for yield and improving quality. Because of this reason, many studies also focused on editing flowering time-related genes. Cai Y. et al. (2018) developed an efficient system using a dual-sgRNA/Cas9 to target deletions in *GmFT2a* and *GmFT5a* genes. Their results showed 15.6% and 15.8% deletion frequencies for target fragments in *GmFT2a* and *GmFT5a*, respectively. They also detected 12.1% exceeding 4.5 kb in *GmFT2a*. In addition, they determined that these deletions can be inherited in T_2 "transgene-free" homozygous ft2a mutants that exhibited the late-flowering phenotype. In another study (Han et al., 2019), soybean maturity gene *E1*, which controls soybean flowering, was edited and 11 bp and 40 bp deletions at the *E1* coding region were generated. These deletions lead to premature translation termination codons and truncated E1 proteins. In addition, Wang L. et al. (2020) created knock-out and overexpression mutations with CRISPR/Cas9 tool in soybean *Pseudo-response regulator* gene (*GmPRR37*), encoding qFT12-2 (flowering time) protein and they demonstrated that *GmPRR37* controls soybean photoperiodic flowering. Cai et al. (2020) also studied *GmFT2a* and *GmFT5a* genes with CRISPR/Cas9 in soybean and showed that these genes collectively regulate flowering time by analyzing the overexpression of ft2a, ft5a, and ft2a/ft5a mutants under short-day SD and long-day conditions.

CRISPR/Cas9 strategy also used to target three GmLox genes (GmLox1, GmLox2, and GmLox3) encoding three lipoxygenases (LOX1, LOX2, and LOX3), which induce a beany flavor that restricts human consumption (Wang J. et al., 2020). They determined that 60 T_0 positive transgenic plants, carrying combinations of sgRNAs and mutations (two of them triple mutant and one of them is a double mutant), had lost the corresponding lipoxygenase activities. Differently, Li et al. (2019) used the CRISPR/Cas9 system in editing *conglycinin* (7S) and glycinin (11S) storage protein genes in soybean and detected 5.8%, 3.8%, and 43.7% gene-editing efficiencies for *Glyma.20g148400*, *Glyma.03g163500,* and *Glyma.19g164900* genes, respectively. Besides, plant architecture is also altered by the application of CRISPR/Cas9 system in soybean. Bao et al. (2019) targeted squamosa promoter binding protein-like genes (*GmSPL9a*, *GmSPL9b*, *GmSPL9c,* and *GmSPL9*) and determined that T2 double homozygous mutant spl9a/spl9b has a shorter plastochron length. In addition, the increased node number on the main stem and branch number is observed in T4 mutant plants.

The cultivation of soybean varieties with higher oleic acid content becomes a major goal in breeding (Wu et al., 2020). In accordance, gene-editing technologies gained an increasingly important role in soybean studies. Although TALENs have not been widely used in legumes, there are a few successful TALEN applications, particularly to increase the oleic acid content and functional studies. Since the oleic acid content is dependent on the activity of *Fatty Acid Desaturase 2* genes, which are the key enzymes for converting oleic acid to linoleic acid that oxidizes readily, most of the studies were focused on introducing mutations at these genes. For instance, Haun et al. (2014) focused on increasing the soybean oleic acid content by targeting *FAD2-1A* and *FAD2-1B* genes. For targeting these two genes, four pairs of TALENs were designed. Only FAD2_T01 and FAD2_T04 were expressed by plants. The mutation rate of FAD2_T04 at both genes was 7.2%; on the other hand, the efficiency of FAD2_T01 was even lower than FAD2_T04, 3.1% at *FAD2-1A,* and 1% at *FAD2-1B*. A decrease in linoleic acid (down to 4%) together with an increase in oleic acid content (up to 80%) was obtained. A similar study was performed by Demorest et al. (2016), targeting *FAD2-1A*, *FAD2-1B,* and *FAD3A* genes. FAD3A_T1, FAD3A_T2, and FAD3A_T3 TALENs were designed to target the *FAD3A* gene of the FAD2-1A and FAD2-1B mutated lines, and they showed 11.2%, 16.0%, and 4.9% mutation rates, respectively. With these mutations, more than 80% increase in oleic acid and a reduction in decreased linoleic acid (2%) were obtained. Moreover, Do et al. (2019) targeted *GmFAD2-1A* and *GmFAD2-1B* genes and created T_0 transgenic plants. The fatty acid profile analysis showed an 80% increase in the oleic acid content, whereas 1.3%–1.7% decrease in linoleic acid in T_1 seeds homozygous for both *GmFAD2* genes. Similarly, Al Amin et al. (2019) applied the CRISPR-Cas9 system for the mutation of the *FAD2-2* gene in soybean and observed an important level change in oleic acid/linoleic acid ratios caused by high-frequency deletions and insertions in the gene. In 2020, Wu and co-workers also used CRISPR/Cas9 in *GmFAD2-1A* and *GmFAD2-2A* genes to create single and double knock-out mutants and showed that their function was slightly changed. Their editing efficiency was determined as 95% for *GmFAD2-1A*, 55.56% for *GmFAD2-2A,* and 66.67% for double mutants. They also determined that the oleic acid content increased up to 73.50%, while the linoleic acid content decreased down to 12.23% in the T_2 generation. In addition, these contents showed similar level changes in the T_3 generation.

Di et al. (2019) enhanced the CRISPR/Cas9 system by using highly active 5 U6 promoters by targeting *Glyma03g36470*, *Glyma14g04180,* and *Glyma06g136900* genes. Results showed that nucleotide insertion, deletion, and substitution mutations occurred. The following year, Bai et al. (2020) constructed 70 CRISPR-Cas9 vectors to target 102 candidate genes and they obtained 407 T_0 mutant lines containing all sgRNAs with 59.2% mutagenesis frequency. In addition to this, 35.6% of lines carried multiplex mutations. As a result, increased nodule numbers in gmric1/gmric2 double mutants and decreased nodulation in gmrdn1-1/1-2/1-3 triple mutant lines were observed.

Zheng et al. (2020) presented easy-to-use binary vector systems with Cas9 driven by egg cell-specific promoters (ECp). They targeted two genes, *GmAGO7a* and *GmAGO7b*, coding ARGONAUTE7 (AGO7), which are key regulators in controlling leaf patterns in soybean. Their results showed that these promoters can induce mutations and multiple, independent mutations can be obtained. In another study, Virdi et al. (2020) generated multiple knockout alleles and also one in-frame allele for the *β-ketoacyl synthase 1* (*KASI*) gene, which has a role in changing sucrose to oil, by using CRISPR/Cas9 mutagenesis and their results indicated that genes lost their function.

Due to the importance of soybean, relatively more CRISPR studies including the modification of its nutrition value and plant architecture (leaf patterns and nodule numbers) were carried out in soybean among other legumes. However, the stable soybean genetic transformation has not yet been established since the soybean is still a recalcitrant crop to transformation. With the improvement of the transformation efficiency, CRISPR studies could advance future genetic studies in soybean with its efficiency, multiplex editing, and high-throughput mutagenesis capability (Bao et al., 2020).

3.4 *V. unguiculata* (Cowpea)

Cowpea (*V. unguiculata* (L.) Walp.) is a legume crop that has a high nutrition content and health benefits. It has an efficient symbiotic nitrogen fixation (SNF) capability, tolerance to low rainfall, and low fertilization requirements. Due to these agronomically important traits, it became one of the most important legumes worldwide (Ji et al., 2019; Che et al., 2021). For these reasons, interest in gene-editing approaches in cowpea is increasing. In 2019, Ji and colleagues demonstrated CRISPR/Cas9-mediated genome editing in cowpea in non-inheritable mutated hairy roots by targeting SNF genes. They observed that nodule formation was completely blocked in the mutants with both alleles disrupted. Following them, Juranic et al. (2020) identified three cowpea meiosis genes; *SPO11-1* (encodes SPO11 protein, which is the initiator of meiotic double-stranded breaks), *REC8* (encodes meiotic recombination protein), and *OSD1* (encodes *Ophiostoma scytalone* dehydratase protein that promotes meiotic progression) used CRISPR/Cas9 gene editing to induce asexual seed formation in cowpea. They determined biallelic mutations in exon 1 and exon 3 of the *SPO11-1* gene resulting in defects in meiosis leading to complete male and female sterilities in the T_0 plants. Recently, Che et al. (2021) also knocked out the cowpea meiosis gene *VuSPO11-1* by using CRISPR/Cas9 and observed mutations at the target.

3.5 *C. arietinum* (Chickpea)

Chickpea (*C. arietinum*) is a commercially important crop worldwide, and gene-editing tools can be used to eliminate the problems in its production. Badhan et al. (2021) performed a study that targeted drought tolerance-associated genes, *4-coumarate ligase* (*4CL*) and *Reveille 7* (*RVE7*), for CRISPR/Cas9 editing in chickpea protoplast and the knock-out of the *RVE7* gene showed high-efficiency editing *in vivo*. These results showed that CRISPR/Cas9 DNA-free gene editing can be used for genes associated with drought tolerance in chickpea by utilizing protoplast. To date, this was the first and only study that used CRISPR/Cas9 gene editing in chickpea.

3.6 *A. hypogaea* (Peanut)

Peanut (*A. hypogaea*) is an important legume crop with a high oleic acid content. A high oleate spontaneous mutant line (F435), which contains 80% oleic acid, has previously been identified by plant breeders *via* a peanut germplasm screening project (Norden et al., 1987). In this line, two types of mutations were reported in the *ahFAD2A* [a "G" to "A" substitution at 448 bp after the start codon (G448A) in the ahFAD2A] and *ahFAD2B* [an "A" insertion between bp 441 and 442 (441_442insA) in the ahFAD2B] genes (Lopez et al., 2000). Yuan et al. (2019) targeted these mutations and in addition, they observed a new mutation, G451T, in *ahFAD2B*. These results suggested that the mutations induced in *ahFAD2B* by CRISPR/Cas9 may be useful in developing high oleate lines. Moreover, TALENs are also used to create targeted *ahFAD2* genes in peanut to increase the oleic acid content (Wen et al., 2018). Two TALEN pairs were constructed, one of them was used to inoculate 216 regenerated roots and the second one was used to inoculate 105 regenerated roots. Observed mutation frequencies were 8.33% and 12.38%, respectively. Most of the mutations occurred as small deletions of 1–10 bps. In the mutant lines, the oleic acid content of seeds was determined as 80.45%, which means a 2-fold increase when compared to wild-type plants. On the other hand, the linoleic acid content was decreased down to 3% in the mutant lines and there was no change in the total fatty acid amount.

Nod factor receptors (NFRs) that initiate a symbiotic relationship with rhizobia also edited with CRISPR/Cas9 to reveal out their functions in peanut nodulation (Shu et al., 2020). The edited mutants with two *AhNFR5* genes showed Nod-phenotype, while two selected *AhNFR1* genes containing mutants still could form nodules after inoculation.

3.7 Non-edited Legume Species

Lentil (*L. culinaris*) is a diploid and a self-pollinated plant in the Fabaceae family, containing rich proteins, minerals, fibers, and carbohydrate sources. Among developing countries and those whose diets are not based on expensive animal protein, it can contribute to the prevention of malnutrition and deficiencies in micronutrients (Kumar et al., 2015). Besides, as lentils add nitrogen to the soil, the quality of the soil improves (Kumar

et al., 2021). A draft lentil whole genome sequence is available at https://knowpulse.usask.ca/lentil-genome, which contains bulk sequencing, gene prediction, and annotation of the assembled 2.6 Gbp of the genome (Bett et al., 2014; Bett et al., 2016). The available draft genome sequences can facilitate the sequence-based targeted candidate genes related to nutrient value, abiotic and biotic stress responses, herbicide resistance, etc. Genomic tools and technologies can help to improve lentil breeding. Some gene transformation efforts are available for lentils. As a successful example of those studies, the *dehydration-responsive element binding* gene (*DREB1A*), which is involved in plant responses to abiotic stresses, was transferred to lentils by *Agrobacterium*, resulting in drought- and salinity-resistant transgenic plants (Khatib et al., 2011). *In vitro* regeneration after transformation is another important issue. It was reported that using decapitated embryos than other tissues was more effective in the generation of shoots (Sarker et al., 2003). Those attempts supply useful tools for new genome-editing research. Many candidate genes related to abiotic, biotic stress factors, and agronomic features have been determined in lentils (Kumar et al., 2021). However, up to date, there is no new genome-editing research studying candidate genes in lentils using ZFNs, TALENs, or CRISPR/Cas9 technologies. Genome-editing technology can be an easy and cheap way to discover the function of those candidate genes to provide cultivars with desired features including stress tolerance or agronomic traits (Bhowmik et al., 2021).

Pea (*P. sativum*) is an important legume crop in the world after the common bean (*Phaseolus vulgaris*) and has rich components including dietary proteins, mineral nutrients, complex starch, and fibers (Bastianelli et al., 1998). Pea's symbiotic nitrogen-fixing ability makes it a valuable crop for the development of systems that improve soil fertility (Mabrouk et al., 2018). In addition, the pea is an original model organism used by Mendel to construct the rules of inheritance (Ellis et al., 2011). The pea genome size is approximately 4.45 Gb. A reference genome was published in 2019, which provides insights into legume genome evolution (Kreplak et al., 2019). The genomic approach has an essential role in determining genes for critical features and developing genomic tools for crop improvement. Although significant progress has been made in pea planting, improving crop yield and quality, crop development must continue to feed the growing world population.

Peas are affected by parasitic weeds, viruses, bacteria, and fungi as much as abiotic stress factors including drought, salinity, heat, and cold stresses, which result in the loss of yield and growth. A stable transformation study was successfully applied in pea transferring *cry1Ac* gene (encoding protoxin) from *Bacillus thuringiensis* (Negawo et al., 2013) and alpha-amylase inhibitor gene from *P. vulgaris* (Schroeder et al., 1995) for insect tolerance *via Agrobacterium*-mediated transformation method. Another transformation attempt was applied for transferring antifungal genes against *Fusarium* spp. to pea, which resulted in a lack of stable expression in 3 years of field trials (Kahlon et al., 2017). *Agrobacterium*-mediated transformation efficiency and regeneration frequency were enhanced from seed-to-seed regeneration using longer infection time and adding zeatin to the selection medium in a recent study (Aftabi et al., 2017). The available genome sequence information, successful transformation, and regeneration applications are factors that affect the usage of genome-editing tools in vegetable crops (Cardi et al., 2017). To date, no studies were conducted using new genome-editing tools in pea, which may be attributed to the insufficient regeneration (Pandey et al., 2021). The development of new genome-editing methodologies can provide new opportunities in breeding to increase yield and produce plants with high nutritional value.

Common bean (*P. vulgaris*) is the most used up grain legume around the world, which has rich nutritious elements including proteins, vitamins, and minerals (Cichy et al., 2009). The common bean whole genome sequence has been available since 2014. In addition, researchers determined genes related to improved leaf and seed mass in common bean (Schmutz et al., 2014). Although successful gene transformation studies are limited in common bean, many indirect and direct gene transfer attempts have been done (reviewed by Nadeem et al. (2021)). One of those attempts involved the construction of transgenic lines that display tolerance to golden mosaic geminivirus (BgMV-BR) *via* transferring *Rep-TrAP-REn* and *BC1* genes to common bean by the biolistic method (Aragão et al., 1998). Xue et al. (2017) transferred the *PvPOX1* gene from a *Fusarium* wilt-resistant genotype to a *Fusarium* wilt-susceptible genotype by *Agrobacterium rhizogenes*. Moreover, variable protocols were applied to improve the regeneration performance of the plant as other legumes (Nadeem et al., 2021). Accordingly, Barraza et al. (2015) reported that the manipulation of *PvTRX1h* gene, which is the ortholog of a plant histone lysine methyltransferase involved in plant hormone synthesis, can help to overcome recalcitrant regeneration problems in the plant as it regulates somatic embryogenesis in common bean callus. Researchers have also tried to improve the tolerance of common bean to major biotic diseases (white mold, bacterial blight, rust, halo blight, anthracnose, and pests) and abiotic stresses. Besides, some other features in the common beans including higher content of minerals (iron and zinc), fast cooking time, canning quality, harvest index, and market class/seed color are significant breeding preferences (Assefa et al., 2019). Although the availability of the whole genome information, CRISPR/Cas, or other genome-editing tools has not been utilized up to now in common bean genetic research, gene-editing technologies can be applied to common bean research to obtain stress-tolerant plants and to meet common bean breeding priorities.

Faba beans have growing advantages over other legumes in cold temperatures; therefore, they are suitable for sustainable farming applications (Temesgen et al., 2015). Furthermore, like other legumes, faba beans have valuable systems to raise soil fertility. Faba bean breeding maintains the need for food and feed, which has a valuable source of protein, fiber, and other nutrients (Khazaei et al., 2021). Publicly available genome sequence data of faba bean are not available, which can be the result of the hardness of assembling the huge genome (Cooper et al., 2017). Faba bean is one of the legume species that transformation and regeneration efficiency are mostly hard although the availability

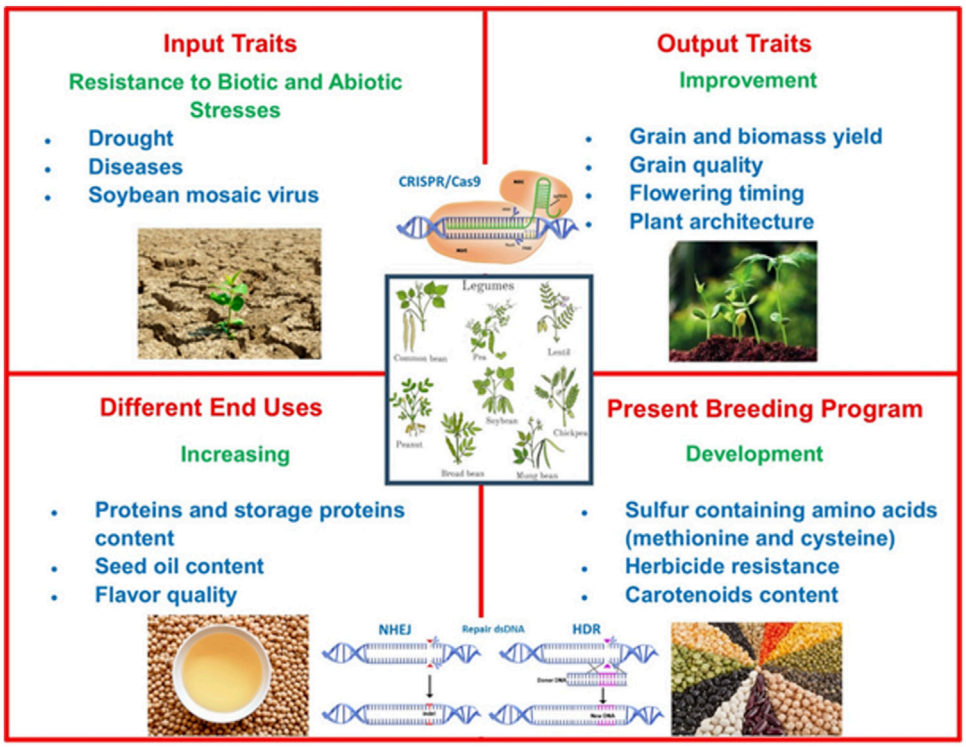


FIGURE 3 | Strategies for improving legume breeding.

of the attempts obtained by *Agrobacterium*-mediated transformation with low transformation efficiency (Cardi et al., 2017; Maalouf et al., 2019). One of these attempts was firstly made by Böttinger et al. (2001) who used *in vitro* development of internode stem segments invaded by *Agrobacterium* and by Hanafy et al. (2005) who cut out embryo axes infiltrated by *Agrobacterium* to obtain stable transgenic lines. Moreover, abiotic stress-tolerant lines were reported by Hanafy et al., 2013 by transforming the potato *PR10a* gene to faba bean by the same transformation method. Many biotic and abiotic stress factors including heat, insects, viruses, and parasitic weeds cause decreased faba bean yields (Cernay et al., 2015; Maalouf et al., 2019). TALEN technology, which is one of the genome-editing tools, has been applied to construct disease tolerance in plants. It was achieved by upregulating resistance gene expression *via* an engineered promoter site, which can bind multiple TALL effectors (Romer et al., 2009) or by changed TALL binding regions of promoters in sugar transporter genes, which are targeted by pathogens to stimulate tolerance in rice (Li et al., 2012; Xiao, 2012). Another advantageous method, namely CRISPR/Cas9, can be used to gain plant virus resistance by destructing the viral genome or by modifying the plant genome for resistance (Lenka et al., 2020). Moreover, vicine and convicine limiting faba bean consumption were defined as pyrimidine glycosides in cotyledons of faba bean, which are toxic effects for humans with a mutation in the enzyme of glucose-6-phosphate dehydrogenase (Luzzatto and Arese, 2018; Khazaei et al., 2021). Those compound levels can be controlled using new genome-editing tools in faba bean.

Mung bean (*Vigna radiata*) is a rapidly growing legume in warm climates in the Fabaceae family. Mung bean has rich dietary protein, folate, and iron in the seeds compared to other legume species (Keatinge et al., 2011). Also, mung bean plants can keep atmospheric nitrogen, which provides improved soil fertility (Maalouf et al., 2019). The whole genome sequence of *V. radiata* var. *sublobata* published in 2014 has enabled genomic research and molecular breeding of mung bean (Kang et al., 2014). However, limited mungbean germplasm and incompatibility with wild relative species affect mung bean breeding (Aasim et al., 2019). In addition, biotic and abiotic stresses also reduce mung bean production. Therefore, progress in transformation technologies supports researchers to develop lines that can cope with abiotic and biotic stress factors. Although genetic transformation efficiency was low in mung bean, many studies *via Agrobacterium*-mediated transformation are available for different agronomic features such as insecticidal (Saini et al., 2007), oxidative stress tolerance (Yadav et al., 2012), and salt stress tolerance (Sahoo et al., 2016). In those studies, the selection of the explants, transformation vectors, and selective agents are the significant factors that affect the success of the transformation efficiency (reviewed by Tripathi et al. (2021)). Although there are no studies using new genome-editing tools in mung bean, the availability of the whole genome sequence and improvement of the *in vitro* regeneration and transformation procedures can

enable the usage of those more sophisticated genome-editing tools to obtain crops with desirable features.

The availability of whole-genome data on the most common non-edited legume species including lentil, pea, common bean, and mung bean, which have abundant nutrients for the human diet worldwide, may allow genome-editing approaches to be developed. Identification and demonstration of the functions of genes related to abiotic and biotic stress tolerance, yield and quality, etc., represent a broad application area for new genome-editing tools to improve those characteristics in the important legume crops that are valuable for human nutrition. Furthermore, legumes are indispensable to soil fertility, so symbiotic nitrogen-fixing pathways could prove useful in genome-editing applications that aim to improve soil quality. Although there have been some promising attempts, the low transformation efficiency and the recalcitrant regeneration problems in those legumes appear to be the most significant limitations, which may be one of the common reasons for the lack of new genome-editing attempts. In order to realize the substantial improvement promised by new genome-editing techniques in those legumes with serious potential, new approaches to improve transformation and regeneration efficiency are critical for the adequate feeding of a large part of the world.

4 FUTURE PROSPECTS

4.1 Future of Genome-Editing Technologies in Legumes

Crop breeding and functional genomics have progressed rapidly with genome-editing research including the Nobel prize awarded system, CRISPR. Recent advancements in genome-editing research have increased the accuracy and efficiency of modifying genes by adding or removing the genetic material. Besides, genome-editing technologies have been performed to facilitate the manipulation of single or multiple genetic loci in different plants. The recently sequenced variable legume genomes are a valuable source of information for researchers with better applications of gene-editing tools. Therefore, a thorough understanding of legume genomic sequences and their functions is essential for efficient genome editing. It is possible to develop new legume varieties with the identification of genes that control certain traits in legumes, such as taste, size, disease resistance, and drought tolerance.

Gene-editing technology is an efficient, precise, and crucial way of meeting the health needs of an increasingly populous world and helping farmers cultivate better crops. In addition to agricultural implications, CRISPR technology could be used more widely in the future to clarify genomic structures and their role in all plants and legumes. For instance, CRISPR technology may improve the understanding of transcriptional regulation of Cas9 and Cpf1, the monitoring of genetic loci and mechanisms, and the regulation of promoter activity in plants (Ahmar et al., 2020). Furthermore, it will cover single-nucleotide polymorphisms (SNPs), as well as genome-wide association studies to change and better understand the epigenetic behavior of legumes. By creating a genome-wide association study in *M. truncatula,* this technique was able to identify the nodulation-associated genes (Curtin et al., 2017) and also the mutational research of five different nodule-related PLAT domains (NPD1-5), and member genes of the nitrate peptide family (NPD) have been identified (Trujillo et al., 2019; Wang, L., et al., 2020).

Hairy root transformation is a long been used technique that enables the production of transgenic roots in a quick and straightforward manner. It is mainly chosen when there are no protocols existing for stable transformation and regeneration, or the desired traits were only seen in roots (Christey, 2001). Recently, a study described the possibility of editing the genomes of transgenic hairy roots using CRISPR/Cas9 (Kiryushkin et al., 2022). This study concluded that by combining these techniques, it is possible to study gene function quickly and efficiently. Therefore, Hairy CRISPR, the term used in Kiryushkin's study, can be considered as another alternative application of genome engineering tools to overcome genome-editing problems in recalcitrant legume species.

CRISPR/Cas9 utilization in legume breeding programs can be implemented in the future to improve different prominent agronomically important features including biotic and abiotic stress resistance, quality and nutritional value increase, augmentation of carotenoid content, and obtaining sulfur-containing amino acids (**Figure 3**). Novel studies discussed in this review show that genome-editing technologies including CRISPR/Cas9 have been widely used for gaining significant traits for legumes, but it is still needed to improve efficient regeneration and transformation systems, reliable screening and selection strategies, and construction of multi-purpose vector systems.

To overcome these problems, several approaches that could be used in near future are summarized as follows:

-Since legumes still have low regenerative capacity, *de novo* meristem induction (Maher et al., 2020) could be used to eliminate tissue culture steps.
-To improve transformation efficiency, new methods could be developed, including the delivery of transformation vectors into germline cells (Zaidi et al., 2020).
-Using inducible promoters (i.e., heat-inducible CRISPR) instead of constitutively active promoters could also help to increase the efficiency of gene targeting in legumes, as it worked well in maize (Barone et al., 2020).
-CRISPR still needs to be improved to reduce off-targeting for its extensive use in legumes.
-Although pea is still not gene-edited, its eIF4E virus-resistance allele (eIF4E1)-encoded N176K substitution with base-editing of the Arabidopsis *eIF4E1* gene generated clover yellow vein virus (ClYVV)-resistant Arabidopsis plants (Bastet et al., 2019). Biomimicking of this natural polymorphism that existed in legumes could be used for the induction of biotic stress tolerance in crops as well as other legumes.
-Even though CRISPR eliminates the possibility of the presence of foreign DNA in the final product, extended field trials should also be carried out to see the performance of the GE plants prior to their commercialization.

4.2 Regulatory Framework of Genome-Edited Legume Crops

In this section, genome-editing application and regulatory framework to all plants were examined in detail, since there is no specific regulation for genome-edited legumes. Both public and private breeders believe that gene editing, the latest innovation in genetic engineering, has great potential to develop new plant varieties. It is possible to edit the genome in several ways resulting in different products: allele replacement, site-directed deletions, site-directed insertions (or site-directed nucleases-SDN-1/2/3 in the terminology of Friedrichs et al. (2019a)), and base conversion (Marzec and Hensel, 2020). CRISPR/Cas9 genome-editing system, one of the most widely used genome-editing systems, has gained wide-scale adoption for its application in biomedicine, agriculture, industrial biotechnology, and other bioeconomy sectors (Friedrichs et al., 2019b). According to the jurisdiction, each of these may be subject to a different regulatory approach. Plant breeders might need to meet different requirements for research, legal, regulatory, and marketing when developing new genome-edited plants using those "genome-editing techniques" (Entine et al., 2021). In many countries, gene-editing regulation initially caused a great deal of confusion, which has been cleared up in the past 4 years. The advent of genome editing has brought new regulatory challenges, particularly in relation to regulatory differences and traceability, which can lead to new types of obligations and trade dilemmas. Across many countries and regions in the world, different regulatory approaches were examined in this section.

GE/GM organisms are regulated in Australia, New Zealand, Europe, and India through a process-driven regulatory trigger. These jurisdictions are revising the content of their regulatory definitions to reveal whether all kinds of genome editing are covered under their existing GE/GM regulatory frameworks. The current situation has shown that all plant varieties produced using the gene-editing technology would have to meet the same standards as GMOs as required by the European Union (EU). A Technical Review of the Gene Regulations had been started by the Office of the Gene Technology Regulator (OGTR) for Australia in October 2016, which resulted in some proposed changes. According to these new regulations, GMO regulations would not apply to organisms modified with site-directed nucleases without templates for genome repair (i.e., SDN-1). As with organisms modified by oligonucleotide-directed mutagenesis, organisms modified using a template to direct genome repair (e.g., SDN-2, SDN-3, etc.) are GMOs. There is still an ongoing discussion among regulatory agencies regarding the regulation of all new technologies according to the existing regulatory framework in India.

GE/GM and genome-editing products are regulated in Canada and the United States according to a product trigger, under which the novelty of a particular trait is evaluated on a case-by-case basis, regardless of the technology used to develop it. Leader in the production of GM crops, the United States has proclaimed that any crop variety containing no foreign genes would be regarded as a conventionally bred crop variety rather than a GM crop. A risk-based, product-triggered regulatory approach is followed in Canada. Biotechnology products derived from gene editing in Canada are subject to a pre-market safety assessment only if they are new (i.e., display a new characteristic) and could pose new risks. Gene-edited products do not need pre-market safety assessments in Canada if they do not exhibit a novel trait (i.e., "novel" refers to "new to the Canadian environment or to the food or feed supply in Canada"). Argentina was one of the first countries to adopt a regulatory solution for new (plant) breeding techniques (N(P)BTs) in 2015, covering the genome-editing subcategory as part of it. If the NBT contains no new genetic material, a non-GM regulatory classification is implemented.

Consequently, genome editing is regulated in different ways in different countries. Because of country-specific regulations, genome-editing regulations are not harmonized globally. Moreover, a variety of regulatory and policy approaches to genome editing need to be understood by different jurisdictions. Although uniform global approaches are not possible, a common understanding is essential for reducing the troubles arising from the use of different regulations. There are many potential applications of genome-editing technology, from medicine and healthcare to food and agricultural production that could help address many of the grand challenges facing the 21st century society. Therefore, market acceptance of genome editing, as well as a transparent discussion of both risks and benefits, will be crucial to any governance activity. Due to the complexity of genome-editing technology, regulators and risk assessors must update their knowledge to respond to escalating information requirements. For genome editing to become a marketable technology, all stakeholders need to prioritize both communication and information exchange. Both advocates and opponents should explain risks based on science without sensationalizing or scaring the public.

5 CONCLUSION

Emmanuelle Charpentier and Jennifer Doudna were awarded the 2020 Nobel prize in Chemistry due to the usage of the world's most popular genetic engineering tools, CRISPR/Cas, in medicine, agriculture, and food industries. Although CRISPR technology has more time for routine use in legumes, it is clear that this new-generation genome-editing technology will make important contributions to legume breeding studies to raise productivity and to improve biotic and abiotic stress tolerance with the improvement of technical (i.e., regeneration and transformation) capability of the legumes together with a reduction of off-targets, generation of multiple PAM site selection system, development of tissue-culture free protocols, enhancement of HDR and viral vector efficiencies for CRISPR, and regulatory and policy environment.

AUTHOR CONTRIBUTIONS

MB wrote the CRISPR part. ABY wrote the TALEN/ZFN part, NT, YCA, and YÖÇ wrote the application part, and PB wrote the regulation and ethical part. All authors read and confirmed the whole article.

REFERENCES

Aasim, M., Barpete, S., Usta, A., and Sevinc, C. (2019). An Insight Into Agrobacterıum Tumafacıens-Mediated Genetic Transformation Studies in Mungbean (Vigna Radiata L. Wilczek). *J. Glob. Innov. Agric. Soc. Sci.* 7 (2), 47–51. doi:10.22194/jgiass/7.855

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., et al. (2017). RNA Targeting with CRISPR-Cas13. *Nature* 550, 280–284. doi:10.1038/nature24049

Aftabi, M., Teressa Negawo, A., and Hassan, F. (2017). Improved Protocol for Agrobacterium-Mediated Transformation of Pea (*Pisum Sativum*). *Mol. Biol.* 7, 1–6. doi:10.4172/2168-9547.1000202

Ahmar, S., Saeed, S., Khan, M. H. U., Ullah Khan, S., Mora-Poblete, F., Kamran, M., et al. (2020). A Revolution toward Gene-Editing Technology and its Application to Crop Improvement. *Int. J. Mol. Sci.* 21 (26), 5665. doi:10.3390/ijms21165665

Al Amin, N., Ahmad, N., Wu, N., Pu, X., Ma, T., Du, Y., et al. (2019). CRISPR-Cas9 Mediated Targeted Disruption of FAD2-2 Microsomal Omega-6 Desaturase in Soybean (Glycine max.L). *BMC Biotechnol.* 19, 9. doi:10.1186/s12896-019-0501-2

Amack, S. C., and Antunes, M. S. (2020). CaMV35S Promoter - A Plant Biology and Biotechnology Workhorse in the Era of Synthetic Biology. *Curr. Plant Biol.* 24, 100179. doi:10.1016/j.cpb.2020.100179

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* 576, 149–157. doi:10.1038/s41586-019-1711-4

Aragão, F. J. L., Ribeiro, S. G., Barros, L. M. G., Brasileiro, A. C. M., Maxwell, D. P., Rech, E. L., et al. (1998). Transgenic Beans (*Phaseolus vulgaris* L.) Engineered to Express Viral Antisense RNAs Show Delayed and Attenuated Symptoms to Bean Golden Mosaic Geminivirus. *Mol. Breed.* 4 (6), 491–499. doi:10.1023/A:1009613607559

Arora, L., and Narula, A. (2017). Gene Editing and Crop Improvement Using CRISPR-Cas9 System. *Front. Plant Sci.* 8, 1932. doi:10.3389/fpls.2017.01932

Assefa, T., Assibi Mahama, 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

Atkins, P. A., and Voytas, D. F. (2020). Overcoming Bottlenecks in Plant Gene Editing. *Curr. Opin. Plant Biol.* 54, 79–84. doi:10.1016/j.pbi.2020.01.002

Badhan, S., Ball, A. S., and Mantri, N. (2021). First Report of CRISPR/Cas9 Mediated DNA-free Editing of *4CL* and *RVE7* Genes in Chickpea Protoplasts. *Int. J. Mol. Sci.* 22, 396. doi:10.3390/ijms22010396

Bai, M., Yuan, J., Kuang, H., Gong, P., Li, S., Zhang, Z., et al. (2020). Generation of a Multiplex Mutagenesis Population via Pooled CRISPR -Cas9 in Soya Bean. *Plant Biotechnol. J.* 18 (3), 721–731. Epub 2019 Sep 9. PMID: 31452351; PMCID: PMC7004907. doi:10.1111/pbi.13239

Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., et al. (2019). CRISPR/Cas9-mediated Targeted Mutagenesis of *GmSPL9* Genes Alters Plant Architecture in Soybean. *BMC Plant Biol.* 19, 131. doi:10.1186/s12870-019-1746-6

Bao, A., Zhang, C., Huang, Y., Chen, H., Zhou, X., and Cao, D. (2020). Genome Editing Technology and Application in Soybean Improvement. *Oil Crop Sci.* 5 (1), 31–40. ISSN 2096-2428. doi:10.1016/j.ocsci.2020.03.001

Barone, P., Wu, E., Lenderts, B., Anand, A., Gordon-Kamm, W., Svitashev, S., et al. (2020). Efficient Gene Targeting in Maize Using Inducible CRISPR-Cas9 and Marker-free Donor Template. *Mol. Plant* 13 (8), 1219–1227. Epub 2020 Jun 20. PMID: 32574856. doi:10.1016/j.molp.2020.06.008

Barraza, A., Cabrera-Ponce, J. L., Gamboa-Becerra, R., Luna-Martínez, F., Winkler, R., and Álvarez-Venegas, R. (2015). The *Phaseolus vulgaris* PvTRX1h Gene Regulates Plant Hormone Biosynthesis in Embryogenic Callus from Common Bean. *Front. Plant Sci.* 6, 577. doi:10.3389/fpls.2015.00577

Bastet, A., Zafirov, D., Giovinazzo, N., Guyon-Debast, A., Nogué, F., Robaglia, C., et al. (2019). Mimicking Natural Polymorphism in *eIF4E* by CRISPR-Cas9 Base Editing Is Associated with Resistance to Potyviruses. *Plant Biotechnol. J.* 17 (9), 1736–1750. doi:10.1111/pbi.13096

Bastianelli, D., Grosjean, F., Peyronnet, C., Duparque, M., and Régnier, J. M. (1998). Feeding Value of Pea (*Pisum Sativum*, L.) 1. Chemical Composition of Different Categories of Pea. *Anim. Sci.* 67 (3), 609–619. doi:10.1017/S1357729800033051

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug II, R. G., et al. (2012). *In Vivo* genome Editing Using a High-Efficiency TALEN System. *Nature* 491, 114–118. doi:10.1038/nature11537

Bett, K., Ramsay, L., Chan, C., Sharpe, A. G., Cook, D. R., Penmetsa, R. V., et al. (2016). "Lentil 1.0 and beyond," in Plant & Animal Genome Conference XXIV (San Diego, CA: TSNN), W222.

Bett, K., Ramsay, L., Sharpe, A., Cook, D., Penmetsa, R. V., and Verma, N. (2014). "Lentil Genome Sequencing: Establishing a Comprehensive Platform for Molecular Breeding," in Proceedings of International Food Legumes Research Conference (IFLRC-VI) and ICCLG-VII (SK, Saskatoon: Crop Development Center), 19.

Bhowmik, P., Konkin, D., Polowick, P., Hodgins, C. L., Subedi, M., Xiang, D., et al. (2021). CRISPR/Cas9 Gene Editing in Legume Crops: Opportunities and Challenges. *Legume Sci.* 3 (Issue 3). doi:10.1002/leg3.96

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the Code of DNA Binding Specificity of TAL-type III Effectors. *Science* 326 (5959), 1509–1512. doi:10.1126/science.1178811

Bottero, E., Massa, G., González, M., Stritzler, M., Tajima, H., Gómez, C., et al. (2021). Efficient CRISPR/Cas9 Genome Editing in Alfalfa Using a Public Germplasm. *Front. Agron.* 3, 661526. doi:10.3389/fagro.2021.661526

Böttinger, P., Steinmetz, A., Schieder, O., and Pickardt, T. (2001). Agrobacterium Mediated Transformation of *Vicia faba*. *Mol. Breed.* 8, 243–254. doi:10.1023/A:1013711210433

Cai K, K., Yin, J., Chao, H., Ren, Y., Jin, L., Cao, Y., et al. (2018). A C3HC4-type RING Finger Protein Regulates Rhizobial Infection and Nodule Organogenesis in *Lotus Japonicus*. *J. Integr. Plant Biol.* 60 (9), 878–896. PMID: 30047576. doi:10.1111/jipb.12703

Cai Y, Y., Chen, L., Sun, S., Wu, C., Yao, W., Jiang, B., et al. (2018). CRISPR/Cas9-Mediated Deletion of Large Genomic Fragments in Soybean. *Int. J. Mol. Sci.* 19 (12), 3835. doi:10.3390/ijms19123835

Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., et al. (2015). CRISPR/Cas9-Mediated Genome Editing in Soybean Hairy Roots. *PLoS ONE* 10 (8), e0136064. doi:10.1371/journal.pone.0136064

Cai, Y., Wang, L., Chen, L., Wu, T., Liu, L., Sun, S., et al. (2020). Mutagenesis of *GmFT2a* and *GmFT5a* Mediated by CRISPR/Cas9 Contributes for Expanding the Regional Adaptability of Soybean. *Plant Biotechnol. J.* 18 (1), 298–309. Epub 2019 Jul 5. PMID: 31240772; PMCID: PMC6920152. doi:10.1111/pbi.13199

Cardi, T., D'Agostino, N., and Tripodi, P. (2017). Genetic Transformation and Genomic Resources for Next-Generation Precise Genome Engineering in Vegetable Crops. *Front. Plant Sci.* 8, 241. doi:10.3389/fpls.2017.00241

Cathomen, T., and Keith Joung, J. (2008). Zinc-finger Nucleases: the Next Generation Emerges. *Mol. Ther.* 16 (7), 1200–1207. PMID: 18545224. doi:10.1038/mt.2008.114

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.* 39 (12), e82. doi:10.1093/nar/gkr218

Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M., and Makowski, D. (2015). Estimating Variability in Grain Legume Yields Across Europe and the Americas. *Sci. Rep.* 5, 11171. doi:10.1038/srep11171

Che, P., Chang, S., Simon, M. K., Zhang, Z., Shaharyar, A., Ourada, J., et al. (2021). Developing a Rapid and Highly Efficient Cowpea Regeneration, Transformation and Genome Editing System Using Embryonic axis Explants. *Plant J.* 106 (3), 817–830. Epub 2021 Mar 16. PMID: 33595147; PMCID: PMC8252785. doi:10.1111/tpj.15202

Chow, R. D., Chen, J. S., Shen, J., and Chen, S. (2021). A Web Tool for the Design of Prime-Editing Guide RNAs. *Nat. Biomed. Eng.* 5 (2), 190–194. doi:10.1038/s41551-020-00622-8

Christey, M. C. (2001). Use of Ri-Mediated Transformation for Production of Transgenic plantsVitro Cell. *Vitro Cell.Dev.Biol.-Plant* 37, 687–700. doi:10.1007/s11627-001-0120-0

Cichy, K. A., Caldas, G. V., Snapp, S. S., and Blair, M. W. (2009). QTL Analysis of Seed Iron, Zinc, and Phosphorus Levels in an Andean Bean Population. *Crop Sci.* 49, 1742–1750. doi:10.2135/cropsci2008.10.0605

Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: Intuitive Guide Selection for CRISPR/Cas9 Genome Editing Experiments and Screens. *Nucleic Acids Res.* 46

(W1), W242–W245. PMID: 29762716; PMCID: PMC6030908. doi:10.1093/nar/gky354

Confalonieri, M., Carelli, M., Gianoglio, S., Moglia, A., Biazzi, E., and Tava, A. (2021). CRISPR/Cas9-Mediated Targeted Mutagenesis of *CYP93E2* Modulates the Triterpene Saponin Biosynthesis in *Medicago Truncatula*. *Front. Plant Sci.* 12, 690231. doi:10.3389/fpls.2021.690231

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 (6121), 819–823. doi:10.1126/science.1231143

Cooper, J. W., Wilson, M. H., Derks, M. F. L., Smit, S., Kunert, K. J., Cullis, C., et al. (2017). Enhancing Faba Bean (*Vicia faba* L.) Genome Resources. *J. Exp. Bot.* 68 (8), 1941–1953. doi:10.1093/jxb/erx117

Curtin, S. J., Michno, J.-M., Campbell, B. W., Gil-Humanes, J., Mathioni, S. M., Hammond, R., et al. (2016). MicroRNA Maturation and MicroRNA Target Gene Expression Regulation Are Severely Disrupted in Soybean *Dicer-Like1* Double Mutants. *G3 (Bethesda)* 6 (2), 423–433. doi:10.1534/g3.115.022137

Curtin, S. J., Tiffin, P., Guhlin, J., Trujillo, D. I., Burghardt, L. T., Atkins, P., et al. (2017). Validating Genome-wide Association Candidates Controlling Quantitative Variation in Nodulation. *Plant Physiol.* 173 (2), 921–931. Epub 2017 Jan 5. PMID: 28057894; PMCID: PMC5291020. doi:10.1104/pp.16.01923

Curtin, S. J., Xiong, Y., Michno, J.-M., Campbell, B. W., Stec, A. O., Čermák, T., et al. (2018). CRISPR/Cas9 and TALENs Generate Heritable Mutations for Genes Involved in Small RNA Processing ofGlycine maxandMedicago Truncatula. *Plant Biotechnol. J.* 16, 1125–1137. doi:10.1111/pbi.12857

Curtin, S. J., Zhang, F., Sander, J. D., Haun, W. J., Starker, C., Baltes, N. J., et al. (2011). Targeted Mutagenesis of Duplicated Genes in Soybean with Zinc-Finger Nucleases. *Plant physiol.* 156 (2), 466–473. doi:10.1104/pp.111.172981

Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., et al. (2016). Direct Stacking of Sequence-specific Nuclease-Induced Mutations to Produce High Oleic and Low Linolenic Soybean Oil. *BMC Plant Biol.* 16, 1–8. doi:10.1186/s12870-016-0906-1

Di, Y.-H., Sun, X.-J., Hu, Z., Jiang, Q.-Y., Song, G.-H., Zhang, B., et al. (2019). Enhancing the CRISPR/Cas9 System Based on Multiple GmU6 Promoters in Soybean. *Biochem. Biophysical Res. Commun.* 519 (4), 819–823. ISSN 0006-91X. doi:10.1016/j.bbrc.2019.09.074

Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T. N., Stacey, G., Gillman, J. D., et al. (2019). Demonstration of Highly Efficient Dual gRNA CRISPR/Cas9 Editing of the Homeologous GmFAD2-1A and GmFAD2-1B Genes to Yield a High Oleic, Low Linoleic and α-linolenic Acid Phenotype in Soybean. *BMC Plant Biol.* 19, 311. doi:10.1186/s12870-019-1906-8

Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., et al. (2016). Efficient Targeted Mutagenesis in Soybean by TALENs and CRISPR/Cas9. *J. Biotechnol.* 217, 90–97. doi:10.1016/j.jbiotec.2015.11.005

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H., and Chandrasegaran, S. (2005). Zinc Finger Nucleases: Custom-Designed Molecular Scissors for Genome Engineering of Plant and Mammalian Cells. *Nucleic Acids Res.* 33 (18), 5978–5990. PMID: 16251401; PMCID: PMC1270952. doi:10.1093/nar/gki912

Ellis, T. H. N., Hofer, J. M. I., Timmerman-Vaughan, G. M., Coyne, C. J., and Hellens, R. P. (2011). Mendel, 150 Years on. *Trends Plant Sci.* 16, 590–596. doi:10.1016/j.tplants.2011.06.006

Entine, J., Felipe, M. S. S., Groenewald, J.-H., Kershen, D. L., Lema, M., McHughen, A., et al. (2021). Regulatory Approaches for Genome Edited Agricultural Plants in Select Countries and Jurisdictions Around the World. *Transgenic Res.* 30, 551–584. doi:10.1007/s11248-021-00257-8

Friedrichs, S., Takasu, Y., Kearns, P., Dagallier, B., Oshima, R., Schofield, J., et al. (2019b). An Overview of Regulatory Approaches to Genome Editing in Agriculture. *Biotechnol. Res. Innovation* 3 (3), 208–220. doi:10.1016/j.biori.2019.07.001

Friedrichs, S., Takasu, Y., Kearns, P., Dagallier, B., Oshima, R., Schofield, J., et al. (2019a). Policy Considerations Regarding Genome Editing. *Trends Biotechnol.* 37 (10), 1029–1032. doi:10.1016/j.tibtech.2019.05.005

Gaj, T., Gersbach, C. A., and Barbas, C. F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based Methods for Genome Engineering. *Trends Biotechnol.* 31 (7), 397–405. Epub 2013 May 9. PMID: 23664777; PMCID: PMC3694601. doi:10.1016/j.tibtech.2013.04.004

Garg, V., Dudchenko, O., Wang, J., Khan, A. W., Gupta, S., Kaur, P., et al. (2021). Chromosome-length Genome Assemblies of Six Legume Species Provide Insights into Genome Organization, Evolution, and Agronomic Traits for Crop Improvement. *J. Adv. Res.* 2021. doi:10.1016/j.jare.2021.10.009

Han, J., Guo, B., Guo, Y., Zhang, B., Wang, X., and Qiu, L.-J. (2019). Creation of Early Flowering Germplasm of Soybean by CRISPR/Cas9 Technology. *Front. Plant Sci.* 10, 1446. doi:10.3389/fpls.2019.01446

Hanafy, M., Pickardt, T., Kiesecker, H., and Jacobsen, H.-J. (2005). Agrobacterium-mediated Transformation of Faba Bean (Vicia faba L.) Using Embryo Axes. *Euphytica* 142, 227–236. doi:10.1007/s10681-005-1690-4

Hanafy, M. S., El-Banna, A., Schumacher, H. M., Jacobsen, H.-J., and Hassan, F. S. (2013). Enhanced Tolerance to Drought and Salt Stresses in Transgenic Faba Bean (Vicia faba L.) Plants by Heterologous Expression of the PR10a Gene from Potato. *Plant Cell Rep.* 32, 663–674. doi:10.1007/s00299-013-1401-x

Hassan, M. M., Zhang, Y., Yuan, G., De, K., Chen, J.-G., Muchero, W., et al. (2021). Construct Design for CRISPR/Cas-based Genome Editing in Plants. *Trends Plant Sci.* 26 (11), 1133–1152. doi:10.1016/j.tplants.2021.06.015

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

Hegazy, W. A., and Youns, M. (2016). TALENs Construction: Slowly but Surely. *Asian Pac J. Cancer Prev.* 17, 3329–3334. doi:10.14456/apjcp.2016.97/APJCP.2016.17.7.3329

Hsu, J. Y., Grünewald, J., Szalay, R., Shih, J., Anzalone, A. V., Lam, K. C., et al. (2021). PrimeDesign Software for Rapid and Simplified Design of Prime Editing Guide RNAs. *Nat. Commun.* 12 (1), 1034–1036. doi:10.1038/s41467-021-21337-7

Huang, T.-K., and Puchta, H. (2019). CRISPR/Cas-mediated Gene Targeting in Plants: Finally a Turn for the Better for Homologous Recombination. *Plant Cell Rep.* 38, 443–453. doi:10.1007/s00299-019-02379-0

Hurt, J. A., Thibodeau, S. A., Hirsh, A. S., Pabo, C. O., and Joung, J. K. (2003). Highly Specific Zinc Finger Proteins Obtained by Directed Domain Shuffling and Cell-Based Selection. *Proc. Natl. Acad. Sci. U.S.A.* 100 (21), 12271–12276. Epub 2003 Oct 3. PMID: 14527993; PMCID: PMC218748. doi:10.1073/pnas.2135381100

Hwang, G.-H., Park, J., Lim, K., Kim, S., Yu, J., Yu, E., et al. (2018). Web-based Design and Analysis Tools for CRISPR Base Editing. *Bmc Bioinforma.* 19 (1), 1–7. doi:10.1186/s12859-018-2585-4

Ji, J., Zhang, C., Sun, Z., Wang, L., Duanmu, D., and Fan, Q. (2019). Genome Editing in Cowpea *Vigna Unguiculata* Using CRISPR-Cas9. *Int. J. Mol. Sci.* 20 (10), 2471. doi:10.3390/ijms20102471

Joung, J. K., and Sander, J. D. (2013). TALENs: a Widely Applicable Technology for Targeted Genome Editing. *Nat. Rev. Mol. Cell Biol.* 14, 49–55. doi:10.1038/nrm3486

Juranić, M., Nagahatenna, D. S. K., Salinas-Gamboa, R., Hand, M. L., Sánchez-León, N., Leong, W. H., et al. (2020). A Detached Leaf Assay for Testing Transient Gene Expression and Gene Editing in Cowpea (*Vigna Unguiculata* [L.] Walp.). *Plant Methods* 16, 88. doi:10.1186/s13007-020-00630-4

Kahlon, J. G., Jacobsen, H.-J., Chatterton, S., Hassan, F., Bowness, R., and Hall, L. M. (2018). Lack of Efficacy of Transgenic Pea (Pisum sativumL.) Stably Expressing Antifungal Genes againstFusarium spp.in Three Years of Confined Field Trials. *GM Crops Food* 9, 90–108. doi:10.1080/21645698.2018.1445471

Kang, B.-C., Yun, J.-Y., Kim, S.-T., 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

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., et al. (2014). Genome Sequence of Mungbean and Insights into Evolution within Vigna Species. *Nat. Commun.* 5 (1), 1–9. doi:10.1038/ncomms6443

Keatinge, J. D. H., Easdown, W. J., Yang, R. Y., Chadha, M. L., and Shanmugasundaram, S. (2011). Overcoming Chronic Malnutrition in a Future Warming World: the Key Importance of Mungbean and Vegetable Soybean. *Euphytica* 180, 129–141. doi:10.1007/s10681-011-0401-6

Khatib, F., Makris, A., Yamaguchi-Shinozaki, K., Kumar, S., Sarker, A., Erskine, W., et al. (2011). Expression of the *DREB1A* Gene in Lentil (*Lens Culinaris* Medik. Subsp. *Culinaris*) Transformed with the Agrobacterium System. *Crop Pasture Sci.* 62 (6), 488–495. doi:10.1071/CP10351

Khazaei, H., O'Sullivan, D. M., Stoddard, F. L., Adhikari, K. N., Paull, J. G., Schulman, A. H., et al. (2021). Recent Advances in Faba Bean Genetic and

Genomic Tools for Crop Improvement. *Legume Sci.* 3 (3), e75. doi:10.1002/leg3.75

Kim, H., Kim, S.-T., Ryu, J., Kang, B.-C., Kim, J.-S., and Kim, S.-G. (2017). CRISPR/Cpf1-mediated DNA-free Plant Genome Editing. *Nat. Commun.* 8 (1), 1–7. doi:10.1038/ncomms14406

Kiryushkin, A. S., Ilina, E. L., Guseva, E. D., Pawlowski, K., and Demchenko, K. N. (2022). Hairy CRISPR: Genome Editing in Plants Using Hairy Root Transformation. *Plants* 11 (1), 51. doi:10.3390/plants11010051

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019). Transposon-encoded CRISPR-Cas Systems Direct RNA-Guided DNA Integration. *Nature* 571 (7764), 219–225. doi:10.1038/s41586-019-1323-z

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 (9), 1411–1422. doi:10.1038/s41588-019-0480-1

Kumar, J., Sen Gupta, D., Baum, M., Varshney, R. K., and Kumar, S. (2021). Genomics-assisted Lentil Breeding: Current Status and Future Strategies. *Legume Sci.* 3 (3), e71. doi:10.1002/leg3.71

Kumar, S., Rajendran, K., Kumar, J., Hamwieh, A., and Baum, M. (2015). Current Knowledge in Lentil Genomics and its Application for Crop Improvement. *Front. Plant Sci.* 6, 78. doi:10.3389/fpls.2015.00078

Kumlehn, J., Pietralla, J., Hensel, G., Pacher, M., and Puchta, H. (2018). The CRISPR/Cas Revolution Continues: From Efficient Gene Editing for Crop Breeding to Plant Synthetic Biology. *J. Integr. Plant Biol.* 60 (12), 1127–1153. PMID: 30387552. doi:10.1111/jipb.12734

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP V3: Expanding the CRISPR Web Toolbox beyond Genome Editing. *Nucleic acids Res.* 47 (W1), W171–W174. doi:10.1093/nar/gkz365

Lenka, B., Satapathy, S. N., and Senapati, M. (2020). Engineering Plant Virus Resistance: Gene Silencing to Genome Editing. *Int. J. Curr. Microbiol. App. Sci.* 9 (10), 3086–3096. doi:10.20546/ijcmas.2020.910.371

Lewis, G. P., Schrire, B., Mackinder, B., and Lock, M. (2005). *Legumes of the World*. Richmond, UK: Royal Botanic Gardens, Kew.

Li, C., Nguyen, V., Liu, J., Fu, W., Chen, C., Yu, K., et al. (2019). Mutagenesis of Seed Storage Protein Genes in Soybean Using CRISPR/Cas9. *BMC Res. Notes* 12, 176. doi:10.1186/s13104-019-4207-2

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., et al. (2018). Expanded Base Editing in Rice and Wheat Using a Cas9-Adenosine Deaminase Fusion. *Genome Biol.* 19, 59. doi:10.1186/s13059-018-1443-z

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, Z., Liu, Z.-B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-Guide RNA Directed Genome Editing in Soybean. *Plant Physiol.* 169, 960–970. doi:10.1104/pp.15.00783

Lin, Q., Jin, S., Zong, Y., Yu, H., Zhu, Z., Liu, G., et al. (2021). High-efficiency Prime Editing with Optimized, Paired pegRNAs in Plants. *Nat. Biotechnol.* 39, 923–927. doi:10.1038/s41587-021-00868-w

Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime Genome Editing in Rice and Wheat. *Nat. Biotechnol.* 38, 582–585. doi:10.1038/s41587-020-0455-x

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., and Chen, L.-L. (2017). CRISPR-P 2.0: an Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Mol. plant* 10 (3), 530–532. PMID: 28089950. doi:10.1016/j.molp.2017.01.003

López, Y., Nadaf, H. L., Smith, O. D., Connell, J. P., Reddy, A. S., and Fritz, A. K. (2000). Isolation and Characterization of the Δ12-fatty Acid Desaturase in Peanut (*Arachis hypogaea* L.) and Search for Polymorphisms for the High Oleate Trait in Spanish Market-type Lines. *Theor. Appl. Genet.* 101, 1131–1138. doi:10.1007/s001220051589

Luzzatto, L., and Arese, P. (2018). Favism and Glucose-6-Phosphate Dehydrogenase Deficiency. *N. Engl. J. Med.* 378, 60–71. doi:10.1056/NEJMra17081110.1056/nejmra1708111

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol. Plant* 8, 1274–1284. doi:10.1016/j.molp.2015.04.007

Maalouf, F., Hu, J., O'Sullivan, D. M., Zong, Z., Hamwieh, A., Kumar, S., et al. (2019). Breeding and Genomics Status in Faba Bean (Vicia faba). *Plant Breed.* 138 (4), 465–473. doi:10.1111/pbr.12644

Mabrouk, Y., Hemissi, I., Salem, I. B., Mejri, S., Saidi, M., and Belhadj, O. (2018). Potential of Rhizobia in Improving Nitrogen Fixation and Yields of Legumes. *Symbiosis* 33, 107. doi:10.5772/intechopen.73495

Maher, M. F., Nasti, R. A., Vollbrecht, M., Starker, C. G., Clark, M. D., and Voytas, D. F. (2020). Plant Gene Editing through *De Novo* Induction of Meristems. *Nat. Biotechnol.* 38, 84–89. doi:10.1038/s41587-019-0337-2

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An Updated Evolutionary Classification of CRISPR-Cas Systems. *Nat. Rev. Microbiol.* 13 (11), 722–736. doi:10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary Classification of CRISPR-Cas Systems: a Burst of Class 2 and Derived Variants. *Nat. Rev. Microbiol.* 18 (2), 67–83. doi:10.1038/s41579-019-0299-x

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). CAS9 Transcriptional Activators for Target Specificity Screening and Paired Nickases for Cooperative Genome Engineering. *Nat. Biotechnol.* 31 (9), 833–838. doi:10.1038/nbt.2675

Malzahn, A., Lowder, L., and Qi, Y. (2017). Plant Genome Editing with TALEN and CRISPR. *Cell Biosci.* 7, 21. doi:10.1186/s13578-017-0148-4

Marzec, M., and Hensel, G. (2020). Prime Editing: Game Changer for Modifying Plant Genomes. *Trends Plant Sci.* 25 (8), 722–724. doi:10.1016/j.tplants.2020.05.008

Meng, Y., Hou, Y., Wang, H., Ji, R., Liu, B., Wen, J., et al. (2017). Targeted Mutagenesis by CRISPR/Cas9 System in the Model Legume *Medicago Truncatula*. *Plant Cell Rep.* 36, 371–374. doi:10.1007/s00299-016-2069-9

Michno, J.-M., Wang, X., Liu, J., Curtin, S. J., Kono, T. J., and Stupar, R. M. (2015). CRISPR/Cas Mutagenesis of Soybean and *Medicago Truncatula* Using a New Web-Tool and a Modified Cas9 Enzyme. *GM crops food.* 6 (4), 243–252. doi:10.1080/21645698.2015.1106063

Nadeem, M. A., Yeken, M. Z., Shahid, M. Q., Habyarimana, E., Yılmaz, H., Alsaleh, A., et al. (2021). Common Bean as a Potential Crop for Future Food Security: an Overview of Past, Current and Future Contributions in Genomics, Transcriptomics, Transgenics and Proteomics. *Biotechnol. Biotechnol. Equip.* 35 (1), 759–787. doi:10.1080/13102818.2021.1920462

Negawo, A. T., Aftabi, M., Jacobsen, H.-J., Altosaar, I., and Hassan, F. S. (2013). Insect Resistant Transgenic Pea Expressing *cry1Ac* Gene Product from *Bacillus Thuringiensis*. *Biol. Control* 67, 293–300. doi:10.1016/j.biocontrol.2013.09.016

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., et al. (2017). Type III CRISPR-Cas Systems Produce Cyclic Oligoadenylate Second Messengers. *Nature* 548 (7669), 543–548. doi:10.1038/nature23467

Norden, A. J., Gorbet, D. W., Knauft, D. A., and Young, C. T. (1987). Variability in Oil Quality Among Peanut Genotypes in the Florida Breeding Program1. *Peanut Sci.* 14, 7–11. doi:10.3146/i0095-3679-14-1-3

Oz, M. T., Altpeter, A., Karan, R., Merotto, A., and Altpeter, F. (2021). CRISPR/Cas9-Mediated Multi-Allelic Gene Targeting in Sugarcane Confers Herbicide Tolerance. *Front. Genome Ed.* 3. doi:10.3389/fgeed.2021.673566

Pandey, A. K., Rubiales, D., Wang, Y., Fang, P., Sun, T., Liu, N., et al. (2021). Omics Resources and Omics-Enabled Approaches for Achieving High Productivity and Improved Quality in Pea (*Pisum Sativum* L.). *Theor. Appl. Genet.* 134, 755–776. doi:10.1007/s00122-020-03751-5

Park, J., Bae, S., and Kim, J.-S. (2015). Cas-Designer: a Web-Based Tool for Choice of CRISPR-Cas9 Target Sites. *Bioinformatics* 31 (24), btv537–4016. doi:10.1093/bioinformatics/btv537

Qi, X., Dong, L., Liu, C., Mao, L., Liu, F., Zhang, X., et al. (2018). Systematic Identification of Endogenous RNA Polymerase III Promoters for Efficient RNA Guide-Based Genome Editing Technologies in Maize. *Crop J.* 6, 314–320. doi:10.1016/j.cj.2018.02.005

Rodas, A. L., Roque, E., Hamza, R., Gómez-Mena, C., Minguet, E. G., Wen, J., et al. (2021). MtSUPERMAN Plays a Key Role in Compound Inflorescence and Flower Development in *Medicago Truncatula*. *Plant J.* 105, 816–830. doi:10.1111/tpj.15075

Römer, P., Recht, S., and Lahaye, T. (2009). A Single Plant Resistance Gene Promoter Engineered to Recognize Multiple TAL Effectors from Disparate

Pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20526–20531. doi:10.1073/pnas.0908812106

Sahoo, D. P., Kumar, S., Mishra, S., Kobayashi, Y., Panda, S. K., and Sahoo, L. (2016). Enhanced Salinity Tolerance in Transgenic Mungbean Overexpressing *Arabidopsis Antiporter* (*NHX1*) Gene. *Mol. Breed.* 36 (10). doi:10.1007/s11032-016-0564-x

Sarker, R. H., Biswas, A., Mustafa, B. M., Mahbub, S., and Hoque, M. I. (2003). Agrobacterium-mediated Transformation of Lentil (*Lens Culinaris* Medik.). *Plant Tissue Cult.* 13 (1), 1–12.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome Structure of the Legume, *Lotus Japonicus*. *DNA Res.* 15 (4), 227–239. doi:10.1093/dnares/dsn008

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome Sequence of the Palaeopolyploid Soybean. *Nature* 463, 178–183. doi:10.1038/nature08670

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

Schroeder, H. E., Gollasch, S., Moore, A., Tabe, L. M., Craig, S., Hardie, D. C., et al. (1995). Bean [alpha]-Amylase Inhibitor Confers Resistance to the Pea Weevil (*Bruchus Pisorum*) in Transgenic Peas (*Pisum Sativum* L.). *Plant Physiol.* 107, 1233–1239. doi:10.1104/107.4.123310.1104/pp.107.4.1233

Scott, C. T. (2005). The Zinc Finger Nuclease Monopoly. *Nat. Biotechnol.* 23, 915–918. doi:10.1038/nbt0805-915

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted Genome Modification of Crop Plants Using a CRISPR-Cas System. *Nat. Biotechnol.* 31 (8), 686–688. doi:10.1038/nbt.2650

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted Base Editing in Rice and Tomato Using a CRISPR-Cas9 Cytidine Deaminase Fusion. *Nat. Biotechnol.* 35, 441–443. doi:10.1038/nbt.3833

Shu, H., Luo, Z., Peng, Z., and Wang, J. (2020). The Application of CRISPR/Cas9 in Hairy Roots to Explore the Functions of *AhNFR1* and *AhNFR5* Genes during Peanut Nodulation. *BMC Plant Biol.* 20, 417. doi:10.1186/s12870-020-02614-x

Siegner, S. M., Karasu, M. E., Schröder, M. S., Kontarakis, Z., and Corn, J. E. (2021). PnB Designer: a Web Application to Design Prime and Base Editor Guide RNAs for Animals and Plants. *BMC Bioinforma.* 22 (1), 1–12. doi:10.1186/s12859-021-04034-6

Saini, R., Singh, R. P., and Jaiwal, P. K. (2007). Agrobacterium Tumefaciens Mediated Transfer of Phaseolus vulgaris α-amylase Inhibitor-1 Gene into Mungbean Vigna Radiata (L.) Wilczek Using Bar as Selectable Marker. *Plant Cell Rep.* 26 (2), 187–198. Epub 2006 Sep 16. PMID: 16983450. doi:10.1007/s00299-006-0224-4

Stella, S., Alcón, P., and Montoya, G. (2017). Class 2 CRISPR-Cas RNA-Guided Endonucleases: Swiss Army Knives of Genome Editing. *Nat. Struct. Mol. Biol.* 24 (11), 882–892. doi:10.1038/nsmb.3486

Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., et al. (2015). Targeted Mutagenesis in Soybean Using the CRISPR-Cas9 System. *Sci. Rep.* 5, 10342. doi:10.1038/srep10342

Suzuki, H., Fukushima, E. O., Shimizu, Y., Seki, H., Fujisawa, Y., Ishimoto, M., et al. (2019). *Lotus Japonicus* Triterpenoid Profile and Characterization of the *CYP716A51* and *LjCYP93E1* Genes Involved in Their Biosynthesis in Planta. *Plant Cell Physiology* 60 (11), 2496–2509. doi:10.1093/pcp/pcz145

Svitashev, S., Schwartz, C., Lenderts, B., Young, J. K., and Mark Cigan, A. (2016). Genome Editing in Maize Directed by CRISPR-Cas9 Ribonucleoprotein Complexes. *Nat. Commun.* 7, 13274. doi:10.1038/ncomms13274

Temesgen, T., Keneni, G., Sefera, T., and Jarso, M. (2015). Yield Stability and Relationships Among Stability Parameters in Faba Bean (*Vicia faba* L.) Genotypes. *Crop J.* 3 (3), 258–268. ISSN 2214-5141. doi:10.1016/j.cj.2015.03.004

Tripathi, A., Debnath, S., and Shukla, S. (2021). Review on Biotechnological Advances in *Vigna Radiata* and Their Future Prospects. *J. Crop Sci. Biotechnol.* 24, 245–258. doi:10.1007/s12892-021-00086-5

Trujillo, D. I., Silverstein, K. A. T., and Young, N. D. (2019). Nodule-specific PLAT Domain Proteins Are Expanded in the Medicago Lineage and Required for Nodulation. *New Phytol.* 222, 1538–1550. Epub 2019 Feb 27. PMID: 30664233. doi:10.1111/nph.15697

Tsutsui, H., and Higashiyama, T. (2016). pKAMA-ITACHI Vectors for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout inArabidopsis Thaliana. *Plant Cell Physiol.* 2016, pcw191. doi:10.1093/pcp/pcw191

Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., et al. (2005). Highly Efficient Endogenous Human Gene Correction Using Designed Zinc-Finger Nucleases. *Nature* 435 (7042), 646–651. Epub 2005 Apr 3. PMID: 15806097. doi:10.1038/nature03556

Varshney, R. K., Thudi, M., Pandey, M. K., Tardieu, F., Ojiewo, C., Vadez, V., et al. (2018). Accelerating Genetic Gains in Legumes for the Development of Prosperous Smallholder Agriculture: Integrating Genomics, Phenotyping, Systems Modelling and Agronomy. *J. Exp. Bot.* 69 (13), 3293–3312. doi:10.1093/jxb/ery088

Virdi, K. S., Spencer, M., Stec, A. O., Xiong, Y., Merry, R., Muehlbauer, G. J., et al. (2020). Similar Seed Composition Phenotypes Are Observed from CRISPR-Generated In-Frame and Knockout Alleles of a Soybean KASI Ortholog. *Front. Plant Sci.* 11, 1006. doi:10.3389/fpls.2020.01005

Vu, T. V., Sivankalyani, V., Kim, E. J., Doan, D. T. H., Tran, M. T., Kim, J., et al. (2020). Highly Efficient Homology-directed Repair Using CRISPR/Cpf1-geminiviral Replicon in Tomato. *Plant Biotechnol. J.* 18, 2133–2143. doi:10.1111/pbi.13373

Wang J, J., Kuang, H., Zhang, Z., Yang, Y., Yan, L., Zhang, M., et al. (2020). Generation of Seed Lipoxygenase-free Soybean Using CRISPR-Cas9. *Crop J.* 8 (3), 432–439. ISSN 2214-5141. doi:10.1016/j.cj.2019.08.008

Wang L, L., Sun, S., Wu, T., Liu, L., Sun, X., Cai, Y., et al. (2020). Natural Variation and CRISPR/Cas9-mediated Mutation in GmPRR37 Affect Photoperiodic Flowering and Contribute to Regional Adaptation of Soybean. *Plant Biotechnol. J.* 18, 1869–1881. doi:10.1111/pbi.13346

Wang, L., Rubio, M. C., Xin, X., Zhang, B., Fan, Q., Wang, Q., et al. (2019). CRISPR/Cas9 Knockout of Leghemoglobin Genes inLotus Japonicusuncovers Their Synergistic Roles in Symbiotic Nitrogen Fixation. *New Phytol.* 224, 818–832. doi:10.1111/nph.16077

Wang, L., Wang, L., Tan, Q., Fan, Q., Zhu, H., Hong, Z., et al. (2016). Efficient Inactivation of Symbiotic Nitrogen Fixation Related Genes in *Lotus Japonicus* Using CRISPR-Cas9. *Front. Plant Sci.* 7, 1333. doi:10.3389/fpls.2016.01333

Wen, S., Liu, H., Li, X., Chen, X., Hong, Y., Li, H., et al. (2018). TALEN-mediated Targeted Mutagenesis of *Fatty Acid Desaturase 2* (*FAD2*) in Peanut (*Arachis hypogaea* L.) Promotes the Accumulation of Oleic Acid. *Plant Mol. Biol.* 97, 177–185. doi:10.1007/s11103-018-0731-z

Wolabu, T. W., Park, J. J., Chen, M., Cong, L., Ge, Y., Jiang, Q., et al. (2020). Improving the Genome Editing Efficiency of CRISPR/Cas9 in Arabidopsis and *Medicago Truncatula*. *Planta* 252, 15. doi:10.1007/s00425-020-03415-0

Wu, N., Lu, Q., Wang, P., Zhang, Q., Zhang, J., Qu, J., et al. (2020). Construction and Analysis of *GmFAD2-1A* and *GmFAD2-2A* Soybean Fatty Acid Desaturase Mutants Based on CRISPR/Cas9 Technology. *Int. J. Mol. Sci.* 21 (3), 1104. doi:10.3390/ijms21031104

Xiao, S. (2012). Protecting Crops from Pathogens: Novel Approaches to an Old Problem. *Gene Technol.* 01. doi:10.4172/2329-6682.1000e103

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 Multiplex Editing Capability with the Endogenous tRNA-Processing System. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3570–3575. doi:10.1073/pnas.1420294112

Xing, H.-L., Dong, L., Wang, Z.-P., Zhang, H.-Y., Han, C.-Y., Liu, B., et al. (2014). A CRISPR/Cas9 Toolkit for Multiplex Genome Editing in Plants. *BMC Plant Biol.* 14, 327. doi:10.1186/s12870-014-0327-y

Xu, R., Li, J., Liu, X., Shan, T., Qin, R., and Wei, P. (2020). Development of Plant Prime-Editing Systems for Precise Genome Editing. *Plant Commun.* 1, 100043. doi:10.1016/j.xplc.2020.100043

Xu, W., Zhang, C., Yang, Y., Zhao, S., Kang, G., He, X., et al. (2020). Versatile Nucleotides Substitution in Plant Using an Improved Prime Editing System. *Mol. Plant* 13, 675–678. doi:10.1016/j.molp.2020.03.012

Xue, R., Wu, X., Wang, Y., Zhuang, Y., Chen, J., Wu, J., et al. (2017). Hairy Root Transgene Expression Analysis of a Secretory Peroxidase (PvPOX1) from Common Bean Infected by Fusarium Wilt. *Plant Sci.* 260, 1–7. doi:10.1016/j.plantsci.2017.03.011

Yadav, S. K., Katikala, S., Yellisetty, V., Kannepalle, A., Narayana, J. L., Maddi, V., et al. (2012). Optimization of Agrobacterium Mediated Genetic Transformation of Cotyledonary Node Explants of *Vigna Radiata*. *SpringerPlus* 1 (1), 1–8. doi:10.1186/2193-1801-1-59

Young, N. D., Debellé, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The Medicago Genome Provides Insight into the Evolution of Rhizobial Symbioses. *Nature* 480, 520–524. doi:10.1038/nature10625

Yuan, M., Zhu, J., Gong, L., He, L., Lee, C., Han, S., et al. (2019). Mutagenesis of *FAD2* Genes in Peanut with CRISPR/Cas9 Based Gene Editing. *BMC Biotechnol.* 19, 24. doi:10.1186/s12896-019-0516-8

Yue, J.-J., HongWei, C.-Y. P., Wei, P., Tsai, Y.-C., and Lin, C.-S. (2020). How to Start Your Monocot CRISPR/Cas Project: Plasmid Design, Efficiency Detection, and Offspring Analysis. *Rice* 13, 9. doi:10.1186/s12284-019-0354-2

Zaidi, S. S.-e. -A., Mahas, A., Vanderschuren, H., and Mahfouz, M. M. (2020). Engineering Crops of the Future: CRISPR Approaches to Develop Climate-Resilient and Disease-Resistant Plants. *Genome Biol.* 21, 289. doi:10.1186/s13059-020-02204-y

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163 (3), 759–771. doi:10.1016/j.cell.2015.09.038

Zhan, X., Lu, Y., Zhu, J. K., and Botella, J. R. (2021). Genome Editing for Plant Research and Crop Improvement. *J. Integr. Plant Biol.* 63 (1), 3–33. doi:10.1111/jipb.13063

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., et al. (2014). The CRISPR/Cas9 System Produces Specific and Homozygous Targeted Gene Editing in Rice in One Generation. *Plant Biotechnol. J.* 12 (6), 797–807. Epub 2014 May 23. doi:10.1111/pbi.12200

Zhang, H., Cao, Y., Zhang, H., Xu, Y., Zhou, C., Liu, W., et al. (2020). Efficient Generation of CRISPR/Cas9-Mediated Homozygous/Biallelic *Medicago Truncatula* Mutants Using a Hairy Root System. *Front. Plant Sci.* 11, 294. doi:10.3389/fpls.2020.00294

Zhang, X., Ferreira, I. R. S., and Schnorrer, F. (2014). A Simple TALEN-Based Protocol for Efficient Genome-Editing in Drosophila. *Methods* 69 (1), 32–37. doi:10.1016/j.ymeth.2014.03.020

Zheng, N., Li, T., Dittman, J. D., Su, J., Li, R., Gassmann, W., et al. (2020). CRISPR/Cas9-Based Gene Editing Using Egg Cell-specific Promoters in Arabidopsis and Soybean. *Front. Plant Sci.* 11, 800. doi:10.3389/fpls.2020.00800

Zhu, H., Li, C., and Gao, C. (2020). Applications of CRISPR-Cas in Agriculture and Plant Biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677. doi:10.1038/s41580-020-00288-9

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., et al. (2018). Efficient C-To-T Base Editing in Plants Using a Fusion of nCas9 and Human APOBEC3A. *Nat. Biotechnol.* 36, 950–953. doi:10.1038/nbt.4261

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise Base Editing in Rice, Wheat and Maize with a Cas9-Cytidine Deaminase Fusion. *Nat. Biotechnol.* 35, 438–440. doi:10.1038/nbt.3811

Zuker, M. (2003). Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* 31 (13), 3406–3415. doi:10.1093/nar/gkg595

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 © 2022 Baloglu, Celik Altunoglu, Baloglu, Yildiz, Türkölmez and Özden Çiftçi. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 11 July 2022
doi: 10.3389/fgene.2022.849016

Genome-wide Association Study for Yield and Yield-Related Traits in Diverse Blackgram Panel (*Vigna mungo* L. Hepper) Reveals Novel Putative Alleles for Future Breeding Programs

Lovejit Singh[1†], Guriqbal Singh Dhillon[2†], Sarabjit Kaur[2†], Sandeep Kaur Dhaliwal[1], Amandeep Kaur[2†], Palvi Malik[2†], Ashok Kumar[3], Ranjit Kaur Gill[1] and Satinder Kaur[2†]*

[1]Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India, [2]School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, [3]Regional Research Station, Punjab Agricultural University, Gurdaspur, India

Blackgram (*Vigna mungo* L. Hepper) is an important tropical and sub-tropical short-duration legume that is rich in dietary protein and micronutrients. Producing high-yielding blackgram varieties is hampered by insufficient genetic variability, absence of suitable ideotypes, low harvest index and susceptibility to biotic-abiotic stresses. Seed yield, a complex trait resulting from the expression and interaction of multiple genes, necessitates the evaluation of diverse germplasm for the identification of novel yield contributing traits. Henceforth, a panel of 100 blackgram genotypes was evaluated at two locations (Ludhiana and Gurdaspur) across two seasons (*Spring* 2019 and *Spring* 2020) for 14 different yield related traits. A wide range of variability, high broad-sense heritability and a high correlation of grain yield were observed for 12 out of 14 traits studied among all environments. Investigation of population structure in the panel using a set of 4,623 filtered SNPs led to identification of four sub-populations based on ad-hoc delta K and Cross entropy value. Using Farm CPU model and Mixed Linear Model algorithms, a total of 49 significant SNP associations representing 42 QTLs were identified. Allelic effects were found to be statistically significant at 37 out of 42 QTLs and 50 known candidate genes were identified in 24 of QTLs.

Keywords: *Vigna mungo*, heritability, yield, GWAS, MTA, QTL, allelic effects

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Senthil Natesan,
Tamil Nadu Agricultural University, India
Aditya Pratap,
Indian Institute of Pulses Research (ICAR), India

***Correspondence:**
Satinder Kaur
satinder.biotech@pau.edu

***ORCID:**
Lovejit Singh
orcid.org/0000-0003-2804-0429
Guriqbal Singh Dhillon
orcid.org/0000-0001-6766-810X
Sarabjit Kaur
orcid.org/0000-0003-0672-9268
Amandeep Kaur
orcid.org/0000-0002-9140-3424
Palvi Malik
orcid.org/0000-0003-1215-7120
Satinder Kaur
orcid.org/0000-0003-3704-3074

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

Received: *05 January 2022*
Accepted: *20 June 2022*
Published: *11 July 2022*

Citation:
Singh L, Dhillon GS, Kaur S, Dhaliwal SK, Kaur A, Malik P, Kumar A, Gill RK and Kaur S (2022) Genome-wide Association Study for Yield and Yield-Related Traits in Diverse Blackgram Panel (Vigna mungo L. Hepper) Reveals Novel Putative Alleles for Future Breeding Programs. Front. Genet. 13:849016. doi: 10.3389/fgene.2022.849016

INTRODUCTION

Blackgram (*Vigna mungo* L. Hepper), a diploid (2n = 2X = 22), short duration legume crop of family Leguminaceae, was domesticated in Northern South Asia from progenitor *Vigna mungo* var. *silvestris* (Lukoki et al., 1980). It is cultivated throughout Southeast Asia because of its multiple benefits to soil and human health. It is nutritionally important crop with about 25% protein—nearly three times that of cereals, 60% carbohydrates, 1.3% fat (Das et al., 2021) as well as important vitamins and minerals (Ghafoor et al., 2001), making it a balanced vegan diet when supplemented with cereals. The ability of its roots to fix atmospheric nitrogen (42 kg/ha/year)

(Dey et al., 2020) contribute towards soil health while deep-roots prevents soil erosion by binding soil particles. Short duration of blackgram makes it suitable for intercropping with corn or millet or rotation with cereals like rice or wheat (Muthusamy and Pandiyan, 2018), adding another benefit for farmer.

India is the largest producer and consumer of blackgram as it is colossally grown in almost every agro-climatic zone (Raizada and Souframanien, 2021). However, the crop accounts for only 13% of the total area (56.02 lakh hectares) and 10% of total pulses production (30.60 lakh tons) in India (Muthusamy and Pandiyan, 2018) with productivity of 5.46 quintals per hectare (Singh et al., 2020). Moreover, around 2-3 million tons of pulses are imported annually to fulfill the domestic consumption requirement. Despite the economic and nutritional value of black Gram, the sluggish growth in production is due to lack of commercialized market setup, multiple biotic stresses (mosaic, seedling blight, leaf blight, leaf crinkle virus, leaf folder, Bihar hairy caterpillar, whitefly) and abiotic stresses (drought, salinity, waterlogging). Photo- and thermo-sensitivity of crop with indeterminate habit of flowering and fruiting leads to competition of assimilates between vegetative and reproductive sinks throughout the growth period causing low harvest index and poor grain yield (Somta et al., 2019; Sahu et al., 2020).

The expansion of the crop is constrained by lack of genetic and genomic resources along with limited diversity (Chaitieng et al., 2006; Gupta et al., 2008; Somta et al., 2019 A systematic program of identification, genetic and genomic characterization and utilization of diverse germplasm is required for successfully decoding the genetic architecture of agronomically important traits for blackgram improvement. Genome and transcriptome sequencing (Pootakham et al., 2021; Raizada and Souframanien, 2021), developing dense molecular linkage maps, and using high-throughput genotyping techniques can widen the horizons improvement of this crop. Genotyping–by- sequencing (GBS) is one of the cost-efficient genomic techniques which includes multiplex sequencing of a subset of the genome and generates numerous SNP markers for linkage mapping (Varshney et al., 2009; Elshire et al., 2011; Noble et al., 2018). Genome wide association studies (GWAS) coupled with GBS have been promising tool for estimating the genetic diversity in different crops of soybean (Hwang et al., 2014), chickpea (Plekhanova et al., 2017), cowpea (Xu et al., 2017), pigeonpea (Varshney et al., 2009), and mungbean (Sokolkova et al., 2020) providing insights to underlying genetic architecture of complex traits (Lorenz et al., 2010; Scherer and Christensen, 2016).

In the present study we performed the GWAS on diverse blackgram germplasm panel to assess its genetic diversity and population structure, and to identify MTA (Marker trait associations) involved in yield and yield related traits using GBS. This study provides a unique genomic resource for the genetic dissection of important traits aimed at blackgram improvement.

MATERIALS AND METHODS

Plant Material and Field Trials

A panel consisting of 100 blackgram (*V. mungo*) genotypes was used for the present study. These included 54 genotypes procured from National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India, while, the remaining genotypes were from germplasm collection of Punjab Agricultural University (PAU), Ludhiana, India (**Supplementary Table S1**). The genotypes were evaluated using a simple lattice design (10 × 10), in two replications at two locations (Ludhiana and Gurdaspur) across two seasons (*Spring* 2019 and *Spring* 2020). The seeds were sown in a 2 m long row with 10 cm plant to plant spacing and 30 cm row to row spacing. Ludhiana (30.9°N, 75.85°E) lies in a sub-tropical zone characterized by relatively high temperatures and low precipitation while Gurdaspur region (32.02°N, 75.24°E) is characterized by lower temperature and high humidity coupled with abundant rainfall. The weekly mean temperature, relative humidity and rainfall for Ludhiana and Gurdaspur have been given in **Supplementary Figure S1**.

Phenotypic Evaluation and Statistical Analysis

Data was collected in three replicates from five randomly selected plants of each genotype in each replicate for plant height at 90% pod maturity (PHM), branches per plant (BpP), nodes per plant (NpP), internodal length (IL = PHM/NpP), clusters per plant (CpP), pods per plant (PpP), pod length (PL), seeds per pod (SpP), biological yield per plant (BYpP), yield per plant (YpP), harvest index (HI) and hundred seed weight (HSW). Days to 50% flowering (DtF) and days to 90% pod maturity (DtM) were recorded based on the entire plot. For phenotypic analysis, environments were designated as E1 (Ludhiana, year 2019), E2 (Ludhiana, year 2020), E3 (Ludhiana combined for years 2019 & 2020), E4 (Gurdaspur, year 2019), E5 (Gurdaspur, year 2020) and E6 (Gurdaspur combined for years 2019 & 2020). Due to the significant differences between two selected locations, combined analysis over two selected locations has not been done. Descriptive statistical analysis across all the environments was done using Meta-R v6.0 software (Alvarado et al., 2020). Statistical analysis for individual and multi-environment was performed using "lme4" (Bates et al., 2015) and "Residual Maximum Likelihood (REML)" (Laird and Ware, 1982) methods. The linear model for analyzing individual environments for simple lattice design was done using the formula:

$$Y_{ijk} = \mu + Rep_i + Block_j(Rep_i) + Gen_k + \varepsilon_{ijk} \quad \text{(across replicates, within environment)}$$

$$Y_{ijk1} = \mu + Year_1 + Rep_i(Year_1) + Block_j(Year_1 Rep_i) + Gen_k + Gen_k \times Year_1 + \varepsilon_{ijk1} \quad \text{(across replicates, across environment)}$$

where Y_{ijk} and Y_{ijkl} represent the trait of interest, μ is the overall mean effect, Rep_i is the effect of *i*th replicate, $Block_j$ (Rep_i) is the effect of *j*th incomplete block within the *i*th replicate, Gen_k is the

effect of the *k*th genotype and ε_{ijk} is the error effect associated with the *i*th replication, *j*th incomplete block and *k*th genotype, assumed to be normally distributed with zero mean and variance $\sigma^2\varepsilon$ (Alvarado et al., 2020). $Year_l$ and Gen_l x $Year_i$ are the effects of the *l*th year and Genotype x Year (G x Y) interactions represented by effect on the *i*th genotype in the *l*th year in the linear model for integrated analysis for multi-environment (across the years). The resulting analysis produced the adjusted trait phenotypic values as BLUPs (Best linear unbiased predictions) within and across environments. The genotypes are considered random effects in the BLUPs model, minimizing/eliminating the effect of the environment from phenotypic effects. The broad-sense heritability of trait at individual environment and across environments was calculated as

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/n\,reps} \text{ (Across replicates, within environment)}$$

$$H^2 = \frac{\sigma_g^2}{\frac{\sigma_g^2 + \frac{\sigma_{ge}^2}{n\,env} + \sigma_e^2}{(n\,reps \times n\,env)}} \text{ (Across replicates, across environment)}$$

Where σ_g^2 and σ_e^2 are the genotype, and error variance components, respectively, σ_{ge}^2 is genotype by environment interaction variance, n env is the number of environments, and n reps is the number of replicates. The estimated broad-sense heritability provides valuable insight into the breeding program's quality, with all effects considered as random effects (Alvarado et al., 2020). The LSD at 5% level of significance was calculated as

$$LSD = t(1 - 0.05,\ dfErr)\ X\ ASED$$

where t is the cumulative Student's t-distribution, 0.05 is the selected α level (5%), dfErr is the degrees of freedom for error in the linear mixed model, and ASED is the average standard error of the differences of the means. The coefficient of variation (%) was calculated as:

$$CV = \frac{\sqrt{MSE}}{Grand\ Mean}\ X\ 100$$

where MSE is the mean squared error, and Grand mean is the mean of the trait. BLUPs for recorded traits in all environments were plotted using the ggplot2 v3.3.2 package (Wickham, 2016) in R v4.0.3 (Core R Team 2019).

DNA Isolation and Genotyping

Total genomic DNA was isolated from fresh leaves of single plant of each genotype using modified cetyl trimethyl ammonium bromide (CTAB) method (Saghai-Maroof et al., 1984). Genotyping-by-sequencing (GBS) of DNA samples was outsourced to Novogene Co. Ltd., China. The GBS library was prepared using double digest restriction enzyme DNA (ddRAD) and sequencing was done with Illumina HiSeq 2000. The raw FASTQ files (obtained from Illumina pipeline CASAVA v1.8.2) were processed for quality control and filtered using Trimmomaticv0.39 software (Bolger et al., 2014) Reads with no matching barcode or cut sites overhangs, having more than 10% unidentified bases (N), with adapter dimers, with lower quality reads, and with Q_{phred} score less than 34, were excluded from further analysis. High-quality paired end reads were aligned using Burrows-Wheeler Aligner (BWA) (Li et al., 2009) to *Vigna radiata* reference genome (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/741/045/GCF_000741045.1_Vradiata_ver6). A quality threshold score of 10 was applied to validate SNP loci (Wu et al., 2020). Sorted binary alignment map (BAM) files were converted to variant caller format (vcf) files using mpileup function of bcftools v1.10.2 in samtools v1.10 software package (Li et al., 2009) with minimum read depth ≥4. Haplotype map (hapmap) format files were generated from vcf files using Tassel v5.0 software (Bradbury et al., 2007). After SNP calling, raw hapmap file was filtered by removing indels, minor allele frequency (maf) > 0.05, genotype missing data less than 10% and heterozygosity less than 30% (Torkamaneh et al., 2020). The generated raw reads were submitted to the NCBI sequence read archives (SRA) with accession number PRJNA802066.

Population Structure and Phylogeny Analysis

Bayesian-based approach in STRUCTURE v2.3.4 software (Pritchard et al., 2000) using a burn-in period of 10,000 and Markov chain Monte Carlo iterations of 100,000 for k ranging from 1 to 8 was done to investigate the population structure of germplasm. Evanno's method (Evanno et al., 2005) and cross-entropy method (Chan and Kroese, 2012) were used to obtain an optimum number of sub-populations. Filtered SNPs were used to calculate genetic distance among genotypes and the phylogenetic tree was constructed using neighbour-joining tree in TASSEL v5.0 (Bradbury et al., 2007) and visualized in iTOL v5 (Letunic and Bork, 2016).

Genome-Wide Association Analysis

Filtered SNPs and BLUPs were used to perform association analysis using the Mixed Linear Model (MLM) (Zhang et al., 2010) and FarmCPU algorithms (Liu X. et al., 2016) with GAPIT v3 (Lipka et al., 2012) in R v4.0.3 (Core R Team 2019). The FarmCPU method was used to control false positives and false negatives by iteratively using a fixed-effect model (FEM) and random effect model (REM), testing marker associations from FEM as covariates in REM. *p*-value of 0.001 or $-\log_{10}$ *p*-value of 3.00 was used as a threshold to determine the significance of association (Ikram et al., 2020). The marker-trait associations (MTAs) identified for the same trait within a region of 100bp was considered as part of one QTL. The phenotypic variation explained (PVE) by each significant SNP was calculated as the squared correlation between the phenotype and genotype of the associated SNP (Bhandari et al., 2020). MTAs were considered stable QTLs if they were identified across all the environments of the respective locations with $-\log_{10}$ *p*-value ≥ 3 and PVEW ≥10%. *t*-test based determination of significance based on phenotypic data in two allelic groups was estimated at *p*-value ≤ 0.05 (Xiong et al., 2019).

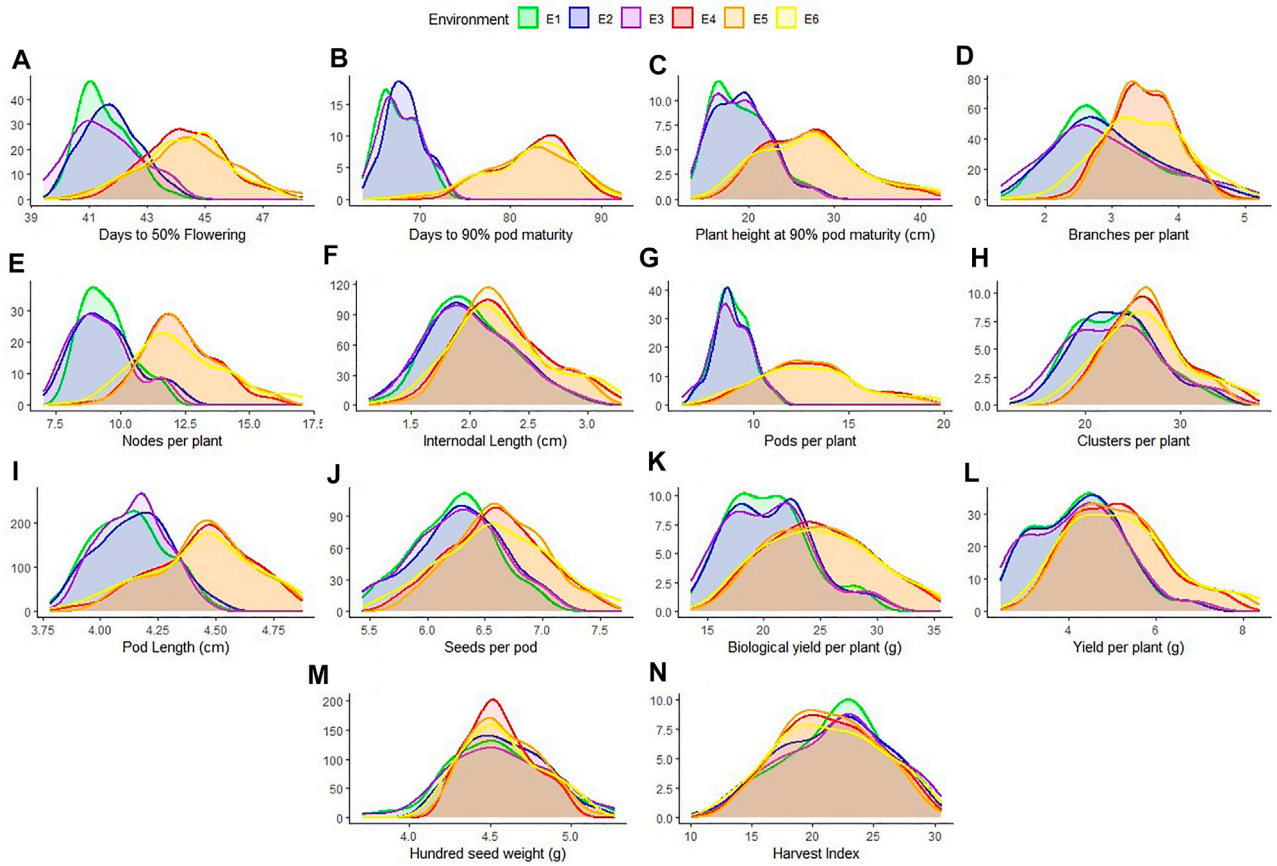


FIGURE 1 | Distribution of 14 characters measured for 100 blackgram germplasm lines across all environments- Ludhiana 2019 (E1); Ludhiana 2020 (E2); Ludhiana combined across years (E3); Gurdaspur 2019 (E4); Gurdaspur 2020 (E5) and Gurdaspur combined across years (E6).

Postulation of Candidate Genes and KEGG Pathway Analysis

Candidate genes were postulated using the functional gene annotations of *Vigna radiata* reference genome (www.ncbi.nlm.nih.gov/assembly/GCF_000741045.1). A window of 200-kb region, upstream and downstream of the associated SNPs was searched to identify already reported candidate genes related to different traits studied (Park et al., 2019). Selected candidate genes were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis using Omics box 2.0.36 combined pathway analysis plugin.

RESULTS

Phenotypic Evaluation

All the recorded traits showed high variability across the six given environments (**Figure 1**). All the traits followed normal distribution if considered in environments of each location however if compared across the two locations, the distribution of traits showed skewness. All the genotypes performed better in Gurdaspur than in Ludhiana. The negative skewness of trait data for the Gurdaspur location indicated an overall better performance of genotypes there, as compared to Ludhiana (**Figure 1**; **Table 1**). Since BLUPs accounted for the variation across the years for individual locations, the phenotypic evaluation is explained only for E3 and E6 to explain the overall variation at a particular location, for ease of understanding the variations (**Table 1**; **Supplementary Table S2**). DtM, and PL at Gurdaspur, and HI at Ludhiana were negatively skewed while all the other traits were either positively skewed or normally distributed (**Figure 1**). Genotypes sown in Gurdaspur showed delayed flowering (E6: 40.65–49.27 days) as compared to Ludhiana (E3: 39.42–47.08 days). The variation in DtM at Gurdaspur (E6: 68.88–90.73 days) were higher than at Ludhiana (E3: 63.62–73.33 ± 2.3 days) with delayed maturation at Gurdaspur. The observed range of CpP and PpP was 5.59–17.08, and 12.07–40.43, respectively, at Ludhiana (E3) whereas 7.62–23.35, and 17–50.28, respectively, at Gurdaspur (E6), indicating better plant phenology at Gurdaspur. The range of PL was slightly more in Gurdaspur (E6: 03.85–05.16 cm) than Ludhiana (E3: 03.66–04.46 cm), eventually leading to higher SpP in Gurdaspur (E6: 05.63–07.68) than Ludhiana (E3: 05.45–07.08). Larger variability along with higher values were observed for

TABLE 1 | Phenotypic evaluation of 100 *V. mungo* genotypes for 14 traits recorded at two locations of Ludhiana and Gurdaspur as BLUPs of 2 years.

Trait	Env	Mean ± SD	Range	Geno.Sign	LSD	CV	H^2
DtF	E3	41.81 ± 1.3	39.42–47.08	2E-28	01.97	03.61	0.85
	E6	44.57 ± 1.7	40.65–49.27	1E-26	01.81	02.64	0.91
DtM	E3	68.22 ± 2.3	63.62–73.33	2E-39	02.56	02.78	0.92
	E6	82.29 ± 4.7	68.88–90.73	1E-35	03.78	02.71	0.94
PHM	E3	20.93 ± 5.7	13.70–59.37	3E-82	02.85	09.77	0.99
	E6	29.90 ± 10.3	17.07–94.34	1E-73	05.66	12.60	0.98
BpP	E3	02.98 ± 0.9	01.33–05.05	2E-54	05.76	07.28	0.95
	E6	03.49 ± 0.7	01.82–05.86	2E-63	04.44	04.45	0.97
NpP	E3	09.84 ± 1.8	07.03–16.41	3E-41	01.59	12.00	0.92
	E6	12.74 ± 2.3	07.98–20.73	8E-43	02.14	11.57	0.93
IL	E3	02.18 ± 0.7	01.14–06.69	8E-71	00.37	12.19	0.98
	E6	02.35 ± 0.5	01.21–04.78	4E-47	00.45	13.37	0.94
CpP	E3	09.19 ± 1.7	05.59–17.08	2E-47	01.53	12.28	0.93
	E6	13.23 ± 3.2	07.62–23.35	6E-46	02.48	13.85	0.93
PpP	E3	23.89 ± 5.4	12.07–40.43	7E-46	04.80	14.45	0.93
	E6	27.43 ± 6.0	17.00–50.28	7E-46	04.93	13.32	0.92
PL	E3	04.19 ± 0.1	03.66–04.46	9E-16	00.48	05.02	0.81
	E6	04.53 ± 0.3	03.85–05.16	4E-55	00.32	04.94	0.95
SpP	E3	06.30 ± 0.4	05.45–07.08	3E-16	00.68	06.65	0.82
	E6	06.68 ± 0.5	05.63–07.68	6E-42	00.57	06.22	0.92
BYpP	E3	21.08 ± 4.0	13.57–30.83	4E-52	03.54	11.48	0.96
	E6	25.82 ± 5.3	14.94–45.85	2E-64	03.60	09.80	0.97
YpP	E3	04.45 ± 1.3	02.45–08.44	9E-62	00.72	11.45	0.96
	E6	05.28 ± 1.4	02.83–09.83	5E-57	00.82	11.35	0.96
HSW	E3	04.53 ± 0.3	03.71–05.26	5E-48	00.29	04.58	0.94
	E6	04.56 ± 0.3	04.12–05.33	4E-38	00.30	04.61	0.92
HI	E3	21.53 ± 4.6	09.49–30.55	2E-52	02.99	09.82	0.95
	E6	21.08 ± 4.3	11.92–29.62	4E-48	03.32	11.44	0.94

[a]*SD -standard deviation, LSD- least significant difference, CV- coefficient of variation, H^2—Broad sense heritability, Geno. Sign.-Genotype Significance.*
[b]*Ludhiana BLUPs, 2 years (E3); and Gurdaspur BLUPs, 2 years (E6).*
[c]*Days to 50% flowering (DtF); Days to 90% pod maturity (DtM); Plant height at 90% pod maturity (PHM); Branches per plant (BpP), Nodes per plant (NpP); Internodal length (IL), Clusters per plant (CpP); Pods per plant (PpP); Pod length (PL), Seeds per pod (SpP); Biological yield per plant (BYpP); Yield per plant (YpP); hundred seed weight (HSW) and Harvest index (HI).*

BYpP at Gurdaspur (E6: 14.94–45.85 g) relative to Ludhiana (E3: 13.57–30.83 g). At Ludhiana (E3), YpP ranged from 2.45 to 8.44 g whereas it varied from 2.83 to 9.83 g at Gurdaspur (E6). HSW ranged from 3.71 to 5.26 g at Ludhiana (E3) and 4.12–5.33 g at Gurdaspur (E6). The five best performing genotypes representing best of all the traits at Ludhiana and Gurdaspur are presented in **Supplementary Table S3**.

Higher broad sense heritability (H^2) for all traits under all environments suggested strong genetic control (**Table 1**). Highest H^2 was observed for PHM (0.99 and 0.98), whereas lowest for DtF (0.61 and 0.91) both at Ludhiana and Gurdaspur. CpP (E3—0.93; E6—0.93), PpP (E3—0.93; E6—0.92), PL (E3—0.81; E6—0.95), SpP (E3—0.82; E6—0.92), BYpP (E3—0.96; E6—0.97), YpP (E3—0.96; E6—0.96), HSW (E3—0.94; E6—0.92) and HI (E3—0.95; E4 - 0.94) also recorded high H^2.

SNP Calling

A total of 35, 49,948 raw physically mapped SNPs were obtained through GBS using the genome sequence of *Vigna radiata* as reference (Kang et al., 2014). Of these SNPs, 26, 39,464 were mapped onto 11 chromosomes while 9, 10,484 were mapped to non - chromosomal contigs. After filtering only 6,967 SNPs were retained of which 2,344 SNPs mapped to non-chromosomal contigs were removed and 4,623 on chromosomal regions used for further analysis (**Supplementary Table S4**, **Figure 2**). The highest density of SNPs was observed in Chr 4 (18.60 markers per Mb), whereas the lowest density was observed in Chr 3 (4.86) with an average density of 15.23 markers per Mb.

Population Structure and Phylogenetic Analysis

Four sub-populations (K = 4) were depicted by both the methods of second-order rate of change of the likelihood (**Figure 3A**) as well as cross-entropy value (**Figure 3B**). The graphical representation of four sub-populations against the admixture coefficient showed that more than 50% of the genotypes contributed to one sub population (**Figure 3C**). The allele frequency divergence was highest between sub-population two and four and the lowest between 3 and 4 (**Supplementary Table S5**). Average distances or expected heterozygosity of individuals within the same cluster were 0.3780 (sub-population 1), 0.2719 (sub-population 2), 0.0551 (sub-population 3) and 0.0250 (sub-population 4) (**Supplementary Table S5**). The mean Fst value of sub-populations 1, 2, 3 and four were 0.0203, 0.4537, 0.8040 and 0.9098, respectively (**Supplementary Table S5**). Multivariate analysis also classified the germplasm into four clusters affirming the results of structure analysis. The germplasm's phylogenetic structure using an unweighted

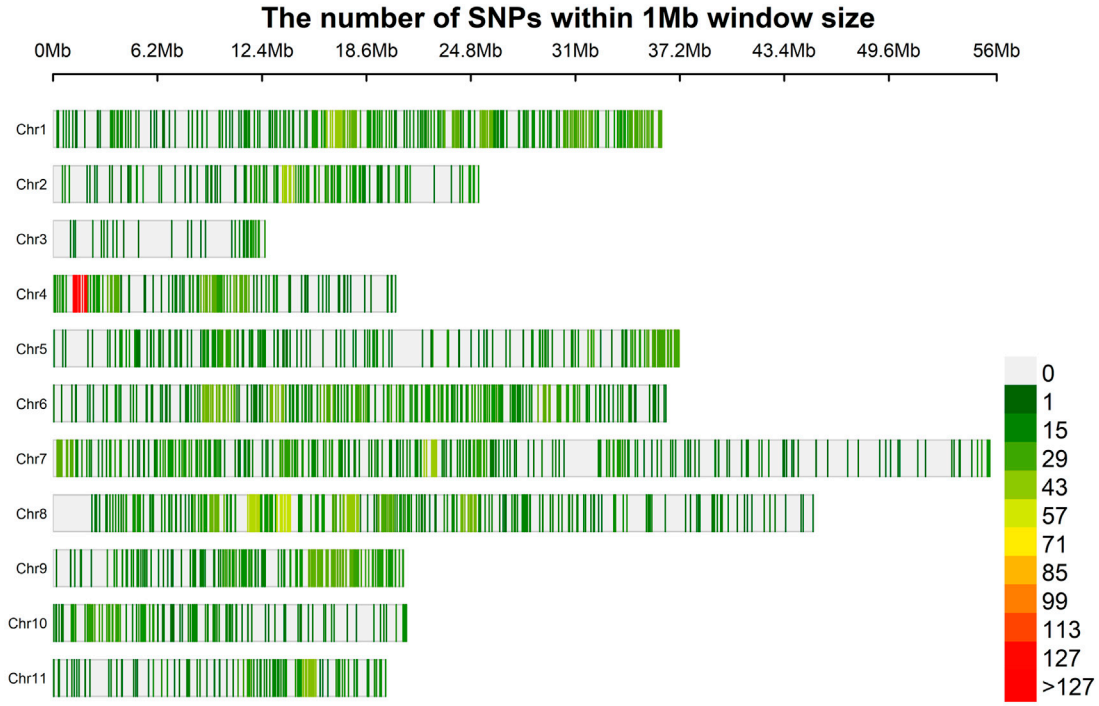


FIGURE 2 | Physical map of 4,623 SNPs identified by GBS of 100 blackgram germplasm lines showing all the 11 chromosomes. Physical position is also shown in million of base pairs (Mb) based. SNP density is also provided in colours Dark Green (1) to Red (127) to reveal the distribution among chromosomes.

neighbour-joining tree showed the distribution of the different genotypes among the sub-populations (**Figure 3D**).

Genome Wide Association Study

A total of 49 stable MTAs contributing to 42 QTLs were found to be significantly associated ($-\log_{10}$ *p*-value ≥ 3, PVE >10%) with 12 of the 14 traits studied, across the three environments of the each locations (**Figure 4**; **Table 2**; **Supplementary Table S6**). GWAS conducted using FarmCPU and MLM algorithms identified 31 and 27 QTLs, respectively of which 16 QTLs were common between two methods. However, only two MTAs *Q. PHM.4* & *Q. PHM.8* were significantly associated with PHM across both the

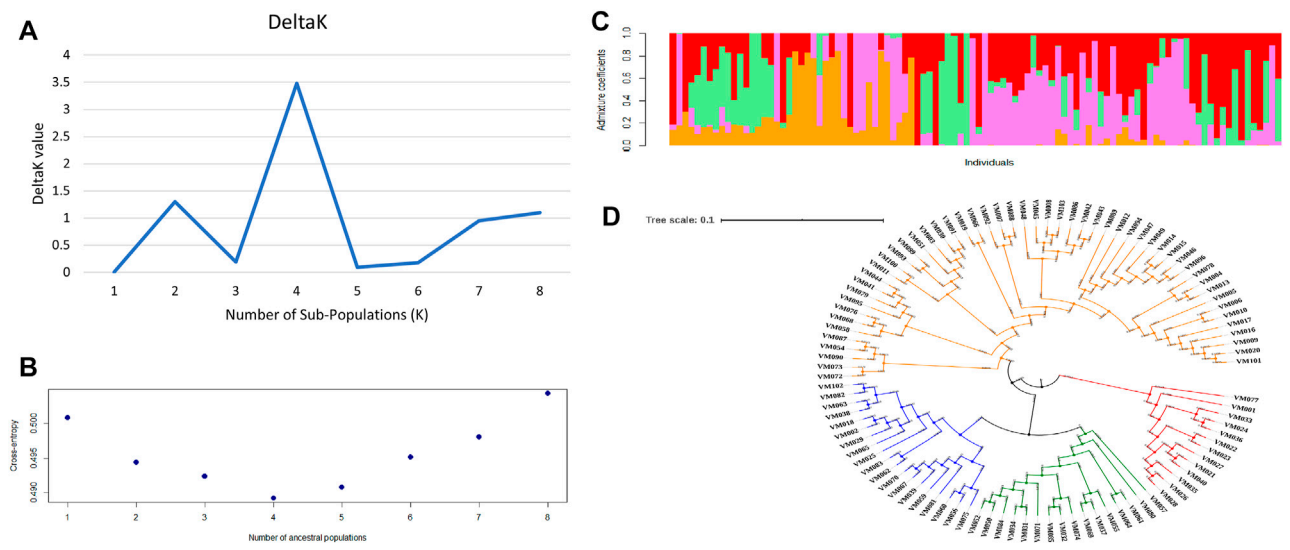


FIGURE 3 | Population structure of 100 blackgram germplasm lines using 4623 SNP markers **(A)** determination of number of sub-populations by DeltaK method by Evanno et al., 2005 **(B)** determination of number of sub-populations using cross entropy value method by Chan and Kroese, 2012 **(C)** population structure at k = 4 **(D)** phylogenetic analysis of 100 blackgram germplasm lines depicting four sub-populations.

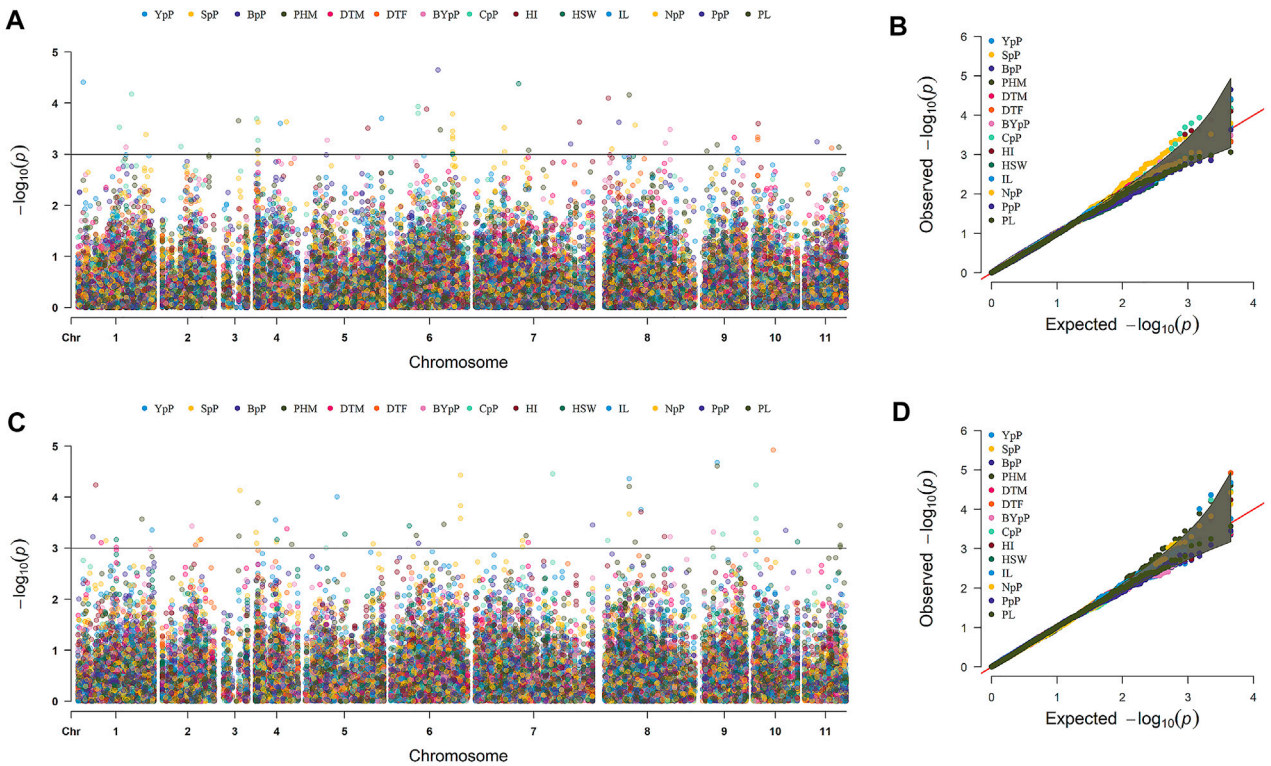


FIGURE 4 | Marker trait associations for different traits detected by genome wide association study across two different locations Ludhiana **(A,B)** and Gurdaspur **(C,D)** as manhattan plots and QQ-plots using 4623 SNP markers. #Days to 50% flowering (DtF); Days to 90% pod maturity (DtM); Plant height at 90% pod maturity (PHM); Nodes per plant (NpP); Internodal length (IL), Clusters per plant (CpP); Pods per plant (PpP); Seeds per pod (SpP); Biological yield per plant (BYpP); Yield per plant (YpP); hundred seed weight (HSW) and Harvest index (HI).

locations. The negative $\log_{10}$ *p*-value of QTLs ranged from 3.009 to 5.112, whereas the PVE ranged from 10.06–24.26%. Among 42 QTLs identified in the study, four QTLs were associated with YpP, eight with PHM, one with BpP, two with NpP, four with IL, five with CpP, one with PL, four with SpP two with BYpP, seven with HI and three with HSW. However, no significant associations were obtained for DtM and PpP. SNPs S8.1.13991269 and S8.1.19533014, on chromosome 8, were found to be associated with two traits each i.e., PHM, IL (*Q. PHM.8* and *Q. IL.8*) and YpP, HI (*Q.YpP.8* and *Q. HI.8.2*), respectively (**Figure 5**).

A few genomic regions harbored multiple significantly associated SNPs, as *Q. PHM.11.2* had four SNPs (-$\log_{10}$ *p*-value 3.0114–3.5826 and PVE 15.193–16.439%), *Q. NpP.6* (-$\log_{10}$ *p*-value 3.15–4.77, PVE 15.69–23.89) and *Q. CpP.10* both had three SNPs (-$\log_{10}$ *p*-value 3.11–4.66 and PVE 11.29–18.07%), whereas other QTLs were defined by a single SNP. Two QTLs for plant height at 90% pod maturity *Q. PHM.4* and *Q. PHM.8* showed consistent effect across all environments of both the locations. Among four QTLs identified for YpP, two (QTLs *Q. YpP.4.1* and *Q. YpP.8*) were detected at Ludhiana, whereas *Q. YpP.4.2* and *Q. YpP.5* were detected at Gurdaspur. Three QTLs for HSW were detected, two for Ludhiana and one for Gurdaspur (*Q.HSW.7*) which was detected by both the FarmCPU (-$\log_{10}$ *p*-value 4.38 and PVE 12.2%) and MLM (-$\log_{10}$ *p*-value 3.59 and PVE 12.34%) algorithms. Out of seven QTLs for HI, two were detected for Ludhiana (*Q. HI.1* and *Q. HI.8.2*) and five were detected for Gurdaspur (*Q. HI.5, Q. HI.6, Q. HI.7, Q. HI.8.1* and *Q. HI.10*). The phenotypic variation explained (PVE) by significant SNPs under FarmCPU method ranged from 10.751% (*Q.YpP.8*) to 22.292% (*Q.NpP.6*) at Ludhiana and 10.47% (*Q. PHM.6.1*) to 24.26% (*Q. CpP.1*) at Gurdaspur, whereas under MLM method, it ranged from 10.73% (*Q.SpP.8.1*) to 16.80% (*Q.PHM.8*) at Ludhiana and 10.08% (*Q.PHM.4*) to 14.69% (*Q.CpP.4*) at Gurdaspur.

Allelic Effects

Out of 49 SNPs/MTAs, 44 SNPs representing 37 QTLs showed significant difference, whereas five SNPs representing five QTLs were found to be non-significant (using the *t*-test statistic) for the respective traits (**Supplementary Table S7**, **Supplementary Figures S2A–I**. A total of 32, 32, 32, 28, 28 and 28 SNPs were statistically and significantly different for respective traits under E1, E2, E3, E4, E5 and E6, respectively. A total of 15 SNPs associated with 12 QTLs namely, *Q. IL.1.1*, *Q. IL.8*, *Q. NpP.6*, *Q. PHM.11.1*, *Q. PHM.11.2*, *Q. PHM.3.1*, *Q. PHM.3.2*, *Q. PHM.4*, *Q. PHM.6.1*, *Q. PHM.6.2*, *Q. PHM.8*, and *Q. SpP.8.2* were found to be significantly different in all of the six environments studied. The 28 SNP associations were significant in three of the six environments. The superior and inferior alleles along with allele

TABLE 2 | MTAs identified through Genome wide association study for 100 *Vigna mungo* germplasm lines in all the enviornments of two locations Ludhiana (L) or Gurdaspur (G) using FarmCPU and MLM algorithms.

MTA	Trait	Loc	Method	SNP	Chr	PosMb	-log_{10}P	MAF	Effect	PVE
Q.DtF.10	DtF	L	FarmCPU	S10.1.9527186	10	9.5272	4.9208	0.07	1.68	17.947
	DtF	L	MLM	S10.1.9527186	10	9.5272	4.1457	0.07	1.5481	16.803
Q.PHM.3.1	PHM	G	FarmCPU	S3.1.7993147	3	7.9931	3.6566	0.115	-6.3321	16.506
Q.PHM.3.2	PHM	L	FarmCPU	S3.1.8219594	3	8.2196	3.2337	0.15	-3.0667	19.919
Q.PHM.4	PHM	L	FarmCPU	S4.1.1259800	4	1.2598	3.8912	0.065	7.009	17.084
	PHM	G	FarmCPU	S4.1.1259800	4	1.2598	3.0761	0.065	10.8299	16.46
	PHM	G	MLM	S4.1.1259800	4	1.2598	3.1226	0.065	11.6496	10.076
Q.PHM.6.1	PHM	G	FarmCPU	S6.1.23358875	6	23.3589	3.4745	0.105	8.5575	10.469
Q.PHM.6.2	PHM	L	FarmCPU	S6.1.25029431	6	25.0294	3.4664	0.205	-3.3382	14.698
Q.PHM.8	PHM	L	FarmCPU	S8.1.13991269	8	13.9913	4.2062	0.055	7.0402	17.697
	PHM	G	FarmCPU	S8.1.13991269	8	13.9913	4.1599	0.055	12.2652	11.234
	PHM	L	MLM	S8.1.13991269	8	13.9913	3.3945	0.055	6.216	10.728
	PHM	G	MLM	S8.1.13991269	8	13.9913	3.8356	0.055	11.9305	13.02
Q.PHM.11.1	PHM	G	FarmCPU	S11.1.16313748	11	16.3137	3.1353	0.19	5.1138	15.263
Q.PHM.11.2	PHM	L	FarmCPU	S11.1.16898133	11	16.8981	3.055	0.17	4.4754	15.99
	PHM	L	FarmCPU	S11.1.16898169	11	16.8982	3.0204	0.165	4.4335	16.003
	PHM	L	FarmCPU	S11.1.16898170	11	16.8982	3.0204	0.165	4.4335	16.003
	PHM	L	FarmCPU	S11.1.16898225	11	16.8982	3.4443	0.155	4.6729	15.565
Q.BpP.6	BpP	G	MLM	S6.1.22219189	6	22.2192	3.7615	0.19	-0.6003	14.341
Q.NpP.4	NpP	G	FarmCPU	S4.1.1343019	4	1.343	3.6303	0.065	2.2705	15.003
	NpP	G	MLM	S4.1.1343019	4	1.343	3.4822	0.065	2.2851	10.85
Q.NpP.6	NpP	L	FarmCPU	S6.1.32970231	6	32.9702	3.5832	0.095	1.7559	18.857
	NpP	L	FarmCPU	S6.1.32970252	6	32.9703	3.8288	0.055	2.267	19.048
	NpP	L	FarmCPU	S6.1.32970258	6	32.9703	4.4295	0.085	2.0396	22.292
	NpP	L	MLM	S6.1.32970258	6	32.9703	3.4366	0.085	1.8555	12.377
Q.IL.1.1	IL	G	FarmCPU	S1.1.2636313	1	2.6363	4.4078	0.115	0.4871	14.843
	IL	G	MLM	S1.1.2636313	1	2.6363	3.7944	0.115	0.4871	14.693
Q.IL.1.2	IL	L	FarmCPU	S1.1.35117546	1	35.1175	3.356	0.06	0.7035	14.089
Q.IL.5	IL	L	MLM	S5.1.15033908	5	15.0339	3.4791	0.08	0.8146	13.589
Q.IL.8	IL	L	FarmCPU	S8.1.13991269	8	13.9913	4.3615	0.055	0.8372	14.825
	IL	L	MLM	S8.1.13991269	8	13.9913	3.6602	0.055	0.8254	14.476
Q.CpP.1	CpP	G	FarmCPU	S1.1.25545554	1	25.5456	4.1785	0.1	-2.1507	24.26
Q.CpP.4	CpP	G	MLM	S4.1.774872	4	0.7749	3.5768	0.07	-3.1131	10.448
Q.CpP.7	CpP	L	FarmCPU	S7.1.36781633	7	36.7816	4.4572	0.06	1.9962	15.922
	CpP	L	MLM	S7.1.36781633	7	36.7816	3.8541	0.06	1.9962	14.712
Q.CpP.9	CpP	L	FarmCPU	S9.1.9823250	9	9.8233	3.2717	0.065	1.6571	14.218
Q.CpP.10	CpP	L	FarmCPU	S10.1.1346198	10	1.3462	3.5792	0.185	0.7477	12.487
	CpP	L	MLM	S10.1.1346198	10	1.3462	3.1982	0.185	0.7477	11.66
	CpP	L	FarmCPU	S10.1.1346205	10	1.3462	4.2381	0.06	1.9314	16.817
	CpP	L	MLM	S10.1.1346205	10	1.3462	3.6948	0.06	1.9314	13.959
	CpP	L	FarmCPU	S10.1.1346249	10	1.3462	3.3004	0.095	1.4055	12.019
Q.PL.1	PL	L	MLM	S1.1.30381382	1	30.3814	3.2522	0.075	-0.1458	11.747
Q.SpP.3	SpP	L	FarmCPU	S3.1.8767931	3	8.7679	4.1296	0.09	0.4397	15.931
	SpP	L	MLM	S3.1.8767931	3	8.7679	3.7625	0.09	0.4589	15.654
Q.SpP.6	SpP	G	MLM	S6.1.29235889	6	29.2359	3.501	0.13	-0.4234	14.263
Q.SpP.8.1	SpP	L	MLM	S8.1.13990816	8	13.9908	3.0808	0.165	-0.3413	12.201
Q.SpP.8.2	SpP	G	MLM	S8.1.16767849	8	16.7678	3.1283	0.145	0.3782	12.395
Q.BYpP.8	BYpP	G	FarmCPU	S8.1.33339056	8	33.3391	3.4848	0.15	3.7945	13.775
Q.BYpP.9	BYpP	L	FarmCPU	S9.1.5335251	9	5.3353	3.3148	0.175	-2.0795	13.007
Q.YpP.4.1	YpP	L	MLM	S4.1.9615705	4	9.6157	3.2364	0.07	1.3938	12.669
Q.YpP.4.2	YpP	G	FarmCPU	S4.1.11912711	4	11.9127	3.6025	0.095	1.1945	12.541
	YpP	G	MLM	S4.1.11912711	4	11.9127	3.4516	0.095	1.214	11.667
Q.YpP.5	YpP	G	FarmCPU	S5.1.35989113	5	35.9891	3.7002	0.08	-1.2605	11.018
	YpP	G	MLM	S5.1.35989113	5	35.9891	3.3001	0.08	-1.2365	11.032
Q.YpP.8	YpP	L	FarmCPU	S8.1.19533014	8	19.533	3.7596	0.17	-1.0298	10.751
	YpP	L	MLM	S8.1.19533014	8	19.533	3.3417	0.17	-1.0257	13.184
Q.HI.1	HI	L	FarmCPU	S1.1.8549788	1	8.5498	4.2388	0.095	-4.7838	11.425
	HI	L	MLM	S1.1.8549788	1	8.5498	3.8081	0.095	-4.8079	15.933
Q.HI.5	HI	G	FarmCPU	S5.1.29623709	5	29.6237	3.5094	0.145	-3.4666	11.234
Q.HI.6	HI	G	MLM	S6.1.16902154	6	16.9022	3.5694	0.08	4.5656	14.564
Q.HI.7	HI	G	FarmCPU	S7.1.49416375	7	49.4164	3.6277	0.16	-3.4401	18.045
Q.HI.8.1	HI	G	MLM	S8.1.4181215	8	4.1812	3.3758	0.095	4.5191	13.586
Q.HI.8.2	HI	L	FarmCPU	S8.1.19533014	8	19.533	3.7094	0.17	-3.5888	12.072

(Continued on following page)

TABLE 2 | (*Continued*) MTAs identified through Genome wide association study for 100 *Vigna mungo* germplasm lines in all the enviornments of two locations Ludhiana (L) or Gurdaspur (G) using FarmCPU and MLM algorithms.

MTA	Trait	Loc	Method	SNP	Chr	PosMb	-log$_{10}$P	MAF	Effect	PVE
	HI	L	MLM	S8.1.19533014	8	19.533	3.2618	0.17	-3.5455	13.138
Q.HI.10	HI	G	MLM	S10.1.2365261	10	2.3653	3.1461	0.085	-4.2602	12.443
Q.HSW.6	HSW	L	FarmCPU	S6.1.8707985	6	8.708	3.4374	0.065	-0.3472	10.822
Q.HSW.7	HSW	G	FarmCPU	S7.1.20599131	7	20.5991	4.3788	0.08	0.2654	12.201
	HSW	G	MLM	S7.1.20599131	7	20.5991	3.5899	0.08	0.2455	12.339
Q.HSW.10	HSW	L	FarmCPU	S10.1.20952550	10	20.9526	3.1235	0.135	0.2536	11.731

[a]*Marker trait associations (MTAs, Environment (Env), Chromosome (Chr), Position in million basepairs (PosMb), minor allele frequency (MAF), Phenotypic variation explained in percentage (PVE).*

[b]*Days to 50% flowering (DtF); Plant height at 90% pod maturity (PHM); Branches per plant (BpP), Nodes per plant (NpP); Internodal length (IL), Clusters per plant (CpP); Pod length (PL), Seeds per pod (SpP); Biological yield per plant (BYpP); Yield per plant (YpP); Harvest index (HI) and Hundred seed weight (HSW).*

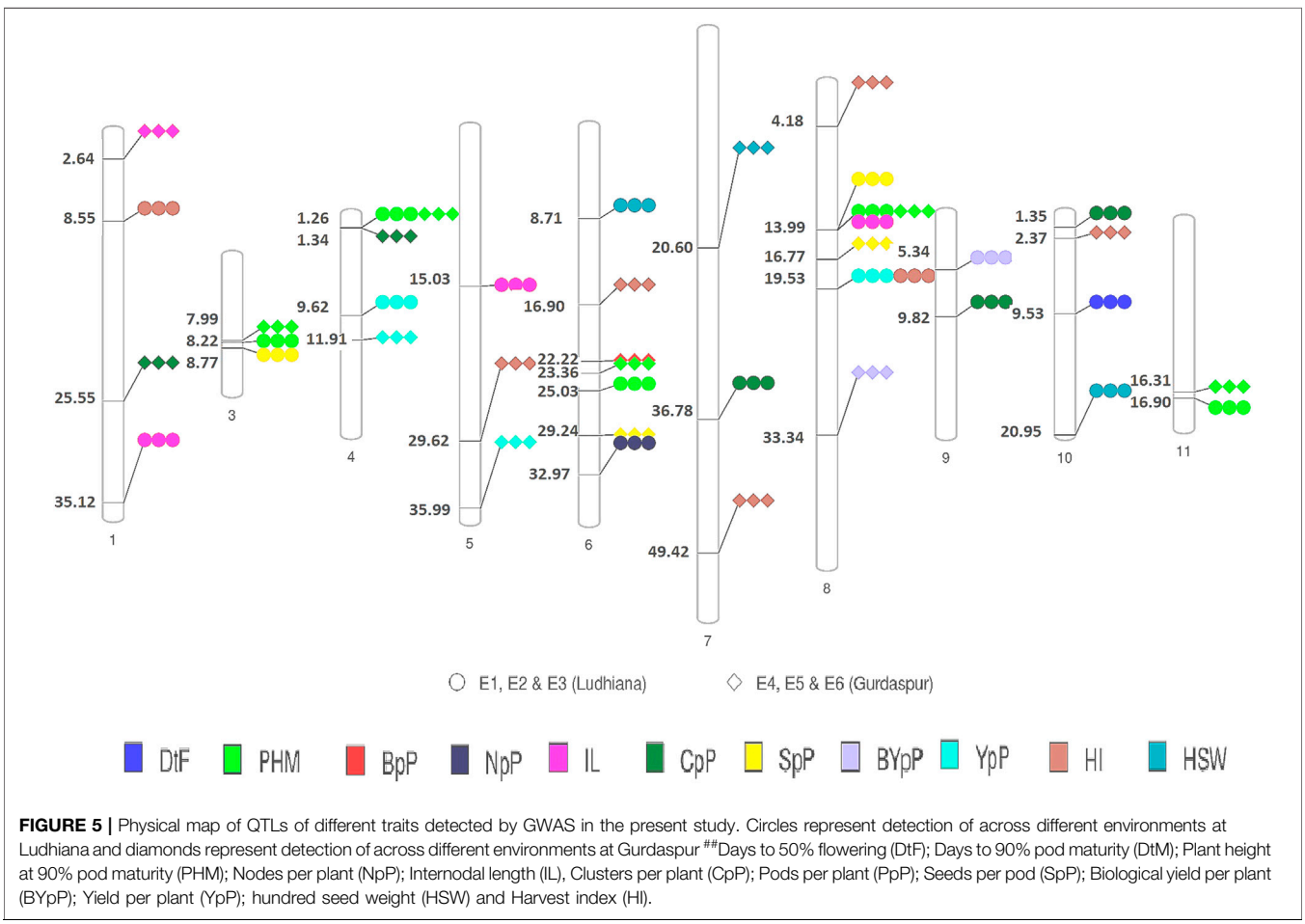


FIGURE 5 | Physical map of QTLs of different traits detected by GWAS in the present study. Circles represent detection of across different environments at Ludhiana and diamonds represent detection of across different environments at Gurdaspur [##]Days to 50% flowering (DtF); Days to 90% pod maturity (DtM); Plant height at 90% pod maturity (PHM); Nodes per plant (NpP); Internodal length (IL), Clusters per plant (CpP); Pods per plant (PpP); Seeds per pod (SpP); Biological yield per plant (BYpP); Yield per plant (YpP); hundred seed weight (HSW) and Harvest index (HI).

count (%) and significantly different mean values observed at Ludhiana (E3) and Gurdaspur (E6) and are presented in **Supplementary Table S8**. The allele with SNP GG associated with *Q. SpP.8.2* was found to be superior at Ludhiana, whereas alternative allele was found to be superior at Gurdaspur. Allelic effects of YpP, HI and HSW are presented in **Figure 6**. The significant phenotypic differences produced by superior and inferior alleles for YpP were 1.3 g at Ludhiana (E3) by *Q. YpP.4.1*; 1.38 g at Gurdaspur (E6) by *Q. YpP.4.2*; 1.26 g at Gurdaspur (E6) by *Q. YpP.5* and 0.94 g at Ludhiana (E3) by *Q. YpP.8*; whereas significant phenotypic differences produced by superior and inferior alleles for HI were 3.91% at Ludhiana (E3) by *Q. HI.1*; 3.19% at Gurdaspur (E6) by *Q. HI.5*; 3.89% at Gurdaspur (E6) by *Q. HI.6*; 4.38% at Gurdaspur (E6) by *Q. HI.7.1*; 3.26% at Gurdaspur (E6) by *Q. HI.8.1*; 3.43% at Ludhiana (E3) by *Q. HI.8.2* and 3.66% at Gurdaspur (E6) by *Q. HI.10*. Likewise, significant phenotypic differences produced by superior and inferior alleles for HSW were 0.32 g at Ludhiana (E3) by *Q.*

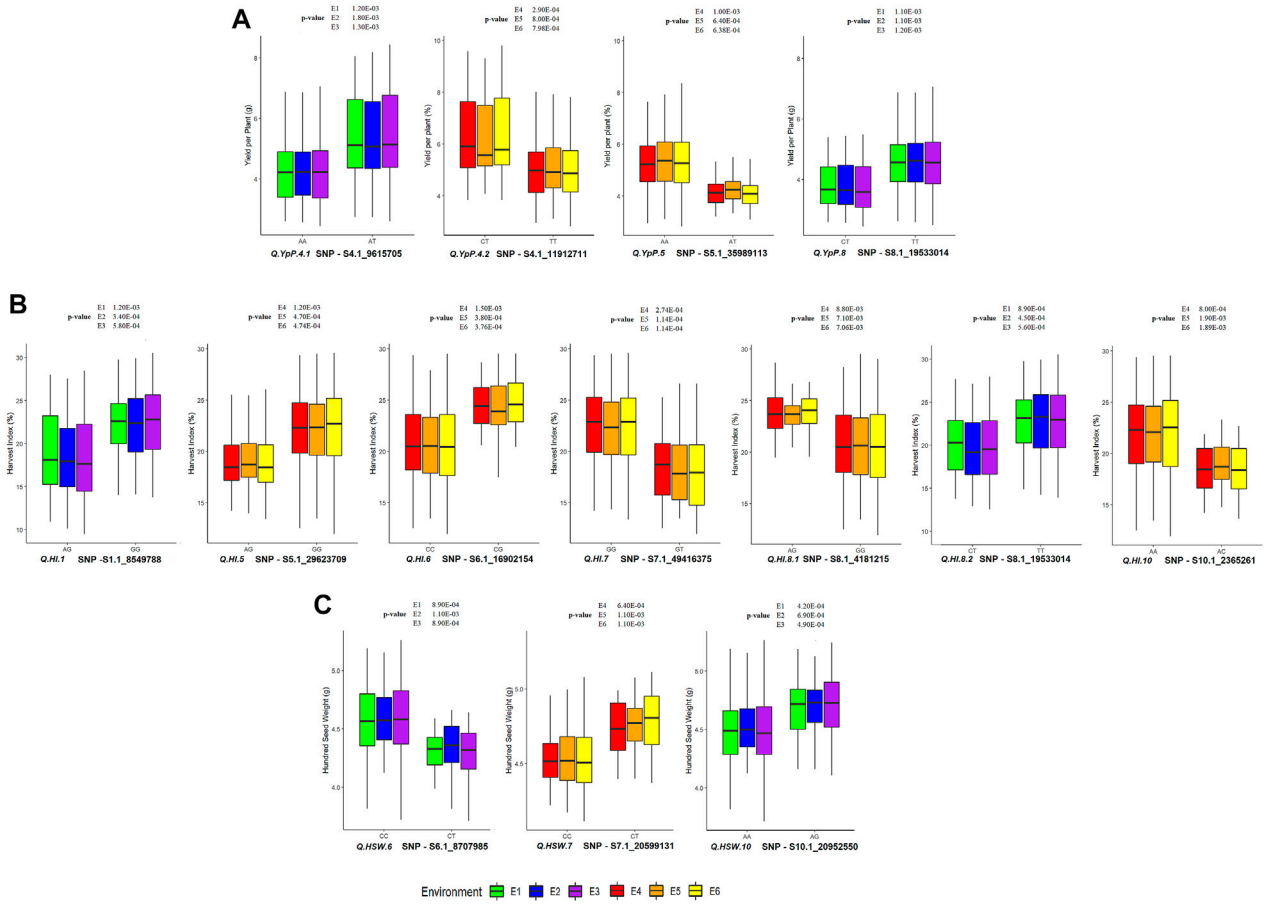


FIGURE 6 | Allelic effect of QTLs associated with **(A)** YpP (Yield per Plant); **(B)** HI (Harvest Index) and **(C)** HSW (Hundred Seed Weight) at Ludhiana and Gurdaspur. *p*-value is provided from *t*-test for the respective environment.

HSW.6, 0.25 g at Gurdaspur (E6) by *Q. HSW.7* and 0.24 g at Ludhiana (E3) by *Q. HSW.10*. A total of six genotypes (MASH218, IC274597, IC370938, IC557431, KUG673 and IC328783) were selected carrying superior alleles for all the traits under study and respective QTLs presented in **Supplementary Table S9**.

Postulation of Candidate Genes

A total of 50 genes for 24 QTLs were identified with different functions for different traits, whereas no known gene was found for the remaining 18 QTLs (**Table 3**). Maximum genes with known function linked to the trait were identified for PHM 15) followed by CpP (9), YpP (7), SpP (5), IL (5), HI (5), HSW (3), and DtF (1), whereas no genes with known function for BpP, NpP, PL and BYpP were identified.

The gene *LOC106774489* was found 73 kb upstream of the *Q. DtF.10* coding PHD finger-like domain-containing enzyme 5B. Nine genes *LOC106757287* (-100.237 kb), *LOC106757069* (-65.973 kb), *LOC106757136* (-47.55 kb), *LOC106756978* (-43.026 kb), *LOC111241394* (-33.113 kb), *LOC106757804* (-2.328 kb), *LOC106757666* (50.731 kb), *LOC106756983* (98.775 kb), *LOC106756984* (118.512 kb) coding for already known enzyme E3 ubiquitin-enzyme ligase MARCH1, bZIP transcription factor 53, enzyme trichome birefringence-like 6, histone-lysine N-methyltransferase EZ2-like, DELLA enzyme RGL1-like, DEAD-box ATP-dependent RNA helicase 24, probable WRKY transcription factor 23, gibberellin 2-beta-dioxygenase two and transcription factor JAMYB-like for plant height were located near *Q. PHM.3.1*.

For internodal length, two genes *LOC106766854* (-33.322 kb) and *LOC106765724* (-33.083 kb), were found close to the *Q. IL.1.1* coding for pectate lyase-like and pectate lyase enzymes, respectively. Another gene *LOC106762425* (-99.425 kb-), with enzyme cytokinin hydroxylase, was found in proximity with *Q. IL.1.2*. The QTL *Q. IL.5* was near the genes *LOC106760883* (-189.1 kb) and *LOC106762422* (16.957 kb) coding for purine permease 1 and ethylene-responsive transcription factor RAP2-4 enzymes. For clusters per plant, three genes *LOC106768944*, *LOC106760064* and *LOC106760083* were in proximity of *Q. CpP.1* (-108.532kb, -32.774 and 68.491 kb) coding SKP1-interacting partner 15, receptor-like protein 12 and polygalacturonase-like enzyme, respectively.

For seeds per pod, three genes *LOC106757271* (71.381 kb), *LOC106757661* (95.044 kb) and *LOC106756994* (100.043 kb) were found close to the *Q. SpP.3* with enzyme galactinol

TABLE 3 | List of candidate genes with their previously reported biological pathway function obtained in the putative QTL regions.

QTL	Trait	Gene ID	Dist (Kb)	Function
Q.DtF.10	DtF	*LOC106774489*	73.878	PHD finger-like domain-containing protein 5B
Q.PHM.3.1	PHM	*LOC106757287*	-100.237	E3 ubiquitin-protein ligase MARCH1
		LOC106757069	-65.973	bZIP transcription factor 53
		LOC106757136	-47.55	protein trichome birefringence-like 6
		LOC106756978	-43.026	histone-lysine N-methyltransferase EZ2-like
		LOC111241394	-33.113	DELLA protein RGL1-like
		LOC106757804	-2.328	DEAD-box ATP-dependent RNA helicase 24
		LOC106757666	50.731	probable WRKY transcription factor 23
		LOC106756983	98.775	gibberellin 2-beta-dioxygenase 2
		LOC106756984	118.512	transcription factor JAMYB-like
Q.PHM.4	PHM	*LOC106759452*	-178.201	tropinone reductase homolog
		LOC106758588	-69.349	tropinone reductase homolog At5g06060
Q.PHM.6.2	PHM	*LOC106764341*	-82.244	steroid 5-alpha-reductase DET2
Q.PHM.11.1	PHM	*LOC106777611*	-171.802	squamosa promoter-binding-like protein 14
		LOC106777237	-112.535	cytochrome P450 71D11
		LOC106777539	-82.108	pentatricopeptide repeat-containing protein At3g48810
Q.IL.1.1	IL	*LOC106766854*	-33.322	pectate lyase-like
		LOC106765724	-33.083	pectate lyase
Q.IL.1.2	IL	*LOC106762425*	-99.425	cytokinin hydroxylase
Q.IL.5	IL	*LOC106760883*	-189.1	purine permease 1
		LOC106762422	16.957	ethylene-responsive transcription factor RAP2-4
Q.CpP.1	CpP	*LOC106768944*	-108.532	SKP1-interacting partner 15
		LOC106760064	-32.774	receptor-like protein 12
		LOC106760083	68.491	polygalacturonase-like
Q.CpP.7	CpP	*LOC106765735*	-115.443	protein POLLEN DEFECTIVE IN GUIDANCE 1
		LOC106766388	42.507	LRR receptor-like serine/threonine-protein kinase RPK2
Q.CpP.9	CpP	*LOC111242573*	-79.009	eukaryotic translation initiation factor 3 subunit H-like
		LOC106773784	166.402	MLO-like protein 1
Q.CpP.10	CpP	*LOC106775061*	131.587	DDB1- and CUL4-associated factor 13
Q.SpP.3	SpP	*LOC106757271*	71.381	galactinol synthase 2
		LOC106757661	95.044	Golgi apparatus membrane protein-like protein ECHIDNA
		LOC106756994	100.043	alkaline/neutral invertase A, mitochondrial
Q.SpP.6	SpP	*LOC106765120*	64.962	dihydrofolate synthetase
Q.SpP.8.2	SpP	*LOC106770299*	147.684	ethylene-responsive transcription factor 1B-like
Q.YpP.4.1	YpP	*LOC106759105*	7.311	myb-related protein 305-like
Q.YpP.5	YpP	*LOC106761836*	-177.436	CLAVATA3/ESR (CLE)-related protein 5-like
		LOC106760678	-107.67	transcription factor PCF5
		LOC106762074	-50.026	sodium/calcium exchanger NCL
		LOC106759995	141.103	basic leucine zipper 34 isoform X1
Q.YpP.8	YpP	*LOC111242272*	116.438	alpha-mannosidase-like
		LOC106771274	129.392	putative 12-oxophytodienoate reductase 11
Q.HI.1	HI	*LOC106758323*	-174.544	UV-B-induced protein, chloroplastic isoform X1
Q.HI.5	HI	*LOC106760579*	-189.197	cytochrome P450 CYP72A219-like
Q.HI.7	HI	*LOC106769438*	105.344	protein root UVB sensitive 6
Q.HI.8.2	HI	*LOC111242272*	116.438	alpha-mannosidase-like
		LOC106771274	129.392	putative 12-oxophytodienoate reductase 11
Q.HSW.6	HSW	*LOC106764301*	-15.52	putative pentatricopeptide repeat-containing protein At1g12700, mitochondrial isoform X1
		LOC106765194	12.233	peroxidase 4
Q.HSW.10	HSW	*LOC106776199*	-175.528	bromodomain-containing protein 4B

[a]*Distance of 5′ position of the gene from SNP, identified associated with the QTL, where–sign shows that the gene was located upstream of the SNP, and +sign shows that the gene was located downstream of the SNP.*

[b]*Days to 50% flowering (DtF); Plant height at 90% pod maturity (PHM); internodal length (IL), clusters per plant (CpP); seeds per pod (SpP); yield per plant (YpP); harvest index (HI) and hundred seed weight (HSW).*

synthase 2, Golgi apparatus membrane enzyme-like enzyme ECHIDNA and alkaline/neutral invertase A respectively. A gene *LOC106765120* (64.962 kb) with enzyme dihydrofolate synthetase, was found in the vicinity of *Q. SpP.6*. Another ethylene-responsive transcription factor 1B-like enzyme coding gene *LOC106770299* (147.684 kb) was found close to the *Q. SpP.8.2*. For yield per plant, a gene *LOC106759105* (7.311kb) with MYB-related protein 305-like enzyme was close to the QTL: *Q. YpP.4.1*. Four genes *LOC106761836* (-177.436 kb), *LOC106760678* (-107.67 kb), *LOC106762074* (-50.026 kb) and *LOC106759995* (141.103 kb) coding CLAVATA3/ESR (CLE)-related protein 5-like, transcription factor PCF5, sodium/calcium exchanger NCL and basic leucine zipper 34 isoform X1 enzymes were in vicinity of *Q. YpP.5*.

Two genes *LOC111242272* (116.438 kb) and *LOC106771274* (129.392 kb) with alpha-mannosidase-like and putative 12-

oxophytodienoate reductase 11 enzymes was close to *Q. YpP.8*. For harvest index, gene *LOC106758323* (-174.544 kb) coding for UV-B-induced protein At3g17800, chloroplastic isoform X1 was related to *Q. HI.1*. The gene *LOC106760579* (-189.197 kb) was found in the proximity of *Q. HI.5* coding cytochrome P450 CYP72A219-like enzyme. The gene *LOC106769438* with protein root UVB sensitive six enzyme function was lying 105.344 kb of *Q. HI.7*. Two genes *LOC111242272* (116.438 kb) and *LOC106771274* (129.392 kb), with enzymes alpha-mannosidase-like and putative 12-oxophytodienoate reductase 11 respectively, were close to the *Q. HI.8.2*. Two genes *LOC106764301* (-15.52 kb) and *LOC106765194* (12.233 kb) with enzymes putative pentatricopeptide repeat-containing protein At1g12700 and peroxidase four, respectively, were close to the *Q. HSW.6*. The gene *LOC106776199* (-175.528 kb) encoding bromodomain-containing protein 4B enzyme was related to *Q. HSW.10*. The nodes per plant (NpP) had two QTLs, but no genes already known for nodes per plant around the vicinity of 200 kb of significant QTLs were found.

KEGG Pathway Analysis

Some of the genomic regions significantly associated with the trait loci (**Supplementary Table S10**) such as Gibberellin 2-beta-dioxygenase (*LOC106756983*), cytochrome P450 CYP72A219 (*LOC106760579*) are involved in Diterpenoid biosynthesis (ko00904) pathway (**Supplementary Figure S3**). PHD finger-like domain-containing protein 5B (*LOC106774489*), and DEAD-box ATP-dependent RNA helicase 24 (*LOC106757804*) play role in Spliceosome (ko03040) process. Moreover, histone-lysine N-methyltransferase EZ2 found to be involved in Lysine degradation (Amino acid metabolism) ko0310. Besides that, pectate lyase-like (*LOC106766854*), pectate lyase (*LOC106765724*), and polygalacturonase-like (*LOC106760083*) have been found to play important role in Pentose and glucuronate interconversions (ko00040). LRR receptor-like serine/threonine-protein kinase RPK2 (*LOC106766388*), alkaline/neutral invertase A (*LOC106756994*) participates in Starch and sucrose metabolism (k000500). Galactinol synthase 2 (*LOC106757271*), alkaline/neutral invertase A (*LOC106756994*) have been shown to play role in Galactose metabolism (ko00052). Peroxidase 4 (*LOC106765194*) was found to play role in Phenylpropanoid biosynthesis (ko00940) (**Supplementary Figure S4**).

DISCUSSION

Blackgram is one of the most popular pulses in Southeast Asia, with India contributing 54% of world production (Singh et al., 2016). Inspite of having high nutritional value, short duration, and photo insensitivity the crop has not been exploited to its full potential. Multiple biotic and abiotic stresses and narrow genetic base of this crop is major hindrance in its expansion. Thus, a systematic approach is required to exploit untapped genetic diversity present within the country so that the germplasm could be exploited for improving breeding potential. . Present study is an effort to exploit the collection of diverse blackgram genotypes for important yield related component traits.

Phenotypic Evaluation

Environment factors highly influence the complex traits; therefore, the material was replicated both in time and space. Previous studies have also reported a wide range of phenotypic variability for traits such as PHM, BpP, IL, CpP, PpP and HI (Panigrahi et al., 2014; Panda et al., 2017; Senthamizhselvi et al., 2019). High heritability was observed for all the traits indicating high genetic control and an effective phenotypic selection for these traits. Different studies have reported high broad-sense heritability for traits, i.e., DtF, DtM, PHM, CpP, PL and YpP (Panigrahi et al., 2014; Kumar et al., 2015); BYpP, HI and HSW (Rolaniya al., 2017; Kuralarasan et al., 2018). High heritability with high selection intensity helps breeders to shorten the breeding cycles of the program leading to faster and higher genetic gains. Many genotypes harbored combination of superior alleles for different yield related traits. For instance, the genotype IC274597 for CpP, PpP, SpP and BYpP; IC370938 for PpP, YpP, HSW and HI; KUG673 for DtF, DtM and HI; IC328783 for PHM, NpP and BYpP; These genotypes can be used in further breeding programs to improve desirable characters.

Population Structure

Ad-hoc delta K and Cross entropy values suggested presence of four sub-populations in the blackgram germplasm which was further supported by phylogenetic analysis which in turn indicated significant diversity in the panel. High Fst values in 2, 3 and 4 sub-populations indicated them to be highly differentiated. A diverse germplasm panel can be a good source for a wide range of traits for breeding and research purpose (Govindaraj et al., 2015). Modelling of genetic structures as covariates helps in controlling the false positives while conducting GWAS. This is the first report on population structure analysis in blackgram; however, four sub-populations in mungbean germplasm have been reported earlier (Breria et al., 2020).

Genome Wide Association Studies (GWAS)

GWAS for 14 yield associated traits identified 49 SNPs contributing to 42 QTLs to be strongly associated with 12 traits except DtM and PpP. Among 42 significant genomic regions identified in the study, the number of QTLs for each trait were as YpP (4), DtF (1), PHM (8), BpP (1), NpP (2), IL (4), CpP (5), PL (1), SpP (4), BYpP (2), HI (7) and HSW (3). High $-\log_{10} p$-value and PVE suggested presence of major QTLs i.e. *Q. DtF.10*, *Q. PHM.3.2*, *Q. NpP.6*, *Q. IL.8*, *Q. CpP.10*, *Q. SpP.3*, *Q. BYpP.9*, *Q. YpP.8,* and *Q. HI.8.2* at Ludhiana, whereas *Q. PHM.3.1*, *Q. PHM.6.1*, *Q. NpP.4*, *Q. IL.1.1*, *Q. CpP.4*, *Q. SpP.6, Q. BYpP.8, Q. YpP.4.2,* and *Q. HI.6* at Gurdaspur were found significant for trait regulation. QTLs *Q. PHM.8,* and *Q. IL.8* were found to be located at same position on eighth chromosome controlling PHM and IL, respectively. The pleiotropy of increased PHM with increased IL has been earlier reported in faba bean (Hughes et al., 2020).

Allelic Effects

The significant allelic effects of MTAs suggested selection of superior alleles could substantially lead to significant improvement of crop. At Ludhiana, QTL *Q. DtF.10* with superior allele TT resulted in earlier flowering with a mean of 41.42 days as compared to 43.05 with alternative allele. *Q. PHM.3.1* with superior allele AA recorded higher mean plant height of 28.54 cm, and 43.23 cm as compared to 19.21 cm, and 27.21 cm with alternative allele GG at Ludhiana, and Gurdaspur respectively. Superior allele TT of QTL *Q. IL.8* decreased internode length with mean value of 2.01 cm, whereas alternate allele observed higher internodal length with mean value of 2.83 cm. For clusters per plant, *Q. CpP.1* with superior allele AA exhibited mean of 18.84 clusters, whereas alternate allele TT exhibited mean of 12.73 clusters with The SNP S6.1.29235837 of *Q. SpP.6* having allele AA produced more average seeds per pod of 6.71 in contrast to 6.45 by alternate allele. Higher yield per plant with average of 5.47 g was observed by presence of AA allele of QTL *Q. YpP.5* while 4.21 g yield was observed with alternate allele. *Q. HI.1* with allele GG exhibited an increased mean value of harvest index (22.53%) as compared to a lower harvest index due to alternate allele (18.62%). *Q. HI.8.2* with allele TT showed a higher harvest index (22.87%) as compared to alternate allele (19.44%). Allelic effects with superior alleles and alternative alleles have been reported for significantly associated markers for root, nutrient uptake and yield related traits in rice (Subedi et al., 2019), for spike ethylene and spike dry weight in wheat (Valluru et al., 2017), and for yield related and heat tolernce traits in wheat (Dhillon et., 2021).

Postulation of Genes

Blackgram is a highly self-pollinated crop and with the given narrow genetic base, is expected to have large LD blocks and large chunks of haplotypes being transferred over the generations without recombination. Keeping this in view, the SNPs identified for traits can be searched upstream and downstream for candidate genes governing those traits (Dhillon et al., 2020). A total of 50 genes for 24 QTLs were identified with different functions with respect to different traits. The QTL *Q. DtF.10* identified for days to flowering was present in the vicinity of the gene coding plant homeodomain finger-like domain-containing enzyme 5B can be a putative gene as PHD finger-like genes are reported to delay flowering Arabidopsis (Greb et al., 2007).

The enzymes encoded by genes found for QTLs of PHM were involved in controlling the plant height as supported by earlier studies of enzyme E3 ubiquitin-enzyme ligase MARCH1 in rice (Hu et al., 2013), bZIP transcription factor 53 in soybean (Ali et al., 2016), enzyme trichome birefringence-like six in Rice (Gao et al., 2017), histone-lysine N-methyltransferase EZ2-like (https://www.uniprot.org/uniprot/O23372), DELLA enzyme RGL1-like in Arabidopsis (Serrano-Mislata et al., 2017), DEAD-box ATP-dependent RNA helicase 24 in rice (Wang et al., 2016), probable WRKY transcription factor 23 in rice (Cai et al., 2014), gibberellin 2-beta-dioxygenase two in Arabidopsis (Rosin et al., 2003), transcription factor JAMYB-like in rice (Zhang et al., 2017), acyl transferase four in rice (Basu et al., 2017), tropinone reductase homolog (Stead, 1989), in grapevine (Guillaumie et al., 2020), steroid 5-alpha-reductase DET2 in soybean (Huo et al., 2018), squamosa promoter-binding-like enzyme 14 in rice (Lu et al., 2013), cytochrome P450 71D11 in cucumber (Wang et al., 2017), pentatricopeptide repeat-containing enzyme At3g48810 in Arabidopsis (Lee et al., 2019), and pentatricopeptide repeat-containing enzyme At1g31430 in Arabidopsis (Lee et al., 2019).

For internodal length, genes coding for pectate lyase-like (PLL), and pectate lyase (PL) in rice (Leng et al., 2017), cytokinin hydroxylase in Arabidopsis (Kiba et al., 2013), purine permease 1 in cotton (Wang et al., 2020), ethylene-responsive transcription factor RAP2-4 enzymes Arabidopsis (Hinz et al., 2010) have been earlier reported. For clusters per plant, genes SKP1-interacting partner 15 (Lu et al., 2016), receptor-like protein 12 (Wang et al., 2008), polygalacturonase-like enzyme (Xiao et al., 2014) in Arabidopsis, probable trehalose-phosphate phosphatase C isoform X2 in *Nelumbo nucifera* (Jin et al., 2016), lysine-specific demethylase JMJ25 isoform X1 in Arabidopsis (Jiang et al., 2007), pollen defective in guidance 1 (Li et al., 2011), LRR receptor-like serine/threonine-protein kinase RPK2 enzymes in Arabidopsis (Mizuno et al., 2007), eukaryotic translation initiation factor 3 subunit H-like in Arabidopsis (Zhou et al., 2014), MLO-like protein 1 reported in peach (Ruperti et al., 2002), and DDB1- and CUL4-associated factor 13 enzyme (Bjerkan and Grini, 2013) has been earlier reported to be involved in controlling cell elongation and flower development.

For seeds per pod, galactinol synthase two in Chickpea (Salvi et al., 2016), Golgi apparatus membrane enzyme-like enzyme ECHIDNA in Arabidopsis (Jia et al., 2018), 60S ribosomal protein L28-2 in soybean (Jones et al., 2020), dihydrofolate synthetase in Arabidopsis (Corral et al., 2018), and ethylene-responsive transcription factor 1B-like enzyme in brassica (Kaur et al., 2020) were known to have a role in increasing number of seeds per pod/siliqua. For yield per plant, myb-related protein 305-like in Arabidopsis (Ambawat et al., 2013), CLAVATA3/ESR (CLE)-related protein 5-like in Arabidopsis (Fletcher, 2018), transcription factor PCF5 in wheat (Zhao et al., 2018), sodium/calcium exchanger NCL in soybean (Liu Y. et al., 2016), basic leucine zipper 34 isoform X1 enzymes in wheat (Sornaraj et al., 2016), alpha-mannosidase-like in wheat (Dal Cortivo et al., 2020), and putative 12-oxophytodienoate reductase 11 in wheat (Pigolev et al., 2018) were previously known for yield improvement. For harvest index, UV-B-induced protein At3g17800 in basil (Dou et al., 2019), cytochrome P450 CYP72A219-like enzyme in Arabidopsis (Bak et al., 2011), root UV-B sensitive in wheat (Agrawal et al., 2004), alpha-mannosidase-like in wheat (Dal Cortivo et al., 2020), and putative 12-oxophytodienoate reductase 11 in wheat (Pigolev et al., 2018) are known for triggering reductions in biomass, yield and harvest index.

The genes detected for HSW, pentatricopeptide repeat-containing protein At1g80550, have been earlier reported for kernel development in maize (Dai et al., 2018; Ren et al., 2019); peroxidase four in soybean (Zhang et al., 2015). The clusters per plant had several QTLs, but no known genes were identified in 200 kb genomic region.

KEGG Pathway Analysis

KEGG analysis revealed role of candidate genes in biological pathways related to respective traits. Gibberellins (GAs) are endogenous phytohormones that are involved in the regulation of the life cycle of plants. It has been identified that the locus encoding gibberellin 2-beta-dioxygenase/GA 2-oxidase present in vicinity of SNP S3.1.7993147 on chromosome 3 significantly associated with plant height at 90% pod maturity participates in Diterpenoid biosynthesis (ko00904) pathway. The role of this locus in the regulation of plant growth in rice has been demonstrated by Sakamoto et al., 2001. KEGG pathway established role of another gene coding for steroid 5-alpha-reductase DET2 with QTL associated with PHM on chromosome six to participate in Brassinosteroid biosynthesis. Ortholog of this gene in cotton (*Gossypium hirsutum* L.), GhDET2 along with BRs are known to play a crucial role in the initiation and elongation of cotton fiber cells (Luo et al., 2007). Similarly, DET2 steroid 5d-reductase in Arabidopsis catalyzes a major rate-limiting in brassinosteroid (BR) biosynthesis (Nakaya et al., 2002). Additionally, Cytochrome P450 CYP72A219-like has also been reported to participate in Diterpenoid biosynthesis pathway (Bathe and Tissier, 2019). Using KEGG pathway analysis, function of only two QTLs could be established with diterpenoid pathway and brassinosteroid biosynthesis pathway. However, the function of remaining significant genomic regions could not be established using this analysis.

CONCLUSION

GWAS analysis led to identification of novel MTA and few putative candidate genes. Though candidate genes need to be examined further and detailed investigations would validate their roles in governing agronomically important traits but the MTA will really help in selecting the genotypes with positive associations. The information derived from this study can be used in the generation of SNP based molecular markers to select traits of interest and accelerate blackgram breeding programme with a better ideotype. Since the blackgram is neglected crop in term of number of genetic and genomic resources, the current study, first of its kind in this crop will definitely open new avenues for broadening its base.

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: NCBI Bioproject, accession number PRJNA802066.

AUTHOR CONTRIBUTIONS

SdK concept design and coordination of research work and manuscript writing, LS, RG, SD, and AKu contributed to phenotyping; LS, GD, AKa, PM, and SbK contributed to data analysis; GD, PM, and SdK supervised and edited the manuscript; All authors read and approved the manuscript.

ACKNOWLEDGMENTS

Authors are thankful to Kuldeep Singh, Cluster Leader-Genebank, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) for useful guidance and support. This work was supported by the Department of Biotechnology (DBT) under the grant number BT/Ag/Network/Pulses-I/2017-18.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.849016/full#supplementary-material

REFERENCES

Agrawal, S. B., Rathore, D., and Singh, A. (2004). Effects of Supplemental Ultraviolet-B and Mineral Nutrients on Growth, Biomass Allocation and Yield of Wheat (*Triticum aestivum* L.). *Trop. Ecol.* 45, 315–325.

Ali, Z., Sarwat, S. S., Karim, I., Faridi, R., Jaskani, M. J., and Khan, A. A. (2016). Functions of Plant's Bzip Transcription Factors. *Pakjas* 53, 303–314. doi:10.21162/PAKJAS/16.2043

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

Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB Transcription Factor Genes as Regulators for Plant Responses: An Overview. *Physiol. Mol. Biol. Plants* 19, 307–321. doi:10.1007/s12298-013-0179-1

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

Basu, D., Majhi, B. B., Sripriya, R., and Veluthambi, K. (2017). Agrobacterium T-DNA Insertion in the Rice DWARF SHOOT and DEFECTIVE PANICLE1 (DSDP1) Gene Causes a Severe Dwarf Phenotype, Reduces Plant Vigour, and Affects Seed Germination. *J. Plant Biochem. Biotechnol.* 26, 395–405. doi:10.1007/s13562-017-0400-3

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *J. Stat. Soft.* 67, 1–48. doi:10.18637/jss.v067.i01

Bathe, U., and Tissier, A. (2019). Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity. *Phytochemistry* 161, 149–162. doi:10.1016/j.phytochem.2018.12.003

Bhandari, A., Sandhu, N., Bartholome, J., Cao-Hamadoun, T.-V., Ahmadi, N., Kumari, N., et al. (2020). Genome-Wide Association Study for Yield and Yield Related Traits under Reproductive Stage Drought in a Diverse Indica-Aus Rice Panel. *Rice* 13. doi:10.1186/s12284-020-00406-3

Bjerkan, K. N., and Grini, P. E. (2013). TheArabidopsisDDB1 Interacting Protein WDR55 Is Required for Vegetative Development. *Plant Signal. Behav.* 8, e25347–6. doi:10.4161/psb.25347

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

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

Breria, C. M., Hsieh, C. H., Yen, J.-Y., Nair, R., Lin, C.-Y., Huang, S.-M., et al. (2020). Population Structure of the World Vegetable Center Mungbean Mini Core Collection and Genome-wide Association Mapping of Loci Associated with Variation of Seed Coat Luster. *Trop. Plant Biol.* 13, 1–12. doi:10.1007/s12042-019-09236-0

Cai, Y., Chen, X., Xie, K., Xing, Q., Wu, Y., Li, J., et al. (2014). Dlf1, a WRKY Transcription Factor, Is Involved in the Control of Flowering Time and Plant Height in Rice. *PLoS One* 9, e102529–13. doi:10.1371/journal.pone.0102529

Chaitieng, B., Kaga, A., Tomooka, N., Isemura, T., Kuroda, Y., and Vaughan, D. A. (2006). Development of a Black Gram [Vigna Mungo (L.) Hepper] Linkage Map and its Comparison with an Azuki Bean [Vigna Angularis (Willd.) Ohwi and Ohashi] Linkage Map. *Theor. Appl. Genet.* 113, 1261–1269. doi:10.1007/s00122-006-0380-5

Chan, J. C. C., and Kroese, D. P. (2012). Improved Cross-Entropy Method for Estimation. *Stat. Comput.* 22, 1031–1040. doi:10.1007/s11222-011-9275-7

Corral, M. G., Haywood, J., Stehl, L. H., Stubbs, K. A., Murcha, M. W., and Mylne, J. S. (2018). Targeting Plant DIHYDROFOLATE REDUCTASE with Antifolates and Mechanisms for Genetic Resistance. *Plant J.* 95, 727–742. doi:10.1111/tpj.13983

Dai, D., Luan, S., Chen, X., Wang, Q., Feng, Y., Zhu, C., et al. (2018). Maize Dek37 Encodes a P-type PPR Protein that Affects Cis-Splicing of Mitochondrial Nad2 Intron 1 and Seed Development. *Genetics* 208, 1069–1082. doi:10.1534/genetics.117.300602

Dal Cortivo, C., Ferrari, M., Visioli, G., Lauro, M., Fornasier, F., Barion, G., et al. (2020). Effects of Seed-Applied Biofertilizers on Rhizosphere Biodiversity and Growth of Common Wheat (*Triticum aestivum* L.) in the Field. *Front. Plant Sci.* 11, 1–14. doi:10.3389/fpls.2020.00072

Das, D., Baruah, I. K., Panda, D., Paswan, R. R., Acharjee, S., and Sarmah, B. K. (2021). Bruchid Beetle Ovipositioning Mediated Defense Responses in Black Gram Pods. *BMC Plant Biol.* 21, 1–22. doi:10.1186/s12870-020-02796-4

Dey, S., Chowardhara, B., Regon, P., Kar, S., Saha, B., and Panda, S. K. (2020). Iron Deficiency in Blackgram (Vigna Mungo L.): Redox Status and Antioxidant Activity. *Plant Biosyst. - Int. J. Deal. all Aspects Plant Biol.*, 1, 16. doi:10.1080/11263504.2020.1866093

Dhillon, G. S., Das, N., Kaur, S., Srivastava, P., Bains, N. S., and Chhuneja, P. (2021). Marker Assisted Mobilization of Heat Tolerance QTLs from Triticum Durum-Aegilops Speltoides Introgression Lines to Hexaploid Wheat. *Indian J. Genet.* 81 (2), 186–198. doi:10.31742/IJGPB.81.2.2

Dhillon, G. S., Kaur, S., Das, N., Singh, R., Poland, J., Kaur, J., et al. (2020). QTL Mapping for Stripe Rust and Powdery Mildew Resistance in Triticum Durum-Aegilops Speltoides Backcross Introgression Lines. *Plant Genet. Resour.* 18, 211–221. doi:10.1017/S1479262120000222

Dou, H., Niu, G., and Gu, M. (2019). Pre-harvest UV-B Radiation and Photosynthetic Photon Flux Density Interactively Affect Plant Photosynthesis, Growth, and Secondary Metabolites Accumulation in Basil (Ocimum Basilicum) Plants. *Agronomy* 9, 434. doi:10.3390/agronomy9080434

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

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

Fletcher, J. (2018). The CLV-WUS Stem Cell Signaling Pathway: A Roadmap to Crop Yield Optimization. *Plants* 7, 87–11. doi:10.3390/plants7040087

Gao, Y., He, C., Zhang, D., Liu, X., Xu, Z., Tian, Y., et al. (2017). Two Trichome Birefringence-like Proteins Mediate Xylan Acetylation, Which Is Essential for Leaf Blight Resistance in Rice. *Plant Physiol.* 173, 470–481. doi:10.1104/pp.16.01618

Ghafoor, A., Sharif, A., Ahmad, Z., Zahid, M. A., and Rabbani, M. A. (2001). Genetic Diversity in Blackgram (Vigna Mungo L. Hepper). *Field Crops Res.* 69, 183–190. doi:10.1016/S0378-4290(00)00141-6

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of Genetic Diversity Assessment in Crop Plants and its Recent Advances: An Overview of its Analytical Perspectives. *Genet. Res. Int.* 2015, 1–14. doi:10.1155/2015/431487

Greb, T., Mylne, J. S., Crevillen, P., Geraldo, N., An, H., Gendall, A. R., et al. (2007). The PHD Finger Protein VRN5 Functions in the Epigenetic Silencing of Arabidopsis FLC. *Curr. Biol.* 17, 73–78. doi:10.1016/j.cub.2006.11.052

Guillaumie, S., Decroocq, S., Ollat, N., Delrot, S., Gomès, E., and Cookson, S. J. (2020). Dissecting the Control of Shoot Development in Grapevine: Genetics and Genomics Identify Potential Regulators. *BMC Plant Biol.* 20, 1–15. doi:10.1186/s12870-020-2258-0

Gupta, S. K., Souframanien, J., and Gopalakrishna, T. (2008). Construction of a Genetic Linkage Map of Black gram,Vigna mungo(L.) Hepper, Based on Molecular Markers and Comparative Studies. *Genome* 51, 628–637. doi:10.1139/G08-050

Hinz, M., Wilson, I. W., Yang, J., Buerstenbinder, K., Llewellyn, D., Dennis, E. S., et al. (2010). ArabidopsisRAP2.2: An Ethylene Response Transcription Factor that Is Important for Hypoxia Survival. *Plant Physiol.* 153, 757–772. doi:10.1104/pp.110.155077

Hu, X., Qian, Q., Xu, T., Zhang, Y. e., Dong, G., Gao, T., et al. (2013). The U-Box E3 Ubiquitin Ligase TUD1 Functions with a Heterotrimeric G α Subunit to Regulate Brassinosteroid-Mediated Growth in Rice. *PLoS Genet.* 9, e1003391. doi:10.1371/journal.pgen.1003391

Hughes, J., Khazaei, H., and Vandenberg, A. (2020). Genetics of Height and Branching in Faba Bean (Vicia faba). *Agronomy* 10, 1191. doi:10.3390/agronomy10081191

Huo, W., Li, B., Kuang, J., He, P., Xu, Z., and Wang, J. (2018). Functional Characterization of the Steroid Reductase Genes GmDET2a and GmDET2b from Glycine Max. *Ijms* 19, 726–821. doi:10.3390/ijms19030726

Hwang, E.-Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A Genome-wide Association Study of Seed Protein and Oil Content in Soybean. *BMC Genomics* 15, 1. doi:10.1186/1471-2164-15-1

Ikram, M., Han, X., Zuo, J.-F., Song, J., Han, C.-Y., Zhang, Y.-W., et al. (2020). Identification of QTNs and Their Candidate Genes for 100-seed Weight in Soybean (Glycine Max L.) Using Multi-Locus Genome-wide Association Studies. *Genes* 11, 714–722. doi:10.3390/genes11070714

Jia, P.-F., Xue, Y., Li, H.-J., and Yang, W.-C. (2018). Golgi-localized LOT Regulates Trans -Golgi Network Biogenesis and Pollen Tube Growth. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12307–12312. doi:10.1073/pnas.1809206115

Jiang, D., Yang, W., He, Y., and Amasino, R. M. (2007). ArabidopsisRelatives of the Human Lysine-specific Demethylase1 Repress the Expression ofFWAandFLOWERING LOCUS Cand Thus Promote the Floral Transition. *Plant Cell.* 19, 2975–2987. doi:10.1105/tpc.107.052373

Jin, Q., Hu, X., Li, X., Wang, B., Wang, Y., Jiang, H., et al. (2016). Genome-wide Identification and Evolution Analysis of Trehalose-6-Phosphate Synthase Gene Family in *Nelumbo nucifera*. *Front. Plant Sci.* 7, 1–12. doi:10.3389/fpls.2016.01445

Jones, S. I., Hunt, M. R., and Vodkin, L. O. (2020). An Embryo Lethal Transgenic Line Manifests Global Expression Changes and Elevated Protein/oil Ratios in Heterozygous Soybean Plants. *PLoS One* 15, e0233721–23. doi:10.1371/journal.pone.0233721

Kang, Y., Kim, S., and Kim, M. (2014). Genome Sequence of Mungbean and Insights Into Evolution Within Vigna Species. *Nat. Commun.* 5 (5443), 1–9. doi:10.1038/ncomms6443

Kaur, J., Akhatar, J., Goyal, A., Kaur, N., Kaur, S., Mittal, M., et al. (2020). Genome Wide Association Mapping and Candidate Gene Analysis for Pod Shatter Resistance in Brassica Juncea and its Progenitor Species. *Mol. Biol. Rep.* 47, 2963–2974. doi:10.1007/s11033-020-05384-9

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

Kumar, G. V., Vanaja, M., Lakshmi, N. J., and Maheswari, M. (2015). Studies on Variability, Heritability and Genetic Advance for Quantitative Traits in Black Gram [Vigna mungo(L.) Hepper]. *Agricul. Res. Jour.* 52, 28. doi:10.5958/2395-146x.2015.00056.3

Kuralarasan, V., Vanniarajan, C., Kanchana, S., Veni, K., and Lavanya, A. (2018). Genetic Divergence, Heritability and Genetic Advance in Mutant Lines of Urdbean [Vigna Mungo (L.) Hepper. *Lr* 41, 833–836. doi:10.18805/LR-3794

Laird, N. M., and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*. 38, 963–974. doi:10.2307/2529876

Lee, K., Park, S. J., Han, J. H., Jeon, Y., Pai, H.-S., and Kang, H. (2019). A Chloroplast-Targeted Pentatricopeptide Repeat Protein PPR287 Is Crucial for Chloroplast Function and Arabidopsis Development. *BMC Plant Biol.* 19, 1–10. doi:10.1186/s12870-019-1857-0

Leng, Y., Yang, Y., Ren, D., Huang, L., Dai, L., Wang, Y., et al. (2017). A Rice PECTATE LYASE-LIKE Gene Is Required for Plant Growth and Leaf Senescence. *Plant Physiol.* 174, 1151–1166. doi:10.1104/pp.16.01625

Letunic, I., and Bork, P. (2016). Interactive Tree of Life (iTOL) V3: an Online Tool for the Display and Annotation of Phylogenetic and Other Trees. *Nucleic Acids Res.* 44, W242–W245. doi:10.1093/nar/gkw290

Li, H.-J., Xue, Y., Jia, D.-J., Wang, T., hi, D.-Q., Liu, J., et al. (2011). POD1 Regulates Pollen Tube Guidance in Response to Micropylar Female Signaling and Acts in Early Embryo Patterning inArabidopsis. *Plant Cell.* 23, 3288–3302. doi:10.1105/tpc.111.088914

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

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, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016a). 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

Liu, Y., Yu, L., Qu, Y., Chen, J., Liu, X., Hong, H., et al. (2016b). GmSALT3, Which Confers Improved Soybean Salt Tolerance in the Field, Increases Leaf Cl-Exclusion Prior to Na+ Exclusion but Does Not Improve Early Vigor under Salinity. *Front. Plant Sci.* 7, 1–14. doi:10.3389/fpls.2016.01485

Lorenz, A. J., Hamblin, M. T., and Jannink, J.-L. (2010). Performance of Single Nucleotide Polymorphisms versus Haplotypes for Genome-wide Association Analysis in Barley. *PLoS One* 5, e14079. doi:10.1371/journal.pone.0014079

Lu, D., Ni, W., Stanley, B. A., and Ma, H. (2016). Proteomics and Transcriptomics Analyses of Arabidopsis Floral Buds Uncover Important Functions of ARABIDOPSIS SKP1-LIKE1. *BMC Plant Biol.* 16, 1–18. doi:10.1186/s12870-015-0571-9

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

Lukoki, L., Marechal, R., and Otoul, E. (1980). Les ancetres sauvages des haricots cultives: Vigna radiata (L.) Wilczek et V. mungo (L.) Hepper. *Bull. Jard. Bot. Natl. Belg./Bull. Natl. Plantentuin Belg.* 50, 385–391. doi:10.2307/3667837

Luo, M., Xiao, Y., Li, X., Lu, X., Deng, W., Li, D., et al. (2007). GhDET2, a Steroid 5α-Reductase, Plays an Important Role in Cotton Fiber Cell Initiation and Elongation. *Plant J.* 51, 419–430. doi:10.1111/j.1365-313X.2007.03144.x

Mizuno, S., Osakabe, Y., Maruyama, K., Ito, T., Osakabe, K., Sato, T., et al. (2007). Receptor-like Protein Kinase 2 (RPK 2) Is a Novel Factor Controlling Anther Development in *Arabidopsis thaliana*. *Plant J.* 50, 751–766. doi:10.1111/j.1365-313X.2007.03083.x

Muthusamy, S., and Pandiyan, M. (2018). Realization of Facts and Profiteering of Black Gram through Different Breeding Methods. Available at: https://www.researchgate.net/publication/327386671 (Accessed March 15, 2021).

Nakaya, M., Tsukaya, H., Murakami, N., and Kato, M. (2002). Brassinosteroids Control the Proliferation of Leaf Cells of *Arabidopsis thaliana*. *Plant Cell. Physiol.* 43, 239–244. doi:10.1093/pcp/pcf024

Noble, T. J., Tao, Y., Mace, E. S., Williams, B., Jordan, D. R., Douglas, C. A., et al. (2018). Characterization of Linkage Disequilibrium and Population Structure in a Mungbean Diversity Panel. *Front. Plant Sci.* 8, 2102. doi:10.3389/fpls.2017.02102

Panda, D. P., Lenka, D., Dash, A. P., Tripathy, S. K., Behera, C., and Baisakh, B. (2017). Genetic Variability and Heritability Studies in Relation to Seed Yield and its Component Traits in Black Gram (*Vigna mungo* L). *Biosci. Trends*, 10 (6), 1412–1414.

Panigrahi, K. K., Mohanty, A., and Baisakh, B. (2014). Genetic Divergence, Variability and Character Association in Landraces of Blackgram (Vigna Mungo [l.] Hepper) from Odisha. *J. Crop Weed* 10, 155–165.

Park, C., Dwiyanti, M. S., Nagano, A. J., Liu, B., Yamada, T., and Abe, J. (2019). Identification of Quantitative Trait Loci for Increased α-tocopherol Biosynthesis in Wild Soybean Using a High-Density Genetic Map. *BMC Plant Biol.* 19, 1–15. doi:10.1186/s12870-019-2117-z

Pigolev, A., Miroshnichenko, D., Pushin, A., Terentyev, V., Boutanayev, A., Dolgov, S., et al. (2018). Overexpression of Arabidopsis OPR3 in Hexaploid Wheat (*Triticum aestivum* L.) Alters Plant Development and Freezing Tolerance. *Ijms* 19, 3989. doi:10.3390/ijms19123989

Plekhanova, E., Vishnyakova, M. A., Bulyntsev, S., Chang, P. L., Carrasquilla-Garcia, N., Negash, K., et al. (2017). Genomic and Phenotypic Analysis of Vavilov's Historic Landraces Reveals the Impact of Environment and Genomic Islands of Agronomic Traits. *Sci. Rep.* 7. doi:10.1038/S41598-017-05087-5

Pootakham, W., Nawae, W., Naktang, C., Sonthirod, C., Yoocha, T., Kongkachana, W., et al. (2021). A Chromosome-scale Assembly of the Black Gram (Vigna Mungo) Genome. *Mol. Ecol. Resour.* 21, 238–250. doi:10.1111/1755-0998.13243

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

Raizada, A., and Souframanien, J. (2021). SNP Genotyping and Diversity Analysis Based on Genic-SNPs through High Resolution Melting (HRM) Analysis in Blackgram [Vigna Mungo (L.) Hepper]. *Genet. Resour. Crop Evol.* 68, 1331–1343. doi:10.1007/s10722-020-01064-6

Ren, R. C., Lu, X., Zhao, Y. J., Wei, Y. M., Wang, L. L., Zhang, L., et al. (2019). Pentatricopeptide Repeat Protein DEK40 Is Required for Mitochondrial Function and Kernel Development in Maize. *J. Exp. Bot.* 70, 6163–6179. doi:10.1093/jxb/erz391

Rolaniya, D. K., Jinjwadiya, M. K., Meghawal, D. R., and Lal, G. M. (2017). Studies on Genetic Variability in Blackgram (Vigna Mungo (L.) Hepper). *J. Pharmacogn. Phytochem.* 7, 1506–1508.

Rosin, F. M., Hart, J. K., Horner, H. T., Davies, P. J., and Hannapel, D. J. (2003). Overexpression of a Knotted-like Homeobox Gene of Potato Alters Vegetative Development by Decreasing Gibberellin Accumulation. *Plant Physiol.* 132, 106–117. doi:10.1104/pp.102.015560.ser

Ruperti, B., Bonghi, C., Ziliotto, F., Pagni, S., Rasori, A., Varotto, S., et al. (2002). Characterization of a Major Latex Protein (MLP) Gene Down-Regulated by Ethylene during Peach Fruitlet Abscission. *Plant Sci.* 163, 265–272. doi:10.1016/S0168-9452(02)00094-8

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, 8014–8018. doi:10.1073/PNAS.81.24.8014

Sahu, B., Thakur, C. L., Markam, S. K., and Nag, U. K. (2020). Performance of Frontline Demonstrations on Productivity and Profitability of Black Gram (Vigna Mungo) in Rainfed Condition of Uttar-Bastar Kanker. *Int. J. Curr. Microbiol. App. Sci.* 9, 1898–1903. doi:10.20546/ijcmas.2020.904.224

Sakamoto, T., Kobayashi, M., Itoh, H., Tagiri, A., Kayano, T., Tanaka, H., et al. (2001). Expression of a Gibberellin 2-oxidase Gene Around the Shoot Apex Is Related to Phase Transition in Rice. *Plant Physiol.* 125, 1508–1516. doi:10.1104/pp.125.3.1508

Salvi, P., Saxena, S. C., Petla, B. P., Kamble, N. U., Kaur, H., Verma, P., et al. (2016). Differentially Expressed Galactinol Synthase(s) in Chickpea Are Implicated in Seed Vigor and Longevity by Limiting the Age Induced ROS Accumulation. *Sci. Rep.* 6, 1–15. doi:10.1038/srep35088

Scherer, A., and Christensen, G. B. (2016). Concepts and Relevance of Genome-wide Association Studies. *Sci. Prog.* 99, 59–67. doi:10.3184/003685016X14558068452913

Senthamizhselvi, S., Muthuswamy, A., and Shunmugavalli, N. (2019). Genetic Variability, Correlation and Path Coefficient Analysis for Yield and Yield Components in Blackgram (Vigna Mungo (L.) Hepper). *Electron. Journ. Plan. Breed.* 10, 1600–1605. doi:10.5958/0975-928X.2019.00206.0

Serrano-Mislata, A., Bencivenga, S., Bush, M., Schiessl, K., Boden, S., and Sablowski, R. (2017). DELLA Genes Restrict Inflorescence Meristem Function Independently of Plant Height. *Nat. Plants* 3, 749–754. doi:10.1038/s41477-017-0003-y

Singh, D. P., Singh, B. B., and Pratap, A. (2016). Genetic Improvement of Mungbean and Urdbean and Their Role in Enhancing Pulse Production in India. *Ind. Jrnl. Gen. Plnt. Bree.* 76, 550–567. doi:10.5958/0975-6906.2016.00072.9

Singh, L., Kumar, A., Kaur, S., and Gill, R. K. (2020). Multivariate Analysis of Yield Contributory Traits for Selection Criteria in Urdbean (V. Mungo L. Hepper). *Ejpb* 11, 1134–1142. doi:10.37992/2020.1104.183

Sokolkova, A., Burlyaeva, M., Valiannikova, T., Vishnyakova, M., Schafleitner, R., Lee, C.-R., et al. (2020). Genome-wide Association Study in Accessions of the

Mini-Core Collection of Mungbean (Vigna Radiata) from the World Vegetable Gene Bank (Taiwan). *BMC Plant Biol.* 20, 363. doi:10.1186/S12870-020-02579-X

Somta, P., Chen, J., Yundaeng, C., Yuan, X., Yimram, T., Tomooka, N., et al. (2019). Development of an SNP-Based High-Density Linkage Map and QTL Analysis for Bruchid (Callosobruchus Maculatus F.) Resistance in Black Gram (Vigna Mungo (L.) Hepper). *Sci. Rep.* 9, 3930. doi:10.1038/s41598-019-40669-5

Sornaraj, P., Luang, S., Lopato, S., and Hrmova, M. (2016). Basic Leucine Zipper (bZIP) Transcription Factors Involved in Abiotic Stresses: A Molecular Model of a Wheat bZIP Factor and Implications of its Structure in Function. *Biochimica Biophysica Acta (BBA) - General Subj.* 1860, 46–56. doi:10.1016/j.bbagen.2015.10.014

Stead, A. D. (1989). Book Reviews. *Ann. Bot.* 64, 369–370. doi:10.1093/oxfordjournals.aob.a087853

Subedi, S. R., Sandhu, N., Singh, V. K., Sinha, P., Kumar, S., Singh, S. P., et al. (2019). Genome-wide Association Study Reveals Significant Genomic Regions for Improving Yield, Adaptability of Rice under Dry Direct Seeded Cultivation Condition. *BMC Genomics* 20, 471. doi:10.1186/s12864-019-5840-9

Torkamaneh, D., Chalifour, F.-P., Beauchamp, C. J., Agrama, H., Boahen, S., Maaroufi, H., et al. (2020). Genome-wide Association Analyses Reveal the Genetic Basis of Biomass Accumulation under Symbiotic Nitrogen Fixation in African Soybean. *Theor. Appl. Genet.* 133, 665–676. doi:10.1007/s00122-019-03499-7

Valluru, R., Reynolds, M. P., Davies, W. J., and Sukumaran, S. (2017). Phenotypic and Genome-wide Association Analysis of Spike Ethylene in Diverse Wheat Genotypes under Heat Stress. *New Phytol.* 214, 271–283. doi:10.1111/nph.14367

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation Sequencing Technologies and Their Implications for Crop Genetics and Breeding. *Trends Biotechnol.* 27, 522–530. doi:10.1016/J.TIBTECH.2009.05.006

Wang, D., Qin, B., Li, X., Tang, D., Zhang, Y. e., Cheng, Z., et al. (2016). Nucleolar DEAD-Box RNA Helicase TOGR1 Regulates Thermotolerant Growth as a Pre-rRNA Chaperone in Rice. *PLoS Genet.* 12, e1005844–23. doi:10.1371/journal.pgen.1005844

Wang, G., Ellendorff, U., Kemp, B., Mansfield, J. W., Forsyth, A., Mitchell, K., et al. (2008). A Genome-wide Functional Investigation into the Roles of Receptor-like Proteins in Arabidopsis. *Plant Physiol.* 147, 503–517. doi:10.1104/pp.108.119487

Wang, H., Li, W., Qin, Y., Pan, Y., Wang, X., Weng, Y., et al. (2017). The Cytochrome P450 Gene CsCYP85A1 Is a Putative Candidate for Super Compact-1 (Scp-1) Plant Architecture Mutation in Cucumber (Cucumis Sativus L.). *Front. Plant Sci.* 8, 1–13. doi:10.3389/fpls.2017.00266

Wang, L., Yin, Y., Wang, L.-F., Wang, M., Zhao, M., Tian, Y., et al. (2020). Transcriptome Profiling of the Elongating Internode of Cotton (Gossypium Hirsutum L.) Seedlings in Response to Mepiquat Chloride. *Front. Plant Sci.* 10, 1–18. doi:10.3389/fpls.2019.01751

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. New York: Springer-Verlag. AvailableAt: https://ggplot2.tidyverse.org (accessed January 05, 2021).

Wu, X., Islam, A. S. M. F., Limpot, N., Mackasmiel, L., Mierzwa, J., Cortés, A. J., et al. (2020). Genome-Wide SNP Identification and Association Mapping for Seed Mineral Concentration in Mung Bean (Vigna Radiata L.). *Front. Genet.* 11, 1–17. doi:10.3389/fgene.2020.00656

Xiao, C., Somerville, C., and Anderson, C. T. (2014). POLYGALACTURONASE INVOLVED IN EXPANSION1 Functions in Cell Elongation and Flower Development in Arabidopsis. *Plant Cell.* 26, 1018–1035. doi:10.1105/tpc.114.123968

Xiong, H., Guo, H., Zhou, C., Guo, X., Xie, Y., Zhao, L., et al. (2019). A Combined Association Mapping and T-Test Analysis of SNP Loci and Candidate Genes Involving in Resistance to Low Nitrogen Traits by a Wheat Mutant Population. *PLoS One* 14, e0211492–15. doi:10.1371/journal.pone.0211492

Xu, P., Wu, X., Muñoz-Amatriaín, M., Wang, B., Wu, X., Hu, Y., et al. (2017). Genomic Regions, Cellular Components and Gene Regulatory Basis Underlying Pod Length Variations in Cowpea (V. unguiculataL. Walp). *Plant Biotechnol. J.* 15, 547–557. doi:10.1111/PBI.12639

Zhang, Y., He, J., Wang, Y., Xing, G., Zhao, J., Li, Y., et al. (2015). Establishment of a 100-seed Weight Quantitative Trait Locus-Allele Matrix of the Germplasm Population for Optimal Recombination Design in Soybean Breeding Programmes. *Exbotj* 66, 6311–6325. doi:10.1093/jxb/erv342

Zhang, Y., Yu, C., Lin, J., Liu, J., Liu, B., Wang, J., et al. (2017). OsMPH1 Regulates Plant Height and Improves Grain Yield in Rice. *PLoS One* 12, e0180825–17. doi:10.1371/journal.pone.0180825

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. *Nat. Genet.* 42, 355–360. doi:10.1038/ng.546

Zhao, J., Zhai, Z., Li, Y., Geng, S., Song, G., Guan, J., et al. (2018). Genome-wide Identification and Expression Profiling of the Tcp Family Genes in Spike and Grain Development of Wheat (triticum Aestivum l.). *Front. Plant Sci.* 9, 1–23. doi:10.3389/fpls.2018.01282

Zhou, F., Roy, B., Dunlap, J. R., Enganti, R., and von Arnim, A. G. (2014). Translational Control of Arabidopsis Meristem Stability and Organogenesis by the Eukaryotic Translation Factor eIF3h. *PLoS One* 9, e95396. doi:10.1371/journal.pone.0095396

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 © 2022 Singh, Dhillon, Kaur, Dhaliwal, Kaur, Malik, Kumar, Gill and Kaur. 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 | Frontiers in Genetics

SYSTEMATIC REVIEW
published: 04 August 2022
doi: 10.3389/fgene.2022.905771

Exploring Chickpea Germplasm Diversity for Broadening the Genetic Base Utilizing Genomic Resourses

*Rajesh Kumar Singh[1], Charul Singh[2], Ambika[3], B. S. Chandana[1], Rohit K. Mahto[1], Ranjana Patial[4], Astha Gupta[5], Vijay Gahlaut[6], Gayacharan[7], Aladdin Hamwieh[8], H. D. Upadhyaya[9,10] and Rajendra Kumar[1]**

[1]Indian Agricultural Research Institute (ICAR), New Delhi, India, [2]University School of Biotechnology, Guru Gobind Singh Indraprastha University, New Delhi, India, [3]Department of Genetics and Plant Breeding, University of Agricultural Sciences, Bangalore, Bangalore, India, [4]Department of Agricultural Sciences, Chandigarh University, Mohali, India, [5]School of Agricultural Sciences, Sharda University, Greater Noida, India, [6]Institute of Himalayan Bioresource Technology (CSIR), Pālampur, India, [7]National Bureau of Plant Genetic Resources (ICAR), New Delhi, India, [8]International Center for Agriculture Research in the Dry Areas (ICARDA), Giza, Egypt, [9]Department of Entomology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [10]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, United States

OPEN ACCESS

Edited by:
Nasya Borisova Tomlekova, Maritsa Vegetable Crops Research Institute (MVCRI), Bulgaria

Reviewed by:
Rakesh Singh, National Bureau of Plant Genetic Resources (ICAR), India
Muhammad Azhar Nadeem, Sivas University of Science and Technology, Turkey

***Correspondence:**
Rajendra Kumar
rajendrak64@yahoo.co.in

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

Received: 27 March 2022
Accepted: 24 June 2022
Published: 04 August 2022

Citation:
Singh RK, Singh C, Ambika, Chandana BS, Mahto RK, Patial R, Gupta A, Gahlaut V, Gayacharan, Hamwieh A, Upadhyaya HD and Kumar R (2022) Exploring Chickpea Germplasm Diversity for Broadening the Genetic Base Utilizing Genomic Resourses. Front. Genet. 13:905771. doi: 10.3389/fgene.2022.905771

Legume crops provide significant nutrition to humans as a source of protein, omega-3 fatty acids as well as specific macro and micronutrients. Additionally, legumes improve the cropping environment by replenishing the soil nitrogen content. Chickpeas are the second most significant staple legume food crop worldwide behind dry bean which contains 17%–24% protein, 41%–51% carbohydrate, and other important essential minerals, vitamins, dietary fiber, folate, β-carotene, anti-oxidants, micronutrients (phosphorus, calcium, magnesium, iron, and zinc) as well as linoleic and oleic unsaturated fatty acids. Despite these advantages, legumes are far behind cereals in terms of genetic improvement mainly due to far less effort, the bottlenecks of the narrow genetic base, and several biotic and abiotic factors in the scenario of changing climatic conditions. Measures are now called for beyond conventional breeding practices to strategically broadening of narrow genetic base utilizing chickpea wild relatives and improvement of cultivars through advanced breeding approaches with a focus on high yield productivity, biotic and abiotic stresses including climate resilience, and enhanced nutritional values. Desirable donors having such multiple traits have been identified using core and mini core collections from the cultivated gene pool and wild relatives of Chickpea. Several methods have been developed to address cross-species fertilization obstacles and to aid in inter-specific hybridization and introgression of the target gene sequences from wild *Cicer* species. Additionally, recent advances in "Omics" sciences along with high-throughput and precise phenotyping tools have made it easier to identify genes that regulate traits of interest. Next-generation sequencing technologies, whole-genome sequencing, transcriptomics, and differential genes expression profiling along with a plethora of novel techniques like single nucleotide polymorphism exploiting high-density genotyping by sequencing assays, simple sequence repeat markers, diversity array technology platform, and whole-genome re-sequencing technique led to the identification and development of QTLs and high-density trait mapping

of the global chickpea germplasm. These altogether have helped in broadening the narrow genetic base of chickpeas.

Keywords: broadening the genetic base, cicer, genetic diversity (GD), gene editing, multiple resistance, omics, QTL mapping, wild chickpea utilization

1 INTRODUCTION

Grain legumes are a key component of the agricultural ecosystem. These plants are a chief member of the most diverse and ecologically crucial botanical families. Legumes play a vital role in crop rotations or intercropping schemes as these plants are capable of nitrogen assimilation through symbiotic relationship with rhizobia. Chickpea (*Cicer arietinum*) is the second most important grain legume after dry bean (*Phaseolus vulgaris* L.). Chickpeas have eight pairs of homologous chromosomes (2n = 16) with an estimated genome size of 738 Mb and 28,269 annotated genes (Varshney et al., 2013). The cultivated chickpea is believed to be originated in the Anatolia of Turkey (Van der Maesen, 1984). Vavilov denominated two primary centers of origin for chickpea viz., southwest Asia (Afghanistan) and the Mediterranean with the secondary center of origin as Ethiopia. Since ancient's times, legumes have been grown for human subsistence. Globally, India is the largest producer and consumer of pulse crops. Pulses are the major source of carbohydrates, proteins, lipids, vitamins, and minerals for people across the globe (Aykroyd and Doughty, 1982). Pulses complement the nutritional quality, bioavailability of nutrients, when consumed along with cereals. Pulses provide 22–24% of protein, which is about twice the amount of wheat and three times the rice. Pulses are one of the cheapest sources of protein and play a very significant role in sustaining nutritional requirements in developing and economically poor countries. They have a low glycemic index (GI) and elicit only a moderate postprandial glycemic response after consumption. As a result, incorporating legumes into one's diet is advised for glycemic-influenced diabetes control (Rizkalla et al., 2002).

Chickpea is the major source of food and nutrition in the semi-arid tropics. In comparison to other pulses, chickpeas are a rich source of protein and carbohydrates, accounting 80% to the whole mass of dried seeds (Geervani, 1991; Chibbar et al., 2010). Chickpea is high in dietary fiber (DF), vitamins, and minerals and is known to lower low-density lipoprotein (Wood and Grusak, 2007). Chickpea has the highest quantity of total DF amongst pulses, which ranges from 18 to 22 g/100 g of raw seed (Aguilera et al., 2009). The soluble and insoluble DF contents of chickpea raw seeds are about 4–8 and 10–18 g/100 g, respectively (Dalgetty and Baik, 2003). It has been demonstrated that chickpeas have more bioavailable protein than other legumes (Sánchez-Vioque et al., 1999; Yust et al., 2003). The changes in protein content of pre- and post-dehulled chickpea dried seeds are observed which range from 17%–22% and 25.3%–28.9%, respectively (Hulse, 1991; Badshah et al., 2003). Raw chickpea seeds have a total fat content ranging from 2.70 to 6.48% (Kaur et al., 2005; Alajaji and El-Adawy, 2006). On an average, raw chickpea seeds give 5.0 mg/100 g Fe, 4.1 mg/100 g Zn, 138 mg/100 g Mg, and 160 mg/100 g Ca. Chickpea is an inexpensive, rich source of folate and tocopherol (Ciftci et al., 2010). The major carotenoids, viz., β-carotene, lutein, zeaxanthin, β-cryptoxanthin, lycopene, and α-carotene are also found in chickpea.

Globally two types of chickpea cultivars desi or microsperma and Kabuli or macrosperma are cultivated. Generally, Kabuli chickpea is predominantly cultivated in temperate regions like the Mediterranean region that includes Western Asia, Southern Europe, and Northern Africa. However, desi chickpea is raised mainly in the semi-arid tropics (Malhotra et al., 1987; Muehlbauer and Singh, 1987) such as Ethiopia and the Indian sub-continent. In general, desi types are characterized by small seeds, angular shape with a rough surface having a dark seed coat and flowers of pink or purple color due to the presence of anthocyanin pigment, whereas Kabuli types are bold seeded owl shaped with smooth surface have beige seed coat and bear white color flowers because of lack of anthocyanin pigment (Pundir et al., 1985). Desi-type chickpeas are generally early maturing and high yielding than the Kabuli type. The desi chickpea is the predominant form cultivated in India occupying approximately 80–85% and the Kabuli chickpea occupies the remaining 15–20% of the total area and production. The chickpea draft genome sequences are already reported for desi (Jain et al., 2013) and Kabuli (Varshney et al., 2013) types.

Chickpeas are majorly grown as rainfed crops since they require less irrigation than other competitive crops such as cereals. However, it can be grown in a wide range of soils and agro-climatic conditions. Chickpea contributes to farming systems' long-term survival as it plays important roles in crop rotation, mixed and intercropping, soil fertility maintenance through nitrogen fixation, and the release of soil-bound phosphorus; overall it improves the soil ecosystem. Globally, chickpea is grown on 14.842 m ha with an annual production volume of 15.083 m tones having a productivity average of 1,016 kg/ha. Indian contribution to the globe is 73.769% (10.949 m ha) in terms of area and 73.456% (11.080 m tones) production as depicted in **Figures 1A,B** with average productivity of 1,012 kg/ha (FAOSTAT 2020). Pakistan, Turkey, Australia, Myanmar, Ethiopia, Iran, Mexico, Canada, China, and the United States are among the other significant chickpea producers.

Rajasthan, Maharashtra, Madhya Pradesh (MP), Uttar Pradesh (UP), Karnataka, and Andhra Pradesh (AP) are the major states growing chickpea and other pulses in India. Rajasthan is also the highest producer of chickpea in India followed by Maharashtra, MP, UP, and Karnataka; and together contribute to 83% of production and 82% of the area in India (**Figures 2A,B**).

Although the productivity is a little higher than average global productivity, it is lesser than the estimated potential yield,

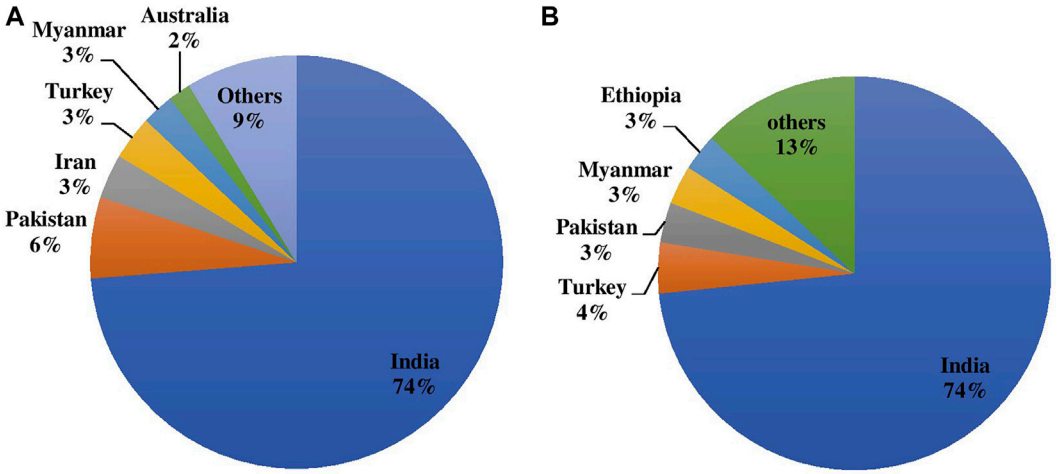


FIGURE 1 | (A) Area and **(B)** Production of chickpea during 2020 in major producing countries in the world.

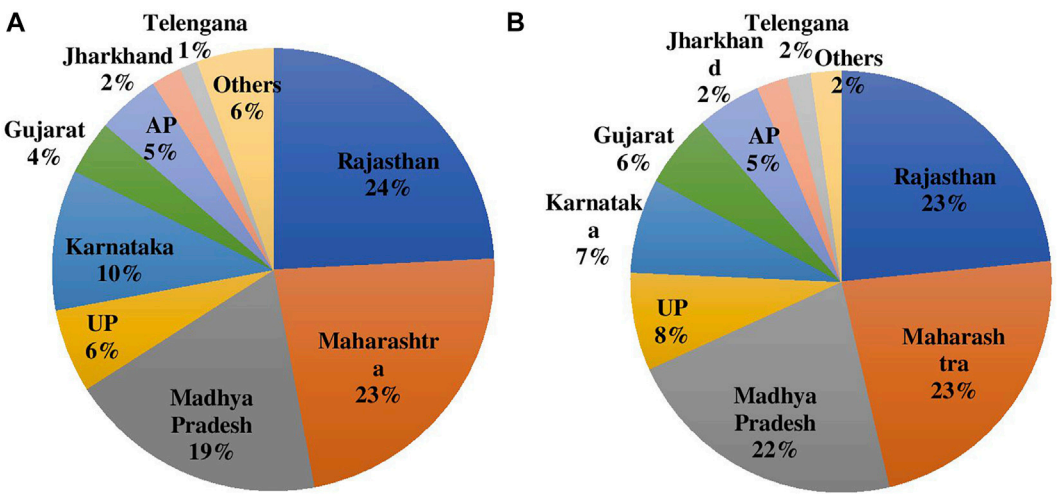


FIGURE 2 | (A) Area and **(B)** Production of chickpea during 2020 in major producing States in India.

i.e., 6 tones/ha under optimum conditions for the crop (Thudi et al., 2016). Ever-increasing the human population linked with climate change and limited arable land poses a challenge to meet the demands of growing malnutrition and hunger. A lot of efforts had been made by the national and international scientific community to enhance the productivity of chickpeas, but unable to enhance up to a significant level. The reasons underlying are a narrow genetic base and as a result poor genetic gains in the breeding of improved varieties which, leads to the reduction in the yield and its adaptation (Varshney et al., 2012). Devastating pests, pathogens, and increased incidences and severity of abiotic stress amid climate change are the major factors adversely affecting chickpea yield and production. Therefore, diverse sources of variations including wild *Cicer* species need to be explored for the genetic enhancement of chickpeas.

Chickpea performs better in cooler areas since it is a C-3 plant, implying that C-3 plants are better for the winter season. However, the harvest index (HI) in pulses (15%–20%) is low when compared to cereals (45–50%), which is a concerning issue. It is caused by excessive vegetative growth and can be countered by early dry matter partitioning into seeds (Saxena and Johansen, 1990). Despite continued efforts by national and international chickpea improvement programs for the last several decades, the production and productivity of chickpeas have not increased significantly. Probably, this has happened due to the lack of variability for desired plant ideotypes, resistance sources for devastating pests and pathogens, and less responsive behavior

of pulses toward modern agricultural practices and inputs. In general, chickpeas and other pulses are grown as a residual or alternative crop in marginal areas, only if the farmers have met their food/income requirements from high productivity- high input responsive crops such as paddy and wheat. After the onset of the green revolution, pulses were further marginalized in their traditional farming systems and local landrace variability in the farmer's field was lost. Furthermore, chickpea is subjected to various types of biotic and abiotic stresses, which are blamed for much of the crop's unstable and low yields (Reddy, 2016).

In the production of chickpea, there has been a considerable risk of abiotic stresses. Crop failure is frequently attributed to moisture and temperature stresses, which leave the greatest impact on grain yield. Drought and heat stresses cause forced maturity, resulting in reduced yield. For example, the terminal drought stress in the Mediterranean region when chickpea is grown in the spring season. Drought along with heat stress alone annually reduces productivity by up to 70%. Another major problem in chickpea production is soil salinity and alkalinity. High levels of salinity and alkalinity in both semi-arid tropics and irrigated sections of the Indo-Gangetic plains are a major problem, as most of the pulses are highly sensitive to salinity and alkalinity. Another abiotic factor that limits chickpea grain yield is cold, particularly in temperate regions. Yield is further affected by lack of highly resistant sources in the cultivated gene pool for many of the devastating pathogens and biotic stresses such as dry root rot, ascochyta blight, collar rot, botrytis grey mold (BGM) and *Helicoverpa* species further aggravate the situation (Reddy, 2016). In India, more than 250 insect species have been documented to be harmful to pulses including the chickpea crop.

To achieve higher and stable productivity, it is crucial to breed superior crop varieties with high yield, improved nutrition, disease, and pest resistance to meet the rising global demands. The genetic gains of chickpea and other legume crops are very less as compared to other crops, the reason behind this is the narrow genetic base. To meet the future demand, we have to accelerate genetic gains which are a cyclic process of identifying new variants, carrying selection, and fixing desirable traits. Further, to sustain higher genetic gain for a longer duration, infusion of genetic diversity in modern varieties from landraces and wild *Cicer* species is required. Genomics, high throughput precision phenotyping tools, and artificial intelligence can help in making a desired selection, and in achieving accelerated genetic gain while reducing genetic diversity loss (Varshney et al., 2018).

2 NARROW GENETIC BASE—A MAJOR BOTTLENECK IN CHICKPEA

Chickpeas have an inherently narrow genetic base as the crop had been subjected to a series of major genetic bottlenecks such as natural selection driven by biotic and abiotic stresses, farmers' selection pressure (domestication syndrome effect), the introduction of a small set of variability (founder effect), utilization of a very small proportion of variability in the breeding of modern cultivars, etc. (Abbo et al., 2003). Chickpea is a self-fertilization crop, which enhances the probability of loss of variability particularly rare alleles/traits in a population during the selection processes, leading to further narrowing of the chickpea genetic base. Some of the other major factors causing narrowed genetic base of chickpea are areas given below:

- Restricted distribution of wild progenitors of chickpea (*C. reticulatum* is restricted to a small area in SE turkey) (Abbo et al., 2003), which obstructs the gene flow from the wild to the cultivated types.
- Founder effect: similar to any other Neolithic crops, chickpea crop is of monophyletic origin from its wild progenitor and only a limited amount of variability is spread to other parts of the world, causing a genetic bottleneck and narrowed genetic base (Ladizinsky, 1985).
- Domestication syndrome: wild progenitors have ordained to cultivated forms after passing through various genetic modifications and acquiring a combination of traits which might have led to the disappearance of many genes/alleles responsible for input response and higher gain yield (Jain et al., 2014).
- The change from autumn to spring sowing in chickpea: in the Early Bronze Age, the shift of chickpea sowing from autumn to spring to avoid certain biotic stresses, i.e., ascochyta blight. This was possible through the selection for vernalization response in chickpea wild progenitor species; which must have caused a drastic loss of genetic diversity (Abbo et al., 2003).
- The replacement of the land races by elite cultivars produced by modern plant breeding methods which are often developed by genetically similar parental lines and most of the breeding programs shares a limited set of parental lines (Tanksley and McCouch, 1997).

Crop improvement mainly relies on the genetic matter available for exploration through the methods of plant breeding, i.e., classical and molecular breeding. The repeated use of the same germplasm has made very less contribution to the development of the new cultivars. Hence, it could be inferred that chickpea has a narrow genetic base and prompt measures for the transfer of targeted traits from wild *Cicer* species to cultivated one should be taken up by properly evaluating, characterizing, identifying, and utilizing the available germplasm during hybridization programs (Varshney et al., 2021).

In cereals, the amount of yield improvement achieved by breeding is substantially more than chickpea and other pulses. This is probably because the crops have not faced such a harsh bottleneck, and have a comparative broader genetic base (Abbo et al., 2003). The drawback of chickpea breeding programs is their narrow genetic base and unavailability of high input responsive cultivars. In order to develop high-yielding lines, chickpea genetic resources are needed to be explored to broaden the genetic base. Genetic diversity is a major contributor to selection-induced genetic gain, therefore, poor genetic diversity in chickpeas is the major limiting factor in enhancing chickpea yield. As a result, expanding the genetic base of chickpeas is

TABLE 1 | List of Cicer species and their distribution.

Sl. No.	Cicer species	Distribution
	Annuals	
1.	*C. arietinum*	Mediterranean region to Myanmar, Ethiopia, Mexico, Chile
2.	*C. bijugum*	Turkey, Syria, Iraq
3.	*C. chorassanlcum*	Afghanistan, Iran
4.	*C. cuneatum*	Ethiopia, Egypt, Sudan, Saudi Arabia
5.	*C. echinospermum*	Turkey, Anatolia, Iraq
6.	*C. judaicum*	Palestine, Lebanon
7.	*C. pinnatifidum*	Cyprus, Iraq, Syria, Turkey, Armenia
8.	*C. reticulatum*	Turkey
9.	*C. yamashitae*	Afghanistan
10.	*C. turcicum*	Southeast Anatolia (Turkey)
	Perennials	
11.	*C. acanthophyllum*	Afghanistan, Pakistan, Tadzhik SSR
12.	*C. anaiolicum*	Turkey, Iran, Iraq, Armenia
13.	*C. atlanticum*	Morocco
14.	*C. balcaricum*	Caucasus
15.	*C. baldshuanicum*	Tadzhik SSR
16.	*C. canariense*	Canary Islands, Tenerife and La palma
17.	*C. fedtschenkoi*	KirghizSSR, Tadzhik SSR, NE Afghanistan
18.	*C. flexuosum*	KirghizSSR, Tadzhik SSR: Tian-shan
19.	*C. floribundum*	Turkey
20.	*C. graecum*	Greece
21.	*C. grande*	Uzbek SSR, Naratau
22.	*C. heterophyllum*	Turkey
23.	*C. incanwn*	Former USSR
24.	*C. incisum*	Greece, Turkey, Iran, Lebanon, Georgian SSR
25.	*C. isauricum*	Turkey
26.	*C. kermanense*	Iran
27.	*C. Korshinskyi*	Tadzhik SSSR
28	*C. laetum*	Description not traced
29.	*C. macracanthum*	Afghanistan, India, Pakistan, Tadzhik SSR
30.	*C. microphyllum*	Afghanistan, Tibet, India, Pakistan, Pamir USSR
31.	*C. mogoltavicum*	Tadzhik SSR
32.	*C. montbrettii*	Albania, Bulgaria, Turkey
33.	*C. multijugum*	Afghanistan
34.	*C. nuristanicum*	Afghanistan, India, Pakistan
35.	*C. oxyodon*	Iran, Afghanistan, Iraq
36.	*C. paucijugum*	Tadzhik SSR
37.	*C. pungens*	Afghanistan, Former USSR
38.	*C. rassuloviae*	Description not traced
39.	*C. rechingeri*	Afghanistan
40.	*C. songaricum*	Tadzhik SSR, Kazakh SSR
41.	*C. spiroceras*	Iran
42.	*C. stapfianum*	Iran
43.	*C. subaphyllum*	Iran
44.	*C. tragacanthoides*	Iran, Turkmen SSR

critical for enhancing breeding efficiency. Chickpea wild species are an important genetic resource, especially for biotic and abiotic stress resistance and nutritional quality. Chickpea mutants with novel features like brachytic growing behavior (Gaur et al., 2008), more than three flowers per node–the cymose inflorescence (Gaur and Gour, 2002), determinate (Hegde, 2011), and semi-determinate growth habit (Harshavardhan et al., 2019; Ambika et al., 2021) with the potential to generate futuristic plant types have been identified. In addition, several relevant agro-morphological features and key biotic factors in a variety of wild annual *Cicer* species have been discovered and proposed for their introgressions into the cultivated gene pool to expand the genetic basis (Singh et al., 2014). Therefore, there is an emergent need to strengthen research efforts for identifying useful breeding techniques to enhance the genetic base of chickpea for enhancing genetic gains and finally chickpea yield. One of the greatest challenges in boosting grain legume output is the availability of high-quality seed and other inputs, which is lagging in the chickpea crop and only possible through infusing more and more variability in seed chain systems (David et al., 2002).

3 SOURCES OF GENETIC DIVERSITY AND BROADENING OF CHICKPEAGENETIC BASE

In the past, crop improvement has led to narrowing down of the genetic base resulting in low genetic gains and increased risk of genetic vulnerability. In order to overcome the genetic bottlenecks and create superior gene pools, broadening the genetic base through pre-breeding is required to enhance the utility of germplasm. To attain sustainable growth in chickpeas, new sources of genes need to be identified and incorporated into high-yielding cultivars. The systematic evaluation, characterization, and utilization of wild species-specific targeted genes, to overcome the drawbacks of the abiotic and biotic stresses by broadening the genetic base of chickpea cultivars, are the emergent and immediate requirements. Broadening of the genetic base is now necessary and useful and it is well recognized in all crops mainly in chickpeas and other pulse crops.

The genetic base of cultivated chickpeas is limited (Kumar and Gugita, 2004). Breeders are unwilling to employ exotic germplasm because of linkage drag and/or loss of adaptive gene complex, which necessitates a prolonged time for developing cultivars. As a result, breeders prefer to focus on adapted and improved materials; while ignoring wild relatives, landraces, and exotic germplasm accessible in gene banks (Nass and Paterniani, 2000); thus, further narrowing the genetic base and expanding the gap between available genetic resources and their use in breeding programs (Marshall, 1989). However, substantial diversity among specified parental lines is critical for the success of any breeding program, particularly when the traits to be improved are quantitative, highly variable, and exhibit high G × E interactions.

3.1 Sources for Broadening of Genetic Base

There are several sources that could be used for broadening of the genetic base in chickpea to overcome the bottleneck of biotic and abiotic stress in the scenario of changing climatic conditions. Tolerance may be contained in the wild relatives, landraces, advanced breeding materials, initial breeding material, and high-yielding cultivars (Meena et al., 2017). Landraces and wild progenitors have been used for the introgression of various abiotic and biotic stress tolerant gene(s). Mini core germplasm (Upadhyaya et al., 2013) along with several varieties and cultivars have been screened intensively for various biotic and abiotic stresses and used for numerous tolerances in chickpeas.

3.1.1 Sources of Chickpea Genetic Diversity: Cicer Wild Relatives

The genus *Cicer* currently comprises 44 species (**Table 1**) containing 10 annuals and 34 perennials (van der Maesenet al., 2007). *C. turcicum*is the recent most identified wild *Cicer* species endemic to Southeast Anatolia (Turkey) (Toker et al., 2021). This is an annual species, and with sequence similarity based on the internal transcribed spacer (ITS) region, it appears that *C. turcicum* is a sister species of *C. reticulatum* and *C. echinospermum*, both of which gives fertile progenies when crossed with the cultivated species. Utilization of the new species in the chickpea improvement program will have a great impact on the genetic base broadening. *C. arietinum* is the only species that is extensively recognized as cultivated species. *Cicer reticulatum* is identified as a probable ancestor of chickpea (Ladizinsky and Adler, 1976a). The cultivated chickpea is believed to be originated in the Anatolia of Turkey (Van der Maesen, 1984). Vavilov specified two primary centres of origin for chickpea, southwest Asia and the Mediterranean with the secondary center of origin as Ethiopia. The chickpea closely associated species viz.; *C. bijugum*, *C. echinospermum*, and *C. reticulatum* are widely distributed across southeastern Turkey and neighboring Syria (Ladizinsky and Adler, 1975; Ladizinsky, 1998). However, several *Cicer* species are restricted to particular geographic areas such as *C. bijugum* in Syria and Turkey, *C. anatolicum* in Armenia and Turkey, *C. macracanthum* in Pakistan, *C. microphyllum* in India and Pakistan, and so on. *C. arietinum* is a cultivated species that can't colonize without human assistance. *C. reticulatum* and *C. bijugum* grow naturally in weedy habitats (fallow lands, road sides, cultivated fields of wheat, and other territories not grabbed by human beings or livestock), *C. pungens* and *C. yamashitae* are found in mountain slopes among rubbles, *C. montbretia* and *C. floribundum* are distributed on forest soils, in broad leaf or pine forests and *C. microphyllum* grows naturally in stony and desert areas of the Himalayas in India (Chandel, 1984). Different *Cicer* species and their distributions are presented in Table 1.

The primary gene pool constitutes domesticated chickpea, *C. arietinum,* and the immediate progenitor, *C. reticulatum*, the species which are easily crossable with regular gene exchange. They differ either by a reciprocal inversion, a paracentric inversion or by the location of chromosomal satellites (Ladizinsky, 1998). The *C. echinospermum* represents a secondary gene pool and is crossable with cultivated chickpea, but gives reduced pollen fertility in the hybrids and their advanced generations. The tertiary gene pool contained remnant 6 annual and 34 perennial species having poor crossing compatibility with cultivated chickpea and requiring advanced approaches for gene transfer. Wild lines of chickpeas are very good sources of the genes/QTLs for the development of varieties which could be climate-resilient and tolerant to most of the biotic and abiotic stresses (**Table 2**). These lines consist of different species of chickpea of the primary, secondary, and tertiary gene pool (**Figure 3**). The resistance transfer from wild species poses several problems such as cross incompatibility, hybrid sterility, hybrid inevitability, and linkage of undesirable traits.

3.1.2 Sources of Chickpea Genetic Diversity: Gene Bank Collections and Introductions

The primary goal of a germplasm collection is to capture a significant amount of genetic variation, conserve, and enhance utilization (Singh and Singh, 1997). The first exploration expedition, led by the United States Department of Agriculture's Regional Pulse Improvement, was conducted in India in the 1970s, collecting almost 7,000 chickpea accessions. In

TABLE 2 | Sources of desirable traits in Cicer species for introgression into elite genetic background of chickpea to broaden genetic base.

S. No.	Trait of interest	Cicer species	References
	Biotic stresses		
1.	Aschochyta blight resistance	*C. arietinum, C. judaicum, C. reticulatum, C. montbretii, C. bijugam, C. pinnnatifidum, C. cuneatum, C. echinospermum*	Vander Maesen and Pundir (1984), Singh and Reddy (1993), Singh et al. (1994), Singh et al. (1998), Collard et al. (2001), Collard et al. (2003), Ahmad et al. (2013), Shah et al. (2005), Pande et al. (2005), Pande et al. (2006), Pande et al. (2010), Kaur et al. (2013), Singh et al. (2014), Benzohra et al. (2014), Li et al. (2017)
2.	Botrytis grey mouldresistance	*C. judaicum, C. bijugam, C. pinnnatifidum, C. reticulatum*	Singh (1982), Van der Maesen and Pundir (1984), Haware et al. (1992), Pande et al. (2006), Basandrai et al. (2006), Basandrai et al. (2008), Kaur et al. (2013), Singh et al. (2014), Hegde et al. (2018)
3.	Cyst nematode resistance	*C. bijugam, C. pinnnatifidum, C. reticulatum*	Greco and Vito (1993), Singh et al. (1994), Ahmad et al. (2013), Singh et al., 2010
4.	Fusarium wilt resistance	*C. arietinum, C. reticulatum, C. bijugam, C. judaicum, C. pinnnatifidum, C. echinospermum, C. cuneatum*	Nene and Haware (1980), Van der Maesen and Pundir (1984), Kaiser et al. (1994), Infantino et al. (1996), Nguyen et al. (2004), Singh et al. (1994), Singh et al. (2005), Ahmad et al. (2013)
5.	Phytophthora root rot resistance	*C. reticulatum, C. bijugum, C. pinnnatifidum, C. Echinospermum*	Knights et al. (2008)
6.	Root-knot nematode resistance	*C. bijugum, C. judaicum, C. pinnnatifidum, C. reticulatum, C. echinospermum*	Singh et al. (2014)
7.	Root-lesion nematode resistance	*C. echinospermum, C. reticulatum*	Thompson et al. (2011)
8.	Rust resistance	*C. bijugam, C. reticulatum, C.echinospermum*	Sillero et al. (2012)
9.	Stem rot resistance	*C. reticulatum, C. pinnatifidum, C. judaicum, C. yamashitae*	Kaur et al. (2008)
10.	Bruchids tolerance	*C. reticulatum*	Singh et al. (2010), Eker et al. (2018)
11.	Helicoverpa pod borer tolerance	*C. bijugum, C. reticulatum, C. echinospermum, C. cuneatum, C. pinnatifidum, C. Microphyllum*	Kaur et al. (1999), Sharma (2004), Sharma et al. (2006), Golla et al. (2018)
12.	Leaf miner tolerance	*C. reticulatum, C. judaicum, C. bijugam, C. cuneatum*	Singh and Weigand (1994), Singh et al. (1994)
13.	Seed beetle tolerance	*C. cuneatum, C. judaicum, C. reticulatum, C. echinospermum*	Gupta and Parihar, (2015)
	Abiotic stress		
14.	Cold tolerance	*C. echinospermum, C.reticulatum, C. bijugum, C. pinnnatifidum, C. judaicum*	Singh et al. (1990), Singh et al. (1995), Sandhu (2004), Toker (2005), Berger et al. (2005), Saeed and Darvishzadeh (2017)
15.	Drought tolerance	*C. anatolicum, C. reticulatum C. microphyllum, C. oxydon, C. montbrettii, C. pinnnatifidium, C. songaricum, C. echinospermum*	Toker et al. (2007), Canci and Toker (2009), Maqbool et al. (2017)
16.	Heat resistance	*C. pinnatifidum, C. reticulatum*	Canci and Toker (2009), Devasirvatham et al. (2012)
17.	Salinity resistance	*C. microphyllum*	Srivastava et al. (2016)
	Yield parameters		
18.	High no. of seeds plant−1	*C. cuneatum, C. montbretii*	Robertson et al. (1995), Robertson et al. (1997), Gupta et al. (2017)
19.	Yield attributes	*C. reticulatum, C. pinnatifidum*	Jaiswal and Singh (1989), Singh et al. (2005), Singh et al. (2012); Singh et al. (2014), Gupta et al. (2017)

India, systematic explorations to expand chickpea germplasm began only after the establishment of the National Bureau of Plant Genetic Resources (NBPGR) in 1976. In India, the area surveyed for chickpea germplasm collection included regions of Rajasthan, Odisha, Maharashtra, Gujarat, eastern parts of Arunachal Pradesh, Bihar, and southern parts of Tamil Nadu and Karnataka (Singh and Singh, 1997). The awareness about the wild *Cicer* species as rich sources of genes/alleles not just for biotic and abiotic stresses, but also for superior agro-morphological features, has sparked a lot of interest in the researchers (Van der Maesen and Pundir, 1984). Chickpea collection displays variations in plant height, foliage color, pod size, pod bearing habit, seed coat texture, seed coat surface, seed color, and seed size (Singh et al., 2001; Archak et al., 2016). Madhya Pradesh collections were double podded, large-seeded (kabuli type), and tuberculated seeded (desi type) with short and medium duration (Pundir and Reddy, 1989; Pundir et al., 1990). NBPGR has introduced valuable germplasm material from many agroecological zones throughout the world. Some of the potential exotic *Cicer arietinum* germplasm exhibit significant levels of resilience to biotic and abiotic stresses. The imports of *Cicer* wild species (*C. canariense, C. anatolicum, C. oxyodon, C. bijugum, C. reticulatum, C. pinnatifidum* and *C. judaicuni*) have received special attention for use in breeding programs. The majority of the introductions came from International Center for Agricultural Research in the Dry Areas (ICARDA). Other important introduction sources included Spain, Afghanistan, The Former Soviet Union, Iran, United States, Morocco, and Greece. Some of the introduced chickpea lines made significant contributions to the genetic enhancement and pre-breeding,

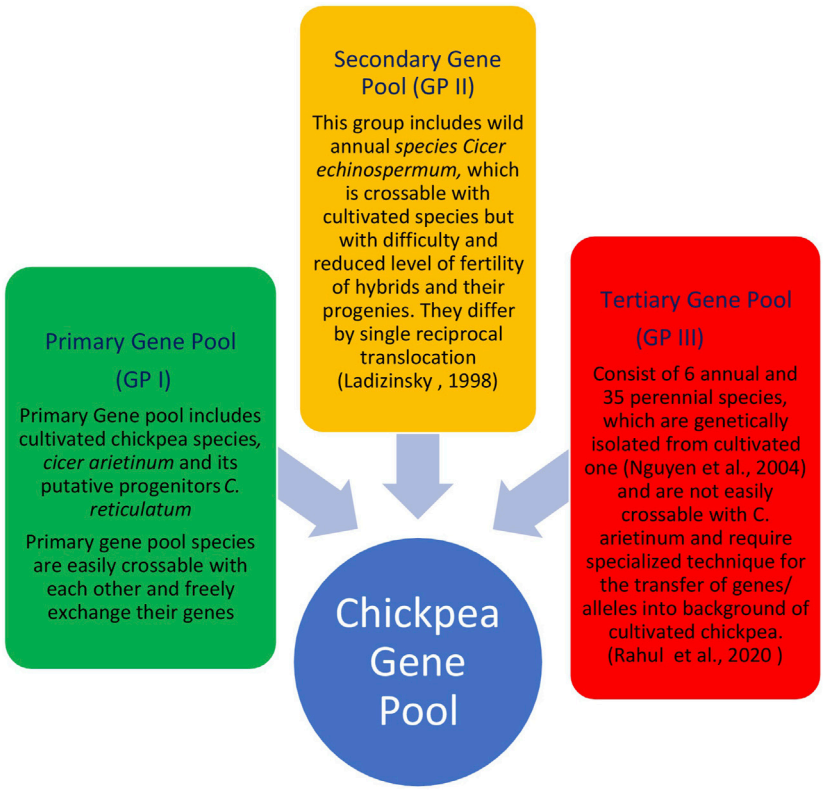


FIGURE 3 | Chickpea gene pool concept and their crossing compatibility.

TABLE 3 | *Ex-situ* conservation of Cicer accessions in the world.

Sl. No.	Country	Gene bank name	Cultivated	Wild relatives	Breeding materials	Others	Total number of accessions
1.	Global	International Crop Research Institute for the Semi-Arid Tropics (ICRISAT)	18,842	308	1,317	297	20,764
2.	Global	International Centre for Agricultural Research in Dry Areas (ICARDA)	6,816	547	5,903	2,102	15,368
3.	India	National Bureau of Plant Genetic Resources (NBPGR), New Delhi	14,635	69	—	—	14,704
4.	Australia	Australian Temperate Field Crops Collection (ATFCC)	8,409	246	—	—	8,655
5.	United States	Western Regional Plant Introduction Station, USDA-ARS, Washington State University	7,742	194	102	—	8,038
6.	Iran	National Plant Gene Bank of Iran, Seed and Plant Improvement Institute (NPGBI-SPII)	5,700	—	—	—	5,700
7.	Russia	N.I. Vavilov Research Institute of Plant Industry	1,628	—	558	581	2,767
8.	Pakistan	Plant Genetic Resources Program (PGRP)	2,057	89	—	—	2,146
9.	Turkey	Plant Genetic Resources Department, Aegean Agricultural Research Institute (AARI)	2,047	21	—	7	2,075
10.	Ukraine	Institute of Plant Production nd. a. V. Ya. Yuryev of NAAS	182	24	9	1,542	1,757
11.	Mexico	Estacio´ n de Iguala, Instituto Nacional de InvestigacionesAgrı´ colas, Iguala	1,600	—	—	—	1,600
12.	Ethiopia	Institute of Biodiversity Conservation (IBC	1,173	—	—	—	1,173
13.	Hungary	Centre for Plant Diversity	23	5	167	972	1,167
14.	Uzbekistan	Uzbek Research Institute of Plant Industry (UzRIPI)	1,055	—	—	—	1,055
		Total	71,909	1,503	8,056	5,501	86,969

Source:http://www.fao.org/wiews-archive/germplasm_query.htm?i_l¼EN.

mainly for resistance to Fusarium wilt, Ascochyta blight, leaf miner, cyst nematode, cold, drought, earliness, tall stature, and bold seeds. The important chickpea germplasm collections, including wild species that have been preserved in *ex-situ* collections in various gene banks around the world (**Table 3**).

3.1.3 Sources of Chickpea Genetic Diversity: Landraces and Cultivated Varieties

Landraces are locally adapted cultivars that evolved in a diverse range of environmental conditions and are maintained generation after generation by farmers and local seed systems. The landraces are the goldmines for trait identification for various biotic and abiotic stresses viz.; drought, salinity and cold. These land races could be exploited in breeding programs for introgression of useful genes/QTLs and enhancing the genetic variability in the modern chickpea cultivars.

The tolerance variation depends on various factors viz.; climatic factors, genotypes, seed attributes, and seed compositions. The most important prerequisite is seedling salinity tolerance since this attribute facilitates the establishment and growth of tolerant genotypes in saline soils. The roles of seed yield, yield components, pods per plant, number of seeds, *in vitro* pollen germination, pollen viability, and *in vivo* pollen tube development to assess the reproductive successful outcome of chickpea under saline stress were investigated (Turner et al., 2013). The increased salt tolerance, as measured under salty ambient by relative yield, was correlated positively with increased shoot biomass, number of pods, and seeds. Pollen viability, *in vitro* pollen germination, and *in vivo* pollen tube growth were uninfluenced by salty ambient in either of the tolerant or sensitive genotypes but pod abortion was relatively higher in salt-sensitive genotypes. Genotypes ICCV-00104, ICCV-06101, CSG-8962, and JG-62 showed a minimum reduction in seedling characters in salt stress conditions. Similar findings were reported by Samineni et al., 2011, while studying chickpea seedlings under saline stress. Flowering terminates at temperatures below 15°C as reported in Australia (Siddique and Sedgley, 1986), India (Savithri et al., 1980; Srinivasan et al., 1999) and the Mediterranean (Singh and Ocampo 1993). It was observed that, when average daily temperature remained below 15°C, plants produced flowers but did not set pods. However, scientists at International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) could develop numerous breeding materials (e.g., ICCV series 88502, 88503, 88506, 88510, and 88516) that are capable to set pods at 12°C–15°C lower average daily temperatures. A pollen selection was applied in Australia to transfer chilling tolerance from ICCV 88516 to chilling sensitive cultivars, leading to the development and release of two chilling tolerant cultivars namely Sonali and Rupali (Clarke and Siddique, 2004). Minicore germplasm was screened for drought tolerance and a few germplasm accessions viz.; ICC series 1356, 3512, 4872, 13523, and 15697 with deeper root systems were identified. The Germplasm accession ICC8261 had the highest root length density, an extremely high root/shoot ratio and rooting depth in both Rabi and Kharif seasons. ICC4958, which is a source used as a deep and large root system parent or check in most drought avoidance studies, was reported to be an extremely prolific rooting genotype. The new genotypes identified could be used as valuable alternative sources for diversification of mapping populations with varying characters and growth durations to obtain the required polymorphism for successfully mapping root traits in chickpeas.

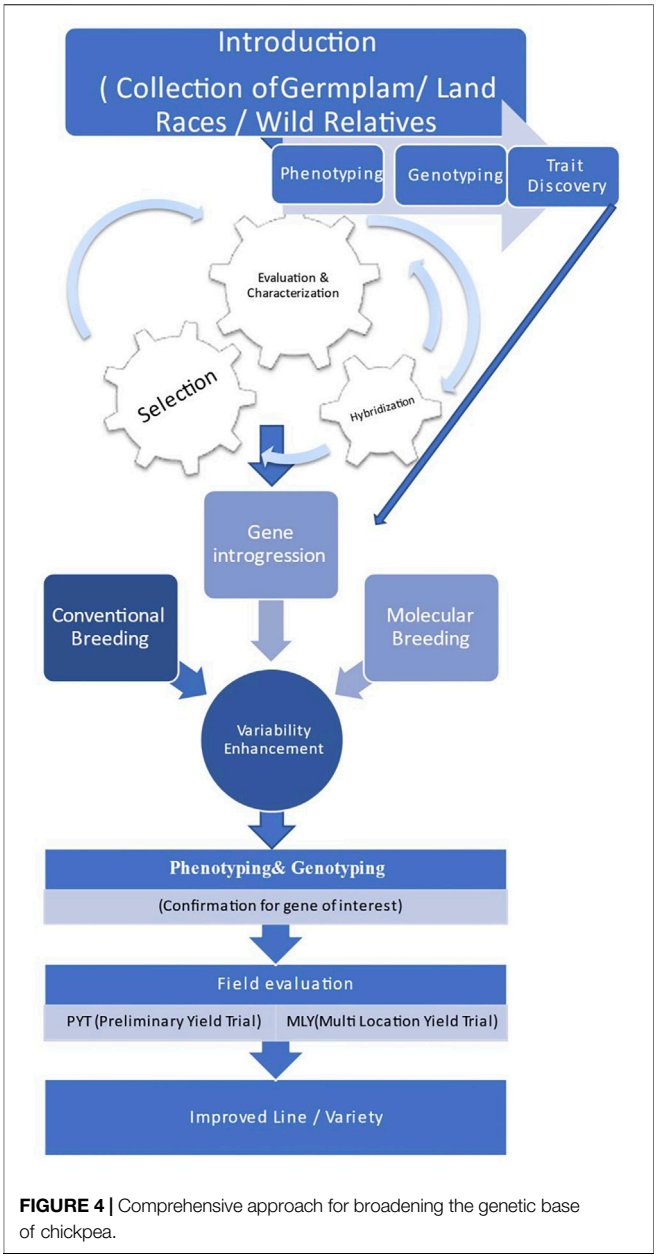


FIGURE 4 | Comprehensive approach for broadening the genetic base of chickpea.

3.2 Approaches for Broadening the Genetic Base

Broadening of the genetic base, up to now, has utilized the techniques of classical breeding viz.; hybridization, segregation, back crossing, cyclic population improvement, pedigree selection among selfed progenies. However, wild relatives couldn't be

TABLE 4 | Sources of resistance to abiotic and biotic stresses as reported by various workers after evaluating the chickpea mini core collection.

Stress	Resistant genotype		References
	Desi	Kabuli	
Drought	ICC- 283, 456, 637, 708, 867, 1205, 1422, 1431, 1882, 2263, 2580, 3325, 4495, 4593, 5613, 5878, 6874, 7441, 8950, 10399, 10945, 11121, 11944, 12155, 12947, 13124, 14402, 14778, 14799, 14815, 15868, 16524	ICC- 4872, 5337, 7272, 7323, 8261, 16796	Kashiwagi et al. (2005), Kashiwagi et al. (2006b), Kashiwagi et al. (2008), Kashiwagi et al. (2010), Parameshwarappa and Salimath (2008), Krishnamurthy et al. (2010), Mulwa et al. (2010), Zaman-Allah et al. (2011a), Zaman-Allah et al. (2011b)
Salinity	ICC- 283, 456, 708, 867, 1431, 2263, 2580, 3325, 4495, 4593, 5613, 5878, 6279, 6874, 7441, 9942, 10399, 10945, 11121, 11944, 12155, 13124, 14402, 14778, 14799, 15868, 16524	ICC- 4872, 7272, 8261, 16796	Serraj et al. (2004), Vadez et al. (2007), Krishnamurthy et al. (2011b)
Heat	ICC- 283, 456, 637, 708, 1205, 1882, 2263, 4495, 5613, 5878, 6874, 7441, 10945, 11121, 11944, 12155, 13124, 14402, 14778, 14799, 14815, 15868	—	Krishnamurthy et al. (2011a), Upadhyaya et al. (2011)
Ascochyta blight	ICC- 1915, 7184, 11284	—	Pande et al. (2006)
Botrytis gray mold	ICC- 2990, 4533, 6279, 7554, 7819, 11284, 12028, 12155, 13219, 13599, 15606, 15610	ICC- 9848, 11764, 12037, 12328, 13816, 14199, 15406	Pande et al. (2006)
Dry root rot	ICC- 1710, 2242	ICC- 2277, 11764, 12328, 13441	Pande et al. (2006)
Fusarium wilt	ICC- 1710, 1915, 2242, 2990, 3325, 4533, 5135, 6279, 6874, 7184, 7554, 7819, 12028, 12155, 13219, 13599, 14402, 14831, 15606, 15610	ICC- 2277, 9848, 12037, 13441, 13816, 14199	Pande et al. (2006)
Pod borer	ICC- 3325, 5135, 6874, 14402, 14831, 15606	ICC- 15406	ICRISAT (2009), Mulwa et al. (2010)
Herbicide	ICC- 2242, 2580, 3325	—	Taran et al. (2010)

utilized because of inter-specific hybridization barriers, limited data for specific traits, and linkage drag. With the advent of molecular breeding techniques, new biotechnological methods, which are being applied for the identification of the QTLs for the traits of interest and needs to be incorporated through various techniques of pre-breeding which are used in transferring useful genes from the exotic or wild species into the high-yielding cultivars. The halted speed of chickpea breeding due to narrow genetic diversity could be fastened by employing wild relatives as a valuable source of new genes and alleles to be further exploited by breeders for allelic richness and broadening of chickpea germplasm. Thus, comprehensive approaches could be utilized for broadening the genetic base in chickpea and other grain legume crops as depicted (**Figure 4**).

Chickpea's limited genetic base is a major source of anxiety for chickpea breeding programs, as genetic variability is a major contributor to selection-induced genetic gain. As a result, expanding the genetic base of chickpeas is critical for enhancing breeding efficiency. Chickpea wild species are an important genetic resource, especially for biotic and abiotic stress resistance and nutritional quality. Chickpea mutants with novel features like brachytic growing behavior (Gaur et al., 2008), more than three flowers per node—the cymose inflorescence (Gaur and Gour, 2002), determinate (Hegde, 2011), upright peduncle podding (Singh et al., 2013) and semi-determinate growth habit (Harshavardhan et al., 2019; Ambika et al., 2021) with the potential to generate futuristic plant types have been identified. In addition, several relevant agro-morphological features and key biotic factors in a variety of wild annual *Cicer* species have been discovered and proposed for their introgressions into the cultivated gene pool to expand the genetic basis (Singh et al., 2014). Some of the useful agro-morphological traits including major biotic and abiotic stresses are presented in **Tables 2**, **4**. There is an emergent need to strengthen research efforts for identifying useful breeding techniques to enhance the genetic base of chickpeas.

3.2.1 Utilization of Adapted and Un-Adapted Germplasm for Traits Discovery and Broadening the Genetic Base

Pre-breeding offers an unparallel opportunity for the introgression of desired genes and gene combinations from exotic germplasm into genetic backgrounds easily employed by breeders with minimal linkage drag (Sharma et al., 2013). Comprehensive broadening of the genetic base through incorporation is the most suitable method when new genetic variabilities for quantitative traits are required, latest and most reliable methods could be optical contribution selection (OCS) based pre-breeding, haplotype-based genomic approaches, and genomic predictions (Varshney et al., 2021). To achieve the highest level of yield, the existing variability among indigenous germplasm has been used. Wild *Cicer* species and exotic germplasm lines include valuable alleles that, if discovered, can aid in breaking yield barriers and improving resistance to various stresses for crop yield stability (Labdi et al., 1996; Tayyar and Waines, 1996; Ahmad and Slinkard, 2003; Ahmad et al., 2005).

Several inter-specific crosses between *Cicer arietinum* and its annual wild relatives have been attempted in the context of wild *Cicer* species usage. There is no evidence of successful hybridization between a perennial *Cicer* species and *Cicer*

arietinum. Ladizinsky and Adler (1976b) reported inter-specific crosses amongst *C. arietinum, C. reticulatum* and *C. cuneatum* for the first time. Several researchers have successfully attempted inter-specific hybrids between *Cicer arietinum* and *Cicer echinospermum* (Verma et al., 1990; Singh and Ocampo, 1993; Pundir and Mengesha, 1995). Numerous crossings between *Cicer arietinum* as the female parent and *Cicer reticulatum, C. echinospermum, C. judaicum, C. bijugum*, and *C. pinnatifidum* as the male parent have been conducted (Verma et al., 1990). Van Dorrestein et al., 1998 aimed to cross *C. arietinum* with *C. judaicum* and *C. bijugum*. Badami et al., 1997 used an embryo rescue strategy to successfully hybridize *C. arietinum* with *C. pinnatifidum*. Inter-specific crosses have resulted in the development of certain pre-breeding lines at IIPR, Kanpur, and PAU, Ludhiana (Singh et al., 2012). Singh et al. (2015) attempted inter-specific crosses and the results revealed a high level of heterosis for the number of pods and seed yield per plant in the F_1 generation. Three cross-combinations viz.; Pusa 1103 x ILWC 46, Pusa 256 x ILWC 46, and Pusa 256 x ILWC 239 demonstrated significantly increased variability for crucial yield related characteristics.

Adoption and harmonizing conventional and modern approaches like molecular breeding, physiological breeding, biotechnological methods, high throughput genomics, and phenomics will aid in the broadening of the genetic base and release of high-yielding varieties which will be tolerant to various biotic and abiotic stresses. Several mapping populations could be developed for the identification of trait-specific QTLs and can be introgressed into high-yielding cultivars for enhancing the gene pool of chickpea.

3.2.2 Bi-Parental Populations for Broadening Genetic Bases

Two inbred lineages are generally crossed in bi-parental populations to generate one or more segregating progenies (Xu et al., 2017). This is the basic approach of combining desired traits in a genotype through ongoing breeding programs. Parents are chosen for a trait of interest based on their genetic and phenotypic diversity allowing the reconstruction of progeny genomes from founder haplotypes to find genomic areas related to the target trait (Dell'Acqua et al., 2015). Bi-parental crosses derived populations capture only a modest impression of the genetic determinants that influence targeted traits in the species and suffer from a lack of diversity owing to the limited genetic base of both parents. Therefore, while the approach is indispensable for any breeding program, genetic diversity must not be reduced in the selection process, to sustain genetic gains for a longer duration. Molecular tools such as re-sequencing technologies and other cost-effective genotyping technologies, which can scan the whole genome, may be useful in the identification of diverse parental lines having the target traits of interest. The utilization of such parental lines will enhance the genetic diversity in the released varieties without compromising the desired yield gain. High-throughput precision phenotyping, genomic selection, and identification of superior haplotypes may further accelerate the breeding cycle and boost the genetic diversity in farmers' fields to enhance the crop resilience toward the biotic and abiotic stresses. In addition, the QTLs detected in the two-parent population may not be expressed in other genetic origins (Rakshit et al., 2012). Mallikarjuna et al., 2017, utilized F2 populations derived from four crosses (ICCV96029 x CDC frontier, ICC5810 x CDC frontier, BGD 132 x CDC frontier, ICC 16641 x CDC frontier) and found major QTLs corresponding to flowering time genes.

3.2.3 Multi-Parent Populations for Broadening Genetic Bases

Multi-parental and germplasm populations, on the other hand, may offer solutions to bi-parental and germplasm populations' major flaws. Throughout the history of scientific crop improvement multi-parental populations or multi-parental cross designs (MpCD) have been generated in a range of crop species. Adaptation to crops that are difficult to artificially hybridize, multi-parental populations are created by making crossings amongst more than two inbred founder lines, which serve as a link between association mapping (GWAS) and traditional bi-parental crosses. While such populations are able to combine and reveal better allelic combinations, transgressive segregants, and simultaneously genetic diversity in the progenies are also enhanced. Multi-parent populations also are more efficient in increasing mapping resolution, if they are used for high-density genotyping using advanced high-throughput genomic technologies (Rakshit et al., 2012). This unique technique dramatically improves mapping resolution by merging numerous founder parents with higher phenotypic and genetic diversity. Thanks to the evolution of more powerful techniques, multi-parental populations can now be utilized in numerous genetic mapping studies (Mackay and Powell, 2007; Huang et al., 2015). Here, the emphasis is on MAGIC populations, which are RILs of fine-scale mosaic panels, although numerous MpCD other forms are also available. Thus, MAGIC populations are considered as a growing and next-generation powerful resource for plant genetics mapping, combining variation and high genetic recombination to analyze complex traits' structure and enhance crop improvement techniques. In various model crop species, MAGIC populations have been generated illustrating their potential to find polymorphisms for underlying QTLs or genes of importance for useful complex traits. There are already MAGIC like or MAGIC populations obtainable in numerous crop species, viz., cereals, legumes, vegetables, fruit trees, and industrial crops with many more in the other works and because of their large genetic foundation, MAGIC populations could be used for discovery of QTL(s) and gene (s), enhancement of breeding populations, introduction and development and of novel genotypes (Pascual et al., 2015). Multi-parent populations such as multiparent advanced generation intercross (MAGIC) populations have gained a tremendous popularity among researchers and breeders. Such populations, along with enhancing genetic diversity, also make it easier to examine the genomic framework and their relationships with phenotypic traits.

3.2.4 Molecular Markers Based Approaches for Broadening Genetic Bases

Since the advent of molecular markers, these tools have played an indispensable role in understanding genetic diversity,

phylogenetic relationship, background, and foreground selection in molecular and conventional breeding programs. Recent advances in genomics, coupled with high throughput and precise phenotyping, have made it easier to identify genes that regulate important agronomic attributes. Genetic variability such as multiple podding per peduncle, multiple seeds per pod, upright podding, tall and erect genotypes, and several other traits for biotic stress tolerance are rare, and incorporating these traits to the major cultivars helps in enhancing the variability in the gene pool. These traits could be used in combination with tools for genomics to expedite the generation of crops with higher genetic variability with better agronomic traits, improved resilience to climate change, and nutritional values (Pourkheirandish et al., 2020). Exploring the marker-assisted selection (MAS) technique along with other biotechnological tools can boost genetic diversity and simultaneously enhancing the yield in chickpeas (Varshney et al., 2005; Varshney et al., 2009).

Genomic advancements have aided in understanding the complex trait's mechanisms affecting chickpeas economically important characters' genetic architecture as well as productivity in order to speed up breeding programs (Roorkiwal et al., 2020). In chickpea, a number of markers and trait relationships and dense genetic maps have allowed MAS to become a routine practice in crop breeding programs (Kulwal et al., 2011; Madrid et al., 2013; Ali et al., 2016; Caballo et al., 2019). Single nucleotide polymorphism (SNP) allelic variants on 27 ortholog candidate genes were utilized for the GWAS study, and potential candidate genes such as *PIN1*, *TB1*, *BA1/LAX1*, *GRAS8*, and *MAX2* were identified for branch number in chickpea utilizing highly diverse chickpea germplasm (Bajaj et al., 2016). The gene for double podding per peduncle was linked to Tr44 and Tr35 on linkage group 6 (Cho et al., 2002). Saxena et al. (2014) has mapped four traits viz. 100-seed weight, pod, number of branches per plant and plant hairiness, using simple sequence repeats (SSRs) and SNP markers. There are several other examples of utilization of molecular makers for the identification of traits and underlying genes/QTLs in chickpea such as 100-seed weight (Das et al., 2015; Kujur et al., 2015b), resistance to *Helicoverpa armigera* (Sharma et al., 2005), pod number (Das et al., 2016), flowering time (Srivastava et al., 2017), plant height (Parida et al., 2017), photosynthetic efficiency traits (Basu et al., 2019), etc. Furthermore, comprehending the chickpea developmental processes' regulations has been facilitated by the framework offered due to discoveries of new microRNAs (miRNAs) and their expression patterns (Jain et al., 2014).

For genomic investigations and crop improvement, numerous polymorphic molecular markers that could be exposed to high-throughput analysis are sought. On the basis of isozyme analysis, *Cicer arietinum* is most closely related to *C. reticulatum*, followed by *C. echinospermum, C. bijugum, C. pinnatifidum, C. judaicum, C. chorassanicum, C. yamashitae* and *C. cuneatum* (Ahmad et al., 1992). *Cicer reticulatum* and *Cicer echinospermum* were grouped together in the same cluster; *Cicer chorassanicum* and *Cicer yamashitae* were grouped together in another cluster; *Cicer bijugum*, *Cicer judaicum*, and *Cicer pinnatifidum* were grouped together in the third different cluster; and *Cicer cuneatum*alone formed the fourth different cluster based on the analysis of RAPD markers (Ahmad, 1999; Sudupak et al., 2002). An AFLP analysis for the same *Cicer* species also confirmed the same pattern (Sudupak et al., 2004). RAPD and ISSR fingerprinting demonstrate that *C. arietinum* cultivars had the narrowest genetic variation while its wild *C. reticulatum* accessions had much greater genetic variation, which could be used in chickpea improvement (Rao et al., 2007). The widespread use of molecular markers in chickpea genetics and breeding began with the introduction of SSR markers. The draft genome sequence of chickpea identified approximately 48,000 SSRs appropriate for PCR primer design for use as genetic markers (Varshney et al., 2013), whereas a draft sequence of *C. reticulatum* (PI 4889777) spanning 327.07 Mb was assembled to the eight linkage groups with 25,680 protein-coding genes (Gupta et al., 2017).

A variety of comparatively new marker systems have recently been introduced including sequence-based SNP and hybridization-based diversity array technology (DArT) markers which offer medium to high-throughput genotyping and are simple to automate. Two sets of Axiom®CicerSNP array have been developed in chickpea, one was including 50,590 probes distributed on all eight linkage groups as described by Roorkiwal et al. (2014) and the second multispecies SNP chip includes chickpea along with other pulses using markers that can be imputed up to whole-genome (800,000 markers) was developed by AgriBio, Centre for AgriBioscience Melbourne, Australia (personal communication).

To date, several studies have been published using DArT and SNP chips. We highlight the 5397 polymorphic DArT markers identified from a pool of 15,360 developed markers utilizing 94 different chickpea genotypes (Thudi et al., 2011). The low genetic diversity was unraveled between wild *Cicer* and cultivated species through DArT markers (Roorkiwal et al., 2014). Although transcriptome investigation of chickpea and its wild progenitors detected thousands of SNPs (Coram and Pang, 2005; Varshney et al., 2009; Gujaria et al., 2011; Agarwal et al., 2012; Bajaj et al., 2015b; Kujur et al., 2015a). These SNPs and markers can be utilized by chickpea breeders in MAS-assisted breeding programs.

3.2.5 Trait Identificationin Legumes for Broadening Genetic Bases

3.2.5.1 Trait Identification Through Sequencing

With the advancement in the next-generation sequencing (NGS)-based approaches, trait mapping has become an easy job to do. Not only are these technologies time-saving but also cutting the cost at basal levels. The genetic mapping is based on recombination (the exchange of DNA sequence between sister chromatids during meiosis) and the distance between the markers measured by cM representing approximately 1% of the recombination frequency, while the physical map is based on the alignment of the DNA sequences, with the distance between markers measured in base pairs. However, the high-resolution physical maps serve as the scaffold for genome sequence assembly to identify the most accurate distance between the markers and the genes linked in addition to exploring the potential candidate

gene(s) linked to desired traits. The trait mapping through sequencing approaches may be categorized into two classes 1) Sequencing of complete populations for trait mapping and 2) Sequencing of pooled samples for trait mapping. Using composite interval mapping a high-density genetic map consisting of 788 SNP markers spanning through 1125cMalong with the identification of 77 QTLs for 12 traits was reported (Jha et al., 2021). Similarly, several QTLs were mapped for several other traits like flowering time (Mallikarjuna et al., 2017; Jha et al., 2021), plant height (Kujur et al., 2016; Barmukh et al., 2021), and primary branches (Barmukh et al., 2021).

3.2.5.2 Trait Identification Through Sequencing of Complete Populations

It primarily consists of the genotyping by sequencing (GBS) and whole-genome re-sequencing (WGRS) mapping populations, both of which yield genome-wide SNPs. GBS is popular because it is inexpensive and provides a lot of genetic data. The discovery of a large number of genome-wide SNPs has facilitated rapid diversity assessment, trait mapping, GS and GWAS in a variety of crop by employing GBS—a potential strategy. A chickpea genetic variation map was developed using whole-genome sequencing technique and genomes were characterized at the sequence level, observing variations in 3,171 cultivated and 195 wild accessions and construction of a pan-genome to explain the genomic diversity across wild progenitors and cultivated chickpea (Varsheny et al., 2021). The 16 mapping populations segregating for different abiotic (drought, heat, salinity), biotic stress (Fusarium wilt, Aschochyta blight, BGM & *Helicoverpa armigera*) and protein contents along with their 35 chickpea parental genotypes were re-sequenced in order to exploit the genetic potential for chickpea improvement (Thudi et al., 2016). Genetic analysis, fine-tuning of genomic areas, and production of genetic maps are facilitated by re-sequencing (Kujur et al., 2015b; Li et al., 2015). Chickpea is one of the best examples of crops in which GBS was used to identify 828 SNPs in addition to the previously mapped SSRs. The creation of these detailed genetic maps aids in the discovery of QTLs in chickpea that controls yield, drought tolerance, and seed weight. It is quite useful for locating QTL hotspots. Moving on to the second promising strategy, WGRS has been found to be more useful in finding candidate genes than GWAS (Jaganathan et al., 2015; Varshney et al., 2014).

3.2.5.3 Trait Identification Through Pooled Sequencing

The analysis is done on the basis of the pooled population through the inclusion of BSR-Seq, Indel-Seq, Mut-Map, QTL-Seq, and Seq-BSA the five major approaches. The "QTL-Seq" is the first and foremost promising technique to have been successfully employed with larger crop plant genomes. This strategy has been used to pinpoint the blast resistance and seedling vigor governing genomic areas in rice, flowering QTLs in cucumber, fruit weight and locule number loci in tomatoes and successfully applied for localization of QTLs/candidate genes for 100 seed weight in chickpea (Takagi et al., 2013; Li et al., 2015). The "MutMap" is a robust and simple NGS-based approach, first of all which was applied for the identification of EMS-induced interesting candidate genes in rice. Crossing of selected mutant plants with wild types, which reduces background noise—the fundamental benefit, is the necessity of mapping the population created for the MutMap experimental strategies. Consequently, using extreme pool samples derived from segregating populations coupled to a wild parent the genome-wide SNP index is calculated. The third method, known as "Seq-BSA," is a straightforward and reliable NGS-based strategy for identifying potential SNPs in specific genomic regions (Takagi et al., 2013). Employing QTL-seq pipelines utilizing parent with high-value trait as reference parent assemblage, genome-wide SNP indexes of both extreme bulks are calculated in the third method. The fourth strategy, "Indel-Seq" which is mostly focused on insertions and deletions, has also emerged as a potential trait mapping approach. To date, the proposed methodologies for identifying genomic regions have relied on the discovery of SNPs followed by the use of various statistical approaches to recognize candidate genomic gene/regions. However, in all approaches, the relevant genomic region-specific existing Indels have not been targeted for trait mapping but ignored. The fact that the Indels reported in the candidate genes are found in most of the cloned genes in rice and other crops and makes this strategy more practicable. The strength of the RNA-seq and BSA were combined for enhancing the strength to find candidate genes for the targeted characteristic—a novel genetic mapping approach as the fifth strategy, dubbed as "Bulked segregant RNA-Seq (BSR-Seq)". This method has been used to successfully identify the glossy3 genes in maize. RNA-seq-based investigations will be cheaper than WGRS at higher coverage; hence, this strategy has more cost savings. We believe that, given the benefits of RNA-Seq, this approach will be effective for legumes with larger genomes (Liu et al., 2012; Trick et al., 2012). Thus, chickpea breeders utilize these generated informations in chickpea MAS-assisted breeding programs.

3.2.6 Transcriptomics Utilization for Broadening the Genetic Bases

Work on legumes focused on building libraries of cDNAs, gene expression profiling, the manufacture of expressed sequence tags (EST), and *in silico* extraction of EST data sets' functional information even before sequences of the genome achievability. Transcriptome sequencing has been employed in other functional genomics methodologies, viz., genome annotation, gene expression profiling, and non-coding RNA identification employed transcriptome sequencing (Morozova and Marra, 2008). In recent years, for generating a large number of transcript reads from a variety of developing and distress-responsive tissues in several leguminous crops through several low-cost sequencing systems has already been established, viz., an improved transcriptome assembly, utilizing FLX/454 sequencing together with Sanger ESTs comprised 103,215 Transcript Assembly Contigs (TACs) with an average contig length of 459 base pairs in chickpea (Hiremath et al., 2011). Employing various sequencing technologies or a combination of two or more sequencing technologies created by transcriptome assemblies provides useful transcriptomic resources such as functional markers, EST-SSRs, Spanning Regions (ISRs), SNPs,

Introns, and so on in soybean and common bean 1,682 and 4,099 SNPs, respectively (Deschamps and Campbell, 2012), ESTs comprising of 103,215 Transcript Assembly Contigs (TACs) in chickpea (Hiremath et al., 2011) can be utilized by the breeders to achieve a better grasping of the molecular underpinnings of distress tolerance and as a result more stress-tolerant beans as well chickpea cultivars may be produced and narrow genetic base may be broadened.

3.2.7 Proteomics and Metabolomics for Broadening the Genetic Bases

New datasets for crop plants can be created by exploiting the opportunities of advancement in "omics" technologies. The advancements will result in a greater integrated association of "omics" data and crop improvement resulting in the evolution from genomic assisted breeding (GAB) to omics assisted breeding (OAB) in the future (Langridge and Fleury, 2011) that can also be utilized for broadening the genetic bases in chickpea.

3.2.7.1 Proteomics Approaches

Increased proteome coverage and advancements in quantitative evaluations have benefitted plant proteome composition, modulation, and alterations of developmental phases including stress–response mechanisms. Proteomic pipelines are rapidly being used in crop research notably to investigate crop-specific features and stress response mechanisms. Proteome mapping, comparative proteomics, discovery of post-translational modifications (PTMs), and protein–protein interaction networks are key topics of plant proteomics (Vanderschuren et al., 2013). In chickpea the comparative root proteomic analysis for the effect of drought and its tolerance in hydroponics using 2D gel electrophoresis coupled with MALDI-TOF revealed eight categories of protein-based on their functional annotation viz.; proteins involved in carbon and energy metabolism, proteins involved in stress response, ROS metabolism, signal transduction, secondary metabolism, nitrogen and amino acid metabolism (Gupta and Laxman, 2020). High-throughput protein quantification has benefited from advancements in accuracy, speed, mass spectrometry (MS) utilizations in terms of sensitivity, and software tools. Gel-based or gel-free, shot-gun, and label-based (isotopic/isobaric) or label-free quantitative proteomics platforms have emerged as a result of developments in MS technology for high-throughput protein quantifications (Abdallah et al., 2012; Hu et al., 2015). In legume crops, comparative proteomics approaches and differential expression analyses have given understanding of distress responses including dehydration, and early phases of cold stress in chickpeas (Pandey et al., 2008) and can be effectively integrated into genomic-assisted breeding programs for broadening the narrow genetic bases.

3.2.7.2 Metabolomics Approaches

In plant metabolic engineering, targeted reverse genetic methods and high-throughput metabolite screening have the advantage of providing a better understanding of metabolic networks on a larger scale in relation to developmental stages of phenotypes and the ability to screen out undesirable traits (Fernie and Schauer, 2009). The literature describes two major metabolomics profiling methodologies that use nuclear magnetic resonance (NMR) and MS. A combination of many analytical techniques generated from one of the MS was frequently used to obtain a larger range of numerous metabolites in plants (Arbona et al., 2013). Flow injection-based analysis with Fourier Transform Infrared spectroscopy and MS (FIA/MS) are two further approaches. The identification of new metabolic QTLs and candidates for the desired traits are made possible by combining metabolomics data, transcriptomics data, high-throughput phenotypes, and bioinformatics platforms to profile large genetically varied populations and increase the accuracy of targeted gene identification. To boost yields and broaden the narrow genetic bases, metabolomics is utilized in conjunction with a genomic-assisted selection and introgression techniques, minimizing the time spent in uncovering new characteristics and allelic mutations (Fernie and Schauer, 2009).

3.2.8 Pan Genomics

Recent developments in genome sequencing technologies have revolutionized the crop improvement programs. Now the whole-genome sequencing (WGS) is not limited to one or two individuals, but a large set of accessions of a species (pangenome) including their crop wild relatives (super-pangenome) are the whole genome sequenced to unravel the full potential of the species for the crop improvement programs. Once the pangenome information is available, the genomic segments/genes lacking in cultivated germplasm can be identified and introgressed in cultivated germplasm to enhance the genetic variability. The total number of genes of a species are collectively known as its pan-genome. It was observed from several evidences that a sole organism can't contain all the genes of a species due to variability present in the genomic sequences. The desirable features of an ideal pan-genome are completeness (i.e., contains all functional genes), stability (i.e., unique catechistic features), comprehensibility (i.e., contains all the genomic information of all the species or individuals), and efficacy (i.e., organized data structure). Pangenome information of a species helps in the identification of desired alleles, rare alleles, presence or absence of a traits in a species. Recently a chickpea pangenome of 592.58 Mb was constructed which containsa total of 29,870 genes (Varshney et al., 2021). The pan-genome was constructed using whole-genome sequencing using 3,366 comprising 3,171 cultivated and 195 wild accessions. Assembly was done by combining the CDC frontier reference genome including 53.60 Mb from cultivated chickpea inclusive of 2.93 Mb from ICC 4958 and 5.28 Mb from 28 accessions of *C. reticulatum*. This pan-genome analysis revealed useful information on genomic regions more often selected during the domestication process, superior haplotypes, and targets for purging deleterious alleles. The new genes identified encoding responses to oxidative stress, response to stimuli, heat shock proteins, cellular response to acidic pH, and response to cold, which could have a possible contribution to the adaptation of chickpea.

3.2.9 QTL Mappings, Their Introgression and Utilization for Broadening the Genetic Bases

The utility of the fundamental assumption of locus finding by co-segregation of characteristics with markers is enhanced by new

TABLE 5 | List of QTLs for various traits in chickpea.

S. No.	Trait	Linkage group	QTL	Position	Reference
Phenological traits					
1.	Plant height	*LG01*	*qPH1.1*	56.984–57.223	Kujur et al., 2016
2.	Plant height	*LG02*	*qPH2.1*	24.496–29.852	
3.	Plant height	*LG03*	*qPH3.1*	26.358–26.536	
4.	Plant height	*LG04*	*qPH4.1*	28.738–28.796	
5.	Plant height	*LG07*	*qPH7.1*	28.738–28.796	
6.	Plant height	*LG08*	*qPH8.1*	44.194–44.882	
7.	Plant height	*LG04*	*qPLHT4.1*	216.23–223.07	
8.	Plant height	*LG01*	*qPLHT1.1*	12.70–13.40	Barmukh et al., 2021
9.	Plant height	*LG04*	*qPLHT4.1*	216.23–223.07	
10.	Plant height	*LG05*	*qPLHT5.1*	1.07–7.62	
11.	Plant height	*LG08*	*qPLHT8.1*	13.74–14.30	
12.	No of primary branches	*LG02*	*qPB.2.1*	111.10–111-40	
13.	No of primary branches	*LG03*	*qPB.3.1*	14.3014.40	
14.	Flowering time	*LG03*	*Qefl1*-1	0.00	Mallikarjuna et al., 2017
15.	Flowering time	*LG04*	*Qefl1*-2	41.00	
16.	Flowering time	*LG01*	*Qefl2*-1	15.00	
17.	Flowering time	*LG03*	*Qefl2*-2	21.00	
18.	Flowering time	*LG04*	*Qefl2*-2	55.00	
19.	Flowering time	*LG08*	*Qefl1*-3	15.00	
20.	Flowering time	*LG03*	*Qefl2*-4	5.00	
21.	Flowering time	*LG03*	*Qefl3*-1	31.00	
22.	Flowering time	*LG08*	*Qefl3*-2	2.00	
23.	Flowering time	*LG06*	*Qefl4*-1	9.00	
24.	Days to flowering initiation (DFI)	*LG06*	*CaDFI_LS6.1*	37.11	Jha et al., 2021
25.	DFI	*LG08*	*CaDFI_LS8.1*	42.71	
26.	DFI	*LG06*	*CaDFI_LS6.1*	37.11	
27.	DFI	*LG08*	*CaDFI_LS6.1*	42.71	
28.	Days to maturity (DM)	*LG01*	*CaDMI_LS1.1*	7.11	
29.	DM	*LG01*	*CaDMI_LS1.2*	152.61	
30.	DM	*LG01*	*CaDMI_LS1.3*	154.81	
Yield and related traits					
31.	Days to pod initiation (DPI)	*LG07*	*CaDPI_LS7.2*	98.01	Jha et al., 2021
32.	DPI	*LG07*	*CaDPI_LS7.1*	97.01	
33.	DPI	*LG06*	*CaDPI_LS6.1*	37.11	
34.	DPI	*LG06*	*CaDPI_LS6.1*	37.11	
35.	DPI	*LG06*	*CaDPI_LS6.1*	37.11	
36.	DPI	*LG01*	*CaDPI_LS1.1*	153.61	
37.	Days to pod filling (DPF)	*LG08*	*CaDPF_LS8.1*	67.41	Jha et al., 2021
38.	DPF	*LG04*	*CaDPF_NS4.2*	136.61	
39.	DPF	*LG04*	*CaDPF_NS4.1*	138.11	
40.	No of filled pods (FP)	*LG06*	*CaFP_NS6.1*	141.40	
41.	100 seed weight (g)	*LG06*	*Ca100SW_LS7.1*	97.01	Jha et al., 2021
42.	100 seed weight (g)	*LG01*	*Ca100SW_LS1.1*	46.21	
43.	100 seed weight (g)	*LG04*	*Ca100SW_LS4.1*	159.71	
44.	100 seed weight (g)	*LG07*	*Ca100SW_LS7.1*	97.01	
45.	100 seed weight (g)	*LG06*	Q100SW6.1	43.66–43.70	Barmukh et al., 2021
46.	100 seed weight (g)	*LG07*	Q100SW7.1	47.61–47.77	
47.	100 seed weight (g)	*LG03*	Q100SW3.1	153.40–167.6	
48.	100 seed weight (g)	*LG06*	Q100SW6.2	87.91–88.02	
49.	100 seed weight (g)	*LG07*	Q100SW7.2	139.78–140.04	
50.	100 seed weight (g)	*LG04*	Q100SW4.1	216.23–223.07	
51.	Seed yield/plant (g)	*LG02*	*CaSYPP_LS2.1*	22.51	Jha et al., 2021
52.	Seed yield/plant (g)	*LG06*	*CaSYPP_LS6.1*	12.21	
53.	Seed yield/plant (g)	*LG06*	*CaSYPP_NS6.2*	52.31	
54.	Seed yield/plant (g)	*LG06*	*CaSYPP_NS6.3*	53.01	
55.	Seed yield/plant (g)	*LG04*	qYPP4.1	86.44–87.52	Barmukh et al., 2021
56.	Seed yield/plant (g)	*LG01*	qYPP1.1	15.00–46.80	
57.	Pods per plant	LG06	qPPP6.1	0.75–1.27	Barmukh et al., 2021
58.	Biological yield/plant		CaBYPP_NS6.1	52.31	Jha et al., 2021
59.	Biological yield/plant	*LG06*	CaBYPP_NS6.1	52.31	

(Continued on following page)

TABLE 5 | (*Continued*) List of QTLs for various traits in chickpea.

S. No.	Trait	Linkage group	QTL	Position	Reference
60.	Biological yield/plant	*LG06*	CaBYPP_LS6.3	114.01	
61.	Biological yield/plant	*LG02*	CaBYPP_LS2,1	55.91	
62.	Biological yield/plant	*LG06*	CaBYPP_LS6.4	115.01	
63.	Biological yield/plant	*LG06*	CaBYPP_LS6.5	115.31	
64.	Biological yield/plant	*LG06*	CaBYPP_NS6.2	58.71	
65.	Harvest index (HI %)	*LG05*	CaHI_NS5.1	42.11	
66.	HI %	*LG07*	CaHI_NS7.1	35.81	
67.	HI %	*LG06*	CaBYPP_NS6.3	170.81	
68.	HI %	*LG06*	CaHI_LS6.2	100.21	
69.	HI %	*LG08*	CaHI_LS8.1	43.11	
70.	HI %	*LG07*	CaHI_NS7.2	142.71	
71.	HI %	*LG06*	CaHI_NS6.1	84.21	
72.	HI %	*LG07*	CaHI_NS7.1	35.81	
Physiological traits					
73.	Chlorophyll Content (CHL, ng/mm^2)	*LG04*	CaCHL_NS4.3	151.51	
74.		*LG04*	CaCHL_NS4.3	151.51	
75.		*LG02*	CaCHL_LS2.1	38.31	
76.	CHL, ng/mm^2	*LG05*	CaCHL_LS5. 1	44.01	
77.	CHL, ng/mm^2	*LG05*	CaCHL_LS5.2	44.31	
78.	CHL, ng/mm^2	*LG04*	CaCHL_NS4.1	142.91	
79.	CHL, ng/mm^2	*LG04*	CaCHL_NS4.2	150.11	
80.	Cell membrane stability (CMS %)	*LG04*	CaCMS_NS4.1	133.61	Jha et al., 2021
81.	CMS %	*LG06*	CaCMS_LS6.1	67.21	
82.	CMS %	*LG03*	CaCMS_NB3.1	0.01	
83.	Nitrogen balance index (NBI)	*LG08*	CaNBI_LS8.3	3.81	Jha et al., 2021
84.	NBI	*LG08*	CaNBI_LS8.1	0.01	
85.	NBI	*LG08*	CcNBI_LS8.2	1.01	
86.	NBI	*LG07*	CaNBI_LS7.2	97.01	
87.	NBI	*LG07*	CaNBI_LS7.1	34.61	
88.	NBI	*LG08*	CaNBI_LS8.2	1.01	
89.	NBI	*LG06*	CaNBI_LS6.1	69.71	
90.	NBI	*LG08*	CaNBI_LS8. 1	0.01	
91.	NBI	*LG06*	CaNBI_LS6.2	70.71	
92.	NBI	*LG07*	CaNBI_LS7.1	34.61	
93.	Normalized difference vegetation index (NDVI)	*LG02*	CaNDVI_LS2.2	66.01	Jha et al., 2021
94.	NDVI	*LG04*	CaNDVI_NS4.1	68.31	
95.	NDVI	*LG04*	CaNDVI_NS4.2	69.21	
96.	NDVI	*LG02*	CaNDVI_LS2.1	65.41	
97.	NDVI	*LG02*	CaNDVI_LS2.2	66.01	
98.	NDVI	*LG03*	CaNDVI_NS3.1	48.41	
99.	NDVI	*LG08*	CaNDVI_NS8.1	18.61	
100.	NDVI	*LG08*	CaNDVI_NS8.2	18.91	
101.	NDVI	*LG03*	CaNDVI_NS3.1	48.41	
102.	NDVI	*LG06*	CaNDVI_NS6.1	20.01	
103.	NDVI	*LG01*	CaNDVI_LS1.2	44.21	
104.	NDVI	*LG01*	CaNDVI_LS1.1	42.21	
105.	NDVI	*LG05*	CaNDVI_NS5.2	36.11	
106.	NDVI	*LG05*	CaNDVI_NS5.1	35.11	
107.	NDVI	*LG04*	CaNDVI_NS4.1	68.31	

permutations of QTL mapping (**Table 5**). However, the definition of a trait can now be expanded beyond whole-organism phenotypes to include phenotypes like the amount of RNA transcript or protein produced by a specific gene because these phenotypes have more typical organismal characteristics viz.; yield in corn are polygenic and QTL mapping works in these situations. Transcript abundance is regulated not only by cis-acting regions like the promoter but also by Transacting transcription factors that may or may not be related. Similarly, local variation at the coding gene and distant variation mapping to other areas of the genome control protein abundance. Local variation is most likely made up of cis variations that regulate transcript levels. Polymorphisms for the protein's stability or control could be another local mechanism. Distant variation, on the other hand, could comprise upstream regulatory control areas (Upadhyaya et al., 2016).

Quantitative trait loci (QTLs) conferring resistance to biotic and abiotic stresses have been applied in chickpeas in the last 2 decades and the molecular markers closely associated with these loci are also located (Santra et al., 2000). For example, several

QTLs conferring Ascochyta blight resistance are identified, and several MAS (SCY17 and SCAE19) were reported as the best markers linked to AB-resistant genes. These two markers were validated on different populations (Iruela et al., 2006; Imtiaz et al., 2008; Madrid et al., 2014). More recently, three major conserved quantitative trait loci (QTLs) that confer AB resistance have been reported, two on chromosome Ca2 and one on chromosome Ca4. These QTLs explained a maximum of 18.5%, and 25% of the total variation. In total, 27 predicted genes were located in chromosome IV close to these QTL (Hamwieh et al., Unpublished data).

The 20 QTLs and candidate genes associated with seed traits were also identified in chickpeas using the GBS approach (Pavan et al., 2017). In pigeon pea, the GBS-based mapping of two RIL populations led to the identification of QTLs and candidate genes for resistance to fusarium wilt (FW) and sterility mosaic disease (SMD) (Saxena et al., 2017) in addition to restoration of fertility (Rf) (Saxena et al., 2018), using GWAS drought tolerance-related traits in chickpea (Kale et al., 2015), flowering time control, seed development and pod dehiscence in pigeon pea (Varshney et al., 2017) have been mapped. The GBS has been utilized in the fine mapping of the "*QTL-hotspot*" region for drought tolerance-related traits in chickpeas (Kale et al., 2015). In the case of chickpea, QTL seq approach has successfully identified a major genomic region (836,859–872,247 bp) on Ca1 chromosome which was further narrowed down to a 35-kb region harboring six candidate genes for 100 seed weight (Das et al., 2015).

Plant breeding can help in solving the global problem of micronutrient deficiencies in a cost-effective and long-term manner. The development of biofortified chickpea varieties is aided by evaluating cultivars for micronutrient contents and identifying quantitative trait loci (QTLs)/genes and markers. The $F_{2:3}$ derived population resulting from a cross between MNK-1 and Annigeri-1 was dissected employing the GBS technique and concentrations of Fe and Zn were examined with the goal of determining the responsible genetic areas (Vandemark et al., 2018). The researchers mapped 839 SNPs on an intra-specific genetic linkage map covering a total distance of 1,088.04 cM with a marker density of 1.30 cM. By combining linkage map data with phenotypic data from the F2:3 populations a total of 11 QTLs for seed Fe concentration on CaLG03, CaLG04, and CaLG05 with phenotypic variance varying from 7.2% (CaqFe3.4) to 13.4% (CaqFe3.4; CaqFe4.2). On CaLG04, CaLG05, and CaLG08 along with eight QTLs for seed Zn concentration with explained phenotypic variances ranging from 5.7% (CaqZn8.1) to 13.7% (CaqZn4.3) were discovered (Pandey et al., 2016).

The identification of marker-trait association between a genetic marker and a trait of interest is the initial stride in crop breeding utilizing molecular breeding/genomics assisted breeding. For initial experiments, linkage maps were created employing F_2 populations. The inter-specific cross *C. arietinum* (ICC 4958) x *C. reticulatum* (PI 489777) was employed to create the first recombinant inbred lines (RILS) mapping population which is now being used as a chickpea reference mapping population for genome mapping (Nayak et al., 2010). Maps created from intra-specific mapping populations have a smaller number of markers (<250 markers) and poorer genome coverage (<800 cM) due to minimal variation in the cultivated chickpea. Consensus genetic maps were also created utilizing both inter and intra-specific mapping populations.

The genetic mapping of QTLs affecting resistance to various diseases, and also vital agronomical traits, in chickpea are extensively documented. Santra et al. (2000) identified two quantitative trait loci (QTL1 and QTL2) that give resistance to Ascochyta blight. These QTLs were predicted to be responsible for overall phenotypic variance (34.4%, 14.6%), respectively (Santra et al., 2000; Tekeoglu et al., 2002). Comparative protein profiling of wild chickpeas and induced mutants was carried out in order to measure genetic diversity between mutants and parental genotypes (Patil and Kamble, 2014). Kujur et al. (2016) reported candidate genes and natural allelic variations for QTLs determining plant height, which was followed by the discovery of QTLs for heat distress response (Paul et al., 2018) as well as photosynthetic efficiency attributes for boosting seed yield in chickpea using GWAS and expression profiling (Basu et al., 2019). These discoveries have opened up new paths for analysis and comprehensive characterization of wild *Cicer* species, which will help in harnessing unidentified allelic variations to extend the genetic foundation of cultivars.

Molecular markers have been discovered for gene(s)/QTL(s) linked to abiotic stress resistances, viz., drought tolerance (Molina et al., 2008; Rehman et al., 2012), salinity resilience (Vadez et al.,2012), biotic stresses, viz., Ascochyta blight (Milla´n et al., 2003; Iruela et al., 2006; Aryamanesh et al., 2010; Garg et al., 2019), Fusarium wilt (Cobos et al., 2005; Gowda et al., 2009; Sabbavarapu et al., 2013) and botrytis gray mold (Anuradha et al., 2011) along with seed characteristics (Gowda et al., 2009) in chickpea. These technologies can be employed to improve chickpea genetics and breeding as well as to explain the variety of the chickpea genome and domestication events. Furthermore, genomic selection has been presented as a promising strategy for enhancing traits that are influenced by a large number of gene (s)/QTL (s) (Bajaj et al., 2015a; Bajaj et al., 2015b). Both phenotypic and genotypic data sets are employed in this approach to determine genomic estimated breeding values (GEBV) of improved progenies.

3.2.10 Genome-Wide Association Studies for Broadening the Genetic Bases

GWAS have become one of the most important genetic methods for analyzing complicated trait QTLs and underlying genes. Many studies have shown that GWAS can be used to map more authentically new genes implicated in complex agronomic variables in plants. Given this, linkage disequilibrium (LD), population substructure, and imbalanced allele frequencies are the key drawbacks of GWAS. Many markers associated with tolerance to abiotic stresses have been also reported in chickpea. In brief, the germplasm of 186 chickpea genotypes has been genotyped with 1856 DArTseq markers. The association with the salinity tolerance in the field (Arish, Sinai, Egypt) and the greenhouse by using hydroponic system at 100 mM NaCl concentration indicated one locus on chromosome Ca4 at 10,618,070 bp associated with salinity tolerance, in addition to another locus-specific to the hydroponic system on chromosome Ca2 at 30,537,619 bp. The gene annotation analysis revealed the location of rs5825813 within the

TABLE 6 | List of engineered genes/traits in chickpea.

Crops	Genotype	Explant	Transgene	Promoter	Gene delivery system	Aim	References
Chickpea	C 235, BG 256, Pusa 362 and Pusa 372	Cotyledonary node	*cry1Ac*	*CaMV35S*	Agrobacterium-mediated	Insect resistance against *H. armigera*	Sanyal et al. (2005)
	ICCC37	Epicotyl	*crylAc*	*CaMV35S*	Agrobacterium-mediated	Insect resistance against *H. armigera*	Indurker et al. (2007)
	Annigeri	Cotyledonary node	*P5CS*	*CaMV35S*	Agrobacterium-mediated	Salinity tolerance	Ghanti et al. (2011)
	P-362	Cotyledonary node	*cry1Ab* and *cry1Ac*	*CaMV35S* and synthetic constitutive expression promoter (*Pcec*)	Agrobacterium-mediated	Insect resistance	Mehrotra et al. (2011)
	DCP 92–3	Embryonic axis	*cry1Ab/cry1Ac*	Rice *actin1* and soybean *msg*	Agrobacterium-mediated	Insect resistance	Ganguly et al. (2014)
	Gokce	Mature embryo	*miR408*	CaMV35S	Agrobacterium-mediated	Drought tolerance	Hajyzadeh et al. (2015)
	ICCV 89,314	Single cotyledon with half embryo	*cry1Ac*	*RuBisCO* small subunit and ubiquitin	Agrobacterium-mediated	Insect resistance to target *H. armigera*	Chakraborty et al. (2016)
	DCP 92–3	Axillary meristem	*cry1Aabc*	*CaMV35S*	Agrobacterium-mediated	Insect resistance	Das et al. (2017)
	PBA HatTrick	Half-embryonic axis	nicotianamine synthase 2 and ferritin	*CaMV35S* and nopaline synthase	Agrobacterium-mediated	Iron biofortifcation	Tan et al. (2018)

TABLE 7 | Genetic transformation of chickpea.

Genotype	Explant	Transgene	Promoter	Gene delivery system	Aim	References
C 235, BG 256, Pusa 362 and Pusa 372	Cotyledonary node	*cry1Ac*	*CaMV35S*	Agrobacterium-mediated	Insect resistance against *H. armigera*	Sanyal et al. (2005)
ICCC37	Epicotyl	*crylAc*	*CaMV35S*	Agrobacterium-mediated	Insect resistance against *H. armigera*	Indurker et al. (2010)
Annigeri	Cotyledonary node	*P5CS*	*CaMV35S*	Agrobacterium-mediated	Salinity tolerance	Ghanti et al. (2011)
P-362	Cotyledonary node	*cry1Ab* and *cry1Ac*	*CaMV35S* and synthetic constitutive expression promoter (*Pcec*)	Agrobacterium-mediated	Insect resistance	Mehrotra et al. (2011)
DCP 92–3	Embryonic axis	*cry1Ab/cry1Ac*	Rice *actin1* and soybean *msg*	Agrobacterium-mediated	Insect resistance	Ganguly et al. (2014)
Gokce	Mature embryo	*miR408*	CaMV35S	Agrobacterium-mediated	Drought tolerance	Hajyzadeh et al. (2015)
ICCV 89,314	Single cotyledon with half embryo	*cry1Ac*	*RuBisCO* small subunit and ubiquitin	Agrobacterium-mediated	Insect resistance to target *H. armigera*	Chakraborty et al. (2016)
DCP 92–3	Axillary meristem	*cry1Aabc*	*CaMV35S*	Agrobacterium-mediated	Insect resistance	Das et al. (2017)
PBA HatTrick	Half-embryonic axis	nicotianamine synthase 2 and ferritin	*CaMV35S* and nopaline synthase	Agrobacterium-mediated	Iron biofortification	Tan et al. (2018)

Embryogenesis-associated protein (EMB8-like), while the location of rs5825939 is within the Ribosomal Protein Large P0 (RPLP0) (Ahmed et al., 2021). Utilizing such markers in practical breeding programs can effectively improve the adaptability of current chickpea cultivars in saline soil.

Besides the above-mentioned reports, GWAS has also been conducted for yield and related traits in chickpea (Li et al., 2021), root morphological traits (Thudi et al., 2021), nutrient content (Diapari et al., 2014; Sab et al., 2020) and abiotic tolerance traits (Thudi et al., 2014; Samineni et al., 2022). Thus, the associated genomic regions identified through GWAS could be used for breeding programs to improve yield-related traits, nutrient content, and biotic and abiotic stress tolerance in chickpea. Recently, in other studies, we have accomplished GWAS for nodule numbers in chickpea by conducting multi-locational phenotypic evaluations and have identified seven significant SNP IDs (Kumar et al. unpublished data).

3.2.11 Genetic Engineering for Broadening Genetic Bases

Genetic engineering has been widely utilized to select resistant gene(s) (**Table 6**) from various resources and transmit them to selected plants to introgress resistance to various abiotic as well as

TABLE 8 | Bioinformatics resources for chickpea.

Bioinformatics resources for chickpea	Description
1. CicArMiSatDB (https://cegresources.icrisat.org/CicArMiSatDB/)	CicArMiSatDB is a web resource for learning about Chickpea microsatellite (Simple Sequence Repeat) markers. It gives the chickpea breeding community useful marker information. This database can be used to find marker information and examine it using the BLAST and Genome Browser implementations
2. PulseDB (https://www.pulsedb.org/organism/641)	The Pulse Crop Database (PCD), formerly the Cool Season Food Legume Database (CSFL), is being developed by Washington State University's Main Bioinformatics Laboratory in collaboration with the USDA-ARS Grain Legume Genetics and Physiology Research Unit, the USDA-ARS Plant Germplasm Introduction and Testing Unit, the United States Dry Pea and Lentil Council, Northern Pulse Growers, and allied scientists in the United States and around the world, to serve as a resource for (GAB). By providing relevant genomic, genetic, and breeding information and analysis, GAB provides tools to find genes associated with features of interest, as well as other approaches to increase plant breeding efficiency and research
3. ACPFG Bioinformatics TAGdb (http://sequencetagdb.info/tagdb/cgi-bin/index)	This service performs BLAST alignment between a single query and short pair reads of selected species
4. The chickpea portal (http://www.cicer.info/)	In collaboration with partners in India (ICRISAT), this AISRF-funded project is focused on the development of efficient selection methods for tolerance to abiotic stress and the application of molecular tools to assist chickpea breeding
5. LIS ChickpeaMine (https://mines.legumeinfo.org/chickpeamine/begin.do)	This mine integrates data for chickpea varieties desi and kabuli. It is developed by LIS/NCGR and sourced from LIS datastore files
6. Chickpea Transcriptome Database (CTDB) (http://nipgr.res.in/ctdb.html)	It provides a full web interface for visualizing and retrieving chickpea transcriptome data. Many tools for similarity searches, functional annotation (putative function, PFAM domain, and gene ontology) searches, and comparative gene expression analyses are included in the database. The latest version of CTDB (v2.0) contains transcriptome datasets from farmed (desi and Kabuli kinds) and wild chickpea with high-quality functional annotation

biotic challenges. Various genes are now being deployed in pulses using *Agrobacterium*-mediated (Eapen et al., 1987; Krishnamurthy et al., 2000; Sharma K. K. et al., 2006), particle gun bombardment (Kamble et al., 2003; Indurker et al., 2007), electroporation of intact axillary buds (Chowrira et al., 1996) electroporation and PEG mediated transformation using protoplasts (Köhler et al., 1987a; Köhler et al., 1987b). The most widely used method for developing transgenics in pulse crops is Agrobacterium mediated explant transformation. To generate transgenic plants, numerous transgenes from various sources have been introduced into pulse crops.

Transgenic chickpea is developed either by gene gun (Kar et al., 1997; Husnain et al., 2000; Tewari-Singh et al., 2004; Indurker et al., 2007) or *Agrobacterium*-mediated method (Kar et al., 1997; Sanyal et al., 2005; Biradar et al., 2009; Acharjee et al., 2010; Asharani et al., 2011; Mehrotra et al., 2011; Ganguly et al., 2014). Important target traits for transgenic plant development in chickpea are insect pest resistance including α amylase inhibitor genes and lectin genes (Dita et al., 2006), Cry genes from *Bacillus thuringiensis*, protease inhibitor genes, disease resistance including transfer of genes such as chitinase gene, antifungal protein genes or stilbene synthase gene for fungal resistance, coat protein genes of viruses for viral resistance and bacterial resistance from T_4 lysozyme gene (Eapen, 2008), various abiotic stresses like salinity, drought, mineral toxicities, cold, temperature, etc., seed proteins, plant architecture, and RNA interference technology could be used to increase carotenoids and flavanoids by engineering metabolic pathways to decrease the effect of endogenous genes (Eapen, 2008).

As presented in **Table 7** transformation through Agrobacterium with the *cry1Ab/Ac* gene in chickpea has resulted in resistance to *Helicoverpa armigera* (Lawo et al., 2008; Ganguly et al., 2014). Bombardment of calli with DNA-coated tungsten particles resulted in somatic embryogenesis and the subsequent generation of transgenic chickpea (Husnain et al., 2000). Other researchers have also reported on the use of transgenic chickpea as a drought-tolerant and pest-resistant cultivar (Bhatnagar-Mathur et al., 2009; Khatodia et al., 2014; Kumar et al., 2014).

3.2.12 Bioinformatic Molecular Data Bases/Resources for Broadening Genetic Bases

The recent data reports on leguminous genomics and transcriptomics have forced the creation of an exhaustive model of legume genomics and transcriptomics databases. Readily available data through online database portals are playing a significant role in research and development. LegumeIP (http://plantgrn.noble.org/LegumeIP/), an integrative database for comparative genomics and transcriptomics of model legumes, for use in studying gene function and genome evolution in this center-stage plant family including the genome sequences of *M. truncatula*, *G. max* and *L. japonicas* and two reference plant species, i.e., *A. thaliana* and *Populus trichocarpa* were employed (Li et al., 2012). The Legume Information System (LIS; https://legumeinfo.org) (Dash et al., 2016) gives users access to genetic and genomic data for model legumes. KnowPulse (https://knowpulse.usask.ca) for chickpea, common bean, field pea, fababean, and lentil, focuses on diversity data and gives

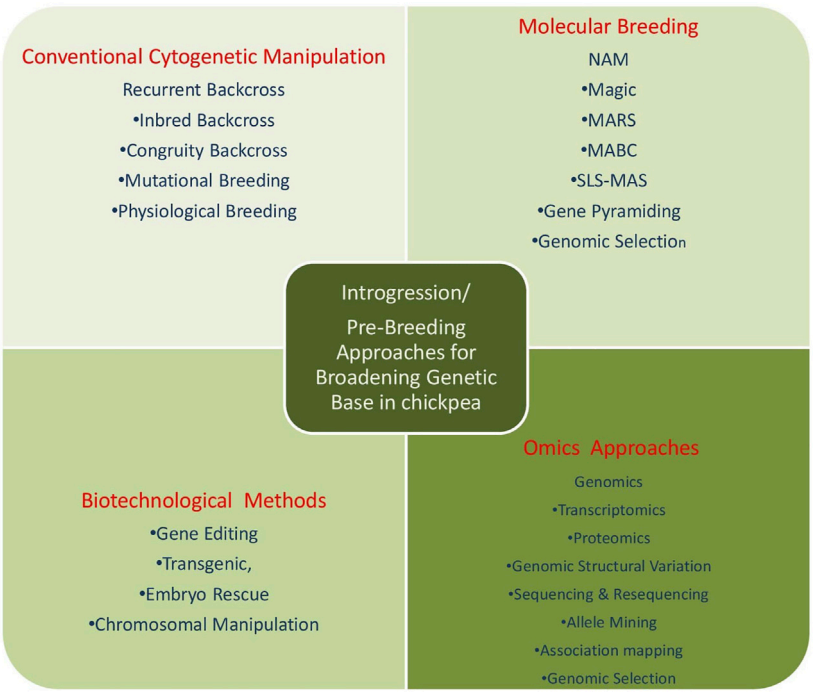


FIGURE 5 | Integrating various approaches for broadening the genetic base.

information on germplasm, genetic markers, sequence variants, and phenotypic traits (Sanderson et al., 2019).

The construction of bioinformatics databases (**Table 8**) for the chickpea gene pool, according to recent breakthroughs in computational genomics, will permit users to visualize and extract chickpea genomics data in order to learn comparative genomics, annotate gene function, and investigate novel transcription factors (Doddamani et al., 2015; Verma et al., 2015; Gayali et al., 2016). Many databases have been built for chickpea, including CicArMiSatDB (https://cegresources.icrisat.org/CicArMiSatDB/) for SSR markers (Doddamani et al., 2014), CicArVarDB (https://cegresources.icrisat.org/cicarvardb/) for SNPs and QTLs, and Chickpea Transcriptome Database (Verma et al., 2015). Furthermore, a few years ago, the PLncPRO tool was developed to acquire unique insights into the rising importance of long noncoding RNAs in response to various abiotic challenges in chickpea (Singh et al., 2017).

There are also other molecular databases developed in other pulse crops which are useful in comparative genomics studies. Some of the important databases are highlighted as further. The PIgeonPEa Microsatellite DataBase (PIPEMicroDB) program (http://cabindb.iasri.res.in/pigeonpea/) stores a catalogue of microsatellites retrieved from the pigeon pea genome (Sarika et al., 2013). The adaptation of this program for chromosome-based search may be utilized for QTL markers for crop improvement and mapping of genes. With the fast development of publicly available Affymetrix GeneChip Medicago Genome Array Gene Chip data from cell types, a wide range of tissues, growth conditions, and stress treatments, the legume research group is in need of an efficient bioinformatics system to assist efforts to analyze the *Medicago* genome through functional genomics. The MtGEA (*Medicago truncatula* Gene Expression Atlas) website (http://bioinfo.noble.org/gene-atlas/) now includes additional gene expression data and genome annotation (He et al., 2009). The *Medicago truncatula* Genome Database (http://www.medicagogenome.org) houses a diverse collection of genomic data sets (Krishnakumar et al., 2015). RNA-Seq Atlas (Seq-Atlas) for *Glycine max* (http://www.soybase.org/soyseq) gathers RNASeq data from a variety of tissues and offers new techniques for analyzing huge transcriptome data sets produced from next-generation sequencing (Severin et al., 2010). SoyBase (https://www.soybase.org/), the USDA-ARS soybean genetic database, is a comprehensive library of professionally maintained soybean genetics, genomics, and related data resources (Grant et al., 2010). The *Lotus japonicus* Gene Expression Atlas (LjGEA: http://ljgea.noble.org/) provides a global picture of gene expression in organ systems of the species including roots, nodules, stems, petioles, leaves, flowers, pods, and seeds. It enables versatile, multifaceted transcriptome analysis (Verdier et al., 2013).

3.2.13 Genome Editing for Broadening Genetic Bases

Genome editing promises giant leaps forward in broadening the genetic bases research. Targeted DNA integration into known locations in the genome has potential advantages over the random insertional events typically achieved using conventional means of genetic modification. The gene of interest is positioned near the T-DNA left border which is responsible for the insertion of plant cell. Molecular biologists can now more accurately target any gene of interest because advances in genome editing tools such as zinc-finger nucleases (ZFNs), homing endonuclease and transcription activator-like effector nucleases (TALENs) could

possibly be exploited for genomics-assisted selection toward accelerated genetic gains (Shan et al., 2013; Bortesi and Fischer, 2015), while more advancements in chickpea enhancement using these cutting-edge approaches are still awaited. In chickpea, the 4-coumarate ligase (4CL) and Reveille 7 (RVE7) genes were selected as genes associated with drought tolerance for CRISPR/Cas9 editing in chickpea protoplast. The knockout of these selected genes in the chickpea protoplast showed high-efficiency editing was achieved for RVE7 gene *in vivo* compared to the 4CL gene (Badhan et al., 2021). These methods, however, are costly and time-consuming since they need complex procedures that require protein engineering. Unlike first-generation genome editing techniques, CRISPR/Cas9 genome editing is straightforward to design and clone and the same Cas9 can theoretically be used with various guide RNAs targeting many places in the genome. Several proof-of-concept demonstrations in crop plants using the primary CRISPR-Cas9 module, and numerous customized Cas9 cassettes have been used to improve target selectivity and reduce off-target cleavage. Thus, the applications of genome editing techniques in chickpea research have great potential (Mahto et al., 2022).

4 INTEGRATING VARIOUS OMICS APPROACHES FOR BROADENING THE CHCKPEA GENETIC BASE

The technological advances that transformed chickpea from an orphan crop to a genomic resource enriched crop in the post-genomics era, Re-sequencing efforts using WGRS have led to the dissection of genetic diversity, population structure, domestication patterns, linkage disequilibrium and the unexploited genetic potential for chickpea improvement (Varshney et al., 2019). Modern genomics technologies have the potential to speed up the process for trait mapping, gene discovery, marker development and molecular breeding, in addition to enhancing the rate of productivity gains in chickpea. Integration of genome-wide sequence information with precise phenotypic variation allows capturing accessions with low-frequency variants that may be responsible for essential phenotypes such as yield components, abiotic stress tolerance, or disease resistance (Roorkiwal et al., 2020). NGS technology has resulted in the development and application of a wide variety of molecular markers for chickpea improvement (Kale et al., 2015; Varshney et al., 2018). Over the past decade, more than 2000 simple sequence repeat (SSR) markers, 15,000 features-based diversity array technology (DArT) platform, and millions of SNP markers have been developed for chickpea (Varshney 2016). The revolution in NGS technologies has enabled sequencing to be performed at a higher depth (whole-genome re-sequencing), mid-depth (skim sequencing), or lower depth (genotyping by sequencing, RAD-Seq). Integrating omics data from multiple platforms such as transcriptomics, proteomics and metabolomics are paramount to bridging the genome-to-phenome gap in crop plants and ultimately identifying the phenotype based on their genetics. applications of genomic technologies for bridging the genotype–phenotype gap in chickpea (**Figure 5**). With the availability of the reference genome, these genetic resources can be subjected to whole-genome re-sequencing (WGRS) or high- to low-density genotyping, based on the objective of the study, using the available genotyping platforms (e.g., genotyping by sequencing, GBS; array-based genotyping). Analysis at the transcriptome, proteome, and metabolome levels can be performed to gain novel insights into the candidate genes and biological processes involved. Using a genomics approach Fusarium wilt resistance WR 315 Annigeri 1 foc4 has been Released as "Super Annigeri 1′ for commercial cultivation in India Mannur et al. (2019).

5 CONCLUSION AND FUTURE PERSPECTIVE

With the employment of modern "Omics" technologies in combination with traditional methods, it is now possible to overcome yield limits, and achieve higher genetic gains ensuring high output for chickpea production and quality features. Chickpea land races and wild *Cicer* species are the goldmines of beneficial genes influencing desired traits of interest for biotic, abiotic, and yield component features. Identification of novel sources of desired traits, QTLs or alleles through extensive evaluation and utilization of landraces and wild *Cicer* species will have a greater impact on developing chickpeas for better climate resilience and higher yield. Many desirable features from primary and secondary gene pools in wild *Cicer* species have been successfully transmitted into cultivated cultivars using both traditional and modern procedures. The wealth of new omics approaches and growing resources offer great potential to transform chickpea breeding in the near future. An integrated application of chickpea "Omics", classical and modern breeding methods, marker-assisted selection, and biotechnological application promises for the broadening of the chickpea genetic base and introgression of new genes for crop traits for higher productivity will lead to next-generation chickpea varieties.

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.

AUTHOR CONTRIBUTIONS

RK conceptualized and supervised the manuscript writing. RS, CS, Ambika, BC, RM, RP, and AG collected the related literature and contributed to the original writing. VG, Gayacharan, AH, HU, and RK extended their help in the inference, review, and editing of the manuscript. All authors went through the final manuscript draft and approved it.

ACKNOWLEDGMENTS

ICARDA authors received support from CWANA integrated initiative.

REFERENCES

Abbo, S., Berger, J., and Turner, N. C. (2003). Viewpoint: Evolution of Cultivated Chickpea: Four Bottlenecks Limit Diversity Andconstrain Adaptation. *Funct. Plant Biol.* 30 (10), 1081–1087. doi:10.1071/fp03084

Abdallah, C., Dumas-Gaudot, E., Renaut, J., and Sergeant, K. (2012). Gel-based and Gel-free Quantitative Proteomics Approaches at a Glance. *Int. J. Plant Genomics* 2012, 494572. doi:10.1155/2012/494572

Acharjee, S., Sarmah, B. K., Kumar, P. A., Olsen, K., Mahon, R., Moar, W. J., et al. (2010). Transgenic Chickpeas (*Cicer Arietinum* L.) Expressing a Sequence-Modified cry2Aa Gene. *Plant Sci.* 178 (3), 333–339. doi:10.1016/j.plantsci.2010.02.001

Agarwal, G., Jhanwar, S., Priya, P., Singh, V. K., Saxena, M. S., Parida, S. K., et al. (2012). Comparative Analysis of Kabuli Chickpea Transcriptome with Desi and Wild Chickpea Provides a Rich Resource for Development of Functional Markers. *PLoS ONE* 7, e52443. doi:10.1371/journal.pone.0052443

Aguilera, Y., Esteban, R. M., Benítez, V., Mollá, E., and Martín-Cabrejas, M. A. (2009). Starch, Functional Properties, and Microstructural Characteristics in Chickpea and Lentil as Affected by Thermal Processing. *J. Agric. Food Chem.* 57 (22), 10682–10688. doi:10.1021/jf902042r

Ahmad, F., and Slinkard, A. E. (2003). Limitations to Bridge-Species Facilitated Alien Gene Transfers in Chickpea: Pre-fertilization Events [*Cicer Arietinum* L.]. *J. Genet. Breed. (Italy)*.

Ahmad, F., Gaur, P. M., and Slinkard, A. E. (1992). Isozyme Polymorphism and Phylogenetic Interpretations in the Genus Cicer L. *Theor. Appl. Genet.* 83 (5), 620–627. doi:10.1007/bf00226907

Ahmad, F., Gaur, P. M., and Croser, J. (2005). Chickpea (*Cicer Arietinum* L.). *Genet. Resour. Chromosome Eng. crop Improve.-Grain Legumes* 1, 187–217. doi:10.1201/9780203489284.ch7

Ahmad, S., Khan, M. A., Sahi, S. T., and Ahmad, R. (2013). Evaluation of Chickpea Germplasm against Ascochyta Rabiei (Pass) Lab. *J. Anim. Plant Sci.* 23 (2), 440–443.

Ahmad, F. (1999). Random Amplified Polymorphic DNA (RAPD) Analysis Reveals Genetic Relationships Among the Annual Cicer Species. *Theor. Appl. Genet.* 98 (3), 657–663. doi:10.1007/s001220051117

Ahmed, S. M., Alsamman, A. M., Jighly, A., Mubarak, M. H., Al-Shamaa, K., et al. (2021). Genome-wide Association Analysis of Chickpea Germplasms Differing for Salinity Tolerance Based on DArTseq Markers. *Plos one* 16 (12), e0260709. doi:10.1371/journal.pone.0260709

Alajaji, S. A., and El-Adawy, T. A. (2006). Nutritional Composition of Chickpea (*Cicer Arietinum* L.) as Affected by Microwave Cooking and Other Traditional Cooking Methods. *J. Food Compos. Anal.* 19, 806–812. doi:10.1016/j.jfca.2006.03.015

Ali, L., Deokar, A., Caballo, C., Tar'an, B., Gil, J., Chen, W., et al. (2016). Fine Mapping for Double Podding Gene in Chickpea. *Theor. Appl. Genet.* 129 (1), 77–86. doi:10.1007/s00122-015-2610-1

Ambika, Hegde, V., Nimmy, M. S., Bharadwaj, C., Tripathi, S., Singh, R. K., et al. (2021). Unraveling Genetics of Semi-determinacy and Identification of Markers for Indeterminate Stem Growth Habit in Chickpea (*Cicer Arietinum* L.). *Sci. Rep.* 11 (1), 1–8. doi:10.1038/s41598-021-01464-3

Anuradha, C., Gaur, P. M., Pande, S., Gali, K. K., Ganesh, M., Kumar, J., et al. (2011). Mapping QTLs for Resistance to Botrytis Grey Mould in Chickpea. *Euphytica* 182 (1), 1–9. doi:10.1007/s10681-011-0394-1

Arbona, V., Manzi, M., Ollas, C. D., and Gómez-Cadenas, A. (2013). Metabolomics as a Tool to Investigate Abiotic Stress Tolerance in Plants. *Int. J. Mol. Sci.* 14 (3), 4885–4911. doi:10.3390/ijms14034885

Archak, S., Tyagi, R. K., Harer, P. N., Mahase, L. B., Singh, N., Dahiya, O. P., et al. (2016). Characterization of Chickpea Germplasm Conserved in the Indian National Genebank and Development of a Core Set Using Qualitative and Quantitative Trait Data. *Crop J* 4 (5), 417–424. doi:10.1016/j.cj.2016.06.013

Aryamanesh, N., Nelson, M. N., Yan, G., Clarke, H. J., and Siddique, K. H. M. (2010). Mapping a Major Gene for Growth Habit and QTLs for Ascochyta Blight Resistance and Flowering Time in a Population between *C. Arietinum* and *C. reticulatum*. *Euphytica* 173, 307–331. doi:10.1007/s10681-009-0086-2

Asharani, B. M., Ganeshaiah, K. N., Raja, A., Kumar, V., and Makarla, U. K. (2011). Transformation of Chickpea Lines with *Cry1X* Using in Planta Transformation and Characterization of Putative Transformants T1 Lines for Molecular and Biochemical Characters. *J. Plant Breed. Crop Sci.* 3 (16), 413–423. doi:10.5897/jpbcs11.074

Aykroyd, W. R., and Doughty, J. (1982). "Legumes in Human Nutrition," in *FAO Food and Nutrition Paper* (FAO of the United Nations), 20, 1–152.

Badami, P. S., Mallikarjuna, N., and Moss, J. P. (1997). Interspecific Hybridization between I and *C. Pinnatifidum*. *Plant Breed.* 116 (4), 393–395. doi:10.1111/j.1439-0523.1997.tb01019.x

Badhan, S., Ball, A. S., and Mantri, N. (2021). First Report of CRISPR/Cas9 Mediated DNA-free Editing of 4CL and RVE7 Genes in Chickpea Protoplasts. *Int. J. Mol. Sci.* 22 (1), 396. doi:10.3390/ijms22010396

Badshah, A., Khan, M., Bibi, N., Khan, M., and Ali, S. (2003). Quality Studies of Newly Evolved Chickpea Cultivars. *Adv. FoodSci.* 25 (3), 96–99.

Bajaj, D., Saxena, M. S., Kujur, A., Das, S., Badoni, S., Tripathi, S., et al. (2015a). Genome-wide Conserved Non-coding Microsatellite (CNMS) Marker-Based Integrative Genetical Genomics for Quantitative Dissection of Seed Weight in Chickpea. *J. Exp. Bot.* 66, 1271–1290. doi:10.1093/jxb/eru478

Bajaj, D., Das, S., Badoni, S., Kumar, V., Singh, M., Bansal, K. C., et al. (2015b). Genome-wide High-Throughput SNP Discovery and Genotyping for Understanding Natural (Functional) Allelic Diversity and Domestication Patterns in Wild Chickpea. *Sci. Rep.* 5, 12468. doi:10.1038/srep12468

Bajaj, D., Upadhyaya, H. D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Identification of Candidate Genes for Dissecting Complex Branch Number Trait in Chickpea. *Plant Sci.* 245, 61–70. doi:10.1016/j.plantsci.2016.01.004

Barmukh, R., Soren, K. R., Madugula, P., Gangwar, P., Shanmugavadivel, P. S., Bharatwaj, C., et al. (2021). Construction of a High-Density Genetic Map and QTL Analysis for Yield, Yield Components and Agronomic Traits in Chickpea (*Cicer Arietinum* L.). *Plos one* 16 (5), e0251669. doi:10.1371/journal.pone.0251669

Basandrai, A. K., Macleod, W. J., Siddique, K., Pande, S., and Payne, P. (2006). Evaluation of Wild Chickpea Germplasm against Australian Isolates of *Botrytis Cinerea* the Casual Organism of Botrytis Grey Mold of Chickpea. BGM Newsl.

Basandrai, A. K., Pande, S., and Basandrai, D. (2008). "Recent Advance in Host-Parasite Interaction, Chickpea *Ascochyta Rabaie* and *Botrytis Cinerea* Systems," in *Crop Improvement, Strategies and Applications*. Editors R. C. Setia, H. Nayyar, and N. Setia (New Delhi India: IK International), 265–301.

Basu, U., Bajaj, D., Sharma, A., Malik, N., Daware, A., Narnoliya, L., et al. (2019). Genetic Dissection of Photosynthetic Efficiency Traits for Enhancing Seed Yield in Chickpea. *Plant, Cell Environ.* 42 (1), 158–173. doi:10.1111/pce.13319

Benzohra, I. E., Bendahmane, B. S., Benkada, M. Y., and Labdi, M. (2014). Evaluation of Wild *Cicer* Species Accessions for Resistance to Three Pathotypes of *Ascochyta Rabiei* (Pass.) Labr.in Algeria. *Afr. J. Micro. Res.* 8 (20), 2022–2029. doi:10.5897/AJMR2014.6628

Berger, J. D., Buck, R., Henzell, J. M., and Turner, N. C. (2005). Evolution in the Genus Cicer—Vernalisation Response and Low Temperature Pod Set in Chickpea (C. Arietinum L.) and its Annual Wild Relatives. *Aust. J. Agric. Res.* 56 (11), 1191–1200. doi:10.1071/ar05089

Bhatnagar-Mathur, P., Vadez, V., Jyotsna, D. M., Lavanya, M., Vani, G., and Sharma, K. K. (2009). Genetic Engineering of Chickpea with the *P5CSF129A* Gene for Osmoregulation with Implications on Drought Tolerance. *Mol. Breed.* 23, 591–606. doi:10.1007/s11032-009-9258-y

Biradar, S. S., Sridevi, O., and Salimath, P. M. (2009). Genetic Enhancement of Chickpea for Pod Borer Resistance through Expression of *CryIAc* Protein. *Karnataka J. Agric. Sci.* 22 (3), 467–470.

Bortesi, L., and Fischer, R. B. (2015). The CRISPR/Cas9 System for Plant Genome Editing and beyond. *Biotech. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006

Caballo, C., Castro, P., Gil, J., Millan, T., Rubio, J., and Die, J. V. (2019). Candidate Genes Expression Profiling during Wilting in Chickpea Caused by *Fusarium Oxysporum* F. Sp. Ciceris Race 5. *PLoSONE* 14 (10), e0224212. doi:10.1371/journal.pone.0224212

Canci, H., and Toker, C. (2009). Evaluation of Annual Wild Cicer Species for Drought and Heat Resistance under Field Conditions. *Genet. Resour. Crop Evol.* 56 (1), 1–6. doi:10.1007/s10722-008-9335-9

Chakraborty, J., Sen, S., Ghosh, P., Sengupta, A., Basu, D., and Das, S. (2016). Homologous Promoter Derived Constitutive and Chloroplast Targeted Expression of Synthetic *cry1Ac* in Transgenic Chickpea Confers Resistance

against *Helicoverpa Armigera*. *Plant Cell Tiss. Org. Cult.* 125, 521–535. doi:10.1007/s11240-016-0968-7

Chandel, K. P. S. (1984). A Note on the Occurrence of Wild Cicer *Microphyllum* Benth. And its Nutrient Status. *Intl. Chickpea Newsl.* 10, 4–5.

Chibbar, R. N., Ambigaipalan, P., and Hoover, R. (2010). Molecular Diversity in Pulse Seed Starch and Complex Carbohydrates and its Role in Human Nutrition and Health. *Cereal Chem.* 87, 342–352. doi:10.1094/cchem-87-4-0342

Cho, S., Kumar, J., Shultz, J. L., Anupama, K., Tefera, F., and Muehlbauer, F. J. (2002). Mapping Genes for Double Podding and Other Morphological Traits in Chickpea. *Euphytica* 128 (2), 285–292. doi:10.1023/a:1020872009306

Chowrira, G. M., Akella, V., Fuerst, P. E., and Lurquin, P. F. (1996). Transgenic Grain Legumes Obtained by in Planta Electroporation-Mediated Gene Transfer. *Mol. Biotechnol.* 5 (2), 85–96. doi:10.1007/bf02789058

Ciftci, H., Ozkaya, A., Cevrimli, B. S., and Bakoglu, A. (2010). Levels of Fat-Soluble Vitamins in Some Foods. *Asian J. Chem.* 22, 1251–1256.

Clarke, H. J., and Siddique, K. H. M. (2004). Response of Chickpea Genotypes to Low Temperature Stress during Reproductive Development. *Field Crops Res.* 90 (2-3), 323–334. doi:10.1016/j.fcr.2004.04.001

Cobos, M. J., Fernandez, M., Rubio, J., Kharrat, M., Moreno, M. T., Gil, J., et al. (2005). A Linkage Map of Chickpea (*Cicer Arietinum* L.) Based on Populations from Kabuli X 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

Collard, B. C. Y., Ades, P. K., Pang, E. C. K., Brouwer, J. B., and Taylor, P. W. J. (2001). Prospecting for Sources of Resistance to Ascochyta Blight in Wild Cicer Species. *Australas. Plant Pathol.* 30 (3), 271–276. doi:10.1071/ap01036

Collard, B. C. Y., Pang, E. C. K., Ades, P. K., and Taylor, P. W. J. (2003). Preliminary Investigation of QTLs Associated with Seedling Resistance to Ascochyta Blight from *Cicer Echinospermum*, a Wild Relative of Chickpea. *Theor. Appl. Genet.* 107 (4), 719–729. doi:10.1007/s00122-003-1297-x

Coram, T. E., and Pang, E. C. K. (2005). Isolation and Analysis of Candidate Ascochyta Blight Defence Genes in Chickpea. Part I. Generation and Analysis of an Expressed Sequence Tag (EST) Library. *Physiol. Mol. Plant Pathol.* 66, 192–200. doi:10.1016/j.pmpp.2005.08.003

Dalgetty, D. D., and Baik, B. K. (2003). Isolation and Characterization of Cotyledon Fibres from Peas, Lentils, and Chickpea. *Cereal Chem.* 80, 310–315. doi:10.1094/cchem.2003.80.3.310

Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Kumar, V., et al. (2015). Deploying QTL-Seq for Rapid Delineation of a Potential Candidate Gene Underlying Major Trait-Associated QTL in Chickpea. *DNA Res.* 22 (3), 193–203. doi:10.1093/dnares/dsv004

Das, S., Singh, M., Srivastava, R., Bajaj, D., Saxena, M. S., Rana, J. C., et al. (2016). mQTL-Seq Delineates Functionally Relevant Candidate Gene Harbouring a Major QTL Regulating Pod Number in Chickpea. *DNA Res.* 23 (1), 53–65. doi:10.1093/dnares/dsv036

Das, A., Datta, S., Thakur, S., Shukla, A., Ansari, J., Sujayanand, G. K., et al. (2017). Expression of a Chimeric Gene Encoding Insecticidal Crystal Protein *Cry1Aabc* of *Bacillus Thuringiensis* in Chickpea (*Cicer Arietinum* L.) Confers Resistance to Gram Pod Borer (*Helicoverpa Armigera* Hubner). *Front. Plant Sci.* 8, 1423. doi:10.3389/fpls.2017.01423

Dash, S., Campbell, J. D., Cannon, E. K., Cleary, A. M., Huang, W., Kalberer, S. R., et al. (2016). Legume Information System (LegumeInfo. Org): a Key Component of a Set of Federated Data Resources for the Legume Family. *Nucl. Acids Res.* 44 (D1), D1181–D1188. doi:10.1093/nar/gkv1159

David, S., Mukandala, L., and Mafuru, J. (2002). Seed Availability, an Ignored Factor in Crop Varietal Adoption Studies: a Case Study of Beans in Tanzania. *J. Sustain. Agric.* 21, 5–20. doi:10.1300/j064v21n02_03

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., et al. (2015). Genetic Properties of the MAGIC Maize Population: A New Platform for High-Definition QTL Mapping in *Zea mays*. *Genome Biol.* 16, 1–23. doi:10.1186/s13059-015-0716-z

Deschamps, S., and Campbell, M. A. (2012). Genetic Variant Discovery and its Use in Genome Characterization of Agronomically Important Crop Species. *Mol. markers plants*, 137–167. doi:10.1002/9781118473023.ch9

Devasirvatham, V., Gaur, P. M., Mallikarjuna, N., Tokachichu, R. N., Trethowan, R. M., and Tan, D. K. (2012). Effect of High Temperature on the Reproductive Development of Chickpea Genotypes under Controlled Environments. *Funct. Plant Biol.* 39 (12), 1009–1018. doi:10.1071/fp12033

Diapari, M., Sindhu, A., Bett, K., Deokar, A., Warkentin, T. D., and Tar'an, B. (2014). Genetic Diversity and Association Mapping of Iron and Zinc Concentrations in Chickpea (*Cicer Arietinum* L.). *Genome* 57 (8), 459–468. doi:10.1139/gen-2014-0108

Dita, M. A., Rispail, N., Prats, E., Rubiales, D., and Singh, K. B. (2006). Biotechnology Approaches to Overcome Biotic and Abiotic Stress Constraints in Legumes. *Euphytica* 147 (1), 1–24. doi:10.1007/s10681-006-6156-9

Doddamani, D., Mohan, A. V. S. K., Katta, M. A. V. S. K., Khan, A. W., Agarwal, G., Shah, T. M., et al. (2014). CicArMiSatDB: the Chickpea Microsatellite Database. *BMC Bioinform.* 15, 212. doi:10.1186/1471-2105-15-212

Doddamani, D., Khan, A. W., Katta, M. A. V. S. K., Agarwal, G., Thudi, M., Ruperao, P., et al. (2015). CicArVarDB: SNP and InDel Database for Advancing Genetics Research and Breeding Applications in Chickpea. *Database* 19, 2015. doi:10.1093/database/bav078

Eapen, S., Köhler, F., Gerdemann, M., and Schieder, O. (1987). Cultivar Dependence of Transformation Rates in Moth Bean after Co-cultivation of Protoplasts with *Agrobacterium Tumefaciens*. *Theor. Appl. Genet.* 75 (1), 207–210. doi:10.1007/bf00249165

Eapen, S. (2008). Advances in Development of Transgenic Pulse Crops. *Biotechnol. Adv.* 26 (2), 162–168. doi:10.1016/j.biotechadv.2007.11.001

Eker, T., Erler, F., Adak, A., Imrek, B., Guven, H., Tosun, H. S., et al. (2018). Screening of Chickpea Accessions for Resistance against the Pulse Beetle, Callosobruchus Chinensis L. (Coleoptera: Bruchidae). *J. Stored Prod. Res.* 76, 51–57. doi:10.1016/j.jspr.2017.12.007

FAOSTAT (2020). FAOSTAT data base, 2020. Available at: http://faostat.fao.org/site (Accessed March 26, 2022)

Fernie, A. R., and Schauer, N. (2009). Metabolomics-assisted Breeding: a Viable Option for Crop Improvement? *Trends Genet.* 25 (1), 39–48. doi:10.1016/j.tig.2008.10.010

Ganguly, M., Molla, K. A., Karmakar, S., Datta, K., and Datta, S. K. (2014). Development of Pod Borer-Resistant Transgenic Chickpea Using a Pod-specific and a Constitutive Promoter-Driven Fused cry1Ab/Ac Gene. *Theor. Appl. Genet.* 127 (12), 2555–2565. doi:10.1007/s00122-014-2397-5

Garg, V., Khan, A. W., Kudapa, H., Kale, S. M., Chitikineni, A., Qiwei, S., et al. (2019). Integrated Transcriptome, Small RNA and Degradome Sequencing Approaches Provide Insights into Ascochyta Blight Resistance in Chickpea. *Plant Biotechnol. J.* 17, 914–931. doi:10.1111/pbi.13026

Gaur, P. M., and Gour, V. K. (2002). A Gene Producing One to Nine Flowers Per Flowering Node in Chickpea. *Euphytica* 128, 231–235. doi:10.1023/a:1020845815319

Gaur, P. M., Gour, V. K., and Srinivasan, S. (2008). An Induced Brachytic Mutant of Chickpea and its Possible Use in Ideotype Breeding. *Euphytica* 159, 35–41. doi:10.1007/s10681-007-9454-y

Gayali, S., Acharya, S., Lande, N. V., Pandey, A., Chakraborty, S., and Chakraborty, N. (2016). CicerTransDB 1.0: a Resource for Expression and Functional Study of Chickpea Transcription Factors. *BMC Plant Biol.* 16, 169. doi:10.1186/s12870-016-0860-y

Geervani, P. (1991). "Utilization of Chickpea in India and Scope for Novel and Alternative Uses," in Proceedings of a Consultants Meeting, Patancheru, AP, 27–30 March 1989 (ICRISAT.), 47–54.

Ghanti, S. K. K., Sujata, K. G., Kumar, B. V., Karba, N. N., Janardhan Reddy, K., Rao, M. S., et al. (2011). Heterologous Expression of *P5CS* Gene in Chickpea Enhances Salt Tolerance without Affecting Yield. *Biol. Plant.* 55, 634. doi:10.1007/s10535-011-0161-0

Golla, S. K., Rajasekhar, P., Akbar, S. M. D., and Sharma, H. C. (2018). Proteolytic Activity in the Midgut of Helicoverpa Armigera (Noctuidae: Lepidoptera) Larvae Fed on Wild Relatives of Chickpea, *Cicer Arietinum*. *J. Econ. Entomol.* 111 (5), 2409–2415. doi:10.1093/jee/toy160

Gowda, S. J. M., Radhika, P., Kadoo, N. Y., Mhase, L. B., and Gupta, V. S. (2009). Molecular Mapping of Wilt Resistance Genes in Chickpea. *Mol. Breed.* 24, 177–183. doi:10.1007/s11032-009-9282-y

Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2010). SoyBase, the USDA-ARS Soybean Genetics and Genomics Database. *Nucl. Acids Res.* 38 (Suppl. l_1), D843–D846. doi:10.1093/nar/gkp798

Greco, N., and Vito, M. D. (1993). *Breeding for Stres Tolerance in Cool-Season Food Legumes*, Editor K. B. Singh K. B. and M. C. Saxena (Chichester, United Kingdom: John Wiley & Sons), 157–166.

Gujaria, N., Kumar, A., Dauthal, P., Dubey, A., Hiremath, P., Bhanu Prakash, A., et al. (2011). Development and Use of Genic Molecular Markers (GMMs) for Construction of a Transcript Map of Chickpea (*Cicer Arietinum* L.). *Theor. Appl. Genet.* 122, 1577–1589. doi:10.1007/s00122-011-1556-1

Gupta, R., and Laxman, S. (2020). tRNA Wobble-Uridine Modifications as Amino Acid Sensors and Regulators of Cellular Metabolic State. *Curr. Genet.* 66 (3), 475–480. doi:10.1007/s00294-019-01045-y

Gupta, S., and Parihar, A. K. (2015). "Broadening the Genetic Base of Pulse Crops," in *Pulses Challenges and Opportunities Under Changing Climatic Scenario,* (Kanpur, UP, India: ISPRD), 86–101.

Gupta, S., Nawaz, K., Parween, S., Roy, R., Sahu, K., Kumar, P. A., et al. (2017). Draft Genome Sequence of *Cicer Reticulatum* L., the Wild Progenitor of Chickpea Provides a Resource for Agronomic Trait Improvement. *DNA Res.* 24, 1–10. doi:10.1093/dnares/dsw042

Hajyzadeh, M., Turktas, M., Khawar, K. M., and Unver, T. (2015). miR408 Overexpression Causes Increased Drought Tolerance in Chickpea. *Gene* 555, 186–193. doi:10.1016/j.gene.2014.11.002

Harshavardhana, Y. S., Hegde, V., Tripathi, S., Raje, R. S., Jain, P. K., Gaikwad, K., et al. (2019). Genetics of Semi-determinacy and Identification of Molecular Marker Linked to Dt1 Locus in Chickpea (*Cicer Arietinum* L.). *Indian J. Genet. Pl. Br.* 79 (Suppl. 01), 270–275. doi:10.31742/ijgpb.79s.1.18

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 (1-2), 147–154. doi:10.1016/0378-4290(92)90063-f

He, J., Benedito, V. A., Wang, M., Murray, J. D., Zhao, P. X., Tang, Y., et al. (2009). The *Medicago Truncatula* Gene Expression Atlas Web Server. *BMC Bioinfo.* 10 (1), 1–9. doi:10.1186/1471-2105-10-441

Hegde, V. S., Tripathi, S., Bharadwaj, C., Agrawal, P. K., and Choudhary, A. K. (2018). Genetics and Genomics Approaches to Enhance Adaptation and Yield of Chickpea (*Cicer Arietinum* L.) in Semi-arid Environments. *Sabrao J. Breed. Genet.* 50 (2), 217–241.

Hegde, V. S. (2011). Morphology and Genetics of a New Found Determinate Genotype in Chickpea. *Euphytica* 182, 35–42. doi:10.1007/s10681-011-0447-5

Hiremath, P. J., Farmer, A., Cannon, S. B., Woodward, J., Kudapa, H., Tuteja, R., et al. (2011). Large-Scale Transcriptome Analysis in Chickpea (*Cicer Arietinum* L.), an Orphan Legume Crop of the Semi-arid Tropics of Asia and Africa. *Plant Biotechnol. J.* 9 (8), 922–931. doi:10.1111/j.1467-7652.2011.00625.x

Hu, J., Rampitsch, C., and Bykova, N. V. (2015). Advances in Plant Proteomics toward Improvement of Crop Productivity and Stress Resistancex. *Front. Plant Sci.* 6, 209. doi:10.3389/fpls.2015.00209

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., et al. (2015). MAGIC Populations in Crops: Current Status and Future Prospects. *Theor. Appl. Genet.* 128 (6), 999–1017. doi:10.1007/s00122-015-2506-0

Hulse, J. H. (1991). "Nature, Composition and Utilization of Pulses," in Uses of Tropical Grain Legumes, Proceedings of a Consultants Meeting, Patancheru, AP, 27–30 March 1989 (ICRISAT), 11–27.

Husnain, T., Fatima, T., Islam, R., and Riazuddin, S. (2000). Plant Regeneration and Expression of Beta-Glucorinidase Gene in Hypocotyls Tissue of Chickpea (*Cicer Arietinum* L.). *Pak. J. Biol. Sci.* 3, 842–845. doi:10.3923/pjbs.2000.842.845

ICRISAT (2009). *ICISAT Archival Report*. Hyderabad, India: International Crop Research Institute for the Semi-Arid Tropics. sci985.0011183X002500030047x.

Imtiaz, M., Materne, M., Hobson, K., van Ginkel, M., and Malhotra, R. S. (2008). Molecular Genetic Diversity and Linked Resistance to Ascochyta Blight in Australian Chickpea Breeding Materials and Their Wild Relatives. *Aust. J. Agric. Res.* 59, 554–560. doi:10.1071/ar07386

Indurker, S., Misra, H. S., and Eapen, S. (2007). Genetic Transformation of Chickpea (*Cicer Arietinum* L.) with Insecticidal Crystal Protein Gene Using Particle Gun Bombardment. *Plant Cell Rep.* 26, 755–763. doi:10.1007/s00299-006-0283-6

Infantino, A., Porta-Puglia, A., and Singh, K. B. (1996). Screening Wild Cicer Species for Resistance to fusarium Wilt. *Plant Dis.* 80, 42–44. doi:10.1094/pd-80-0042

Iruela, M., Rubio, J., Barro, F., Cubero, J. I., Millan, T., and Gil, J. (2006). Detection of Two Quantitative Trait Loci for Resistance to Ascochyta Blight in an Intra-specific Cross of Chickpea (*Cicer Arietinum* L.): Development of SCAR Markers Associated with Resistance. *Theor. Appl. Genet.* 112, 278–287. doi:10.1007/s00122-005-0126-9

Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2015). Genotyping-by-sequencing Based Intra-specific Genetic Map Refines a "QTL-Hotspot" Region for Drought Tolerance in Chickpea. *Mol. Genet. Genom.* 290 (2), 559–571. doi:10.1007/s00438-014-0932-3

Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A Draft Genome Sequence of the Pulse Crop Chickpea (*Cicer Arietinum* L.). *Plant J* 74 (5), 715–729. doi:10.1111/tpj.12173

Jain, M., Chevala, V. V., and Garg, R. (2014). Genome-wide Discovery and Differential Regulation of Conserved and Novel microRNAs in Chickpea via Deep Sequencing. *J. Exp. Bot.* 65, 5945–5958. doi:10.1093/jxb/eru333

Jaiswal, H. K., and Singh, B. D. (1989). Analysis of Gene Effects for Yield and Certain Yield Traits in Crosses between *Cicer Arietinum* L. And *C. reticulatum* Ladz. *Indian J. Genet. Plant Breed.* 49 (1), 9–17.

Jha, U. C., Nayyar, H., Palakurthi, R., Jha, R., Valluri, V., Bajaj, P., et al. (2021). Major QTLs and Potential Candidate Genes for Heat Stress Tolerance Identified in Chickpea (*Cicer Arietinum* L.). *Front. Plant Sci.* 12, 655103. doi:10.3389/fpls.2021.655103

Kaiser, W. J., Alcala-Jimenez, A. R., Hervas-Vargas, A., Trapero-Casas, J. L., and Jimenez-Diaz, R. M. (1994). Screening of Wild Cicer Species for Resistance to Races 0 to 5 of *Fusarium Oxysporum* F. Sp. Ciceris. *Plant Dis.* 78, 962–967. doi:10.1094/pd-78-0962

Kale, S. M., Jaganathan, J., Ruperao, P., Chen, C., Punna, R., Kudapa, H., et al. (2015). Prioritization of Candidate Genes in QTL-Hotspot Region for Drought Tolerance in Chickpea (Cicer Arietinum L.). *Sci. Rep.* 5, 15296. doi:10.1038/srep15296

Kamble, S., Misra, H. S., Mahajan, S. K., and Eapen, S. (2003). A Protocol for Efficient Biolistic Transformation of Mothbean (*Vigna Aconitifolia* L. Jacq. Marechal.). *Plant Mol. Biol. Rep.* 21 (4), 457–458. doi:10.1007/bf02772595

Kar, S., Basu, D., Das, S., Ramkrishnan, N. A., Mukherjee, P., Nayak, P., et al. (1997). Expression of *cry1Ac* Gene of *Bacillus Thuringiensis* in Transgenic Chickpea Plants Inhibits Development of Pod Borer (*Heliothis Armigera*) Larvae. *Transgenic Res.* 6, 177–185. doi:10.1023/a:1018433922766

Kashiwagi, J., Krishnamurthy, L., Upadhyaya, H. D., Krishna, H., Chandra, S., Vadez, V., et al. (2005). Genetic Variability of Drought-Avoidance Root Traits in the Mini-Core Germplasm Collection of Chickpeas (*Cicer Arietinum* L.). *Euphytica* 146 (3), 213–222. doi:10.1007/s10681-005-9007-1

Kashiwagi, J., Krishnamurthy, L., Singh, S., and Upadhyaya, H. D. (2006b). Variation of SPAD Chlorophyll Meter Readings (SCMR) in the Mini-Core Germplasm Collection of Chickpeas. *J. SAT Agric. Res.* 2 (1), 1–3.

Kashiwagi, J., Krishnamurthy, L., Upadhyaya, H. D., and Gaur, P. M. (2008). Rapid Screening Technique for Canopy Temperature Status and its Relevance to Drought Tolerance Improvement in Chickpea. *J. SAT Agric. Res.* 6, 105–114.

Kashiwagi, J., Upadhyaya, H. D., and Krishnamurthy, L. (2010). Significance and Genetic Diversity of SPAD Chlorophyll Meter Reading in Chickpea Germplasm in the Semi-arid Environments. *J. Food Legumes* 23 (2), 99–105.

Kaur, S., Chhabra, K. S., and Arora, B. S. (1999). Incidence of *Helicoverpa Armigera* (Hubner) on Wild and Cultivated Species of Chickpea. *Int. Chickpea Pigeonpea Newsl.* 6, 18–19.

Kaur, M., Singh, N., and Sodhi, N. S. (2005). Physicochemical, Cooking, Textural and Roasting Characteristics of Chickpea (*Cicer Arietinum* L.) Cultivars. *J. Food Eng.* 69, 511–517. doi:10.1016/j.jfoodeng.2004.09.002

Kaur, L., Tripathi, H. S., Vishwa, D., Reddy, M. V., Singh, G., and Kharkwal, M. C. (2008). "Management of Diseases in Food Legumes," in Proc. fourth Int. Food Legumes Res. Confer., New Delhi, India, 608–637. 1.

Kaur, L., Sirari, A., Kumar, D., Sandhu, J. S., Singh, S., Kapoor, K., et al. (2013). Combining Ascochyta Blight and Botrytis Grey Mould Resistance in Chickpea through Interspecific Hybridization. *Phytopathol. Mediterr.* 52 (1), 157–165. doi:10.14601/Phytopathol_Mediterr-11651

Khatodia, S., Kharb, P., Batra, P., and Chowdhury, V. K. (2014). Development and Characterization of Transgenic Chickpea (*Cicer Arietinum* L.) Plants with *cry1Ac* Gene Using Tissue Culture Independent Protocol. *Int. J. Adv. Res.* 2, 323–331.

Knights, E. J., Southwell, R. J., Schwinghamer, M. W., and Harden, S. (2008). Resistance to Phytophthora Medicaginis Hansen and Maxwell in Wild Cicer Species and its Use in Breeding Root Rot Resistant Chickpea (*Cicer Arietinum* L.). *Aust. J. Agric. Res.* 59 (4), 383–387. doi:10.1071/ar07175

Köhler, F., Golz, C., Eapen, S., Kohn, H., and Schieder, O. (1987a). Stable Transformation of Moth Bean *Vigna Aconitifolia* via Direct Gene Transfer. *Plant Cell Rep.* 6 (4), 313–317. doi:10.1007/BF00272007

Köhler, F., Golz, C., Eapen, S., and Schieder, O. (1987b). Influence of Plant Cultivar and Plasmid-DNA on Transformation Rates in Tobacco and Moth Bean. *Plant Sci.* 53 (1), 87–91.

Krishnakumar, V., Kim, M., Rosen, B. D., Karamycheva, S., Bidwell, S. L., Tang, H., et al. (2015). MTGD: The *Medicago Truncatula* Genome Database. *Plant Cell Physiol.* 56, 1e1. doi:10.1093/pcp/pcu179

Krishnamurthy, K. V., Suhasini, K., Sagare, A. P., Meixner, M., De Kathen, A., Pickardt, T., et al. (2000). Agrobacterium Mediated Transformation of Chickpea (*Cicer Arietinum* L.) Embryo Axes. *Plant Cell Rep.* 19 (3), 235–240. doi:10.1007/s002990050005

Krishnamurthy, L., Kashiwagi, J., Gaur, P. M., Upadhyaya, H. D., and Vadez, V. (2010). Sources of Tolerance to Terminal Drought in the Chickpea (*Cicer Arietinum* L.) Minicore Germplasm. *Field Crops Res.* 119 (2-3), 322–330. doi:10.1016/j.fcr.2010.08.002

Krishnamurthy, L., Gaur, P. M., Basu, P. S., Chaturvedi, S. K., Tripathi, S., Vadez, V., et al. (2011a). Large Genetic Variation for Heat Tolerance in the Reference Collection of Chickpeas (*Cicer Arietinum* L.) Germplasm. *Plant Genet. Resour.* 9 (1), 59–69. doi:10.1017/s1479262110000407

Krishnamurthy, L., Turner, N. C., Gaur, P. M., Upadhyaya, H. D., Varshney, R. K., Siddique, K. H. M., et al. (2011b). Consistent Variation across Soil Types in Salinity Resistance of a Diverse Range of Chickpea (*Cicer Arietinum* L.) Genotypes. *J. Agron. Crop Sci.* 197 (3), 214–227. doi:10.1111/j.1439-037x.2010.00456.x

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015a). Employing Genome-wide SNP Discovery and Genotyping Strategy to Extrapolate the Natural Allelic Diversity and Domestication Patterns in Chickpea. *Front. Plant Sci.* 6, 162. doi:10.3389/fpls.2015.00162

Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M. S., et al. (2015b). Ultra-high Density Intra-specific Genetic Linkage Maps Accelerate Identification of Functionally Relevant Molecular Tags Governing Important Agronomic Traits in Chickpea. *Sci. Rep.* 5 (1), 1–13. doi:10.1038/srep09468

Kujur, A., Upadhyaya, H. D., Bajaj, D., Gowda, C. L. L., Sharma, S., Tyagi, A. K., et al. (2016). Identification of Candidate Genes and Natural Allelic Variants for QTLs Governing Plant Height in Chickpea. *Sci. Rep.* 6, 27968. doi:10.1038/srep27968

Kulwal, P. L., Thudi, M., and Varshney, R. K. (2011). "Genomics Interventions in Crop Breeding for Sustainable Agriculture," in *Encyclopaedia of Sustainability Science and Technology* (Springer). doi:10.1007/978-1-4419-0851-3

Kumar, S., and Gugita, S. (2004). "How Wide the Genetic Base of Pulse Crops Is," in Proc. Nat. Simp. Crop Diversification Natur. Resour. Manag., Kanpur, India. Editors M. Ali, B. B. Singh, S. Kumar, and V. Dhar, 211–217.*Pulses in New Perspective*

Kumar, K. B., Mahalakshmi, S. L., Kumar, S. M., and Devi, B. P. (2014). Optimized Agrobacterium-Mediated Genetic Transformation in Chickpea (*Cicer Arietinum* L.) Cultivar Swetha (ICCV-2). *Trends Biosci.* 7, 2237–2244.

Labdi, M., Robertson, L. D., Singh, K. B., and Charrier, A. (1996). Genetic Diversity and Phylogenetic Relationships Among the Annual Cicer Species as Revealed by Isozyme Polymorphism. *Euphytica* 88 (3), 181–188. doi:10.1007/bf00023889

Ladizinsky, G., and Adler, A. (1975). Origin of Chickpea as Indicated by Seed Protein Electrophoresis. *Israel J. Bot.* 24, 183–189. doi:10.1007/bf00132919

Ladizinsky, G., and Adler, A. (1976a). The Origin of Chickpea *Cicer Arietinum* L. *Euphytica* 25 (1), 211–217. doi:10.1007/bf00041547

Ladizinsky, G., and Adler, A. (1976b). Genetic Relationships Among the Annual Species of *Cicer L. Theor. Appl. Genet.* 48 (4), 197–203. doi:10.1007/bf00527371

Ladizinsky, G. (1985). Founder Effect in Crop-Plant Evolution. *Econ. Bot.* 39 (2), 191–199. doi:10.1007/bf02907844

Ladizinsky, G. (1998). Plant Evaluation under Domestication. *Dordrecht Springer*, 1–254. doi:10.1007/978-94-011-4429-2

Langridge, P., and Fleury, D. (2011). Making the Most of 'omics' for Crop Breeding. *TRENDS Biotechnol.* 29 (1), 33–40. doi:10.1016/j.tibtech.2010.09.006

Lawo, N. C., Mahon, R. J., Milner, R. J., Sarmah, B. K., Higgins, T. J. V., and Romeis, J. (2008). Effectiveness of *Bacillus Thuringiensis* Transgenic Chickpeas and the Entomopathogenic Fungus *Metarhizium Anisopliae* in Controlling *Helicoverpa Armigera* (Lepidoptera: Noctuidae). *Appl. Environ. Microbiol.* 74, 4381–4389. doi:10.1128/aem.00484-08

Li, J., Dai, X., Liu, T., and Zhao, P. X. (2012). LegumeIP: an Integrative Database for Comparative Genomics and Transcriptomics of Model Legumes. *Nucl. Acids Res.* 40 (D1), D1221–D1229. doi:10.1093/nar/gkr939

Li, H., Rodda, M., Gnanasambandam, A., Aftab, M., Redden, R., Hobson, K., et al. (2015). Breeding for Biotic Stress Resistance in Chickpea: Progress and Prospects. *Euphytica* 204 (2), 257–288. doi:10.1007/s10681-015-1462-8

Li, Y., Ruperao, P., Batley, J., Edwards, D., Davidson, J., Hobson, K., et al. (2017). Genome Analysis Identified Novel Candidate Genes for Ascochyta Blight Resistance in Chickpea Using Whole Genome Re-sequencing Data. *Front. Plant Sci.* 8, 359. doi:10.3389/fpls.2017.00359

Li, Y., Ruperao, P., Batley, J., Edwards, D., Martin, W., Hobson, K., et al. (2021). Genomic Prediction of Preliminary Yield Trials in Chickpea: Effect of Functional Annotation of SNPs and Environment. *Plant Genome* 15, e20166. doi:10.1002/tpg2.20166

Liu, Z., Crampton, M., Todd, A., and Kalavacharla, V. (2012). Identification of Expressed Resistance Gene-like Sequences by Data Mining in 454-derived Transcriptomic Sequences of Common Bean (Phaseolus vulgaris L.). *BMC plant Biol.* 12 (1), 1–9. doi:10.1186/1471-2229-12-42

Mackay, I., and Powell, W. (2007). Methods for Linkage Disequilibrium Mapping in Crops. *Trends plant Sci.* 12 (2), 57–63. doi:10.1016/j.tplants.2006.12.001

Madrid, E., Chen, W., Rajesh, P. N., Castro, P., Millan, T., and Gil, J. (2013). Allele-specific Amplification for the Detection of Ascochyta Blight Resistance in Chickpea. *Euphytica* 189, 183–190. doi:10.1007/s10681-012-0753-6

Madrid, E., Seoane, P., Claros, M. G., Barro, F., Rubio, J., Gil, J., et al. (2014). Genetic and Physical Mapping of the QTL AR3 Controlling Blight Resistance in Chickpea (*Cicer Arietinum* L). *Euphytica* 198 (1), 69–78. doi:10.1007/s10681-014-1084-6

Mahto, R. K., Ambika, Singh, C., Chandana, B. S., Singh, R. K., Verma, S., et al. (2022). Chickpea Biofortification for Cytokinin Dehydrogenase *via* Genome Editing to Enhance Abiotic-Biotic Stress Tolerance and Food Security. *Front. Genet.* 13, 900324. doi:10.3389/fgene.2022.900324

Malhotra, R. S., Pundir, R. P. S., and Slinkard, A. E. (1987). "Genetic Resources of Chickpea," in *The Chickpea*. Editors M. C. Saxena and K. B. Singh (Aberystwyth: CAB Int. Cambrian News Ltd), 67–81.

Mallikarjuna, B. P., Samineni, S., Thudi, M., Sajja, S. B., Khan, A. W., Patil, A., et al. (2017). Molecular Mapping of Flowering Time Major Genes and QTLs in Chickpea (*Cicer Arietinum* L.). *Front. Plant Sci.* 8, 1140. doi:10.3389/fpls.2017.01140

Mannur, D. M., Babbar, A., Thudi, M., Sabbavarapu, M. M., Roorkiwal, M., Yeri, S. B., 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 (1), 1–13. doi:10.1007/s11032-018-0908-9

Maqbool, M. A., Aslam, M., and Ali, H. (2017). Breeding for Improved Drought Tolerance in Chickpea (*Cicer Arietinum* L.). *Plant Breed.* 136 (3), 300–318. doi:10.1111/pbr.12477

Marshall, D. R. (1989). "Limitations to the Use of Germplasm Collections," in *The Use of Plant Genetic Resources*. Editor A. H. D. Brown, O. H. Frankel, D. R. Marshall, and J. T. Williams (New York: Cambridge University Press), 105–120.

Meena, A. K., Gurjar, D., and Kumhar, B. L. (2017). Pre-breeding Is a Bridge between Wild Species and Improved Genotypes—A Review. *Chem. Sci. Rev. Lett.* 6 (22), 1141–1151. doi:10.20546/ijcmas.2017.606.007

Mehrotra, M., Singh, A. K., Sanyal, I., Altosaar, I., and Amla, D. V. (2011). Pyramiding of Modified *cry1Ab* and *cry1Ac* Genes of *Bacillus Thuringiensis* in Transgenic Chickpea (*Cicer Arietinum* L.) for Improved Resistance to Pod Borer Insect *Helicoverpa Armigera*. *Euphytica* 182 (1), 87–102. doi:10.1007/s10681-011-0501-3

Milla´n, T., Rubio, J., Iruela, M., Daly, K., Cubero, J. I., and Gil, J. (2003). Markers Associated with Ascochyta Blight Resistance in Chickpea and Their Potential in Marker-Assisted Selection. *Field Crops Res.* 84, 373–384.

Molina, C., Rotter, B., Horres, R., Udupa, S., Besser, B., Bellarmino, L., et al. (2008). SuperSAGE: the Drought Stress-Responsive Transcriptome of Chickpea Roots. *BMC Genom.* 9 (1), 553. doi:10.1186/1471-2164-9-553

Morozova, O., and Marra, M. A. (2008). Applications of Next-Generation Sequencing Technologies in Functional Genomics. *Genomics* 92 (5), 255–264. doi:10.1016/j.ygeno.2008.07.001

Muehlbauer, F. J., and Singh, K. B. (1987). "Genetics of Chickpea," in *The Chickpea*. Editors M. C. Saxena and K. B. Singh (Wallingford: CAB Int.), 99–126.

Mulwa, R. M. S., Kimurto, P. K., and Towett, B. K. (2010). "Evaluation and Selection of Drought and Pod Borer (*Helicoverpa Armigera*) Tolerant Chickpea Genotypes for Introduction in Semi-arid Areas of Kenya.," in Second RUFORUM Biennial Meeting, Entebbe, Uganda, 20 - 24 September 2010.

Nass, L. L., and Paterniani, E. (2000). Pre-breeding: a Link between Genetic Resources and Maize Breeding. *Sci. Agric.* 57, 581–587. doi:10.1590/s0103-90162000000300035

Nayak, S. N., Zhu, H., Varghese, N., Datta, S., Choi, H., Horres, R., et al. (2010). Integration of Novel SSR and Gene-Based SNP Marker Loci in the Chickpea Genetic Map and Establishment of New Anchor Points with *Medicago Truncatula* Genome. *Theor. Appl. Genet.* 120, 1415–1441. doi:10.1007/s00122-010-1265-1

Nene, Y. L., and Haware, M. P. (1980). Screening Chickpea for Resistance to Wilt. *Plant Dis.* 64 (4), 379–380. doi:10.1094/pd-64-379

Nguyen, T. T., Taylor, P. W. J., Redden, R. J., and Ford, R. (2004). Genetic Diversity Estimates in Cicer Using AFLP Analysis. *Plant Breed.* 123 (2), 173–179. doi:10.1046/j.1439-0523.2003.00942.x

Pande, S. K. H. M. S., Siddique, K. H. M., Kishore, G. K., Bayaa, B., Gaur, P. M., Gowda, et al. (2005). Ascochyta Blight of Chickpea (*Cicer Arietinum* L.): a Review of Biology, Pathogenicity, and Disease Management. *Aust. J. Agric. Res.* 56 (4), 317–332. doi:10.1071/ar04143

Pande, S., Galloway, J., Gaur, P. M., Siddique, K. H. M., Tripathi, H. S., Taylor, P., et al. (2006). Botrytis Grey Mould of Chickpea: a Review of Biology, Epidemiology, and Disease Management. *Aust. J. Agric. Res.* 57 (11), 1137–1150. doi:10.1071/ar06120

Pande, S., Sharma, M., Gaur, P. M., and Gowda, C. L. L. (2010). *Host Plant Resistance to Ascochyta Blight of Chickpea*. Hyderabad, India: ICRISAT. Information Bulletin No.82.

Pandey, A., Chakraborty, S., Datta, A., and Chakraborty, N. (2008). Proteomics Approach to Identify Dehydration Responsive Nuclear Proteins from Chickpea (*Cicer Arietinum* L.). *Mol. Cell. Proteomics* 7 (1), 88–107. doi:10.1074/mcp.m700314-mcp200

Pandey, M. K., Roorkiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., et al. (2016). Emerging Genomic Tools for Legume Breeding: Current Status and Future Prospects. *Front. Plant Sci.* 7, 455. doi:10.3389/fpls.2016.00455

Parameshwarappa, S. G., and Salimath, P. M. (2008). Field Screening of Chickpea Genotypes for Drought Resistance. *Karnataka J. Agric. Sci.* 21 (1), 113–114.

Parida, S. K., Srivastava, R., and Bajaj, D. (2017). A Genome-wide mQTL-Seq Scan Identifies Potential Molecular Signatures Regulating Plant Height in Chickpea. *Plant Mol. Biol. Report.* 35 (2), 273–286. doi:10.1007/s11105-016-1021-z

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., et al. (2015). Potential of a Tomato MAGIC Population to Decipher the Genetic Control of Quantitative Traits and Detect Causal Variants in the Resequencing Era. *Plant Biotechnol. J.* 13 (4), 565–577. doi:10.1111/pbi.12282

Patil, A. S., and Kamble, G. C. (2014). Comparative Protein Profiling Study in Wild Chickpea and its Induced Mutants. *Experiment* 18, 1297–1303.

Paul, P. J., Samineni, S., Thudi, M., Sajja, S. B., Rathore, A., Das, R. R., et al. (2018). Molecular Mapping of QTLs for Heat Tolerance in Chickpea. *Int. J. Mol. Sci.* 19, 2166. doi:10.3390/ijms19082166

Pavan, S., Lotti, C., Marcotrigiano, A. R., Mazzeo, R., Bardaro, N., Bracuto, V., et al. (2017). A Distinct Genetic Cluster in Cultivated Chickpea as Revealed by Genome-wide Marker Discovery and Genotyping. *Plant Genome* 10 (2), 1–9. doi:10.3835/plantgenome2016.11.0115

Pourkheirandish, M., Golicz, A. A., Bhalla, P. L., and Singh, M. B. (2020). Global Role of Crop Genomics in the Face of Climate Change. *Front. Plant Sci.* 11, 922. doi:10.3389/fpls.2020.00922

Pundir, R. P. S., and Mengesha, M. H. (1995). Cross Compatibility between Chickpea and its Wild Relative, *Cicer Echinospermum* Davis. *Euphytica* 83 (3), 241–245. doi:10.1007/bf01678136

Pundir, R. P. S., and Reddy, K. N. (1989). Induction, Genetics and Possible Use of Glabrousness in Chickpea. *Euphytica* 42 (1), 141–144. doi:10.1007/bf00042625

Pundir, R. P. S., Rao, N. K., and van der Maesen, L. J. G. (1985). Distribution of Qualitative Traits in the World Germplasm of Chickpea (*Cicer Arietinum* L.). *Euphytica* 34, 697–703. doi:10.1007/bf00035406

Pundir, R. P. S., Mengesha, M. H., and Reddy, K. N. (1990). Leaf Types and Their Genetics in Chickpea (*Cicer Arietinum* L.). *Euphytica* 45 (3), 197–200. doi:10.1007/bf00032985

Rakshit, S., Ganapathy, K. N., Gomashe, S. S., Rathore, A., Ghorade, R. B., Nagesh Kumar, M. V., et al. (2012). GGE Biplot Analysis to Evaluate Genotype, Environment and Their Interactions in Sorghum Multi-Location Data. *Euphytica* 185 (3), 465–479. doi:10.1007/s10681-012-0648-6

Rao, L. S., Usha Rani, P., Deshmukh, P. S., Kumar, P. A., and Panguluri, S. K. (2007). RAPD and ISSR Fingerprinting in Cultivated Chickpea (*Cicer Arietinum* L.) and its Wild Progenitor *Cicer Reticulatum* Ladizinsky. *Genet. Resour. Crop Evol.* 54 (6), 1235–1244. doi:10.1007/s10722-006-9104-6

Reddy, A. A. (2016). Pulse Production Technology: Status and Way Forward. *Econ. Political Wkly.* 44 (52), 73–80. doi:10.2139/ssrn.1537540

Rehman, A. U., Malhotra, R. S., Bett, K., Tar'an, B., Bueckert, R., and Warkentin, T. D. (2012). Mapping QTL Associated with Traits Affecting Grain Yield in Chickpea (*Cicer Arietinum* L.) under Terminal Drought Stress. *Crop Sci.* 51, 450–463.

Rizkalla, S. W., Bellisle, F., and Slama, G. (2002). Health Benefits of Low Glycaemic Index Foods, Such as Pulses, in Diabetic Patients and Healthy Individuals. *Br. J. Nutr.* 88, 255–262. doi:10.1079/bjn2002715

Robertson, L. D., Singh, K. B., and Ocampo, B. (1995). *A Catalog of Annual Wild Cicer Species*. Beirut, Lebanon: International Center for Agricultural Research in the Dry Areas ICARDA.

Robertson, L. D., Ocampo, B., and Singh, K. B. (1997). Morphological Variation in Wild Annual Cicer Species in Comparison to the Cultigen. *Euphytica* 95 (3), 309–319. doi:10.1023/a:1003004516921

Roorkiwal, M., Von Wettberg, E. J., Upadhyaya, H. D., Warschefsky, E., Rathore, A., and Varshney, R. K. (2014). Exploring Germplasm Diversity to Understand the Domestication Process in Cicer Spp. Using SNP and DArT Markers. *PLoS One* 9 (7), e102016. doi:10.1371/journal.pone.0102016

Roorkiwal, M., Bharadwaj, C., Barmukh, R., Dixit, G. P., Thudi, M., Gaur, P. M., et al. (2020). Integrating Genomics for Chickpea Improvement: Achievements and Opportunities. *Theor.Appl. Genet.* 133, 1703–1720. doi:10.1007/s00122-020-03584-2

Sab, S., Lokesha, R., and Mannur, D. M. (2020). Genome-Wide SNP Discovery and Mapping QTLs for Seed Iron and Zinc Concentrations in Chickpea (Cicer Arietinum L.). *Front. Nutr.* 7, 559120. doi:10.3389/fnut.2020.559120

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 Ascochyta Blight in Chickpea (*Cicer Arietinum* L.). *Euphytica* 193 (1), 121–133. doi:10.1007/s10681-013-0959-2

Saeed, A., and Darvishzadeh, R. (2017). Association Analysis of Biotic and Abiotic Stresses Resistance in Chickpea (Cicer spp.) Using AFLP Markers. *Biotechnol. Biotechnol. Equip.* 31 (4), 698–708. doi:10.1080/13102818.2017.1333455

Samineni, S., Siddique, K. H., Gaur, P. M., and Colmer, T. D. (2011). Salt Sensitivity of the Vegetative and Reproductive Stages in Chickpea (*Cicer Arietinum* L.): Podding Is a Particularly Sensitive Stage. *Environ. Exp. Bot.* 71 (2), 260–268. doi:10.1016/j.envexpbot.2010.12.014

Samineni, S., Mahendrakar, M. D., Hotti, A., Chand, U., Rathore, A., and Gaur, P. M. (2022). Impact of Heat and Drought Stresses on Grain Nutrient Content in Chickpea: Genome-wide Marker-Trait Associations for Protein, Fe and Zn. *Environ. Exp. Bot.* 194, 104688. doi:10.1016/j.envexpbot.2021.104688

Sánchez-Vioque, R., Clemente, A., Vioque, J., Bautista, J., and Millán, F. (1999). Protein Isolates From Chickpea (*Cicer arietinum* L.): Chemical Composition, Functional Properties and Protein Characterization. *Food Chem.* 64 (2), 237–243.

Sandhu, J. S. (2004). *Pers. Communication*. Punjab Agric. Univ., Ludhiana, Punjab.

Sanderson, L. A., Caron, C. T., Tan, R., Shen, Y., Liu, R., and Bett, K. E. (2019). KnowPulse: A Web-Resource Focused on Diversity Data for Pulse Crop Improvement. *Front. Plant Sci.* 10, 965. doi:10.3389/fpls.2019.00965

Santra, D. K., Tekeoglu, M., Ratnaparkhe, M., Kaiser, W. J., and Muehlbauer, F. J. (2000). Identification and Mapping of QTLs Conferring Resistance to Ascochyta Blight in Chickpea. *Crop Sci.* 40, 1606–1612. doi:10.2135/cropsci2000.4061606x

Sanyal, I., Singh, A. K., Kaushik, M., and Amla, D. V. (2005). Agrobacterium-mediated Transformation of Chickpea (*Cicer Arietinum* L.) with *Bacillus Thuringiensis* cry1Ac Gene for Resistance against Pod Borer Insect *Helicoverpa Armigera*. *Plant Sci.* 168 (4), 1135–1146. doi:10.1016/j.plantsci.2004.12.015

Sarika, S., Arora, V., Iquebal, M. A., Rai, A., and Kumar, D. (2013). PIPEMicroDB: Microsatellite Database and Primer Generation Tool for Pigeonpea Genome. *Database* 2013, bas054. doi:10.1093/database/bas054

Savithri, K. S., Ganapathy, P. S., and Sinha, S. K. (1980). Sensitivity to Low Temperature in Pollen Germination and Fruit-Set in *Cicer Arietinum* L. *J. Exp. Bot.* 31 (2), 475–481. doi:10.1093/jxb/31.2.475

Saxena, N. P., and Johansen, C. (1990). "Realized Yield Potential in Chickpea and Physiological Considerations for Further Genetic Improvement," in Proceedings of the International Congress of Plant Physiology (New Delhi: Indian Society for Plant Physiology and Biochemistry, Indian Agricultural Research Institute), 279–288.1

Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014). An Integrated Genomic Approach for Rapid Delineation of Candidate Genes Regulating Agro-Morphological Traits in Chickpea. *DNA Res.* 21, 695–710. doi:10.1093/dnares/dsu031

Saxena, R., K., Kale, S., M., Kumar, V., Parupalli, S., Joshi, S., Singh, V., K., et al. (2017). Genotyping-by-sequencing of Three Mapping Populations for Identification of Candidate Genomic Regions for Resistance to Sterility Mosaic Disease in Pigeonpea. *Sci. Rep.* 7 (1), 1–10. doi:10.1038/s41598-017-01535-4

Saxena, R., K., Patel, K., Sameer Kumar, C., V., Tyagi, K., Saxena, K.,B., and Varshney, R., K. (2018). Molecular Mapping and Inheritance of Restoration of Fertility (Rf) in A4 Hybrid System in Pigeonpea (Cajanus Cajan (L.) Millsp.). *Theor. Appl. Genet.* 131, 1605–1614. doi:10.1007/s00122-018-3101-y

Serraj, R., Krishanamurthy, L., and Upadhyay, H. D. (2004). Screening Chickpea Minicore Germplasm for Tolerance to Soil Salinity. *Int. Chickpea Pigeonpea Newsl. (ICPN)* 11, 29–32.

Severin, A. J., Woody, J. L., Bolon, Y. T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-seq Atlas of Glycine Max: a Guide to the Soybean Transcriptome. *BMC Plant Biol.* 10 (1), 1–16. doi:10.1186/1471-2229-10-160

Shah, T. M., Haq, M. A., Atta, B. M., Alam, S. S., and Ali, H. (2005). Evaluation of Cicer Species for Resistance to Ascochyta Blight. *Pak. J. Bot.* 37 (2), 431.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted Genome Modification of Crop Plants Using a CRISPR-Cas System. *Nat. Biotechnol.* 31, 686–688. doi:10.1038/nbt.2650

Sharma, H. C., Pampapathy, G., Lanka, S. K., and andRidsdill-Smith, T. J. (2005). Exploitation of Wild *Cicer Reticulatum* Germplasm for Resistance to *Helicoverpa Armigera*. *J. Econ. Entomol.* 98 (6), 2246–2253. doi:10.1093/jee/98.6.2246

Sharma, S., Upadhyaya, H. D., Roorkiwal, M., Varshney, R. K., and Gowda, C. L. (2013). "Chickpea," in *Genetic and Genomic Resources of Grain Legume Improvement*, 81–111. doi:10.1016/b978-0-12-397935-3.00004-9

Sharma, H. C., Bhagwat, M. P., Pampapathy, G., Sharma, J. P., and Ridsdill-Smith, T. J. (2006). Perennial Wild Relatives of Chickpea as Potential Sources of Resistance to *Helicoverpa Armigera*. *Genet. Resour. Crop Evol.* 53 (1), 131–138. doi:10.1007/s10722-004-1951-4

Sharma, K. K., Lavanya, M., and Anjaiah, V. (2006). Agrobacterium-mediated Production of Transgenic Pigeonpea (*Cajanus Cajan* L. Millsp.) Expressing the Synthetic Bt cry1Ab Gene. *In Vitro* Cellular. *Dev. Biol. Plant.* 42 (2), 165–173. doi:10.1079/ivp2005730

Sharma, H. C. (2004). *A Little Help from Wild: Exploiting Wild Relatives of Chickpea for Resistance to* Helicoverpa Armigera. Patancheru, India: ICRISAT.

Siddique, K., and Sedgley, R. (1986). Chickpea (Cicer-Arietinum-L), a Potential Grain Legume for Southwestern Australia - Seasonal Growth and Yield. *Aust. J. Agric. Res.* 37, 245–261. doi:10.1071/Ar9860245

Sillero, J. C., Moreno-Alías, I., and Rubiales, D. (2012). Identification and Characterization of Resistance to Rust (Uromyces Ciceris-Arietini (Grognot) Jacz. And Boyd) in a Germplasm Collection of Cicer Spp. *Euphytica* 188 (2), 229–238. doi:10.1007/s10681-012-0709-x

Singh, K. B., and Ocampo, B. (1993). Interspecific Hybridization in Annual Cicer Species. *J. Genet. Breed.* 47 (3), 199–204.

Singh, K. B., and Reddy, M. V. (1993). Resistance to Six Races of *Ascochyta Rabiei* in the World Germplasm Collection of Chickpeas. *Crop Sci.* 33 (1), 186–189. doi:10.2135/cropsci1993.0011183x003300010033x

Singh, K. B., and Weigand, S. (1994). Identification of Resistant Sources in Cicer Species to *Liriomyza cicerina*. *Genet. Res. Crop Evol.* 41, 75–79.

Singh, K. B., and Singh, O. (1997). "Prospects of Creating Higher Yield Potential in Chickpea," in *Recent Advances in Pulses Research* (Kanpur: Indian Society of Pulses Research and Development, IIPR), 247, 9.

Singh, K. B., and Weigand, S. (1994). Identification of Resistant Sources in *Cicer* Species to Liriomyza Cicerina. *Genet. Resour. Crop Evol.* 41 (2), 75–79. doi:10.1007/bf00053051

Singh, K. B., Malhotra, R. S., and Saxena, M. C. (1990). Sources for Tolerance to Cold in Cicer Species. *Crop Sci.* 30 (5), 1136–1138. doi:10.2135/cropsci1990.0011183x003000050036x

Singh, K. B., Malhotra, R. S., and Saxena, M. C. (1995). Additional Sources of Tolerance to Cold in Cultivated and Wild Cicer Species. *Crop Sci.* 35 (5), 1491–1497. doi:10.2135/cropsci1995.0011183x003500050037x

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 Evol.* 45 (1), 9–17. doi:10.1023/a:1008620002136

Singh, N. P., Krishna, R., and Kumar, R. (2001). An Assay of Effects of Different Traits on Chickpea Grain Yield. *Ann. Agric. Res. New Ser.* 22 (4), 564–569.

Singh, S., Gumber, R. K., Joshi, N., and Singh, K. (2005). Introgression from Wild *Cicer Reticulatum* to Cultivated Chickpea for Productivity and Disease Resistance. *Plant Breed.* 124 (5), 477–480. doi:10.1111/j.1439-0523.2005.01146.x

Singh, A. K., Singh, R. V., Bharati, R. C., Chandra, N., Kumar, C., and Dimree, S. K. (2010). Introduction of Wild and Weedy Relatives of Crop Plants in India. *Environ. Ecol.* 28 (3A), 1715–1721.

Singh, R. P., Singh, I., Singh, S., and Sandhu, J. S. (2012). Assessment of Genetic Diversity Among Interspecific Derivatives in Chickpea. *J. Food Legumes* 25 (2), 150–152.

Singh, M., Dahiya, O. P., Singh, N., Nizar, A., Dutta, M., Tyagi, R. K., et al. (2013). Identification of an Upright Peduncle and Podding Genotype in Chickpea Germplasm Conserved in the National Gene Bank. *Indian J. Plant Genet. Resour.* 26 (2), 166–168.

Singh, M., Bisht, I. S., Dutta, M., Kumar, K., Basandrai, A. K., Kaur, L., et al. (2014). Characterization and Evaluation of Wild Annual Cicer Species for Agro-Morphological Traits and Major Biotic Stresses under Northwestern Indian Conditions. *Crop Sci.* 54, 229–239. doi:10.2135/cropsci2013.04.0225

Singh, M., Kumar, K., Bisht, I. S., Dutta, M., Rana, M. K., Rana, J. C., et al. (2015). Exploitation of Wild Annual Cicer Species for Widening the Gene Pool of Chickpea Cultivars. *Plant Breed.* 134 (2), 186–192. doi:10.1111/pbr.12254

Singh, U., Khemka, N., Rajkumar, M. S., Garg, R., and Jain, M. (2017). PLncPRO for Prediction of Long Non-coding RNAs (lncRNAs) in Plants and its Application for Discovery of Abiotic Stress-Responsive lncRNAs in Rice and Chickpea. *Nucl. Acids Res.* 45, e183. doi:10.1093/nar/gkx866

Singh, G. (1982). Screening Chickpea for Gray Mold Resistance. *Inter. Chickpea Newsl.* 7, 13–14.

Srinivasan, A., Saxena, N. P., and Johansen, C. (1999). Cold Tolerance during Early Reproductive Growth of Chickpea (*Cicer Arietinum* L.): Genetic Variation in Gamete Development and Function. *Field Crops Res.* 60 (3), 209–222. doi:10.1016/s0378-4290(98)00126-9

Srivastava, R., Bajaj, D., Malik, A., Singh, M., and Parida, S. K. (2016). Transcriptome Landscape of Perennial Wild *Cicer Microphyllum* Uncovers Functionally Relevant Molecular Tags Regulating Agronomic Traits in Chickpea. *Sci. Rep.* 6 (1), 1–17. doi:10.1038/srep33616

Srivastava, R., Upadhyaya, H. D., Kumar, R., Daware, A., Basu, U., Shimray, P. W., et al. (2017). A Multiple QTL-Seq Strategy Delineates Potential Genomic Loci Governing Flowering Time in Chickpea. *Front. Plant Sci.* 8, 1105. doi:10.3389/fpls.2017.01105

Sudupak, M., Akkaya, S., and Kence, A. (2002). Analysis of Genetic Relationships Among Perennial and Annual Cicer Species Growing in Turkey Using RAPD Markers. *Theor. Appl. Genet.* 105 (8), 1220–1228. doi:10.1007/s00122-002-1060-8

Sudupak, M. A., Akkaya, M. S., and Kence, A. (2004). Genetic Relationships Among Perennial and Annual Cicer Species Growing in Turkey Assessed by

AFLP Fingerprinting. *Theor.Appl. Genet.* 108, 937–944. doi:10.1007/s00122-003-1505-8

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 (1), 174–183. doi:10.1111/tpj.12105

Tan, G. Z., Das Bhowmik, S. S., Hoang, T. M., Karbaschi, M. R., Long, H., Cheng, A., et al. (2018). Investigation of Baseline Iron Levels in Australian Chickpea and Evaluation of a Transgenic Biofortifcation Approach. *Front. Plant Sci.* 9, 788. doi:10.3389/fpls.2018.00788

Tanksley, S. D., and McCouch, S. R. (1997). Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *science* 277 (5329), 1063–1066. doi:10.1126/science.277.5329.1063

Taran, B., Warkentin, T. D., Vandenberg, A., and Holm, F. A. (2010). Variation in Chickpea Germplasm for Tolerance to Imazethapyr and Imazamox Herbicides. *Can. J. Plant Sci.* 90 (1), 139–142. doi:10.4141/cjps09061

Tayyar, R. I., and Waines, J. G. (1996). Genetic Relationships Among Annual Species of Cicer (Fabaceae) Using Isozyme Variation. *Theor. Appl. Genet.* 92 (2), 245–254. doi:10.1007/bf00223381

Tekeoglu, M., Rajesh, P. N., and Muehlbauer, F. J. (2002). Integration of Sequence Tagged Microsatellite Sites to Chickpea Genetic Map. *Theor. Appl. Genet.* 105, 847–854. doi:10.1007/s00122-002-0993-2

Tewari-Singh, N., Sen, J., Kiesecker, H., Reddy, V. S., Jacobsen, H. J., and Mukherjee, S. G. (2004). Use of a Herbicide or Lysine Plus Threonine for Non Antibiotic Selection of Transgenic Chickpea. *Plant Cell Rep.* 22, 576–583. doi:10.1007/s00299-003-0730-6

Thompson, J. P., Reen, R. A., Clewett, T. G., Sheedy, J. G., Kelly, A. M., Gogel, B. J., et al. (2011). Hybridisation of Australian Chickpea Cultivars with Wild Cicer Spp. Increases Resistance to Root-Lesion Nematodes (*Pratylenchus Thornei* and *P. Neglectus*). *Australas. Plant Pathol.* 40 (6), 601–611. doi:10.1007/s13313-011-0089-z

Thudi, M., Bohra, A., Nayak, S. N., Varghese, N., Shah, T. M., Penmetsa, R. V., et al. (2011). Novel SSR Markers from BAC-End Sequences, DArT Arrays and a Comprehensive Genetic Map with 1291 Marker Loci for Chickpea (*Cicer Arietinum* L.). *PLoS ONE* 6 (11), e27275. doi:10.1371/journal.pone.0027275

Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic Dissection of Drought and Heat Tolerance in Chickpea through Genome-wide and Candidate Gene-Based Association Mapping Approaches. *Plos One* 9 (5), e96758. doi:10.1371/journal.pone.0096758

Thudi, M., Khan, A. W., Kumar, V., Gaur, P. M., Katta, K., Garg, V., et al. (2016). Whole Genome Re-sequencing Reveals Genome-wide Variations Among Parental Lines of 16 Mapping Populations in Chickpea (*Cicer Arietinum* L.). *BMC plant Biol.* 16 (1), 53–64. doi:10.1186/s12870-015-0690-3

Thudi, M., Chen, Y., Pang, J., Kalavikatte, D., Bajaj, P., Roorkiwal, M., et al. (2021). Novel Genes and Genetic Loci Associated with Root Morphological Traits, Phosphorus-Acquisition Efficiency and Phosphorus-Use Efficiency in Chickpea. *Front. Plant Sci.* 12, 636973. doi:10.3389/fpls.2021.636973

Toker, C., Canci, H., and Yildirim, T. O. L. G. A. (2007). Evaluation of Perennial Wild Cicer Species for Drought Resistance. *Genet. Resour. crop Evol.* 54 (8), 1781–1786. doi:10.1007/s10722-006-9197-y

Toker, C., Berger, J., Eker, T., Sari, D., Sari, H., Gokturk, R. S., et al. (2021). *Cicer Turcicum*: a New *Cicer* Species and its Potential to Improve Chickpea. *Front. Plant Sci.* 12, 587. doi:10.3389/fpls.2021.662891

Toker, C. (2005). Preliminary Screening and Selection for Cold Tolerance in Annual Wild Cicer Species. *Genet. Resour. Crop Evol.* 52 (1), 1–5. doi:10.1007/s10722-005-1743-5

Trick, M., Adamsk, N., M., Mugford, S., G., Jiang, C., C., Febrer, M., and Uauy, C. (2012). Combining SNP Discovery from Next-Generation Sequencing Data with Bulked Segregant Analysis (BSA) to Fine-Map Genes in Polyploidy Wheat. *BMC Plant Biol.* 12 (1), 1–17. doi:10.1186/1471-2229-12-14

Turner, N. C., Colmer, T. D., Quealy, J., Pushpavalli, R., Krishnamurthy, L., Kaur, J., et al. (2013). Salinity Tolerance and Ion Accumulation in Chickpea (*Cicer Arietinum* L.) Subjected to Salt Stress. *Plant Soil* 365 (1), 347–361. doi:10.1007/s11104-012-1387-0

Upadhyaya, H. D., Dronavalli, N., Gowda, C. L. L., and Singh, S. (2011). Identification and Evaluation of Chickpea Germplasm for Tolerance to Heat Stress. *Crop Sci.* 51 (5), 2079–2094. doi:10.2135/cropsci2011.01.0018

Upadhyaya, H. D., Dronavalli, N., Dwivedi, S. L., Kashiwagi, J., Krishnamurthy, L., Pande, S., et al. (2013). Mini Core Collection as a Resource to Identify New Sources of Variation. *Crop Sci.* 53 (6), 2506e2517. doi:10.2135/cropsci2013.04.0259

Upadhyaya, H. D., Bajaj, D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Genetic Dissection of Seed-Iron and Zinc Concentrations in Chickpea. *Sci. Rep.* 6 (1), 1–12. doi:10.1038/srep24050

Vadez, V., Krishnamurthy, L., Serraj, R., Gaur, P. M., Upadhyaya, H. D., Hoisington, D. A., et al. (2007). Large Variation in Salinity Tolerance in Chickpea Is Explained by Differences in Sensitivity at the Reproductive Stage. *Field crops Res.* 104 (1-3), 123–129. doi:10.1016/j.fcr.2007.05.014

Vadez, V., Krishnamurthy, L., Thudi, M., Anuradha, C., Colmer, T. D., and Turner, N. C. (2012). Assessment of ICCV 2 X JG 62 Chickpea Progenies Shows Sensitivity of Reproduction to Salt Stress and Reveals QTL for Seed Yield and Yield Components. *Mol. Breed.* 30, 9–21. doi:10.1007/s11032-011-9594-6

Van der Maesen, L. J. G., and Pundir, R. P. S. (1984). Availability and Use of Wild Cicer Germplasm. *FAO / IBPGR Plant Gen. Resour. Newsl.* 57, 19–24.

Van der Maesen, L. J. G., Maxted, N., Javadi, F., Coles, S., and Davies, A. M. R. (2007). "Taxonomy of the Genus Cicer Revisited," in *Chickpea Breed. Management* (Wallingford: CAB Int.), 14–46.

Van der Maesen, L. J. G. (1984). "Taxonomy, Distribution and Evolution," in *Genetic Resources and Their Exploitation: Chickpea, Faba Bean and Lentils*. Editors J. R. Witcombe and W. Erskine (The Hague, Netherlands, 95–101. doi:10.1007/978-94-009-6131-9_9

Van Dorrestein, B., Baum, M., and Malhotra, R. S. (1998). "Interspecific Hybridization between Cultivated Chickpea (*Cicer Arietinum* L.) and the Wild Annual Species *C. Judaicum* Boiss, *C. Pinnatifidum* Jaub. And Sp. and *C. Bijugum* KH. Rech," in Proceedings of the 3rdEuropean Grain Legume Conference, 362.

Vandemark, G. J., Grusak, M. A., and McGee, R. J. (2018). Mineral Concentrations of Chickpea and Lentil Cultivars and Breeding Lines Grown in the US Pacific Northwest. *Crop J.* 6 (3), 253–262. doi:10.1016/j.cj.2017.12.003

Vanderschuren, H., Lentz, E., Zainuddin, I., and FGruissem, W. (2013). Proteomics of Model and Crop Plant Species: Status, Current Limitations and Strategic Advances for Crop Improvement. *J. Proteome.* 93, 5–19. doi:10.1016/j.jprot.2013.05.036

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genomic-assisted Breeding for Crop Improvement. *Trends Plant Sci.* 10, 621–630. doi:10.1016/j.tplants.2005.10.004

Varshney, R. K., Hiremath, P. J., Lekha, P., Kashiwagi, J., Balaji, J., Deokar, A. A., et al. (2009). A Comprehensive Resource of Drought- and Salinity-Responsive ESTs for Gene Discovery and Marker Development in Chickpea (*Cicer Arietinum* L.). *BMC Genom* 10 (1), 1–18. doi:10.1186/1471-2164-10-523

Varshney, R. K., Ribaut, J. M., Buckler, E. S., Tuberosa, R., Rafalski, J. A., and Langridge, P. (2012). Can Genomics Boost Productivity of Orphan Crops? *Nat. Biotechnol.* 30 (12), 1172–1176. doi:10.1038/nbt.2440

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft Genome Sequence of Chickpea (*Cicer Arietinum* L.) Provides a Resource for Trait Improvement. *Nat. Biotechnol.* 31 (3), 240–246. doi:10.1038/nbt.2491

Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., et al. (2014). Genetic Dissection of Drought Tolerance in Chickpea (*Cicer Arietinum* L.). *Theor. Appl. Genet.* 127 (2), 445–462. doi:10.1007/s00122-013-2230-6

Varshney, R., K., Saxena, R., K., Upadhyaya, H., D., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome Resequencing of 292 Pigeonpea Accessions Identifies Genomic Regions Associated with Domestication and Agronomic Traits. *Nat. Genet.* 49, 1082–1088. doi:10.1038/ng.3872

Varshney, R. K., Thudi, M., Pandey, M. K., Tardieu, F., Ojiewo, C., Vadez, V., et al. (2018). Accelerating Genetic Gains in Legumes for the Development of Prosperous Smallholder Agriculture: Integrating Genomics, Phenotyping, Systems Modelling and Agronomy. *J. Exp. Bot.* 69 (13), 3293–3312. doi:10.1093/jxb/ery088

Varshney, R., K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). 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

Varshney, R. K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., et al. (2021). A Chickpea Genetic Variation Map Based on the Sequencing of 3,366. *Genomes. Nat.* 599 (7886), 622–627. doi:10.1038/s41586-021-04066-1

Varshney, R. K. (2016). Exciting Journey of 10 Years from Genomes to Fields and Markets: Some Success Stories of Genomics-Assisted Breeding in Chickpea, Pigeonpea and Groundnut. *Plant Sci.* 242, 98–107. doi:10.1016/j.plantsci.2015.09.009

Verdier, J., Torres-Jerez, I., Wang, M., Andriankaja, A., Allen, S. N., He, J., et al. (2013). Establishment of the Lotus Japonicus Gene Expression Atlas (LjGEA) and its Use to Explore Legume Seed Maturation. *Plant J.* 74 (2), 351–362. doi:10.1111/tpj.12119

Verma, M. M., Sandhu, J. S., Rraar, H. S., and Brar, J. S. (1990). Crossability Studies in Different Species of Cicer (L.). *Crop Improv.* 17 (2), 179–181.

Verma, M., Kumar, V., Patel, R. K., Garg, R., and Jain, M. (2015). CTDB: an Integrated Chickpea Transcriptome Database for Functional and Applied Genomics. *PLoS ONE* 10, e0136880. doi:10.1371/journal.pone.0136880

Wood, J. A., and Grusak, M. A. (2007). "Nutritional Value of Chickpea," in *Chickpea Breeding and Management*. Editors S. S. Yadav, R. Redden, W. Chen, and B. Sharma (Wallingford: CAB International), 101–142.

Xu, Y., Obielodan, M., Sismour, E., Arnett, A., Alzahrani, S., and Zhang, B. (2017). Physicochemical, Functional, Thermal and Structural Properties of Isolated Kabuli Chickpea Proteins as Affected by Processing Approaches. *Int. J. Food Sci.* 52 (5), 1147–1154. doi:10.1111/ijfs.13400

Yust, M. M., Pedroche, J., and Giron-Calle, J. (2003). Production of ACE Inhibitory Peptides by Digestion of Chickpea Legumin with Alcalase. *J. Food Chem.* 81, 363–369. doi:10.1016/s0308-8146(02)00431-4

Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011a). Chickpea Genotypes Contrasting for Seed Yield under Terminal Drought Stress in the Field Differ for Traits Related to the Control of Water Use. *Funct. Plant Biol.* 38 (4), 270–281. doi:10.1071/fp10244

Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011b). A Conservative Pattern of Water Use, rather Than Deep or Profuse Rooting, Is Critical for the Terminal Drought Tolerance of Chickpea. *J. Exp. Bot.* 62 (12), 4239–4252. doi:10.1093/jxb/err139

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 RS declared a shared affiliation with the author Gayacharan to the handling editor at the time of review.

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 © 2022 Singh, Singh, Ambika, Chandana, Mahto, Patial, Gupta, Gahlaut, Gayacharan, Hamwieh, Upadhyaya 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 04 August 2022
doi: 10.3389/fgene.2022.876763

Multilocation Yield Trials and Yield Stability Evaluation by GGE Biplot Analysis of Promising Large-Seeded Peanut Lines

Nipatcha Pobkhunthod[1,2], *Jetsada Authapun*[2], *Songyos Chotchutima*[2], *Sarawut Rungmekarat*[2], *Piya Kittipadakul*[2], *Jaungjun Duangpatra*[2] *and Tanapon Chaisan*[2]*

[1]Agricultural Research and Technology Transfer Center, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand, [2]Department of Agronomy, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand

OPEN ACCESS

Edited by:
Mahendar Thudi,
Dr. Rajendra Prasad Central Agricultural University, India

Reviewed by:
Cemal Kurt,
University of Çukurova, Turkey
Edmore Gasura,
University of Zimbabwe, Zimbabwe
Naiyin Xu,
Jiangsu Academy of Agricultural Sciences (JAAS), China
Ahmed Amri,
International Center for Agriculture Research in the Dry Areas (ICARDA), Ethiopia
Leonardo Lopes Bhering,
Federal University of Viçosa, Brazil

***Correspondence:**
Tanapon Chaisan
fagrtpc@ku.ac.th

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

Received: 15 February 2022
Accepted: 23 June 2022
Published: 04 August 2022

Citation:
Pobkhunthod N, Authapun J, Chotchutima S, Rungmekarat S, Kittipadakul P, Duangpatra J and Chaisan T (2022) Multilocation Yield Trials and Yield Stability Evaluation by GGE Biplot Analysis of Promising Large-Seeded Peanut Lines. Front. Genet. 13:876763. doi: 10.3389/fgene.2022.876763

The demand by industries for large-seeded peanuts is increasing in Thailand and Southeast Asia. New large-seeded peanut lines were recently developed in Thailand to respond to the demand. In this study, a multilocation yield trial was performed to identify the best genotype(s) in Thailand's central region and investigate the genotype–environment interaction (GEI) on peanut production. Twelve promising large-seeded peanut lines and two check varieties (KU50 and KK6) were planted at 12 different planting locations during the dry and rainy seasons of 2018 and the dry season of 2019. This study found significant yield potential variability in the promising lines of peanuts evaluated at different planting locations. A combined analysis of variance presented that the environment and genotypes had a considerable impact ($p < 0.001$) on the pod and seed yield. The GEI showed a high impact ($p < 0.01$) on pod yield and an effect ($p < 0.05$) on seed yield. The environment presented the most significant influence on pod and seed yield variations, followed by genetics and GEI. The total variation in seed yield was 64.22%, composed of PC1 and PC2 values at 45.71% and 18.51%, respectively. The GGE biplot analysis of the yield potentials at each location indicated that KUP12BS029-1-1-3 was the ideal genotype, with a high yield potential and most stability at multilocations, followed by KUP12BS030-3-4-1 and KUP12BS030-1-4-3. These promising lines will be released as new peanut varieties in central Thailand and are recommended as parental lines in breeding programs for large-seeded and yield potential in Thailand and Southeast Asia.

Keywords: yield stability, GGE biplot, GEI, pod yield, seed yield

INTRODUCTION

Peanut (*Arachis hypogaea L.*) is an important legume crop in Southeast Asia. The market demand for large-seeded peanuts has been increasing, and better quality large-seeded peanut varieties are needed for peanut products, particularly by farmers. Therefore, breeding for yield has been the primary strategy to improve peanut productivity in peanut-growing countries (Nigam et al., 1991).

The peanut seed size is essential for the processed industry and can be used to measure the quality and price of peanuts. The size of the peanut seed indicates the quality of the peanut product. Peanut products made of large seeds taste better than those made of small seeds (Haruthaithanasan, 2002). The standard for grading peanut seeds are the following—1) large-seeded: 100-seeds weight of over

60 g, 2) medium-seeded: 100-seeds weight between 35 and 60 g, and 3) small-seeded: 100-kernels weight of less than 35 g (Waranyawat, 1999).

Most large-seeded peanut varieties have a late harvesting date, which is unsuitable for Thailand's cropping system. The new large-seeded peanut varieties with resistance to peanut bud necrosis virus and early maturity date and suitable for the central region of Thailand were developed in 2010 under a peanut breeding program conducted by the Department of Agronomy, the Faculty of Agriculture, Kasetsart University, Thailand. Several large-seeded peanut genotypes were selected and evaluated at the research station in this peanut breeding program. A yield trial for the selected large-seeded peanut genotypes is required to determine their yield potential and stability at various planting locations (Authapun et al., 2016). This process is conducted in different environments over many years to measure the adaptability of the plants, identifying those that are potentially adaptable to specific environments and those that are generally adaptable (Yan and Hunt, 2001).

Selection for high yield and stability with critical economic traits, such as yield and adaptability, is essential in successful breeding programs. The necessary factors for consideration in breeding studies are variety, location, season, and environment (De lacy, 1981). Multi-environment or multilocation yield trials conducted in plant breeding programs are essential in evaluating genotypes and hybrids for yield and stability (Alwala et al., 2010). An essential factor in the stability studies of peanuts is the yield potential at multiple locations (Shorter and Norman, 1983). Generally, the yield of plants is influenced by the environment (E) more than the genotype–environment interaction (GEI) and genotype (G). Different peanut lines therefore show the highest yield potential at different locations (Tai and Hammons, 1978; Oliveira and Godoy, 2006; Kasno and Trustinah, 2015; De Moura et al., 2017). Testing plant varieties in different testing environments, where the responses would reflect the adaptability of the peanut genotypes to the inherent and persistent natural environmental factors of the different peanut production areas, provides valuable information for cultivar selection and final release (Banterng et al., 2006). Determining the most promising multilocation lines based on the findings of a preliminary yield trial, demonstrating yield potential and stability, is vital for selecting and releasing new varieties.

Different methods of analysis have been proposed to determine GEI, such as using a regression coefficient (Finlay and Wilkinson, 1963), calculating the sum of squared deviations from the regression (Eberhart and Russel, 1966), and the additive main effects and multiplicative interaction (AMMI) (Gauch and Zobel, 1988; Annicchiarico, 1997). Yan et al. (2000) proposed a method known as genotype plus genotype by environment interaction (GGE) or a multi-environment experiment in which the G + GE graph (GGE biplot) is displayed as a graph to facilitate the evaluation of the visible traits and the mega-environment identification of each genotype. The GGE biplot identifies the winning genotypes by crop every year. Therefore, using the GGE graph compares genotypes in different environments according to the best genotype and environmental performance, such as genotype grouping and environment. The polygon view of a GGE biplot explicitly displays the "which-won-where" pattern and hence provides a succinct summary of the GE pattern of a multi-environment trial data set. The polygon is formed by connecting the markers of the genotypes that are further away from the biplot origin such that all other genotypes are contained in the polygon (Yan, 2002).

The GGE biplot presents a graphic illustration of the data collected and aids in evaluating the comparative results. This biplot provided highly reliably graphical results to identify high yields and stability in genotypes in hybrid maize (Alwala et al., 2010). The biplot identified the mega-environment that influenced the variability of grain yield in barley (Jalata, 2011). It was used to select peanut germplasm for development as a specific genotype depending on environmental and management conditions (Zurweller et al., 2018). The objective of the current study was to apply a multilocation evaluation of large-seeded peanuts for yield stability. Specifically, the goals were to identify large-seeded peanut line(s) with high yield potential and stability suitable for multilocations.

MATERIALS AND METHODS

Plant Materials

The peanut genotypes used in this study were 12 promising lines with a 100-seeds weight of more than 60 g. These lines were crossed from the Khon Kean 5 (KK5) variety (a Virginia-type peanut of medium seed size, high yield, a decumbent type canopy, and low incidence of bud necrosis virus disease) and Khon Kean 6 (KK6) variety (a Virginia-type, large-seeded peanut, with large pods, an erect plant type canopy, dark green leaves, and resistant to bud necrosis virus disease).

The two large-seeded check varieties were KK6 and Kasetsart 50 (KU50), recommended by the Thailand Department of Agriculture and often used by Thai farmers. KU50 is resistant to drought and foliar diseases and has a high yield and high seed dormancy.

Multilocation Yield Trials

Twelve promising large-seeded peanut lines and two check varieties (KU50 and KK6) (**Supplementary Table S1**) were planted at twelve locations in central Thailand during the dry and rainy seasons of 2018 and the dry season of 2019 (**Supplementary Table S2**). The experiment at each location was conducted using a randomized complete block design with three replications. Crop management in this experiment included planting date, plant spacing, and irrigation at each location; each plot consisted of six rows and was 4 m long. Preemergent herbicides were used to control weeds in the experiments in all fields. All plots were fertilized during flowering at 156.25 $kg{\cdot}ha^{-1}$ of NPK (13-13-21). Gypsum ($CaSO_4{\cdot}2H_2O$) was applied at early pegging at 312.50 $kg{\cdot}ha^{-1}$. The peanut fields were irrigated under drought conditions (**Supplementary Table S2**). The pod yield and seed yield were collected and reported at 8% moisture content.

Weather and Soil Fertility

In the dry season of 2018, the average temperature ranged from 30.5°C to 30.9°C, and the total rainfall was approximately 166.3–380.6 mm. During the rainy season of 2018, the average temperature ranged from 27.7°C to 28.5°C, and the total rainfall was approximately 459.0–732.1 mm. During the dry season of 2019, the average temperature ranged from 30.7°C to 30.9°C, and the total rainfall was approximately 166.6–262.5 mm (**Supplementary Table S2**).

The planting area trials were conducted at 12 locations with known environmental differences, namely, 1) plains, 2) river plains, 3) piedmont plains, and 4) upland (**Supplementary Table S2**). The soil series in the planted areas were Nakhon Sawan, Lop Buri, Khok Samrong, Choke Chai, and Mae Sai. The Nakhon Sawan soil series is a shallow soil group to the rock wall layer. The soil reaction is acidic to neutral, with good drainage and low fertility. The Lop Buri soil series is an intense black clay soil group with deep and wide cracks when dry. The soil reaction in this soil series is neutral to alkaline, with moderate to good drainage and moderate to high fertility. The Khok Samrong soil series is an enthusiastic fine loam soil group. The soil reaction of this soil series is neutral to alkaline, drainage is quite inadequate to worse and fertility is moderate to low. The Choke Chai soil series is a deep to intense clay soil group produced from the fine mass parent material. The soil reaction of this soil series is strongly acidic and the drainage quite bad, with low fertility. The Maesai soil series is siltstone sandy soil that has risen from distributary sedimentation. The soil reaction of this soil series is neutral or basic, and drainage is quite bad and fertility is moderate to low.

Data Collection

The pod and seed yields were determined based on sampling in the two center rows of each plot, excluding plants at the head and end of each row. The pod yield was based on sun-dried measurements to reduce the moisture content. Dry pod and seed weights were measured and recorded at 8% moisture content.

Statistical Data Analysis

The collected data were analyzed using the R program version 3.6.1 [R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/] for pod and seed yield variance at multiple locations. The analysis of variance was used to compare the means in each environment. Significant differences among the means of yields were compared using Fisher's protected least significant difference (LSD) at the 5% level of probability. Evaluation for the yield potential and stability was performed based on a GGE biplot graphical user interface package (GGE biplot GUI) using principal component 1 (PC1) and PC2 scores. Visualization of "which-won-where/what" patterns for the multilocation yield trials was used to study genotypes (G) and genotype–environment interactions (GEI).

RESULTS

Combined Analysis of Variance for Pod Yield and Seed Yield

The results showed that environment (E), genotype–environment interaction (GEI), and genotype (G) significantly affected pod and seed yields (**Table 1**). In particular, GEI highly affected the pod yield ($p < 0.01$) and had a significant effect on the seed yield ($p < 0.05$). The highest variations based on the percentage of total variability represented by the total sum of squares in pod and seed yields were 55.19% and 64.97%, respectively. The E factor had the most significant influence on pod and seed yield variations. The GEI variation had the second greatest significance with 15.07% (pod) and 10.17% (seed). The G variation was the least influential, with 4.31% (pod) and 2.81% (seed).

Correlation Coefficient Analysis

Correlations of yield components on the pod and seed yields were analyzed for the 12 planting locations. All characteristics significantly affected the yield. The number of seeds per plant and pods per plant showed a high correlation with the pod yield (r = 0.73, 0.70) and seed yield (r = 0.70, 0.53), respectively. However, the number of seeds per pod had a low correlation with the pod yield and seed yield (r = 0.04, 0.05). Similarly, 100-seeds weight had a low correlation with the pod yield (r = 0.03) (**Table 2**).

Yield Performance at Multilocations

The number of pods, seeds, 100-seeds weight, pod yield, and seed yield of the 14 large-seeded peanut lines in the 12 planting locations differed significantly (**Table 3**). The average pod yield differed significantly, ranging from 2.62 to 3.89 t·ha^{-1}. The average pod yield of 14 large-seeded peanut lines at each planting location ranged from 0.90 [Wang Thong (WT)] to 4.95 t·ha^{-1} [Chon Phrai (CP)] (**Table 4**). The pod yield of the 14 large-seeded peanut lines was significantly different for 5 out of the 12 planting locations. The pod yield of KUP12BS029-1-1-3 (5.92 t·ha^{-1}) was higher than that of KU50 (2.42 t·ha^{-1}) at Lam Sonthi (LST). The pod yield of the KUP12BS030-4-2-1 (1.42 t·ha^{-1}) peanut line was higher than that of KK6 (0.82 t·ha^{-1}) and KU50 (0.53 t·ha^{-1}) at WT.

KUP12BS029-1-1-3 showed a high performance for the number of pods, number of seeds, pod yield, and seed yield. The average seed yields from the 12 planting locations were significantly different, ranging from 1.45 to 2.17 t·ha^{-1}. The averaged seed yield of the KUP12BS029-1-1-3 line was higher than for that of KK6 (20.70%) and KU 50 (26.80%) (**Table 3**). At each planting location, the average seed yield ranged from 2.19 (WT) to 3.64 t·ha^{-1} (CP) (**Table 5**). The seed yield of the 14 large-seeded peanut lines in each planting area was significantly different in 3 out of the 12 planting locations at WP, Sa Bot (SB), and Wang Phloeng (WP). The seed yield of KUP12BS030-1-4-3 (1.26 t·ha^{-1}) was higher than that for KU50 (0.82 t·ha^{-1}) and KK6 (0.53 t·ha^{-1}) at SB. The seed yield of KUP12BS030-4-2-1 (0.48 t·ha^{-1}) was higher than that for KU50 (0.18 t·ha^{-1}) and KK6

TABLE 1 | Combined analysis of pod yield and seed yield of 14 peanut lines tested across 12 locations.

Source of variation	Df	Pod yield (t·ha^{-1})				Seed yield (t·ha^{-1})			
		SS	MS	%SS	Pr (>F)	SS	MS	%SS	Pr (>F)
Environment (E)	11	799.24	72.658***	55.19	<2.2e-16***	363.43	33.039***	64.97	<2.2e-16***
Rep: E	12	49.70	4.141***	3.43	6.096e-06***	19.75	1.645***	3.53	6.647e-08***
Genotype (G)	13	62.43	4.802***	4.31	1.558e-07***	15.75	1.212***	2.81	101e-05***
G×E	143	208.18	1.456**	14.38	0.004446**	56.94	0.398*	10.17	0.04893*
Pooled error	324	328.58	1.014	22.69		103.52	0.320	18.51	
Total	503	1448.13				559.39			

****, **, *: significant at 0.001, 0.01, and 0.05 probability levels, respectively.*
Df, degrees of freedom; SS, sum of squares; MS, mean square.

TABLE 2 | Correlation coefficients of yield components with pod yield and seed yield of 14 peanut lines testing across 12 locations.

Yield component	Correlation coefficient	
	Pod yield	Seed yield
No. of pods per plant	0.70***	0.53***
No. of seeds per pod	0.04***	0.05***
No. of seeds per plant*	0.73***	0.70***
100-seeds weight (g)	0.19***	0.19***
Shelling percentage (%)	0.03***	0.15***

**; significant at 0.001 probability level.*

(0.25 t·ha^{-1}) at WT. However, the seed yield of KUP12BS029-1-1-3 was not different from that of KK6 and KU 50 at each location.

Genotype Plus Genotype by Environment Interaction Biplot Analysis

The results of the GGE biplot showed that the total variation of the pod yields was 70.96%, composed of PC1 (47.04%) and PC2 (23.92%) (**Figure 1A**). The PC1 score indicates the yield of the lines: PC1 > 0 indicates the high yield lines, whereas PC1 < 0 indicates the low yield lines. The PC2 score derived from the multilocation tests indicates a line's stability. If the PC2 score approaches zero, the lines are stable. Based on the GGE biplot analysis, the peanut lines showed PC1 > 0 and low PC2 scores were KUP12BS030-1-4-3 (G7) and KUP12BS029-1-1-3 (G6), indicating high yield and high stability.

The total variation in the seed yield was 64.26%, composed of PC1 and PC2 values of 45.73% and 18.53%, respectively (**Figure 1B**). The lines that showed high stability and stable seed yield (PC1 > 0 and low PC2 score) were KUP12BS029-1-1-3 (G6), KUP12BS030-1-4-3 (G7), and KUP12BS030-3-4-1 (G8). KUP12BS029-1-1-3 (G6) had the highest mean pod yield and seed yield. Identification of the ideal genotype for pod and seed yields showed that KUP12BS029-1-1-3 (G6) was positioned the closest to the ideal pod and seed yield lines (**Figure 2**). Then, the KUP12BS029-1-1-3 (G6) was indicated as the ideal large-seeded peanut line for high stability and high pod and seed yields.

The analysis for the pod yield of the peanut lines studied at multilocations was facilitated using a "which-won-where" pattern, showing the interaction of genotype to the data sets of different environments at the multilocation yield trials (Yan, 2002). The polygon view of this biplot showed the test locations in six sectors, where the lines at the corner of each section had the highest yield. The polygon view showed that the KUP12BS029-1-1-3 (G6) was the highest pod yield in Tak Fa (TF), Wang Phloeng (WP), Si Thep (ST), and Lam Sonthi (LST). KUP12BS030-4-2-1 (G9), KUP12BS001-3-4-3 (G3), KUP12BS014-3-4-1 (G4), and KK6 (G2) had the highest pod yields in Sa Bot (SB) and Tha Luang (TL), Phatthana Nikhom (PN), Wang Muang (WM), and Chon Phrai (CP), respectively (**Figure 3A**).

Considering the seed yield of the lines at each location, the KUP12BS029-1-1-3 (G6) had the highest seed yield at Lam Sonthi (LST), Si Thep (ST), Wang Phloeng (WP), and Khok Charoen (KC) locations (**Figure 3B**). KUP12BS030-4-2-1 (G9) produced the highest seed yield in Tha Luang (TL), Tak Fa (TF), Wang Thong (WT), Phatthana Nikhom (PN), and Sa Bot (SB). KUP12BS014-3-4-1 (G4) showed the highest seed yield in Wang Muang (WM). Of the 12 testing locations, Lam Sonthi (LST) and Khok Samrong (KS), which had long vectors, were the most discriminating in pod yield, followed by Si Thep (ST) and Khok Charoen (KC) (**Figure 4A**), whereas Si Thep (ST) and Lam Sonthi (LST) were the most discriminating in seed yield (**Figure 4B**).

The average environment axis (AEA) of the biplot (**Figure 4**) is the line that passes through all locations represented by the circles and the center of the circles. The angle between the vectors at each location and AEA can be used to identify the representative location. A tested location with the smallest angle between its vector and AEA is the best representative of the test location (Yan and Tinker, 2006). Thus, ST (Si Thep) was the most representative location for both the pod yield and seed yield (**Figure 4**).

DISCUSSION

The combined analysis of variance showed significant effects for all sources of variation for both pod and seed yields, indicating the differential behavior of genotypes that were not consistent with the different environments. These results were consistent with other published reports of peanut yield trials (Dos Santo et al., 2012; Kasno and Trustinah, 2015). In the current study,

TABLE 3 | Yield and yield components of 12 peanut lines and 2 check varieties.

Code	Line/variety	No. pods per plant	No. seeds per plant	100-Seeds weight (g)	Pod yield (t·ha^{-1})	Relative to check		Relative to mean (%)	Seed yield (t·ha^{-1})	Relative to check		Relative to mean (%)
						KK6 (%)	KU50 (%)			KK6 (%)	KU50 (%)	
G1	KU50	30^{bc}	47^{bc}	56.14^{d}	3.19^{def}	92.78	100.00	95.15	1.41$^{b–e}$	95.19	100.00	96.38
G2	KK6	27^{cd}	40^{cd}	64.81^{bc}	3.41$^{a–e}$	100.00	107.79	102.56	KK6	100.00^{cd}	105.05^{cd}	101.25^{bc}
G3	KUP12BS001-3-4-3	31^{b}	48^{ab}	53.44^{d}	3.00^{efg}	87.35	94.15	89.59	1.62^{cde}	90.19	94.75	91.32
G4	KUP12BS014-3-4-1	25^{d}	41^{cd}	55.37^{d}	2.79^{fg}	81.12	87.44	83.20	1.45^{e}	80.54	84.61	81.55
G5	KUP12BS014-5-1-3	29^{bc}	44^{bcd}	52.71^{d}	2.62^{g}	76.32	82.26	78.27	1.50^{de}	83.28	87.48	84.32
G6	KUP12BS029-1-1-3	35^{a}	53^{a}	63.64^{bc}	3.89^{a}	113.11	121.92	116.01	2.17^{a}	120.70	126.80	122.21
G7	KUP12BS030-1-4-3	30^{bc}	46^{bc}	64.61^{bc}	3.66^{abc}	106.45	114.73	109.17	1.92^{ab}	106.72	112.11	108.06
G8	KUP12BS030-3-4-1	30^{bc}	43^{bcd}	68.88^{a}	3.52$^{a–d}$	102.38	110.35	105.00	1.89^{bc}	103.79	109.03	105.09
G9	KUP12BS030-4-2-1	30^{bc}	45^{bcd}	66.97^{ab}	3.64$^{a–d}$	105.82	114.06	108.53	1.95^{ab}	108.59	114.07	109.94
G10	KUP12BS031-2-4-2	29^{bc}	41^{cd}	63.96^{bc}	3.72^{ab}	108.22	116.65	111.00	1.86^{bc}	103.44	108.67	104.74
G11	KUP12BS031-5-2-1	26^{cd}	39^{d}	65.58^{abc}	3.29^{b-e}	95.59	103.03	98.04	1.73^{bcd}	96.25	101.11	97.45
G12	KUP12BS036-4-2-3	30^{bc}	39^{d}	65.72^{abc}	3.64^{a-d}	105.77	114.00	108.48	1.75$^{b–e}$	95.26	100.07	96.45
G13	KUP12BS050-2-4-2	29^{bc}	42^{bcd}	66.71^{ab}	3.25$^{c–f}$	94.49	101.84	96.91	1.78^{bc}	99.37	104.39	100.61
G14	KUP12BS054-2-4-3	31^{b}	45^{bcd}	61.88^{c}	3.29$^{b–e}$	95.65	103.10	98.10	1.78^{bc}	98.83	103.82	100.07
	Mean	29.43	43.79	62.17	3.35	97.50	105.09	100.00	1.77	98.73	103.71	100.00
	F-test	**	***	**	***				***			
	CV (%)	31.04	34.74	15.57	32.84				34.31			
	LSD (0.05)	4.48	5.66	4.07	0.46				0.26			

****, **: significant at 0.001 and 0.01 probability levels.*
Means with the different lowercase superscripts (a–g) in the same column represent significant differences.

the location influenced the large-seeded peanut lines. The pod yield and seed yield were the most critical factors affecting peanut yields. The influence of environmental variation on yield was more significant than GEI and genotype variations. Variations in the pod and seed yields of lines at the 12 locations were influenced mainly by the E effect, with minor influences by GEI and G. These results are consistent with reported GEIs in peanuts (Oliveira and Godoy, 2006; Kasno and Trustinah, 2015; De Moura et al., 2017). The variation of GEI is three times greater than the variation of G. Compared to G, the magnitude of GEI suggests that different environments might exist, which makes the breeder's work with a selection more difficult. This result is consistent with earlier reports from a study of grain yield variation in barley caused by GEI (Jalata, 2011).

As the primary target for breeding in peanuts, the yield received more attention than the other examined traits. Correlation analyses revealed a pattern of association of traits and their direct contributions to yield. The correlation showed that the number of pods per plant and the number of seeds per plant had the most substantial positive direct effects on pod yield and seed yield, indicating that improvements in these traits should improve the total yield. These results follow reports regarding the number of pods per plant (Nandini and Savithramma, 2012; Thirumala et al., 2014). The 100-seeds weight and shelling percentage trait had a positive direct effect on yield, with weaker effects observed for the number of pods per plant and the number of seeds per plant trait. This information suggests that special attention should be paid to these traits when selecting high-yielding genotypes of large-seeded peanuts. The 100-seeds weight, based on the test planting data at 12 locations, found that some lines had 100-seeds weight less than 60 g, indicating the additive genetic effects and an opportunity to improve this trait *via* selection. Similar results were found for the peanut pod weight (Patil et al., 2014).

The current study identified that the rankings of the genotypes, based on yield, were changed for specific locations. For example, the pod yield of KUP12BS029-1-1-3, KUP12BS030-4-2-1, and KUP12BS036-4-2-3 presented the highest pod yield at LST, WT, and KS, respectively. These results presented the influence of GEI on the pod yield. GEI reduced the usefulness of lines by reducing their yield performance (Pham and Kang, 1988). Moreover, the pod and seed yield evaluations under multilocations showed different peanut lines presenting the highest pod yield at different locations. These results indicate

TABLE 4 | Pod yield (t·ha^{-1}) of 12 peanut lines and 2 check varieties planted in 12 locations.

Code	Line/variety	Location												
		CP	KC	KS	LST	PN	SB	ST	TF	TL	WM	WP	WT	Mean
G1	KU50	4.67	3.73^{abc}	5.96$^{a–d}$	2.42^{c}	3.25	1.56	3.83	3.42	2.83	2.37	2.89$^{a–d}$	0.53^{d}	3.16
G2	KK6	6.04	5.38^{ab}	7.05^{ab}	3.88^{abc}	2.05	1.58	3.77	3.62	1.35	2.82	3.40^{ab}	0.82^{bcd}	3.48
G3	KUP12BS001-3-4-3	5.65	4.38^{abc}	5.20^{-e}	2.43^{c}	2.75	1.33	3.11	3.73	3.16	2.36	1.29^{e}	0.69^{cd}	3.01
G4	KUP12BS014-3-4-1	3.88	2.38^{c}	3.30$^{a–e}$	2.56^{c}	2.53	2.49	2.64	3.56	2.81	2.97	2.00^{cde}	0.42^{d}	2.63
G5	KUP12BS014-5-1-3	4.87	3.13^{bc}	3.99^{de}	2.83^{c}	2.62	1.27	3.82	2.77	2.71	2.87	1.78^{cde}	0.87^{bcd}	2.79
G6	KUP12BS029-1-1-3	5.26	4.71^{abc}	6.18^{abc}	5.92^{a}	2.87	2.16	6.44	3.64	3.15	2.59	2.91$^{a–d}$	0.89^{bcd}	3.89
G7	KUP12BS030-1-4-3	5.18	3.67^{bc}	6.11^{a-d}	4.83^{abc}	2.72	3.69	4.51	3.67	2.14	2.83	3.42^{ab}	1.20^{abc}	3.66
G8	KUP12BS030-3-4-1	4.79	5.05^{ab}	5.22^{a-e}	5.85^{a}	2.26	1.89	4.21	3.06	3.30	2.46	2.96^{a-d}	1.25^{ab}	3.52
G9	KUP12BS030-4-2-1	4.95	4.58^{abc}	4.23^{de}	5.70^{ab}	2.38	1.91	5.20	4.84	3.26	2.18	3.07^{abc}	1.42^{a}	3.64
G10	KUP12BS031-2-4-2	4.90	6.07^{a}	7.06^{ab}	3.46^{c}	2.55	1.73	5.35	4.08	2.84	2.29	3.58^{a}	0.80^{bcd}	3.73
G11	KUP12BS031-5-2-1	5.41	4.56^{abc}	4.84$^{b–e}$	3.03^{c}	2.36	2.18	4.27	3.64	2.97	2.15	3.40^{b}	0.69^{cd}	3.29
G12	KUP12BS036-4-2-3	5.06	5.11^{ab}	7.46^{a}	4.49^{c}	2.79	2.29	4.49	3.49	2.64	2.85	2.09^{cde}	0.94$^{a–d}$	3.64
G13	KUP12BS050-2-4-2	4.29	4.62^{abc}	4.84$^{b–e}$	3.99^{c}	2.50	1.93	3.97	4.35	2.88	2.40	2.47$^{a–e}$	0.78^{bcd}	3.25
G14	KUP12BS054-2-4-3	4.35	3.91^{abc}	4.42^{cde}	4.58^{c}	2.44	2.53	4.72	3.77	2.73	2.53	2.25$^{b–e}$	1.27^{ab}	3.29
	Mean	4.95	4.38	5.42	4.00	2.58	2.07	4.31	3.69	2.77	2.55	2.69	0.90	
	F-test	ns	*	*	*	ns	ns	ns	ns	ns	ns	*	*	
	CV (%)	9.34	32.60	25.01	37.14	18.83	43.79	32.86	24.75	31.54	16.3	26.75	34.27	
	LSD (0.05)	1.39	2.35	2.25	2.45	0.71	1.59	2.33	1.58	1.57	0.69	1.59	0.50	

**: significant at 0.05 probability level; ns, not significant.*

Means with the different lowercase superscripts (a–e) in the same column represent significant differences.

CP, Chon Phrai; KC, Khok Charoen; KS, Khok Samrong; LST, Lam Sonthi; PN, Phatthana Nikhom; SB, Sa Bot; ST, Si Thep; TF, Tak Fa; Tl, Tha Luang; WM, Wang Muang; WT, Wang Thong; WP, Wang Phloeng.

TABLE 5 | Seed yield (t·ha^{-1}) of 12 peanut lines and 2 check varieties planted in 12 locations.

Code	Line/variety	Location												
		CP	KC	KS	LST	PN	SB	ST	TF	TL	WM	WP	WT	Mean
G1	KU50	4.06	1.59	3.57	2.32	1.15	0.53^{b}	2.40	2.12	0.81	0.82	1.40^{ab}	0.25^{bcd}	1.75
G2	KK6	2.69	2.18	3.30	1.42	1.57	0.82^{b}	2.46	2.05	1.66	1.56	1.25^{abc}	0.18^{cd}	1.76
G3	KUP12BS001-3-4-3	3.10	1.72	2.83	1.64	1.41	0.50^{b}	1.92	2.03	2.05	1.49	0.57^{d}	0.24^{bcd}	1.62
G4	KUP12BS014-3-4-1	2.93	0.81	1.90	1.67	1.40	0.73^{b}	1.70	2.16	1.78	1.41	0.92^{bcd}	0.13^{d}	1.46
G5	KUP12BS014-5-1-3	3.04	1.37	2.21	1.38	1.72	0.55^{b}	2.49	1.60	1.44	1.23	0.67^{d}	0.29$^{a–d}$	1.50
G6	KUP12BS029-1-1-3	3.64	2.03	3.62	3.22	1.72	0.63^{b}	4.36	2.30	1.92	1.15	1.22^{abc}	0.29$^{a–d}$	2.18
G7	KUP12BS030-1-4-3	3.13	1.46	3.18	2.68	1.65	1.26^{a}	2.78	2.52	1.19	1.36	1.49^{a}	0.38^{abc}	1.92
G8	KUP12BS030-3-4-1	3.49	1.89	2.82	2.92	1.29	0.72^{b}	2.76	1.75	1.91	1.26	1.31^{abc}	0.33$^{a–d}$	1.87
G9	KUP12BS030-4-2-1	3.05	1.74	2.23	3.17	1.52	0.54^{b}	3.19	2.93	2.04	1.27	1.31^{abc}	0.48^{a}	1.96
G10	KUP12BS031-2-4-2	3.42	2.05	3.19	1.75	1.44	0.68^{b}	3.14	2.42	1.52	1.11	1.40^{ab}	0.24^{bcd}	1.86
G11	KUP12BS031-5-2-1	3.74	1.65	2.10	1.61	1.09	0.70^{b}	2.90	2.28	1.73	1.28	1.56^{a}	0.17^{cd}	1.73
G12	KUP12BS036-4-2-3	2.67	1.70	3.22	2.20	1.70	0.51^{b}	2.66	2.01	1.55	1.32	0.80^{cd}	0.24^{bcd}	1.72
G13	KUP12BS050-2-4-2	3.02	1.76	2.88	2.49	1.53	0.72^{b}	2.39	2.68	1.52	1.19	1.07$^{a–d}$	0.22^{bcd}	1.79
G14	KUP12BS054-2-4-3	2.99	1.64	2.40	2.54	1.40	0.84^{b}	3.06	2.27	1.57	1.29	0.95^{bcd}	0.42^{ab}	1.78
	Mean	3.21	1.69	2.82	2.21	1.47	0.70	2.73	2.22	1.62	.1.27	1.14	0.28	
	F-test	ns	ns	ns	ns	ns	*	ns	ns	ns	ns	**	*	
	CV (%)	14.85	34.47	29.63	38.27	18.83	27.36	35.28	20.50	32.71	28.10	26.99	46.42	
	LSD (0.05)	0.80	0.95	1.49	1.46	0.54	0.36	1.53	0.77	0.93	0.58	0.36	0.21	

***, *: significant at 0.01 and 0.05 probability levels; ns, not significant.*

Means with the different lowercase superscripts (a–d) in the same column represent significant differences.

CP, Chon Phrai; KC, Khok Charoen; KS, Khok Samrong; LST, Lam Sonthi; PN, Phatthana Nikhom; SB, Sa Bot; ST, Si Thep; TF, Tak Fa; Tl, Tha Luang; WM, Wang Muang; WT, Wang Thong; WP, Wang Phloeng.

that the influence of GEI had a crossover effect on pod and seed yields of these peanut lines. These results are consistent with earlier published reports of peanut yield trials (Banterng et al., 2006; Abhinandan et al., 2018; Zurweller et al., 2018; Okori et al., 2019). The mean yields observed in Chon Phrai were higher than in other sites. This result is not surprising when we consider that the Maesai soil series in the north of Chon Phrai is naturally more advantageous for this promising population.

The GGE analysis was performed to examine the performance and stability of promising lines and standard checks. This analysis was designed to provide an insight into the effect of G and GEI on yield and identify genotypes that are particularly well suited for a particular environment. It has been valuable for characterizing

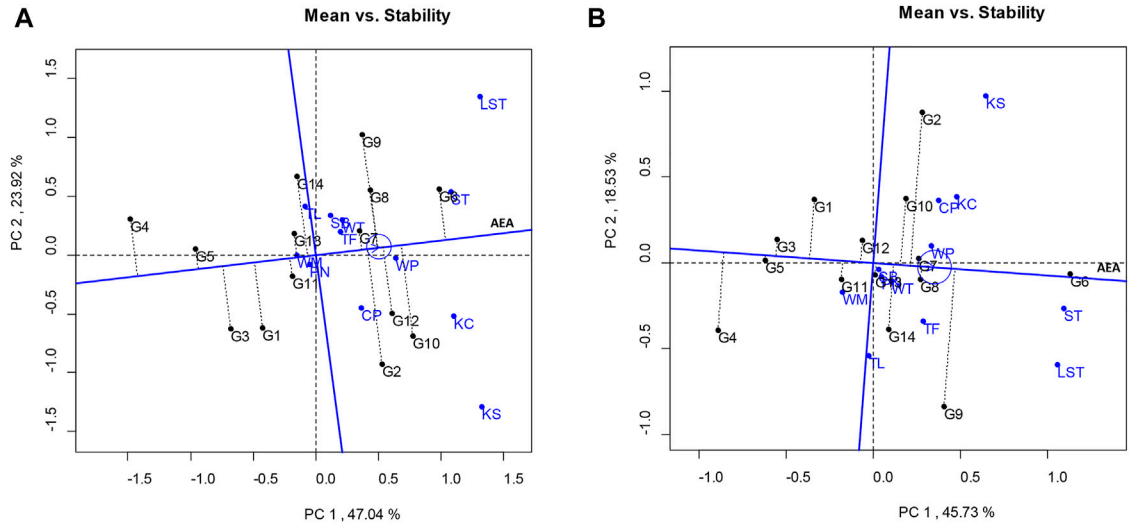


FIGURE 1 | Average environment coordination views of the GGE biplot based on location-focused scaling of mean performance and stability of genotypes. **(A)** Pod yield. **(B)** Seed yield.

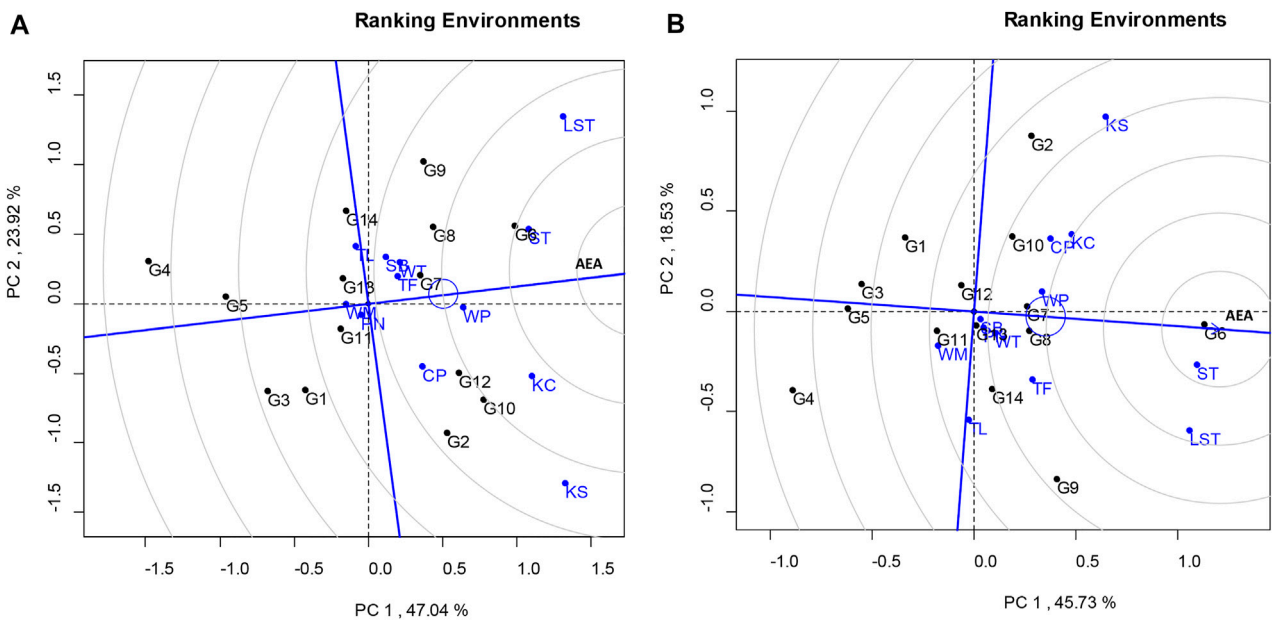


FIGURE 2 | GGE biplot based on genotype-focused scaling comparing tested genotypes with ideal genotypes. **(A)** Pod yield. **(B)** Seed yield.

the broadly suitable locations for growing a specific line or group of genotypes. These GGE results indicated that the peanut lines KUP12BS029-1-1-3, KUP12BS030-4-2-1, and KUP12BS031-2-4-2 had the highest pod yields, whereas KUP12BS030-1-4-3 was the most stable peanut line. These results revealed that the line with the highest pod yield was not always the most stable line at each location. Thus, some large-seeded peanut lines showed high stability, but not all highly stable genotypes had a high mean pod and seed yield. Stability is only significant to farmers when this trait is associated with high mean performance (Jompuk, 2016).

Reviewing the biplot graph, KUP12BS029-1-1-3 had the highest yield at LST, ST, WP, TF, and WT, but KUP12BS029-1-1-3 had the highest yield at Lam Sonthi (LST) and Si Thep (ST) (**Table 4**). These results indicate that considering the multilocation data analysis using the GGE biplot method was more reliable than considering the data collected at each location (Yan, 2002). Lam Sonthi (LST) had

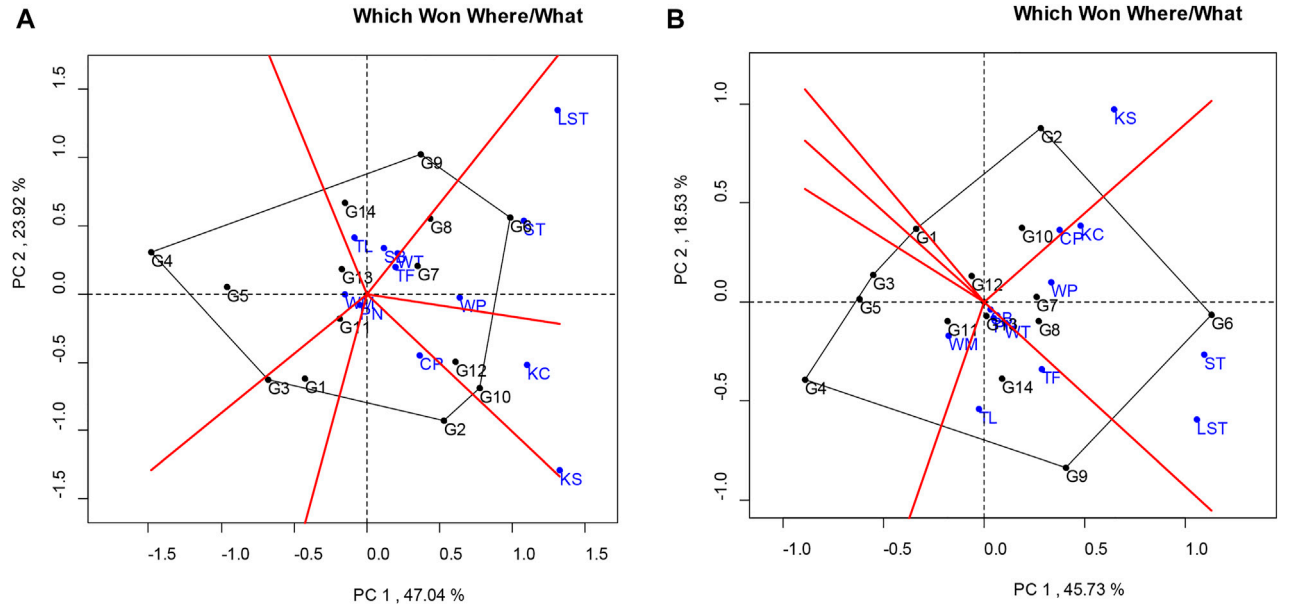


FIGURE 3 | Polygon views of the GGE biplot of the "which-won-where/what" pattern of genotypes and locations. **(A)** Pod yield. **(B)** Seed yield.

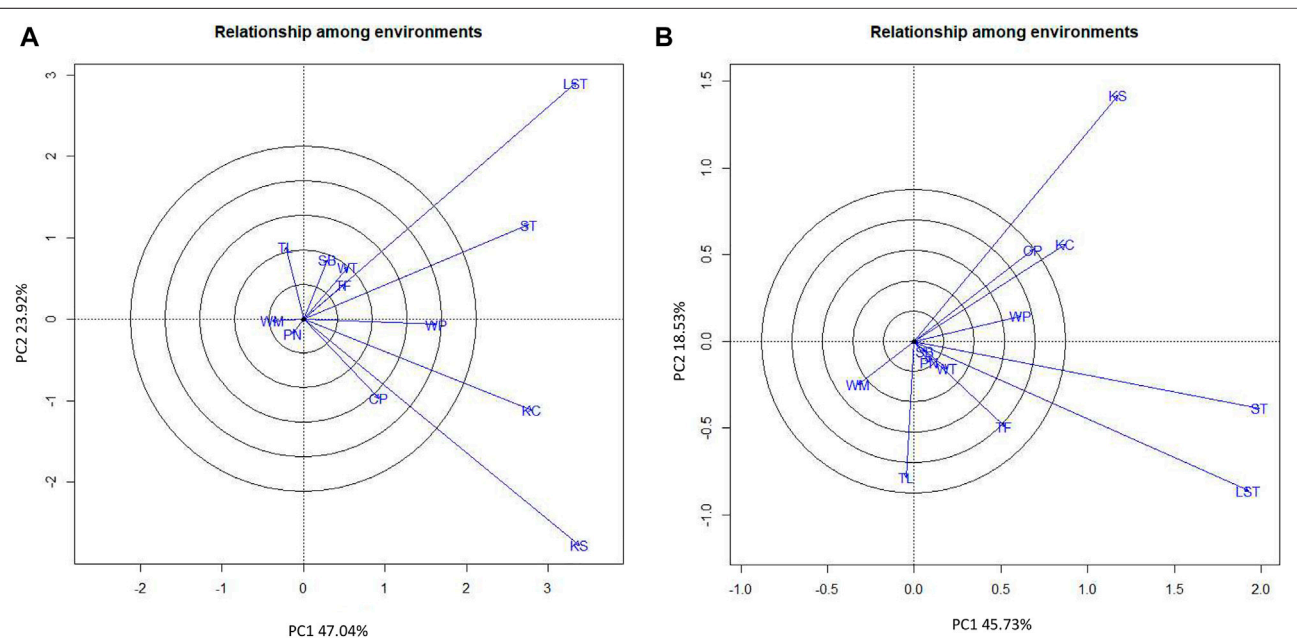


FIGURE 4 | GGE biplot showing discriminating ability and representativeness of test locations. **(A)** Pod yield. **(B)** Seed yield.

long vectors and was the most discriminating for pod and seed yields. The nonrepresentative location can be considered a suitable testing location for selecting specifically adapted genotypes. However, the other locations with short vectors were not suitable for identifying peanut lines with high pod or seed yield (Yan and Tinker, 2006). For regional yield trials, the GGE biplot is an excellent analytical data tool for identifying the best genotypes at each location and the most stable genotypes for production. Moreover, this study reveals that the genetic and environment interaction (G×E) is the major barrier for increasing the yield of peanuts. According to the high effect of G×E, it is not easy to breed a new variety for use at different locations. For increasing the yield of large-seed peanuts, the breeder should focus on breeding a specific peanut variety for large-seed and high yields in a specific area.

CONCLUSION

This study shows the variability of yields in large-seeded peanut lines at multiple locations. Locational variations influenced yield more than variations of GEI and G. For the testing locations, Si Thep (ST) was both discriminating and a representative location that can be considered a suitable testing location for selecting genotypes. The KUP12BS029-1-1-3 line presented a high yield at multilocations. Therefore, the KUP12BS029-1-1-3 line is the large-seeded peanut genotype that is most suitable and should be introduced to farmers because its genotype shows high yield and high stability at multiple locations. These promising large-seeded peanut lines are released as new peanut varieties and used as parental lines in breeding programs for large-seeded yields in Thailand and Southeast Asia.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TC, NP, and JA concept design, coordination of research work, data analysis, and manuscript writing; SC, SR, NP, and PK experimented in planting areas and collected data. All authors read and approved the manuscript.

FUNDING

This study was supported by the Kasetsart University Research and Development Institute (KURDI), Bangkok, Thailand, and Agricultural Research Development Agency (Public Organization).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.876763/full#supplementary-material

REFERENCES

Alwala, S., Kwolek, T., McPherson, M., Pellow, J., and Meyer, D. (2010). A Comprehensive Comparison between Eberhart and Russell Joint Regression and GGE Biplot Analyses to Identify Stable and High Yielding Maize Hybrids. *Field Crops Res.* 119 (2-3), 225–230. doi:10.1016/j.fcr.2010.07.010

Annicchiarico, P. (1997). Joint Regression vs. AMMI Analysis of Genotype-Environment Interactions for Cereals in Italy. *Euphytica* 94, 53–62. doi:10.1023/A:1002954824178

Authapun, J., Lertsuchatavanich, U., Rungmekarat, S., Rajchanuwong, P., Promchote, P., Jindaluang, W., et al. (2016). Yield Stability Analysis of Multi-Environment Yield Trials in Peanut Breeding." in Proceedings of the 54th Kasetsart University Annual Conference. Bankok

Banterng, P., Patanothai, A., Pannangpetch, K., Jogloy, S., and Hoogenboom, G. (2006). Yield Stability Evaluation of Peanut Lines: A Comparison of an Experimental versus a Simulation Approach. *Field Crops Res.* 96 (1), 168–175. doi:10.1016/j.fcr.2005.06.008

Copeland, S. C., Balota, M., and Isleib, T. G. (2013). "Interaction between Genotypes and the NCSU Advanced Yield Test and Multistate Peanut Variety and Quality Evaluation Testing Program," in *Proceedings Of the 45th APRES Annual Meeting*. Editor S. K. TubbsCutchins (College, Georgia: Young Harris), 32.

De lacy, I. H. (1981). "Analysis and Interpretation of Pattern of Response in Regional Variety Trials," in *Interpretation of Plant Response and Adaptation to Agricultural Environments*. Editor V. E. DE BythMontgomery (Brisbane: University of Queensland), 27–50.

Eberhart, S. A., and Russell, W. A. (1966). Stability Parameters for Comparing Varieties 1. *Crop Sci.* 6 (1), 36–40. doi:10.2135/cropsci1966.0011183X000600010011x

Finlay, K., and Wilkinson, G. (1963). The Analysis of Adaptation in a Plant-Breeding Programme. *Aust. J. Agric. Res.* 14 (6), 742–754. doi:10.1071/AR9630742

Gauch, H. G., and Zobel, R. W. (1988). Predictive and Postdictive Success of Statistical Analyses of Yield Trials. *Theor. Appl. Genet.* 76 (1), 1–10. doi:10.1007/BF00288824

Haruthaithanasan, W. (2002). *Yield Stability and Seed Size of Large-Seeded Peanuts*. Bangkok: Department of Product Development Faculty of Agro-Industry, Kasetsart University.

Jalata, Z. (2010). GGE-Biplot Analysis of Multi-Environment Yield Trials of Barley (*Hordeium Vulgare* L.) Genotypes in Southeastern Ethiopia Highlands. *Int. J. Plant Breed. Genet.* 5 (1), 59–75. doi:10.3923/ijpbg.2011.59.75

Jompuk, C. (2016). *Methods of Analysis in Quantity Genetics in Plant Breeding*. Bangkok. Kasetsart University.

Kasno, A., and Trustinah (2015). Genotype-environment Interaction Analysis of Peanut in Indonesia. *Sabrao J. Breed. Genet.* 47 (4), 482–492. http://sabraojournal.org/wp-content/uploads/2018/01/SABRAO-J-Breed-Genet-474-482-492-Kasno.pdf.

Matus-Cádiz, M. A., Hucl, P., Perron, C. E., and Tyler, R. T. (2003). Genotype × Environment Interaction for Grain Color in Hard White Spring Wheat. *Crop Sci.* 43 (1), 219–226. doi:10.2135/cropscience2003

Nandini, C., and Savithramma, D. L. (2012). Character Association and Path Analysis in F8 Recombinant Inbred Line Population of the Cross NRCG 12568 9 NRCG 12326 in Groundnut (*Arachis hypogaea* L.). *Asian J. Bio Sci.* 7 (1), 55

Nigam, S. N., Dwivedi, S. L., and Gibbons, R. W. (1991). Groundnut Breeding: Constraints, Achievements and Future Possibilities. *Plant Breed. Abstr.* 61, 1127

Okori, P., Charlie, H., Mwololo, J., Munthali, W., Kachulu, L., Monyo, E., et al. (2019). Genotype-by-environment Interactions for Grain Yield of Valencia Groundnut Genotypes in East and Southern Africa. *Aust. J. Crop Sci.* 13 (12), 2030–2037. doi:10.21475/ajcs.19.13.12.p2039

Oliveira, E. J., and Godoy, I. J. (2006). Pod Yield Stability Analysis of Runner Peanut Lines Using AMMI. *Cbab* 6, 310–317. doi:10.12702/1984-7033.v06n04a09

Patil, A. S., Hedvat, I., Levy, Y., Galili, S., and Hovav, R. (2018). Genotype-by-environment Effects on the Performance of Recombinant Inbred Lines of Virginia-type Peanut. *Euphytica* 214, 83. doi:10.1007/s10681-018-2159-6

Patil, A. S., Nandanwar, H. R., Punewar, A. A., and Shah, K. P. (2014). Stability for Yield and its Component Traits in Groundnut (Arachis hypogaeaL.). *Inter. Jour. Bio-reso. Stress Manag.* 5 (2), 240–245. doi:10.5958/0976-4038.2014.00562.4

Pham, H. N., and Kang, M. S. (1988). Interrelationships Among and Repeatability of Several Stability Statistics Estimated from International Maize Trials. *Crop Sci.* 28 (6), 925–928. doi:10.2135/cropsci1988.0011183x002800060010x

Rathnakumar, A. L., Manohar, S. S., Nadaf, H. L., Patil, S. C., Deshmukh, M. P., Thirumalaisamy, P. P., et al. (2020). G × E Interactions in QTL Introgression Lines of Spanish-type Groundnut (*Arachis hypogaea* L.). *Euphytica* 216, 85. doi:10.1007/s10681-020-02613-x

Rocha, M. d. M., Damasceno-Silva, K. J., Menezes-Júnior, J. Â. N. d., Carvalho, H. W. L. d., Costa, A. F. d., Lima, J. M. P. d., et al. (2017). Yield Adaptability and Stability of Semi-erect Cowpea Genotypes in the Northeast Region of Brazil by REML/BLUP. *Rev. Cienc. Agron.* 48 (5), 862–871. doi:10.5935/1806-6690.20170102

Santos, R. C. d., Silva, A. F., Gondim, T. M. S., Oliveira Júnior, J. O. L. d., Araújo Neto, R. B. d., Sagrilo, E., et al. (2012). Stability and Adaptability of Runner Peanut Genotypes Based on Nonlinear Regression and AMMI Analysis. *Pesq. Agropec. Bras.* 47 (8), 1118–1124. doi:10.1590/S0100-204X2012000800012

Shorter, R., and Norman, R. (1983). Cultivar Cross Environment Interactions for Kernel Yield in Virginia Type Peanuts (*Arachis hypogaea* L.) in Queensland. *Aust. J. Agric. Res.* 34 (4), 415–426. doi:10.1071/ar9830415

Tai, P. Y. P., and Hammons, R. O. (1978). Genotype-Environment Interaction Effects in Peanut Variety Evaluation1. *Peanut Sci.* 5 (2), 72–74. doi:10.3146/i0095-3679-5-2-4

Thirumala, R. V., Venkanna, V., Bhadru, D., and Bharathi, D. (2014). Studies on Variability, Character Association and Path Analysis on Groundnut (*Arachis hypogaea* L.). *Int. J. Pure Appl. Biosci.* 2 (2), 194

Waranyawat, A. (1999). *Oil Crops, Soybeans, Peanuts, castor*. 1 ed. Bangkok: Department of Agronomy. Faculty of Agriculture, Kasetsart University.

Yan, W., and Hunt, L. A. (2001). Interpretation of Genotype × Environment Interaction for Winter Wheat Yield in Ontario. *Crop Sci.* 41 (1), 19–25. doi:10.2135/cropsci2001.41119x

Yan, W., Hunt, L. A., Sheng, Q., and Szlavnics, Z. (2000). Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Sci.* 40 (3), 597–605. doi:10.2135/cropsci2000.403597x

Yan, W. (2002). Singular-Value Partitioning in Biplot Analysis of Multienvironment Trial Data. *Agron. J.* 94 (5), 990–996. doi:10.2134/agronj2002.990010.2134/agronj2002.0990

Yan, W., and Tinker, N. A. (2006). Biplot Analysis of Multi-Environment Trial Data: Principles and Applications. *Can. J. Plant Sci.* 86, 623–645. doi:10.4141/P05-169

Zurweller, B. A., Xavier, A., Tillman, B. L., Mahan, J. R., Payton, P. R., Puppala, N., et al. (2018). Pod Yield Performance and Stability of Peanut Genotypes under Differing Soil Water and Regional Conditions. *J. Crop Improv.* 32 (4), 532–551. doi:10.1080/15427528.2018.1458674

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 © 2022 Pobkhunthod, Authapun, Chotchutima, Rungmekarat, Kittipadakul, Duangpatra and Chaisan. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 05 August 2022
doi: 10.3389/fgene.2022.924287

Development of High Yielding Fusarium Wilt Resistant Cultivar by Pyramiding of "Genes" Through Marker-Assisted Backcrossing in Chickpea (*Cicer arietinum* L.)

C. Bharadwaj[1†], J. Jorben[1†], Apoorva Rao[1], Manish Roorkiwal[2,3], B. S. Patil[1], Jayalakshmi[4], S. Khayum Ahammed[4], D. R. Saxena[5], M. Yasin[5], J. E. Jahagirdar[6], P. L. Sontakke[6], M. S. Pithia[7], M. K. Chudasama[7], Indu Swarup[8], R. K. Singh[8], S. D. Nitesh[9], Annapurna Chitikineni[2,10], Sarvjeet Singh[11], Inderjit Singh[11], Aditya Pratap[12], G. P. Dixit[12], A. K. Srivastava[12] and Rajeev K. Varshney[2,10]**

[1]Division of Genetics, ICAR- Indian Agricultural Research Institute, New Delhi, India, [2]Center of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [3]Genomic Breeding Lead, Khalifa Center for Genetic Engineering and Biotechnology (KCGEB), UAE University, Al Ain, United Arab Emirates, [4]Regional Agricultural Research Station, Kurnool, India, [5]RAK College of Agriculture, RVSKVV, Sehore, India, [6]Agricultural Research Station, Badnapur, India, [7]Pulse Research Station, Gujarat Agriculture University, Junagadh, India, [8]AICRIP on Chickpea, College of Agriculture Indore, Indore, India, [9]Department of Plant Breeding and Genetics, CSAUAT, Kanpur, India, [10]State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, WA, Australia, [11]Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India, [12]Crop Protection Division, Indian Institute of Pulse Research, Kanpur, India

Pusa 391, a mega *desi* chickpea variety with medium maturity duration is extensively cultivated in the Central Zone of India. Of late, this variety has become susceptible to Fusarium wilt (FW), which has drastic impact on its yield. Presence of variability in the wilt causing pathogen, *Fusarium oxysporum* f.sp. *ciceri* (*foc*) across geographical locations necessitates the role of pyramiding for FW resistance for different races (*foc* 1,2,3,4 and 5). Subsequently, the introgression lines developed in Pusa 391 genetic background were subjected to foreground selection using three SSR markers (GA16, TA 27 and TA 96) while 48 SSR markers uniformly distributed on all chromosomes, were used for background selection to observe the recovery of recurrent parent genome (RPG). BC_1F_1 lines with 75–85% RPG recovery were used to generate BC_2F_1. The plants that showed more than 90% RPG recovery in BC_2F_1 were used for generating BC_3F_1. The plants that showed more than 96% RPG recovery were selected and selfed to generate BC_3F_3. Multi-location evaluation of advanced introgression lines (BC_2F_3) in six locations for grain yield (kg/ha), days to fifty percent flowering, days to maturity, 100 seed weight and disease incidence was done. In case of disease incidence, the genotype IL1 (BGM 20211) was highly resistant to FW in Junagarh, Indore, New Delhi, Badnapur and moderately resistant at Sehore and Nandyal. GGE biplot analysis revealed that IL1(BGM20211) was the most stable genotype at Junagadh, Sehore and Nandyal. GGE biplot analysis revealed that IL1(BGM 20211) and IL4(BGM 20212) were the top performers in yield and highly stable across six environments and were nominated for Advanced Varietal Trials (AVT) of AICRP

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Afroz Alam,
Banasthali University, India
Rafiul Amin Laskar,
Bahona College, India

***Correspondence:**
C. Bharadwaj
drchbharadwaj@gmail.com
Rajeev K. Varshney
Rajeev.Varshney@murdoch.edu.au

[†]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: *20 April 2022*
Accepted: *17 June 2022*
Published: *05 August 2022*

Citation:
Bharadwaj C, Jorben J, Rao A, Roorkiwal M, Patil BS, Jayalakshmi, Ahammed SK, Saxena DR, Yasin M, Jahagirdar JE, Sontakke PL, Pithia MS, Chudasama MK, Swarup I, Singh RK, Nitesh SD, Chitikineni A, Singh S, Singh I, Pratap A, Dixit GP, Srivastava AK and Varshney RK (2022) Development of High Yielding Fusarium Wilt Resistant Cultivar by Pyramiding of "Genes" Through Marker-Assisted Backcrossing in Chickpea (Cicer arietinum L.). Front. Genet. 13:924287. doi: 10.3389/fgene.2022.924287

(All India Coordinated Research Project on Chickpea) in 2018–19. BGM20211 and BGM 20212 recorded 29 and 28.5% average yield gain over the recurrent parent Pusa 391, in the AVT-1 and AVT-2 over five environments. Thus, BGM20211 was identified for release and notified as Pusa Manav/Pusa Chickpea 20211 for Madhya Pradesh, Gujarat and Maharashtra, Southern Rajasthan, Bundhelkhand region of Uttar Pradesh states by the Central Sub-Committees on Crop Standards, Notification and Release of Varieties of Agricultural Crops, Ministry of Agriculture and Farmers Welfare, Government of India, for commercial cultivation in India (Gazette notification number S.O.500 (E) dt. 29-1-2021).Such pyramided lines give resilience to multiple races of fusarium wilt with added yield advantage.

Keywords: Fusarium Wilt, MABC, Pusa 391, GGE biplot analysis, recurrent parent genome recovery

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is a rich source of nutrition and is ranked second amongst food legumes after common bean (Bharadwaj et al., 2010). It is a self-pollinated diploid crop with genome size 740 Mbp (Varshney et al., 2013), 2n = 2x = 16 and is grown extensively in about 57 countries under varied environmental conditions. Globally it is grown in an area of 13.72 million hectares (M ha) with an annual production of 14.25 million tons (MT) (Faostat, 2020). South and South-East Asia dominate in chickpea production contributing 80% of global contribution. The largest share of chickpea production (65%, 9.0 MT) is by India followed by Australia (14%) (Merga and Haji, 2019). To attain self-sufficiency by 2050, the total pulse production in the country needs to reach 39 MT (Vision 2050; IIPR) and amongst all pulses, chickpea production alone needs to reach about 16–17.5 MT from a limited area of about 10.5 m ha with an average productivity of 15–17 q/ha (Dixit et al., 2019).

The productivity of chickpea has progressively increased from 1961, although its vulnerability to biotic and abiotic stresses has also steadily decreased, mostly because of the repeated cultivation of limited number of cultivars and use of only a few prominent donor parents in breeding programmes (Muehlbauer and Sarker, 2017). In case of biotic stresses, wilt caused by *Fusarium oxysporum* (Schlechtend.: Fr.) f. Sp. *ciceris* (Padwick) Matuo and K. Sato is a serious problem in most of the chickpea growing regions of India. The pathogen is a soil-borne, facultative fungus and leads to vascular wilting leading to an average annual yield loss of 10–30%, which can sometimes even cause complete yield loss (Sunkad et al., 2019).

Multi environmental trials are less effective to select stable genotypes across locations, especially under the effect of genotype X environment interaction (G x E x I), leading to the development of wide and specifically adaptable lines. This understanding is even more important owning to race complexity of FW in multiple-environmental trails. Management strategy for FW using chemical formulations as well as the biological control techniques is challenging as the pathogen is harbored in seed and soil. Also, cost involvement and the hazardous nature of chemicals make their use ineffective. Therefore, there is a need for cost effective management strategies which involve developing wilt resistant lines that are widely adapted (Haware and Nene, 1982; Sharma et al., 2005). Conventional breeding methodology takes a long time and is considered less effective to pyramid multiple genes conferring resistance against various races of same pathogen in a single variety. The availability of sufficient genomic information in terms of the physical map, genetic, and draft sequencing of *kabuli* chickpea genotype CDC Frontier has improved the marker-assisted backcrossing (MABC) breeding approach. Mapping studies revealed that all *foc* resistance genes (1, 3,4 and 5) are present on LG02, and tightly linked markers to these races have been identified (Millan et al., 2006; Gowda et al., 2009). Successful application of the MABC approach in chickpea for introgressing resistance/tolerance to drought (Bharadwaj et al., 2021), FW (Varshney et al., 2014a; Pratap et al., 2017), and Ascochyta blight (Varshney et al., 2014b) into popular chickpea varieties have immensely proved the application of MABC in developing elite varieties. Thus, MABC can quickly aid in developing wilt-resistant varieties and pyramid numerous genes in a single introgression line by using foreground and background selection using genome-wide SSR markers (Bharadwaj et al., 2021).This will help in developing multi-race resistant introgression lines that can be widely cultivated across the country.

Pusa 391, a mega variety that is widely grown in central India for its excellent grain quality, has become susceptible to FW since past few years and subsequently its yield had reduced drastically. Thus, in the present study, pyramiding of multiple races against *foc* 1,2,3,4 and 5 was undertaken at ICAR-Indian Agricultural Research Institute (IARI) in collaboration with International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). We also report the release of Pusa Chickpea 20211, a variety developed by molecular breeding, which is highly resistant to FW and is superior to national check (JG 16).

MATERIALS AND METHODS

Pusa 391, a high yielding *desi* chickpea variety with high grain quality and market preference, was released in 1997 for Central zone of India, was chosen as the recurrent parent for

TABLE 1 | Mean yield- Yield performance (Kg/ha), days to Fifty percent Flowering (DFF), days to maturity (DM), 100 Seed Weight (HSW) and disease Index (DI) of chickpea at six locations during 2017–2018.

Genotype	Plant number	Yield (kg/ha)	DFF(d)	DM(d)	100SW(g)	DI (%)
IL1	P391*WR315-3	2,717.42	56.25	106.58	20.33	8.67
IL2	P391*WR315-5	2,512.25	57.33	108.08	20.23	9.93
IL3	P391*WR315-7	2,691.42	57.08	107.17	20.46	9.33
IL4	P391*WR315-9	2,705.42	57.00	107.08	20.77	8.99
IL5	P391*WR315-10	2,410.92	58.42	108.75	19.62	11.10
IL6	P391*WR315-12	2,385.42	58.92	108.75	19.77	12.18
IL7	P391*WR315-15	2,369.83	59.08	109.25	20.04	10.79
IL8	P391*WR315-17	2,474.50	57.00	109.58	20.33	10.84
IL9	P391*WR315-20	2,414.25	58.00	109.33	19.90	9.77
IL10	P391*WR315-22	2,386.17	58.33	108.83	20.14	12.11
WR 315	—	2,173.58	58.25	110.50	18.03	5.70
Pusa 391	—	1993.92	59.42	112.50	24.81	29.57
JG 16 (National Check)	—	2,274.83	57.17	112.33	18.34	17.26

TABLE 2 | Analysis of joint variance for the trait yield (Per hectare), days to fifty per cent flowering (DFF), days to maturity (DM), 100 seed weight (100SW) and disease incidence (DI) evaluated in 13 genotypes in a different environment (**, significant at $p \leq 0.01$; ***, significant at $p \leq 0.001$.)

		Mean sum of square				
Sources of var	DF	Yield	DFF	DM	100SW	DI
E	5	19,252,199.94***	1808.88***	1,346.41***	113.36***	321.04***
G	12	523,189.506***	11.17***	36.75***	14.19***	6.61***
GE	60	38,466.062 ***	9.11***	15.90***	4.12**	416.05***
RESID	72	6,581.482***	1.92***	1.26***	3.05***	28.66***
CV(%)		3.34	2.40	1.03	8.72	21.52

introgression of FW resistance by using WR 315 as donor parent, which harbors resistant genes *foc*1, 3, 4 and 5. It has a seed size of about 20–25 g per 100 seeds, matures in 110–120 days, and has an average yield of 17–18 Q/ha (http://farmer.gov. in/imagedefault/pestanddiseasescrops/pulses.pdf). Each introgression line was planted in two replications comprising two rows of 4 m length and planted at 30 × 10 cm spacing.

DNA Extraction and Marker Analysis

DNA was isolated from tender leaf tissues of 18–20 days old seedlings of parents, F_1 and backcross generations, using the protocol described by Tapan et al. (2013). A total of six SSR markers, namely TA110, TA37, TR19, GA16, TA27, and TA96 reported to be in the cluster containing genes for conferring FW resistance on the linkage group CaLG02 (Millan et al., 2006) were subjected to parental polymorphism to identify polymorphic markers. PCR was carried out in the Chickpea Molecular Breeding Laboratory, Division of Genetics, ICAR-IARI using a G-STORM thermal cycler (Labtech, France). The PCR amplicons were resolved on a 1.2% agarose gel or with an ABI 3730 (Applied Biosystems). Polymorphic markers with donor and recipient cross-combinations from the hotspot region of the LG02 were used for foreground selection (**Supplementary Table S1**).

Based on previous studies, a panel of 365 highly polymorphic SSR markers (Thudi et al., 2011; Bharadwaj et al., 2021) were tested at the Centre of Excellence in Genomics and Systems Biology (CEGSB), ICRISAT for parental polymorphism between the donor and recurrent parents for potential use in background selection. Background selection was based on polymorphic markers identified for each donor and recipient parent cross-combination (**Supplementary Table S1**). Recurrent parent genome (RPG) recovery was calculated for selection using the formula suggested by Sundaram et al. (2008).

Phenotypic Screening for Fusarium Wilt Resistance at Multiple Locations

Ten advanced MABC lines (BC_3F_3 progenies) were sown in wilt sick plot, along with parent Pusa 391, resistant check (WR 315), susceptible check (JG 62) and National check (JG 16) in RCBD (randomized complete block design) with two replications in the years 2017–18 across six diverse locations viz., Indore (22.7196° N, 75.8577° E), Badnapur (19.8682° N, 75.7256° E), Junagadh (21.52° N, 70.45° E) Nandyal (15.47° N, 78.48° E), Sehore (23.2032° N, 77.0844° E) and IARI-New Delhi (28°70′N, 76°58′E) during 2017-18for morphological characters such as yield (per hectare), days to 50% flowering (DFF), days to maturity (DM), and hundred seed weight (HSW) (**Table 1**). There was sufficient inoculum load (spore concentration of 5–6 × 10^6 conidia/ml/g of soil) in the wilt sick plot as manifested by the 100% mortality of susceptible check JG 62. Visual observations were taken at seedling to flowering stage of crop based on mortality rate as

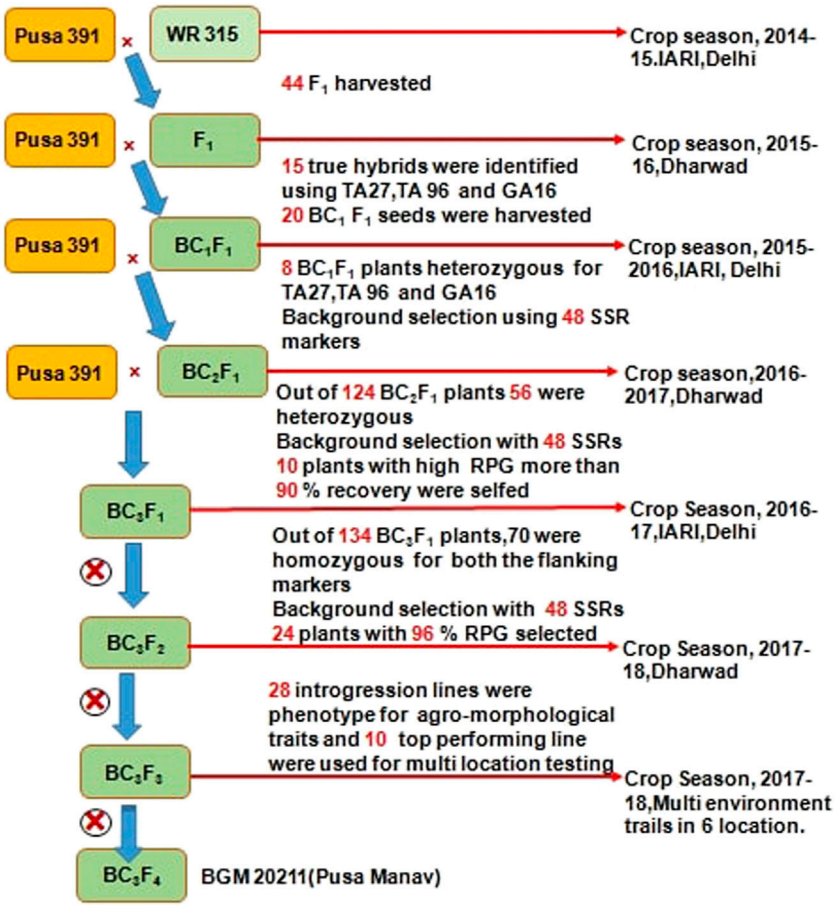


FIGURE 1 | Detailed representation of marker assisted back cross bred lines of Chickpea for Fusarium wilt at ICAR-IARI.

per the classification of Sharma et al. (2019), designated as highly resistant (less than 10% plant mortality), moderately resistant (10.1–20.0% plant mortality), susceptible (20.1–40.0% plant mortality), and highly susceptible (more than 40.0% plant mortality).The superior introgression lines (IL1, IL3 and IL4) having yield superiority and reaction to FW across locations were recommend for AICRP trials. BGM20211 (IL1) and BGM 20212 (IL4) were tested in AVT-1 and AVT-2 along with checks in 2018–2019 and 2019-2020, respectively, for adaptability and yield advantage and BGM 20213 (IL 3) was recommended for AVT-1 in 2019–20.

% Disease incidence calculated as per Sharma et al. (2017) as,

$$\%\ Disease\ incidence = \frac{Total\ number\ of\ infected\ plants}{Total\ number\ of\ plants} \times 100$$

Statistical Analysis

All analyses were performed in the R software. A preliminary variability study was conducted using *TraitStats*:R package, developed by Nitesh et al. (2021), while the GGE Biplots were performed with the metan: R package, developed by Olivoto and Lucio (2020).

RESULTS

Analysis of Variance (ANOVA)

Joint ANOVA results revealed significant differences ($p < 0.05$) for all five traits among the genotypes under the study in each of the environments tested (E1, E2, E3, E4, E5, and E6) (**Table 2**).

Foreground and Background Marker Analysis

Based on parental polymorphism, three markers viz., GA16, TA27 and TA96 were polymorphic and used for foreground selection between Pusa 391 and WR 315. For background selection, among the 365 SSR markers analysed on both parents, 141 were observed to be polymorphic. 48 of these markers that covering the whole genome uniformly, so that each chromosome contains six polymorphic markers distributed evenly were identified for background selection.

Marker-Assisted Introgression

At ICAR-IARI, New Delhi and IARI-Regional Station, Dharwad, marker-assisted backcrossing was carried out during 2014–18. The detailed description used for introgression for FW resistance

TABLE 3 | Descriptive statistics involving Maximum (Max), Mean, Minimum (Min), standard deviation (Sd) and Standard error (Se).

Variable	Max	Mean	Min	Sd	Se
DFF	78	57.87	43	7.95	0.64
DMM	124	109.13	93	7.28	0.58
DI	57	12.02	2.1	7.53	0.6
100 GW (UNIT)	26.9	20.14	14.8	2.79	0.22
Yield (UNIT)	3,902	2,423.82	1,082	822.19	66.04

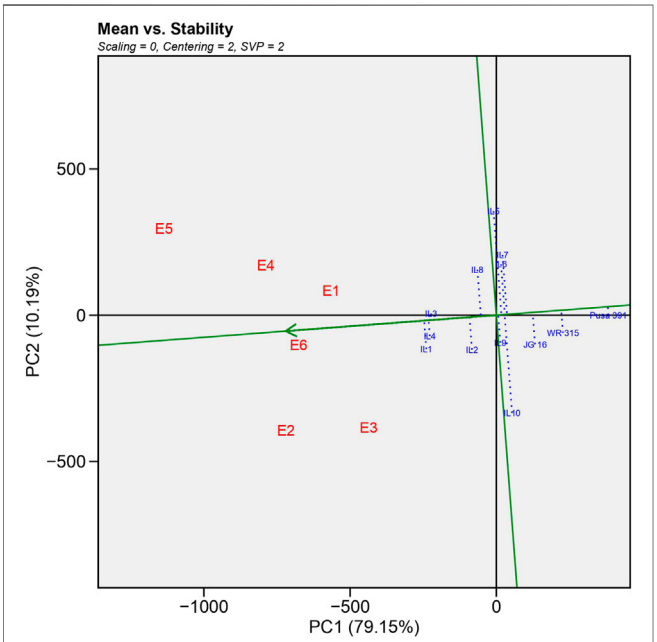


FIGURE 2 | Average-environment coordination (AEC) view showing the mean yield performance and stability of different genotypes based on genotype and genotype interaction (GGE)-biplot analysis.

from WR 315 to Pusa 391 is provided in **Figure 1**. Crosses were carried out between Pusa 391 (recurrent parent) and WR 315 (donor) in crop season 2014-15 to develop F1s. A total of 44 F_1s were obtained, out of which 40 germinated and 15 were confirmed as true hybrids using polymorphic foreground markers (TA27, TA96, GA16). These F_1s were utilized to make backcross with Pusa 391, and 20 BC_1F_1 seeds were generated during the off-season in 2015–16 at Dharwad. Out of 20 BC_1F_1 seeds produced, eight positive plants for foreground marker and had the higher genome recovery of 75–85% were selected for backcrossing to generate 124 BC_2F_1 seeds during the main season IARI, New Delhi in 2015–16. From the 124 BC_2F_1 plants, 56 were heterozygous for foreground markers, of which ten plants with more than 90% genome recovery were used for generation of backcrossing at IARI regional station Dharwad to generate 134 BC_3F_1.Then upon foreground selection, 70 plants were observed to be heterozygous. These 70 plants were subjected for background selection using 48 SSR markers, and 24 BC_3F_1 plants showing more than 96% RPG recovery were selected and selfed. Finally, with two rounds of selfing, 28 best BC_3F_3 plants were analysed with foreground and background markers with 90–97% RPG recovery and showing agronomic superiority were selected. The top ten high yielding, highly disease resistant lines (<10% disease incidence) with plant type similar to Pusa 391 (recurrent parent) were evaluated at multiple locations.

Phenotypic Performance of Pusa 391 MABC Lines in Multi-Location Trails

Introgression lines (BC_3F_3) were phenotyped at six diverse environments, Indore, Badnapur, Junagadh, Nandyal, Sehore and IARI-New Delhi, in the crop season 2017–2018 and transgressive segregants were identified by morphological superiority and with high wilt resistance (**Table 1**). IL1, IL4, and IL3 were the top performers in the multi-location trial with a mean yield advantage of 20, 19 and 18%, respectively, over the national check JG 16 and mean yield gain of 36, 35 and 35% over recurrent parent Pusa 391. Disease incidence index over multiple locations of IL1, IL2, IL3, IL4, and IL 9 were 8.66, 9.93, 9.33, 8.99 and 9.76, respectively, which showed a highly resistant reaction. On the other hand, IL5, IL6, IL7, IL8 and IL10 were 11.1, 12.1, 10.8, 10.8 and 12.1 respectively, which showed moderately resistant reactions.

The hundred seed weight (HSW) of IL1 (BGM 2011) across locations was 20.33 g, IL4 (BGM 20212) was 20.77, and IL3 (BGM 20213) was 20.46, as compared to recurrent parent Pusa391, which recorded 24.81 g IL1 (BGM 20211) had early flowering (56 days) and early maturity (107days) compared to the parent Pusa 391 flowered in 59 days and matured in 112 days. Variation in genotype for different traits studied is presented as descriptive statistics in **Table 3**. IL1 (BGM 20211) and IL4 (BGM 20212) were the top two performers in multilocation trials conducted, and they were nominated for AICRP trials in 2018–19 (AVT-1) and 2019-20 (AVT-2). The next top performer, IL3 (BGM 20213), was nominated for AICRP trials in 2019–20 (AVT-1).

National Advanced Varietal Trial Evaluation

Advanced varietal trial data of BGM 20211 and BGM 20212 indicated that they were early flowering (51 and 53 days, respectively) as well as early maturing (106 and 107 days, respectively) lines. Under ICAR-AICRP on chickpea, special trials were conducted in known as "Wilt resistance introgression lines" (WRIL) trials that can assess lines performance in multiple locations. There was a 23 and 35% overall weighted percentage increase in AVT-1 and AVT-2 respectively with overall mean of 29% over recurrent parent Pusa 391. In the case of IL2 (BGM 20212), the overall weighted percentage increased over the mean of 28.5%, with 24% in AVT1 and 33% in AVT 2, respectively, over the recurrent parent Pusa 391 (**Tables 4**, **5**). Disease incidence in multiple locations as per AICRP trials are presented in **Supplementary Table S2**.

GGE biplot analysis: GGE biplot analysis explained 66.11% of the total variation, where PC1 (wilt incidence) and PC2 (resistance stability) accounted for 51.64 and 14.47% variation, respectively.

TABLE 4 | AVT-1 Comparison of yield (kg ha−1) performance of BGM 20211 and BGM 20212 introgression line developed in the genetic background of Pusa 391, with the recurrent genotype at six locations in the Advanced Varietal Trials−1 of ICAR–All India Coordinated Research Project on Chickpea conducted during 2018–2019 (Source: AICRP Chickpea Annual Report 2018–2019).

Entry	Indore	Junagadh	Sehore	Badnapur	Nandyal	Mean	Freq[a]
BGM 20211	2078	3,526	1,671	1,584	1,131	1998	5/5
BGM 20212	2,156	3,718	1,659	1,437	1,121	2018.2	5/5
Pusa 391(Recurrent parent	1,648	3,064	1,123	1,337	961	1,626.6	—
Critical difference at 5%	75	370	114	188	301	—	—
CV(%)	2.7	7.9	5.9	9.5	16.7	—	—
General mean (kg/ha)	1921	3,197	1,319	1,358	1,237	1806.4	—
state avg. yield (kg/ha)	1,165	1,244	1,165	782	1,051	1,081.4	—

[a]*Frequency, the ratio of a number of locations in which the introgression line performs higher than the recurrent parent to the total number of locations evaluated.*

TABLE 5 | AVT-2 Comparison of yield (kg ha−1) performance of BGM 20211 and BGM 20212 introgression line developed in the genetic background of Pusa 391, with the recurrent genotype at six locations in the Advanced Varietal Trials−2 of ICAR–All India Coordinated Research Project on Chickpea conducted during 2019–2020 (Source: AICRP Chickpea Annual Report 2019–2020).

Entry	Indore	Junagadh	Sehore	Badnapur	Nandyal	Mean	Freq[a]
BGM 20211	2000	3,915	2,182	2,343	1,428	2,373.6	5/5
BGM 20212	2,162	3,699	2,148	2,136	1,539	2,336.8	5/5
Pusa 391	1,643	3,286	1,530	1,494	826	1755.8	—
Critical difference at 5%	368	598	142	473	263	—	—
CV(%)	13.7	11.8	6	16.7	12	—	—
General mean (kg/ha)	1854	3,507	1,649	1963	1,513	2097.2	—
State avg. Yield (kg/ha)	1,165	1,244	1,165	782	1,051	1,081.4	—

[a]*Frequency, the ratio of number of locations in which the introgression line performs higher than the recurrent parent to the total number of locations evaluated.*

1) Mean vs. Stability

As illustrated in **Figure 2**, GGE biplot analysis explained a total of 89.34% variation, the horizontal axis (PC1) accounted for 79.15% of the total variation. It represented the main effect of genotypes, whereas the vertical axis (PC2) accounted for 10.19% of total variation and showed the impact of G X E interaction (Yan and Kang, 2002). The average environment coordinate (AEC) axis is a single arrow line passed from the biplot origin to the average environment, depicted by a dotted circle. On the vertical axis, ILs located to the right of the AEC indicated higher yield than average yield and vice versa. Thus, the biplot organized the yield performance as IL1>IL4>IL3>IL2>IL8 in that order. However, the recurrent parent Pusa 391 and national check JG 16 showed a lower yield than ILs. The AEC vertical axis displayed the stability of genotype yield, which was considered stronger if the horizontal AEC axis line length was shorter (Lakew et al., 2014; Harish et al., 2020). The IL1, IL4 and IL3 were the most stable and high yielding as they were farthest from origin and shortest vector length.

2) Which-Won-Where pattern analysis

The polygon view was generated by interconnecting the markers of the ILs that were farthest from the biplot origin with straight lines, resulting in markers of all cultivars being contained in the polygon (**Figure 3**). To divide the biplot into various sectors, lines perpendicular to each side of the polygon or their extensions were drawn from the biplot origin. The peak

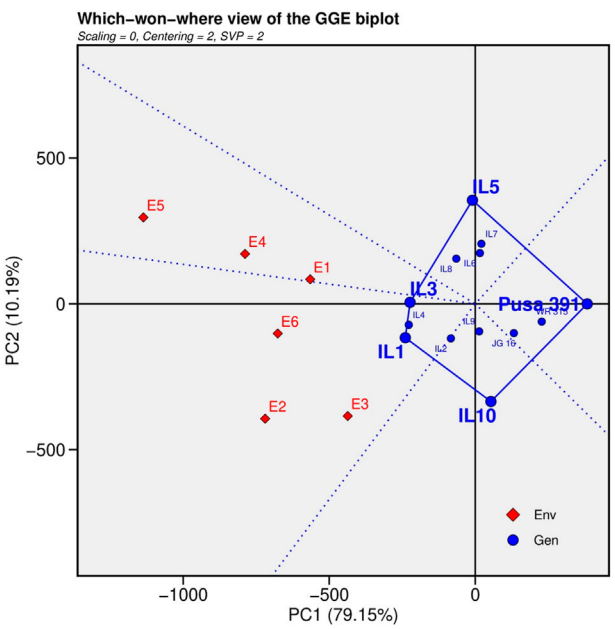


FIGURE 3 | "Which-won-where" view for the primary component of interaction (PC1) and second principal component (PC 2).

cultivar in each sector was the top cultivar for traits found in that section; on the other hand, genotypes found inside the polygon and near the biplot's origin were not sensitive to changing

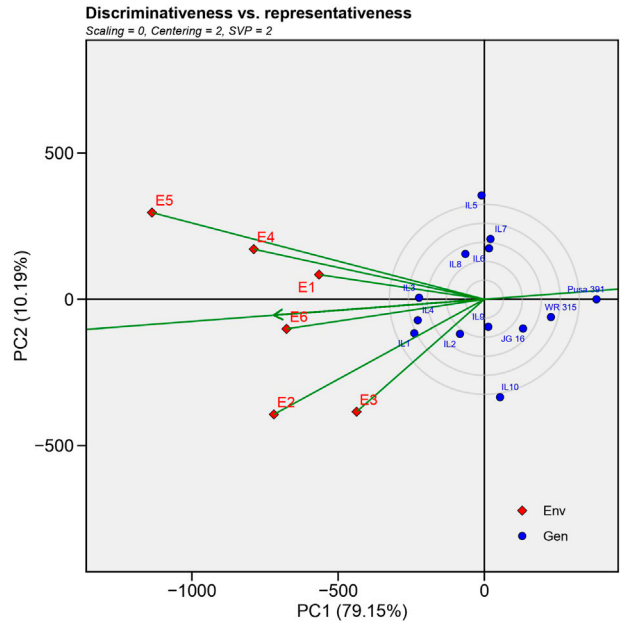


FIGURE 4 | GGE-biplot environment view for yield that shows the correlation between test environments and correlation coefficient between any two environments is approximated by the cosine angle between their vectors.

environmental conditions (Dimitrios et al., 2008).The genotypes positioned at the greatest distance from the biplot origin were the best or worst ILs in particular or every environment. IL1 performed superior under E2 (Junagadh), E3 (Sehore) and E6 (Nandyal) and IL3 performed superior in E1 (Indore), E4 (Badnapur) and E5 (New Delhi) from the which won where pattern analysis.

3) Evaluation of testing locations based on discrimination power vs. representativeness

An ideal location needs to be highly distinctive and represent the target location simultaneously (Yan, 2010). The ability of a place to maximise the diversity among potential cultivars in a study is referred to as discrimination (Blanche and Myers, 2006). The ability to represent, reveals that the study included an environment that was indicative of the conditions in the other locations (Mohammadi and Amri, 2012). An ideal environment will identify genotypes with high and stable yield. The small circle in the GGE-biplot display represents a perfect position determined by the mean coordinates of all testing locations (**Figure 4**). There was a positive association between the length of the location vector and the ability to discriminate between locations, but a negative correlation between the angle of the location vector with the ideal location and the location's representativeness of the target environment (Yan, 2010). The observed angle between E1, E4, E5 and E2, E3, E6 was less than 90^0, indicating a positive correlation among environment sets, and similar results can be expected in these regions. Following analysis, it was observed that E5 (New Delhi) >E2>E3>E4>E6>E1 had the longest environmental vectors among the test environments, making it the most "discriminating location" with the potential to distinguish different genotypes. The ranking of environments in terms of being the best representative locations was E6>E1>E4>E2,E3, E5 were in the order, and thus E6 (Nandyal) can be considered the most representative environment.

DISCUSSION

Fusarium wilt has been a widely distributed disease that can cause upto 100% yield loss based on varietal susceptibility and changing climatic conditions that have resulted in the shift of large chickpea growing area from cool long Northern India to warm short central and southern India (Patil et al., 2015). The presence of eight physiological races of *foc* (0,1A,1B/C, 2, 3, 4, 5 and 6) has been reported across different countries (Haware and Nene, 1982). FW is prevalent in dry and warm semi-arid tropic (SAT) regions of Asia, Africa and South America (Nene and Sheila, 1996). Race 1 is typical in Central and Peninsular India, race 2 in Northern India and 3 and 4 in Punjab and Haryana (Haware et al., 1992). Also, some of the cultivars are susceptible with time, which could be attributed to variability in wilt incidence and genetic differences among genotypes and genotype x environment interactions (Sharma et al., 2012).

The "5Gs" breeding technique (genome assembly, germplasm characterization, gene function identification, genomic breeding, and gene editing) has recently been proposed for obtaining precision and boosting crop improvement to satisfy future demands for nutritious food (Varshney et al., 2019). MABC using genome-wide SSR markers for foreground and background selection for recovery of recurrent parent genome is an environment-independent, precise, and rapid strategy for developing disease-resistant cultivars (Bharadwaj et al., 2021). Deploying resistant variety is one of the key sustainable strategies that breeders can adopt as it is most effective and environmentally safe for integrated disease management (Sharma et al., 2017).

Genetic inheritance studies reveal that resistance to race 0,1A,2 and 4 are either digenic or trigenic, but races 3 and 5 were monogenic (Tullu et al., 1999; Tekeoglu et al., 2000).Based on several inter- and intra-specific crosses, it was reported that wilt resistance genes *foc*1, *foc*3, *foc* 4 and *foc* 5 (Races 1, 3, 4 and 5) are mapped on two gene clusters, i.e., GA16 and TA96 (*foc* 1 and 4 clusters), TA 96 and TA 27 (*foc* 3 and 5 clusters) (Millan et al., 2006). However, resistance genes *per se* and proteins that were reported to be involved in pathogen defense were localized in between the region or in close vicinity of the gene cluster. Also, resistance loci *ar1* and *ar2a* against Ascochyta blight were localized on LG02 and near *foc* gene clusters. Thus, LG02 is considered a hot spot for pathogen defense (Millan et al., 2006).

The major hindrance in chickpea breeding for FW is variation in pathogen races over multiple locations and their interaction with different weather conditions over space and time (Sharma et al., 2014). Stable high yielding lines with high disease resistance are required to develop widely adaptable varieties. (Srivastava

et al., 2021). Thus, GGE Mean vs. Stability analysis recorded IL1>IL4>IL3 were most stable and high yielding introgression line in the order and the worst performing genotype was Pusa 391 > WR 315 > JG16. Which-won-where analysis revealed IL1 performed best under Junagadh, Sehore and Nandyal region and IL3 performed best in Indore, Badnapur and New Delhi from the which-won-where plot. The New Delhi environment was considered the most discriminating location because this location is subjected to distinct climatic conditions compared to other environments. Discrimination and representativeness analyses reveal that E6 (Nandyal) is the most representative location. The genotypes were highly stable in this location, as was reported by Sharma et al. (2019) while screening for wilt-resistant genotypes in wilt sick plot over ten locations. The top three best performing introgression lines (IL1, IL4 and IL3) with more than 30% yield advantage over recurrent parent Pusa 391 were nominated for AVT trials based on the multi-location studies.

MABC approach was used to pyramid races1, 2, 3,4 and 5 for FW and RPG recovery. In the current study, BC_2F_1 and BC_3F_1 generation achieved 90 and 96% RPG recovery for selected MABC lines in the genetic background of Pusa 391. Similar genome recovery was reported by Mannur et al., 2019and Bharadwaj et al., 2021 in chickpea. Thus, MABC reduces the time taken to develop a variety and such genome recovery is usually possible inBC_4F_1 and more generations in conventional breeding.

The introgression line IL1 (BGM 20211) is highly resistant to FW and moderately resistant to stunt, collar rot, dry root, pod borer and possesses excellent grain, colour and shape, as per AICRP report, 2020. The grain protein content was found to be 18.92%. In the case of 100 SW, parent Pusa 391 had a higher 100 SW than introgression line BGM 20211. Also, it is an early flowering and early maturing IL (57 days to flowering and 107 days to maturity), that can fit in the double cropping system and is ideal for the sustainability of the rice-wheat cropping system (Bharadwaj et al., 2021).Further, it can also escape heat stress at harvest in central India compared to Pusa 391, which matures in 112 days.

In the case of disease incidence, BGM 20211 was highly resistant for FW in Junagadh, Indore, IARI-New Delhi and found moderately resistant at Sehore and Nandyal. BGM 20212 was highly resistant in Junagarh and New Delhi and moderately resistant in Indore, Sehore and Nandyal. BGM 20213 was highly resistant in Junagarh, Indore, New Delhi and found to be moderately resistant at Sehore and Nandyal. National check (JG 16) was moderately resistant in Junagarh, Indore, New Delhi and Nandyal and susceptible in Sehore. Also, recurrent parent Pusa 391 was moderately resistant in Badnapur, susceptible in New Delhi, Nandyal, Junagarh and Indore, highly susceptible in Sehore. Superior performance across locations confirms the molecular basis of pyramiding with morphological and wilt sick studies.

In the case of FW, introgression for *foc*1 (Varshney et al., 2014a), *foc*2 (Pratap et al., 2017) and *foc* 4 (Mannur et al., 2019) were developed on the elite genetic background of C214, Pusa 256 and Annigeri, respectively. Mannur et al., 2019 reported a 125% mean yield advantage of superior introgression line over the recurrent parent JG 74 and Super Annigeri 1 reported a8% mean yield advantage over recurrent parent Annigeri 1. In our study, BGM 20211 outperformed parent Pusa 391 by 29% (average over five regions in AVT-1 and AVT-2) and national check (JG16) by 28% (average over 6 locations in multi-location trials), which include southern India (Nandyal), Central India (Badnapur, Indore and Junagarh) and North India (New Delhi).This variety profusely branches and possesses a large number of pods per unit area and has demonstrated an overall weighted mean yield of 2,186 kg/ha and the highest yield potential reported was 3,915 kg/ha in a wilt stress environment, compared to 1,691 kg/ha in case of recurrent parent Pusa 391.

This is the first report in pulses where FW genes are pyramided in recurrent parent background and released for commercial cultivation using the MABC approach.Marker assisted back cross breeding approach was utilized for pyramiding of FW resistance for different races (*foc* 1,2,3,4 and 5). Foreground selection was performed using three SSR markers (GA16, TA 27 and TA 96) and background selection for recovery (RPG) of recurrent parent genome was done using 48 SSR markers that were uniformly distributed on all chromosomes. Multi-location evaluation of advanced introgression lines (BC_2F_3) was done in six locations for grain yield parameters and wilt screening along with GGE biplot analysis. IL1 (BGM 20211) and IL4 (BGM 20212) were the top performers in yield and were highly stable across all environments and were nominated for Advanced Varietal Trials (AVT) of AICRP (All India Coordinated Research Project on Chickpea). BGM 20211 was identified for release and notified as Pusa Manav for Madhya Pradesh, Gujarat and Maharashtra states by the Central Sub-Committees on Crop Standards, Notification and Release of Varieties of Agricultural Crops, Ministry of Agriculture and Farmers Welfare, Government of India, for commercial cultivation in India. High yielding pyramided lines for FW are important to avoid economic losses and for improving Chickpea production across India.

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 authors.

AUTHOR CONTRIBUTIONS

CB and JJ contributed to the analysis, interpretation of data, prepared the manuscript and conceived the study. AR, BP, JL, SA, DS, MY, JJ, PLS, MP, MC, IS, RS, SN, SS, IS, AP, GD, AS contributed to the coordinated evaluation of material. CB designs the study, reviewed and edited the manuscript. MR, AC provided feedback, and edited the manuscript. CB, SN, MR, RV contributed to the statistical analysis, interpretation of data. CB and RV reviewed and edited the manuscript. All the authors read and approved the manuscript.

FUNDING

The study was funded by ICAR-CRP (Code: 12-143D) collaborative project and the institutional fund from ICAR-IARI.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.924287/full#supplementary-material

REFERENCES

Bharadwaj, C., Chauhan, S. K., Rajguru, G., Srivastava, R., Tara Satyavathi, C., Yadav, S., et al. (2010). Diversity Analysis of Chickpea (*Cicer Arietinum*) Cultivars Using STMS Markers. *Indian J. Agric. Sci.* 80 (11), 947. https://epubs.icar.org.in/index.php/IJAgS/article/view/1720.

Bharadwaj, C., Tripathi, S., Soren, K. R., Thudi, M., Singh, R. K., Sheoran, S., et al. (2021). Introgression of " QTL-hotspot " Region Enhances Drought Tolerance and Grain Yield in Three Elite Chickpea Cultivars. *Plant Genome* 14 (1). doi:10.1002/tpg2.20076

Blanche, S. B., and Myers, G. O. (2006). Identifying Discriminating Locations for Cultivar Selection in Louisiana. *Crop Sci.* 46 (2), 946–949. doi:10.2135/cropsci2005.0279

Dimitrios, B., Christos, G., Jesus, R., and Eva, B. (2008). Separation of Cotton Cultivar Testing Sites Based on Representativeness and Discriminating Ability Using GGE Biplots. *Agron. J.* 100, 1230–1236.

Dixit, G. P., Srivastava, A. K., and Singh, N. P. (2019). Marching towards Self-Sufficiency in Chickpea. *Curr. Sci.* 116 (2), 239–242. doi:10.18520/cs/v116/i2/239-242

Faosatat, (2020). Crops. Retrieved from: https://www.fao.org/faostat/en/#data/QC/visualize

Gowda, S. J. M., Radhika, P., Kadoo, N. Y., Mhase, L. B., and Gupta, V. S. (2009). Molecular Mapping of Wilt Resistance Genes in Chickpea. *Mol. Breed.* 24 (2), 177–183. doi:10.1007/s11032-009-9282-y

Harish, D., Bharadwaj, C., Kumar, Tapan, Patil, B. S., Pal, Madan, Hegde, V. S., et al. (2020). Identification of Stable Drought Tolerant Landraces of Chickpea (*Cicer Arietinum*) under Multiple Environments. *Indian J. Agric. Sci.* 90 (8), 1575–1581.

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 (1-2), 147–154. doi:10.1016/0378-4290(92)90063-F

Haware, M. P., and Nene, Y. L. (1982). Races ofFusarium Oxysporumf. sp.Ciceri. *Plant Dis.* 66 (9), 809–810. doi:10.1094/PD-66-809

Lakew, T., Tariku, S., Alem, T., and Bitew, M. (2014). Agronomic Performances and Stability Analysis of Upland Rice Genotypes in North West Ethiopia. *Int. J. Sci. Res. Publ.* 4 (4), 1–9. https://www.ijsrp.org/research-paper-0414/ijsrp-p2838.pdf.

Mannur, D. M., Babbar, A., Thudi, M., Sabbavarapu, M. M., Roorkiwal, M., Yeri, S. B., 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 (1), 1–13. doi:10.1007/s11032-018-0908-9

Merga, B., and Haji, J. (2019). Economic Importance of Chickpea: Production, Value, and World Trade. *Cogent Food & Agric.* 5 (1), 1615718. doi:10.1080/23311932.2019.1615718

Millan, T., Clarke, H. J., Siddique, K. H. M., Buhariwalla, H. K., Gaur, P. M., Kumar, J., et al. (2006). Chickpea Molecular Breeding: New Tools and Concepts. *Euphytica* 147 (1), 81–103. doi:10.1007/s10681-006-4261-4

Mohammadi, R., and Amri, A. (2012). Analysis of Genotype × Environment Interaction in Rain-Fed Durum Wheat of Iran Using GGE-Biplot and Non-parametric Methods. *Can. J. Plant Sci.* 92 (4), 757–770. doi:10.4141/cjps2011-133

Muehlbauer, F. J., and Sarker, A. (2017). "Economic Importance of Chickpea: Production, Value, and World Trade," in *The Chickpea Genome* (Cham: Springer), 5–12. doi:10.1007/978-3-319-66117-9_2

Nene, Y. L., and Sheila, V. K. (1996). *A World List of Chickpea and Pigeonpea Pathogens*. Patancheruvu: ICRISAT.

Nitesh, S. D., Patroti, P., and Parashuram, S. (2021). *TraitStasts: Statistical Data Analysis for Randomized Block Design Experiments*. R Package version 1.0.2. https://www.niteshgpb.in/TraitStats.pdf.

Olivoto, T., and Lúcio, A. D. C. (2020). Metan: An R Package for Multi-environment Trial Analysis. *Methods Ecol. Evol.* 11 (6), 783–789. doi:10.1111/2041-210x.13384

Patil, M., Om, G., Pawar, M., and Chobe, D. R. (2015). Effect of Culture Media, pH and Temperature on the Mycelial Growth and Sporulation of *Fusarium Oxysporum F. Sp. Ciceris* Isolates of Chickpea from Central Zone of India. *JNKVV Res. J.* 49, 54–60. http://jnkvv.org/PDF/14022016112818Res%20Jour%2049-1.pdf.

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. Genet. Genomics* 292 (6), 1237–1245. doi:10.1007/s00438-017-1343-z

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. *Fusarium oxysporum F. Sp. cicerisPlant Dis.* 89 (4), 385–390. doi:10.1094/PD-89-0385

Sharma, M., Ghosh, R., Tarafdar, A., Rathore, A., Chobe, D. R., Kumar, A. V., et al. (2019). Exploring the Genetic Cipher of Chickpea (*Cicer Arietinum L.*) through Identification and Multi-Environment Validation of Resistant Sources against Fusarium Wilt (*Fusarium Oxysporum* F. Sp. Ciceris). *Front. Sustain. Food Syst.* 3, 78. doi:10.3389/fsufs.2019.00078

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. Sp. Ciceris* in Chickpea. *Field Crops 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 (1), 454–514. doi:10.1186/1471-2164-15-454

Sharma, M., Tarafdar, A., Ghosh, R., and Gopalakrishanan, S. (2017). "Biological Control as a Tool for Eco-Friendly Management of Plant Pathogens," in *Advances in Soil Microbiology: Recent Trends and Future Prospects* (Singapore: Springer), 153–188. doi:10.1007/978-981-10-7380-9_8

Srivastava, A. K., Dixit, G. P., Dixit, G. P., Singh, N. P., Saxena, D. R., Saabale, P. R., et al. (2021). Evaluation of Chickpea (*Cicer Arietinum* L.) Accessions for Fusarium Wilt Resistance. *Jeb* 42 (1), 82–89. doi:10.22438/jeb/42/1/mrn-1475

Sundaram, R. M., Vishnupriya, M. R., Biradar, S. K., Laha, G. S., Reddy, G. A., Rani, N. S., et al. (2008). Marker Assisted Introgression of Bacterial Blight Resistance in Samba Mahsuri, an Elite Indica Rice Variety. *Euphytica* 160 (3), 411–422. doi:10.1007/s10681-007-9564-6

Sunkad, G., Deepa, H., Shruthi, T. H., and Singh, D. (2019). Chickpea Wilt: Status, Diagnostics and Management. *Indian Phytopathol.* 72 (4), 619–627. doi:10.1007/s42360-019-00154-5

Tapan, K., Bharadwaj, C., Satyavathi, C. T., and Jain, P. K. (2013). A High Throughput, Improved Rapid and Reliable Genomic DNA Extraction Protocol from Chickpea (*Cicer Arietinum* L.). *Vegetos* 26 (2), 185–190. doi:10.5958/j.2229-4473.26.2.073

Tekeoglu, M., Santra, D. K., Kaiser, W. J., and Muehlbauer, F. J. (2000). Ascochyta Blight Resistance Inheritance in Three Chickpea Recombinant Inbred Line Populations. *Crop Sci.* 40 (5), 1251–1256. doi:10.2135/cropsci2000.4051251x

Thudi, M., Bohra, A., Nayak, S. N., Varghese, N., Shah, T. M., Penmetsa, R. V., et al. (2011). Novel SSR Markers from BAC-End Sequences, DArT Arrays and a Comprehensive Genetic Map with 1,291 Marker Loci for Chickpea (*Cicer Arietinum* L.). *PLoS One* 6 (11), e27275. doi:10.1371/journal.pone.0027275

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 (1), 43–50. doi:10.1023/a:1003604209476

Varshney, R. K., Mir, R. R., Bhatia, S., Thudi, M., Hu, Y., Azam, S., et al. (2014a). Integrated Physical, Genetic and Genome Map of Chickpea (*Cicer Arietinum* L.). *Funct. Integr. Genomics* 14 (1), 59–73. doi:10.1007/s10142-014-0363-6

Varshney, R. K., Mohan, S. M., Gaur, P. M., Chamarthi, S. K., Singh, V. K., Srinivasan, S., et al. (2014b). Marker-assisted Backcrossing to Introgress Resistance to Fusarium Wilt (FW) Race 1 and Ascochyta Blight (AB) in C 214, an Elite Cultivar of Chickpea. *Plant Genome* 7 (1), 1–11. doi:10.3835/plantgenome2013.10.0035

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft Genome Sequence of Chickpea (*Cicer Arietinum*) Provides a Resource for Trait Improvement. *Nat. Biotechnol.* 31 (3), 240–246. doi:10.1038/nbt.2491

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 Chickpea Accessions from 45 Countries Provides Insights into Genome Diversity, Domestication and Agronomic Traits. *Nat. Genet.* 51, 857–864. doi:10.1038/s41588-019-0401-3

Yan, W., and Kang, M. S. (2002). *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. New York: CRC Press. doi:10.1201/9781420040371

Yan, W. (2010). Optimal Use of Biplots in Analysis of Multi-Location Variety Test Data. *Acta Agron. Sin.* 36 (11), 1805–1819. doi:10.1016/s1875-2780(09)60075-4

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 © 2022 Bharadwaj, Jorben, Rao, Roorkiwal, Patil, Jayalakshmi, Ahammed, Saxena, Yasin, Jahagirdar, Sontakke, Pithia, Chudasama, Swarup, Singh, Nitesh, Chitikineni, Singh, Singh, Pratap, Dixit, Srivastava 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 | Frontiers in Genetics

TYPE Review
PUBLISHED 08 August 2022
DOI 10.3389/fgene.2022.944501

Check for updates

OPEN ACCESS

EDITED BY
Frank M. You,
Agriculture and Agri-Food Canada (AAFC), Canada

REVIEWED BY
Sunil S. Gangurde,
University of Georgia, United States
Yelda Özden Çiftçi,
Gebze Technical University, Turkey
Faheem Shehzad Baloch,
Sivas Bilim ve Teknoloji Fakültesi, Turkey
Mehdi Rahimi,
Graduate University of Advanced Technology, Iran

*CORRESPONDENCE
R. K. Salgotra,
rks_2959@rediffmail.com
Xianzhong Feng,
fengxianzhong@iga.ac.cn
Javaid Akhter Bhat,
javid.akhter69@gmail.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 15 May 2022
ACCEPTED 28 June 2022
PUBLISHED 08 August 2022

CITATION
Majeed A, Johar P, Raina A, Salgotra RK, Feng X and Bhat JA (2022), Harnessing the potential of bulk segregant analysis sequencing and its related approaches in crop breeding.
Front. Genet. 13:944501.
doi: 10.3389/fgene.2022.944501

COPYRIGHT
© 2022 Majeed, Johar, Raina, Salgotra, Feng and Bhat. 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.

Harnessing the potential of bulk segregant analysis sequencing and its related approaches in crop breeding

Aasim Majeed[1†], Prerna Johar[2†], Aamir Raina[3], R. K. Salgotra[2*], Xianzhong Feng[4*] and Javaid Akhter Bhat[4,5*]

[1]School of Agricultural Biotechnology, Punjab Agriculture University (PAU), Ludhiana, India, [2]School of Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Jammu, India, [3]Department of Botany, Faculty of Life Sciences, Aligarh Muslim University, Aligarh, India, [4]Zhejiang Lab, Hangzhou, China, [5]International Genome Center, Jiangsu University, Zhenjiang, China

Most plant traits are governed by polygenes including both major and minor genes. Linkage mapping and positional cloning have contributed greatly to mapping genomic loci controlling important traits in crop species. However, they are low-throughput, time-consuming, and have low resolution due to which their efficiency in crop breeding is reduced. In this regard, the bulk segregant analysis sequencing (BSA-seq) and its related approaches, viz., quantitative trait locus (QTL)-seq, bulk segregant RNA-Seq (BSR)-seq, and MutMap, have emerged as efficient methods to identify the genomic loci/QTLs controlling specific traits at high resolution, accuracy, reduced time span, and in a high-throughput manner. These approaches combine BSA with next-generation sequencing (NGS) and enable the rapid identification of genetic loci for qualitative and quantitative assessments. Many previous studies have shown the successful identification of the genetic loci for different plant traits using BSA-seq and its related approaches, as discussed in the text with details. However, the efficiency and accuracy of the BSA-seq depend upon factors like sequencing depth and coverage, which enhance the sequencing cost. Recently, the rapid reduction in the cost of NGS together with the expected cost reduction of third-generation sequencing in the future has further increased the accuracy and commercial applicability of these approaches in crop improvement programs. This review article provides an overview of BSA-seq and its related approaches in crop breeding together with their merits and challenges in trait mapping.

KEYWORDS

QTL-seq, MutMap, next-generation sequencing, crop breeding, fine-mapping

Introduction

Identification and dissection of genetic loci determining a particular trait is a regular process in genetics. Most of the complex quantitative traits are regulated by multiple loci distributed across the genome of a species. So to precisely detect the specific genetic elements linked with the trait of interest, one has to link all the loci with that trait (Bhat and Yu, 2021). A quantitative trait locus (QTL) is defined as a region within the genome that is associated with the genetic variation of a quantitative trait. QTL mapping is a widely accepted and applied approach to identify the genes/QTLs determining a complex quantitative trait. Moreover, positional cloning and QTL mapping are the two powerful approaches to dissect the genetic basis of phenotypic variation of important agronomic traits. Both these approaches investigate the genomes for polymorphic markers, followed by linking the polymorphic markers with a particular trait to identify the most likely candidate genomic regions controlling that trait. At the next level, increasing the marker density across these candidate regions would ensure further refinement of their physical interval (fine-mapping), followed by the evaluation of their actual physical position on the chromosomes (physical mapping). Diverse QTLs and underlying genes for numerous traits across a myriad of species have been successfully deciphered using these approaches. The major limitation to these approaches, however, is that they usually are low-throughput and time-consuming (Song et al., 2017).

The bulk segregant analysis is a high-throughput QTL mapping approach to rapidly identify genomic loci regulating the trait of interest. In contrast to individual segregant analysis (ISA), which classifies segregants according to their marker genotypes, the BSA pools segregants according to their phenotypes. When the former compares trait values of different classes, the latter compares marker allele frequencies in different classes (Huang et al., 2020). Although ISA is more commonly used, however, due to more precision and power of BSA and its simplicity, quickness, and cheaper nature than ISA, it provides additional advantages as compared to ISA. The brisk evolution of sequencing technologies along with the rapid downfall of sequencing costs has put the BSA approach to a newer level by integrating the traditional BSA approach with NGS. The basis of BSA is to generate two phenotypically contrasting groups or populations by crossing two extreme phenotypes. This is followed by creating two bulks from the segregating populations, i.e., F_2 by selecting individuals with contrasting phenotypes; for example, tall and short plants, tolerant and susceptible plants, etc. (Zhang and Panthee, 2020). The key to this approach is that the alleles of a locus controlling the trait would be enriched in either bulk; for example, the allele "A" can occur frequently in the tolerant plants, and the allele "a" frequently exists in the susceptible plants, whereas those not affecting the trait would segregate randomly in both bulks (Zhang and Panthee, 2020). BSA was initially targeted to develop genetic markers for trait dissection at earlier stages (Giovannoni et al., 1991; Michelmore et al., 1991). Both marker development and genetic mapping were time-consuming and labor-intensive. However, the rapid advancement of sequencing technologies has greatly facilitated marker discovery and their associations with traits of interest. Integrating BSA with sequencing has dramatically enhanced the speedy detection of marker-trait association by eliminating the time-consuming marker detection step in the traditional BSA approach. This hybrid approach of BSA combined with sequencing was subsequently termed BSA-seq (Zhang and Panthee, 2020). BSA-seq can be regarded as a selective genotyping in which only the tails (individuals with extreme phenotypes) from a population are selected for genotyping. The tailed concept, originally proposed by Darvasi and Soller (2013), reduces the cost and simplifies the analytical process without compromising the statistical power. Rather than analyzing each individual, bulking all the individuals from each tail to create two pools significantly reduces the sequencing cost. BSA-Seq is comparatively an expeditious approach to accomplish the bulk segregant analysis by NGS. BSA conjugated with NGS ensures the rapid identification of both qualitative and quantitative trait loci (Zhang et al., 2021) and speeds up the recognition of candidate genes controlling relevant agronomic traits in diverse crop species (Liang et al., 2020). It can be applied to any population with significant phenotypic differences (Dakouri et al., 2018). For BSA-seq to be more efficient and fruitful, comparatively high sequencing depth and coverage are needed to distinguish significant SNP-trait associations. This results in a sharp rise in the sequencing cost (Zhang et al., 2021), which curbs the application of BSA-Seq to species with large genomes (Tang et al., 2018). However, BSA-seq requires only two sequencing reactions for two pools, thus compensating for high depth and coverage. Nevertheless, for an efficient and productive BSA-seq experiment, the sequencing must be performed to the deepest affordable level, rather than to construct a large pool.

General overview of the BSA-seq technique

Creation of bulks

For the fast-track identification of QTLs linked with a particular trait of interest, a mapping population has to be constructed from a cross between parents encompassing contrasting attributes (Figure 1A). From the progeny of this cross, the individuals exhibiting contrasting phenotypes for a particular trait are initially identified. These contrasting individuals would form two different groups/bulks. For example, some individuals may be resistant to a disease, thereby forming one group/bulk, whereas the other

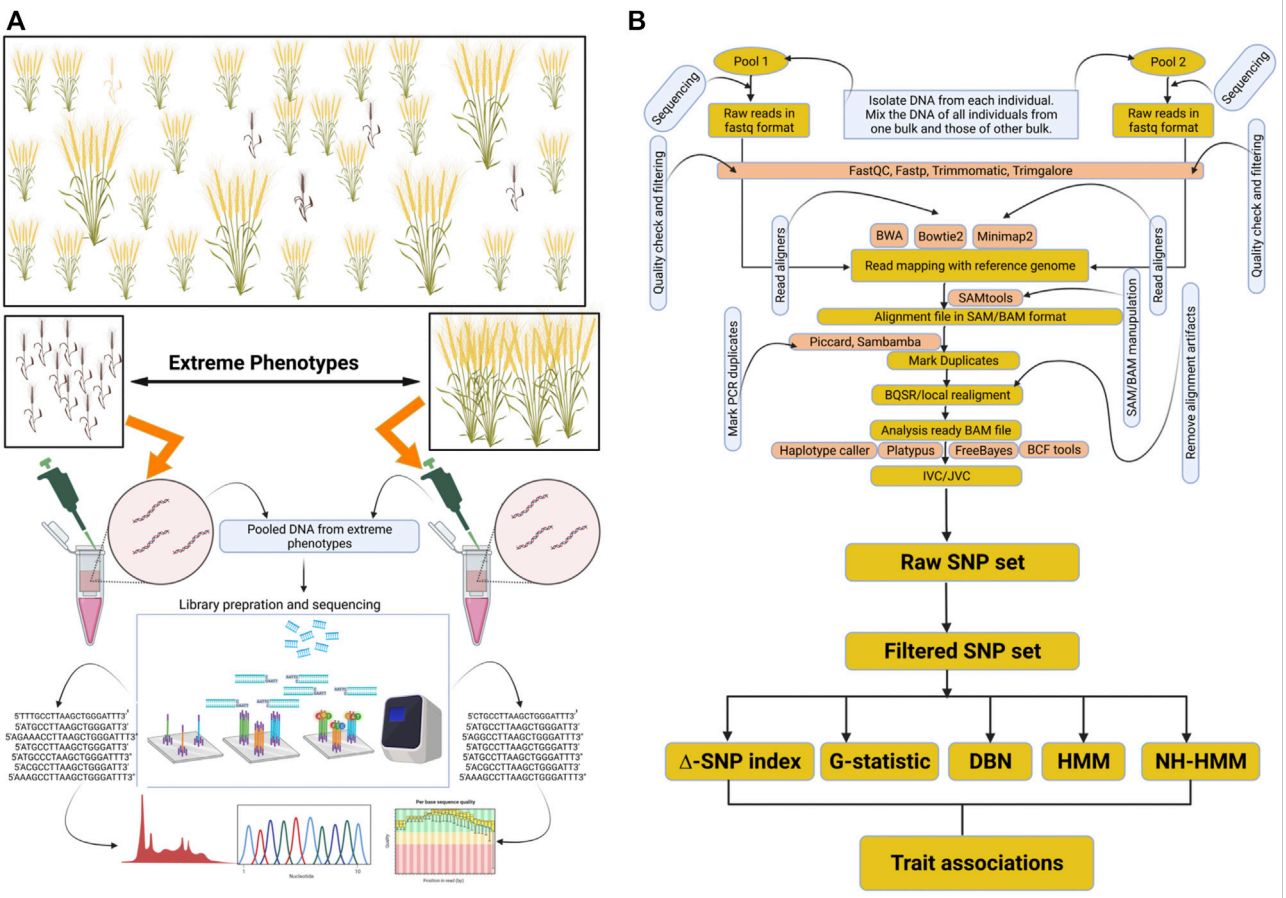


FIGURE 1
Representation of BSA-seq and general data analysis approach for marker trait associations. **(A)** depicts the creation of opposite bulks and their sequencing. **(B)** depicts variant identification and their association with the trait. This figure was created through Biorender https://biorender.com.

individuals showing susceptibility to the disease form another contrasting group/bulk. Then, the DNA of the individuals from each group is extracted, and all the DNA samples of one group are pooled to create one bulk, and those of the other group are pooled separately to create a second bulk (Song et al. (2017)). After that, sequencing libraries are prepared from the pooled DNA samples of each group/bulk, followed by sequencing of the libraries using an appropriate sequencing platform (Yuexiong et al., 2020; Zhan et al., 2020).

Sequencing and variant calling

There are diverse variant calling approaches, with no constraints for appraising a single technique, to call SNPs. Usually, the application of a variant calling technique depends on the organism and the depth and coverage of the sequencing data. The differences in the depth and breadth of sequencing coverage have implications on variant calling. Researchers have a choice to use a particular sequencing strategy, depending on their budget. The bulked samples can be sequenced by using different approaches like whole-genome sequencing, genotyping by sequencing (GBS), restriction site-associated DNA sequencing (RAD-seq), etc. The outcome of the SNP calling depends on the sequencing strategy used. Whatever the sequencing strategy used, the downstream analysis of the sequenced reads, in the fastQ/fasta format, involves aligning them with a reference genome or a *de novo* genome assembly. A standard reference genome of a species is used for this purpose; however, owing to the fact that a single reference genome could not cover all the diversities present within a species, a pangenome concept has emerged to resolve this issue. Therefore, it would be more advantageous to sequence the genome of at least one parent and use it for aligning the reads of two bulks (Luo et al., 2019; Bayer et al., 2020; Kumar et al., 2020). Read alignment represents a critical step of data analysis. Common alignment tools include BWA (Li and Durbin, 2010), Bowtie2 (Langmead and Salzberg, 2012), and Minimap2 (Li, 2018). The resulting alignments are stored and sorted in the

SAM/BAM format. BAM is preferred and has become the standard format due to its compressed size and indexed nature. Manipulation of the BAM file is mostly performed through the SAMtools package (Li et al., 2009). After read alignment with the reference genome, the next step is to identify and remove the duplicated reads, i.e., the reads originating from the same genomic region. The duplicated reads may arise due to the amplification of the same fragment several times during library preparation. Picard (http://broadinstitute.github.io/picard), Sambamba (Tarasov et al., 2015), etc., are used to identify and mark these PCR duplicates in the BAM file for downstream exclusion. Before variant calling, some SNP calling pipelines utilize additional processing steps; for example, the GATK Best Practices workflow (Van der Auwera et al., 2013) performs adjustments to base quality scores of sequencing reads (base quality score recalibration (BQSR)) to remove the alignment artifacts and to reduce false positives through local realignment. However, BQSR/realignment has been found to marginally improve the variant calling accuracy; therefore, these steps may be considered optional (Koboldt, 2020). A myriad of tools have been developed for variant calling accurately like FreeBayes (Garrison and Marth, 2012), GATK HaplotypeCaller (DePristo et al., 2011), Platypus (Rimmer et al., 2014), SAMtools/BCFtools (Li, 2011), TASSEL (Bradbury et al., 2007), etc. Studies have shown that different callers produce similar results (with 90% concordance), and the differences arise only around the low coverage and low confidence regions. Despite these low differences, the differently called genome-wide variants by different callers could amount to 1,000s, necessitating the need for benchmarking and fine-tuning the variant caller (Koboldt, 2020). Choosing a single tool is usually sufficient; however, variants called through different callers can be integrated for sensitive advantage. Various tools like BCFtools are used for this purpose. Whatever the tools used, variant calling can be performed in two ways: individual variant calling (IVC) and joint variant calling (JVC). In IVC, variants are called to create a VCF file for each sample separately, followed by the merging of individual VCF files through BCFtools or other packages. One of the main problems with IVC is that since VCF files contain positional information of variants only, it is not possible to distinguish whether a variation absent in some samples is a wild type or just has low coverage to be called a variant (Koboldt, 2020). In JVC, all the samples are called simultaneously and produce genotypes at each variant position for all samples, which has the potential to resolve the above problem of the IVC. The JVC can also infer the likely genotype of a sample based on the information from the other, which provides a sensitive advantage around low coverage regions (DePristo et al., 2011; Koboldt, 2020). Errors during short-read alignment can produce artifacts during variant calling (Li, 2014). In addition, artifacts may arise due to low-quality base calls, local misalignment around indels, erroneous alignments around low complexity regions, and paralogous alignments of reads not well represented in the reference. These artifacts have been excellently described by Koboldt (2020). Such false positives usually skip during automated filtration, so a visual cross-check using genome browsers is needed to review the alignment of variants.

Approaches for downstream data analysis

When the principle of BSA-seq for the mapping of QTLs is simple, a myriad of statistical methods have been developed to analyze BSA-seq data (Figure 1B). A more convenient and robust pipeline called PyBSASeq was developed by Zhang and Panthee (2020). Once the SNPs are generated by the variant caller, generally, the next step is to filter them based on certain criteria. The unmapped SNPs, missing SNPs, SNPs with more than one alternate (ALT) allele, and the SNPs with low-quality scores are excluded (Zhang and Panthee, 2020). The filtered SNPs are then subjected to Fischer's exact test to obtain a set of significant SNPs. Identification of SNPs is accomplished by matching the bases from sequencing data to the reference genome. Each identified SNP is compared with the reference genome and designated as REF (reference SNP) if identical with the reference genome or ALT (alternate SNP) if not identical with the reference genome. Now, after dividing the number of ALT SNPs by the total number of SNPs (REF + ALT), an allele frequency measure termed the SNP index is obtained. The difference in SNP indices between the two bulks is termed the Δ SNP index. For any SNP, the greater the value of its Δ SNP index, the higher is the probability that SNP is associated with the trait of interest (Zhang and Panthee, 2020). In BSA, the alleles associated with a trait get enriched in either bulk. Therefore, if a gene contributes to a trait, its alleles and, therefore, the SNPs within that gene are enriched in either bulk. For example, in one bulk, there may be more REF allele-containing reads, whereas the other bulk may contain more ALT allele-containing reads. Due to the phenomenon of linkage disequilibrium (LD), the SNPs flanking this gene are also enriched, depending on their closeness to the gene. Based on the quantification of enrichment values of these trait-associated flanking SNPs, a recent python algorithm has been developed to analyze the BSA-seq data more simply and effectively (Zhang and Panthee, 2020). The pipeline can also calculate a G-statistic value for each SNP through the G-test using both REF and ALT SNPs in each bulk. The higher the G-statistic value, the more likely the SNP is linked with the trait. After the calculation of the Δ SNP index or the G-statistic, a sliding window algorithm is utilized to aid the visualization. The sliding windows may contain 100s to 1000s of SNPs among which only a few can be significant. The pipeline developed by Zhang and Panthee (2020) uses the ratio of significant SNPs and the total number of SNPs within a sliding window as an indicator of the trait-associated gene within the sliding window. The greater the ratio, the higher

is the probability of the sliding window containing the trait-associated gene. This approach is referred to as the significant SNP method. When using the Δ SNP index or the G-statistic values during window sliding, the higher Δ SNP index and G-statistic values indicate that the window under observation contains the trait-associated gene (Zhang and Panthee, 2020). Based on the plots derived from the sliding window approach, candidate genomic regions can be distinguished. The candidate genomic regions would be the windows containing the trait-associated gene. These candidate regions or QTLs associated with a particular trait can then be validated using diverse polymorphic markers (Arikit et al., 2019). Next, the SNPs within the candidate regions can be annotated using distinct bioinformatics tools. PyBSASeq is simple, effective, and highly sensitive. It performs better at low sequence coverage; therefore, it has the potential to significantly reduce the sequencing costs. It can calculate significance levels of the detected associations; however, it suffers from the deficiency of the estimation of confidence intervals for the detected QTLs. It may be resolved in future versions.

Block regression mapping (BRM) is another robust approach developed to analyze BSA-seq data comprehensively. This method was developed by Huang et al. (2020). The authors developed this algorithm to solve two key issues in analyzing BSA-seq data: 1) to accurately determine the significance threshold and 2) to determine the confidence interval of the QTLs. These two issues remain associated with QTL-seq, as claimed by Huang et al. (2020). Through the BRM approach, the users can also integrate the results from the BRM pipeline with the Pooled QTL Heritability Estimator (PQHE) (Tang et al., 2018) to estimate the heritability. The method is based on a null hypothesis (Ho), which if an allele is not associated with the trait, the frequency of that allele in two pools is equal. Conversely, the difference in frequencies of an allele between the two pools is equal to zero if the allele is not associated with the trait. However, under this condition, both the pools should be a random sample from the population for the marker. At Ho

$$f_1 = f_2 = f \text{ or } \Delta f = f_1 - f_2 = 0$$

where f = frequency of an allele in a population, f_1 = frequency of an allele in pool-1, and f_2 = frequency of an allele in pool-2. If an allele is linked with a trait, then

$$\Delta f = f_1 - f_2 \neq 0$$

The larger the value of Δf for an allele, the more strongly the allele is associated with the trait (Huang et al., 2020). After calculating the Δf value for each marker, a continuous Δf curve can be plotted across the genome. QTLs can be identified from this curve as their peaks. The local regression method LOESS (Jacoby, 2000) is used to fit the Δf, f_1, and f_2 curves, followed by block regression. To calculate the significance levels, the method relies on the fact that Δf approximately follows a normal distribution under the central limit theorem. Therefore, the significance level of the Δf is calculated using a two-tailed test. If it is significant, the alternative hypothesis is accepted, i.e., there is QTL present in this peak. Then, the confidence interval is derived as the region between the left and right intersection points of a horizontal line (calculated mathematically) with the curve. This region represents the 95% confidence interval of the QTL (Huang et al., 2020). Bonferroni correction is used for multiple testing. Here, f1 and f2 are equivalent to SNP indices, and Δf is the ΔSNP index of the QTL-seq method of Takagi et al. (2013a) and PyBSASeq of Zhang and Panthee (2020). The main advantage of the BRM approach is that it can calculate significance levels through multiple testing and determine the confidence intervals.

Among the other statistical approaches developed to analyze BSA-seq data, a G-statistic-based approach developed by Magwene et al. (2011) is well known. It calculates the G-statistic value for each SNP through a smoothed version of the G-test using both REF and ALT SNPs in each bulk. The higher the G-statistic value, the more likely the SNP is linked with the trait. This method takes into consideration the allele frequency variation due to bulks and variation due to sequencing of bulks. Larger bulk sizes and enough sequencing depth have the potential to detect even weak effect QTLs (Magwene et al., 2011). Although this approach is simple, Huang et al. (2020) have asserted that the method of calculating FDR for multiple testing is not concretely devised, confidence intervals cannot be estimated through this method, and it is less effective under low sequencing depth.

The MULTIPOOL method was developed by Edwards and Gifford (2012) for genetic mapping through the utilization of pooled genotyping. This approach was focused on experiments with model organisms, where the progeny of a cross is grouped and pooled based on phenotypes. Its theme is simple: a marker not linked with a trait shall segregate with equal frequency in both pools, whereas the marker linked with a trait shall be enriched in either pool. It was developed to handle larger data sets containing 1000s of markers. It uses the dynamic Bayesian network (DBN) approach for estimating confidence intervals and statistical accuracy of QTLs. The method can be used for any number of replicates and multiple experimental designs (Edwards and Gifford, 2012). It uses a probabilistic multi-locus dynamic Bayesian network model, wherein a single chromosome is considered at a time to model the influence of pool size and recombination on the frequency of neighboring alleles and describes the allele frequency change across the chromosome. Although MULTIPOOL does not rely on a specific read aligner or SNP calling strategy, however, it suffers from the problem of estimating the LOD threshold and judging the significance of signals accurately.

A simpler and widely accepted method known as QTL-seq was developed by Takagi et al. (2013b) to identify the QTLs in rice recombinant inbred lines (RIL) and F_2 populations but can be applied to any population for detecting genomic regions that

TABLE 1 Key characteristics of different statistical approaches and pipelines used to analyze BSA-seq data.

	Method	Key statistics used	Citations	Limitation	Advantage
1	G-statistic	G-test	200	Based on estimating the G′ threshold; the method for calculating FDR for multiple testing has not been concretely devised; significantly affected by sequencing depth and is less suitable under low sequencing depth; no estimation of confidence intervals	Simplicity
2	MULTIPOOL	Probabilistic multi-locus dynamic Bayesian network model	70	Based on estimating the LOD threshold, judging the significance of signals	Non-reliance on a particular aligner and SNP calling strategy
3	QTL-Seq	SNP index, Δ SNP index	780	Significance threshold estimated in QTL-seq is inappropriate; no estimation of confidence interval	Simplicity and intuition
4	EXPLoRA	Hidden Markov model (HMM), Linkage disequilibrium (LD)	45	No multiple testing correction, sometimes maps a single QTL as two or more adjacent QTLs, no confidence interval estimation	Robust even at a low signal-to-noise ratio
5	Hidden Markov model	HMM	9	Does not take into account that co-segregation of SNPs is affected by the distance between them	
6	Non-homogeneous hidden Markov model	HMM	16	Takes the effect of distance between SNPs during co-segregation into account	
7	QTG-Seq	smooth LOD test, Euclidean distance, and G-statistic	49	Large pool size and high sequencing coverage required	Time- and cost-saving strategy for fine-mapping, suitable for minor-effect QTLs, mapping resolution up to the gene level, and requires only four generations from the first cross of any parent lines for fine-mapping
8	PyBSASeq	Fischer's exact test, Δ SNP index or G-statistic, significant SNP method	4	No estimation of confidence intervals for detected QTLs	Simple and effective, calculates significance, can detect SNP-trait associations at lower sequencing coverage so can reduce up to 80% sequencing cost, high sensitivity
9	Block regression mapping	Δf or Δ SNP index, Δf curve LOESS analysis, block regression, central limit theorem, and Bonferroni correction	10	Not apparent yet	Calculates significance, uses multiple testing, estimates confidence intervals
10	QTLseqr	Δ SNP index and G-statistic	93	Not apparent	Calculates significance, uses multiple testing, and options for better visualization

underwent artificial or natural selection. It can also be applied to populations under different environmental conditions like high and low temperatures. However, this method is not suitable for detecting minor effect QTLs as replicated measurements are not possible for each genotype. The approach uses the Δ SNP index method. It first calculates k (number of reads having an allele different from the reference); then, the SNP index is calculated using the formula

$$\text{SNP} - \text{index} = \text{k}\,1/\text{n},$$

where n = total number of reads.

QTL-seq estimates the contribution of each parent to the variation. If SNP index = 0, there is no variation and all SNPs are the same as reference. If SNP index = 1, all SNPs belong to either parent and if SNP index = 0.5, each parent contributes equally to the variation. Generally, only the SNPs with SNP index > 0.3 in either bulk are retained for downstream analysis. A sliding window is then applied to visualize the graphs based on the SNP index. After that, the Δ SNP index is calculated for all genomic regions, and the regions exhibiting a higher Δ SNP index than the background genome represent the regions associated with the trait of interest. These regions correspond to peaks or the valleys of the SNP index plot (Takagi et al., 2013a). Depending on the type of genotyping, i.e., whether analysis involves allele frequencies of both tails of the phenotypic distribution of a targeted trait (bidirectional selective genotyping) (Zhang et al., 2003) or the allele frequencies from only one tail (unidirectional selective genotyping) (Foolad et al., 2001), the QTL-seq can be termed as bidirectional and unidirectional QTL-seq, respectively. Although this selective genotyping of one or both phenotypic extremes has the potential to detect effective QTLs, a simulation study evaluating the power and precision of unidirectional and

bidirectional approaches revealed that the latter is more powerful than the former (Navabi et al., 2009). QTL-seq has been successfully used in a myriad of species like tomato (Illa-Berenguer et al., 2015; Wen et al., 2019), capsicum (Park et al., 2019), groundnut (Pandey et al., 2017; Clevenger et al., 2018), watermelon (Branham et al., 2018; Cho et al., 2021), bottle gourd (Chanda et al., 2018; Song et al., 2020), pear (Xue et al., 2017), radish (Hu et al., 2022), rice (Lei et al., 2020), soybean (Zhang et al., 2018), etc. QTL-seq is the most popular and widely used tool for BSA-seq analysis and has the most citations (Table 1). However, Huang et al. (2020) opine that the significance threshold estimated in QTL-seq is inappropriate, and there is no estimation of confidence interval.

Several approaches have also been developed that rely on the hidden Markov model (HMM) concept to link SNPs with the phenotype. The HMM is used to explain or derive the probabilistic characteristic of any random process. It is used to describe the observed events that depend on hidden events. HMMs capture the hidden information from observed sequential events. In HMM, the system being modeled is assumed to be a Markov process with unknown parameters, and the observed parameters are used to determine the hidden parameters. The latter is used for further analysis (Hundal et al., 2016; Lan et al., 2017). The EXPLoRA method was developed by Duitama et al. (2014) to precisely distinguish between true and spurious linked regions for a trait of interest. This algorithm relies on linkage disequilibrium and uses HMM to model the relationships between neighboring markers. This algorithm is robust and performs better under a low signal-to-noise ratio. This tool is claimed to give better results when the true linkage signal is diluted by the availability of few segregants, sampling, and screening errors (Duitama et al., 2014). EXPLoRA is effective even at a low signal-to-noise ratio, but no multiple testing correction and confidence interval estimation are carried out. Another approach named the hidden Markov model (HMM) was developed by Calaesen and Burzykowski (2015) to analyze the BSA-seq data. The model assumes different states of a nucleotide, and each state in an offspring being same or different compared to the parent. Transition of nucleotides implies transition in states. By calculating the probabilities of transitions and states, the most probable state of each SNP is selected, which indicates the most probable genomic regions associated with the trait (Calaesen and Burzykowski, 2015). Through this method, each identified SNP is classified into one of the several predefined states having their specific biological interpretation. The HMM identified states allow the identification of genomic regions containing genes governing the trait. This method is based on the assumption that the identified SNPs are equally spaced across the whole genome, which may not always be the case. Furthermore, the co-segregation of SNPs is affected by the distance between them. Taking these two issues into consideration, an extended method of the HMM known as the non-homogeneous hidden Markov model (NH HMM) was developed by Ghavidel et al. (2015), which takes the distance between SNPs into account.

The quantitative trait gene sequencing (QTG-Seq) method was developed by Zhang et al. (2019) to accelerate QTL fine-mapping. The method partitions QTLs to convert a quantitative trait into a near-qualitative trait. The partitioning is performed by selfing the individuals heterozygous for the target QTL and homozygous for other QTLs. This is followed by mining, in which bulked pools are sequenced. In addition to the Euclidean distance and G-statistic, a new statistic called smooth LOD was used to delimit the QTL to a small interval (Zhang et al., 2019). For the determination of minor-effect QTLs and fine-mapping, QTG-Seq is cost-effective and time-saving but at the cost of a large pool size and high sequencing coverage required. The details and key features of these approaches are presented in Table 1.

BSR-Seq approach

SNPs can be deduced from the transcriptomic data also; therefore, it is also possible to use the RNA-sequencing technology to efficiently identify SNPs from bulks. This integration of BSA and transcriptome is known as bulked segregant analysis RNA-seq (BSR-Seq). The fundamental principles of BSR-Seq would remain the same as that of the traditional BSA-seq, with the difference that only the transcribed genome is used as a data source. BSR-Seq has been applied for the elucidation of important genomic regions and SNPs associated with different traits in both plants and animals. For example, Wang et al. (2013) identified 1,255 and 56,419 differentially expressed genes (DEGs) and SNPs, respectively, between resistant and susceptible pools against enteric septicemia in catfish. By pooling the RNA samples from 12 homozygous F_3 resistant lines to the stem rust pathogen (strain Ug99 F_3) and 11 susceptible homozygous lines, Edae and Rouse (2019) could map the stem rust resistance to a 3.2-Mbp region on chromosome 2U of *Ae. umbellulata,* with two nucleotide-binding and leucine-rich repeat (NLR) genes as the potential candidate genes (Edae and Rouse, 2019). Moreover, BSR-Seq was used to clone the *glossy3* (*gl3*) gene of maize (Liu et al., 2012). In addition, the molecular details of wheat powdery mildew resistance through BSR-Seq revealed that a single dominant gene on chromosome 5DS conferred resistance (Zhu et al., 2020). BSR-seq was also used to identify DEGs and SNPs associated with waterlogging tolerance (Du et al., 2017). Through the BSR-seq technique, the regulatory network of melon color was identified by Chayut et al. (2015). The cold tolerance response of *Actinidia arguta* through BSR-Seq revealed that soluble sucrose and β-amylase activity were enhanced in tolerant population compared to susceptible population (Lin et al., 2021). Through BSR-Seq, in addition to the QTL regions, the differential expression of candidate genes is also

achieved. However, for traits affected by the environment and the traits determined by many minor genes, BSR-Seq may not be very effective (Edae and Rouse, 2019).

MutMap approach

With the advent of the sequencing technologies, there has been a rapid progress in deciphering the causative alleles for a particular trait more quickly than traditional QTL mapping approaches. Spontaneous mutations and activation of natural mutagens like transposons and viruses, etc., are the main sources of variation in the natural populations. If variations are not sufficient in a natural population, then artificial mutagens like EMS, acridine dyes, base analogs, and UV, X-, and gamma-rays can be used to induce mutation (Raina et al., 2016, 2020). The mutations caused by these agents lead to altered phenotypes through the generation of SNPs, indels, or segmental breaks (Tribhuvan et al., 2018). The phenotypic variations either existing spontaneously or induced artificially are exploited to map the causative genes/QTLs using an appropriate marker system. When the phenotypic variation is artificially induced to create a mutant phenotype, then the causative mutation can be analyzed and identified through MutMap methods. These methods include MutMap, MutMap+, SHOREmap (SHOrtREad map), MutMap-Gap, and NGM (next-generation mapping). First, a mutant phenotype is created through mutagenesis which is then crossed with the parent to create F_1 and F_2 populations. With the aid of a marker system like SNPs, the mutant phenotype is screened for segregation in the filial generations and thus mapped on the genome. The key to these approaches is the utilization of traditional BSA to generate SNP data. They exploit the power of NGS technologies to map inherited traits across any plant species where the generation of an F_2 mapping populations is possible. SHOREmap was developed by Schneeberger et al. (2009) to identify the causative gene mutation for slow growth and pale green leaves in *Arabidopsis*. The authors first generated mutant lines and crossed it with distant parent to dilute the distribution pattern of non-causal SNPs present throughout the genome. Next, they created a bulk of 500 mutant individuals, and their DNA was pooled and sequenced. SNPs were identified between the parent and mutant. The basic idea of this technique is that among the progenies of *mutant x parent*, those with the mutant phenotype are assumed to have SNP distribution similar to mutants at the loci controlling the mutant phenotype, whereas other loci have a random distribution. NGM mapping was developed by Austin et al. (2011) to identify the cell wall biosynthesis and maintenance genes in *Arabidopsis*. NGM is similar to SHOREmap except that it utilizes less mutant lines (10) to create a bulk, without compromising the power to detect causal mutation. Like SHOREmap and NGM, the MutMap technique, which was developed by Abe et al. (2012), also uses the creation of bulks from mutant F_2 progenies (Etherington et al., 2014). MutMap, however, differs from the aforementioned two techniques in the sense that while the latter utilizes distantly related mapping lines, the former relies on the crossing between the mutant and its wild type itself. This approach of MutMap directly targets the causal SNPs generated during mutagenesis (Tribhuvan et al., 2018). The SNPs, which are associated with the mutant phenotype, would show 0% wild type and 100% mutant reads, whereas the unlinked SNPs would show 50% each. Abe et al. (2012) developed an SNP index as the number of mutant SNP reads divided by the total number of SNP reads. If this index =1, it means that the SNP is highly linked to the mutant phenotype. This method is more likely to map recessive mutations. An advancement to MutMap known as MutMap+ was developed by Fekih et al. (2013) to tackle the lethal or sterile mutations, wherein F_2 cannot be developed. Bulks of around 20–30 individuals for mutant and wild type are created at M_3 generation, sequenced at ~10x coverage followed by SNP identification in both bulks. Then, the SNP index for both is calculated, and the Δ SNP index is derived by subtracting the wild-type SNP index from the mutant SNP index. The positive Δ SNP index values indicate that SNP is linked with the phenotype. A further extension of MutMap, to map a causal mutation with the gaped region of the genome, was developed and named MutMap-Gap (Takagi et al., 2013a, b). Here, if an SNP with index = 1 remains undetected in the reference genome, there is a possibility that such SNPs are present within the gaped regions. Then, the unassembled reads are *de novo* assembled, and the casual SNP is identified using the *de novo* assembly. The key advantages of MutMap include 1) no need of large mapping populations, 2) no need of genetic linkage maps, 3) no need of natural variation in the population, 4) time-saving and labor-effective, and 5) direct identification of casual SNPs. Key disadvantages include 1) availability of a reference genome, 2) artificial mutagenesis required to develop mutant lines, 3) maintenance of mutant lines, and 4) not applicable if a phenotype cannot be scored (Tribhuvan et al., 2018).

Successful application of BSA-Seq in elucidation of trait-associated QTLs

A myriad of studies on diverse species have proven the applicability of BSA-seq in mapping QTLs for different traits of agronomic importance. The details of some important studies on important crop plants are presented in Table 2. BSA-Seq successfully identified the genomic regions controlling the locule number and fruit weight in tomatoes (Illa-Berenguer et al., 2015) that may lead to significant breakthroughs in fruit development in tomatoes. Breeding heat-tolerant cultivars of tomatoes seems appealing. In order to identify the heat stress-responsive QTLs in tomato, Wen et al. (2019) used an integrated approach of conventional QTL mapping, BSA-Seq, and RNA-Seq and

TABLE 2 Details of the studies utilizing BSA-Seq for the elucidation of trait-specific genomic regions in different crop species.

S.No.	Species	Pop type	Pop size	Pool size	Sequencing strategy	Number of SNPs	Bioinformatics approach used	Trait	Key findings	Reference
1	Rice	NIL-F_2	176	35 individuals from extreme phenotypes	Whole-genome sequencing (WGS)	455,262	QTL-seq method	Grain length and weight	One major QTL, 15–20 Mb on chr 5, for grain length and weight identified	Yaobin et al. (2018)
2	Rice	F_3	10,800	385 tolerant pools and 430 sensitive pools	Paired-end Illumina sequencing on Hiseq 2000 platform	450,000	G′ statistic method	Cold tolerance	Six QTLs were mapped on chromosomes 1, 2, 5, 8, and 10	Yang et al. (2013b)
3	Rice	F_2	940	20 extreme phenotypes for heading time (HT) and plant height (PH)	WGS on the Illumina HiSeq X Ten platform	511,393 for HT and 543,319 for PH	ΔSNP-index method	Heading time and plant height	Four QTLs for HT on chromosomes 3, 6, 9 and 10. Three QTLs for PH on chromosomes 1 and 8	Zhang et al. (2021)
4	Rice	RILs	190	20 extreme phenotypes were used for bulking	Paired-end sequencing using the Illumina MiSeq platform	184,917	Euclidean distance and ΔSNP-index method	Cold tolerance	One major QTL on chr6 was identified, which spans 1.81 Mb and harbors 269 genes	Sun et al. (2018)
5	Rice	RILs	151	----------	Paired-end sequencing using Illumina HiSeq 2500	116,993	Euclidean distance and ΔSNP-index method	Grain shape	One major QTL on chr9 was identified, which spans 0.8 Mb and harbors 101 genes	Wu et al. (2020)
6	Cucumber	F_2	258	10 individuals from extreme phenotypes	Illumina paired-end sequencing	234,393	Δ (SNP index)	Early flowering	One major QTL around 890 kb on chr 1 for early flowering. The gene Csa1G651710 was identified as the main flowering switch	Lu et al. (2014)
7	Cucumber	F_3	135	15 resistant and 15 susceptible	Paired-end sequencing using Illumina HiSeq 2000	933,846 and 915,524 for susceptible and resistant bulk	ΔSNP-index method	Vein yellowing virus resistance	A unique region in chromosome 5 containing 24 annotated genes was identified for resistance	Pujol et al. (2019)
8	Maize	RILs	224	46 more extreme plants formed two pools	Paired-end sequencing using Illumina HiSeq 2000	3,301,371	Customized R-script	Flowering time and plant height	Two major QTLs found for FT on chr 5 and chr 8 were 10.8 Mb and 18.9 Mb in size, respectively. Two major QTLs on chr 4 and chr 6 found for PH were 21.2 Mb and 9.7 Mb in size, respectively	Haase et al. (2015)
9	Maize	ILs	400	10 tolerant and 10 sensitive extreme phenotypes	BSR-seq	114,580	Bayes' theorem	Waterlogging	In tolerant and sensitive bulks, 354 and 1,094 genes were differentially expressed, respectively. GRMZM2G055704 on chromosome 1 was identified as	Du et al. (2017)

(Continued on following page)

10.3389/fgene.2022.944501

TABLE 2 (*Continued*) Details of the studies utilizing BSA-Seq for the elucidation of trait-specific genomic regions in different crop species.

S.No.	Species	Pop type	Pop size	Pool size	Sequencing strategy	Number of SNPs	Bioinformatics approach used	Trait	Key findings	Reference
									a candidate gene responsive to waterlogging	
10	Wheat	RILs	244	six low- and six high-TGW	SLAF-seq	132,530	ΔSNP-index method	1,000 grain weight (TGW)	One candidate gene associated with TGW was identified on chr 7A	Hu et al. (2016)
11	Hessian fly (HF), a wheat galling parasite	Non-structured Louisiana field population	---	23 virulent and 19 avirulent	WGS	1.5 million	Fisher's exact test using PoPoolation2	Hessian fly (HF) virulence to wheat R genes H6, Hdic, and H5	One 1.3-Mb region for HF virulence was mapped to HF autosome 2	Navarro-Escalante et al. (2020)
12	Chickpea	F_4	221	10 individuals of each low and high seed weight forming two pools	Paired-end WGS using the Illumina HiSeq 2000 platform	118,321	Δ (SNP index)	100 seed weight	One major QTL of 35 kb on chromosome 1 containing six genes	Das et al. (2015)
13	Chickpea	RILs	92 and 139 for two populations	10 and 14 extreme phenotypes for two populations	WGS	77,938 in one population and 106,907 in the other	G-statistic and ΔSNP-index method	Ascochyta blight resistance	17 QTLs identified and mapped on chromosomes Ca1, Ca2, Ca4, Ca6, and Ca7	Deokar et al. (2019)
14	Tomato	F_2 populations	549	10 individuals of each extreme phenotype	Paired-end WGS using the Illumina HiSeq 2000 platform	----------	Δ (SNP index)	Fruit weight (FW) and locule number (LC)	Three highly significant and newly mapped FW QTLs on chr 1 and chr 11. 66 candidate genes for FW. Three LC QTLs of low significance	Illa-Berenguer et al. (2015)
15	Groundnut	RILs	266	25 individuals with extreme phenotypes	Paired-end WGS using Illumina HiSeq 2000	259,621 for rust and 243,262 for LS	Δ (SNP index)	Rust and late leaf spot disease	One 3.06-Mb region on the A03 pseudomolecule of A-genome harboring 3,136 SNPs was identified for rust resistance. A 2.98 Mb region on A03 pseudomolecule harboring 66 SNPs was identified for LS resistance	Pandey et al. (2017)
1F6	Groundnut	RILs	366	20 individuals with extreme phenotypes	WGS	10,759	ΔSNP-index method	Fresh seed dormancy	Two genomic regions on the B05 and A09 pseudomolecules control seed dormancy	Kumar et al. (2020)
17	Potato	Diploid mapping population	90	10 individuals with extreme phenotypes	Paired-end Illumina HiSeq 2000	6,766,8,152,000	Pearson's chi-squared test	Steroidal glycoalkaloids (GAs)	One region located on chromosome 1 ranging from 63.1 to 73.5 Mb was found the most confident	Kaminski et al. (2016)
18	Pepper	F_2	249	30 individuals with extreme phenotypes	SLAF-seq	106,848	Euclidean distance	first flower node	One QTL on chr 12 was detected, followed by 393 high-	Zhang et al. (2018)

(Continued on following page)

10.3389/fgene.2022.944501

TABLE 2 (*Continued*) Details of the studies utilizing BSA-Seq for the elucidation of trait-specific genomic regions in different crop species.

S.No.	Species	Pop type	Pop size	Pool size	Sequencing strategy	Number of SNPs	Bioinformatics approach used	Trait	Key findings	Reference
									quality SNP markers associated with FFN	
19	Sunflower	F_2 and F_3	300	15 individuals with extreme phenotypes	Genotyping-by-sequencing (GBS)	11,484	G-statistic using QTLseqr	Broomrape resistance	Two major QTLs on chromosome 3	Imerovski et al. (2019)
20	Rapeseed	BC_8	965	36 individuals with extreme phenotypes	Illumina HiSeq 2000 platform	1,830,225	ΔSNP-index method	Plant architecture	Five major QTLs on chromosome 1	Ye et al. (2022)
21	Pigeonpea	F_2	179	15 individuals with extreme phenotypes	WGS	47,429, 46,510, and 54,556 for three different bulk types	ΔSNP-index method	Days to flowering (DTF)	Two significant genomic regions, one on CcLG03 and another on CcLG08 were found controlling DTF	Singh et al. (2022)

found five major QTLs determining the trait of interest. Their results have significance in breeding for improved thermotolerance in tomatoes. A major QTL on chromosome 1 regulating capsaicinoid biosynthesis in the pericarp of capsicum was identified by Park et al. (2019) using BSA-Seq in integration with RNA-Seq. Significant yield losses and deteriorated fodder quality in groundnut are caused due to rust and late leaf spot fungal diseases. To address this issue, Pandey et al. (2017) identified three QTL loci for rust resistance and one for late leaf spot resistance using the BSA-Seq approach. Furthermore, Clevenger et al. (2018) also mapped late leaf spot resistance QTLs in groundnut by BSA-Seq. Identification of these genomic regions controlling rust and late leaf spot resistance and their introgression into elite groundnut cultivars would bring revolution in groundnut breeding. In the case of watermelon, *Fusarium* wilt, caused by the fungus *Fusarium oxysporum* f. sp. niveum (Fon) leads to significant losses in yield. Using BSA-seq coupled to QTL mapping, the genomic region controlling resistance to this fungus was fine-mapped to around the 315-kb region (Branham et al., 2018). In order to map loci determining semi-dwarfism in watermelon, Cho et al. (2021) identified a single recessive gene through BSA-Seq. Watermelons are severely affected by heat stress. Grafting watermelon to heat-tolerant bottle gourd rootstocks is one solution to this problem (Yang Y. et al., 2013a). So understanding of the inheritance and identification of loci controlling heat tolerance in bottle gourd may lead to significant breakthroughs in watermelon breeding for heat tolerance. Using BSA-Seq, seven heat-tolerant QTLs with one as a major effect QTL for heat tolerance in bottle gourd were determined (Song et al., 2020). In addition, bottle gourd rootstock is used to improve cold tolerance and disease resistance in cucurbits. The aphid-transmitted papaya ringspot virus watermelon strain (PRSV-W) and zucchini yellow mosaic virus (ZYMV) are the two most common viruses infecting bottle gourd. Through BSA-seq, fine-mapping of the *Prs* locus and identification of the candidate resistance gene for PRSV-W were elucidated (Chanda et al., 2018). The red peel of pears is more attractive and also provides health benefits to consumers. So genes controlling the skin coloration aid in cultivar selection and enhance pear breeding. Using BSA-seq, a 582.5-kb candidate genomic region associated with red/green skin (R/G) locus, harboring approximately 81 predicted protein-coding genes, was identified (Xue et al., 2017). Further fine-mapping and elucidation of the specific casual genes would enhance the commercial value of pears. An important commercial attribute of radish is its root shape, measured as the ratio of the root length to root diameter. Hu et al. (2022) identified seven QTLs distributed on five chromosomes controlling the root shape. The results of this study are significant for fine-mapping and functional analysis of root-shaped QTLs and cultivar breeding for the root shape in radish (Hu et al., 2022). Among the abiotic stresses, salt stress negatively affects all crop species, leading to compromised plant performance and significant losses in crop

yield. Seedling survival and overall yield in rice are directly affected by salt tolerance at the bud burst stage. Using BSA-seq, a major candidate region on chromosome 7 was identified, which was further narrowed down to a 222-kb genomic interval. Furthermore, five differentially expressed genes (DEGs) were identified in this candidate region through the RNA-seq approach at the bud burst stage under the salt-treated condition. In addition, the expression of one gene, (*OsSAP16*), was enhanced under drought stress, implying that *OsSAP16* is the strong candidate gene (Lei et al., 2020). These results are significant for improving the salt tolerance of rice varieties. Grain size and weight are important traits that determine the overall yield in cereal crops. In order to identify the candidate genomic regions controlling the grain size and weight in rice, Yaobin et al. (2018 used BSA-seq and identified a 15–20 Mb region on chromosome 5. Plant height is closely related to soybean yield. Using QTL-seq, Zhang et al. (2018) identified a 1.73-Mb region on chromosome 13. Linkage mapping was used to confirm this region in the mapping population. Candidate gene analysis revealed that *Glyma.13 g249400* showed significantly higher expression in soybean plants with greater plant height; therefore, it can be a strong candidate gene for this trait (Zhang et al. (2018).

In addition to its importance in crop plants, BSA-seq is also widely used in other species as well like yeast (Wenger et al., 2010; Hu et al., 2015), *Tilapia* fishes (Gu et al., 2018), etc. Using MULTIPOOL, Vogel et al. (2021) was able to dissect the genomic regions controlling the root and crown rot resistance against phytophthora in squash fish, whereas De Witt et al. (2019) and Wang et al. (2019) were successful in elucidating the unique alleles involved in lignocellulosic inhibitor tolerance and genomic variants linked to high-temperature fermentation performance in yeast, respectively. Furthermore, Trindade de Carvalho et al. (2017) successfully used BSA-seq in yeast through EXPLoRA that relies on the hidden Markov model (HMM).

Potential of BSA-seq in medicinal plant genomics

Medicinal plants are less explored at the genomic level as compared to staple crop plants. Even the breeding programs for the genetic improvement of medicinal plants are at the pioneering stage, and the development of trait-specific homogeneous lines is far away from reality. However, BSA-seq can greatly speed up and facilitate their breeding programs by making the use of F_2 generations. Thus, medicinal plant breeders can get a general idea about the nature of the progenies in the context of a specific trait by integrating BSA with NGS technologies. This information can then be used to develop trait-specific homogeneous lines through selfing of the selected lines. Practically, this approach is feasible for only those medicinal plants that have less generation time and flower early in life. Some examples of such medicinal plants that are best suited for which reference genome is available and hence BSA-seq can serve their breeding purpose effectively may include stevia, tea, and tulsi. For longer generation time in medicinal plants, especially tree species, the creation of F_2 generation is almost impossible.

Conclusion and future prospects

BSA-seq and its related approaches have the potential to quickly identify the trait-specific genomic regions/QTLs in a high-throughput manner. It takes the advantage of traditional BSA in integration with rapidly evolving NGS technologies. The most admirable attribute of this approach is that it takes only F_2 generations to precisely identify trait-specific genomic regions/QTLs, thereby saving much time. However, this is achieved at the cost of additional capital investment for deep sequencing. Therefore, there is a trade-off between time and capital investment in using BSA-seq. With the rapid advancement of NGS technologies and a steep decrease in the cost of sequencing, it is expected that in near future, the sequencing depth would not be a matter of concern while estimating the overall cost of BSA-seq.

Author contributions

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

Funding

This work was financially supported by the Zhenjiang International-Joint fund (Grant No. 595 GJ2020010) and by the Zhejiang Lab (Soybean Intelligent Computing Breeding and Applications, 596 Grant No. 2021PE0AC04).

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.

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H. N., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi:10.1038/nbt.2095

Arikit, S., Wanchana, S., Khanthong, S., Saensuk, C., Thianthavon, T., Vanavichit, A., et al. (2019). QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). *Sci. Rep.* 9, 8328. doi:10.1038/s41598-019-44856-2

Austin, R. S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N. J., Bonetta, D., et al. (2011). Next-generation mapping of Arabidopsis genes. *Plant J.* 67, 715–725. doi:10.1111/j.1365-313X.2011.04619.x

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., and Edwards, D. (2020). Plant pan-genomes are the new reference. *Nat. Plants* 6 (8), 914–920. doi:10.1038/s41477-020-0733-0

Bhat, J. A., and Yu, D. (2021). High-throughput NGS-based genotyping and phenotyping: Role in genomics-assisted breeding for soybean improvement. *Legum. Sci.* 3, e81. doi:10.1002/leg3.81

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., Buckler, E. S., et al. (2007). Tassel: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi:10.1093/bioinformatics/btm308

Branham, S. E., Patrick Wechter, W., Lambel, S., Massey, L., Ma, M., Fauve, J., et al. (2018). QTL-seq and marker development for resistance to *Fusarium oxysporum f. sp. niveum* race 1 in cultivated watermelon. *Mol. Breed.* 38 (11), 139. doi:10.1007/s11032-018-0896-9

Chanda, B., Wu, S., Fei, Z., and Ling, K. (2018). "Resequencing of bottle gourd germplasm and using QTL-seq to fine-map PRSV-W resistance in Bottle gourd (*Lagenaria sinceraria*)," in *Cucurbitaceae proceedings*. Available at: https://www.ars.usda.gov/research/publications/publication/?seqNo115=358356.

Chayut, N., Yuan, H., Ohali, S., Meir, A., Yeselson, Y., Portnoy, V., et al. (2015). A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β-carotene accumulation in melon fruit. *BMC Plant Biol.* 15, 274. doi:10.1186/s12870-015-0661-8

Cho, Y., Lee, S., Park, J., Kwon, S., Park, G., Kim, H., et al. (2021). Identification of a candidate gene controlling semi-dwarfism in watermelon, *Citrullus lanatus*, using a combination of genetic linkage mapping and QTL-seq. *Hortic. Environ. Biotechnol.* 62 (3), 447–459. doi:10.1007/s13580-020-00330-x

Claesen, J., and Burzykowski, T. (2015). A hidden Markov-model for gene mapping based on whole-genome next generation sequencing data. *Stat. Appl. Genet. Mol. Biol.* 14 (1), 21–34.

Clevenger, J., Chu, Y., Chavarro, C., Botton, S., Culbreath, A., Isleib, T. G., et al. (2018). Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker-assisted selection. *Front. Plant Sci.* 9, 83. doi:10.3389/fpls.2018.00083

Dakouri, A., Zhang, X., Peng, G., Falk, K. C., Gossen, B. D., Strelkov, S. E., et al. (2018). Analysis of genome-wide variants through bulked segregant RNA sequencing reveals a major gene for resistance to *Plasmodiophora brassicae* in *Brassica oleracea*. *Sci. Rep.* 8, 17657. doi:10.1038/s41598-018-36187-5

Darvasi, A., and Soller, M. (2013). *Advanced intercross lines*, 39–42.

Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Kumar, L. V, et al. (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res.* 22, 193–203. doi:10.1093/dnares/dsv004

De Witt, R. N., Kroukamp, H., Van Zyl, W. H., Paulsen, I. T., and VolscHenk, H. (2019). QTL analysis of natural *Saccharomyces cerevisiae* isolates reveals unique alleles involved in lignocellulosic inhibitor tolerance. *FEMS Yeast Res.* 19, foz047. doi:10.1093/femsyr/foz047

Deokar, A., Sagi, M., Daba, K., and Tar'an, B. (2019). QTL sequencing strategy to map genomic regions associated with resistance to ascochyta blight in chickpea. *Plant Biotechnol. J.* 17, 275–288. doi:10.1111/pbi.12964

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi:10.1038/ng.806

Du, H., Zhu, J., Su, H., Huang, M., Wang, H., and Ding, S. (2017). Bulked segregant RNA-seq reveals differential expression and SNPs of candidate genes associated with waterlogging tolerance in maize. *Front. Plant Sci.* 8, 1022. doi:10.3389/fpls.2017.01022

Duitama, J., Sánchez-Rodríguez, A., Goovaerts, A., Pulido-Tamayo, S., Hubmann, G., Foulquié-Moreno, M. R., et al. (2014). Improved linkage analysis of Quantitative Trait Loci using bulk segregants unveils a novel determinant of high ethanol tolerance in yeast. *BMC Genomics* 15, 207. doi:10.1186/1471-2164-15-207

Edae, E. A., and Rouse, M. N. (2019). Bulked segregant analysis RNA-seq (BSR-Seq) validated a stem resistance locus in *Aegilops umbellulata*, a wild relative of wheat. *PLoS One* 14, e0215492. doi:10.1371/journal.pone.0215492

Edwards, M. D., and Gifford, D. K. (2012). High-resolution genetic mapping with pooled sequencing. *BMC Bioinforma.* 13, S8. doi:10.1186/1471-2105-13-S6-S8

Etherington, G. J., Monaghan, J., Zipfel, C., and MacLean, D. (2014). Mapping mutations in plant genomes with the user-friendly web application CandiSNP. *Plant Methods* 10, 41. doi:10.1186/s13007-014-0041-7

Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., et al. (2013). MutMap+: Genetic mapping and mutant identification without crossing in rice. *PloS one* 8, e68529. doi:10.1371/journal.pone.0068529

Foolad, M. R., Zhang, L. P., and Lin, G. Y. (2001). Identification and validation of QTLs for salt tolerance during vegetative growth in tomato by selective genotyping. *Genome* 44, 444–454. doi:10.1139/g01-030

Garrison, E., and Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*, 3907. arXiv, 1207.

Ghavidel, F. Z., Claesen, J., and Burzykowski, T. (2015). A nonhomogeneous hidden Markov model for gene mapping based on next-generation sequencing data. *J. Comput. Biol.* 22, 178–188. doi:10.1089/cmb.2014.0258

Giovannoni, J. J., Wing, R. A., Ganal, M. W., and Tanksley, S. D. (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Res.* 19, 6553–6558. doi:10.1093/nar/19.23.6553

Gu, X. H., Jiang, D. L., Huang, Y., Li, B. J., Chen, C. H., Lin, H. R., et al. (2018). Identifying a major QTL associated with salinity tolerance in Nile Tilapia using QTL-Seq. *Mar. Biotechnol.* 20, 98–107. doi:10.1007/s10126-017-9790-4

Haase, N. J., Beissinger, T., Hirsch, C. N., Vaillancourt, B., Deshpande, S., Barry, K., et al. (2015). Shared genomic regions between derivatives of a large segregating population of maize identified using bulked segregant analysis sequencing and traditional linkage analysis. *G3* 5, 1593–1602. doi:10.1534/g3.115.017665

Hu, M. J., Zhang, H. P., Liu, K., Cao, J. J., Wang, S. X., Jiang, H., et al. (2016). 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

Hu, T., Wei, Q., Wang, J., Wang, W., Hu, H., Yan, Y., et al. (2022). *Identification of quantitative trait loci controlling radish root shape using QTL-seq* [Preprint from Research Square]. Durham UK: Europe PMC. doi:10.21203/rs.3.rs-1500743/v1

Hu, W., Suo, F., and Du, L. L. (2015). Bulk segregant analysis reveals the genetic basis of a natural trait variation in fission yeast. *Genome Biol. Evol.* 7 (12), 3496–3510. doi:10.1093/gbe/evv238

Huang, L., Tang, W., Bu, S., and Wu, W. (2020). Brm: A statistical method for QTL mapping based on bulked segregant analysis by deep sequencing. *Bioinformatics* 36, 2150–2156. doi:10.1093/bioinformatics/btz861

Hundal, J., Miller, C. A., Griffith, M., Griffith, O. L., Walker, J., Kiwala, S., et al. (2016). Cancer immunogenomics: Computational neoantigen identification and vaccine design. *Cold Spring Harb. symposia quantitative Biol.* 81, 105–111. doi:10.1101/sqb.2016.81.030726

Illa-Berenguer, E., Van Houten, J., Huang, Z., and van der Knaap, E. (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *Theor. Appl. Genet.* 128, 1329–1342. doi:10.1007/s00122-015-2509-x

Imerovski, I., Dedić, B., Cvejić, S., Miladinović, D., Jocić, S., Owens, G. L., et al. (2019). BSA-seq mapping reveals major QTL for broomrape resistance in four sunflower lines. *Mol. Breed.* 39, 41. doi:10.1007/s11032-019-0948-9

Jacoby, W. G. (2000). Loess:. *Elect. Stud.* 19, 577–613. doi:10.1016/s0261-3794(99)00028-1

Kaminski, K. P., Kørup, K., Andersen, M. N., Sønderkær, M., Andersen, M. S., Kirk, H. G., et al. (2016). Next generation sequencing bulk segregant analysis of potato support that differential flux into the cholesterol and stigmasterol metabolite pools is important for steroidal glycoalkaloid content. *Potato Res.* 59, 81–97. doi:10.1007/s11540-015-9314-4

Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Med.* 12, 91. doi:10.1186/s13073-020-00791-w

Kumar, R., Janila, P., Vishwakarma, M. K., Khan, A. W., Manohar, S. S., Gangurde, S. S., et al. (2020). Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. *Plant Biotechnol. J.* 18 (4), 992–1003. doi:10.1111/pbi.13266

Lan, Y., Zhou, D., Zhang, H., and Lai, S. (2017). "Development of early warning models," in *Early warning for infectious disease outbreak* (Academic Press), 35–74.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923

Lei, L., Zheng, H., Bi, Y., Yang, L., Liu, H., Wang, J., et al. (2020). Identification of a major qtl and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* L.) using QTL-Seq and RNA-Seq. *Rice* 13 (1), 55. doi:10.1186/s12284-020-00416-1

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. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26 (5), 589–595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi:10.1093/bioinformatics/bty191

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30, 2843–2851. doi:10.1093/bioinformatics/btu356

Liang, T., Chi, W., Huang, L., Qu, M., Zhang, S., Chen, Z. Q., et al. (2020). Bulked segregant analysis coupled with whole-genome sequencing (BSA-Seq) mapping identifies a novel pi21 haplotype conferring basal resistance to rice blast disease. *Int. J. Mol. Sci.* 21, 2162. doi:10.3390/ijms21062162

Lin, M., Sun, S., Fang, J., Qi, X., Sun, L., Zhong, Y., et al. (20212021). BSR-Seq analysis provides insights into the cold stress response of *Actinidia arguta* F_1 populations. *BMC Genomics* 22, 72. doi:10.1186/s12864-021-07369-9

Liu, S., Yeh, C. T., Tang, H. M., Nettleton, D., and Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PloS One* 7, e36406. doi:10.1371/journal.pone.0036406

Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., et al. (2014). QTL-seq identifies an early flowering QTL located near Flowering Locus T in cucumber. *Theor. Appl. Genet.* 127, 1491–1499. doi:10.1007/s00122-014-2313-z

Luo, H., Pandey, M. K., Khan, A. W., Guo, J., Wu, B., Cai, Y., et al. (2019). Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 17 (7), 1248–1260. doi:10.1111/pbi.13050

Magwene, P. M., Willis, J. H., and Kelly, J. K. (2011). The statistics of bulk Segregant analysis using next generation sequencing. *PLoS Comput. Biol.* 7, e1002255. doi:10.1371/journal.pcbi.1002255

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U. S. A.* 88, 9828–9832. doi:10.1073/pnas.88.21.9828

Navabi, A., Mather, D. E., Bernier, J., Spaner, D. M., and Atlin, G. N. (2009). QTL detection with bidirectional and unidirectional selective genotyping: Marker-based and trait-based analyses. *Theor. Appl. Genet.* 118 (2), 347–358. doi:10.1007/s00122-008-0904-2

Navarro-Escalante, L., Zhao, C., Shukle, R., and Stuart, J. (2020). BSA-seq discovery and functional analysis of candidate Hessian fly (*Mayetiola destructor*) avirulence genes. *Front. Plant Sci.* 11, 956. doi:10.3389/fpls.2020.00956

Pandey, M. K., Khan, A. W., Singh, V. K., Vishwakarma, M. K., Shasidhar, Y., Kumar, V., et al. (2017). QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 15, 927–941. doi:10.1111/pbi.12686

Park, M., Lee, J. H., Han, K., Jang, S., Han, J., Lim, J. H., et al. (2019). A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of *Capsicum chinense* revealed using QTL-seq and RNA-seq. *Theor. Appl. Genet.* 132 (2), 515–529. doi:10.1007/s00122-018-3238-8

Pujol, M., Alexiou, K. G., Fontaine, A. S., Mayor, P., Miras, M., Jahrmann, T., et al. (2019). Mapping cucumber vein yellowing virus resistance in cucumber (*Cucumis sativus* L.) by using BSA-seq analysis. *Front. Plant Sci.* 10, 1583. doi:10.3389/fpls.2019.01583

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., Wilkie, A. O. M., et al. (2014). Integrating mapping-assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* 46, 912–918. doi:10.1038/ng.3036

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., et al. (2009). SHOREmap: Simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* 6, 550–551. doi:10.1038/nmeth0809-550

Singh, V., Sinha, P., Obala, J., Khan, A. W., Chitikineni, A., Saxena, R. K., et al. (2022). QTL-seq for the identification of candidate genes for days to flowering and leaf shape in pigeonpea. *Heredity* 128, 411–419. doi:10.1038/s41437-021-00486-x

Song, H., Huang, Y., and Gu, B. (2020). QTL-Seq identifies quantitative trait loci of relative electrical conductivity associated with heat tolerance in bottle gourd (*Lagenaria siceraria*). *PloS one* 15 (11), e0227663. doi:10.1371/journal.pone.0227663

Song, J., Li, Z., Liu, Z., Guo, Y., and Qiu, L. J. (2017). Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Front. Plant Sci.* 8, 919. doi:10.3389/fpls.2017.00919

Sun, J., Yang, L., Wang, J., Liu, H., Zheng, H., Xie, D., et al. (2018). Identification of a cold-tolerant locus in rice (*Oryza sativa* L.) using bulked segregant analysis with a next-generation sequencing strategy. *Rice* 11, 24. doi:10.1186/s12284-018-0218-1

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013a). 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

Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., et al. (2013b). MutMap-Gap: Whole-genome re-sequencing of mutant F_2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene *Pii*. *New Phytol.* 200, 276–283. doi:10.1111/nph.12369

Tang, W., Huang, L., Bu, S., Tamiru, M., Abe, A., Mitsuoka, C., et al. (2018). Estimation of QTL heritability based on pooled sequencing data. *Bioinformatics* 34, 978–984. doi:10.1093/bioinformatics/btx703

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., and Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034. doi:10.1093/bioinformatics/btv098

Tribhuvan, K. U., Kumar, K., Sevanthi, A. M., and Gaikwad, K. (2018). MutMap: A versatile tool for identification of mutant loci and mapping of genes. *Indian J. Plant Physiol.* 23, 612–621. doi:10.1007/s40502-018-0417-1

Trindade de Carvalho, B., Holt, S., Souffriau, B., Lopes Brandão, R., Foulquié-Moreno, M. R., and Thevelein, J. M. (2017). Identification of novel alleles conferring superior production of rose flavor phenylethyl acetate using polygenic analysis in yeast. *MBio* 8 (6), e01173-17. doi:10.1128/mBio.01173-17

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 43, 01–11. doi:10.1002/0471250953.bi1110s43

Vogel, G., LaPlant, K. E., Mazourek, M., Gore, M. A., and Smart, C. D. (2021). A combined BSA-Seq and linkage mapping approach identifies genomic regions associated with Phytophthora root and crown rot resistance in squash. *Theor. Appl. Genet.* 134 (4), 1015–1031. doi:10.1007/s00122-020-03747-1

Wang, R. J., Sun, Y., Bao, L. S., Zhang, J., Jiang, Y., Yao, J., et al. (2013). Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. *BMC Genomics* 14, 929. doi:10.1186/1471-2164-14-929

Wang, Z., Qi, Q., Lin, Y., Guo, Y., Liu, Y., Wang, Q., et al. (2019). QTL analysis reveals genomic variants linked to high-temperature fermentation performance in the industrial yeast. *Biotechnol. Biofuels* 12 (1), 59. doi:10.1186/s13068-019-1398-7

Wen, J., Jiang, F., Weng, Y., Sun, M., Shi, X., Zhou, Y., et al. (2019). Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. *BMC Plant Biol.* 19 (1), 398. doi:10.1186/s12870-019-2008-3

Wenger, J. W., Schwartz, K., and Sherlock, G. (2010). Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS Genet.* 6, e1000942. doi:10.1371/journal.pgen.1000942

Wu, L., Cui, Y., Xu, Z., and Xu, Q. (2020). Identification of multiple grain shape-related loci in rice using bulked segregant analysis with high-throughput sequencing. *Front. Plant Sci.* 11, 303. doi:10.3389/fpls.2020.00303

Xue, H., Shi, T., Wang, F., Zhou, H., Yang, J., Wang, L., et al. (2017). Interval mapping for red/green skin color in Asian pears using a modified QTL-seq method. *Hortic. Res.* 4, 17053. doi:10.1038/hortres.2017.53

Yang, Y., Lu, X., Yan, B., Li, B., Sun, J., Guo, S., et al. (2013a). Bottle gourd rootstock-grafting affects nitrogen metabolism in NaCl-stressed watermelon leaves and enhances short-term salt tolerance. *J. Plant Physiol.* 170, 653–661. doi:10.1016/j.jplph.2012.12.013

Yang, Z., Huang, D., Tang, W., Zheng, Y., Liang, K., Cutler, A. J., et al. (2013b). Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *Plos one* 8, e68433. doi:10.1371/journal.pone.0068433

Yaobin, Q., Peng, C., Yichen, C., Yue, F., Derun, H., Tingxu, H., et al. (2018). QTL-Seq identified a major QTL for grain length and weight in rice using near isogenic F_2 population. *Rice Sci.* 25 (3), 121–131. doi:10.1016/j.rsci.2018.04.001

Ye, S., Yan, L., Ma, X., Chen, Y., Wu, L., Ma, T., et al. (2022). Combined BSA-seq based mapping and RNA-seq profiling reveal candidate genes associated with plant architecture in *Brassica napus*. *Int. J. Mol. Sci.* 23, 2472. doi:10.3390/ijms23052472

Yuexiong, Z., Gang, Q., Qianqian, M., Minyi, W., Xinghai, Y., Zengfeng, Ma., 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

Zhan, Z., Jiang, Y., Shah, N., Hou, Z., Zhou, Y., Dun, B., et al. (2020). Association of clubroot resistance locus PbBa8. 1 with a linkage drag of high erucic acid content in the seed of the European turnip. *Front. Plant Sci.* 11, 810. doi:10.3389/fpls.2020.00810

Zhang, B., Qi, F., Hu, G., Yang, Y., Zhang, L., Meng, J., et al. (2021). BSA-seq-based identification of a major additive plant height QTL with an effect equivalent to that of Semi-dwarf 1 in a large rice F_2 population. *Crop J.* 9, 1428–1437. doi:10.1016/j.cj.2020.11.011

Zhang, H., Wang, X., Pan, Q., Li, P., Liu, Y., Lu, X., et al. (2019). QTG-Seq accelerates QTL fine mapping through QTL partitioning and whole-genome sequencing of bulked segregant samples. *Mol. Plant* 12 (3), 426–437. doi:10.1016/j.molp.2018.12.018

Zhang, J., and Panthee, D. R. (2020). PyBSASeq: A simple and effective algorithm for bulked segregant analysis with whole-genome sequencing data. *BMC Bioinforma.* 21 (1), 99. doi:10.1186/s12859-020-3435-8

Zhang, L. P., Lin, G. Y., Niño, D., and Foolad, M. R. (2003). Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* × *L. hirsutum* cross by selective genotyping. *Mol. Breed.* 12, 3–19.

Zhang, X., Wang, W., Guo, N., Zhang, Y., Bu, Y., Zhao, J., et al. (2018). Combining QTL-seq and linkage mapping to fine map a wild soybean allele characteristic of greater plant height. *BMC Genomics* 19 (1), 226. doi:10.1186/s12864-018-4582-4

Zhu, T., Wu, L., He, H., Song, J., Jia, M., Liu, L., et al. (2020). Bulked segregant RNA-Seq reveals distinct expression profiling in Chinese wheat cultivar Jimai 23 responding to powdery mildew. *Front. Genet.* 11, 474. doi:10.3389/fgene.2020.00474

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 10 August 2022
DOI 10.3389/fgene.2022.942079

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Vivekanand Tiwari,
Agricultural Research Organization (ARO), Israel
Bhagwat Dadarao Nawade,
Chonnam National University, South Korea

*CORRESPONDENCE
Gyan P. Mishra,
gyan.gene@gmail.com
Shiv Kumar,
SK.Agrawal@cgiar.org
Harsh Kumar Dikshit,
harshgeneticsiari@gmail.com

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

RECEIVED 12 May 2022
ACCEPTED 11 July 2022
PUBLISHED 10 August 2022

CITATION
Dutta H, Mishra GP, Aski MS, Bosamia TC, Mishra DC, Bhati J, Sinha SK, Vijay D, C. T. MP, Das S, Pawar PA-M, Kumar A, Tripathi K, Kumar RR, Yadava DK, Kumar S and Dikshit HK (2022), Comparative transcriptome analysis, unfolding the pathways regulating the seed-size trait in cultivated lentil (Lens culinaris Medik.). *Front. Genet.* 13:942079. doi: 10.3389/fgene.2022.942079

COPYRIGHT
© 2022 Dutta, Mishra, Aski, Bosamia, Mishra, Bhati, Sinha, Vijay, C. T., Das, Pawar, Kumar, Tripathi, Kumar, Yadava, Kumar and Dikshit. 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.

Comparative transcriptome analysis, unfolding the pathways regulating the seed-size trait in cultivated lentil (*Lens culinaris* Medik.)

Haragopal Dutta[1], Gyan P. Mishra[1]*, Muraleedhar S. Aski[1], Tejas C. Bosamia[2], Dwijesh C. Mishra[3], Jyotika Bhati[3], Subodh Kumar Sinha[4], Dunna Vijay[5], Manjunath Prasad C. T.[5], Shouvik Das[6], Prashant Anupama-Mohan Pawar[6], Atul Kumar[5], Kuldeep Tripathi[7], Ranjeet Ranjan Kumar[8], Devendra Kumar Yadava[9], Shiv Kumar[10]* and Harsh Kumar Dikshit[1]*

[1]Division of Genetics, Indian Agricultural Research Institute, New Delhi, India, [2]Plant Omics Division, Central Salt and Marine Chemicals Research Institute, Bhavnagar, India, [3]Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, New Delhi, India, [4]ICAR-National Institute for Plant Biotechnology, New Delhi, India, [5]Division of Seed Science and Technology, Indian Agricultural Research Institute, New Delhi, India, [6]Laboratory of Plant Cell Wall Biology, Regional Centre for Biotechnology, Faridabad, India, [7]Germplasm Evaluation Division, National Bureau of Plant Genetic Resources, New Delhi, India, [8]Division of Biochemistry, Indian Agricultural Research Institute, New Delhi, India, [9]Division of Crop Science, Indian Council of Agricultural Research, New Delhi, India, [10]South Asia and China Program, International Center for Agricultural Research in the Dry Areas, NASC Complex, New Delhi, India

Market class, cooking time, quality, and milled grain yield are largely influenced by the seed size and shape of the lentil (*Lens culinaris* Medik.); thus, they are considered to be important quality traits. To unfold the pathways regulating seed size in lentils, a transcriptomic approach was performed using large-seeded (L4602) and small-seeded (L830) genotypes. The study has generated nearly 375 million high-quality reads, of which 98.70% were properly aligned to the reference genome. Among biological replicates, very high similarity in fragments per kilobase of exon per million mapped fragments values ($R > 0.9$) showed the consistency of RNA-seq results. Various differentially expressed genes associated mainly with the hormone signaling and cell division pathways, transcription factors, kinases, etc. were identified as having a role in cell expansion and seed growth. A total of 106,996 unigenes were used for differential expression (DE) analysis. String analysis identified various modules having certain key proteins like Ser/Thr protein kinase, seed storage protein, DNA-binding protein, microtubule-associated protein, etc. In addition, some growth and cell division–related micro-RNAs like miR3457 (cell wall formation), miR1440 (cell proliferation and cell cycles), and miR1533 (biosynthesis of plant hormones) were identified as having a role in seed size determination. Using RNA-seq data, 5254 EST-SSR primers were generated as a source for future studies aiming for the identification of linked markers. *In silico*

validation using Genevestigator® was done for the Ser/Thr protein kinase, ethylene response factor, and Myb transcription factor genes. It is of interest that the xyloglucan endotransglucosylase gene was found differentially regulated, suggesting their role during seed development; however, at maturity, no significant differences were recorded for various cell wall parameters including cellulose, lignin, and xylose content. This is the first report on lentils that has unfolded the key seed size regulating pathways and unveiled a theoretical way for the development of lentil genotypes having customized seed sizes.

KEYWORDS

RNA-seq, transcription factors, signal transduction pathway, *Lens culinaris*, seed parameters

Introduction

Lentil (*Lens culinaris* ssp. *culinaris* Medik.) or *Masur* (n = x = 7) is a self-pollinated, annual legume crop with a genome size of 4.2 Gbp (Arumuganathan and Earle, 1991; Dikshit et al., 2022). In 2019, the global lentil yield recorded was 5.73 million tons from a 4.8 million hectare area, and Canada was the largest producer (2.17 mt from 1.49 mha), followed by India (1.23 mt from 1.36 mha). Compared to the global productivity (1194.6 kg/ha), the lentil productivity of India is very poor (901 kg/ha) (FAO 2022). Madhya Pradesh and Uttar Pradesh are the largest lentil-producing states in India and are followed by West Bengal and Bihar (Mishra et al., 2022). For a very large section of the population residing mostly in developing countries, lentil is considered an inexpensive nutrition source, especially for proteins, carbohydrates, vitamins, dietary fibers, and micronutrients (Mishra et al., 2020a; Priti et al., 2021; Priti et al., 2022).

The lentil industry always tries to maintain the quality of the lentil, and seed size and shape is considered the key trait for obtaining the optimum quality. Cooking time and dehulling efficiency, which significantly govern the market preference are greatly influenced by the seed size and shape (Wang, 2008; Singh et al., 2022). Moreover, seed size and seed number govern the overall seed yield of any crop, which is mainly determined at the early seed developmental stages (Ruan et al., 2012; Mishra et al., 2020b) and is very precisely controlled by maternal and filial tissues (Weber et al., 2005). Advancements in understanding the seed development-associated pathway–related genes in lentils have been attributed to the availability of several markers, genomic resources, and cost-effective next–generation sequencing (NGS) tools. At present, the latest version (version 2) of lentil genome has been released (Ramsay et al., 2021). Thus, comparative RNA-seq analysis was aimed to identify the DEGs (differentially expressed genes) and pathways from the lentil genotypes differing significantly for the seed size during their early seed developmental stage.

Many reports demonstrating DE of various genes in the genotypes differing for the seed size in crops like peanut (Li et al., 2021), soybean (Du et al., 2017; Lu et al., 2017), etc. are available. Moreover, several complex regulatory networks including transcription factors, sugar, hormone signaling, and metabolic pathways regulating seed development are known in various plant species (Ruan et al., 2012; Du et al., 2017). In addition, molecular markers like EST-SSRs can be developed using the RNA seq data derived from the genotypes differing in the seed size. These novel markers can be used for the identification of linked markers using some mapping population for its application in the marker-assisted selection for seed size traits in lentils (Bosamia et al., 2015). Micro-RNAs (miRNAs) are the class of small noncoding RNAs (~21-nucleotide long) that control the post-transcriptional expression of mRNA in lentils (Mishra et al., 2022). The miRNAs interact with the target mRNAs after getting loaded into the Ago-proteins (Dueck et al., 2012) and thus prevent gene expression (Winter et al., 2009). Using recent bioinformatics tools, putative miRNA target sites can be identified from the RNA-seq dataset, which will help in better understanding their role in the regulation of seed developmental pathways resulting in differential seed sizes in lentils.

The seed size is directly regulated by the chemical composition of cell wall in which the most distinctive constituent is cellulose (Costa and Plazanet, 2016), whereas lignin is the next most abundant biopolymer (Kathirselvam et al., 2019) and xylan is one of the highly abundant hemicelluloses (Scheller and Ulvskov, 2010). Thus, deciphering the complex cell wall dynamics associated with the regulation of seed size will help in understanding their relative function during seed development. Therefore, the RNA-seq approach was used for the identification of regulatory gene networks and pathways controlling seed size in lentils using genotypes differing significantly for the seed size.

Materials and methods

Plant materials, sampling, and seed parameter analysis using VideometerLab

Two lentil genotypes differing significantly in seed size, L830 (small-seeded; mean 1000 seed weight = 20.0g) and L4602 (large-

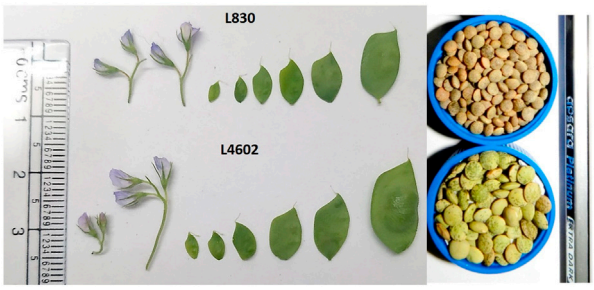


FIGURE 1
Different seed developmental stages and the developed seeds of L830 and L4602 genotypes.

seeded; mean 1000 seed weight = 42.13g) were used for the RNA-seq analysis (Figure 1; Table 1). Detailed seed phenotyping was done using the VideometerLab 4.0 instrument (Videometer A/S, Denmark), which acquires morphological and spectral information using strobed LED technology in the UV, visible, and NIR wavelengths (total 19 wavelengths; 365–970 nm) (Figure 2). The VideometerLab vision system was used to capture the images at 2056 × 2056 pixels of 30 seeds placed in special 3D printed plates, which were customized to fit our seed samples. The data generated were segmented, quantified, and plotted using custom-designed software (VideometerLab software ver. 2.13.83), which ultimately provided a vast array of information such as seed area, length, and width (Shrestha et al., 2015).

The plants were grown in the National Phytotron Facility located at Indian Agricultural Research Institute, New Delhi, India (latitude: 28.6412°N; longitude: 77.1627°E; and altitude: 228.61-m AMSL) in 30 cm (diameter) pots (three seeds per pot) containing grow media consisted of coco peat: vermiculite: sand in 1:2:1 ratio. As we aimed to find the DEGs regulating the seed size at a very early stage, the flower buds were tagged on the day of fertilization and developing pods were collected 15 days after fertilization, and developing seeds were removed very carefully from the pods. These seed samples were then stored in RNAlater at −80°C before RNA extraction. A total of four samples were used which consisted of L830 (small-seeded) and L4602 (large-seeded) samples in two biological replications each.

RNA extraction, cDNA library construction, and Illumina sequencing

As per the manufacturer's protocol, total RNA was extracted from the seed samples of two lentil genotypes L830 and L4602 using the RNeasy plant mini kit (Qiagen, Hilden, Germany). DNAse I (Thermo, United States) treatment was given twice to remove the genomic DNA; afterward, RNA was dissolved in nuclease-free water. The RNA quality and integrity were measured using Bioanalyzer 2100 RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, United States) and on 2% agarose gel, respectively, and then high-quality RNA (1.5 µg; OD 260/280 = 2.0–2.1; OD 260/230 = ≥2.0–2.3; RIN value ≥ 7.0) were taken from two biological replicates (from two lentil genotypes) for the construction of four cDNA libraries using TruSeq mRNA Library Prep kit (Illumina Inc., United States). Using Illumina HiSeq 2500 platform 100 bp paired-end (PE) reads (100 × 2 = 200bp) were generated from the cDNA libraries at AgriGenome Labs Pvt. Ltd., Hyderabad, Telangana, India. The samples were labeled as L830 and L4602, in two replications (Rep-1 and Rep-2) each. The raw sequence data of these four PE libraries were submitted to the National Center for Biotechnology Information SRA database (Accession number PRJNA800200).

TABLE 1 Descriptive statistics for various seed traits of the genotypes L4602 and L830 using t-test.

S. No.	Variables	L4602		L830		Statistical significance
		Mean ± SD	SE mean	Mean ± SD	SE mean	
1	1000 seed weight (g)	42.13 ± 1.21	0.70	20.90 ± 1.82	1.10	HS
2	Area (mm^2)	22.59 ± 1.17	0.262	11.02 ± 0.68	0.152	HS
3	Length (mm)	5.57 ± 0.169	0.038	3.82 ± 0.121	0.027	HS
4	Width (mm)	5.24 ± 0.179	0.04	3.71 ± 0.117	0.0262	HS
5	Width/Length	0.9399 ± 0.027	0.006	0.9723 ± 0.015	0.003	HS
6	Compactness (Circle)	0.9379 ± 0.025	0.006	0.9688 ± 0.014	0.003	HS
7	Compactness (Ellipse)	0.9983 ± 0.002	0.0003	0.999 ± 0.0004	0.00008	NS
8	Width/Area	0.2320 ± 0.006	0.001	0.3377 ± 0.011	0.002	HS
9	Volume (mm^3)	0.007 ± 0.0002	0.00005	0.0048 ± 0.0001	0.00003	HS
10	Perimeter (mm)	15.477 ± 0.452	0.101	10.663 ± 0.369	0.082	HS

Where total count or replication (N) = 20; HS: Highly significant ($p \leq 0.01$); NS, Not significant

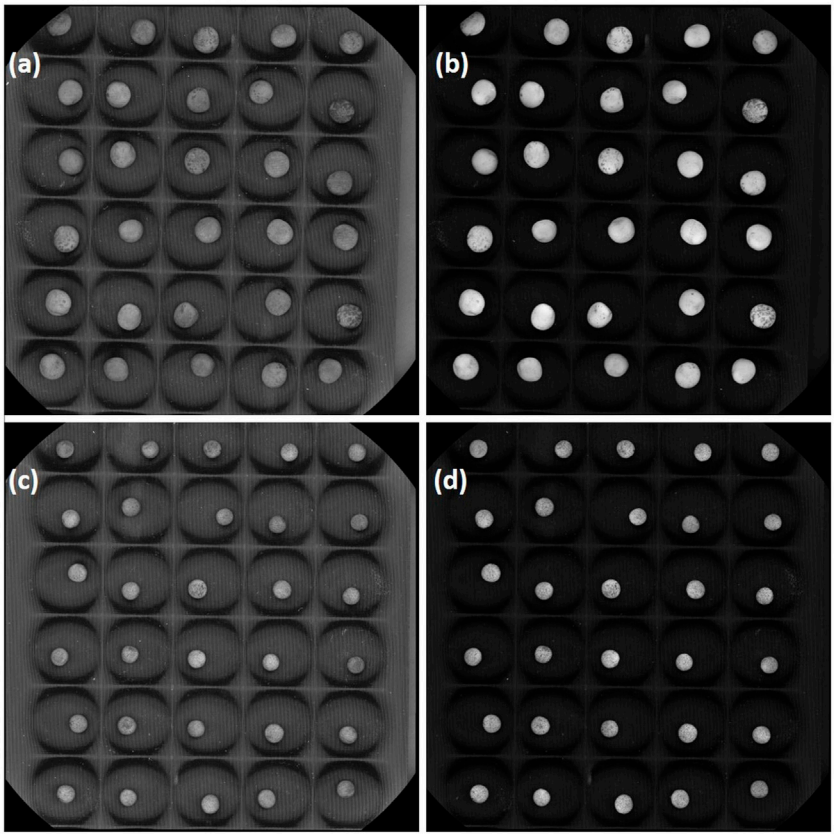


FIGURE 2
A representative mature seed photograph of lentil genotypes **(A,B)** L4602 and **(C,D)** L830 was taken at two different wavelengths [**(A,C)** at 375 nm; **(B,D)** at 590 nm] using VideometerLab.

Transcriptome assembly and annotation, and GO

FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for the analysis of the raw reads, whereas AdapterRemoval2 (ver. 2.2.0; https://github.com/MikkelSchubert/adapterremoval) was used to trim the low-quality bases (Phred score ≤ 30) and adapter sequences. Bowtie2-based SILVA database was used for the removal of rRNA sequences, and then, GC content, Q30, and total clean reads were determined in the high-quality reads (HQRs). For the alignment of trimmed reads, Bowtie2 ver. 2.2.2.9 (http://Bowtie2-bio.sourceforge.net/index.shtml) was used and unigenes with ≥200 bp transcript length were assembled for the estimation of transcript expression and their downstream annotations. For alignment, the lentil reference genome (Ver. 2) was used (Ramsay et al., 2021). Read count normalization was done, and transcript expression levels were determined using FPKM (fragments per kilobase of exon per million mapped fragments) analysis. The annotation of assembled unigenes was done using UniProt Plant Database and Gene Ontology (GO) terms were identified using BLAST2GO (https://www.blast2go.com; Conesa et al., 2005).

Differential gene expression and GO enrichment analysis

The normalization of gene expression values was done using FPKM and fold change (FC), whereas expression was calculated using Cuffdiff (http://cole-trapnelllab.github.io/cufflinks/cuffdiff/). Differential gene expression analysis was performed by considering the individual transcript expression level counts in small- (L830) and large-seeded (L4602) samples using the edgeR (http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html) program (Robinson et al., 2010). The false discovery rate (FDR ≤ 0.05) and Log2 Fold Change (Log2FC +2/−2) were used as a threshold for the identification of DEGs as upregulated or downregulated. The GO terms for all assembled unigenes were extracted wherever possible. “Goseq” of “R package” (https://bioconductor.org/packages/release/bioc/html/goseq.html) was used for the GO enrichment analysis, and the

identified GO terms were grouped into biological process, cellular component, and molecular function categories. Afterward, a scatter plot was generated and visualized for the GO-enriched terms using Revigo (http://revigo.irb.h) (Supek et al., 2011). The information available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (http://www.genome.jp/kegg/) database was used to perform the metabolic pathway analysis of the assembled unigenes (Kanehisa and Goto, 2000).

miRNA and miRNA target sites identification

Precursor miRNA sequences were identified using C-mii software, by searching the sequences against the miRbase database and identifying the hairpin-like secondary structures (Numnark et al., 2012). Then, putative miRNAs were identified having parameters like 1) predicted mature miRNAs length should be 18–24 nt; 2) a maximum of two mismatches over known mature miRNAs; 3) localization of mature miRNA in one arm of the predicted stem-loop structure; 4) not >5 mismatches between miRNA and guide miRNA sequence in the stem-loop structure; 5) presence of 30–70% A + U content in the miRNAs; 6) and a highly negative minimal folding free energy and minimal free energy index value of the secondary structure. Putative target genes were identified using psRNATarget (https://www.zhaolab.org/psRNATarget/) against the available *Arabidopsis thaliana* annotated data to understand the biological functions of the miRNAs (Dai et al., 2018) and miRNA target network was built using Cytoscape software (Shannon et al., 2003).

Protein–protein interaction network

String network analysis was performed to understand the PPI network of the identified DEGs. The identified target list was first uploaded into the STRING database (https://string-db.org/) (Szklarczyk et al., 2015); afterward, highly interconnected proteins were identified using the MCODE plugin of Cytoscape v3.7.0 (Shannon et al., 2003). The final prediction of the PPI networks was based on *Medicago truncatula* as this showed very high sequence similarity.

In silico validation analysis using Genevestigator®

The DEGs were filtered based on growth- and development-related GO terms and were then narrowed down as per their FDR value (<0.05). Peptide sequences of the filtered DEGs were extracted from the Uniprot database using BLASTP. The associated reference genes having maximum stability of expression against *Medicago truncatula* were identified by submitting the shortlisted target accessions (unique) as input in the Genevestigator® software (Zimmermann et al., 2004). For the grouping of target genes having similar profiles across the above conditions, a hierarchical clustering tool was used and a Log2 ratio change value was obtained for the identified genes and the top 15 perturbations were presented as a heat map.

SSR prediction and EST-SSR identification

The unigenes from present RNA-seq data were used for the SSR prediction using the MIcroSAtellite Identification Tool (MISA) (Thiel et al., 2003) and the development of genic-SSR markers. To avoid any already known SSR primers, primer sequences already available in the public domain were searched and unigenes were removed from our RNA-seq data using an in-house Perl script (Bosamia et al., 2015). Standalone primer3 ver. 0.4.0 tool was used for primer designing as per the default parameters (Untergasser et al., 2012). Furthermore, a set of developmental process–related SSRs were siphoned considering the GO terms associated with various developmental processes like embryo development, flower development, endosperm development, cell cycle process, seed growth, and reproductive shoot system development.

Preparation of alcohol insoluble residue sample

The lentil seeds (600 mg) were crushed, flash frozen in liquid nitrogen, and then ground in Qiagen TissueLyser II (at 30 Hz for 2–3 min) to a fine powder. The 100-mg fine powder was incubated in 5 ml of 80% ethanol containing 4.0 mM HEPES buffer (pH 7.5) at 70°C for 30 min, cooled on ice, and centrifuged at 10000 rpm for 15 min. The supernatant was discarded and the residue was washed with 5 ml of 70% ethanol. The residue was further suspended in 5 ml of chloroform: methanol (1:1) for 3 min at room temperature and centrifuged at 14000 rpm for 15 min. The residue was washed with 5-ml acetone, and the pellet was dried in the desiccator and used as an AIR sample for further analysis (Pawar et al., 2017).

Cellulose estimation using the Updegraff method

The Updegraff reagent (acetic acid: nitric acid: water = 8:1:2 v/v) was added to the 2 mg of AIR and incubated at 100°C for 30 min. The mixture was centrifuged (10,000 rpm; 15 min), washed with acetone four times, and the residue was dried overnight. The residue was hydrolyzed in 72% sulphuric acid and glucose was analyzed using anthrone assay (Updegraff 1969). A glucose standard curve was used to calculate the cellulose content.

Estimation of xylose and *O*-acetyl content

A total of 2 mg of AIR was incubated with HCl and NaOH samples for xylose and acetyl content analysis, respectively. The mixture was neutralized with NaOH and HCl, respectively. The xylose and *O*-acetyl content were analyzed using Megazyme K-ACET and K-XYLOSE kits (Rastogi et al., 2022).

Acetyl bromide soluble lignin content

The acetyl bromide solution (25%), which was diluted using acetic acid was incubated at 50°C for 2 h. The solubilized powder was then mixed with NaOH and hydroxylamine hydrochloride. The absorbance was taken at 280 nm, and lignin content was measured as explained in Foster et al. (2010).

Lignin and cellulose estimation through Fourier transform-infrared spectroscopy

The fine lentil seed powder was analyzed for lignin and cellulose content using FT-IR spectroscopy as previously described (Pawar et al., 2017). In a word, a Tensor FT-IR spectrometer (Bruker Optics) equipped with a single-reflectance horizontal ATR cell (ZnSe Optical Crystal, Bruker Optics) was used for the analysis. The spectrum was taken within a range of 600 cm^{-1} to 4000 cm^{-1} with a resolution of 4 cm^{-1}. The standard was prepared with potassium bromide powder. Each sample was measured twice (by removing and putting different aliquots of powder to evaluate their heterogeneity), and each spectrum was the average of 16 scans (Labbe et al., 2005; Canteri et al., 2019).

Statistical analysis

Two sample *t*-test was performed to determine the genotypic variance (among L830 and L4602 genotypes) for the traits such as seed area, length, width, width/length ratio, compactness, width/area ratio, volume, perimeter, lignin, cellulose, and xylose contents. All statistical analyses were conducted using Minitab ver.18 software (Minitab 2020).

Results

Seed parameters

The lentil genotypes L4602 (42.13 g/1000 seeds) and L830 (20.90 g/1000 seeds) selected for the study differed significantly for 1000 seed weight. In addition, many other seed morphological traits were measured using the VideometerLab instrument (like area, length, width, width/length, compactness, width/area, volume, and perimeter) and showed significant variations for these genotypes (Table 1). It is of interest that the mean seed area (mm^2), length (mm), and width (mm) of the genotypes L4602 (22.59, 5.57, 5.24, respectively) and L830 (11.02, 3.82, 3.71, respectively) were found significantly different between these genotypes (Table 1).

RNA sequencing and reference transcriptome assembly

The developing seeds (15 days after pollination) were used for the isolation of RNA from both L830 and L4602 lentil genotypes for analyzing the transcriptional changes during the early seed development stage using the Illumina HiSeq 2500/4000 platform. The samples were named L830-Rep1, L830-Rep2, L4602-Rep1, and L4602-Rep2. More than 448 million reads have been generated from all four sample combinations, having an average GC content of 47.17%. The raw sequence ranged from 81.83 to 127.13 million (for L830 Rep1 and Rep2) and 120.56 to 118.93 million (for L4602 Rep1 and Rep2), respectively. After removal of adapters and rRNA sequences and passing through the low-quality reads nearly 375 million HQRs were obtained having a Phred score of ≥30. The clean reads ranged from 80.01 million to 99.4 million for L830 and from 78.74 million to 117.18 million for L4602 genotypes with an average Q30 quality score of 96.75%. The summary of different RNA-seq parameters of the studied genotypes is presented in Table 2.

The preprocessed and rRNA removed reads were used for reference-based pair–wise alignment and assembly of HQRs with lentil genome using the Trinity program. From the filtered reads, ~98.70% of reads from each sample were properly aligned back to the assembled unigenes, indicating HQR data for the assembly. The total number of PE reads obtained in the small-seeded (L830) sample ranged from 40.91 million to 63.56 million, whereas in the large-seeded (L4602) sample from 59.46 million to 60.28 million. All the assembled transcripts were ≥200bp long with a GC content of 46.24–48.08% (Table 2). The details of transcript length, FPKM range, and GC range are provided in Supplementary Material S1 and Supplementary Table S1.

Differential gene expression analysis and transcript annotation

EdgeR program was used to find the total number of genes expressed in small-seeded (L830) and large-seeded (L4602) genotypes. Out of 512 identified DEGs, 199 showed upregulation and 313 downregulation (Supplementary Table S2). The assembly has generated 120,149 transcripts, and the longest transcript is of 32,136 bp size. A total of 43,369 unigenes were generated, of which 21,038 unigenes showed at least one

TABLE 2 Summary of RNA-seq data in four developing seeds samples of lentils (small- and large-seeded).

Sample name/Attributes	L830-Rep1	L830-Rep2	L4602-Rep1	L4602-Rep2
Mean read quality (Phred score)	39.06	39.32	39.68	39.03
Number of raw reads (PE)	63569309	40918140	60281921	59467429
Number of bases (Gb)	12.71	8.18	12.05	11.89
GC (%)	46.34	46.24	48.02	48.08
% data ≥ Q30	95.39	95.83	97.09	95.02
Read length (bp)	100 × 2	100 × 2	100 × 2	100 × 2
Total read count	127138618	81836280	120563842	118934858
Clean read count	99405074	80015960	78742292	117185022
QC Pass (%)	78.19	97.78	65.31	98.53
Aligned Read Count	94320301	77115384	59243198	112377029
Aligned (%)	94.88	96.38	75.24	95.9

Where Rep1 and Rep2 are the two biological replications.

TABLE 3 Assembled transcript summary.

Annotation Category	Numbers
No. of assembled transcripts (Total: 120,149)	106,996
Longest Read length (bp)	32,136
Mean GC of transcripts (%) (All Assembled Transcripts: 38.96%)	38.83
Total No. of Unigenes considered for BlastX	43,369
No. of Unigenes with a hit in UniProt Plant Database	38,154
No. of Unigenes with a significant hit in UniProt Plant Database	21,038
No. of Unigenes with a significant hit in PMN Database	5,334

significant hit in UniProt Plant Protein Database. The assembled unigenes were annotated using UniProt Plant Database for the identification of lentil gene function using the BLASTX program and 106,996 unigenes were considered for differential gene expression analysis at the threshold level [FDR≤ 0.05; Log2FC +2/−2] (Table 3, Supplementary Table S3). Correlation analysis was done using FPKM values of the two biological replicates of lentil samples and the very high similarity (R > 0.9) verified the consistency among the samples (Supplementary Table S1). The most similar BLASTx matches were recorded for *Medicago truncatula* (12101) followed by *Cicer arietinum* (6470), *Pisum sativum* (638), and *Glycine max* (178) (Supplementary Table S4), and the top 20 BLASTX hits are presented in Supplementary Material S1.

GO enrichment, scatter-plot, and KEGG pathway analysis

The GO terms for all the assembled unigenes were extracted and grouped into biological process (1301), cellular component (403), and molecular function (1521) categories (Supplementary Table S5) using the Blast2GO program (http://www.blast2go.com/). For cellular component, the key enriched GO terms include "integral component of membrane," "nucleus," cytoplasm, "ribosome" etc., whereas for Molecular Functions these include "ATP, metal ion, DNA, and RNA binding," "serine/threonine-protein kinase activity," "hydrolase activity," and transferase activity. DEGs were found enriched for several Biological Processes GO terms like "regulation of transcription," "translation," "protein folding and glycosylation," "DNA replication," "auxin activated signaling pathway," "metal ion transport," and "photosynthesis" (Figure 3; Supplementary Material S1). These signify the presence of substantial differences between large- and small-seeded lentil genotypes during the early seed development stage and also the vital role of identified DEGs resulting in differential seed size.

All the DEGs with GO allotted [3224 Nos] were subjected to gene enrichment analysis and the enriched GO terms are further visualized as a scatter plot (Figure 4). Among various GO terms, "cell division" [GO:0051301], mitotic cell cycle phase transition [GO:0000278], multidimensional cell growth [GO:0009825], and "GDP mannose biosynthetic process" [GO: 0009298] are significantly enriched. To identify the major biological pathways which got altered in lentils when studied in large- and small-seeded genotypes at early seed development stage, the core DEGs were mapped to the KEGG pathways database. The KEGG annotated unigenes [13133 Nos] were found distributed to 25 KEGG pathways and the overrepresented pathways include: "glycolysis/gluconeogenesis," "glyoxylate and dicarboxylate metabolism," "pentose and glucuronate interconversions," "pyruvate metabolism," and "cyanoamino acid metabolism" (Supplementary Tables S6, S7). The details of the upregulated and downregulated KEGG enriched terms (biological process, cellular component, and molecular function) are presented as a pie chart (Supplementary Material S1).

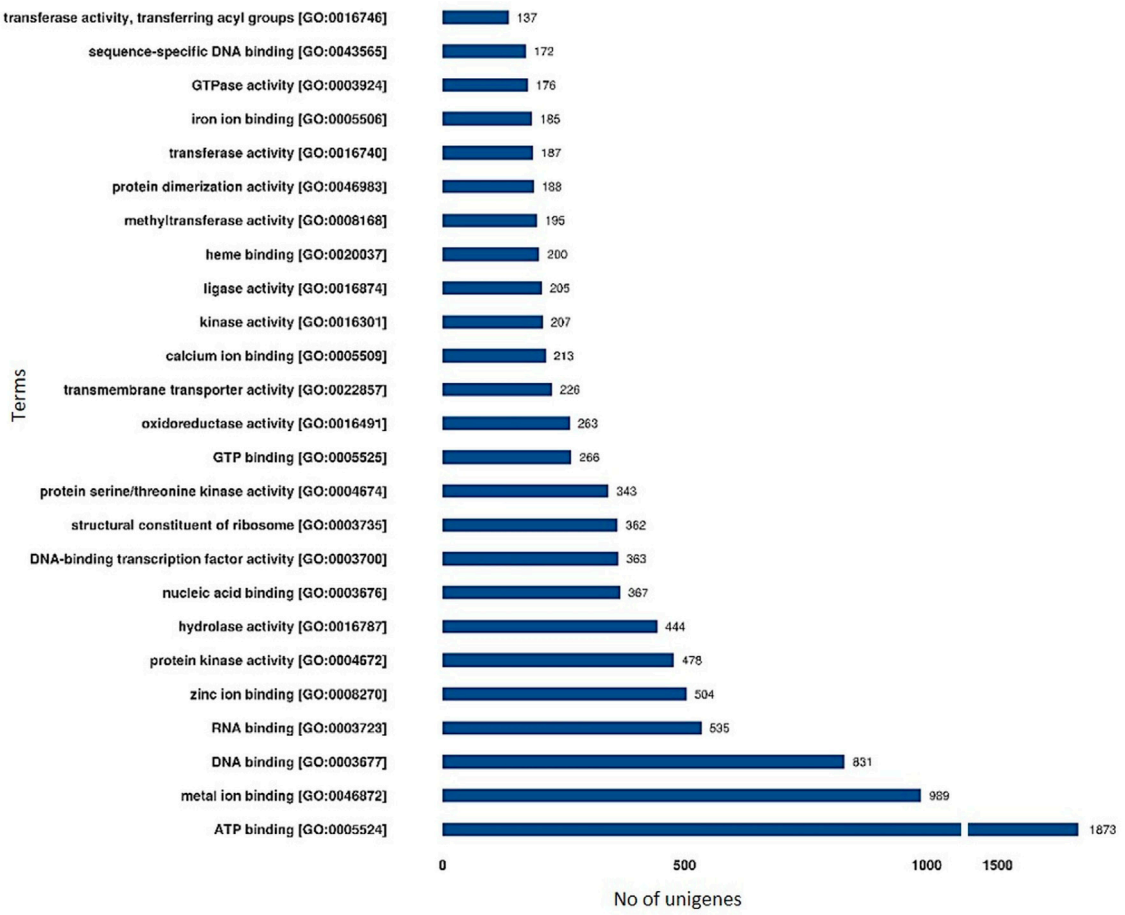


FIGURE 3
Functional annotation of unigenes based on Gene Ontology (GO) categorization of lentil genotypes differing for seed sizes (molecular function). These GO terms are classified into three categories (cellular component, molecular function, and biological processes).

miRNA prediction, miRNA target site identification, and function analysis

Through homology searches of the sequences and miRNA prediction with the c-mii program, we identified 46,195 upregulated and 48,911 downregulated miRNA candidates (Supplementary Table S8). The putative miRNAs varied from 19 to 26 nt in length, with a majority of them being 21 nt in length. Furthermore, based on the multiplicity (≥10 numbers) of the miRNAs identified the key miRNAs identified from upregulated (miR838, miR1440, miR1533, miR3457, miR3637, miR3951, and miR6024) and downregulated DEGs (miR-190, miR902, miR1533, miR3437, miR3520, miR4353, miR5543, miR5576, miR5658, miR5720, miR7511, miR7743, and miR9742) are presented in Table 4. In upregulated category, the esi-miR3457-3p family was the largest family with nearly 400 members followed by gma-miR1533 (256), vvi-miR3637-5p (121), and aly-miR838-3p (100), csi-miR3951 (100), osa-miR1440b (100), sly-miR6024 (100), and stu-miR6024-3p (100). Likewise, from the downregulated DEGs the gma-miR4353 family was the largest (1225) followed by bdi-miR7743-3p (1089), ghr-miR7511 (625), osa-miR5543 (484), ppt-miR902i-3p (441), hme-miR-190 (256), and aly-miR3437-3p (154) (Supplementary Table S8). Among these, a few are growth and cell division–related miRNAs like miR3457 (cell wall formation), miR1440 (cell proliferation and cell cycles), and miR1533 (biosynthesis of plant hormones). This study for the first time could identify the transcriptome-based miRNAs and their targets from the RNA-seq data of two lentil genotypes differing in seed size.

Protein–protein interaction network analysis

The predicted PPI network was broadly categorized into three main functional modules, which corresponded to respective pathways (Supplementary Table S9; Supplementary

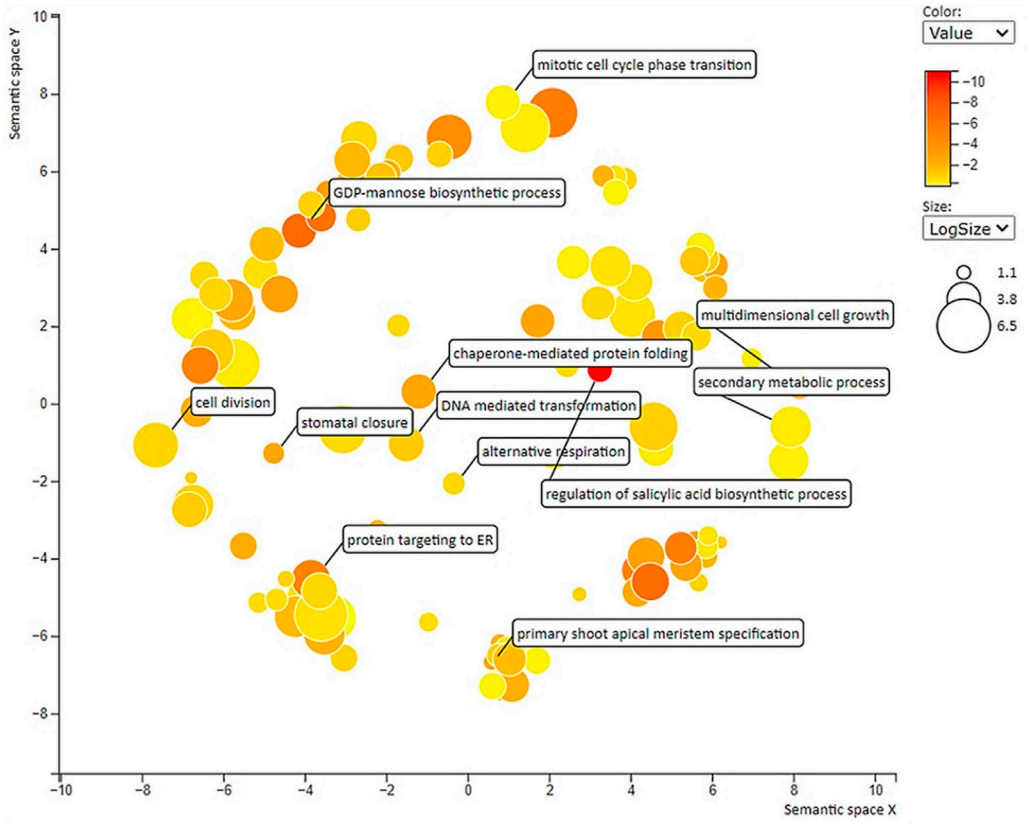


FIGURE 4
Scatter plot showing overrepresented GO term ($p < 0.01$) in all comparisons with labels having seed-size responsive terms. Different shades in circles indicate the difference in *p*-values (as given in the scale), whereas the bubble size indicates the frequency of the GO term.

Material S1). In the upregulated category, module-1 mainly consisted of proteins like Ser/Thr protein kinase, oligopeptide transporter OPT family protein, seed storage protein, etc., whereas module-2 mainly consisted of curved DNA-binding protein, HSP 70 family, Ser/Thr protein kinase family, etc. Module-3 consisted of ABC transporter B family member, somatic embryogenesis receptor kinase, etc. Likewise, from the downregulated category, module-1 consisted of mainly mitochondrial Rho GTPase, ARF GTPase activator proteins, etc., whereas module-2 is comprised mainly of signal recognition particle receptor, aminoacyl-tRNA synthetase, microtubule-associated protein, kinesin family proteins, etc.

Development of novel EST–SSR markers from the lentil transcriptome

Unigenes from RNA-seq data were used for the development of novel and nonredundant EST-SSR markers by searching all the publicly available SSR primers of lentils in unigenes and matched unigenes were removed (Bosamia et al., 2015). SSR prediction was performed using MISA and a total of 14,437 SSRs were predicted from all assembled unigenes belonging to eight classes of microsatellites. Of these, mononucleotide (46.02%) was most abundant followed by trinucleotide (26.38%) and dinucleotide (19.75%) motifs. However, tetranucleotide (1.12%), pentanucleotide (0.28%), and hexanucleotide (0.35%) motifs were least frequent (Table 5). Furthermore, EST-SSR primers were identified by examining 106996 sequences and the total size of examined sequences (bp) was 106969661 bp. Using primer3 software a total of 5254 primer pairs were generated and from that, a set of primers (23 Nos) were identified from the genes having functional relevance to the reproduction process and seed development (Supplementary Table S10).

Validation studies using *in silico* expression analysis

The Genevestigator database is used to study the gene expression details like when, where, and how the gene gets expressed in any living system. This assists in the *in silico*

TABLE 4 List of miRNAs and the target fragment identified from the RNA-seq data generated for seed size variations.

S. No.	miRNA identified	Target details	miRNA aligned fragment	Target aligned fragment	Multiplicity
(a) From downregulated DEGs					
1	gma-miR4353	Lcu.2RBY.Chr1:20126119-20138212	CAAGUCGUAGCCGGUGUU AUUACU	AGACAAGACACCAGCUGC GACUUG	35
2	bdi-miR7743-3p	Lcu.2RBY.Chr7:474327216-474368813	UUUGAACUUUUGUAUUGG AUCUUU	CUAGCUCUAAUACAAAAG UAUGAA	33
3	ghr-miR7511	Lcu.2RBY.Chr7:474327216-474368813	AGAAGUUUUGCAUGUGUA GCUGAG	CAUAACUUCUAAUGCAGA ACUUCU	25
4	osa-miR5543	Lcu.2RBY.Chr5:37183597-37189152	UAUGAAUGGUAUAUUUUCUUG	AAAUAAGAUAUGCCAUUCAAA	22
5	ppt-miR902i-3p	Lcu.2RBY.Chr7:474327216-474368813	UGGAGGAUCUGCAUCGUAAAC	AAUUCUGAUCCAGAUCUUCUA	21
6	aly-miR3437-3p	Lcu.2RBY.Chr5:37183597-37189152	CGGUGGAUCUUGUUUUUUGU	UUGAGAAAUAAGAUCCGCCA	16
7	hme-miR-190	Lcu.2RBY.Chr5:37183597-37189152	AGAUAUGUUUGAUAUUCU UGGU	CAAAAGAAUAUUGAACUGAUCU	16
8	ath-miR5658	Lcu.2RBY.Chr4:218966545-219024978	AUGAUGAUGAUGAUGAUGAAA	UAUUAUUAUUAUUAUUAUUAU	13
9	ahy-miR3520-5p	Lcu.2RBY.Chr6:116449652-116462261	AGGUGAUGGUGAAUAUCU UAUC	UUCAAUAUCUUUACCGUCAGCU	11
10	gma-miR1533	Lcu.2RBY.Chr4:218966545-219024978	AUAAUAAAAAUAAUAAUGA	UUAUUAUUAUUAUUAUUAU	11
11	rgl-miR5576	Lcu.2RBY.Chr4:226243450-226251514	AGAAGUUGGCAUUUGCAA ACACU	ACAAGGAGCAAAUGCCAA CUUCU	11
12	bra-miR5720	Lcu.2RBY.Chr4:218966545-219024978	UUGUGAUUUGGUUGGAAUAUC	CGUAGUUCGACCAAACUACGA	10
13	gma-miR9742	Lcu.2RBY.Chr7:474327216-474368813	UGUGUUGUUUGUUUU GUAGCA	AUAAACAAAACAAACAACAAA	10
(b) From upregulated DEGs					
14	esi-miR3457-3p	Lcu.2RBY.Chr1:538150492-538159831	UAGCUUGUCCUGGGAUUCCGU	AUGGAUUCUCAGCAGAAGCUA	20
15	gma-miR1533	Lcu.2RBY.Chr1:520995772-521054241	AUAAUAAA-AAUAAUAAUGA	UUAUUGUUAUUGUUUGUUGU	16
16	vvi-miR3637-5p	Lcu.2RBY.Chr1:520995772-521054241	AUUUAUGUAUUGUGUUUU GUCGGA	GUGUUCAAAGUUCAAUGU AUGAAA	11
17	aly-miR838-3p	Lcu.2RBY.Chr1:520995772-521054241	UUUUCUUCUUCUUCUUGCACA	AUUGUUAAGACAAGAAGAAAG	10
18	csi-miR3951	Lcu.2RBY.Chr1:520995772-521054241	UAGAUAAAG-AUGAGA GAAAAA	UUUUUUUUUUAUUCUUUA UCUG	10
19	osa-miR1440b	Lcu.2RBY.Chr1:520995772-521054241	UUUAGGAGAGUGGUAUUUGAG	UUUAAAUACCUCUUGCUUAAG	10
20	sly-miR6024	Lcu.2RBY.Chr1:520995772-521054241	UUUUAGCAAGAGUUGUUU UACC	CGUUGAACUACUUUUGCU GGAU	10
21	stu-miR6024-3p	Lcu.2RBY.Chr1:520995772-521054241	UUUUAGCAAGAGUUGUUU UCCC	GUGAAAAUGAUUUUAGAU AAAA	10

validation of the available results and also helps in the formulation of new hypotheses (Zimmermann et al., 2004; Mishra et al., 2022). The genes identified from the developing seeds of the lentil genotypes were of four broad developmental response groups viz., membrane protein, secondary metabolite, signaling molecules, and transcription factors. These genes got validated (Supplementary Table S11) with that of *Medicago truncatula* and the top 15 major perturbations were presented in Figures 5, 6 as a heat map. It is of interest that for membrane protein category: putative transmembrane protein [MTR_5g070330], CASP-like protein [MTR_7g083120], etc. and for secondary metabolite category: glycosyltransferase [MTR_7g076655], ferredoxin-nitrite reductase [MTR_4g086020], etc. could be validated. Likewise, for

TABLE 5 SSR prediction summary as identified from the RNA-seq data.

S. No.	Type of repeat	Number of SSRs predicted
1	Mononucleotide (p1)	6644 (46.02%)
2	Di-nucleotide (p2)	2851 (19.75%)
3	Tri-nucleotide (p3)	3808 (26.38%)
4	Tetra-nucleotide (p4)	162 (1.12%)
5	Penta-nucleotide (p5)	40 (0.28%)
6	Hexa-nucleotide (p6)	50 (0.35%)
7	Compound (c)	882 (6.01%)
8	Compound (c*)	14 (0.10%)
Total		14,437

signaling molecule category: nonspecific serine/threonine-protein kinase [MTR_4g123940], calmodulin-binding protein [MTR_4g106810], etc. and for transcription factor category: ethylene response factor [MTR_1g087920], NAC transcription factor-like protein [MTR_1g069805], Myb transcription factor [MTR_8g042410], etc. are validated (Figures 5, 6).

Estimation of plant cell wall composition

The RNA sequencing data showed changes in gene expression of plant cell wall biosynthetic genes. Moreover, cell wall composition is known to determine the size and shape of the lentil seeds. To validate this, we have analyzed and compared the wall composition in seeds of L4602 and L830. Since lentil cell wall composition was not extensively analyzed, it was measured using wet chemistry methods and FT-IR (Table 6). Cellulose content is known to reinforce the cell wall, and it was 24.07% in L4602 and 25.96% in L830 genotypes; however, the difference was statistically insignificant. Lignin is a phenolic polymer that gives rigidity to cell walls, and it was 21.98% in L4602 and 25.07% in the L830 genotype. The insignificant difference was further validated using FT-IR, a similar amount of lignin and cellulose was observed in both genotypes. A particular type of xylan, i.e., arabinoxylan is abundant in lentils which interacts with both cellulose and lignin and further strengthens the cell wall matrix. The xylan was less than cellulose and lignin in both genotypes. Moreover, there was no difference in xylan content between L4602 and L830. We have also measured the acetylation level of the cell membrane that might be associated with cell extensibility. Although the acetyl content was 4.02% in the L830 genotype and 2.014% in L4602; however, this difference was statistically insignificant. Overall, cellulose was found as one of the most abundant wall components in lentil seeds. However, an increasing trend for all the cell wall components was recorded for the L830 genotype when compared with L4602, but the difference was not significant. Further validation by techniques like mass spectroscopy and nuclear magnetic resonance might reveal more about the lentil cell wall.

Discussion

A detailed understanding of the molecular mechanism regulating the seed size in lentils is of great importance for not only improving the seed yield but also for fetching better market prices in lentils. This study encompasses the identification of genes and pathways involved in seed size and shape determination using the RNA-seq approach. To reach a meaningful conclusion, we have analyzed the transcriptome at the early seed development stage and other seed membrane parameters like cellulose, lignin, and xylose at the seed maturity stage, which is thoroughly discussed in the following sections.

Seed parameters

Seed size determination in lentils is mainly done by measuring the 100 or 1000 seed weight (Tullu et al., 2001), which is considered a very crude method and is unable to distinguish different seed shape parameters like seed thickness or seed plumpness. Shahin and Symons (2001) used the computer-assisted two–dimensional image analysis to measure the lentil seed diameter, whereas Shahin et al. (2012) used cameras to capture the three-dimensional image of lentil seeds to determine the plumpness. Thus, to study various other seed parameters like area, length, width, compactness, volume, and perimeter, this study for the first time used the state-of-the-art VideometerLab instrument and very precise data were generated using 20 replicates. As recorded for this study, previous studies also revealed large variations in seed weight and diameters in lentils (Tullu et al., 2001; Tripathi et al., 2022).

RNA-seq, assembly, DE, and annotation

The sequence information for lentils is still incomplete with limited access and minimal annotation (Sharpe et al., 2013); thus, we opted for a combination of *de novo* and reference-based analysis. As the genome size of the lentil is quite big (nearly 4.2 Gb), approximately 8–12 Gb data per sample (and a total of 448 million reads) have been generated, which is very similar to that of 452 million reads by Mishra et al. (2022) and 404.67 million reads by Hosseini et al. (2021) and is considered enough for detailed RNA-seq analysis in lentil.

Several groups of growth- and development-related genes got differentially expressed with high log FC values (Supplementary Table S2) and were grouped into four main classes, viz., membrane protein, secondary metabolite, signaling molecules, and transcription factors. Likewise, the crucial metabolic

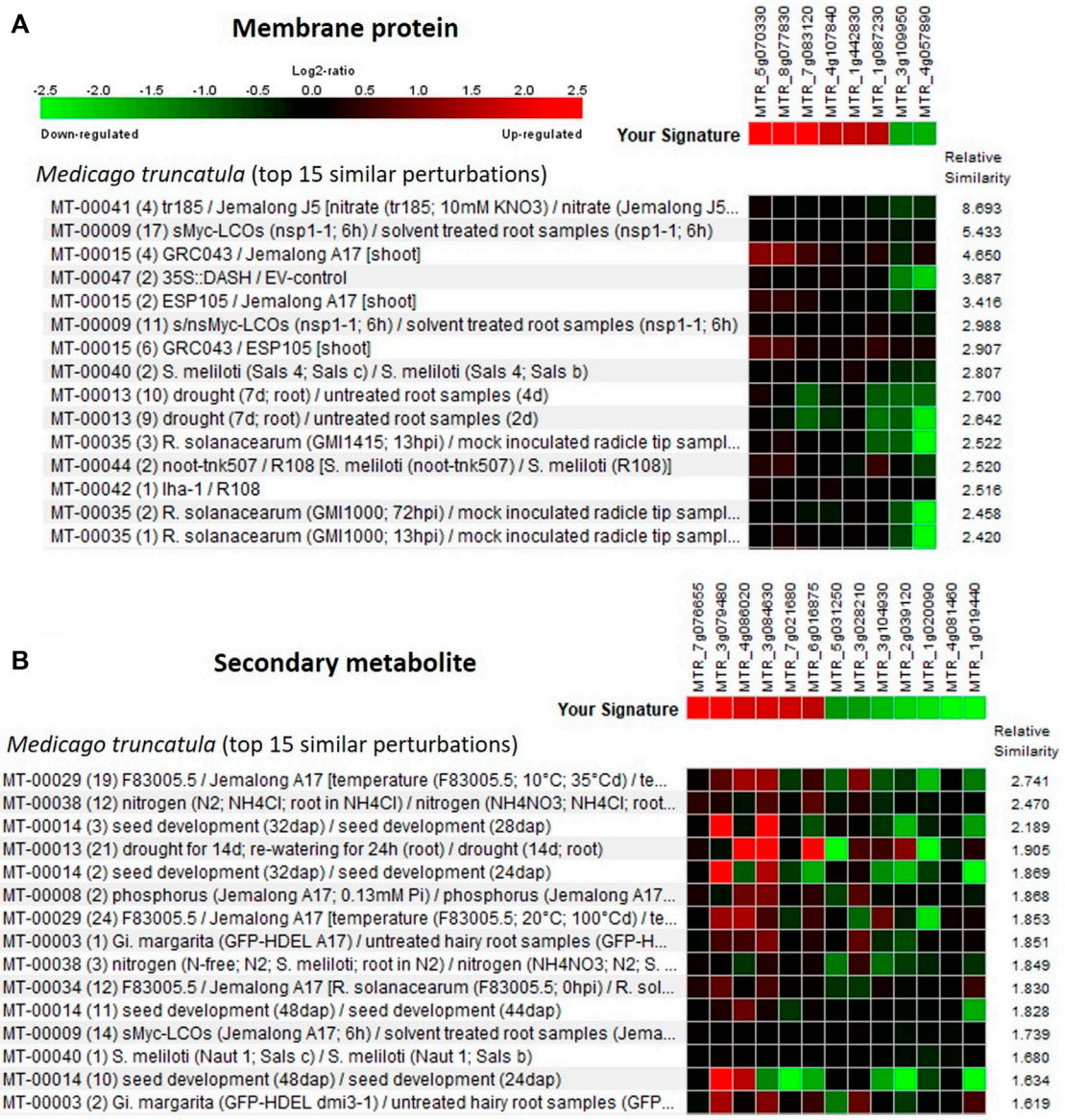


FIGURE 5
Heat map generated using the differentially expressed genes (DEGs) associated with **(A)** membrane protein and **(B)** secondary metabolite. Genevestigator® was used, and the top 15 perturbations are presented.

pathways identified in other crops for seed size variations include hormonal signaling pathways, transcription factors (Du et al., 2017), and cell division pathways (Guo et al., 2018).

GO enrichment, scatter-plot, and KEGG pathway analysis

For GO terms, some common terms like "carbohydrate metabolic process," "protein folding and glycosylation," "DNA replication," "auxin activated signaling pathway," and "photosynthesis" were also recorded by Li et al. (2021) between peanut genotypes differing for seed size. However, very little is known about the direct role of "photosynthesis"-related DEGs in the seed developmental process, and detailed investigations are needed. In addition, some key GO terms, like "G2 phase of mitotic interphase," "nucleosome assembly," "DNA-dependent DNA replication and DNA modification," and "cytokinin-activated signaling pathway" were also found significantly enriched. Similar results were also observed for peanuts (Li et al., 2021).

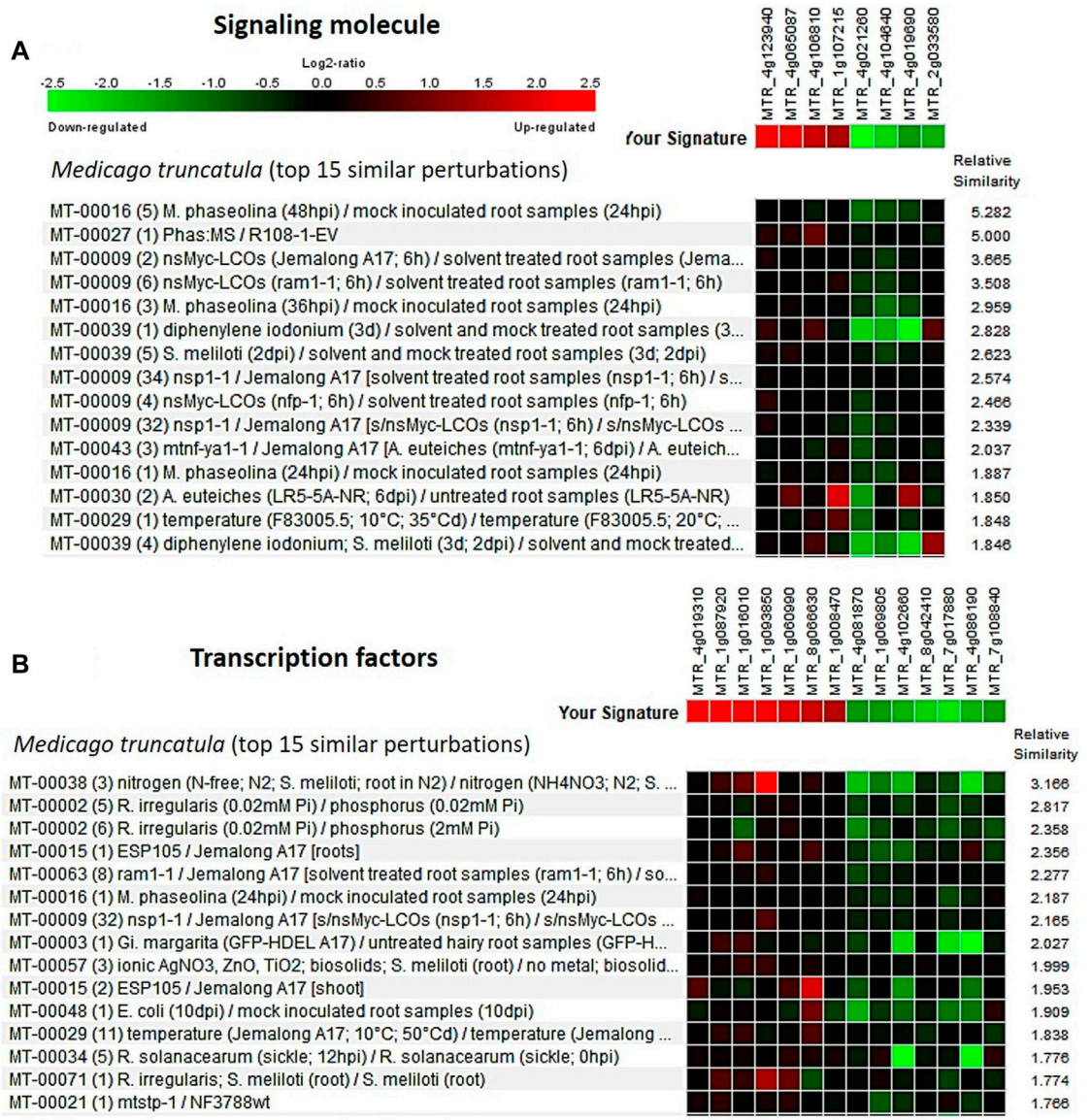


FIGURE 6
Heat map generated using the DEGs associated with **(A)** signaling molecule and **(B)** transcription factors from the lentil genotypes (early seed developmental stage). Genevestigator® was used, and the top 15 perturbations are presented.

The scatter-plot analysis also identified mainly GO terms associated with cell division and cell growth, which also supported the hypothesis of operation of cell division and cell growth mediated response pathway during the lentil seed development process in the genotypes differing for the seed size (Bosamia et al., 2020; Li et al., 2021). The overrepresented KEGG pathways mainly include sugar metabolism pathways (gluconeogenesis, pentose, pyruvate metabolism, etc.), and similar results were also reported for lentils (Mishra et al., 2022) and peanuts (Bosamia et al., 2020).

Protein–protein interaction network and gene expression validation

For any biological function, proteins are the key agents as they control both molecular and cellular mechanisms by interacting with each other and also with other molecules like DNA and RNA (Fionda 2019). Thus, PPI networks were studied to identify their role in the regulation of seed size expression in lentils. The predicted PPI network classified three modules each for both upregulated and downregulated DEGs, which mainly

TABLE 6 Estimation of cellulose, lignin, acetyl bromide soluble lignin content, D-xylose, and acetylated xylose in the mature lentil seeds of the genotypes L4602 and L830.

S. No.	Variables	L4602		L830		*p*-value*
		Mean ± SD	SE mean	Mean ± SD	SE mean	
1	FT-IR cellulose (%)	24.07 ± 0.714	0.50	25.96 ± 0.063	0.045	0.065
2	FT-IR lignin (%)	11.165 ± 0.049	0.035	12.80 ± 0.141	0.10	0.004
3	Xylose content (mg/g)	4.16 ± 0.088	0.063	6.86 ± 1.59	1.1	0.139
4	O-Acetyl content (mg/g)	2.014 ± 0.166	0.12	4.02 ± 1.77	1.3	0.252
5	ABSL lignin content (%)	21.98 ± 5.79	0.41	25.07 ± 5.98	0.42	0.652

Where total count or replication (N) = 3; FT-IR: Fourier transform-infrared spectroscopy; ABSL: acetyl bromide soluble lignin content; * at 95% confidence level

consisted of kinases, hydrolases, seed storage proteins, GTPase, etc. Likewise, the PPI network analysis in other crops also identified many modules encompassing proteins like kinases, hydrolases, etc. (Dasgupta et al., 2021). *In silico* validation studies were performed on the DEGs (small vs. large seeds), using the Genevestigator® tool (Zimmermann et al., 2004) and *Medicago truncatula* as a reference. A number of developmental pathway–related enzymes and proteins like putative transmembrane protein, glycosyltransferase, serine/threonine protein kinase, and NAC and Myb TF are mainly identified and validated. While performing the *in silico* validation studies on lentils, similar observations were also reported by Dam et al. (2018) and Mishra et al. (2022).

Role of miRNAs in seed size determination

The identification of miRNAs using the miRNA target site was previously used by Mishra et al. (2022) in lentils. This study has identified a number of key miRNAs (miR1533, miR3457, miR1440, miR7743, miR902, miR7511, and miR5543) having role in the regulation of various phytochemical synthesis and developmental pathways including cell division. miR1533 was identified as having functions like biosynthesis of plant hormones, starch metabolism, and nutrient response (Yawichai et al., 2019). The miR3457 is known to regulate the WSC domain–containing protein having a function in cell wall formation (Tarver et al., 2015), whereas miR7743 was found associated with dormancy release (Zhang et al., 2018). miR1440 was identified which functions by targeting the RAD (RADIALIS) gene of the MYB gene family and thereby regulates cell proliferation and cell cycles (Kabir et al., 2022). In addition, many developmental pathways are found regulated by miR902 and miR7511 (Berruezo et al., 2017; Zhang et al., 2019) and miR5543, having a role in cell transport (Achakzai et al., 2018). However, miR5576 and miR5658 were identified as having a function in embryo development ending seed dormancy (Zhang et al., 2015). This is the first report wherein we have identified a number of key miRNAs having roles in the determination of seed size in lentils.

Cell wall composition and their role

Cellulose and hemicellulose (like galactomannan, mannan, and xyloglucan) play a key role in providing shape and size to both developing and mature seeds (Buckeridge 2010). Till now, there is no detailed report about the cell wall parameters like cellulose, lignin, and xylose in the regulation of overall seed size parameters in lentils. In our RNA-seq data, many cell wall–associated GO terms like cell wall organization [GO: 0071555] and plant-type cell wall organization or biogenesis [GO:0071669] were found enriched. In this study, using FT-IR we found 25.96% cellulose in lentil seeds, whereas 40%–60% cellulose was reported for different plant species (Costa and Plazanet, 2016). Seeds in general contain relatively less cellulose when compared to other plant tissues, whereas endosperms are rich in storage starch and polysaccharide. In the L830 genotype, 12.8% lignin was recorded, whereas the mean lignin content in soybean was 5.13% (Krzyzanowski et al., 2008). Moreover, the soybean genotypes having >5% lignin content in their seed coat were found less to be prone to mechanical damage (Alvarez et al., 1997). Lentil seeds recorded more lignin content than the soybean, which might be due to the presence of more colored compounds in the lentil seed coat.

It is of interest that the transcripts encoding xyloglucan endotransglucosylase were found differentially regulated in our RNA-seq data, suggesting their role in cell wall formation during cell expansion and seed growth. Similar results were also reported for the seed size expression in soybean when studied using the RNA-seq approach (Du et al., 2017). However, no significant difference was recorded for the D-xylose contents in the seeds of the studied lentil genotypes at maturity (Table 6). Xylose and xyloglucan are known to play a major role as a storage polysaccharide in the developing seed tissues (Buckeridge 2010). In the studied lentil genotypes, the D-xylose content was

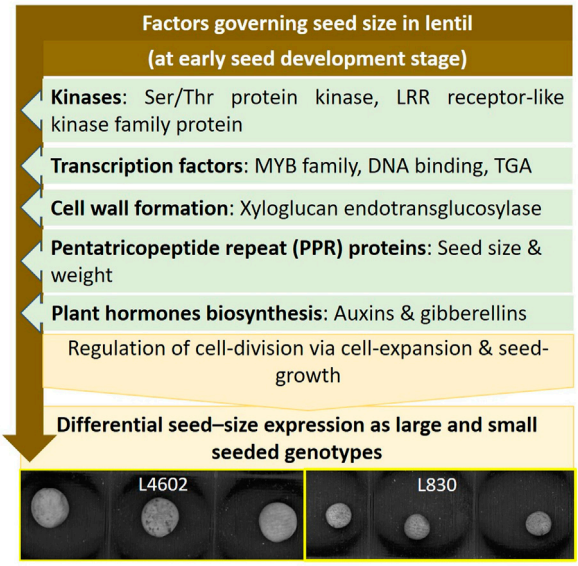


FIGURE 7
Schematic diagram showing the interplay of key genes and pathways regulating the seed size in lentil genotypes.

recorded in the range of 4.16–6.86 mg/100g, whereas similar levels of acetyl–xylose content (3.5–4.5% of dry weight) were recorded in the hardwoods by Teleman et al. (2002). Acetylation of polysaccharides affects their water solubility, interactions with cellulose, and various other physicochemical properties, which might be resulting in different seed sizes in different genotypes (Busse-Wicher et al., 2014).

Genes regulating the overall seed size via various pathways

Several cell division–related genes may influence the seed size in plants, although not yet thoroughly studied and reported in lentils. In general, seed size is greatly influenced by grain filling, cell number, cell size, and cell shape (Chun et al., 2020). More cell division in rice reportedly resulted in a bigger seed size by an increase in the cell numbers (Guo et al., 2018). It is of interest that we have also found DEGs related to the "cell division" like "DNA replication," "microtubule-based movement," and "cell wall organization" as highly enriched. Likewise, Li et al. (2021) reported enriched GO terms related to "cell division" like "nucleosome assembly," "microtubule-based process," and "cytokinesis."

Genes encoding for protein kinases like serine/threonine-protein kinase pakD-like and LRR receptor–like kinase family protein were found differentially regulated in the studied lentil genotypes. In addition, String (https://string-db.org/) and GO analysis also identified serine/threonine-protein kinase as the key player, and this was further validated using Genevestigator® [MTR_4g123940]. Similar types of receptor-like kinases have been reported to regulate the seed size in soybean (Du et al., 2017) and *Arabidopsis* (Yu et al., 2014) during the early seed developmental stage. In cultivated soybean, a PP2C-1 (phosphatase 2C-1) allele was reported to contribute to increased seed size (Lu et al., 2017), as different members of the PP2C family function as the key players in various plant signal transduction processes including RLK signaling pathways and also as a negative regulator of MAPK pathway (Rodriguez 1998).

MYB family of transcription factors in *Arabidopsis* (Zhang et al., 2013) and MYB-like DNA–binding protein [Glyma.12G100600] in soybean (Du et al., 2017) are known to regulate the seed size. Likewise, our study identified the role of the MYB family of TF and many DNA-binding TF–regulating seed size in lentils. TGA transcription factors (members of the bZIP family) are also identified that are reported to have a role in seed size determination in peanuts (Li et al., 2021).

A large family of plant proteins having tandemly repeated degenerate 35 amino acid pentatricopeptide repeat (PPR) motifs are called PPR proteins. SNP-based markers have identified a gene EVM0025654, which encodes for a PPR protein that was found associated with the increased seed size by altered cell division in peanuts (Li et al., 2021). Moreover, in maize, a PPR encoding gene (qKW9) was reported to regulate the kernel size and weight (Huang et al., 2020), whereas in peanut a candidate PPR protein gene regulating the seed weight was identified (Gangurde et al., 2020). Thus, the PPR-containing protein–encoding gene enriched in this study seems an important candidate gene associated with the regulation of seed size in lentils. Moreover, a candidate seed size regulatory gene (CYP78A5 or cytochrome P450 78A5) having a role in the stimulation of cell proliferation and promotion of developing ovules growth was identified (https://www.uniprot.org/uniprot/Q9LMX7). In addition, the overexpression of *GmCYP78A5* in soybean has shown increased seed size and seed weight of the transgenics (Du et al., 2017).

By coincidence, various hormone metabolism–related genes especially for auxin (indole-3-acetic acid [IAA]–amino acid hydrolase ILR1-like 5) and gibberellin hormones were identified in lentils. These genes were known to have a role in the seed size regulation in peanuts (Li et al., 2021). Gibberellin-related genes showed lower expression levels in our study, which suggest that the differences in the levels of phytohormone-related transcripts may affect variations in seed phenotypes. IAA influences different facets of plant growth and development (Ludwig-Müller 2011) and through IAA conjugation it maintains auxin homeostasis, whereas deconjugation occurs through certain amidohydrolases (Bitto et al., 2009). In

peanuts, an EVM0031048 gene encoding for IAA–amino acid hydrolase ILR1–like 5 was found to have a role in the regulation of seed size (Li et al., 2021). In addition, the identified EST-SSRs can be used for the identification of linked markers using an appropriate mapping population developed for the seed size trait in lentils.

Conclusion

Results of the study conclusively showed that the key genes like kinases, TF, cell wall forming enzymes, and hormone biosynthesis pathways are involved in seed size determination in lentils by regulating the cell division via cell expansion and overall seed growth. The same was also corroborated using the miRNA and GO results. The details of the findings of this study are comprehensively presented in Figure 7. The information generated in this study is of immense use and can be used for the development of lentil genotypes having customized seed quality traits.

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

Conceptualization: GM, HKD, DY, and SK; methodology: HD, MA, RK, PP, SD, AK, KT, MP, and SS; formal analysis: HD, TB, DM, JB, DV, and SS; resources: GM, SK, and HKD; data curation: HKD and GM; writing—original draft preparation: HD, GM, HKD, and MA; writing—review and editing: HKD, GM, DY, and SK; supervision: GM and HKD.

Funding

The work was supported and funded by the Indian Council of Agricultural Research (ICAR) and the International Center for Agricultural Research in the Dry Areas (ICARDA).

Acknowledgments

The authors are thankful to Dilip Kumar for his technical support during the study. The authors would like to acknowledge Bharatkumar Mathuradas Davda, Founder, Tara International (Vadodara, Gujarat) for providing access to VideometerLab 4.0 instrument at ICAR-IARI, New Delhi.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.942079/full#supplementary-material

References

Achakzai, H. K., Barozai, M. Y. K., Din, M., Baloch, I. A., and Achakzai, A. K. K. (2018). Identification and annotation of newly conserved microRNAs and their targets in wheat (*Triticum aestivum* L.). *PLoS ONE* 13 (7), e0200033. doi:10.1371/journal.pone.0200033

Alvarez, P. J. C., Krzyzanowski, F. C., Mandarino, J. M. G., and França Neto, J. B. (1997). Relationship between soybean seed coat lignin content and resistance to mechanical damage. *Seed Sci. Technol.* 25, 209–214.

Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9, 208–218. doi:10.1007/BF02672069

Berruezo, F., de Souza, F. S. J., Picca, P. I., Nemirovsky, S. I., Martínez Tosar, L., Rivero, M., et al. (2017). Sequencing of small RNAs of the fern *Pleopeltis minima* (Polypodiaceae) offers insight into the evolution of the microrna repertoire in land plants. *PLoS ONE* 12 (5), e0177573. doi:10.1371/journal.pone.0177573

Bitto, E., Bingman, C. A., Bittova, L., Houston, N. L., Boston, R. S., Fox, B. G., et al. (2009). X-ray structure of ILL2, an auxin-conjugate amidohydrolase from *Arabidopsis thaliana*. *Proteins* 74, 61–71. doi:10.1002/prot.22124

Bosamia, T. C., Dodia, S. M., Mishra, G. P., Ahmad, S., Joshi, B., Thirumalaisamy, P. P., et al. (2020). Unraveling the mechanisms of resistance to Sclerotium rolfsii in peanut (*Arachis hypogaea* L.) using comparative RNA-Seq analysis of resistant and susceptible genotypes. *PLoS ONE* 15 (8), e0236823. doi:10.1371/journal.pone.0236823

Bosamia, T. C., Mishra, G. P., Thankappan, R., and Dobaria, J. R. (2015)). Novel and stress relevant EST derived SSR markers developed and validated in peanut. *PLoS ONE* 10, e0129127. doi:10.1371/journal.pone.0129127

Buckeridge, M. S. (2010). Seed cell wall storage polysaccharides: Models to understand cell wall biosynthesis and degradation. *Plant Physiol.* 154 (3), 1017–1023. doi:10.1104/pp.110.158642

Busse-Wicher, M., Gomes, T. C., Tryfona, T., Nikolovski, N., Stott, K., Grantham, N. J., et al. (2014). The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. *Plant J.* 79 (3), 492–506. doi:10.1111/tpj.12575

Canteri, M. H. G., Renard, C. M. G. C., Le Bourvellec, C., and Bureau, S. (2019). ATR-FTIR spectroscopy to determine cell wall composition: application on a large diversity of fruits and vegetables. *Carbohydr. Polym.* 15 (212), 186–196. doi:10.1016/j.carbpol.2019.02.021

Chun, Y., Fang, J., Zafar, S. A., Shang, J., Zhao, J., Yuan, S., et al. (2020). MINI SEED2 (MIS2) encodes a receptor-like kinase that controls grain size and shape in rice. *Rice* 13, 7. doi:10.1186/s12284-020-0368-9

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., Robles, M., et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi:10.1093/bioinformatics/bti610

Costa, G., and Plazanet, I. (2016). Plant cell wall, a challenge for its characterisation. *Adv. Biol. Chem.* 6, 70–105. doi:10.4236/abc.2016.63008

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* 46, W49-W54. –W54. doi:10.1093/nar/gky316

Dam, S. V., Võsa, U., Graaf, A. V. D., Franke, L., and Magalhães, J. P. D. (2018). Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* 19 (4), 575–592. doi:10.1093/bib/bbw139

Dasgupta, U., Mishra, G. P., Dikshit, H. K., Mishra, D. C., Bosamia, T., Roy, A., et al. (2021). Comparative RNA-Seq analysis unfolds a complex regulatory network imparting yellow mosaic disease resistance in mungbean [*Vigna radiata* (L.) R. Wilczek]. *Plos One* 16 (1), e0244593. doi:10.1371/journal.pone.0244593

Dikshit, H. K., Mishra, G. P., Aski, M. S., Singh, A., Tripathi, T., Bansal, R., et al. (2022) Lentil breeding. In: *Fundamentals of field crop breeding*, Editor D.K. Yadava, et al. pp:1181. doi:10.1007/978-981-16-9257-4_24

Du, J., Wang, S., He, C., Zhou, B., Ruan, Y. L., Shou, H., et al. (2017). Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* 68 (8), 1955–1972. doi:10.1093/jxb/erw460

Dueck, A., Ziegler, C., Eichner, A., Berezikov, E., and Meister, G. (2012). microRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res.* 40, 9850–9862. doi:10.1093/nar/gks705

FAO (2022). *FAOSTAT. Food and agriculture organization of the united nations*. Rome. Available online at: http://faostat.fao.org (Accessed on May 1st, 2022).

Fionda, V. (2019). "Networks in biology," in *Encycl. Bioinforma. Comput. Biol.* Editors S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach (Oxford: Academic Press), 1, 915–921. doi:10.1016/B978-0-12-809633-8.20420-2

Foster, C. E., Martin, T. M., and Pauly, M. (2010). Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) Part I: Lignin. *J. Vis. Exp.* 37, e1745. doi:10.3791/1745

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., et al. (2020). Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 18, 1457–1471. doi:10.1111/pbi.13311

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

Hosseini, S. Z., Ismaili, A., Firouzabadi, F. N., Fallahi, H., Nejad, A. R., Sohrabi, S. S., et al. (2021). Dissecting the molecular responses of lentil to individual and combined drought and heat stresses by comparative transcriptomic analysis. *Genomics* 113, 693–705. doi:10.1016/j.ygeno.2020.12.038

Huang, J., Lu, G., Liu, L., Raihan, M. S., Xu, J., Jian, L., et al. (2020). The kernel size-related quantitative trait locus qKW9 encodes a pentatricopeptide repeat protein that aaffects photosynthesis and grain filling. *Plant Physiol.* 183, 1696–1709. doi:10.1104/pp.20.00374

Kabir, N., Zhang, X., Liu, L., Qanmber, G., Zhang, L., Wang, Y. X., et al. (2022). *RAD* gene family analysis in cotton provides some key genes for flowering and stress tolerance in upland cotton *G. hirsutum*. *BMC Genomics* 23, 40. doi:10.1186/s12864-021-08248-z

Kanehisa, M., and Goto, S. (2000). Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Kathirselvam, M., Kumaravel, A., Arthanarieswaran, V., and Saravanakumar, S. (2019). Assessment of cellulose in bark fibers of *Thespesia populnea*: influence of stem maturity on fiber characterization. *Carbohydr. Polym.* 212, 439–449. doi:10.1016/j.carbpol.2019.02.072

Krzyzanowski, F., Franca Neto, J., Mandarino, J., and Kaster, M. (2008). Evaluation of lignin content of soybean seed coat stored in a controlled environment. *Rev. Bras. sementes* 30 (2), 220–223. doi:10.1590/s0101-31222008000200028

Labbe, N., Rials, T., Kelley, S., Cheng, Z., Kim, J., Li, Y., et al. (2005). FT-IR imaging and pyrolysis-molecular beam mass spectrometry: new tools to investigate wood tissues. *Wood Sci. Technol.* 39 (1), 61–76. doi:10.1007/s00226-004-0274-0

Li, Z., Zhang, X., Zhao, K., Zhao, K., Qu, C., Gao, G., et al. (2021). Comprehensive transcriptome analyses reveal candidate genes for variation in seed size/weight during peanut (*Arachis hypogaea* L.) domestication. *Front. Plant Sci.* 12, 666483. doi:10.3389/fpls.2021.666483

Lu, X., Xiong, Q., Cheng, T., Li, Q. T., Liu, X. L., Bi, Y. D., et al. (2017). A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10, 670–684. doi:10.1016/j.molp.2017.03.006

Ludwig-Müller, J. (2011). Auxin conjugates: Their role for plant development and in the evolution of land plants. *J. Exp. Bot.* 62, 1757–1773. doi:10.1093/jxb/erq412

Minitab, I. (2020). Minitab. Available at: http://www.minitab.com/en-US/products/minitab/ (Accessed on April 15, 2022).

Mishra, G. P., Aski, M. S., Bosamia, T., Chaurasia, S., Mishra, D. C., Bhati, J., et al. (2022). Insights into the host-pathogen interaction pathways through RNA-Seq analysis of *Lens culinaris* Medik. in response to *Rhizoctonia bataticola* Infection. *Genes.* 13 (1), 90. doi:10.3390/genes13010090

Mishra, G. P., Dikshit, H. K., Kumari, J., Priti, Tripathi K., Devi, J., et al. (2020a). Identification and characterization of novel penta-podded genotypes in the cultivated lentil (*Lens culinaris* Medik.). *Crop Sci.* 60 (4), 1974–1985. doi:10.1002/csc2.20156

Mishra, G. P., Dikshit, H. K., Ramesh, S. V., Tripathi, K., Kumar, R. R., Aski, M., et al. (2020b). Yellow mosaic disease (YMD) of mungbean (*Vigna radiata* (L.) Wilczek): current status and management opportunities. *Front. Plant Sci.* 11, 918. doi:10.3389/fpls.2020.00918

Numnark, S., Mhuantong, W., Ingsriswang, S., and Wichadakul, D. (2012). C-Mii: a tool for plant miRNA and target identification. *BMC Genomics* 13, S16. doi:10.1186/1471-2164-13-S7-S16

Pawar, P. M., Derba-Maceluch, M., Chong, S. L., Gandla, M. L., Bashar, S. S., Sparrman, T., et al. (2017). In muro deacetylation of xylan affects lignin properties and improves saccharification of aspen wood. *Biotechnol. Biofuels* 10, 98. doi:10.1186/s13068-017-0782-4

PritiMishra, G. P., Dikshit, H. K., Vinutha, T., Mechiya, T., Stobdan, T., et al. (2021). Diversity in phytochemical composition, antioxidant capacities, and nutrient contents among mungbean and lentil microgreens when grown at plain-altitude region (Delhi) and high-altitude region (Leh-Ladakh), India. *Front. Plant Sci.* 12, 710812. doi:10.3389/fpls.2021.710812

Priti, Sangwan S., Kukreja, B., Mishra, G. P., Dikshit, H. K., Singh, A., et al. (2022). Yield optimization, microbial load analysis, and sensory evaluation of mungbean (*Vigna radiata* L.), lentil (*Lens culinaris* subsp. *culinaris*), and Indian mustard (*Brassica juncea* L.) microgreens grown under greenhouse conditions. *Plos One* 17, e0268085. doi:10.1371/journal.pone.0268085

Ramsay, L., Koh, C. S., Kagale, S., Gao, D., Kaur, S., Haile, T., et al. (2021). Genomic rearrangements have consequences for introgression breeding as revealed by genome assemblies of wild and cultivated lentil species. bioRxiv. 2021 Jul 24Retrieved from https://knowpulse.usask.ca/genome-assembly/Lcu.2RBY.

Rastogi, L., Chaudhari, A. A., Sharma, R., and Pawar, P. A. M. (2022). Arabidopsis GELP7 functions as a plasma membrane-localized acetyl xylan esterase, and its overexpression improves saccharification efficiency. *Plant Mol. Biol.* doi:10.1007/s11103-022-01275-8

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616

Rodriguez, P. L. (1998). Protein phosphatase 2C (PP2C) function in higher plants. *Plant Mol. Biol.* 38 (6), 919–927. doi:10.1023/a:1006054607850

Ruan, Y. L., Patrick, J. W., Bouzayen, M., Osorio, S., and Fernie, A. R. (2012). Molecular regulation of seed and fruit set. *Trends Plant Sci.* 17, 656–665. doi:10.1016/j.tplants.2012.06.005

Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–289. doi:10.1146/annurev-arplant-042809-112315

Shahin, M. A., and Symons, S. J. (2001). A machine vision system for grading lentils. *Can. Biosys Eng.* 43, 7. https://library.csbe-scgab.ca/docs/journal/43/c0019.pdf.

Shahin, M. A., Symons, S. J., and Wang, N. (2012). Predicting dehulling efficiency of lentils based on seed size and shape characteristics measured with image analysis. *Qual. Assur. Saf. Crops Foods* 4, 9–16. doi:10.1111/j.1757-837X.2011.00119.x

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Sharpe, A. G., Ramsay, L., Sanderson, L. A., Fedoruk, M. J., Clarke, W. E., Rong, L., et al. (2013). Ancient orphan crop joins modern era: Gene-based SNP discovery and mapping in lentil. *BMC Genomics* 14, 192. doi:10.1186/1471-2164-14-192

Shrestha, S., Deleuran, L. C., Olesen, M. H., and Gislum, R. (2015) Use of multispectral imaging in varietal identification of tomato. *Sensors*,15(2):4496–4512. doi:10.3390/s150204496

Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., Kumar, S., and Sarker, A. (2022) Breeding for abiotic stress tolerance in lentil in genomic era. In: *Genomic designing for abiotic stress resistant pulse crops*. Editor C. Kole. 145. doi:10.1007/978-3-030-91039-6_5

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *Plos One* 6, e21800. doi:10.1371/journal.pone.0021800

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi:10.1093/nar/gku1003

Tarver, J. E., Cormier, A., Pinzón, N., Taylor, R. S., Carré, W., Strittmatter, M., et al. (2015). microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the Brown alga Ectocarpus. *Nucleic Acids Res.* 43 (13), 6384–6398. doi:10.1093/nar/gkv578

Teleman, A., Tenkanen, M., Jacobs, A., and Dahlman, O. (2002). Characterization of O-acetyl-(4-O-methylglucurono) xylan isolated from birch and beech. *Carbohydr. Res.* 337 (4), 373–377. doi:10.1016/s0008-6215(01)00327-5

Thiel, T., Michalek, W., Varshney, R. K., Graver, A., 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

Tripathi, K., Kumari, J., Gore, P. G., Mishra, D. C., Singh, A. K., Mishra, G. P., et al. (2022). Agro-morphological characterization of lentil germplasm of Indian national genebank and development of a core set for efficient utilization in lentil improvement programs. *Front. Plant Sci.* 12, 751429. doi:10.3389/fpls.2021.751429

Tullu, A., Kusmenoglu, I., McPhee, K. E., and Muehlbauer, F. J. (2001). Characterization of core collection of lentil germplasm for phenology, morphology, seed and straw yields. *Genet. Resour. Crop Evol.* 48, 143–152. doi:10.1023/A:1011254629628

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40 (15), e115. doi:10.1093/nar/gks596

Updegraff, D. M. (1969). Semimicro determination of cellulose in biological materials. *Anal. Biochem.* 3 (2), 420–424. doi:10.1016/S0003-2697(69)80009-6

Wang, N. (2008). Effect of variety and crude protein content on dehulling quality and on the resulting chemical composition of red lentil (*Lens culinaris*). *J. Sci. Food Agric.* 88, 885–890. doi:10.1002/jsfa.3165

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi:10.1146/annurev.arplant.56.032604.144201

Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* 11, 228–234. doi:10.1038/ncb0309-228

Yawichai, A., Kalapanulak, S., Thammarongtham, C., and Saithong, T. (2019). Genome-wide identification of putative microRNAs in cassava (*Manihot esculenta* Crantz) and their functional landscape in cellular regulation. *Biomed. Res. Int.* 19, 2019846. doi:10.1155/2019/2019846

Yu, F., Li, J., Huang, Y., Liu, L., Li, D., Chen, L., et al. (2014). FERONIA receptor kinase controls seed size in *Arabidopsis thaliana*. *Mol. Plant* 7, 920–922. doi:10.1093/mp/ssu010

Zhang, B., Zhang, X., Zhang, M., Guo, L., Qi, T., Wang, H., et al. (2019). Transcriptome analysis implicates involvement of long noncoding RNAs in cytoplasmic male sterility and fertility restoration in cotton. *Int. J. Mol. Sci.* 20 (22), 5530. doi:10.3390/ijms20225530

Zhang, F., Dong, W., Huang, L., Song, A., Wang, H., Fang, W., et al. (2015). Identification of microRNAs and their targets associated with embryo abortion during chrysanthemum cross breeding via high-throughput sequencing. *PLoS One* 10 (4), e0124371. doi:10.1371/journal.pone.0124371

Zhang, Y., Liang, W., Shi, J., Xu, J., and Zhang, D. (2013). MYB56 encoding a R2R3 MYB transcription factor regulates seed size in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* 55, 1166–1178. doi:10.1111/jipb.12094

Zhang, Y., Wang, Y., Gao, X., Liu, C., and Gai, S. (2018). Identification and characterization of microRNAs in tree peony during chilling induced dormancy release by high-throughput sequencing. *Sci. Rep.* 8, 4537. doi:10.1038/s41598-018-22415-5

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* 136, 2621–2632. doi:10.1104/pp.104.046367

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 15 August 2022
doi: 10.3389/fgene.2022.945787

Linkage QTL Mapping and Genome-Wide Association Study on Resistance in Chickpea to *Pythium ultimum*

Chiti Agarwal[1], *Weidong Chen*[2], *Rajeev Kumar Varshney*[3] *and George Vandemark*[2]*

[1]Department of Plant Pathology, Washington State University, Pullman, WA, United States, [2]USDA-ARS, Grain Legume Genetics and Physiology Research Unit, Pullman, WA, United States, [3]Centre for Crop and Food Innovation, State Agricultural Biotechnology Centre, Murdoch University, Murdoch, WA, Australia

The soilborne oomycete plant pathogen *Pythium ultimum* causes seed rot and pre-emergence damping-off of chickpea (*Cicer arietinum* L.). The pathogen has been controlled for several decades using the fungicide metalaxyl as seed treatment but has re-emerged as a severe problem with the detection of metalaxyl-resistant isolates of the pathogen from infested fields in the United States Pacific Northwest. The objective of this study was to identify genetic markers and candidate genes associated with resistance to *P. ultimum* in an interspecific recombinant inbred line population (CRIL-7) derived from a cross between *C. reticulatum* (PI 599072) x *C. arietinum* (FLIP 84-92C) and conduct genome-wide association studies (GWAS) for disease resistance using a chickpea diversity panel consisting of 184 accessions. CRIL-7 was examined using 1029 SNP markers spanning eight linkage groups. A major QTL, "qpsd4-1," was detected on LG 4 that explained 41.8% of phenotypic variance, and a minor QTL, "qpsd8-1," was detected on LG8 that explained 4.5% of phenotypic variance. Seven candidate genes were also detected using composite interval mapping including several genes previously associated with disease resistance in other crop species. A total of 302,902 single nucleotide polymorphic (SNP) markers were used to determine population structure and kinship of the diversity panel. Marker–trait associations were established by employing different combinations of principal components (PC) and kinships (K) in the FarmCPU model. Genome-wide association studies detected 11 significant SNPs and seven candidate genes associated with disease resistance. SNP Ca4_1765418, detected by GWAS on chromosome 4, was located within QTL qpsd4-1 that was revealed in the interspecific CRIL-7 population. The present study provides tools to enable MAS for resistance to *P. ultimum* and identified genomic domains and candidate genes involved in the resistance of chickpea to soilborne diseases.

Keywords: chickpea, disease, pulses, Pythium, resistance

OPEN ACCESS

Edited by:
Waltram Ravelombola, Texas A&M University, United States

Reviewed by:
Tuanjie Zhao, Nanjing Agricultural University, China
Eugenio López-Cortegano, University of Edinburgh, United Kingdom

***Correspondence:**
George Vandemark
george.vandemark@usda.gov

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

Received: *16 May 2022*
Accepted: *20 June 2022*
Published: *15 August 2022*

Citation:
Agarwal C, Chen W, Varshney RK and Vandemark G (2022) Linkage QTL Mapping and Genome-Wide Association Study on Resistance in Chickpea to Pythium ultimum. Front. Genet. 13:945787. doi: 10.3389/fgene.2022.945787

INTRODUCTION

Chickpea (*Cicer arietinum* L) is one of the most historically significant field crops, being among the eight "founder crops" domesticated by Neolithic societies 8,000–12,000 years ago in the "Fertile Crescent" of present-day Iraq, Syria, and Turkey (Zohary et al., 2012). Its importance continues to this day, with more than 14.2 million tonnes of chickpea produced globally in 2019, making it the third most important pulse crop in terms of global production, after dry beans (*Phaseolus vulgaris* L.) and peas (*Pisum sativum* L.) (FAOSTAT, 2022). Chickpea is produced in more than 50 nations, with India being the largest producer, accounting for approximately 68% of global production (FAOSTAT, 2022). There are two major market classes of chickpea based on seed traits; "*Desi*" chickpea, which have a "teardrop"-shaped seed and a pigmented seed coat, and "*Kabuli*" chickpea, which has an "owl head" shape, a light beige or cream-colored seed coat, and is typically larger than the *Desi* chickpea (Toker, 2009).

In the United States, chickpea is primarily grown in dryland production systems in rotation with wheat or barley in the Pacific Northwest (Idaho and Washington) and Northern Plains (Montana and North Dakota). In 2019, more than 194,000 tonnes of chickpea were produced in the United States, with an estimated value greater than $116 million (NASS, 2022). However, several diseases challenge farmers in the United States, including Ascochyta blight caused by *Ascochyta rabiei* (Pass.) Lab. (Bayaa et al., 2011). Recently, seed rot and pre-emergent damping-off of chickpea caused by metalaxyl-resistant *P. ultimum* Trow has re-emerged as a significant disease in the Pacific Northwest.

The genus *Pythium* includes several soilborne species that cause seed and seedling diseases across a wide range of crops (Martin and Loper, 1999). *P. ultimum* causes seed rot, damping-off, and root rot in several legumes, including soybean (*Glycine* max L.), common bean (*Phaseolus vulgaris* L.), and pea (*Pisum sativum* L.) (Plaats-Niterink, 1981). Seed rot and pre-emergence damping-off of chickpea caused by *P. ultimum* were first detected in the United States in 1979 in Washington State (Kaiser and Hannan, 1983). Historically, these diseases have been controlled using pre-plant seed treatments containing metalaxyl or its stereoisomer mefenoxam (Casas et al., 1990). However, in 2014 isolates of *P. ultimum* var. *ultimum* with metalaxyl resistance were collected from unsprouted and rotten chickpea seeds obtained from a field in Washington exhibiting poor sprouting (Chen and Van Vleet, 2016). Subsequently, metalaxyl-resistant isolates of *P. ultimum* have been collected from several chickpea production fields in Idaho and Washington, where poor seedling sprouting was observed (Wang et al., 2020). Greenhouse and field tests showed that ethaboxam effectively manages metalaxyl-resistant *P. ultimum*, and commercial chickpea farmers now commonly apply a seed treatment containing both metalaxyl and ethaboxam for disease control (Wang et al., 2020). However, this increases production costs for farmers, and the effective use of ethaboxam depends on pathogen populations not developing resistance to the fungicide. Although resistance to ethaboxam in *P. ultimum* has not been detected, resistance has been detected in *P. aphanidermatum* (Edson) Fitzp. and *P. deliense* Meurs isolated from soybean (*Glycine* max L.) and dry bean (*Phaseolus vulgaris* L.), respectively. Increased production costs and concerns about the development of metalaxyl-resistant *P. ultimum* suggest that effective disease control in the future may require other approaches, including the use of disease-resistant chickpea cultivars.

Initial studies indicated that only small, dark-seeded *Desi* chickpeas were resistant to *P. ultimum* and all *Kabuli* chickpeas tested were susceptible (Kaiser and Hannan, 1983). We recently evaluated a collection of commercial chickpea cultivars and accessions from the United States National Plant Germplasm System (NPGS) and the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) for resistance to metalaxyl-resistant *P. ultimum* (Agarwal et al., 2020). The great majority of resistant accessions were *Desi* types and accessions with pigmented seed coats. Although the popular *Kabuli* cultivars 'Sierra' (Muehlbauer et al., 2004) and 'Nash' (Vandemark et al., 2015) were susceptible, three *Kabuli* accessions W625864, W625882, and W625884 were identified that were significantly more resistant than Sierra to two different isolates of metalaxyl-resistant *P. ultimum* (Agarwal et al., 2020). These results were promising because chickpea production in the United States is almost entirely composed of *Kabuli* types (Vandemark et al., 2014). Although these partially resistant accessions may be useful as parents for developing *Kabuli* cultivars with improved resistance to metalaxyl-resistant *P. ultimum*, it may also be possible to accelerate the development of resistant cultivars through the use of marker-assisted breeding approaches.

A range of genomic resources is available for chickpeas, including a draft sequence of the cultivated chickpea genome, which has an estimated size of approximately 738 Mb and contains 28,269 genes (Varshney et al., 2013). By using different marker genotyping platforms and molecular mapping approaches, significant associations have been identified between molecular markers and several diseases of chickpea, including Ascochyta blight (Tekeoglu et al., 2004; Sabbavarapu et al., 2013; Jendoubi et al., 2016; Garg et al., 2018; Mannur et al., 2019), *Fusarium* wilt (Winter et al., 2000; Sharma and Muehlbauer, 2007; Jingade and Ravikumar, 2015; Li et al., 2015; Meng et al., 2015; Mannur et al., 2019; Karadi et al., 2021), and dry root rot (Karadi et al., 2021). The objective of this study was to detect significant marker–trait associations and identify candidate genes for resistance in chickpea to metalaxyl-resistant *P. ultimum*.

MATERIALS AND METHODS

Plant Material

A chickpea mapping population (CRIL-7) that included 177 chickpea recombinant inbred lines (RIL) derived from an interspecific cross *C. reticulatum* (PI 599072) x *C. arietinum* (FLIP 84–92C) was used to conduct QTL analysis. PI 599072 is a *Desi* type and is resistant to *P. ultimum*, whereas FLIP 84-92C is a disease-susceptible *Kabuli* type. The RILs were increased under

greenhouse conditions by single seed descent to F_7 (Tekeoglu et al., 2000). Genome-wide association study was carried out on 184 taxonomically, morphologically, and geographically diverse accessions obtained from ICRISAT, Patancheru, India. This collection included 34 *Kabuli*, 144 *Desi*, and 6 pea-shaped accessions (Upadhyaya et al., 2001).

Disease Screening Assay and Resistance Scoring

P. ultimum strain PT410 (Agarwal et al., 2020) was used for all disease screening. The isolate was originally obtained from decaying chickpea seeds collected from a field in Patterson, WA, and its resistance to metalaxyl was confirmed through serial subculturing on media containing 50 ppm metalaxyl (Wang et al., 2020). The isolate was cultured and maintained on sucrose yeast extract agar in Petri plates at room temperature. CRIL-7 recombinant inbred lines along with the parental genotypes were screened for disease reaction to *P. ultimum* isolate (PT410) under controlled growth chamber conditions (12°C night–14°C days, 12 h day length) at Washington State University, Pullman, United States. Five seeds were planted in 10-cm pots containing 70 g of soil mix infested with 25,000 CFU of *P. ultimum* oospores for each entry. Myles and Sierra chickpea cultivars were used as resistant and susceptible checks, respectively (Agarwal et al., 2020). Pots were arranged in a completely randomized design. The number of seedlings that emerged from each pot was counted 14 days after planting. The experiment was repeated once.

Similarly, 184 accessions of mini-core collection from ICRISAT were evaluated for resistance to *P. ultimum* PT410 using the aforementioned methods. Again, the cultivars Sierra and Myles were included as susceptible and resistant controls. The experiment was repeated once. Results of these evaluations were previously reported (Agarwal et al., 2020).

Statistical Analysis

Statistical analysis was conducted using JMP 14 software (MP®, Version <14>. SAS Institute Inc. Cary, NC, 1989–2019). Broad-sense heritability (H^2) was calculated based on the average mean seedling emergence values from repeated experiments with R software (http://www.R-project.org/) using MME-based algorithms (Alexander and Lange, 2011).

Linkage Group Construction and Linkage QTL Mapping

Genomic DNA was extracted from young leaves of both parents and 177 RILs. Single nucleotide polymorphic (SNP) markers were detected using the Genotyping by Sequencing (GBS) approach. The libraries from the parental lines and RILs were prepared using *Msl*I restriction enzymes and sequenced using the Illumina NextSeq 500 V^2 to generate 150 bp paired-end reads by LGC Company (https://www.lgcgroup.com/). For processing reads, demultiplexing of all library groups was done using Illumina bcl2fastq 2.17.1.14 software, followed by demultiplexing of library groups into samples according to their barcodes. Quality trimming was done by discarding low-quality reads with a final length <20 bases, and filtered data were used to call SNPs. The filtered, high-quality reads from each sample were aligned to the chickpea reference genome (*Cicer arietinum* CDC Frontier whole-genome assembly v1.0) (Varshney et al., 2013). The variant discovery was made using Freebayes v1.0.2-16 with the parameters min-base-quality: 10; min-supportingallele-qsum: 10; read-mismatch-limit: 3; min-coverage: 5; min-alternate count: 4, excluding unobserved genotypes; and mismatch-base-quality-threshold: 10. Variant filtering was done by removing markers with missing allele calls and minor allele frequency (MAF) < 0.05. Redundant markers were excluded from the analysis by implementing the BIN function in QTL IciMaping 4.1 (Meng et al., 2015). SNP markers with highly distorted segregation ratios at probability level ($p \leq 0.0001$) were excluded. However, SNP that were slightly distorted ($0.0001 \leq p \leq 0.05$) from the Mendelian ratio were included in the linkage map.

These filtered markers were used to construct linkage groups (LGs) using the "Map" function in QTL IciMaping 4.1 and were assigned numbers (LG1–LG8) based on the genomic position of SNP markers. LGs with unlinked markers were removed from further construction. The remaining SNP were grouped with a logarithm-of-odds (LOD) threshold of 9.0. Recombination counting and ordering (RECORD) and "COUNT" (number of recombination events) algorithms were used in ordering and rippling. The linkage map and the best linear unbiased predictions (BLUP) value of phenotypic data of the CRIL-7 population from repeated experiments were used for QTL analysis. QTL was detected with composite inclusive composite interval mapping of additive (ICIM-ADD) function in QTL IciMaping 4.1 (Li et al., 2007). The threshold used to declare significant QTL was the permutation test with 1,000 permutations at the 0.05 significance level. Mapping parameters to detect additive QTL were set as step = 1.0 cM and PIN = 0.001 (PIN: the largest value for entering variables in stepwise regression of residual phenotype on marker variables). Parents for trait enhancing alleles were detected using the sign of the additive effects; the positive sign denotes that trait enhancing allele is from parent PI599072, whereas the negative sign indicates that the trait-enhancing allele is from FLIP 84-92C.

SNP Panel for GWAS

An SNP dataset from the Center of Excellence in Genomics & Systems Biology, ICRISAT that contained approximately 900,000 SNPs across 184 chickpea accessions was used for GWAS. First, SNP data were filtered by removing markers with more than 80% missing data and minor allele frequency smaller than 0.05. Duplicate markers and duplicate genotype samples in the dataset were then removed along with contigs and scaffolds using VCFtools (Hyun et al., 2008). Next, pairwise r^2 was calculated for all SNPs across each chromosome of the chickpea genome. SNPs with significant r^2 values ($p < 0.001$) were considered informative and were pruned using a linkage disequilibrium (LD) pruning method implemented in PLINK software v1.09 using "—indep-pairwise 50 5 0.5" command line in Linux (Purcell et al., 2007). The pruned set of SNP markers was used for association analysis.

Population Structure and Relatedness

Population structure and kinship were estimated using ADMIXTURE (v1.23) software. The ADMIXTURE tool uses a model-based algorithm to estimate the ancestry of unrelated individuals (Alexander and Lange, 2011). The number of underlying population groups (k) was estimated from 1 to 10 using the maximum likelihood estimation approach with a fast numerical optimization algorithm. A Q-matrix file representing the least number of population groups (k) was used for GWAS. Population structure was further estimated by principal component analysis (PCA) using the PLINK function. The EMMA algorithm embedded in the GAPIT package of R software was used to account for kinship (Hyun et al., 2008; Lipka et al., 2012). Finally, a dendrogram was generated using a neighbor-joining (NJ) algorithm to assess the relationship between mini-core accessions.

Marker–Trait Association Analysis

Association analysis was done using disease scores of the 184 chickpea accessions and 302,902 filtered SNPs. Kinship relatedness (K) was considered a random effect, and population structure based on the number of principal components (PC) that explained 25%–50% of the total phenotypic variance was considered a fixed effect. For marker–trait associations, models with different combinations of the population (PC)/admixture(Q) and family (K) structures were applied using the GAPIT package of R software: FarmCPU with Kinship (K), FarmCPU with Kinship (K) + PC(2), FarmCPU with Kinship (K) + PC(3), FarmCPU with Kinship (K) + PC(4), FarmCPU with Kinship (K) + PC(5), and FarmCPU with Kinship (K) + Admixture (Q) (Lipka et al., 2012). The best model was selected based on the mean squared difference (MSD) value between observed and expected *p*-values of all SNPs (Mamidi et al., 2011). The final Q–Q plots and Manhattan plots were created, and significant SNPs were calculated based on a *p*-value < 10^{-5} and Bonferroni cut-off (*p*-value of 0.05/(the total number of SNP markers) (i.e., $0.05/302902 = 1.65 \times 10^{-7}$). Genes located within a 100-Kb region centered on a significant SNP were selected as candidate genes. The SNPeff tool (http://snpeff.sourceforge.net/) was used to detect genome coordinates of candidate genes and amino acid changes due to the SNPs (Cingolani et al., 2012).

RESULTS

Phenotypic Evaluation for Resistance to *Pythium ultimum*.

A total of 177 CRIL-7 lines and both parents were screened for resistance to *P. ultimum* in a repeated experiment. A summary of descriptive statistics of disease reactions is presented in **Supplementary Table S1**. The mean of the susceptible check Sierra was 0.2 and the mean of the resistant check Myles was 4.1 in both experiments. The mean of PI 599072 was >2.5 in both experiments, indicating partial resistance to *P. ultimum*. The mean of FLIP 84-92C was ≤0.3 in both experiments, indicating susceptibility to *P. ultimum*. The means of all RILs in experiments 1 and 2 were 1.7 and 1.8, respectively. Transgressive segregants were observed among RILs in both experiments. RIL effects, experiment effects, and their interaction effects were all significant (**Supplementary Table S1**), with the greatest magnitude for the RIL effect and the least for the interaction effect. A broad-sense heritability estimate of 0.78 suggests that disease resistance is highly heritable (**Supplementary Table S2**).

SNP-Based Interspecific Genetic Map

A total of 65,112 SNP markers were obtained by GBS on CRIL-7 population [*C. reticulatum* (PI 599072) x *C. arietinum* (FLIP 84–92C)]. After filtering, 1,029 SNP markers were used to construct a linkage map and were assigned to the eight chickpea linkage groups (LG1–LG8) (**Supplementary Figure S3**). The eight linkage groups covered 1,186.30 cM (**Supplementary Table S4**). LG 8 was the smallest linkage group with 65 markers and a length of 50.59 cM, while the largest was LG 5 with 112 markers and a total length of 281.55 cM (**Supplementary Table S3**). Gaps in marker coverage of LGs are due to a high proportion of distorted segregation markers among the interspecific mapping population.

QTLs and Candidate Genes Detected by Linkage Mapping

Two QTLs were detected that were significantly associated with resistance to *P. ultimum*, QTL qpsd4-1 on LG 4, and QTL qpsd8-1 on LG 8. qpsd4-1 explained 46.5% of total variance with LOD = 25.2, whereas qpsd8-1 explained 4.5% of the total phenotypic variance with LOD = 3.3 (**Table 1**). qpsd4-1 had a positive additive effect value, indicating that positive alleles came from PI 599072, while the additive effect value was negative for qpsd8-1, indicating that negative alleles came from FLIP 84-92C.

The physical positions of flanking markers were used to identify putative genes associated with resistance to *P. ultimum* located within two QTLs, qpsd4-1 and qpsd8-1. *Cicer arietinum* cv. CDC Frontier (*Kabuli*) reference genome on the Pulse Crop database was used for identifying candidate genes (Varshney et al., 2013). Three candidate genes related to disease resistance were identified within the region flanking qpsd4-1 on LG 4, and four additional candidate genes were detected within the flanking qpsd8-1 region of LG 8 (**Table 2**).

Marker–Trait Association and Candidate Genes Through GWAS Analysis

For GWAS analysis, the SNP dataset on the ICRISAT accessions was filtered based on the minor allele frequencies (MAF <5%), missing data, and duplicate markers, and 229,965 SNPs were removed. Further linkage disequilibrium (LD) pruning removed 367,133 SNPs. After filtering and pruning, a total of 302,902 polymorphic SNPs remained for GWAS (**Supplementary Figure S4**).

PCA was performed to estimate population structure. The first two principal components (PC) explained 25% of the total variance, and the first five principal components (PC)

TABLE 1 | Statistical summary of QTLs for disease resistance in CRIL-7 [*C. reticulatum* (PI 599072) x *C. arietinum* (FLIP 84-92C)].

QTL	Linkage group	Position	Left marker	Right marker	LOD	Pve%	Additive effect
qpsd4-1	4	98	1793365SNPCa4	1674052SNPCa4	25.24	46.75	0.8783
qpsd8-1	8	36	2171187SNPCa8	2088222SNPCa8	3.31	4.53	–0.3109

[a]*QTL, names represent the traits, the linkage group number;* [b]*PVE, the percentage of phenotypic variance explained by the QTL;* [c]*Additive effect with positive values shows contribution toward greater resistance, while negative values show contribution toward greater susceptibility.*

TABLE 2 | Candidate genes with positions and annotations from QTL analysis.

Candidate gene	[a]Ch	Physical position	Functional annotation	References
Ca_07798	4	Ca4:1741231..1747336	JmjC domain-containing protein D	Hou et al. (2015)
Ca_07797	4	Ca4:1703853..1721273	WD-repeat family	Wang et al. (2009)
Ca_07799	4	Ca4:1759881..1762270	Zinc finger protein family	Gupta et al. (2012)
Ca_02390	8	Ca8:2091020..2093474	1-aminocyclopropane-1-carboxylate synthase	(Tekeoglu et al. (2000); Tekeoglu et al. (2002))
Ca_02384	8	Ca8:2135402..2136340	AT-hook DNA-binding protein	Yadeta et al. (2011)
Ca_02383	8	Ca8:2152196..2153734	Multidrug and toxic compound extrusion (MATE) transporters	Sun et al. (2011)
Ca_02389	8	Ca8:2094005..2096765	Protein kinase family	(Martin et al., 1993; Xu et al., 2013)

[a]*Ch. is the chromosome number of the significant SNP marker.*

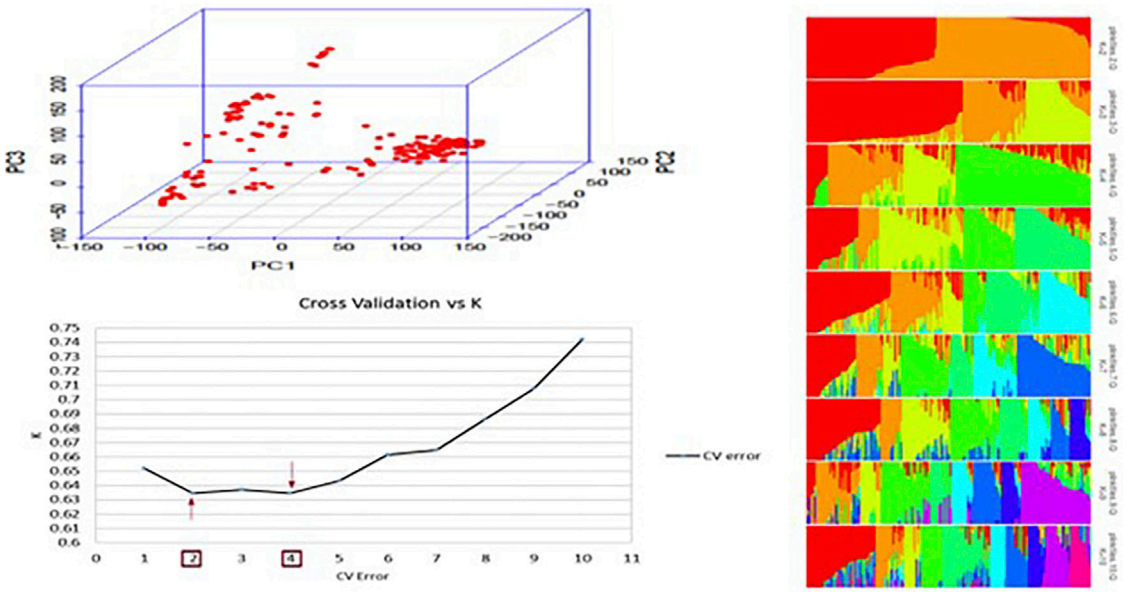


FIGURE 1 | Population structure of 184 chickpea accessions. **(A)** Principal component analysis (PCA) of all accessions based on 302,902 genome-wide SNPs. PCA divided the population into two subgroups shown in the circles. **(B)** Cross-validation plot for the SNP dataset plotted using the ADMIXTURE tool. K represents the number of subpopulations, and CV is the cross-validation error. The red arrows highlight the K value with the lowest CV errors. **(C)** Bar plots for K = 2–10. Each plot was created from 184 genotypes; each single vertical line represents each genotype, and each color represents one cluster.

explained 50% of the total variance. Graphs plotted using the first two PC explained the distribution of genotypes within different subpopulations (**Figure 1A**). In ADMIXTURE analysis with a total of 10 numbers of ancestral populations, the lowest cross-validation error was observed at K = 4, followed by k = 2 (with a minimum difference) were used as cofactors for GWAS (**Figure 1B**). Using K = 2, the panel was split into two subpopulations corresponding to *Desi* and *Kabuli* classes (**Figure 1C**), while with k = 4, genotypes were grouped according to seed shapes; angular, owl, and pea-shaped.

The greatest number of significant SNPs was observed on chromosome 4 (Ca4) (**Figure 2**). Marker Ca4_1765418 on chromosome 4 with a *p*-value of 1.65×10^{-7} was also detected within QTL qpsd4-1 that was revealed in the interspecific CRIL-7 population. Among the 35 significant SNPs, 29 were found in intergenic regions and 6 within genic regions (**Table 3**). Seven candidate genes that could be related to disease resistance were

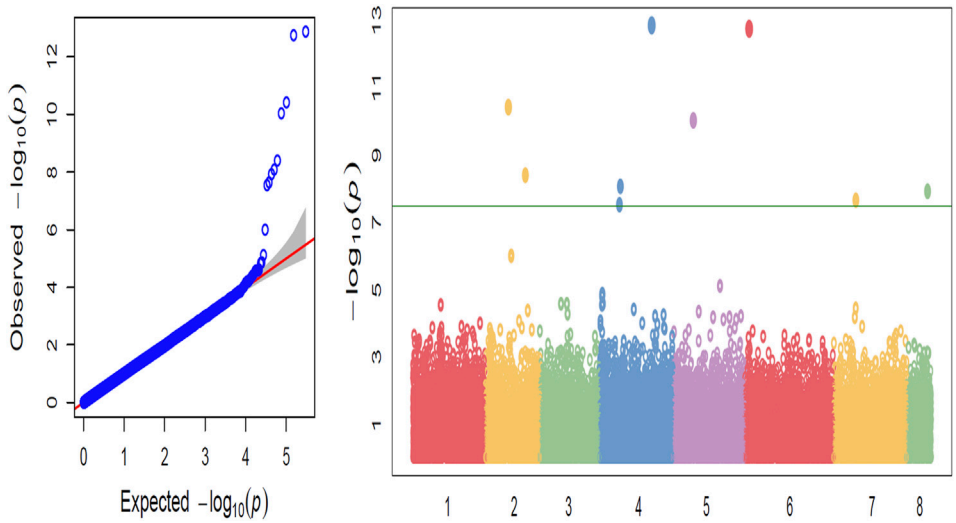


FIGURE 2 | (A) Quantile–quantile plots illustrating the comparison between expected and observed –log10(*p*)-values. **(B)** GWAS-derived Manhattan plot showing significant *p*-values associated with disease resistance using SNPs. The *x*-axis represents the relative density of reference genome-based SNPs physically mapped on 8 chickpea chromosomes, and *y*-axis indicates the –log10(*p*)-value. Colored dots represent individual SNPs, and markers significantly associated with disease resistance are above the Bonferroni cut-off (horizontal line).

found on chromosomes 2, 4, 6, 7, and 8 within 100 kb flanking regions of significant SNPs (**Table 4**).

DISCUSSION

GBS can be used to generate large-scale SNPs that are abundant in the genome allowing for the construction of high-density genetic maps and higher statistical power in association studies (Spencer et al., 2009; Yang et al., 2017). The CRIL-7 population has been used previously to develop linkage maps. These maps covered 981.6 cM and 1,174.4 cM across 9 linkage groups based on isozyme, inter simple sequence repeat (ISSR) loci, and RAPD markers (Santra et al., 2000b; Tekeoglu et al., 2002). In the present study, SNPs were used to resolve eight linkage groups spanning 1,186.30 cM.

QTLs and Candidate Genes Detected by Linkage Mapping

A major QTL for resistance to metalaxyl-resistant *P. ultimum*, qpsd4-1, was detected on LG 4 (**Table 1**). Previously, CRIL-7 has been evaluated for Ascochyta blight resistance, and QTLs associated with disease resistance were detected on LG 1, LG 4, and LG 8 (Santra et al., 2000a; Tekeoglu et al., 2004; Kumar et al., 2018). Two significant clusters of QTLs (QTLAR1 and QTLAR2) associated with resistance to Ascochyta blight have also been detected on LG 4 based on an analysis of RILs from a cross of cultivars Amit and ICCV 96029 (Deokar et al., 2019). Additionally, two QTLs (AB-Q-SR-4-1 and AB-Q-SR-4-2) on LG 4 associated with Ascochyta blight disease resistance were detected using mapping population (C 214' × 'ILC 3279) (Gisi and Sierotzki, 2008). These results and our detection of a major QTL for disease resistance on LG 4 suggest that genes for resistance to diverse chickpea pathogens are located on LG 4.

A total of seven genes associated with disease resistance were found in the QTL region on LG 4 and LG 8. Three candidate genes, Ca_07798, Ca_07797, and Ca_07799, were identified in QTL qpsd4-1. Ca_07798 is a JmjC domain-containing protein D, which is a group of histone lysine demethylases. This protein positively regulates rice defense against bacterial blight pathogen *Xanthomonas oryzae pv.* oryzae by epigenetically suppressing negative defense regulator H3K4me2/3 (Hou et al., 2015). The Ca_07797 gene belongs to a WD-repeat family that is involved in plant innate immune signaling pathway. Studies on maize showed that WDR-containing TTG1 protein–induced resistance against leaf blights (Ibraheem et al., 2010; Ibraheem et al., 2015). In tobacco, the interaction of TTG1-WDR with an elicitin protein (ParA1) from a pathogenic oomycete *Phytophthora parasitica* var. *nicotianae*-activated plant immune responses, including the generation of reactive oxygen species and programmed cell death (Wang et al., 2009). The Ca_07799 gene belongs to a zinc finger protein family that has been shown to play a critical role in disease resistance across many plant species (Gupta et al., 2012).

The Ca_02390 gene found on LG 8 encodes for 1-aminocyclopropane-1-carboxylate synthase. This enzyme initiates the conversion of S-adenosyl-L-methionine (SAM) into 1-aminocyclopropane-1-carboxylate (ACC), which is the precursor of ethylene and acts as a signaling molecule to regulate plant growth and reduce stress response (Polko and Kieber, 2019). The role of ACC and ethylene biosynthesis in plant defense against bacterial pathogens, including *Erwinia carotovora subsp.* carotovora and *Pseudomonas syringae* have been studied in *Arabidopsis thaliana* (L.) Heynh mutants where plants with reduced ACC production showed greater

TABLE 3 | Statistical summary of single nucleotide polymorphisms (SNPs) significantly associated with disease resistance trait.

SNP	[a]Ch	Position (bp)	Intergenic/Genic region	*p*.value	[b]MAF	FDR_Adjusted_*p*-values	Effect
Ca1_18665430	1	18665430	Intergenic	2.71E^{-05}	0.315217	4.69E^{-01}	0.145585
Ca1_44936836	1	44936836	Intergenic	9.94E^{-05}	0.304348	8.63E^{-01}	–0.20701
Ca2_15514982	2	15514982	Intergenic	3.88E^{-11}	0.11413	3.92E^{-06}	0.353739
Ca2_17444963	2	17444963	Genic	1.00E^{-06}	0.076087	3.04E^{-02}	0.221825
Ca2_22761060	2	22761060	Intergenic	8.85E^{-05}	0.168478	8.19E^{-01}	0.181644
Ca2_26901711	2	26901711	Intergenic	4.01E^{-09}	0.057065	2.43E^{-04}	0.413354
Ca2_28579479	2	28579479	Intergenic	3.99E^{-05}	0.372283	5.76E^{-01}	0.242365
Ca3_14221600	3	14221600	Intergenic	2.58E^{-05}	0.201087	4.69E^{-01}	0.197585
Ca3_18156610	3	18156610	Intergenic	2.59E^{-05}	0.05163	4.69E^{-01}	–0.38554
Ca3_18585784	3	18585784	Intergenic	5.18E^{-05}	0.076087	6.75E^{-01}	0.227405
Ca4_666303	4	666303	Intergenic	8.92E^{-05}	0.30163	8.19E^{-01}	0.172426
Ca4_1646485	4	1646485	Genic	2.79E^{-05}	0.078804	4.69E^{-01}	0.25133
Ca4_1765418	4	1765418	Intergenic	1.49E^{-05}	0.350543	3.48E^{-01}	0.15354
Ca4_1840434	4	1840434	Intergenic	1.35E^{-05}	0.222826	3.40E^{-01}	0.178898
Ca4_2249905	4	2249905	Intergenic	2.52E^{-05}	0.195652	4.69E^{-01}	0.197475
Ca4_13331455	4	13331455	Genic	2.93E^{-08}	0.133152	9.88E^{-04}	–0.33236
Ca4_14007934	4	14007934	Intergenic	8.06E^{-09}	0.201087	4.07E^{-04}	–0.18258
Ca4_22925858	4	22925858	Intergenic	3.88E^{-05}	0.269022	5.76E^{-01}	0.151793
Ca4_28639214	4	28639214	Intergenic	9.97E^{-05}	0.288043	8.63E^{-01}	0.150288
Ca4_34906194	4	34906194	Intergenic	1.37E^{-13}	0.076087	2.67E^{-08}	–0.50906
Ca4_37220588	4	37220588	Intergenic	6.10E^{-05}	0.388587	6.86E^{-01}	0.154591
Ca4_42835144	4	42835144	Intergenic	5.34E^{-05}	0.07337	6.75E^{-01}	0.229015
Ca5_13648457	5	13648457	Intergenic	9.30E^{-11}	0.0625	7.04E^{-06}	–0.51466
Ca5_17315201	5	17315201	Intergenic	4.23E^{-05}	0.108696	5.82E^{-01}	0.276777
Ca5_26889766	5	26889766	Intergenic	6.57E^{-05}	0.23913	6.86E^{-01}	0.182016
Ca5_31521962	5	31521962	Genic	7.48E^{-06}	0.201087	2.06E^{-01}	0.185133
Ca5_33795751	5	33795751	Genic	7.28E^{-05}	0.067935	7.11E^{-01}	0.233878
Ca5_37840504	5	37840504	Intergenic	6.28E^{-05}	0.146739	6.86E^{-01}	0.190138
Ca5_42306937	5	42306937	Genic	7.26E^{-05}	0.092391	7.11E^{-01}	0.22412
Ca5_45529647	5	45529647	Intergenic	5.82E^{-05}	0.26087	6.86E^{-01}	0.155,523
Ca6_2943215	6	2943215	Intergenic	1.76E^{-13}	0.146739	2.67E^{-08}	–0.40534
Ca7_14199535	7	14199535	Intergenic	6.34E^{-05}	0.084239	6.86E-^{01}	0.22165
Ca7_14818403	7	14818403	Intergenic	3.57E^{-05}	0.05163	5.69E^{-01}	–0.345
Ca7_14972314	7	14972314	Intergenic	2.25E^{-08}	0.070652	8.51E^{-04}	–0.2963
Ca8_14057710	8	14057710	Intergenic	1.19E^{-08}	0.125	5.15E^{-04}	–0.26813

[a]*Ch. is the chromosome number of the significant SNP marker.*
[b]*MAF is minor allele frequency.*

TABLE 4 | Significant SNP with candidate genes and annotations from GWAS.

SNP	[a]Ch	SNP reference/ alternate allele	Intergenic/ Genic region	Closest candidate genes	Gene position (bp)	Functional annotation	References
Ca4_34906194	4	G/A	Intergenic	Ca_19996	34854043–34857882	Cellulose synthase–like protein	Douchkov et al. (2016)
Ca6_2943215	6	A/G	Intergenic	Ca_10436	2942150–2942952	Calmodulin-binding protein	Upadhyaya et al. (2001)
Ca2_26901711	2	C/T	Intergenic	Ca_17276	26889005–26890367	LUPR1 protein	Dong et al. (2015)
Ca2_26901711	2	C/T	Intergenic	Ca_17277	26904743–26905771	O-methyltransferase family	Yang et al. (2017)
Ca4_14007934	4	T/C	Intergenic	Ca_04625	14016692–14018353	Thiamine thiazole synthase family	(72)
Ca8_14057710	8	G/A	Intergenic	Ca_22742	14006548–14013672	Ethylene-responsive transcription factor 1-like protein	Dong et al. (2015)
Ca7_14972314	7	T/C	Intergenic	Ca_09957	14960121–14966841	Histidine kinase protein	Heo et al. (1999)

[a]*Ch. is the chromosome number of the significant SNP marker.*

disease susceptibility (Norman-Setterblad et al., 2000; Guan et al., 2015). Recently, upregulation of genes involved in ethylene biosynthesis was detected in resistant apple seedling reactions to infection by *P. ultimum* (Zhu and Saltzgiver, 2020). Ca_02384 on LG 8 encodes an AT-hook DNA-binding protein that binds to minor groove DNA and alters gene expression. Genes such as AHL19 encoding an AT-hook DNA-binding protein are associated with enhanced

disease resistance in *A. thaliana* to Verticillium wilt caused by *V. dahliae*, *V. albo-atrum*, *and V. longisporum* (Yadeta et al., 2011). The Ca_02383 gene on LG 8 belongs to the family of multidrug and toxic compound extrusion (MATE) transporters associated with plant disease resistance during pathogen interaction. The expression of MATE genes in plants is induced by pathogen attack (Sun et al., 2011). Members of the MATE family such as enhanced disease susceptibility 5 (EDS5) and activated disease susceptibility 1 (ADS1) function as negative regulators of plant immune systems by reducing basal resistance during pathogen interaction or by negatively regulating the accumulation of salicylic acid and pathogenesis-related 1 (PR1) gene expression (Nawrath et al., 2002; Sun et al., 2011). Another candidate gene on LG 8, Ca_02389, belongs to the protein kinase family. Members of this family have also been shown to be upregulated in response to pathogens, for example, *Xanthomonas oryzae pv. oryzicola* (*Xoc*) in *Oryza sativa* and *Pseudomonas syringae* pv tomato (*pto*) in *Solanum lycopersicum* (Martin et al., 1993; Xu et al., 2013). Recently, GWAS of common bean identified several candidate genes associated with resistance to *P. ultimum*, including genes for protein kinase superfamily proteins and MAPK/ERK kinase 1 (Dramadri et al., 2020).

GWAS Analysis of Resistance in Chickpea to *P. ultimum*

Different statistical models were deployed using FarmCPU to assess population structure and kinship in the chickpea diversity panel. A combination of these models using FarmCPU separates a mixed linear model (MLM) into a random effect and a fixed-effect model, which reduces false positives and false negatives caused by kinship and population structure and gives highly significant SNP markers (Lipka et al., 2012). The best model was selected based on MSD value (**Supplementary Table S5**), with a low MSD value indicating less deviation from the expected distribution of *p*-values, signifying lower type I error of the selected model. This study identified a total of seven candidate gene mapping to 11 loci associated with resistance in chickpea to *P. ultimum*. PCA and ADMIXTURE analyses revealed two major groups within the core collections (**Figure 1**) corresponding to *Desi* and *Kabuli* market classes.

GWAS identified many SNPs associated with disease-resistance–related traits. Gene Ca_19996 encodes cellulose synthase–like protein, which inhibits the progress of the fungal penetration peg during powdery mildew infection caused by *Blumeria graminis f. sp. hordei (Bgh)* in barley (*Hordeum vulgare* L.) (Douchkov et al., 2016). Gene Ca_09957 encodes a histidine kinase protein involved in seed maturation and disease resistance against fungal and bacterial pathogens (Pham et al., 2012). Ca_17277 belongs to the O-methyltransferase family of enzymes that play a significant role in plant stress and disease resistance. Studies on corn (*Zea mays* L.) and wheat (*Triticum aestivum* L.) demonstrated that caffeoyl-CoA O-methyltransferase conferred resistance against southern leaf blight, gray leaf spot, and sharp eyespot disease, respectively (Yang et al., 2017; Wang et al., 2018). Ca_04625 encodes a thiamine thiazole synthase that increases resistance to fungal pathogens by enhancing anti-oxidative capacity and inducing systemic acquired resistance (SAR) in diverse plant species, including *Oryza sativa* L. *A. thaliana* (L.) Heynh., *Nicotiana sp*., and *Cucumis sativus* L. (Goyer, 2010). Ca_22742 encodes an ethylene-responsive transcription factor 1–like protein. Ethylene-responsive transcription factors play a critical role in the plant defense system by regulating pathogenesis-related (PR) gene expression, including effectors GmERF5 and GmERF113, and contribute to resistance against root and seed rot caused by Oomycete pathogens, including *Phytopthora nicotianae* and *Py. sojae* (Goyer, 2010; Yang et al., 2017). The Ca_17276 gene encodes LUPR1 protein that is upregulated in response to the *Hyaloperonospora parasitica* (*LURP*) cluster in part of the *A. thaliana*. The LUPR1 gene has been associated with resistance to oomycetes *Hyaloperonospora parasitica* and *Py. infestans* (Dong et al., 2015). The Ca_10436 gene encodes a calmodulin-binding protein, which has activated and enhanced resistance to a broad spectrum of pathogens in *Nicotiana tabacum* and *A. thaliana* (Zhao et al., 2017). Additionally, calmodulin-binding proteins in *A. thaliana* and *Hordeum vulgare* have been shown to confer resistance to powdery mildew by interacting with MLO (powdery mildew-resistance gene o) protein (Heo et al., 1999).

CONCLUSION

In this study, we used an interspecific chickpea population to identify one major and one minor QTL associated with resistance to *P. ultimum*. We also identified 35 SNPs and 14 candidate genes associated with disease resistance based on the GWAS of a chickpea diversity panel. SNP Ca4_1765418, detected by GWAS on chromosome 4, was located within QTL qpsd4-1 that was revealed in the interspecific CRIL-7 population. These findings suggest this region of the genome should be examined more closely to identify genes conditioning disease resistance. Significant QTLs must be validated in different chickpea populations before the markers can be widely used for breeding. The present study provides tools to enable MAS for resistance to *P. ultimum* and identified genomic domains and candidate genes involved in the resistance of chickpea to soilborne diseases.

DATA AVAILABILITY STATEMENT

The set of 184 genotypes used for GWAS analysis in the present study is part of 300 genotypes that were sequenced in an earlier study (Varshney et al., 2013). Sequencing data for 184 genotypes as part of 300 genotypes have been reported in the earlier study. These data are available in the NCBI under accession code SRA: SRP096939; BioProject: PRJNA362278, and in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb with accession code CNP0000370.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology: CA, WC, RV, and GV. Data collection: CA under the guidance of WC, RV, and GV. Data analysis: CA under the guidance of RV and GV. Original draft preparation: CA. Review and editing: CA, WC, RV, and GV. Project funding acquisition and administration: GV. All authors have read and approved the final version of the manuscript.

FUNDING

This research was supported by the United States Department of Agriculture-Agricultural Research Service project #2090–21000-034–000D (Improving Genetic Resources and Disease Management for Cool Season Food Legumes).

ACKNOWLEDGMENTS

The authors acknowledge Mahendar Thudi and Prasad Bajaj at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India, for their invaluable guidance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.945787/full#supplementary-material

REFERENCES

Agarwal, C., Chen, W., Coyne, C., and Vandemark, G. (2020). Identifying Sources of Resistance in Chickpea to Seed Rot and Seedling Damping-off Caused by Metalaxyl-resistant *Pythium Ultimum*, [Internet]. *Crop Sci.* 61, 1739. [cited 2021 Jan 20];csc2.20424. doi:10.1002/csc2.20424

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinforma.* 12 (1), 246. [cited 2020 Apr 20]Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-246. doi:10.1186/1471-2105-12-246

Bayaa, B., and Chen, W. (2011). "Ascochyta Blight of Chickpea," in *Compendium of Chickpea and Lentil Diseases and Pests*. Editors W. Chen, H. C. Sharma, and F. J. Muehlbauer (St. Paul, MN: APS Press).

Casas, A. T., Kaiser, W. J., and Ingram, D. M. (1990). Control of Pythium Seed Rot and Preemergence Damping-Off of Chickpea in the U.S. Pacific Northwest and Spain. *Plant Dis.* 74, 563–569. doi:10.1094/pd-74-0563

Chen, W., and Van Vleet, S. (2016). *WSU - Chickpea Damping-Off Due to Metalaxyl-Resistant Pythium: An Emerging Disease in the Palouse*. [Internet]. Washington: Washingtom State University Extension. Pullman, Washington : Washington State University. Extension., 1–4. [cited 2020 Apr 20]Available from: https://pubs.wsu.edu/ItemDetail.aspx?ProductID=15853&SeriesCode=&CategoryID=150&Keyword=.

Cingolani, P., Platts, A., Wang, le L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila melanogaster* Strain W1118; Iso-2; Iso-3. *Fly. (Austin)* 6 (2), 80–92. doi:10.4161/fly.19695

Deokar, A., Sagi, M., and Tar'an, B. (2019). Genome-wide SNP Discovery for Development of High-Density Genetic Map and QTL Mapping of Ascochyta Blight Resistance in Chickpea (Cicer Arietinum L.). *Theor. Appl. Genet.* 132, 1861–1872. doi:10.1007/s00122-019-03322-3

Dong, L., Cheng, Y., Wu, J., Cheng, Q., Li, W., Fan, S., et al. (2015). Overexpression of GmERF5, a New Member of the Soybean EAR Motif-Containing ERF Transcription Factor, Enhances Resistance to Phytophthora Sojae in Soybean. *J. Exp. Bot.* 66, 2635–2647. doi:10.1093/jxb/erv078

Douchkov, D., Lueck, S., Hensel, G., Kumlehn, J., Rajaraman, J., Johrde, A., et al. (2016). The Barley (Hordeum Vulgare) Cellulose Synthase-like D2 Gene (HvCslD2) Mediates Penetration Resistance to Host-Adapted and Nonhost Isolates of the Powdery Mildew Fungus, [Internet]. *New Phytol.* 212 (2), 421–433. [cited 2021 Oct 24]. doi:10.1111/nph.14065

Dramadri, I. O., Amongi, W., Kelly, J. D., and Mugisha Mukankusi, C. (2020). "Genome-wide Association Analysis of Resistance to *Pythium Ultimum* in Common Bean (*Phaseolus vulgaris*)," [Internet] in *Plant Breed*. Editor W. Link (New Jersey: Wiley). [cited 2020 Nov 17];pbr.12855. doi:10.1111/pbr.12855

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 (3), 45. [cited 2020 Nov 16]. doi:10.1007/s10681-018-2125-3

Gisi, U., and Sierotzki, H. (2008). Fungicide Modes of Action and Resistance in Downy Mildews. *Eur. J. Plant Pathol.* 122 (1), 157–167. doi:10.1007/s10658-008-9290-5

Goyer, A. (2010). Thiamine in Plants: Aspects of its Metabolism and Functions. *Phytochemistry* 71, 1615–1624. doi:10.1016/j.phytochem.2010.06.022

Guan, R., Su, J., Meng, X., Li, S., Liu, Y., Xu, J., et al. (2015). Multilayered Regulation of Ethylene Induction Plays a Positive Role in Arabidopsis Resistance against *Pseudomonas syringae*. *Plant Physiol.* 169 (1), 299–312. doi:10.1104/pp.15.00659

Gupta, S. K., Rai, A. K., Kanwar, S. S., and Sharma, T. R. (2012). Comparative Analysis of Zinc Finger Proteins Involved in Plant Disease Resistance, [Internet]. *PLoS One* 7 (8), e42578. [cited 2020 Apr 20]Available from: http://www.ncbi.nlm.nih.gov/pubmed/22916136. doi:10.1371/journal.pone.0042578

Heo, W. D., Lee, S. H., Kim, M. C., Kim, J. C., Chung, W. S., Chun, H. J., et al. (1999). Involvement of Specific Calmodulin Isoforms in Salicylic Acid-independent Activation of Plant Disease Resistance Responses. *Proc. Natl. Acad. Sci. U. S. A.* 96 (2), 766–771. doi:10.1073/pnas.96.2.766

Hou, Y., Wang, L., Wang, L., Liu, L., Li, L., Sun, L., et al. (2015). JMJ704 Positively Regulates Rice Defense Response against Xanthomonas Oryzae Pv. Oryzae Infection via Reducing H3K4me2/3 Associated with Negative Disease Resistance Regulators. *BMC Plant Biol.* 15 (1), 286. doi:10.1186/s12870-015-0674-3

Hyun, M. K., 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, [Internet]. *Genetics* 178 (3), 1709–1723. [cited 2020 Apr 20]Available from: http://www.ncbi.nlm.nih.gov/pubmed/18385116. doi:10.1534/genetics.107.080101

Ibraheem, F., Gaffoor, I., and Chopra, S. (2010). Flavonoid Phytoalexin-dependent Resistance to Anthracnose Leaf Blight Requires a Functional Yellow Seed1 in Sorghum Bicolor. *Genetics* 184 (4), 915–926. doi:10.1534/genetics.109.111831

Ibraheem, F., Gaffoor, I., Tan, Q., Shyu, C.-R., and Chopra, S. (2015). A Sorghum MYB Transcription Factor Induces 3-deoxyanthocyanidins and Enhances Resistance against Leaf Blights in Maize. *Molecules* 20 (2), 2388–2404. doi:10.3390/molecules20022388

Jendoubi, W., Bouhadida, M., Millan, T., Kharrat, M., Gil, J., Rubio, J., et al. (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, [Internet]. *Euphytica* 210, 119–133. [cited 2020 Nov 17]. doi:10.1007/s10681-016-1712-4

Jingade, P., and Ravikumar, R. L. (2015). Development of Molecular Map and Identification of QTLs Linked to Fusarium Wilt Resistance in Chickpea, [Internet]. *J. Genet.* 94, 723–729. [cited 2020 Nov 17]Available from: http://www.ias.ac.in/. doi:10.1007/s12041-015-0589-7

Kaiser, W. J., and Hannan, R. M. (1983). Etiology and Control of Seed Decay and Preemergence Damping-Off of Chickpea byPythium Ultimum. *Plant Dis.* 67 (1), 77. [cited 2020 Apr 19]Available from: http://agris.fao.org/agris-search/search.do?recordID=US19900085897. doi:10.1094/pd-67-77

Karadi, A., Samineni, S., Sajja, S., Sharma, M., Thudi, M., Mallikarjuna, B. P., et al. (2021). Molecular Mapping of Dry Root Rot Resistance Genes in Chickpea (Cicer Arietinum L.), [Internet]. *Euphytica* 217, 123. [cited 2021 Oct 24] Available from: https://research-repository.uwa.edu.au/en/publications/molecular-mapping-of-dry-root-rot-resistance-genes-in-chickpea-ci. doi:10.1007/s10681-021-02854-4

Kumar, K., Purayannur, S., Kaladhar, V. C., Parida, S. K., and Verma, P. K. (2018). mQTL-Seq and Classical Mapping Implicates the Role of an AT-HOOK MOTIF CONTAINING NUCLEAR LOCALIZED (AHL) Family Gene in Ascochyta Blight Resistance of Chickpea. *Plant Cell Environ.* 41 (9), 2128–2140. doi:10.1111/pce.13177

Li, H., Ye, G., and Wang, J. (2007). A Modified Algorithm for the Improvement of Composite Interval Mapping, [Internet]. *Genetics* 175 (1), 361–374. [cited 2022 Apr 26]Available from: https://academic.oup.com/genetics/article/175/1/361/6062043. doi:10.1534/genetics.106.066811

Li, H., Rodda, M., Gnanasambandam, A., Aftab, M., Redden, R., Hobson, K., et al. (2015). Breeding for Biotic Stress Resistance in Chickpea: Progress and Prospects. *Euphytica* 204 (2), 257–288. doi:10.1007/s10681-015-1462-8

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: Genome Association and Prediction Integrated Tool. *Bioinforma. [Internet]* 28 (18), 2397–2399. doi:10.1093/bioinformatics/bts444

Mamidi, S., Chikara, S., Goos, R. J., Hyten, D. L., Annam, D., Moghaddam, S. M., et al. (2011). Genome-Wide Association Analysis Identifies Candidate Genes Associated with Iron Deficiency Chlorosis in Soybean, [Internet]. *Plant Genome* 4 (3), 154–164. doi:10.3835/plantgenome2011.04.0011

Mannur, D. M., Babbar, A., Thudi, M., Sabbavarapu, M. M., Roorkiwal, M., Yeri, S. B., 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.), [Internet]. *Mol. Breed.* 39 (1), 2. [cited 2020 Nov 17]. doi:10.1007/s11032-018-0908-9

Martin, F. N., and Loper, J. E. (1999). Soilborne Plant Diseases Caused by Pythium spp.: Ecology, Epidemiology, and Prospects for Biological Control, [Internet]. *CRC Crit. Rev. Plant Sci.* 18, 111–181. [cited 2021 Jan 20]Available from: https://www.tandfonline.com/action/journalInformation?journalCode=bpts20.

Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Frary, A., Ganal, M. W., Spivey, R., et al. (1993). Map-based Cloning of a Protein Kinase Gene Conferring Disease Resistance in Tomato. *Science* 262 (5138), 1432–1436. doi:10.1126/science.7902614

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 (3), 269–283. doi:10.1016/j.cj.2015.01.001

Muehlbauer, F. J., Temple, S. R., and Chen, W. (2004). Registration of 'Sierra' Chickpea, [Internet]. *Crop Sci.* 44, 1864. [cited 2020 Apr 20]. doi:10.2135/cropsci2004.1864

Nawrath, C., Heck, S., Parinthawong, N., and Métraux, J.-P. (2002). EDS5, an Essential Component of Salicylic Acid-dependent Signaling for Disease Resistance in Arabidopsis, Is a Member of the MATE Transporter Family. *Plant Cell* 14 (1), 275–286. doi:10.1105/tpc.010376

Norman-Setterblad, C., Vidal, S., and Palva, E. T. (2000). Interacting Signal Pathways Control Defense Gene Expression in Arabidopsis in Response to Cell Wall-Degrading Enzymes from Erwinia Carotovora, [Internet]. *Mol. Plant-Microbe Interact.* 13, 430–438. [cited 2020 Apr 20]. doi:10.1094/MPMI.2000.13.4.430

Pham, J., Liu, J., Bennett, M. H., Mansfield, J. W., and Desikan, R. (2012). Arabidopsis Histidine Kinase 5 Regulates Salt Sensitivity and Resistance against Bacterial and Fungal Infection. *New Phytol.* 194 (1), 168–180. doi:10.1111/j.1469-8137.2011.04033.x

Plaats-Niterink, A. van der. (1981). Monograph of the Genus Pythium [Internet]. [cited 2020 Apr 18] Available from: http://www.wi.knaw.nl/publications/1021/content_files/content.htm

Polko, J. K., and Kieber, J. J. (2019). 1-Aminocyclopropane 1-Carboxylic Acid and its Emerging Role as an Ethylene-independent Growth Regulator. *Front. Plant Sci.* 10, 1602. doi:10.3389/fpls.2019.01602

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795

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 Ascochyta Blight in Chickpea (Cicer Arietinum L.). *Euphytica* 193 (1), 121–133. doi:10.1007/s10681-013-0959-2

Santra, D. K., A4 - Tekeoglu, M., A4 - Ratnaparkhe, M., A4 - Kaiser, M., Kaiser, W. J., Muehlbauer, F. J., et al. (2000). Identification and Mapping of QTLs Conferring Resistance to Ascochyta Blight in Chickpea. *Crop Sci.* 40 (6), 1606–1612. doi:10.2135/cropsci2000.4061606x

Santra, D. K., Tekeoglu, M., Ratnaparkhe, M., Kaiser, W. J., and Muehlbauer, F. J. (2000). Identification and Mapping of QTLs Conferring Resistance to Ascochyta Blight in Chickpea. *Crop Sci.* 40 (6), 1606–1612. doi:10.2135/cropsci2000.4061606x

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

Spencer, C. C., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing Genome-wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *PLoS Genet.* 5 (5), e1000477. doi:10.1371/journal.pgen.1000477

Sun, X., Gilroy, E. M., Chini, A., Nurmberg, P. L., Hein, I., Lacomme, C., et al. (2011). ADS1 Encodes a MATE-Transporter that Negatively Regulates Plant Disease Resistance. *New Phytol.* 192 (2), 471–482. doi:10.1111/j.1469-8137.2011.03820.x

Tekeoglu, M., Işık, M., and Muehlbauer, F. (2004). QTL Analysis of Ascochyta Blight Resistance in Chickpea. *Turkish J Agric*28, 183–187.

Tekeoglu, M., Rajesh, P. N., and Muehlbauer, F. J. (2002). Integration of Sequence Tagged Microsatellite Sites to the Chickpea Genetic Map, [Internet]. *Theor. Appl. Genet.* 105 (6–7), 847–854. [cited 2020 Apr 19]Available from: http://www.ncbi.nlm.nih.gov/pubmed/12582909. doi:10.1007/s00122-002-0993-2

Tekeoglu, M., Santra, D. K., Kaiser, W. J., and Muehlbauer, F. J. (2000). Ascochyta Blight Resistance Inheritance in Three Chickpea Recombinant Inbred Line Populations. *Crop Sci.* 40, 1251–1256. doi:10.2135/cropsci2000.4051251x

Toker, C. (2009). A Note on the Evolution of Kabuli Chickpeas as Shown by Induced Mutations in Cicer Reticulatum Ladizinsky. *Genet. Resour. Crop Evol.* 56, 7–12. doi:10.1007/s10722-008-933

Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a Chickpea Core Subset Using Geographic Distribution and Quantitative Traits. *Crop Sci.* 41 (1), 206–210. doi:10.2135/cropsci2001.411206x

Vandemark, G., Guy, S. O., Chen, W., McPhee, K., Pfaff, J., Lauver, M., et al. (2015). Registration of 'Nash' Chickpea. *J. Plant Registrations* 9, 275–278. doi:10.3198/jpr2014.07.0047crc

Vandemark, G. J., Brick, M. A., Osorno, J. M., Kelly, J. D., and Urrea, C. A. (2014). "Edible Grain Legumes," in *Madison, WI: Yield Gains in Major U.S. Field Crops*. Editors S. Smith, B. Diers, and J. Specht BC (US: ASA, CSSA, SSSA), 87–124.

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft Genome Sequence of Chickpea (Cicer Arietinum) Provides a Resource for Trait Improvement. *Nat. Biotechnol.* 31 (3), 240–246. doi:10.1038/nbt.2491

Wang, M., Van Vleet, S., McGee, R., Paulitz, T. C., Porter, L. D., Vandemark, G., et al. (2020). Chickpea Seed Rot and Damping-Off Caused by Metalaxyl-Resistant Pythium Ultimum and its Management with Ethaboxam, [Internet]. *Plant Dis.* 105, 1728. [cited 2021 Jan 20]; Available from: https://apsjournals.apsnet.org/doi/abs/10.1094/PDIS-08-20-1659-RE. doi:10.1094/PDIS-08-20-1659-RE

Wang, M., Zhu, X., Wang, K., Lu, C., Luo, M., Shan, T., et al. (2018). A Wheat Caffeic Acid 3-O-Methyltransferase TaCOMT-3D Positively Contributes to Both Resistance to Sharp Eyespot Disease and Stem Mechanical Strength. *Sci. Rep.* 8 (1), 6543–6614. doi:10.1038/s41598-018-24884-0

Wang, Y., Liu, R., Chen, L., Wang, Y., Liang, Y., Wu, X., et al. (2009). Nicotiana tabacumTTG1 Contributes to ParA1-Induced Signalling and Cell Death in Leaf Trichomes. *J. Cell Sci.* 122 (15), 2673–2685. doi:10.1242/jcs.049023

Winter, P., Benko-Iseppon, A.-M., Hüttel, 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. reticulatum* Cross: Localization of Resistance Genes for fusarium Wilt Races 4 and 5. *Theor. Appl. Genet.* 101 (7), 1155–1163. doi:10.1007/s001220051592

Xu, M.-R., Huang, L.-Y., Zhang, F., Zhu, L.-H., Zhou, Y.-L., and Li, Z.-K. (2013). Genome-Wide Phylogenetic Analysis of Stress-Activated Protein Kinase Genes in Rice (OsSAPKs) and Expression Profiling in Response to Xanthomonas Oryzae Pv. Oryzicola Infection. *Plant Mol. Biol. Rep.* 31 (4), 877–885. doi:10.1007/s11105-013-0559-2

Yadeta, K. A., Hanemian, M., Smit, P., Hiemstra, J. A., Pereira, A., Marco, Y., et al. (2011). The *Arabidopsis thaliana* DNA-Binding Protein AHL19 Mediates Verticillium Wilt Resistance. *MPMI* 24 (12), 1582–1591. doi:10.1094/mpmi-04-11-0090

Yang, Q., He, Y., Kabahuma, M., Chaya, T., Kelly, A., Borrego, E., et al. (2017). A Gene Encoding Maize Caffeoyl-CoA O-Methyltransferase Confers Quantitative Resistance to Multiple Pathogens. *Nat. Genet.* 49 (9), 1364–1372. doi:10.1038/ng.3919

Zhao, Y., Chang, X., Qi, D., Dong, L., Wang, G., Fan, S., et al. (2017). A Novel Soybean ERF Transcription Factor, GmERF113, Increases Resistance to Phytophthora Sojae Infection in Soybean. *Front. Plant Sci.* 8, 299. Available from: https://www.frontiersin.org/article/10.3389/fpls.2017.00299. doi:10.3389/fpls.2017.00299

Zhu, Y., and Saltzgiver, M. (2020). A Systematic Analysis of Apple Root Resistance Traits to Pythium Ultimum Infection and the Underpinned Molecular Regulations of Defense Activation, [Internet]. *Hortic. Res.* 7, 62. [cited 2021 Oct 25]Available from: https://www.nature.com/articles/s41438-020-0286-4. doi:10.1038/s41438-020-0286-4

Zohary, D., Hopf, M., and Weiss, E. (2012). *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. [Internet]. 4th ed. Oxford: Oxford University Press, 264. Available from: https://www.oxfordscholarship.com/10.1093/acprof:osobl/9780199549061.001.0001/acprof-9780199549061.

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 © 2022 Agarwal, Chen, Varshney and Vandemark. 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 | Frontiers in Genetics

TYPE Original Research
PUBLISHED 17 August 2022
DOI 10.3389/fgene.2022.953833

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Nisha Singh,
Gujarat Biotechnology University, India
Valerio Hoyos-Villegas,
McGill University, Montreal, Canada
Karansher Singh Sandhu,
Bayer Crop Science, United States

*CORRESPONDENCE
Javaid Akhter Bhat,
javid.akhter69@gmail.com
Rajeev K. Varshney,
rajeev.varshney@murdoch.edu.au
Deyue Yu,
dyyu@njau.edu.cn

†These authors share first authorship

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

RECEIVED 26 May 2022
ACCEPTED 22 July 2022
PUBLISHED 17 August 2022

CITATION
Bhat JA, Adeboye KA, Ganie SA, Barmukh R, Hu D, Varshney RK and Yu D (2022), Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (Glycine max L.).
Front. Genet. 13:953833.
doi: 10.3389/fgene.2022.953833

COPYRIGHT
© 2022 Bhat, Adeboye, Ganie, Barmukh, Hu, Varshney and Yu. 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.

Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (*Glycine max* L.)

Javaid Akhter Bhat[1,2†*], Kehinde Adewole Adeboye[3†], Showkat Ahmad Ganie[4], Rutwik Barmukh[5], Dezhou Hu[1], Rajeev K. Varshney[5,6*] and Deyue Yu[1*]

[1]Soybean Research Institution, National Center for Soybean Improvement, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, [2]International Genome Center, Jiangsu University, Zhenjiang, China, [3]Department of Agricultural Technology, Ekiti State Polytechnic, Isan, Nigeria, [4]Plant Molecular Science and Centre of Systems and Synthetic Biology, Department of Biological Sciences, Royal Holloway University of London, Surrey, United Kingdom, [5]Center of Excellence in Genomics & Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [6]Murdoch's Centre for Crop & Food Innovation, State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Perth, WA, Australia

Identifying the genetic components underlying yield-related traits in soybean is crucial for improving its production and productivity. Here, 211 soybean genotypes were evaluated across six environments for four yield-related traits, including seed yield per plant (SYP), number of pods per plant number of seeds per plant and 100-seed weight (HSW). Genome-wide association study (GWAS) and genomic prediction (GP) analyses were performed using 12,617 single nucleotide polymorphism markers from NJAU 355K SoySNP Array. A total of 57 SNPs were significantly associated with four traits across six environments and a combined environment using five Genome-wide association study models. Out of these, six significant SNPs were consistently identified in more than three environments using multiple GWAS models. The genomic regions (±670 kb) flanking these six consistent SNPs were considered stable QTL regions. Gene annotation and in silico expression analysis revealed 15 putative genes underlying the stable QTLs that might regulate soybean yield. Haplotype analysis using six significant SNPs revealed various allelic combinations regulating diverse phenotypes for the studied traits. Furthermore, the GP analysis revealed that accurate breeding values for the studied soybean traits is attainable at an earlier generation. Our study paved the way for increasing soybean yield performance within a short breeding cycle.

KEYWORDS

haplotype-based breeding, GWAS, legumes, seed yield, candidate genes, crop improvement

Introduction

Soybean (*Glycine* max L. Merr.) is one of the most important food legume crops cultivated globally (Hina et al., 2020). Many countries including China are highly dependent on soybean imports to fulfil their domestic needs (Liu et al., 2018; Karikari et al., 2019). Improving soybean yield has been a primary objective of breeders from China and several other countries across the world (Karikari et al., 2019). Seed yield of soybean is governed by yield-related traits such as the number of pods per plant, number of seeds per plant, and seed weight (Bianchi et al., 2020). However, yield and its component traits are complex, being controlled by multiple genes and are considerably affected by the environment (Liu et al., 2011).

Many studies have revealed the genetic basis of yield-related traits in staple crops such as maize (Badu-Apraku et al., 2020; Yang et al., 2020), rice (Yue et al., 2015; Adeboye K. A. et al., 2021), and soybean (Karikari et al., 2019; Hu et al., 2020) by using bi-parental mapping populations. To date, several hundred QTLs regulating yield and yield-related traits have been mapped across the soybean genome, and many QTLs were consistently identified in different populations (http://www.soybase.org). All these studies have validated the role of linkage mapping as an efficient approach to dissecting the genetic basis of quantitative traits (Karikari et al., 2019; Hina et al., 2020). However, a major limitation of the linkage mapping approach is its dependence on limited diversity existing within segregating populations derived from two contrasting parents. By contrast, GWAS has emerged as an alternative approach, which is more efficient in the identification of QTLs regulating quantitative traits by utilizing natural diversity existing within crop germplasm and the use of high-density genetic markers (Zargar et al., 2015; Yu et al., 2019). Importantly, GWAS has a higher potential to identify candidate genes regulating the trait of interest because of a reduced level of genomic linkage disequilibrium (LD) (Alqudah et al., 2020).

Advances in next-generation sequencing technologies have enabled a wider availability of high-throughput sequencing and genotyping platforms (Bhat et al., 2020; Sahu et al., 2020). As a result, genomics-assisted breeding (GAB) has emerged as the method of choice for crop improvement in plant breeding programs (Varshney et al., 2021). Although, both linkage mapping and GWAS approaches are being successfully used for the identification of QTLs/candidate genes in crop plants, limited allelic diversity and genomic resolution associated with linkage mapping are addressed by GWAS in the gene identification process (Brachi et al., 2011). GWAS has been efficiently used to identify the QTLs/genes underlying various yield-related traits in soybeans such as seed protein and oil content (Zhang et al., 2019), agronomy (Zatybekov et al., 2017), salt tolerance (Zeng et al., 2017), and yield-related traits (Hu et al., 2020). Another genomic-based plant breeding approach is the genomic prediction (GP). Here, phenotypic traits or performance of an individual is predicted based on genomic data. GP is currently being used in multiple plant species to estimate the genetic values (genotypic estimated breeding values (GEBVs)) of the individual genotypes based on the genome-wide genotypic data without the need for phenotypic data (Habier et al., 2007; Bhat et al., 2016; Crossa et al., 2017). The GP allows for the captures of QTLs with minor effects since the model is based on the genome-wide marker data rather than few markers as in the marker-assisted selection model. Thus, it has a great potential for improving the genetic gain associated with yield and other complex traits within a limited time frame in different crop plants (Crossa et al., 2017; Voss-Fels et al., 2019; Lebedev et al., 2020). Moreover, GP has been used in soybean for improving different traits such as cyst nematode infestation (Ravelombola et al., 2020), disease resistance (Rolling et al., 2020), agronomic traits (Beche et al., 2021), and seed yield (Mendonça et al., 2020). The results of these studies have demonstrated the potential of GP for improving complex traits in soybean. Furthermore, advanced sequencing technologies are providing high accuracy in gene and haplotype mining in crop germplasm (Bevan et al., 2017; Bhat et al., 2021).

The present study analyzes the genetic basis of yield-related traits in summer planting soybean genotypes grown in soybean growing areas of China. We evaluated 211 diverse soybean genotypes across six environments for four yield-related traits, including seed yield per plant (SYP), number of pods per plant (PPP), number of seeds per plant (SPP), and 100-seed weight (HSW). Based on the phenotypic performance, genome-wide association analysis was conducted to identify QTLs associated with the studied traits using five different models. The genes underlying the identified QTLs were validated based on RNA-seq data for soybean tissues. Furthermore, superior haplotypes and alleles were identified within the genomic regions associated with the studied traits. Also, genetic values of individual genotypes were estimated based on the studied yield-related traits to facilitate the selection of soybean for improved yield performance.

Materials and methods

Plant materials and field experiment

The GWAS panel of soybean used in the current study consists of 211 diverse genotypes; which include 201 genotypes originating from 25 provinces of China that represents all three ecological habitats of China (Zhang S. et al., 2021) and ten genotypes from the United States, Japan, and Brazil (Zhang et al., 2015). This soybean germplasm was evaluated at three different locations in China viz., Nanjing, Nantong, and Yangzhou, for two consecutive years. This makes a total of six different environments viz., E1 and E2 (Nanjing); E3 and E4 (Nantong); and E5 and E6 (Yangzhou). The study

location was previously described by Bhat et al. (2022). Nanjing (32°12′ N, 118°37′ E) has north subtropical humid climate. It receives an average rainfall of 1,106.5 mm, 76% average relative humidity, and 15.4°C average temperature. Nantong (31°58′ N, 120°53′ E) is located at the lower reaches of Yangtze River in the alluvial plain with mild marine climate, and possesses an average temperature and precipitation of 15.1°C and ~1,040 mm, respectively. Yangzhou (32°23′ N, 119°25′ E) is located in the southern end of Yangtze Huaihe plain, and receives an average precipitation and temperature of 1,020 mm and 14.8°C, respectively. In each environment, all the 211 diverse soybean genotypes were planted in a randomized complete block design with three replications. Each genotype was planted in a plot of three rows with row-length and spacing of 200 and 50 cm, respectively. Normal agronomic practices were followed for the cultivation of soybean germplasm at each location, as previously described by Zhang S. et al. (2021).

Phenotypic data collection and analysis

At maturity, five consecutive plants of each genotype were selected from the middle of each plot for data collection. Phenotypic data were recorded for four yield-related traits including seed yield per plant (SYP), number of pods per plant (PPP), number of seeds per plant (SPP), and 100-seed weight (HSW). The phenotypic data were subjected to analysis of variance with the mixed linear model using *lme4* – an r-package implemented in PBTools v1.4 (IRRI, 2014). BLUPs were generated for GWAS by setting the genotype as random. Pearson correlation coefficient between traits was determined at $p < 0.05$ and visualized using MVApp (Julkowska et al., 2019).

Genotyping, linkage disequilibrium, and genome-wide association study

NJAU 355 K Soy SNP Array previously developed and described by Wang et al. (2016) was used in this study. Quality control analysis was performed using PLINK v1.07 (Purcell et al., 2007) with the following criteria: missing genotype and individual at 0.1; minor allele frequency (MAF) at 0.01, and Hardy-Weinberg exact test at 0.000001. For the genome-wide LD analysis, pairwise squared allele-frequency correlations (r^2) between SNP markers with known genomic positions were calculated using Trait Analysis by Association, Evolution, and Linkage (TASSEL) v5.72 (Bradbury et al., 2007) with 100 sliding window sizes. The expected values of r^2 under drift equilibrium were calculated according to Hill and Weir (1988) and plotted against physical distance (Kbp). The LD decay curve line was fitted on the scatterplot using the smoothing spline regression line at the genome level following the procedure of Remington et al. (2001) in the *R* environment.

The GWAS was performed using the following five models:

1) General linear model (GLM) with principal component analysis (PCA) to reduce false positive association due to population structure (Price et al., 2006) based on the equation as follows:

$$y = S_i + Q + e$$

Where Si = testing marker, and Q = Population structure.

2) The compressed mixed linear model (CMLM) (Zhang et al., 2010), has increased statistical power for marker-trait association detection relative to other methods. CMLM method groups individuals into clusters, and random effects are fitted as genetic values of clusters in a mathematical model:

$$Y = S_i + Q + K + e$$

Where Si = testing marker, Q = Population structure, and K = Kinship by group.

3) The multiple-locus mixed linear model (MLMM) (Segura et al., 2012) incorporates the kinship matrix and Pseudo Quantitative Trait Nucleotide (QTN) to control false discovery rate (FDR) based on the model equation:

$$y = S_i + S + Q + K + e$$

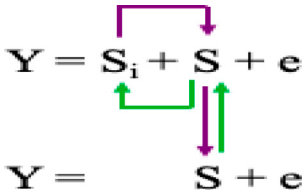
Where Si = testing marker, Q = Population structure, K = Kinship of individuals, and S = Pseudo QTN.

4) The fixed and random model circulating probability unification (FarmCPU) (Liu et al., 2016) iteratively uses the Fixed Effect Model (FEM) and a Random Effect Model (REM) as shown in the model equation:

$$y = S_i + S + e$$
$$y = K + e$$

Where K = kinship derived from only the associated markers or Pseudo QTN (S) using maximum likelihood method.

5) Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) (Huang et al., 2019) is an improved version of FarmCPU model and expressed as below:



S = pseudo QTNs that are not in LD with each other selected in two FEMs and one filtering process and optimization using Bayesian information criterion (BIC).

Among the five GWAS models used, GLM and CMLM are single locus models; whereas, MLMM, FarmCPU, and BLINK are the multi-locus models. All of the models were implemented using GAPIT v3 package (Lipka et al., 2012) in an *R* environment. To correct for the population structure, PCA was employed using BIC to estimate the optimal numbers of PCA (Schwarz, 1978). The population structure was visualized based on soybean ecological regions in China using the "*ggbiplot*" package in the *R* environment. Significant SNP associations were determined according to the inherently included method in GAPIT v3 (Lipka et al., 2012) as the negative logarithm of *p*-value (where *p*-value = 0.01/number of markers, and 0.01 represents MAF cut-off). Furthermore, only SNPs that were commonly detected across a minimum of two models were considered for further analysis.

Prediction of candidate genes

The genes lying upstream and downstream of each stable SNP (within the LD decay distance of the studied population) were obtained from the freely available online RNA-seq data for different soybean tissues on SoyBase website (https://www.soybase.org). Functional descriptions of these genes were also obtained, manually screened, and presented in a heatmap.

Allele-effect and haplotype analysis

Effects of alleles underlying the significant stable SNP markers were analyzed as previously described by Su et al. (2019) and Alemu et al. (2021). Genotypes were grouped into independent groups according to their specific SNP alleles, and means were compared using Turkey's HSD test.

Haplotype analysis was conducted using PLINK v1.07 (Purcell et al., 2007). The stable SNP markers (identified in two or more two environments) were considered reference markers for building haplotype blocks/loci. Besides, all markers that were in close association with the reference SNP markers within the estimated LD decay distance of the studied population formed a haplotype block/locus. Effects of haplotype alleles on the studied traits were tested across all the six environments using the conditional haplotype testing command (*--chap*). Also, the contribution of each haplotype to the observed phenotypic variance across the environments was estimated using the *"--hap-assoc*" command (Purcell et al., 2007) and visualized in Microsoft Excel.

TABLE 1 Combined analysis of variance and broad-sense heritability for four yield and yield-related traits.

Traits	PPP	HSW	SPP	SYP
Variance				
Genotype (G)	320.8***	24.8***	1,144.5***	11.9***
Environment (E)	386.1***	1.3***	861.8***	22.9***
G × E	73.8***	1.7***	329.4***	5.8***
Error	259.5	2.2	919.5	17.7
Broad-sense heritability	0.85	0.99	0.89	0.80

***represents the significance level at $p< 0.0001$.

Genomic prediction

The genomic prediction was explored for each trait in individual and combined environments using Genomic best linear unbiased prediction (gBLUP) and the ridge regression best linear unbiased prediction (rrBLUP) based on the mixed-model:

$$y = X\beta + Z\mu + \varepsilon$$

where β and μ represent the vectors of fixed and random effects, respectively, and ε is the residual error.

The gBLUP was implemented in TASSEL v5.72 (Bradbury et al., 2007). Cross-validation was achieved in five-folds with 20 iterations to test the genomic prediction accuracy and to avoid overfitting of the model. The rrBLUP on the other hand was implemented using the "rrBLUP" package (Endelman, 2011; Endelman and Jannink, 2012) in the R environment. To validate the genomic prediction accuracy, the dataset was randomly divided into training and testing sets at 80 and 20% respectively. To manage the challenges of overfitting, the cross-validation was conducted in five hundred cycles of iterations. The predictive ability was estimated as the Pearson's correlation coefficient between the observed and predicted phenotypic values of the test set based on the effect estimates of genotypes in the training set.

Results

Phenotypic analysis

Analysis of variance for the four yield-related traits evaluated across 211 soybean genotypes is summarized in Table 1. A highly

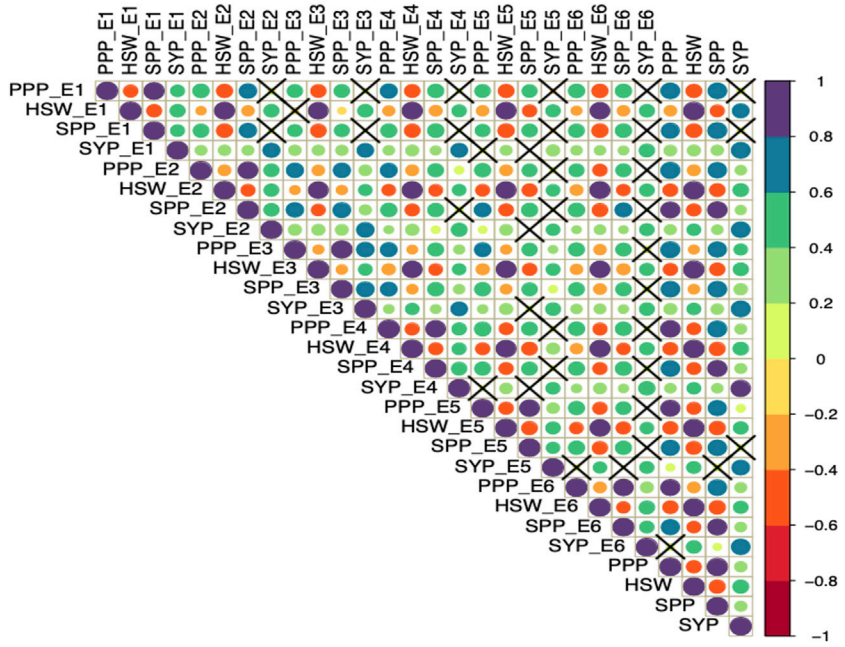


FIGURE 1
Pearson correlation analysis of yield and yield-related traits evaluated across diverse environments. The four traits including seed yield per plant (SYP), number of pods per plant (PPP), number of seeds per plant (SPP), and 100-seed weight in grams (HSW) were evaluated in six environments (E1, E2, E3, E4, E5, and E6) and the combined environment. The color and size of the circle reflect the strength of correlation. The non-significant correlations ($p > 0.05$) are indicated with a cross in individual cells.

significant variation ($p < 0.0001$) was observed for the genotype, environment, and G × E interaction in all the studied traits. However, the estimates of variance components varied across different traits (Table 1). For all the studied traits, the genotype component accounted for the highest proportion of the observed variations. Moreover, medium to high broad-sense heritability (h^2) was observed, ranging from 0.61 (SYP) to 0.99 (HSW) in individual environments (Supplementary Table S1) and from 0.80 (SYP) to 0.99 (HSW) in the combined environment (Table 1).

Furthermore, Pearson correlation analysis revealed that SYP has a positive significant correlation with PPP, SPP and HSW (Figure 1). Whereas, HSW was negatively correlated with PPP and SPP in all the studied environments. Also, PPP showed a positive correlation with SPP across all six environments.

Marker quality control, population structure, and linkage disequilibrium

The quality control analysis retained 12,617 SNPs across 211 soybean genotypes at a genotyping rate of 99% after removing SNPs that failed the missingness, minor allele frequency, and Hardy-Weinberg exact tests. The markers were distributed across the soybean genome, with the highest (1,112) and lowest (352) number of markers present on Chr.06 and Chr.05, respectively (Figure 2A). Heatmaps and dendrograms of the kinship matrix, based on 12,617 polymorphic SNPs for the studied genotypes, indicated that there was no clear clustering among the genotypes (Figure 2B). The population structure based on soybean ecological regions in China also revealed a continuous distribution without any distinct structure (Figure 2C).

The graphical representation of the linkage disequilibrium characteristics of the 211 soybean genotypes is presented in Figure 3. The average r^2 value of the genome was 0.12, and the LD decay was found to initiate at an r^2 value of 0.47 and reached half-decay at 0.24. The LD decay curve intersected with the half-decay at 670 Kbp, which represents the genome-wide critical distance to detect linkage. Hence, markers associated with the same trait within this distance were considered as a single QTL.

Marker-trait associations for yield-related traits

A total of 57 SNPs detected using five different GWAS models were found to possess significant association with four studied traits across six different environments (Supplementary Figure S1; Table 2). These SNPs were

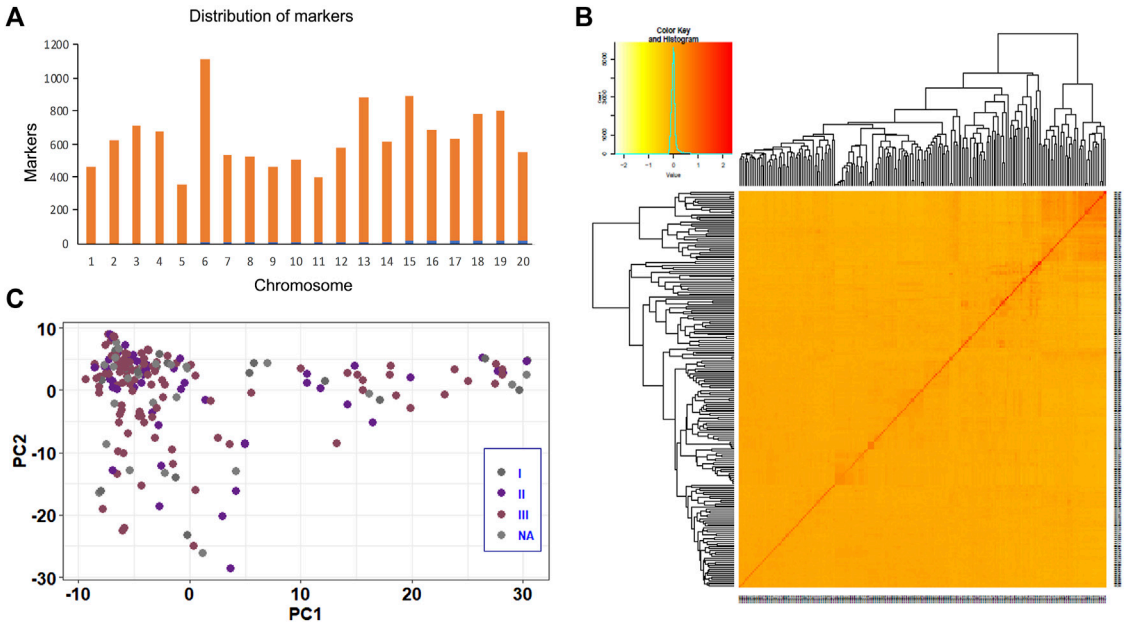


FIGURE 2
Kinship plot and population structure analysis of the soybean panel using 12,617 SNP markers. **(A)** Distribution of 12,617 SNP markers across 20 soybean chromosomes used for GWAS and GP analysis. **(B)** Kinship plot depicting the relationship among 211 soybean genotypes. **(C)** Population structure analysis of 211 soybean genotypes.

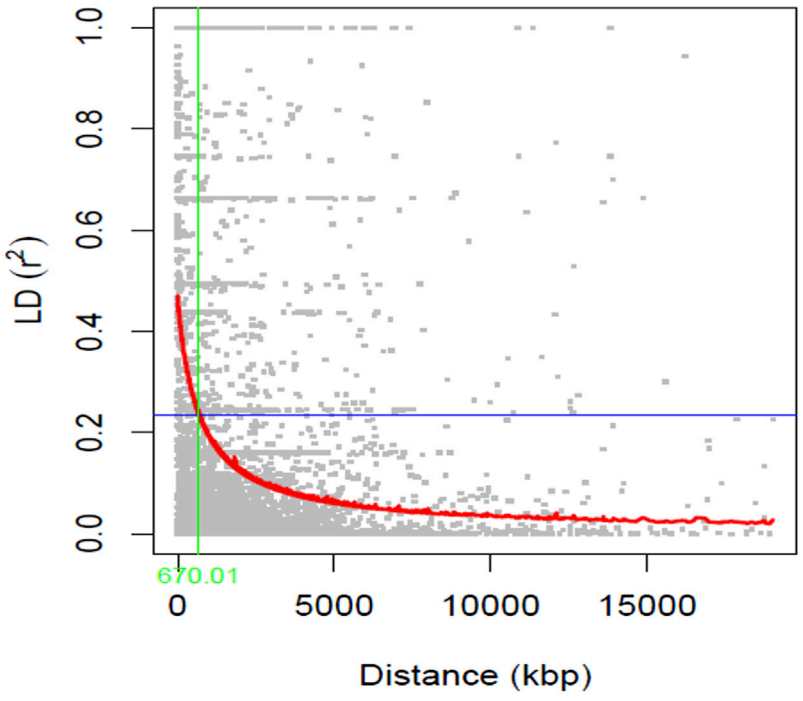


FIGURE 3
A scatter plot (r^2 values) of pairwise SNPs showing genome-wide linkage disequilibrium (LD) decay. The red curve line represents the smoothing spline regression model fitted to LD decay. The vertical green line indicates the genetic distance (670 Kbp) at which the LD half-decay (r^2 = 0.24, the horizontal blue line) intersect with the LD decay curve.

TABLE 2 Significant marker-trait associations identified for four yield and yield-related traits across six environments (E1, E2, E3, E4, E5, and E6) and a combined environment (COM) using five GWAS models.

Sr. No	SPP	Trait	Chr[a]	Pos. (bp)[b]	Environments	Model	*p*-value	MAF
1	AX-93668,616	SYP	1	11,556,172	E3 and COM	BLINK, FarmCPU, GLM	2.68E-06 to 6.87E-14	0.06
2	AX-93973,672	PPP	1	55,502,450	E1	BLINK, GLM	1.69E-08 to 7.65E-08	0.01
3	AX-93991,698	SYP	3	24,910,743	COM	BLINK, FarmCPU	3.23E-07 to 7.38E-07	0.04
4	AX-93703,924	PPP, SPP	4	4,291,705	E3, E6 and COM	BLINK, FarmCPU, GLM	1.33E-06 to 8.45E-07	0.04
5	AX-93707,240	HSW	4	1,611,547	E1 and E5	BLINK, FarmCPU, GLM	1.05E-12 to 9.43E-07	0.17
6	AX-94000,527	PPP	4	5,416,455	E2 and COM	FarmCPU, GLM	1.26E-06 to 4.17E-10	0.35
7	AX-94006,694	HSW	4	39,016,845	E3	BLINK, FarmCPU	5.37E-07 to 9.88E-07	0.45
8	AX-93922099	HSW	5	36,599,702	E1, E5 and COM	BLINK, FarmCPU	2.12E-09 to 8.33E-08	0.09
9	AX-93715,038	HSW	5	2,491,676	E1 and E4	FarmCPU, GLM	1.01E-07 to 6.23E-07	0.09
10	AX-93725,825	PPP	6	5,623,054	E1	BLINK, GLM	3.09E-07 to 3.63E-07	0.13
11	AX-93730,411	HSW	6	19,462,776	E2	BLINK, GLM	1.35E-09 to 2.78E-08	0.09
12	AX-93735,201	HSW	6	43,399,273	E4	BLINK, FarmCPU	1.07E-06 to 1.83E-06	0.14
13	AX-94033,285	SYP	6	48,786,905	E3	BLINK, FarmCPU	6.83E-09 to 9.40E-10	0.48
14	AX-94034,566	HSW, SYP	7	1,626,042	E3 and E3	BLINK, FarmCPU, GLM	1.24E-06 to 2.98E-07	0.32
15	AX-94050,700	SYP	8	15,859,520	COM	BLINK, FarmCPU	5.66E-10 to 9.74E-09	0.04
16	AX-94283,862	PPP	8	19,011,459	COM	BLINK, FarmCPU	1.27E-06 to 7.72E-08	0.04
17	AX-93755,601	PPP	8	19,248,053	E3	BLINK, FarmCPU	1.57E-10 to 5.25E-07	0.12
18	AX-93636,437	SYP	9	6,958,542	E2	BLINK, GLM	6.90E-07 to 9.77E-12	0.05
19	AX-93772,794	SYP	9	4,4,392,036	COM	BLINK, GLM	1.36E-06 to 1.78E-06	0.13
20	AX-93793,210	HSW, SPP	11	29,587,057	E1, E2, E3, E4, E5 and COM	BLINK, FarmCPU, GLM and MLMM	1.13E-08 to 7.01E-11	0.26
21	AX-93792,964	HSW, PPP	11	27,468,886	E5, E6	BLINK, FarmCPU, GLM	1.31E-07 to 5.42E-13	0.19
22	AX-94276,492	PPP, SPP	11	2,211,768	E3	BLINK, GLM	1.34E-08 to 8.75E-10	0.10
23	AX-93792,958	HSW	11	27,459,211	E2 and COM	BLINK, FarmCPU, GLM and MLMM	3.22E-06 to 8.67E-07	0.19
24	AX-94092,104	SYP	11	38,213,308	COM	BLINK, FarmCPU, GLM	1.98E-08 to 4.35E-07	0.09
25	AX-93797,890	PPP, SPP	12	5,802,813	E3 and E5	BLINK, FarmCPU	1.11E-06 to 9.02E-08	0.03
26	AX-93804,315	PPP	12	33,817,396	E1	FarmCPU, GLM	2.29E-08 to 3.72E-07	0.01
27	AX-93805,697	SYP	12	37,339,415	E4	BLINK, GLM	1.49E-07 to 3.29E-07	0.20

(Continued on following page)

TABLE 2 (*Continued*) Significant marker-trait associations identified for four yield and yield-related traits across six environments (E1, E2, E3, E4, E5, and E6) and a combined environment (COM) using five GWAS models.

Sr. No	SPP	Trait	Chr[a]	Pos. (bp)[b]	Environments	Model	*p*-value	MAF
28	AX-93807,406	HSW, SPP	13	1,843,185	E1, E2, E4, E5, E6 and COM	BLINK, FarmCPU, GLM	1.17E-10 to 8.25E-08	0.28
29	AX-94104,132	HSW, SPP	13	1,704,148	E3 and COM	BLINK, FarmCPU, GLM	1.74E-06 to 9.50E-07	0.23
30	AX-94111,538	HSW	13	29,191,712	E5 and COM	BLINK, FarmCPU	1.00E-07 to 4.87E-07	0.03
31	AX-93810,993	HSW	13	13,752,815	E5	BLINK, FarmCPU	1.73E-06 to 4.86E-08	0.05
32	AX-94123,137	PPP, SPP	14	23,593,462	E3	BLINK, GLM	1.31E-10 to 9.43E-09	0.02
33	AX-94137,762	SYP	15	22,710,572	E5	BLINK, FarmCPU, CMLM, GLM and MLMM	1.64E-09 to 2.48E-06	0.04
34	AX-93648,081	HSW	15	1,479,982	E6	BLINK, FarmCPU, GLM	1.84E-07 to 6.39E-07	0.34
35	AX-94138,593	SYP	15	29,628,987	COM	BLINK, FarmCPU, GLM	2.98E-07 to 8.36E-15	0.01
36	AX-93647,998	PPP	15	14,306,514	COM	BLINK, GLM	1.10E-10 to 1.24E-06	0.05
37	AX-94139,057	SPP	15	32,570,960	E6	BLINK, GLM	1.07E-06 to 1.31E-07	0.07
38	AX-94139,404	SYP	15	37,194,980	E3	FarmCPU and MLMM	1.18E-13 to 2.51E-06	0.03
39	AX-94139,741	SPP	15	40,016,648	COM	BLINK, GLM	1.66E-07 to 4.89E-10	0.14
40	AX-94139,803	PPP	15	40,162,413	E2	BLINK, GLM	1.44E-08 to 6.40E-08	0.06
41	AX-93843,622	HSW	15	44,240,130	E5	FarmCPU, GLM	1.03E-06 to 4.93E-08	0.13
42	AX-93843,767	SPP	15	44,966,712	COM	BLINK, GLM	3.31E-06 to 9.03E-11	0.14
43	AX-93650,734	HSW	16	30,750,889	E3	BLINK, FarmCPU, GLM	2.98E-06 to 9.61E-09	0.13
44	AX-93855,303	SPP	16	33,390,841	E5	BLINK, FarmCPU, GLM	2.13E-09 to 7.13E-07	0.02
45	AX-93652076	SYP	17	1,3,931,777	E6	BLINK, FarmCPU, GLM	1.67E-06 to 4.45E-07	0.01
46	AX-93959,968	SPP	17	12,429,289	E2	FarmCPU, GLM	1.52E-06 to 4.87E-07	0.02
47	AX-94176727	HSW, PPP	18	46,137,043	E1, E2 and COM	BLINK, FarmCPU, GLM	1.22E-07 to 5.79E-07	0.04
48	AX-93869,048	PPP	18	3,396,703	E3	BLINK, FarmCPU	2.32E-06 to 3.28E-07	0.04
49	AX-93886,740	PPP	19	3,901,635	E3	FarmCPU, GLM	1.57E-07 to 3.77E-06	0.03
50	AX-94199992	PPP, SPP	20	12,095,298	E1, E3 and COM	BLINK, FarmCPU, GLM	1.27E-08 to 6.40E-10	0.05
51	AX-93901,622	SPP	20	10,355,416	E2 and E3	BLINK, FarmCPU, GLM and MLMM	1.88E-06- to 4.95E-07	0.01
52	AX-93903,055	PPP	20	17,413,084	E1 and E2	BLINK, FarmCPU, GLM	1.10E-07 to 8.65E-09	0.01
53	AX-94198582	PPP	20	4,715,203	E3	BLINK, GLM	2.07E-09 to 2.67E-13	0.08
54	AX-94201014	PPP	20	17,645,281	E3	GLM, MLMM	1.50E-06 to 8.73E-11	0.02

(Continued on following page)

TABLE 2 (*Continued*) Significant marker-trait associations identified for four yield and yield-related traits across six environments (E1, E2, E3, E4, E5, and E6) and a combined environment (COM) using five GWAS models.

Sr. No	SPP	Trait	Chr[a]	Pos. (bp)[b]	Environments	Model	*p*-value	MAF
55	AX-93903,184	PPP	20	18,155,072	E3	BLINK, GLM	4.38E-11 to 7.33E-09	0.02
56	AX-94207999	HSW	20	42,330,677	E2	BLINK, GLM	6.29E-09 to 7.83E-07	0.30
57	AX-94292,257	PPP	20	45,876,916	E5	BLINK, FarmCPU	1.99E-07 to 2.36E-07	0.01

[a]Chromosome.
[b]Physical position; MAF (minor allele frequency).

TABLE 3 Stable QTLs/genomic regions identified for the studied traits in at least three or more environments.

QTL	Chr[a]	Rep.SPP[b]	Pos. (bp)[c]	Environments	Model	Related QTL	References
qPPP-SPP4	4	AX-93703,924	4,291,705	E3, E6 and COM	*BLINK, FarmCPU, GLM*	Novel QTL	Not available
qHSW5	5	AX-93922099	36,599,702	E1, E5 and COM	*BLINK, FarmCPU*	*Seed weight 34–9; Seed-yield 22–10*	Han et al., 2012; Du et al., 2009
qHSW-SPP11	11	AX-93793,210	29,587,057	E1, E2, E3, E4, E5 and COM	*BLINK, FarmCPU, GLM and MLMM*	*Seed weight 35–9*	Han et al. (2012)
qHSW-SPP13	13	AX-93807,406	1,843,185	E1, E2, E4, E5, E6 and COM	BLINK, FarmCPU, GLM	Novel QTL	Not available
qHSW-PPP18	18	AX-94176727	46,137,043	E1, E2 and COM	BLINK, FarmCPU, GLM	Novel QTL	Not available
qPPP-SPP20	20	AX-94199992	12,095,298	E1, E3 and COM	BLINK, FarmCPU, GLM	Novel QTL	Not available

[a]Chromosome.
[b]The representative SPP, with the min *p* value.
[c]Physical position.
The italic values indicate "QTL names" and "Gene IDs".

distributed across 17 soybean chromosomes. Among the significant SNPs identified, the highest number of SNPs (10) are found on Chr.15, followed by eight and five SNPs on Chr.20 and Chr.11, respectively. Four significant SNPs were found each on Chr.04, Chr.06, and Chr.13; whereas, Chr.08 and Chr.12 possessed three significant SNPs each. The remaining ten chromosomes possessed one or two significant SNPs. Further, some of these SNPs were consistently detected in multiple environments, using different GWAS models, and were found to be associated with more than one studied trait. Such SNPs were considered stable MTAs. For example, the significant SNP (AX-93793,210) on Chr.11 was identified consistently in five individual environments (E1, E2, E3, E4, and E5) and the combined environment. In addition, this SNP was also identified through four different GWAS models (BLINK, FarmCPU, GLM, and MLMM), and was found to be associated with two yield-related traits (HSW and SPP). Similarly, another SNP (AX-93807,406) detected on Chr.13 was found to be significantly associated with HSW and SPP across five individual environments (E1, E2, E4, E5, and E6) and the combined environment. This SNP was also identified using three GWAS models (BLINK, FarmCPU, and GLM).

Furthermore, three significant SNPs (AX-94199992, AX-93703,924, and AX-94176727) on Chr.20, Chr.04, and Chr.18, respectively, were consistently detected in the combined environment using three different models (BLINK, FarmCPU, and GLM) and showed association with two of the three traits viz., HSW, PPP, and SPP. The SNP (AX-93922099) was detected in the combined environment using two different models (BLINK and FarmCPU) and was associated with HSW. Few significant SNPs such as AX-93792,964, AX-94034,566, AX-93797,890, and AX-94104,132, present on Chr.11, Chr.07, Chr.12, and Chr.13, respectively, were found to be associated with two of the four studied traits, using up to three GWAS models in one or two individual environments and a combined environment. Moreover, eight SNPs (AX-93668,616, AX-93707,240, AX-

TABLE 4 Putative genes underlying six QTLs and their gene annotation.

QTL name	Gene IDs	Chrom-osome	Gene functional annotation
qPPP-SPP4	*Glyma04g05500*	04	Protein folding; abiotic stress response; positive regulation of transcription
	Glyma04g05520	04	NA
	Glyma04g05580	04	Gluconeogenesis; glycolysis; translational initiation; abiotic stress response
	Glyma04g05690	04	Lipid biosynthetic process
	Glyma04g05720	04	Protein folding; abiotic stress response
	Glyma04g05800	04	Photosynthesis
qHSW5	*Glyma05g31250*	05	Acetyl-CoA metabolic process; abiotic stress response; polysaccharide transport; sterol biosynthetic process; brassinosteroid biosynthetic process
	Glyma05g31260	05	Nuclear division; cytokinesis by cell plate formation; chromatin silencing; nucleolus organization; biological process; cell proliferation; histone phosphorylation; histone H3-K9 methylation
qHSW-SPP11	*Glyma11g28990*	11	NA
	Glyma11g29000	11	Protein N-linked glycosylation; ethylene biosynthetic process; sugar mediated signaling pathway; stem cell division; proteasomal ubiquitin-dependent protein catabolic process; cotyledon development; regulation of post-embryonic root development
qHSW-SPP13	*Glyma13g01950*	13	Carbohydrate metabolic process; regulation of meristem growth
qHSW-PPP18	*Glyma18g38490*	18	Regulation of transcription; gibberellin biosynthetic process; response to auxin stimulus; response to abscisic acid stimulus; gibberellic acid mediated signaling pathway; embryo development; terpenoid biosynthetic process; cotyledon development; cell division
	Glyma18g38570	18	Cell morphogenesis; protein N-linked glycosylation; N-terminal protein myristoylation; cell growth; protein ubiquitination; regulation of protein localization; protein autophosphorylation; Golgi vesicle transport; root hair elongation
	Glyma18g38610	18	Regulation of transport
qPPP-SPP20	*Glyma20g08580*	20	Actin filament organization; regulation of stomatal movement; regulation of protein localization

The italic values indicate "QTL names" and "Gene IDs".

93792,958, AX-94000,527, AX-93715,038, AX-94111,538, AX-93901,622, and AX-93903,055) were detected in one or two individual environment(s) and a combined environment. Each of these eight SNPs was associated with only one trait and identified using up to four different GWAS models. The SNPs (AX-94276,492 and AX-94123,137) were found to be associated with two traits (PPP and SPP) and were identified using two GWAS models (BLINK and GLM), but were identified in only one individual environment. The remaining 37 significant SNPs were found to be associated with only one trait and environment, and most of them were identified using two or three GWAS models.

Based on GWAS, we identified six significant SNPs (AX-93703,924, AX-93922099, AX-93793,210, AX-93807,406, AX-94176727, and AX-94199992) on Chr.4, Chr.5, Chr.11, Chr.13, Chr.18, and Chr.20, respectively, consistently in three or more than three environments and using multiple models (Tables 2, 3). Of these, five SNPs (AX-93807,406, AX-93793,210, AX-94199992, AX-93703,924, and AX-94176727) were found to be associated with two studied traits (Tables 2, 3). However, the SNP AX-93922099 was associated with only HSW (Tables 2, 3). Hence, based on the LD decay, the genomic regions (±670 kb) flanking these significant SNPs (AX-93807,406, AX-93793,210, AX-94199992, AX-93922099, AX-93703,924, and AX-94176727) were considered as QTLs viz., *qHSW-SPP13*, *qHSW-SPP11*, *qPPP-SPP20*, *qHSW5*, *qPPP-SPP4*, and *qHSW-PPP18*, respectively (Table 3). These QTLs/genomic regions represented stable genetic elements potentially regulating soybean yield-related traits.

RNA-seq data revealed 15 putative genes regulating yield-related traits

Six QTLs/genomic regions were identified on Chr.04 (*qPPP-SPP4*), Chr.05 (*qHSW5*), Chr.11 (*qHSW-SPP11*), Chr.13 (*qHSW-SPP13*), Chr.18 (*qHSW-PPP18*), and Chr.20 (*qPPP-SPP20*) were further used to identify putative genes regulating yield-related traits in soybean. The gene models and their annotations underlying the physical intervals of these six QTL regions were downloaded from the SoyBase database to predict putative candidates (Supplementary Table S2). In total, 739 gene models were identified within the physical genomic interval of *qPPP-SPP4*, *qHSW5*, *qHSW-SPP11*, *qHSW-SPP13*, *qHSW-PPP18* and *qPPP-SPP2*. However, by considering gene annotation, we selected 113 gene models within their physical genomic interval, which consisted of 31, 33, 10, 11, 14, and 14 genes underlying *qPPP-SPP4*, *qHSW5*, *qHSW-SPP11*, *qHSW-*

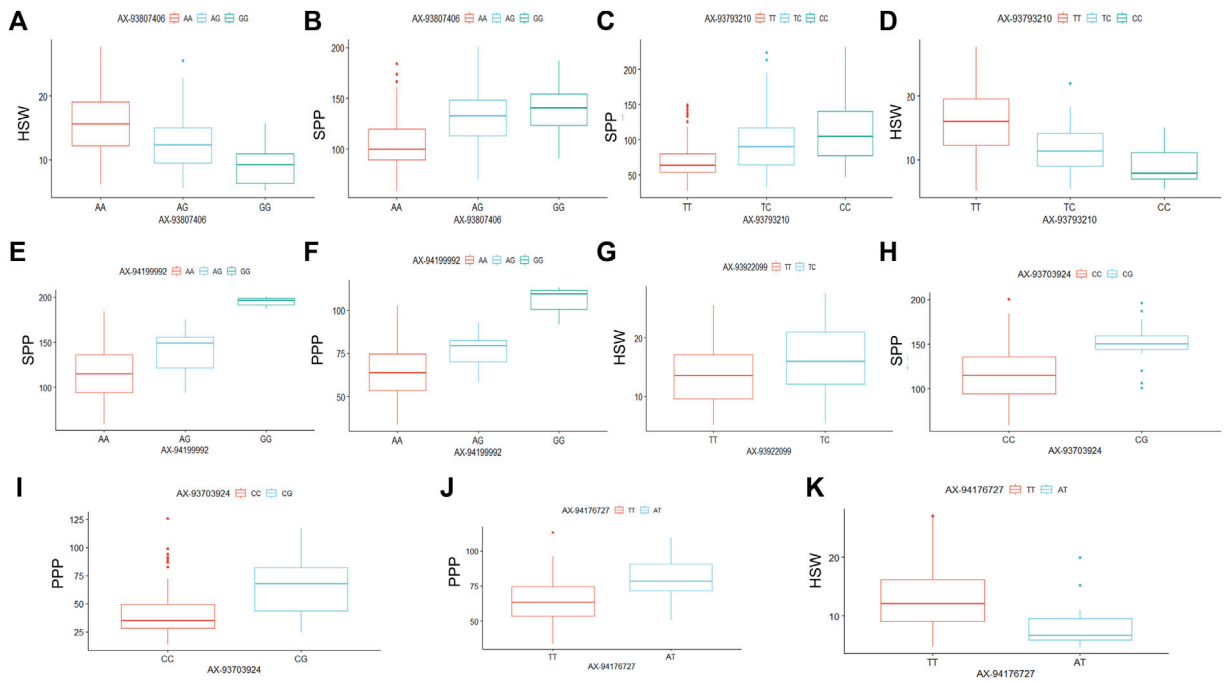


FIGURE 4
Allele-effect analysis for six stable significant SNPs including AX-93807,406 **(A,B)**, AX-93793,210 **(C,D)**, AX-94199992 **(E,F)**, AX-93922099 **(G)**, AX-93703,924 **(H,I)**, and AX-94176727 **(G,K)**. The box plot depicts the number of the alleles for each of the six significant SNPs in 211 soybean genotypes, and the contribution of these alleles to the phenotypic variation observed for yield and yield-related traits.

SPP13, *qHSW-PPP18* and *qPPP-SPP20*, respectively (Supplementary Table S2). In addition, RNA-seq data for samples collected across different stages of soybean growth and development (www.soybase.org) was downloaded and analyzed for identifying putative genes underlying the QTL intervals (Supplementary Table S3). The RNA-seq data of these genes are represented using a heatmap (Supplementary Figure S2). Based on the *in-silico* analysis of gene expression data and gene annotations, we predicted a total of 15 putative candidates underlying six QTLs. These include 6, 2, 2, 1, 3, and 1 gene underlying *qPPP-SPP4*, *qHSW5*, *qHSW-SPP11*, *qHSW-SPP13*, *qHSW-PPP18*, and *qPPP-SPP20*, respectively (Table 4).

Allelic effects of stable marker-trait associations

The six significant SNPs (AX-93703,924, AX-93922099, AX-93793,210, AX-93807,406, AX-94176727, and AX-94199992), showing stable MTAs with yield-related traits were further used to determine the effects of their individual alleles on the studied traits (Figure 4). The alleles of these six SNP markers showed substantial effects on yield-related traits combined from all the six environments (Figure 4). However, the number of alleles for each of these six SNP markers in the whole soybean population varied from two to three. For example, the SNP markers AX-93807,406, AX-93793,210, and AX-94199992 each possessed three different alleles; whereas, AX-93922099, AX-93703,924, and AX-94176727 possessed two alleles each (Figure 4). The SNP marker AX-93807,406 possessed three alleles (AX-93807406-AA, AX-93807406-AG, and AX-93807406-GG), and were found to regulate HSW and SPP (Figures 4A,B). The AX-93807406-AA, AX-93807406-AG, and AX-93807406-GG alleles governed higher, intermediate, and lower HSW, respectively; whereas these same alleles regulated lower, intermediate, and higher SPP, respectively. The SNP marker AX-93793,210 is associated with two yield-related traits (HSW and SPP), and all three alleles of this marker (AX-93793210-TT, AX-93793210-TC, and AX-93793210-CC) showed significantly different allelic effects on both HSW and SPP (Figures 4C,D). The allele AX-93793210-TT was associated with higher HSW, whereas AX-93793210-CC was associated with lower HSW. The effect of AX-93793210-TC on HSW was intermediate between that of AX-93793210-TT and AX-93793210-CC. However, the effect of three alleles of AX-93793,210 on SPP was found to be opposite to that of HSW, which also supports the negative correlation between HSW and SPP.

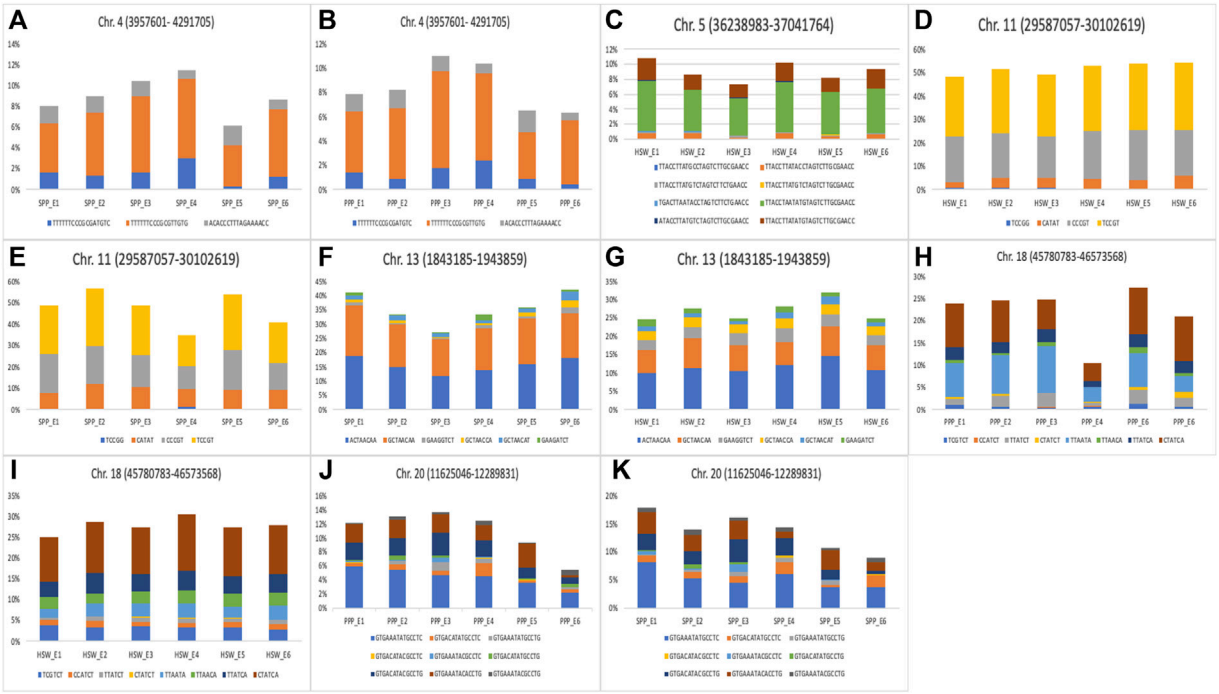


FIGURE 5
Haplotype analysis for yield and yield-related traits in soybean. **(A–K)** The bar plot represents haplotype alleles identified for haplotype block on **(A)** Chr.04 (3,957,601–4291,705 bp) **(B)** Chr.04 (3,957,601–4291,705 bp) **(C)** Chr.05 (36,238,983–3,7,041,764 bp) **(D)** Chr.11 (29,587,057–30102,619 bp) **(E)** Chr.11 (29,587,057–30102,619 bp) **(F)** Chr.13 (1,843,185–1943,859 bp) **(G)** Chr.13 (1,843,185–1943,859 bp) **(H)** Chr.18 (45,780,783–46,573,568 bp) **(I)** Chr.18 (45,780,783–46,573,568 bp) **(J)** Chr.20 (11,625,046–12,289,831 bp), and **(K)** Chr.20 (11,625,046–12,289,831 bp), and their contribution to the regulation of yield and yield-related traits. Six haplotype blocks were identified by considering six stable significant SNPs (AX-93807,406, AX-93793,210, AX-94199992, AX-93922099, AX-93703,924 & AX-94176727) as reference markers.

Moreover, three alleles of SNP marker AX-94199992, including AX-94199992-AA, AX-94199992-AG, and AX-94199992-GG, were found to govern three diverse phenotypes of SPP and PPP (Figures 4E,F). Besides, these three alleles of AX-94199992 regulated SPP and PPP phenotype in the same order as lower, intermediate and higher, respectively, which further supports the positive correlation of SPP and PPP traits. The two alleles of marker AX-93922099, including AX-93922099-TT and AX-93922099-TC, showed a significant difference in the regulation of HSW. For instance, the allele AX-93922099-TT governed lower HSW; whereas, the allele AX-93922099-TC was associated with higher HSW (Figure 4G). Similarly, the AX-93703,924 marker governed SPP and PPP traits, and the two alleles of this marker (AX-93703924-CC and AX-93703924-CG) regulated lower and higher trait values, respectively, for both the traits (Figures 4H,I). The SNP marker AX-94176727, possessing two alleles (AX-94176727-TT and AX-94176727-AT), regulated contrasting phenotypes of HSW and PPP (Figures 4J,K). For instance, AX-94176727-TT and AX-94176727-AT regulated lower and higher HSW, respectively; whereas, the same alleles governed higher and lower PPP values, respectively. These results are per the negative correlation observed between HSW and PPP.

Haplotypes for yield-related traits

The six stable markers mentioned above were used as a reference for the identification of haplotypes for yield-related traits. These stable markers were located on Chr.04 (AX-93703,924), Chr.05 (AX-93922099), Chr.11 (AX-93793,210), Chr.13 (AX-93807,406), Chr.18 (AX-94176727), and Chr.20 (AX-94199992). All the markers that were in strong LD (within ±670 kbp) with these six SNP markers, represented a haplotype block/locus (Figure 5; Table S4). For example, 17 SNP markers were in strong LD with the reference marker AX-93703,924 (3,957,601–4291,705) and formed a haplotype block. Three haplotype alleles were identified within this haplotype block, in the soybean population (Figure 5A). These three haplotype alleles identified on Chr.04 showed significant

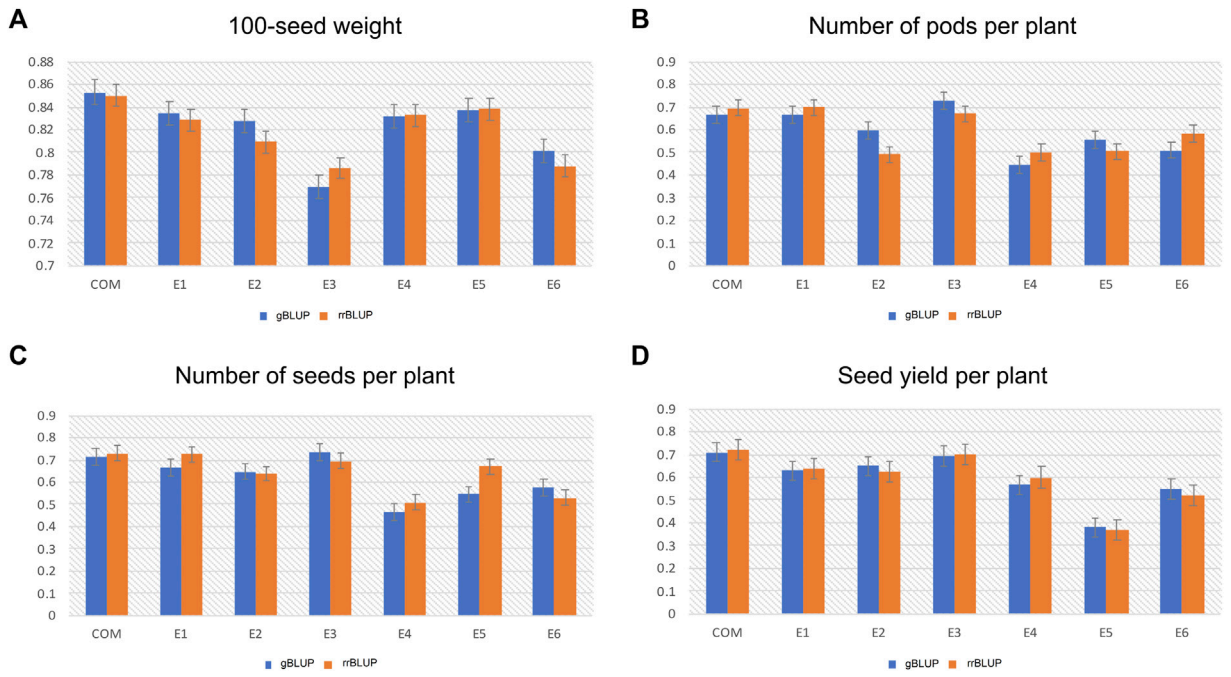


FIGURE 6
Histograms showing the genomic prediction accuracy of the genomic best linear unbiased prediction (gBLUP) and ridge regression best linear unbiased prediction (rrBLUP) models for **(A)** 100-seed weight (HSW) **(B)** Number of pods per plant (PPP) **(C)** Number of seeds per plant (SPP), and **(D)** Seed yield per plant (SYP), across six different environments (E1, E2, E3, E4, E5, and E6).

differences in the phenotypes of SPP and PPP. Further, the reference marker AX-93922099 (36,238,983–3,7,041,764) formed a haplotype block with 26 markers, which consisted of eight haplotype alleles (Figure 5B). Substantial phenotypic variance in HSW was observed for haplotype alleles present within this haplotype block on Chr.05 (Figure 5C). The marker AX-93793,210 (29,587,057–30102,619) constituted a haplotype block with five SNP markers constituting four haplotype alleles. Variation in these four alleles led to significant phenotypic variation in HSW and SPP traits (Figures 5D,E).

Eight SNP markers were associated with AX-93807,406 (1,843,185–1943,859), which represented a haplotype block and constituted six haplotype alleles. All the six alleles identified within this haplotype block showed significant differences in the phenotypes of HSW and SPP (Figures 5F,G). Similarly, six SNPs were in association with the reference marker AX-94176727 (45,780,783–46,573,568), which together formed a haplotype block on Chr.18. Eight alleles identified within this haplotype block showed substantial variation in the phenotypes of HSW and PPP (Figures 5H,I). Further, 13 SNP markers were in close association with AX-94199992 (11,625,046–12,289,831) and formed a haplotype block representing nine haplotype alleles (Figures 5J,K). The haplotype alleles of the AX-94199992 (11,625,046–12,289,831) block showed significant phenotypic variation in PNP and SNP (Figures 5H,I). Although haplotype alleles for most haplotype blocks showed significant differences in the phenotypes of different yield-related traits in six individual environments, a few exceptions were also observed. For example, the haplotype alleles of AX-93922099 (36,238,983–3,7,041,764) block on Chr.05 did not show significant phenotypic variation in HSW for E3 (NT1) and E6 (YZ2). Moreover, the haplotype alleles of all other haplotype blocks showed a significant phenotypic difference ($p < 0.05$) in their associated traits across all six environments. The phenotypic variance contributed by the alleles of different haplotype blocks to the associated traits across six environments is shown in Figure 5. The list of markers that are in close association with the reference markers and the effects of the common haplotypes are provided in Supplementary Table S4.

Genomic prediction

The genome-wide prediction accuracy values obtained from the gBLUP and rrBLUP approaches for the studied yield-related traits are presented in Figure 6. Based on the gBLUP approach, the GP accuracy of HSW among different environments ranged between 0.76 and 0.85 (Figure 6A). The E3 environment showed the lowest GP accuracy (0.76), while the combined environment displayed the highest GP accuracy (0.85) for HSW (Figure 6A). For the PPP trait, the E4 environment had the lowest GP accuracy with 0.44, while the highest accuracy was recorded in the E3 environment (0.72)

(Figure 6B). Moreover, the GP accuracy for SPP varied from 0.46 to 0.72 in E4 and E3 environments, respectively (Figure 6C). Similarly, in the case of SYP, the GP accuracy was found to be highest (0.70) in the combined environment, whereas the lowest GP accuracy of 0.37 was observed in the E5 environment (Figure 6D). A similar trend in the genome-wide prediction accuracy was observed using the rrBLUP approach: HSW ranged between 0.78 and 0.85 for E3 and combined environment, respectively (Figure 6A); and PPP varied from 0.49 to 0.69. The PPP trait possessed the highest GP accuracy in the E1 environment and the lowest accuracy in the E2 environment (Figure 6B). Also, the GP accuracy for SPP and SYP respectively ranged from 0.50 (E4) to 0.73 (combined environment) and 0.36 (E5) to 0.72 (combined environment) (Figures 6C,D).

Discussion

Yield-related traits are important characters associated with yield and directly govern the productivity and quality of crops (Hina et al., 2020). They also represent selection targets in plant breeding programs when direct selection for yield is complex. Hence, crop germplasm collections are characterized for yield-related traits to facilitate crop improvement (Adeboye K. A. et al., 2021). In soybean, a complex inheritance pattern of yield and its sensitivity to the environment have been documented (Lee et al., 2015). Therefore, it has remained the long-term goal of breeders to improve soybean yield by manipulating yield-related traits. Determining the genetic basis of yield-related traits is a key step in soybean improvement strategy for developing varieties with higher yield potential.

In the present study, 211 soybean germplasm accessions were characterized in six field trials for four yield-related traits. The genotypic performance of the soybean germplasm based on all these traits varied significantly, indicating the possibility of genetic improvement. Moreover, the medium to high broad-sense heritability observed for these traits across environments is an indication that the same phenotypic performance is achievable if grown in the same environment. However, the significant environmental variations observed for all the four yield-related traits studied suggest possible complexity in their inheritance pattern which may lead to difficulty in breeding efforts. These findings corroborate many earlier reports on these yield-related traits in soybean germplasm (Hu et al., 2014; Diers et al., 2018; Klein et al., 2020; Li et al., 2020). Moreover, correlation analysis revealed a positive association of SYP with the three yield-related traits (PPP, SPP, and HSW) consistent with the findings of Malik et al. (2006). The HSW was negatively correlated with PPP and SPP; while PPP was positively associated with SPP. These results are in accordance with previous reports (Malik et al., 2006; Liu et al., 2011; Li et al., 2020).

Based on the high genetic variability observed, the studied yield-related traits were subjected to further analysis to unravel their genetic basis and pave the way for their improvement through marker-assisted breeding (MAB). MAB involves the identification of genetic markers that are associated with the trait of interest in a defined germplasm collection, such as bi-parental population or a panel of diverse accessions as used in this study. Marker-trait association using a panel of diverse accessions is often confounded by several factors, including genotyping error, population structure, and linkage disequilibrium. These factors are responsible for the high rate of false associations that are not useful in MAB. In this study, the discovery of false association was reduced by ensuring only quality markers which included a total of 12.617 SNPs at genotyping rate of 99%, which were retained for further analysis. Although the population structure of the germplasm collection used in this study appears continuous with no definite stratification, the LD decay distance of 670 kb within which significant association may be defined as relatively large. In self-pollinating plants such as soybean, LD may range over several hundred Kbp leading to the inclusion of many candidate genes in a single LD block exhibiting a significant signal (Gupta et al., 2005; Yano et al., 2016). Moreover, several studies have revealed that the discovery of false positives arising from population structure in crops may not be completely controlled (Myles et al., 2009; Hamblin et al., 2011; Lipka et al., 2015). Based on this, we explored various statistically robust models for the genome-wide marker-trait association analysis.

Hundreds of QTLs for yield and yield-related traits have been reported in soybean mainly by using the low-resolution approach of linkage mapping. Rarely any of these QTLs have been used for breeding high-yielding soybean varieties (Karikari et al., 2019). In this context, GWAS facilitates enhanced resolution and accuracy for mining genetic loci for four yield-related traits (Assefa et al., 2019; Zhang W. et al., 2021). In the current study, we report a total of 57 MTAs associated with four traits. These MTAs were detected on all chromosomes of soybean, except Chr.02 and Chr.10, indicating the complex genetic regulation of soybean yield, which is in agreement with previous reports (Li et al., 2019; Hu et al., 2020). Many significant MTAs were detected in one environment and some in five environments, suggesting the presence of environmental influence on these traits. Per the present results, the interaction of QTLs with the environment has been previously documented (Fang et al., 2020; Hu et al., 2020).

Significant SNPs reported in more than three environments and using different models were considered stable MTAs. The regions within ±670 kb flanking six significant SNPs were referred to as stable QTLs, based on LD decay (Wang et al., 2016). The QTL on Chr.11 associated with HSW has been previously reported in the genomic region between 27,790,963–32,194,459 bp (Han et al., 2012), and the genomic region underlying *qHSW-SPP11* was found to co-locate with the same physical interval. Therefore, *qHSW-SPP11* might be similar to *Seed weight 35–9*, as reported earlier by Han et al. (2012). Notably, compared to *Seed weight 35–9*, the physical interval of

qHSW-SPP11 has been considerably decreased in the present study. Furthermore, *qHSW5* identified in the present study was found to co-locate with two previously identified QTLs viz., Seed weight 34–9 (8,665,543–40,414,305 bp) and Seed-yield 22–10 (35,536,817–37,612,231 bp) on Chr.05 (Kraakman et al., 2004; Du et al., 2009; Han et al., 2012). However, the remaining four QTLs (*qHSW-SPP13*, *qPPP-SNP20*, *qPPP-SPP4*, and *qHSW-PPP18*) identified in the current study represented novel QTLs. The physical intervals of *qHSW-SPP11* and *qHSW5* were considerably reduced in the present study. This ability of the GWAS allows for its increased utility in crop breeding programs for developing high-yielding stress-tolerant soybean varieties (Zargar et al., 2015; Yu et al., 2019).

The favourable and unfavourable alleles can be easily determined either with or without considering the heterozygous SNPs in plant species (Soltani et al., 2017; Su et al., 2019). For instance, Wu et al. (2016) set the heterozygous SNPs as missing and only used the homozygous SNPs for a GWA study in *Brassica napus*. Soybean is a highly heterozygous crop species with a complex background, in which the presence of heterozygous loci is very common. A locus is considered to be in a heterozygous state if the depth of the minor allele is larger than one-third of the total sample depth during SNP calling (Chong et al., 2016). The above-mentioned six stable significant SNP loci (AX-93807,406, AX-93793,210, AX-94199992, AX-93922099, AX-93703,924, and AX-94176727) associated with yield-related traits were not all in a heterozygous state in the GWAS population. For example, the SNP markers AX-93807,406, AX-93793,210, and AX-94199992 were heterozygous, whereas the remaining three markers AX-93922099, AX-93703,924, and AX-94176727 were homozygous.

Trait values governed by the heterozygous alleles of three SNP markers (AX-93807,406, AX-93793,210, and AX-94199992) were intermediate between two homozygous alleles which control the extreme phenotypes of yield-related traits. However, two homozygous alleles identified for the remaining three SNP markers (AX-93922099, AX-93703,924, and AX-94176727) regulate contrasting/extreme trait values of the corresponding traits. As a result, the SNP alleles with higher trait effect, i.e., which increase the target trait, were defined as “favourable alleles”; whereas, SNP alleles regulating the lowest trait value were defined as “unfavourable alleles”. However, the heterozygous alleles possessing trait effect between favourable and unfavourable alleles were referred to as “intermediate alleles”. This classification assisted in the use of these alleles for yield modulation in soybean. To date, the effectiveness of marker-based gene pyramiding strategies in soybean has been demonstrated for soybean mosaic virus (Wang D. G. et al., 2017), Phytophthora rot and powdery mildew resistance (Ramalingam et al., 2020), and rust resistance (Yamanaka and Hossain, 2019). Hence, the elite alleles identified for yield-related traits within six significant SNP markers can be effectively used for developing high-yielding soybean varieties through MAB efforts. Negligible efforts have been made toward mining candidate genes for yield-related traits in soybean (Karikari et al., 2019; Li et al., 2019; Qi et al., 2020); except for two genes that have been reported viz., in (Jeong et al., 2012) and *PP2C-1* (Lu et al., 2011). In the present study, we predicted some putative genes underlying six QTLs identified based on the gene expression data and annotations. We selected only those genes whose functions were directly or indirectly related to regulating the seed yield of soybeans, such as seed oil, seed protein, photosynthesis, cell division or elongation, and phytohormones. The putative genes identified in the present study need further functional validation for their deployment in soybean breeding programs.

Recent studies have documented desirable haplotype alleles for important traits such as salinity tolerance in soybean (Patil et al., 2016), grain quality traits in rice (Wang X. et al., 2017), and drought stress tolerance in pigeon pea (Sinha et al., 2020). In the present study, haplotypes were constructed by using six significant SNP markers as a reference. These six stable markers possessed multiple significant SNP markers within the LD range of 670 kbp. Our results revealed that haplotype alleles identified within the haplotype blocks/loci regulated a diverse range of phenotypes of yield-related traits in soybean. The haplotype analysis revealed that, compared to individual SNP markers, the haplotype-based markers possessed a considerably higher number of alleles regulating a diverse range of phenotypic variation for the trait of interest, similar to previous studies (Zaitlen et al., 2005; Qian et al., 2017). Hence, haplotype-based markers provided more options to modulate the desired yield potential in soybean (Meuwissen et al., 2001). In the case of significant SNP markers identified, we found a maximum of three alleles in the GWAS population, which allowed to modify the yield of soybean at three levels. The incorporation of these haplotype alleles in soybean breeding programs can effectively improve yield potential in soybean. We propose that the haplotype-based breeding approach will assist in the selection of desirable plant genotypes possessing superior haplotype alleles (Varshney et al., 2019). Parent accessions with diverse haplotypes can be used for the generation of novel superior haplotypes. However, it is important to identify the interactive effects of diverse haplotypes of various genes regulating the trait of interest.

Genomic Prediction (GP) is a modern breeding approach that involves the use of genome-wide markers to estimate the breeding value of the genotypes at the genomic level (Meuwissen et al., 2001; Varshney et al., 2014). For the past 2 decades, GP has emerged as a powerful tool to select favourable genetic material for traits of interest (Bhat et al., 2016; Crossa et al., 2017). In the present study, GWAS identified minor-to-moderate effect QTLs. Thus, it was hypothesized that the GP method could be more appropriate to select high-yielding genotypes based on the overall marker effect (Varshney et al., 2014). Different statistical models have been developed and used for GP analysis (Daetwyler et al., 2010;

Wang et al., 2018; Merrick and Carter, 2021). However, these methods mainly differ in the assumptions of the distribution and variances of marker effects (Wang et al., 2015). In the present study, we explored two approaches, including the gBLUP and rrBLUP, both of which are based on the mixed linear model statistical functions. Therefore, our results show a similar trend in their prediction accuracy as expected, suggesting their equal potential and efficiency in the prediction of yield-related traits in soybean.

The range of moderate to high GP accuracy of 100-seed weight and seed yield observed in our study is similar to the observation of Ravelombola et al. (2021) and Matei et al. (2018) in soybean based on rrBLUP approach. Similarly, Duhnen et al. (2017) reported a moderate GP accuracy for the seed yield of soybean based on the gBLUP approach. Moreover, moderate to high GP accuracy has been reported for yield-related traits from both rrBLUP and gBLUP approaches in other crops such as wheat (Ali et al., 2020), tea (Lubanga et al., 2021), rice (Xu et al., 2018), and chickpea (Roorkiwal et al., 2016). The genomic prediction results from our study revealed that accurate breeding values for the studied yield-related traits can be estimated at an earlier generation of soybean, which allows for yield improvement within a short breeding cycle.

Conclusion

The present study used GWAS, haplotype analysis, and GP for studying the genetic architecture of four yield and yield-related traits in soybean. GWAS identified a total of 57 significant SNPs and six stable QTL regions (*qPPP-SPP4*, *qHSW5*, *qHSW-SPP11*, *qHSW-SPP13*, *qHSW-PPP18*, and *qPPP-SPP20*). Among these six QTLs, four QTLs (*qPPP-SPP4*, *qHSW-SPP13*, *qHSW-PPP18*, and *qPPP-SPP20*) were novel; whereas, the remaining two QTLs (*qHSW5* and *qHSW-SPP11*) were reported previously. Besides, a total of 15 genes underlying these six QTLs were prioritized as putative candidates. Allele-effect analysis of the six significant SNPs showed the presence of two or three alleles within each of these SNPs that regulated contrasting phenotypes of the associated traits. Moreover, multiple haplotype alleles detected within each of the six haplotype blocks regulated a diverse range of phenotypic variation for yield and yield-related traits. The GP analysis for four studied traits showed moderate to high accuracy using two methods (gBLUP and rrBLUP). The stable QTLs as well as the desirable SNP alleles and haplotype alleles (underlying these stable QTLs) identified for the yield-related traits can serve as potential resource for the improvement of soybean yield. After proper validation of these QTLs and alleles/haplotypes in different genetic backgrounds of soybean, they can be introduced into marker-assisted breeding programs for developing high-yielding soybean varieties. Besides, the putative candidate genes underlying these stable QTLs, after proper functional validation using overexpression or gene knockout studies, can be deployed in the development of high-yielding soybean varieties. Our study provided critical analyses of cultivated soybean genetic resources and identified novel genomic resources (QTLs and haplotype alleles) for soybean yield improvement programs.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: "https://www.soybase.org/projects/SoyBase.C2021.03.php".

Author contributions

JB, DY, RKV conceived and planned the experiments. JB, DH performed the experiments. KA, analyzed the data. KA, JB, SG, RB, and DH interpreted, validated and visualized the results. JB, KA wrote the initial draft of the manuscript. JB, DY, RKV supervised the project. All the authors read and revised the manuscript.

Funding

This work was supported in part by National Natural Science Foundation of China (No. 32090065, and Ministry of Science and Technology (No. 2017YFE0111000). JB also thank International Genome Center, Jiangsu University for their research assistance and facilities provided. RKV is thankful to Food Futures Institute, Murdoch University for supporting this research in part.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.953833/full#supplementary-material

References

Adeboye, K. A., Oduwaye, O. A., Daniel, I. O., Fofana, M., and Semon, M. (2021b). Characterization of flowering time response among recombinant inbred lines of WAB638-1/PRIMAVERA rice under reproductive stage drought stress. *Plant Genet. Resour.* 19 (1), 1–8. doi:10.1017/S1479262121000010

Adeboye, K. A., Semon, M., Oyetunde, O. A., Oduwaye, O. A., Adebambo, A. O., Fofana, M., et al. (2021a). Diversity array technology (DArT)-based mapping of phenotypic variations among recombinant inbred lines of WAB638-1/PRIMAVERA under drought stress. *Euphytica* 217, 130. doi:10.1007/s10681-021-02860-6

Alemu, A., Feyissa, T., Maccaferri, M., Sciara, G., Tuberosa, R., Ammar, K., et al. (2021). Genome-wide association analysis unveils novel QTLs for seminal root system architecture traits in Ethiopian durum wheat. *BMC Genomics* 22, 20–16. doi:10.1186/s12864-020-07320-4

Ali, M., Zhang, Y., Rasheed, A., Wang, J., and Zhang, L. (2020). Genomic prediction for grain yield and yield-related traits in Chinese winter wheat. *Int. J. Mol. Sci.* 21, 1342. doi:10.3390/ijms21041342

Alqudah, A. M., Sallam, A., Baenziger, P. S., and Börner, A. (2020). Gwas: Fast-Forwarding gene identification and characterization in temperate cereals: Lessons from barley–A review. *J. Adv. Res.* 22, 119–135. doi:10.1016/j.jare.2019.10.013

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 Genomics* 20, 527. doi:10.1186/s12864-019-5907-7

Badu-Apraku, B., Adewale, S., Paterne, A. A., Gedil, M., Toyinbo, J., and Asiedu, R. (2020). Identification of QTLs for grain yield and other traits in tropical maize under *Striga* infestation. *PLoS ONE* 15, e0239205. doi:10.1371/journal.pone.0239205

Beche, E., Gillman, J. D., Song, Q., Nelson, R., Beissinger, T., Decker, J., et al. (2021). Genomic prediction using training population design in interspecific soybean populations. *Mol. Breed.* 41, 15. doi:10.1007/s11032-021-01203-6

Bevan, M. W., Uauy, C., Wulff, B. B., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature* 543, 346–354. doi:10.1038/nature22011

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., et al. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* 7, 221. doi:10.3389/fgene.2016.00221

Bhat, J. A., Deshmukh, R., Zhao, T., Patil, G., Deokar, A., Shinde, S., et al. (2020). Harnessing high-throughput phenotyping and genotyping for enhanced drought tolerance in crop plants. *J. Biotechnol.* 324, 248–260. doi:10.1016/j.jbiotec.2020.11.010

Bhat, J. A., Karikari, B., Adeboye, K. A., Ganie, S. A., Barmukh, R., Hu, D., et al. (2022). Identification of superior haplotypes in a diverse natural population for breeding desirable plant height in soybean. *Theor. Appl. Genet.* 135, 2407–2422. doi:10.1007/s00122-022-04120-0

Bhat, J. A., Yu, D., Bohra, A., Ganie, S. A., and Varshney, R. K. (2021). Features and applications of haplotypes in crop breeding. *Commun. Biol.* 4, 1266. doi:10.1038/s42003-021-02782-y

Bianchi, J. S., Quijano, A., Gosparini, C. O., and Morandi, E. N. (2020). Changes in leaflet shape and seeds per pod modify crop growth parameters, canopy light environment, and yield components in soybean. *Crop J.* 8, 351–364. doi:10.1016/j.cj.2019.09.011

Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biol.* 12, 232–238. doi:10.1186/gb-2011-12-10-232

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

Chong, X., Zhang, F., Wu, Y., Yang, X., Zhao, N., Wang, H., et al. (2016). A SNP-enabled assessment of genetic diversity, evolutionary relationships and the identification of candidate genes in chrysanthemum. *Genome Biol. Evol.* 8, 3661–3671. doi:10.1093/gbe/evw270

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

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi:10.1534/genetics.110.116855

Diers, B. W., Specht, J., Rainey, K. M., Cregan, P., Song, Q., Ramasubramanian, V., et al. (2018). Genetic architecture of soybean yield and agronomic traits. *G3* 8, 3367–3375. doi:10.1534/g3.118.200332

Du, W., Wang, M., Fu, S., and Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. *J. Genet. Genom.* 36, 721–731. doi:10.1016/S1673-8527(08)60165-4

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., et al. (2017). Genomic selection for yield and seed protein content in soybean: A study of breeding program data and assessment of prediction accuracy. *Crop Sci.* 57, 1325–1337. doi:10.2135/cropsci2016.06.0496

Endelman, J. B., and Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *G3* 2, 1405–1413. doi:10.1534/g3.112.004259

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

Fang, Y., Liu, S., Dong, Q., Zhang, K., Tian, Z., Li, X., et al. (2020). Linkage analysis and multi-locus genome-wide association studies identify QTNs controlling soybean plant height. *Front. Plant Sci.* 11, 9. doi:10.3389/fpls.2020.00009

Gupta, P. K., Rustgi, S., and Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol. Biol.* 57, 461–485. doi:10.1007/s11103-005-0257-z

Habier, D., Fernando, R. L., and Dekkers, J. C. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi:10.1534/genetics.107.081190

Hamblin, M. T., Buckler, E. S., and Jannink, J.-L. (2011). Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27, 98–106. doi:10.1016/j.tig.2010.12.003

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., et al. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theor. Appl. Genet.* 125, 671–683. doi:10.1007/s00122-012-1859-x

Hill, W. G., and Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* 33, 54–78. doi:10.1016/0040-5809(88)90004-4

Hina, A., Cao, Y., Song, S., Li, S., Sharmin, R. A., Elattar, M. A., et al. (2020). High-resolution mapping in two RIL populations refines major "QTL Hotspot" regions for seed size and shape in soybean (*Glycine max* L.). *Int. J. Mol. Sci.* 21, 1040. doi:10.3390/ijms21031040

Hu, D., Zhang, H., Du, Q., Hu, Z., Yang, Z., Li, X., et al. (2020). Genetic dissection of yield-related traits via genome-wide association analysis across multiple environments in wild soybean (*Glycine soja* Sieb. and Zucc.). *Planta* 251, 39–17. doi:10.1007/s00425-019-03329-6

Hu, Z., Zhang, D., Zhang, G., Kan, G., Hong, D., and Yu, D. (2014). Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). *Breed. Sci.* 63, 441–449. doi:10.1270/jsbbs.63.441

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). Blink: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* 8, giy154. doi:10.1093/gigascience/giy154

Jeong, N., Suh, S. J., Kim, M. H., Lee, S., Moon, J. K., Kim, H. S., et al. (2012). *Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* 24, 4807–4818. doi:10.1105/tpc.112.104968

Julkowska, M. M., Saade, S., Agarwal, G., Gao, G., Pailles, Y., Morton, M., et al. (2019). MVApp—Multivariate analysis application for streamlined data analysis and curation. *Plant Physiol.* 180, 1261–1276. doi:10.1104/pp.19.00235

Karikari, B., Chen, S., Xiao, Y., Chang, F., Zhou, Y., Kong, J., et al. (2019). Utilization of interspecific high-density genetic map of RIL population for the QTL detection and candidate gene mining for 100-seed weight in soybean. *Front. Plant Sci.* 10, 1001. doi:10.3389/fpls.2019.01001

Klein, A., Houtin, H., Rond-Coissieux, C., Naudet-Huart, M., Touratier, M., Marget, P., et al. (2020). Meta-analysis of QTL reveals the genetic control of yield-related traits and seed protein content in pea. *Sci. Rep.* 10, 15925. doi:10.1038/s41598-020-72548-9

Kraakman, A. T., Niks, R. E., Van den Berg, P. M., Stam, P., and Van Eeuwijk, F. A. (2004). Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168, 435–446. doi:10.1534/genetics.104.026831

Lebedev, V. G., Lebedeva, T. N., Chernodubov, A. I., and Shestibratov, K. A. (2020). Genomic selection for forest tree improvement: Methods, achievements and perspectives. *Forests* 11, 1190. doi:10.3390/f11111190

Lee, C., Choi, M. S., Kim, H. T., Yun, H. T., Lee, B., Chung, Y. S., et al. (2015). Soybean [*Glycine max* (L.) Merrill]: Importance as a crop and pedigree reconstruction of Korean varieties. *Plant Breed. Biotechnol.* 2, 179–196. doi:10.9787/pbb.2015.3.3.179

Li, D., Zhao, X., Han, Y., Li, W., and Xie, F. (2019). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* 111, 90–95. doi:10.1016/j.ygeno.2018.01.004

Li, M., Liu, Y., Wang, C., Yang, X., Li, D., Zhang, X., et al. (2020). Identification of traits contributing to high and stable yields in different soybean varieties across three Chinese latitudes. *Front. Plant Sci.* 10, 1642. doi:10.3389/fpls.2019.01642

Lipka, A. E., Kandianis, C. B., Hudson, M. E., Yu, J., Drnevich, J., Bradbury, P. J., et al. (2015). From association to prediction: Statistical methods for the dissection and selection of complex traits in plants. *Curr. Opin. Plant Biol.* 24, 110–118. doi:10.1016/j.pbi.2015.02.010

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, W., Kim, M. Y., Van, K., Lee, Y. H., Li, H., Liu, X., et al. (2011). QTL identification of yield-related traits and their association with flowering and maturity in soybean. *J. Crop Sci. Biotechnol.* 14, 65–70. doi:10.1007/s12892-010-0115-7

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

Liu, X., Rahman, T., Song, C., Yang, F., Su, B., Cui, L., et al. (2018). Relationships among light distribution, radiation use efficiency and land equivalent ratio in maize-soybean strip intercropping. *Field Crops Res.* 224, 91–101. doi:10.1016/j.fcr.2018.05.010

Lu, X., Xiong, Q., Cheng, T., Li, Q. T., Liu, X. L., Bi, Y. D., et al. (2011). A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10, 670–684. doi:10.1016/j.molp.2017.03.006

Lubanga, N., Massawe, F., and Mayes, S. (2021). Genomic and pedigree-based predictive ability for quality traits in tea (*Camellia sinensis* (L.) O. Kuntze). *Euphytica* 217, 32–15. doi:10.1007/s10681-021-02774-3

Malik, M. F. A., Qureshi, A. S., Ashraf, M., and Ghafoor, A. (2006). Genetic variability of the main yield related characters in soybean. *Int. J. Agri. Biol.* 8, 815–819.

Matei, G., Woyann, L. G., Milioli, A. S., de Bem Oliveira, I., Zdziarski, A. D., Zanella, R., et al. (2018). Genomic selection in soybean: Accuracy and time gain in relation to phenotypic selection. *Mol. Breed.* 38, 117. doi:10.1007/s11032-018-0872-4

Mendonça, L. D. F., Galli, G., Malone, G., and Fritsche-Neto, R. (2020). Genomic prediction enables early but low-intensity selection in soybean segregating progenies. *Crop Sci.* 60, 1346–1361. doi:10.1002/csc2.20072

Merrick, L. F., and Carter, A. H. (2021). *Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. Plant Genome* 14, e20158. doi:10.1002/tpg2.20158

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. doi:10.1093/genetics/157.4.1819

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

Patil, G., Do, T., Vuong, T. D., Valliyodan, B., Lee, J. D., Chaudhary, J., et al. (2016). Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Sci. Rep.* 6, 19199. doi:10.1038/srep19199

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Qi, Z., Song, J., Zhang, K., Liu, S., Tian, X., Wang, Y., et al. (2020). Identification of QTNs controlling 100-seed weight in soybean using multilocus genome-wide association studies. *Front. Genet.* 11, 689. doi:10.3389/fgene.2020.00689

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

Ramalingam, J., Alagarasan, G., Savitha, P., Lydia, K., Pothiraj, G., Vijayakumar, E., et al. (2020). Improved host-plant resistance to Phytophthora rot and powdery mildew in soybean (*Glycine max* (L.) Merr.). *Sci. Rep.* 10, 13928. doi:10.1038/s41598-020-70702-x

Ravelombola, W., Qin, J., Shi, A., Song, Q., Yuan, J., Wang, F., et al. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS ONE* 16, e0255761. doi:10.1371/journal.pone.0255761

Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2020). Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. *PLoS ONE* 15, e0235089. doi:10.1371/journal.pone.0235089

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11479–11484. doi:10.1073/pnas.201394398

Rolling, W. R., Dorrance, A. E., and McHale, L. K. (2020). Testing methods and statistical models of genomic prediction for quantitative disease resistance to *Phytophthora sojae* in soybean [*Glycine max* (L.) Merr] germplasm collections. *Theor. Appl. Genet.* 133, 3441–3454. doi:10.1007/s00122-020-03679-w

Roorkiwal, M., Rathore, A., Das, R. R., Singh, M. K., Jain, A., Srinivasan, S., et al. (2016). Genome-enabled prediction models for yield related traits in chickpea. *Front. Plant Sci.* 7, 1666. doi:10.3389/fpls.2016.01666

Sahu, P. K., Mondal, S., Sao, R., Vishwakarma, G., Kumar, V., Das, B. K., et al. (2020). Genome-wide association mapping revealed numerous novel genomic loci for grain nutritional and yield-related traits in rice (*Oryza sativa* L.) landraces. *3 Biotech.* 10, 487. doi:10.1007/s13205-020-02467-z

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi:10.1214/aos/1176344136

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi:10.1038/ng.2314

Sinha, P., Singh, V. K., Saxena, R. K., Khan, A. W., Abbai, R., Chitikineni, A., et al. (2020). Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (*Cajanus cajan* L.). *Plant Biotechnol. J.* 18, 2482–2490. doi:10.1111/pbi.13422

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

Su, J., Zhang, F., Chong, X., Song, A., Guan, Z., Fang, W., et al. (2019). Genome-wide association study identifies favorable SNP alleles and candidate genes for waterlogging tolerance in chrysanthemums. *Hortic. Res.* 6, 21–13. doi:10.1038/s41438-018-0101-7

Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. E. (2021). Designing future crops: Genomics-assisted breeding comes of age. *Trends Plant Sci.* 26, 631–649. doi:10.1016/j.tplants.2021.03.010

Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* 132, 1797–1816. doi:10.1007/s00122-018-3252-x

Varshney, R. K., Terauchi, R., and McCouch, S. R. (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12, e1001883. doi:10.1371/journal.pbio.1001883

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132, 669–686. doi:10.1007/s00122-018-3270-8

Wang, D. G., Lin, Z., Kai, L. I., Ying, M. A., Wang, L. Q., Yang, Y. Q., et al. (2017a). Marker-assisted pyramiding of soybean resistance genes *RSC4, RSC8*, and *RSC14Q* to soybean mosaic virus. *J. Integ. Agri.* 16, 2413–2420.

Wang, J., Chu, S., Zhang, H., Zhu, Y., Cheng, H., and Yu, D. (2016). Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci. Rep.* 6, 20728. doi:10.1038/srep20728

Wang, X., Pang, Y., Wang, C., Chen, K., Zhu, Y., Shen, C., et al. (2017b). New candidate genes affecting rice grain appearance and milling quality detected by genome-wide and gene-based association analyses. *Front. Plant Sci.* 7, 1998. doi:10.3389/fpls.2016.01998

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, X., Yang, Z., and Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. *Sci. Bull. (Beijing).* 60, 925–935. doi:10.1007/s11434-015-0791-2

Wu, J., Zhao, Q., Liu, S., Shahid, M., Lan, L., Cai, G., et al. (2016). Genome-wide association study identifies new loci for resistance to Sclerotinia stem rot in *Brassica napus*. *Front. Plant Sci.* 7, 1418. doi:10.3389/fpls.2016.01418

Xu, Y., Wang, X., Ding, X., Zheng, X., Yang, Z., Xu, C., et al. (2018). Genomic selection of agronomic traits in hybrid rice using an NCII population. *Rice* 11, 32–10. doi:10.1186/s12284-018-0223-4

Yamanaka, N., and Hossain, M. M. (2019). Pyramiding three rust-resistance genes confers a high level of resistance in soybean (*Glycine max*). *Plant Breed.* 138, 686–695. doi:10.1111/pbr.12720

Yang, J., Liu, Z., Chen, Q., Qu, Y., Tang, J., Lübberstedt, T., et al. (2020). Publisher correction: Mapping of QTL for grain yield components based on a DH population in maize. *Sci. Rep.* 10, 13218. doi:10.1038/s41598-020-70123-w

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934. doi:10.1038/ng.3596

Yu, Z., Chang, F., Lv, W., Sharmin, R. A., Wang, Z., Kong, J., et al. (2019). Identification of QTN and candidate gene for seed-flooding tolerance in soybean [*Glycine max* (L.) Merr.] using genome-wide association study (GWAS). *Genes* 10, 957. doi:10.3390/genes10120957

Yue, F., Rong-rong, Z., Ze-chuan, L., Li-yong, C., Xing-hua, W., and Shi-hua, C. (2015). Quantitative trait locus analysis for rice yield traits under two nitrogen levels. *Rice Sci.* 22, 108–115. doi:10.1016/j.rsci.2015.05.014

Zaitlen, N. A., Kang, H. M., Feolo, M. L., Sherry, S. T., Halperin, E., and Eskin, E. (2005). Inference and analysis of haplotypes from combined genotyping studies deposited in dbSNP. *Genome Res.* 15, 1594–1600. doi:10.1101/gr.4297805

Zargar, S. M., Raatz, B., Sonah, H., Bhat, J. A., Dar, Z. A., Agrawal, G. K., et al. (2015). Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *J. Crop Sci. Biotechnol.* 18, 293–308. doi:10.1007/s12892-015-0037-5

Zatybekov, A., Abugalieva, S., Didorenko, S., Gerasimova, Y., Sidorik, I., Anuarbek, S., et al. (2017). GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. *BMC Plant Biol.* 17, 179. doi:10.1186/s12870-017-1125-0

Zeng, A., Chen, P., Korth, K., Hancock, F., Pereira, A., Brye, K., et al. (2017). Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. *Mol. Breed.* 37, 30–14. doi:10.1007/s11032-017-0634-8

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16, 217. doi:10.1186/s12864-015-1441-4

Zhang, S., Hao, D., Zhang, S., Zhang, D., Wang, H., Du, H., et al. (2021a). Genome-wide association mapping for protein, oil and water-soluble protein contents in soybean. *Mol. Genet. Genomics* 296, 91–102. doi:10.1007/s00438-020-01704-7

Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., et al. (2019). A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* 20, 5915. doi:10.3390/ijms20235915

Zhang, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021b). Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134, 1329–1341. doi:10.1007/s00122-021-03774-6

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. *Nat. Genet.* 42, 355–360. doi:10.1038/ng.546

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 August 2022
DOI 10.3389/fgene.2022.907267

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Cengiz Toker,
Akdeniz University, Turkey
Muhammad Azhar Nadeem,
Sivas University of Science and Technology, Turkey

*CORRESPONDENCE
Ahmed Sallam,
sallam@ipk-gatersleben.de,
amsallam@aun.edu.eg,
Yasser S. Moursi,
ysm01@fayoum.edu.eg
Shamseldeen Eltaher,
shams.eltaher@gebri.usc.edu.eg

†These authors share first authorship

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

RECEIVED 29 March 2022
ACCEPTED 18 July 2022
PUBLISHED 24 August 2022

CITATION
Sallam A, Moursi YS, Martsch R and Eltaher S (2022), Genome-wide association mapping for root traits associated with frost tolerance in faba beans using KASP-SNP markers.
Front. Genet. 13:907267.
doi: 10.3389/fgene.2022.907267

COPYRIGHT
© 2022 Sallam, Moursi, Martsch and Eltaher. 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.

Genome-wide association mapping for root traits associated with frost tolerance in faba beans using KASP-SNP markers

Ahmed Sallam[1,2*†], Yasser S. Moursi[3*†], Regina Martsch[4] and Shamseldeen Eltaher[5*]

[1]Department Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, [2]Department of Genetics, Faculty of Agriculture, Assiut University, Assiut, Egypt, [3]Department of Botany, Faculty of Science, Fayoum University, Fayoum, Egypt, [4]Department of Crop Sciences, Georg-August-Universität Göttingen, Göttingen, Germany, [5]Department of Plant Biotechnology, Genetic Engineering and Biotechnology Research Institute (GEBRI), University of Sadat City (USC), Sadat, Egypt

Frost is an abiotic stress factor that threatens plant development and crop productivity not only in cold regions but also in temperate zones. Roots play an important role in plant growth during frost stress. Therefore, variation in root characteristics could be studied to improve frost tolerance in winter faba bean. The present study aimed to identify the genomic regions that control frost tolerance in a winter faba bean population by focusing on root-related traits. A set of 185 genotypes were tested for frost tolerance under artificial frost growth conditions at −16°C, −18°C, and −19°C in a growth chamber. Frost stress reduced the root-related parameters in all genotypes, with a wide variation among genotypes. A genome-wide association study identified nine novel single-nucleotide polymorphisms that are associated with the root-related traits. The most frost-tolerant genotypes were identified; two genotypes, S_028 and S_220, exhibited remarkable performance under frost stress. Moreover, they harbored all four of the alleles favorable for frost tolerance. Remarkably, two markers showed genetic pleiotropic effects with positive allele effects on root fresh matter and root dry matter. Thus, both genotypes can be implemented in a breeding program to provide the alleles for healthier roots under frost conditions to develop more frost-tolerant varieties, and the two markers can be used to screen large collections to select for frost tolerance. These results may provide novel insights for improving frost tolerance in faba beans and in other legume crops.

KEYWORDS

***Vicia faba*, KASP, freezing temperature, root traits, GWAS**

Introduction

The faba bean (*Vicia faba* L.) is the only edible pulse crop among *Vicia* species (Lyu et al., 2021). Relative to other grain legumes, the faba bean is ranked fourth after chickpea, field pea, and lentil (Rebaa et al., 2017; Alharbi and Adhikari, 2020), with a global production of approximately 4.5 million tons produced from a cultivated area of approximately 2.5 Mha (Food and agriculture organization of the United Nations, 2019). The high protein content and well-balanced amino acid profile of the faba bean distinguish it from other legumes (Roy et al., 2010; Lyu et al., 2021). Owing to its high protein content, the faba bean is grown as a food crop in developing countries but principally as a livestock feed in Europe (Sallam et al., 2016a; Nadeem, 2021). Faba bean are less likely to generate off-flavors than soybean and pea owing to their low lipid content and low endogenous lipoxygenase activity, allowing them to be incorporated into everyday meals (Chang and McCurdy, 1985). As the demand for organic food grows, legume-based crops should be emphasized in crop rotations alongside cereals and oilseeds. Similar to other leguminous plants, the faba bean crop fixes atmospheric nitrogen into biologically usable ammonia, facilitating natural nitrogen fertilization of soils (Barton et al., 2014). Another advantage of the faba bean over other legumes is its ability to adapt to a wide range of climatic and soil conditions (Pociecha et al., 2008; Katerji et al., 2011; Zhou et al., 2018). Nevertheless, faba bean are also susceptible to a range of biotic and abiotic challenges, which contribute to loss due to environmental factors, including frost (Sallam et al., 2016b; O'Sullivan and Angra, 2016).

Frost is an important climatic factor that affects agricultural production and the agroforestry economy in temperate and subtropical areas worldwide (Snyder and Melo-abreu, 2005; Papagiannaki et al., 2014; Ambroise et al., 2020). Moreover, because of global warming and climate change, it is predicted that frost will become more problematic (Gu et al., 2008). In Northern and Central Europe, severe winter frost is a major abiotic stress factor affecting beans, and owing to the insufficient winter hardiness of the genotypes in use, the faba bean is primarily cultivated as a spring crop (Arbaoui et al., 2008; Sallam et al., 2016b). Frost stress not only decreases productivity but also lowers the diet value as a result of seed staining (Hawthorne, 2007). Additionally, frost can destroy the N-fixing bacteria by reducing the activity of nitrogenase, which threatens the symbiotic relationship (Stoddard et al., 2006).

To avoid the severe effects of summer drought on grain filling and ultimate yield, farmers have to sow the crops earlier and thus seek more frost-tolerant varieties (Rezaei et al., 2015). Although winter-type faba beans have a higher yield and protein content than spring-type faba beans, they are primarily grown as a spring crop in cool-temperate locations (Arbaoui et al., 2008; Sallam et al., 2015; Sallam et al., 2016a). Therefore, breeding for improving frost tolerance and developing cultivars with high frost tolerance is urgently needed. The development of frost-tolerant varieties *via* conventional breeding is challenging as frost tolerance is a polygenic trait (Li et al., 2011; Avia et al., 2013; Sallam et al., 2016a). Three parameters are critical in determining a genotype's winter hardiness: (i) frost tolerance, (ii) ability to survive biotic stresses such as snow mold, and (iii) resistance to abiotic stress factors such as high water saturation in the soil (Arbaoui, 2007; Sallam and Martsch, 2015; Sallam et al., 2016b). Thus, in addition to testing under simulated frost growth conditions in growth chambers, faba bean genotypes should be tested in the field to determine the overall winter hardiness of the most tolerant genotypes. In multi-location trials, many traits associated with winter hardiness should be assessed and analyzed. According to the climate prediction models, frost conditions will become more severe, causing a marked reduction in plant production due to a lack of root growth (Ambroise et al., 2020). Most studies on frost tolerance have focused on the examination of the aboveground organs; this is attributed to the fact that roots are protected under the soil surface (Lee et al., 2009; Smékalová et al., 2014). A previous study showed that, at the seedling stage, the faba bean is considered less cold tolerant than its wild relatives belonging to the western Mediterranean regions (Inci and Toker, 2011). Recently, Alharbi et al. (2021) have studied the effects of frost on pod setting in faba bean under field conditions, assessing several traits at the vegetative and reproductive stages. In both studies, the authors evaluated several morphological and yield traits but no root traits.

The faba bean is a diploid plant with 2n = 2x = 12 chromosomes and has one of the largest genomes among crops (approximately 13,000 Mb) (Satovic et al., 2013) compared with the genomes in soybean (1,200 Mb) and in field pea (4,000 Mb) (Torres et al., 2006; Satovic et al., 2013). This may make physical and genetic mapping, as well as map-based cloning, more challenging (Ellwood et al., 2008; Lavania et al., 2015). Because of the complexity of the faba bean genome, it is difficult to find a reference genome that can be used for genotyping by sequencing (GBS), for example, other crops such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), maize (*Zea mays*) and rice (*Oryza sativa*). Webb et al. (2016) constructed a faba bean consensus map (FBCM) using single-nucleotide polymorphism (SNP) markers developed from *Medicago truncatula* (legume model). The SNPs in this map were used in many studies to identify alleles associated with the target traits in the faba bean.

Using the recent advances in molecular genetics tools and methods along with a breeding program will accelerate the genetic improvement in frost tolerance in faba bean. One of these tools, marker-assisted selection (MAS), has been identified as a powerful approach for accelerating breeding operations to improve certain traits. Association mapping is one of the most successful strategies for identifying quantitative trait loci (QTLs) that underlie trait variation among the different methods used for

MAS. Moreover, it is a tool for mapping QTLs that regulate complex characteristics on the basis of marker–trait linkage disequilibrium. Previously, a genome-wide association study (GWAS) has been successfully employed to identify causative alleles associated with frost stress in faba bean. The GWAS can be conducted using any type of DNA molecular marker. Recently, kompetitive allele-specific PCR (KASP) genotyping has been introduced by LGC limited to be applied in QTL mapping and the GWAS (Semagn et al., 2014). The KASP is designed for a single-nucleotide polymorphism which differs between genotypes of the same species. KASP markers have many advantages over other DNA molecular markers including high sensitivity, reliability, reproducibility, and genotyping accuracy (Thomson, 2014). Webb et al. (2016) produced a set of 757 KASP markers that were used to genotype many faba bean populations for GWAS and QTL mapping studies.

Several studies have been published on frost tolerance in faba bean plants using the same genotypes, and several morphological characteristics such as stem and leaf characteristics, as well as physiological characteristics, have been studied (Sallam and Martsch, 2015; Sallam et al., 2015; Sallam et al., 2016a; Sallam et al., 2016b). Likewise, using the GWAS, alleles associated with several morphological parameters including loss of color and turgidity, freezing survival, and regrowth after freezing have been identified in the same collection of genotypes (Ali et al., 2016).

The effects of frost on root traits have not been extensively studied, and our study is one of the first reports, to the best of our knowledge, to investigate this essential issue. These screenings focused on the morphological traits of faba bean roots, such as root length (RL), root frost susceptibility (RFS), root fresh matter (RFM), and root dry matter (RDM). Thus, the present study aimed to i) investigate the genetic variation in root traits under frost stress; ii) study the correlations between frost tolerance and root traits; and iii) identify marker–trait associations (MTAs) and QTLs associated with root traits.

Materials and methods

A set of 183 single-seed-descent (SSD) lines of the faba bean from the Göttingen Winter Bean Population (GWBP) were used as plant materials (Supplementary Table S1). These SSD lines were developed from eleven founder lines: six lines from Germany (Webo, Wibo, Hiverna/1, 79/79, L977/88, and L979/S1), two lines from France (Côte d'Or/1 and Arrisot), and three lines from the UK (Banner, Bourdon, and Bulldo). These 11 founder lines were first sown in open pollination conditions and natural selection in order to produce a freely recombining population. After eight generations of ongoing maintenance of this population, < 400 SSD were drawn and inbred to generation > F9 and beyond (Stelling, 1989; unpublished; Gasim, 2003).

Two additional spring faba bean lines were used as controls (checks) in the frost experiments: Hedin/2 and Minica (Sallam and Martsch, 2015; Sallam et al., 2015; Sallam et al., 2016b).

Experimental layout and artificial frost experiment

The 185 genotypes (183 SSD lines + 2 controls) were assessed for frost tolerance in a Vötsch frost growth chamber (FGCh; 4 m^2). All genotypes were assessed for frost tolerance in eight replications. The use of the 4 m^2 area of this chamber once corresponded to one replication. Each replication included all genotypes in an alpha lattice design. In each replication, all genotypes were sown in pots filled with a mixture of compost soil and sand (3:1, respectively). Each pot contained four different genotypes with two seeds per genotype (Supplementary Figure S1). When all plants reached the two-leaf stage (at room temperature), all pots were moved to the well-isolated FGCh. The growth conditions (program) in the FGCh were set at a light exposure of 200 μmol s^{-1} m^{-2} for 10 h per day and 80–90% humidity. The seedlings of all genotypes were exposed to two phases described as follows: (I) a hardening phase in which all genotypes were exposed to a temperature of 4°C during the day and 0°C at night for 10 days and (II) the true frost test for 3 days at −16°C, −18°C, and −19°C, respectively, during the night and thawing during the hardening and frost days (Supplementary Figure S2). Plants were irrigated only during the hardening phase to keep the pots at approximately 70% of the soil's water capacity.

Scoring and measurement of phenotypic traits

In a previous study by Sallam et al. (2015), shoot traits in a set of 216 genotypes, including our 185 genotypes, were scored: loss of leaf color + loss of leaf turgidity (LC + LT) and frost tolerance index (FTI). The low values of LC + LT and the high values of FTI indicated frost tolerance according to Sallam et al. (2015). In addition to the shoot traits, this study also assessed the root traits but did not report them. In the present study, we evaluated the root traits and studied their relationship with frost tolerance.

After the frost test, the soil in each pot was carefully washed and removed to avoid root damage. Then, the roots of plants belonging to each genotype were cautiously washed again and prepared for the assessment of four root traits. RFS was visually scored from 1 (healthy roots) to 9 (withered black dead roots; Figure 1). RL was measured in cm, whereas RFM and RDM were measured in grams. Lower values of RFS and higher values of RL, RFM, and RDM indicated tolerance to frost.

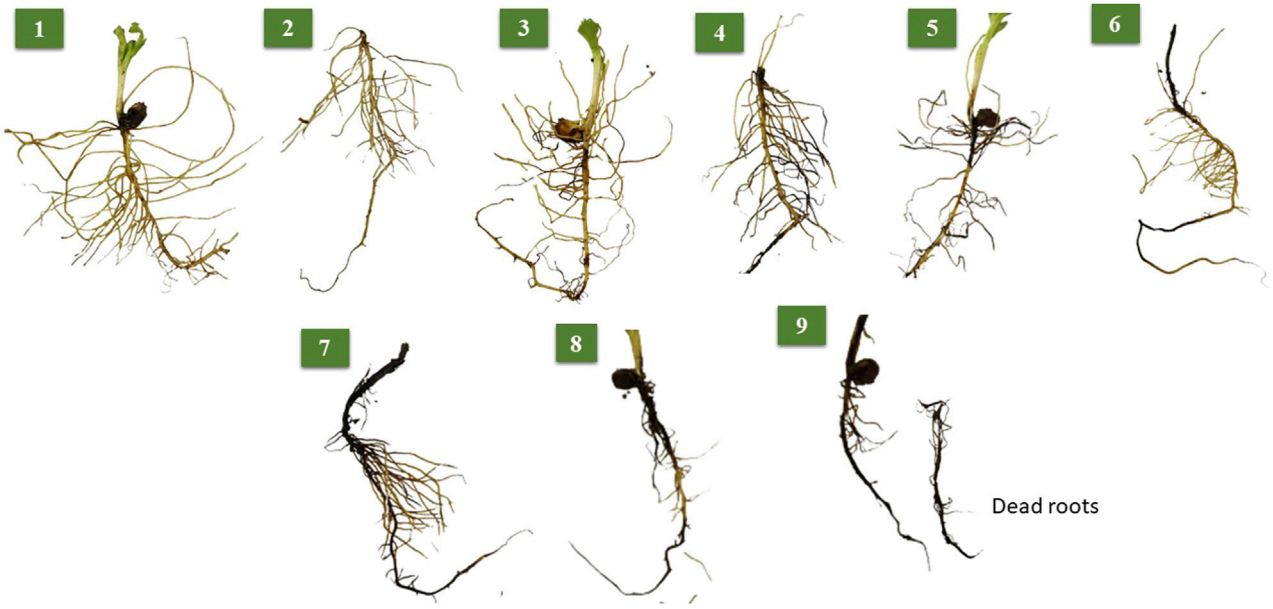


FIGURE 1
Visual scoring of root frost susceptibly scored on the 189 genotypes.

Statistical analyses for root traits

The analysis of variance was performed with PLABSTAT software (Utz, 1997) using the following equation:

$$Y_{ij} = \mu + g_i + r_j + g_{rij},$$

where Y_{ij} is the observation of the genotype i in replication j; μ is the general mean; g_i and r_j are the main effects of the genotype and replication, respectively; g_{rij} is the genotype × replication interaction of the genotype i with replication j. Broad-sense heritability (H^2) of each trait was calculated as the ratio of the genotypic variance with the phenotypic variance: $H^2 = \sigma^2 g/\sigma^2 p$ (Rasmusson and Lambert, 1960). Pearson's correlation coefficient was calculated among all traits by PLABSTAT software (Utz, 1997).

DNA extraction and KASP genotyping

Using Illustra Nucleon Phytopure Genomic DNA Extraction kits (GE Healthcare Life Sciences, UK), DNA was isolated from the 11 founder lines and GWBP. The DNA was genotyped utilizing the KASPar™ (kompetitive allele-specific PCR) test platform (KBioscience, United Kingdom), a single-plex SNP genotyping technology that uses allele-specific amplification followed by fluorescence detection. Of a total of 687 SNPs identified in the FBCM developed by Webb et al. (2016), 189 SNP markers were polymorphic among the 11 founder lines and were chosen to genotype the 189 genotypes (Sallam et al., 2016a; Webb et al., 2016). These SNP markers were mapped to six linkage groups using the legume model, *Medicago truncatula*. Each linkage group corresponded to one chromosome. The list of 189 SNPs and the sequence of KASP markers are presented in Supplementary Table S2.

Population structure and GWAS

Population structure analysis was performed using the 189 KASP markers on the whole population (Sallam and Martsch, 2015; Sallam et al., 2016a). The analysis revealed no population structure among the genotypes. Therefore, a GWAS was performed between markers and root traits using a general linear model (GLM). Significant markers associated with the root traits were detected using a threshold of 1% Bonferroni correction (Duggal et al., 2008). The GWAS was performed by TASSEL v5.2.40 (Bradbury et al., 2007). The phenotypic variation explained by marker (R^2) and allele effects was estimated for each trait by TASSEL v5.2.40. The quantile–quantile (QQ) plot was presented for each trait using the R package "qqman" (Turner, 2018). MapChart v2.2 (Voorrips, 2002) was used to illustrate the QTLs and their positions in the linkage group in the FBCM. The linkage disequilibrium (r^2) was calculated between each pair of significant SNPs located on the same chromosome by TASSEL v5.0.

TABLE 1 Ranges including minimum, maximum, and mean values, least-square difference, F-value, and heritability estimates (H^2) for root length, root fresh susceptibility, root fresh matter, and root dry matter of faba bean genotypes evaluated under frost conditions.

	Minimum	Maximum	Mean	L.S.D (p = 0.05)	F-value	Heritability
Root fresh susceptibility	2.06	9.00	6.36	1.83	4.59**	78.20
Root length	1.94	30.88	14.33	8.30	3.90**	74.37
Root fresh matter	0.11	3.44	1.13	1.11	2.93**	65.82
Root dry matter	0.01	0.26	0.096???	0.08	3.31**	69.81

F-value shows the genetic variation among genotypes. **$p < 0.01$.

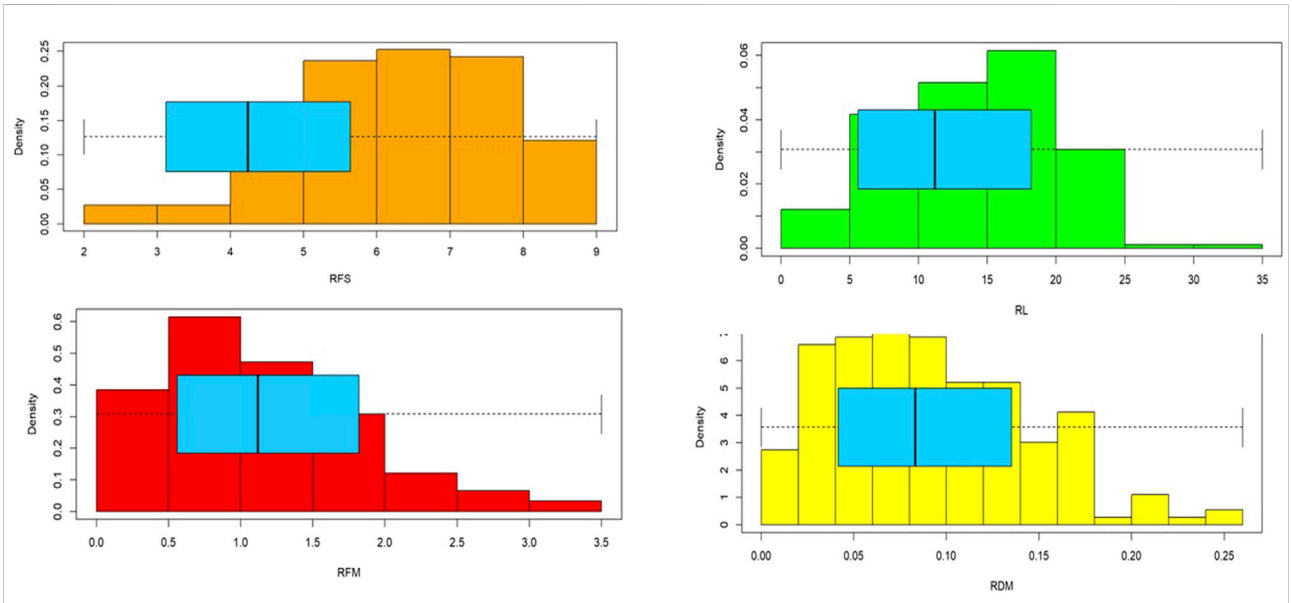


FIGURE 2
Histogram showing the distribution of all genotypes in each trait and box plot analysis illustrating the minimum, maximum, and mean values for each trait.

Results

Genetic variation

All the spring genotypes died under the imposed frost stress. Highly significant differences were found among all genotypes for all root traits scored after frost stress (Table 1). The H^2 values were 0.78, 0.74, 0.70, and 0.66 for RFS, RL, RDM, and RFM, respectively. The minimum, maximum, and mean values for all genotypes are presented in Table 1. The distribution of all genotypes for all traits is presented in Figure 2. Each trait had a wide range: 2.06–9.00 for RFS, 1.94–30.88 cm for RL, 0.11–3.44 g for RFM, and 0.01–0.26 g for RDM (Table 1; Figure 2). RFM and RDM exhibited a right-skewed distribution, whereas RFS had a left-skewed distribution. RL tended to have a normal distribution with some skew. The most tolerant genotype differed by traits. The S_002 genotype was the most frost-tolerant genotype with respect to RFS, with a score of 2.06. The genotype S_132 showed the highest RL of 30.88 cm. S_028 and S_052 genotypes showed the highest RFM and RDM, respectively.

Phenotypic correlation

Pearson's correlation analysis revealed very high and significant correlations among root traits (Figure 3A). A significant positive correlation was found among RL, RFM, and RDM. The highest correlation was found between RFM and RDM (r = 0.89***). RFS was negatively and significantly correlated with the other root traits, with r = −0.54**, −0.52**,

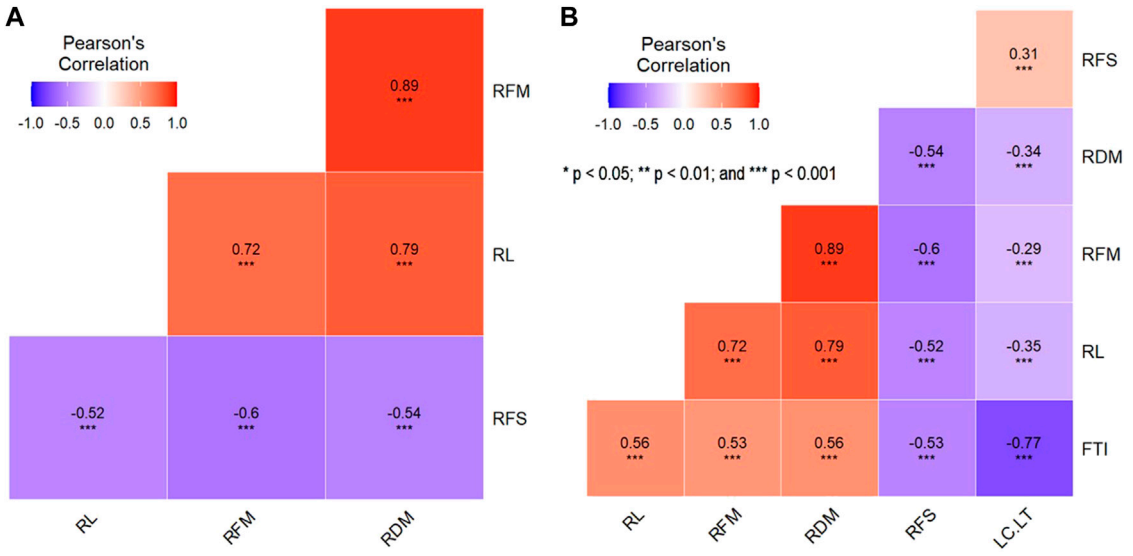


FIGURE 3
(A) Phenotypic correlation analysis among root traits and **(B)** correlation between both LC + LT and FTI scored on seedling shoots and root traits.

TABLE 2 List of significant SNPs associated with root traits using the GLM model including the *p*-value, chromosome number, phenotypic variation, target allele, and allele effect.

Trait	Marker	*p*-value	Chromosome no.	Position	Marker R^2%	Allele	Allele effect
RL	Vf_Mt1g072640_001	0.00496	3	67.73	7.34	C	3.34
RL	Vf_Mt7g051360_001	0.00261	5	33.55	6.75	A	3.32
RFS	Vf_Mt4g091530_001	0.00389	6	90.81	5.81	A	−0.78
RFS	Vf_Mt5g009720_001	0.000439	1	77.56	4.89	G	−0.84
RFM	Vf_Mt1g072640_001	0.000159	3	67.73	4.87	C	0.51
RFM	Vf_Mt7g073970_001	0.00304	5	62.92	4.78	C	0.31
RDM	Vf_Mt1g072640_001	0.000928	3	67.73	4.64	C	0.035
RDM	Vf_Mt1g082210_001	0.0027	3	83.25	4.49	A	0.029
RDM	Vf_Mt7g073970_001	0.00252	5	62.92	4.21	C	0.024

and -0.60^{**} between RFS and RDM, RL, and RFM, respectively.

As mentioned in the materials and methods, the FTI and LC + LT scored by Sallam et al. (2015) were used in this study as key frost-tolerant traits to test their correlation with root traits (Figure 3B). The correlation between root traits and both FTI and LC + LT is presented in Figure 3B. A significant negative correlation was found between LC + LT and RL, RDM, and RFM, whereas a significant positive correlation was found between LC + LT and RFS ($r = 0.31^{**}$).

The FTI was significantly correlated with all root traits; it was positively and significantly correlated with RDM ($r = 0.56^{***}$), RFM ($r = 0.53^{***}$), and RL ($r = 0.56^{***}$) and negatively and significantly correlated with RFS ($r = -0.53^{**}$).

Genome-wide association mapping

A marker–trait association analysis was performed between the 189 SNP markers and root traits for all genotypes (Table 2). Nine SNPs were found to be significantly associated with all root traits. The Manhattan plot of all the nine significant SNPs is presented in Supplementary Figure S3. The significant markers were distributed on chr1, chr3, chr5, and chr6. Chromosome 3 had the highest number of QTLs ($n = 4$), followed by chromosome 5 with three QTLs. A single QTL was identified on chromosomes 1 and 6. Of the nine SNP markers, eight markers had positive allele effects (i.e., enhancing the respective trait), but Vf_Mt4g091530_001 showed a negative allele effect on RFS (i.e., reducing the respective trait;

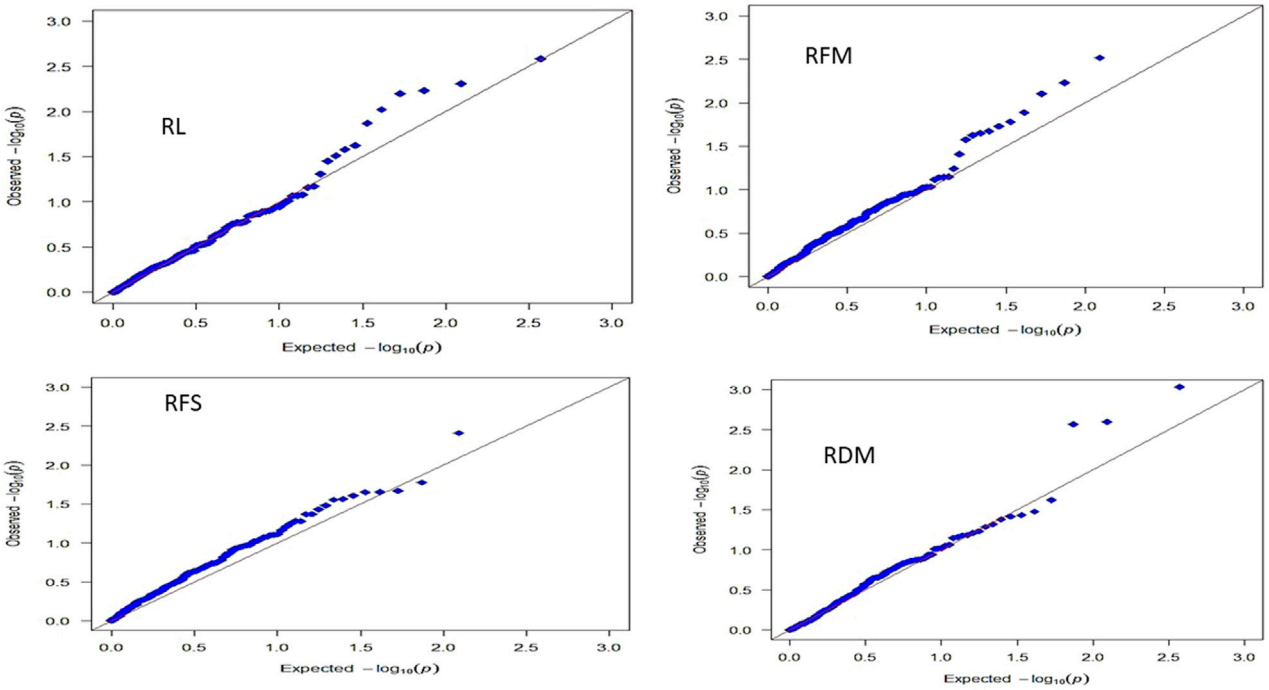


FIGURE 4
QQ plots resulted from the GWAS using the GLM model for each trait.

Table 2). The resultant QQ plots for each trait are presented in Figure 4.

Two SNPs (Vf_Mt1g072640_001 and Vf_Mt7g051360_001) located on chromosomes 3 and 5, respectively, were found to be significantly associated with RL. The phenotypic variation explained (R^2) was 7.34 and 6.75% for the SNP markers Vf_Mt1g072640_001 and Vf_Mt7g051360_001, respectively. The target allele (C) of Vf_Mt1g072640_001 and allele (A) of Vf_Mt7g051360_001 had allele effects of 3.34 and 3.32, respectively, on RL (Table 2).

Similarly, two SNPs (Vf_Mt4g091530_001 and Vf_Mt5g009720_001) located on chromosomes 6 and 1, respectively, were found to be significantly associated with RFS. R^2 was 5.81 and 4.89% for Vf_Mt4g091530_001 and Vf_Mt5g009720_001 markers, respectively. The target allele (A) of Vf_Mt4g091530_001 had a negative allele effect of −0.78 on RFS, whereas the target allele (G) of Vf_Mt5g009720_001 had an allele effect of −0.84 on RFS (Table 2).

Vf_Mt1g072640_001 and Vf_Mt7g073970_001 were found to be significantly associated with RFM. The phenotypic variation explained R^2 was 4.87 and 4.78% for Vf_Mt1g072640_001 and Vf_Mt7g073970_001 markers, respectively. The target allele (C) of Vf_Mt1g072640_001 and allele (C) of Vf_Mt7g073970_001 had allele effects of 0.51 and 0.31 on RFM, respectively (Table 2).

Three SNPs (Vf_Mt1g072640_001, Vf_Mt1g082210_001, and Vf_Mt7g073970_001) were found to be significantly associated with RDM. R^2 was 4.64, 4.49, and 4.21% for Vf_Mt1g072640_001, Vf_Mt1g082210_001, and Vf_Mt7g073970_001, respectively. The target alleles of these three markers had positive allele effects. The target alleles of Vf_Mt1g072640_001 (C), Vf_Mt1g082210_001 (A), and Vf_Mt7g073970_001 (C) had allele effects of 0.035, 0.029, and 0.024 on RDM, respectively (Table 2).

The linkage disequilibrium (LD; r^2) between each SNP pair located on the same chromosome was tested. No significant LD was found between markers located on chromosome 3 or 6 (Figure 5).

Notably, two markers showed a kind of pleiotropy as each marker was associated with more than one trait. The Vf_Mt1g072640_001 marker was found to be significantly associated with RL, RFM, and RDM with positive allele effects on all these traits, and Vf_Mt7g073970_001 had a significant association with RDM and RFM.

Discussion

Genetic variation

The high genetic variation among genotypes in the GWBP for all root traits is valuable for the improvement in frost tolerance in winter faba beans. The genetic variation was very useful for discriminating between the tolerant and susceptible

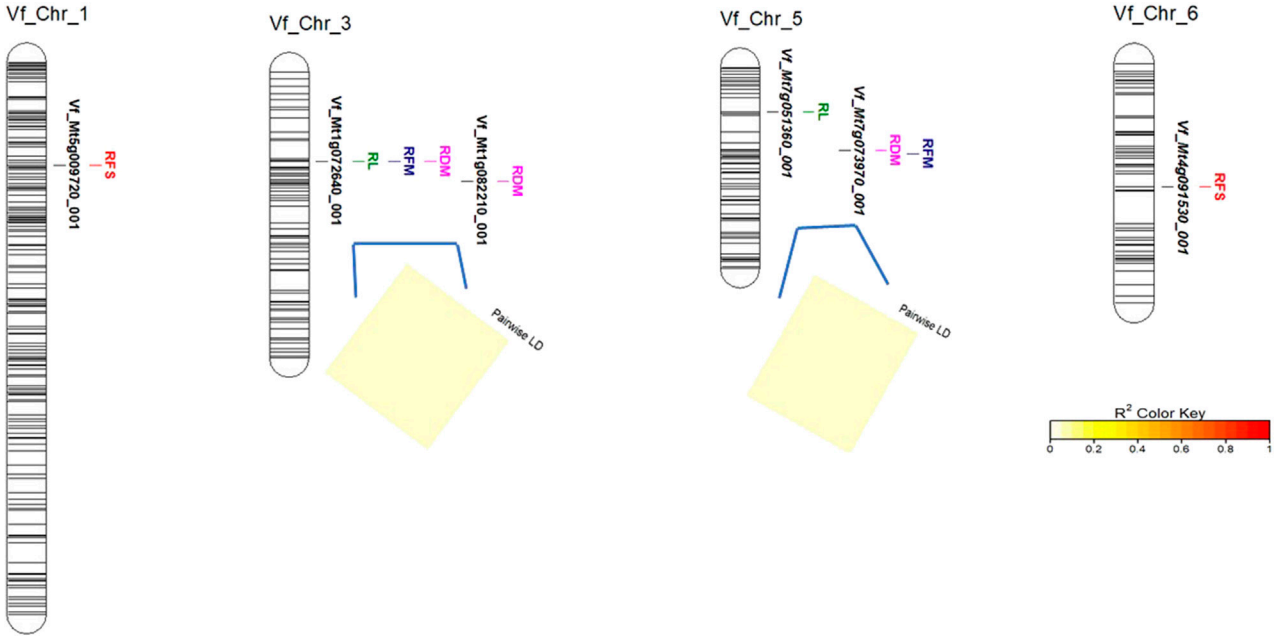


FIGURE 5
Distribution of QTLs associated with root traits detected by the GWAS and LD analysis (r^2) between SNPs located on the same chromosome.

genotypes in each trait. The high genetic variation among genotypes for all traits was expected as the population was derived from a natural cross of eight founder lines. This high genetic variation makes the GWBP ideal for various studies. The same population was used for studying frost tolerance on shoots at the seedling stage (Sallam et al., 2015), winter hardiness (Sallam et al., 2016a), drought tolerance (Ali et al., 2016), and agronomic performance (Gasim and Link, 2007).

In this study, all plants of the two spring control lines died during frost stress, indicating that the freezing temperature used in this study was suitable to determine the true performance of each genotype for each trait, that is, to distinguish the tolerant genotypes from the susceptible ones. The 10 most frost-tolerant genotypes for each trait were determined. Only two genotypes, S_028 and S_220, were found to be common among the 10 genotypes most tolerant with respect to the four root traits. These two genotypes experienced less injury from frost stress (low RFS score) and showed the highest values of RL, RFM, and RDM. Moreover, the two genotypes had been previously identified as frost-tolerant genotypes for shoot traits and had high survival after frost (Sallam et al., 2015). Notably, Sallam et al. (2016a) reported the same two genotypes with high winter survival under natural freezing temperatures in field conditions (high winter hardiness).

As yield is the ultimate goal of breeding programs and the aboveground organs experience higher frost stress than roots also because studying roots in the soil in field conditions is challenging. Thus, most studies on frost resistance have focused on the aboveground organs; however, roots play a crucial role in plant growth and final productivity. Roots need to stay biologically active for a longer time than the aboveground organs to ensure sufficient water and mineral supply to avoid reduction in plant growth and productivity (Ryyppö et al., 1998). Belowground frost damages fine roots, which are important for water and mineral uptake; regeneration of these damaged fine roots would occur at the expense of the shoot (Gaul et al., 2008). Root development was demonstrated to play a key role in the survival and establishment of sorghum grown under chilling conditions (Bekele et al., 2014). This observation is in agreement with our findings as frost damage to fine roots, particularly in the susceptible genotypes, results in plant death. Studying root traits at the seedling stage may provide good and useful information on the root characteristics of each genotype as roots can be easily removed from the pot and scored.

RFS, an important trait related to frost tolerance in faba beans, signifies the effect of frost on roots (Figure 1). In the present study, many genotypes, including the spring genotypes, showed withered black roots after the frost test. Similarly, a visual screening scale for the aboveground organs has been successfully used to select for frost tolerance in faba beans. The most tolerant genotypes showed less injury, whereas the most susceptible genotypes had withered black leaves and stems (Inci and Toker, 2011). It has been reported that the roots of susceptible genotypes did not die during the frost period but rather were affected during the cold acclimation period

(Ambroise et al., 2020). This trait was studied by Sallam et al. (2015), but an analysis of other root traits such as RL, RFM, and RDM under frost stress was reported for the first time in the present study to the best of our knowledge. The high heritability estimates for all traits were due to the non-significant environment. Such high heritability estimates indicate that selecting root traits to improve frost tolerance in a breeding program is feasible and effective.

Phenotypic correlation

Highly significant phenotypic correlations were found among all root traits under frost stress (Figures 3A, B). This result indicates the possibility of improving these traits together through a breeding program by selecting the best genotypes for all four root traits. Similar results were reported in sorghum under chilling conditions; root traits showed significant positive correlations under both normal and chilling conditions (Bekele et al., 2014). RFS was found to be negatively correlated with all other root traits, indicating that the less the frost injury in roots, the longer the roots (RL) and the higher the fresh and dry matter yield (RFM and RDM; Figures 3A, B). Kreyling et al. (2012) reported that freezing temperature causes severe damage to fine roots, resulting in a significant decrease in root biomass (by 50%). Similarly, a negative correlation was found between the total root density and frost susceptibility (Laughlin et al., 2021). Moreover, they found that plants with longer roots were found to tolerate frost better than those with shorter roots (Laughlin et al., 2021). These findings are consistent with our results as a high negative correlation was found between RFS and RFM ($r = -0.60^{**}$). To understand the interplay between root traits and aboveground organ (shoot and leaf) traits that were scored earlier in the same population by Sallam et al. (2015), a correlation analysis among these traits was performed. LC + LT refers to the symptoms of frost stress on the leaves of faba bean seedlings, whereas the FTI comprises survival traits (regrowth after frost and tendency to survive) scored after frost (Sallam et al., 2015). The highly significant phenotypic correlation between shoot and root traits highlighted the importance of healthy roots in frost tolerance. The correlation of FTI with all root traits showed a higher significance than LC + LT with all root traits, indicating the role of roots in the survival mechanism after the frost test (Figure 3B). However, shoots exhibited different physiological aspects and stress responses than those exhibited by roots, indicating two different frost hardening and resistance mechanisms (Ryyppö et al., 1998). In the present study, highly significant positive correlations were found between shoot and root traits. Likewise, in sorghum, under chilling conditions, root biomass and RL were positively correlated with plant survival (Balota et al., 2010; Bekele et al., 2014). Similarly, root weight has been used successfully as a selection parameter for cold tolerance in Cyprus Vetch under field growth conditions (Ratinam et al., 1994). Taken together, these findings suggested that it is crucial for plants to have highly biologically active roots for a longer time than any other organs (e.g., shoots) as the roots are the main source of water and nutrient supply to aboveground tissues.

Therefore, breeding for root frost tolerance at the seedling stage is very important to produce cultivars with high frost tolerance. It has been previously reported that there is a significant correlation between frost tolerance at the seedling stage and under field conditions (Sallam et al., 2016a; Sallam et al., 2017). Because it is difficult to study the roots in field conditions, studying the root traits at the seedling stage is a suitable alternative to obtain information on the degree of frost tolerance of each tested genotype.

Genome-wide association

Association mapping is regarded as a powerful approach for identifying variations responsible for key features in various crops (Pritchard and Przeworski, 2001; Reich et al., 2001). The association mapping in GWBP was performed on various genetic backgrounds of 189 SSD lines obtained from a natural cross between 11 genotypes collected from different parts of Europe (Sallam et al., 2015). A GWAS using the same number of SNPs (189) has been performed on a subset of this population ($n = 182$ genotypes) in earlier studies to identify alleles associated with frost tolerance in faba bean shoots (Sallam and Martsch, 2015; Sallam et al., 2016b) and determine leaf fatty acid composition under hardening conditions (Sallam et al., 2016b), winter hardiness in field conditions (Sallam et al., 2016a), drought tolerance at the seedling stage (Ali et al., 2016), and *Ascochyta fabae* resistance (Faridi et al., 2021). These studies indicate that this population has a wide genetic variation, which allows breeders and geneticists to identify markers associated with the target traits. However, in the present study, the number of KASP markers was lower than that in other GWAS studies in other crops because of the higher complexity of the faba bean genome relative to that of other legume crops and the lack of a faba bean reference genome. Thus, genotyping by sequencing in faba beans remains a challenging task. The KASP markers used in the present study may be the best option for genetic studies to identify markers associated with target traits in faba beans (Sallam and Ul-Allah, 2019). This set of KASP markers was initially developed by Webb et al. (2016). The GWBP is a highly diverse population ideal for the GWAS because (1) it has no population structure, which causes spurious association, and (2) it has an extremely low degree of LD (Sallam et al., 2015). A long-range LD increases the number of false associations in the marker–trait association analysis by the GWAS (Alqudah et al., 2020).

As there is no population structure among genotypes in the GWBP, a GLM model was the appropriate analysis approach that

fits with association mapping (Alqudah et al., 2020). The results of QQ plots resulting from the GWAS for each trait further supported the efficiency of the GLM model in detecting the marker–trait associations. Most of the observed and expected *p*-values lie on the diagonal line except for the true significant markers associated with the traits. The QQ plot analysis indicates the efficiency of GWAS results (Alqudah et al., 2020).

In the present study, the association analysis revealed that nine different significant markers were associated with RL, RFS, RFM, and RDM (Table 2; Figure 5). In the same population, 52 SNPs associated with frost tolerance in faba bean shoots at the seedling stage were previously reported (Sallam et al., 2015). Using a biparental winter faba bean population (101 recombinant inbred lines) and a set of 113 KASP markers developed by Webb et al. (2016), 27 QTLs were found to be associated with frost tolerance-related shoot traits at the seedling stage (Sallam et al., 2016b). Thus, the present study reports novel SNPs associated with root traits under frost stress. All SNPs detected in the two aforementioned studies were located on chromosomes 1, 2, 3, 4, and 6 in the FBCM (Webb et al., 2016). Most of the SNPs associated with frost tolerance in the shoots were located on chromosome 1 (Sallam and Ul-Allah, 2019). In the present study, the nine identified SNPs were located on chromosomes 1, 3, 5, and 6. Notably, two novel SNPs associated with root traits were located on chromosome 5, which had not been previously reported to be associated with shoot traits either in the GWBP or the biparental populations.

Remarkably, when the same KSAP markers were used to identify the markers associated with frost tolerance in the shoots of two different populations, no marker was found to be significantly associated with any of the shoot traits. This finding indicates that the markers significantly associated with root traits were entirely different from those associated with shoot traits under frost stress. This also supports the idea that the mechanism of frost resistance in the leaves or shoots is different from that in the roots. This is partially supported by the negative and significant correlations between the leaf traits (LC + LT) and the root traits (RL, RFM, and RDM; Figure 3B).

Notably, two markers were found to be associated with more than one trait: Vf_Mt1g072640_001 associated with RL, RFM, and RDM and Vf_Mt7g073970_001 associated with RFM and RDM (Table 2; Figure 5). It is important to note that these two markers had pleiotropic genetic effects (i.e., for each marker, the same allele controls the variation in both traits RFM and RDM simultaneously), and their alleles have positive allele effects. It is highly likely that the alleles of these two markers play a crucial role in frost tolerance by increasing these advantageous traits. Another valuable marker is Vf_Mt4g091530_001, which had a negative allele effect on RFS, indicating that it increases frost tolerance by decreasing RFS and should be validated in different genetic backgrounds to be used for improving RFM and RDM because the allele effects for both markers (Vf_Mt1g072640_001 and Vf_Mt7g073970_001) were positive. Moreover, all markers detected in this study had $R^2 < 0.10$, which indicated that all these QTLs had minor effects on frost tolerance. Non-significant LD between marker pairs located on chromosomes 3 and 5 indicates that these SNPs represent individual QTL associated with root frost tolerance.

Notably, the two markers associated with RL, RFM, and RDM could be considered very informative markers as they had pleiotropic genetic effects (Table 2; Figure 5), and their target alleles have positive allele effects on these traits. Genome-wide association study was very useful to identify alllele associated with root traits under frost stress. However, more studies are need to understand the genetic control of frost tolerance in faba bean. Moreover, genomic selection in fabe bean should be utilized to improve target trait such as frost tolerance. Genomic selection is the best approach for breeding target traits after considering seed amount, cost, labor and time (Sandhu et al., 2021).

Conclusion

This study elucidates the vital role of roots and their association with frost tolerance in winter faba beans. Nine novel SNPs and genomic regions controlling root traits under frost stress were reported. The highly significant correlations found between the FTI and root traits are promising for the selection of truly frost-tolerant genotypes. The two genotypes S_028 and S_220 are candidate parent genotypes because both of them show tolerance to frost stress at the level of shoots and roots at the seedling stage and winter hardiness under field conditions. Notably, both genotypes possess four alleles out of the six frost tolerance-associated alleles. The two markers Vf_Mt1g072640_001 and Vf_Mt7g073970_001 are of high significance because they show pleiotropic effects on RFM and RDM, and their alleles have positive additive effects on both traits. Notably, the allele C of the marker Vf_Mt1g072640_001 can be used to select for improving RL, RFM, and RDM, and the allele A of the marker Vf_Mt4g091530_001 decreases RFS. Thus, employing these two genotypes and these markers may help in developing or selecting faba bean cultivars with high frost tolerance and high winter hardiness.

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.

Author contributions

AS conceptualization, designed the study, conducted the experiments, helped score the data in the statistical analyses, and drafted the manuscript. YM helped in scoring the data, reviewed the manuscript, and discussed the data. RM helped in the experiments and scored the data, and SE helped in the statistical analyses and helped in drafting and reviewing the manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.907267/full#supplementary-material

References

Alharbi, N. H., and Adhikari, K. N. (2020). Factors of yield determination in faba bean (*Vicia faba*). *Crop Pasture Sci.* 71 (4), 305–321. doi:10.1071/CP19103

Alharbi, N. H., Alghamdi, S. S., Migdadi, H. M., El-harty, E. H., and Adhikari, K. N. (2021). Evaluation of frost damage and pod set in faba bean (*Vicia faba* L.) under field conditions. *Plants* 10 (9), 1925. doi:10.3390/plants10091925

Ali, M. B. M., Welna, G. C., Sallam, A., Martsch, R., Balko, C., Gebser, B., et al. (2016). Association analyses to genetically improve drought and freezing tolerance of faba bean (*Vicia faba* L.). *Crop Sci.* 56 (3), 1036–1048. doi:10.2135/cropsci2015.08.0503

Alqudah, A. M., Sallam, A., Baenziger, P. S., and Börner, A. (2020). GWAS: Fast-forwarding gene identification and characterization in temperate cereals: lessons from barley--A review. *J. Adv. Res.* 22, 119–135. doi:10.1016/j.jare.2019.10.013

Ambroise, V., Legay, S., Guerriero, G., Hausman, J.-F., Cuypers, A., and Sergeant, K. (2020). The roots of plant frost hardiness and tolerance. *Plant Cell Physiol.* 61 (1), 3–20. doi:10.1093/pcp/pcz196

Arbaoui, M. (2007). Detailed genetic analysis of faba bean *(Vicia faba L.)* winter-hardiness and related traits. Dissertation.

Arbaoui, M., Link, W., Satovic, Z., and Torres, A.-M. (2008). Quantitative trait loci of frost tolerance and physiologically related trait in faba bean (*Vicia faba* L.). *Euphytica* 164 (1), 93–104. doi:10.1007/s10681-008-9654-0

Avia, K., Pilet-Nayel, M. L., Bahrman, N., Baranger, A., Delbreil, B., Fontaine, V., et al. (2013). Genetic variability and QTL mapping of freezing tolerance and related traits in medicago truncatula. *Theor. Appl. Genet.* 126 (9), 2353–2366. doi:10.1007/S00122-013-2140-7

Balota, M., Payne, W. A., Veeragoni, S. K., Stewart, B. A., and Rosenow, D. T. (2010). Respiration and its relationship to germination, emergence, and early growth under cool temperatures in sorghum. *Crop Sci.* 50 (4), 1414–1422. doi:10.2135/cropsci2009.08.0448

Barton, L., Thamo, T., Engelbrecht, D., and Biswas, W. K. (2014). Does growing grain legumes or applying lime cost effectively lower greenhouse gas emissions from wheat production in a semi-arid climate? *J. Clean. Prod.* 83, 194–203. doi:10.1016/j.jclepro.2014.07.020

Bekele, W. A., Fiedler, K., Shiringani, A., Schnaubelt, D., Windpassinger, S., Uptmoor, R., et al. (2014). Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions. *Plant Cell Environ.* 37 (3), 707–723. doi:10.1111/pce.12189

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 (19), 2633–2635. doi:10.1093/bioinformatics/btm308

Chang, P. R. Q., and McCurdy, A. R. (1985). Lipoxygenase activity in fourteen legumes. *Can. Inst. Food Sci. Technol. J.* 18 (1), 94–96. doi:10.1016/s0315-5463(85)71727-0

Duggal, P., Gillanders, E. M., Holmes, T. N., and Bailey-Wilson, J. E. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9 (1), 516. doi:10.1186/1471-2164-9-516

Ellwood, S. R., Phan, H. T. T., Jordan, M., Hane, J., Torres, A. M., Avila, C. M., et al. (2008). Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with Lens culinaris. *BMC Genomics* 9 (1), 380. doi:10.1186/1471-2164-9-380

Faridi, R., Koopman, B., Schierholt, A., Ali, M. B., Apel, S., and Link, W. (2021). Genetic study of the resistance of faba bean (*Vicia faba*) against the fungus ascochyta fabae through a genome-wide association analysis. *Plant Breed.* 140 (3), 442–452. doi:10.1111/pbr.12918

Food and agriculture organization of the United Nations (2019). Food and agriculture organization of the United Nations. *Int. Organ.* 1 (2), 350–353. doi:10.1017/S0020818300006160

Gasim, S., and Link, W. (2007). Agronomic performance and the effect of self-fertilization on German winter faba beans. *J. Central Eur. Agric.* 8 (1), 121–128.

Gasim, S. M. (2003). *Winter-hardy faba bean: applied genetic research on the reproductive mode of the European gene pool* Ph.D thesis. Germany: University of Göttingen.

Gaul, D., Hertel, D., and Leuschner, C. (2008). Effects of experimental soil frost on the fine-root system of mature Norway spruce. *J. Plant Nutr. Soil Sci.* 171 (5), 690–698. doi:10.1002/jpln.200700284

Gu, L., Hanson, P. J., Post, W. M, Kaiser, D. P., Yang, B., Nemani, R., et al. (2008). The 2007 eastern US spring freeze: increased cold damage in a warming world? *BioScience* 58 (3), 253–262. doi:10.1641/B580311

Hawthorne, W. (2007). A summary of pulse management options to minimize frost damage. *Pulse Aust. Bull.* 2, 1–4.

Inci, N. E., and Toker, C. (2011). Screening and selection of faba beans (*Vicia faba* L.) for cold tolerance and comparison to wild relatives. *Genet. Resour. Crop Evol.* 58 (8), 1169–1175. doi:10.1007/s10722-010-9649-2

Katerji, N., Mastrorilli, M., Lahmer, F. Z., Maalouf, F., and Oweis, T. (2011). Faba bean productivity in saline--drought conditions. *Eur. J. Agron.* 35 (1), 2–12. doi:10.1016/j.eja.2011.03.001

Kreyling, J., Haei, M., and Laudon, H. (2012). Absence of snow cover reduces understory plant cover and alters plant community composition in boreal forests. *Oecologia* 168 (2), 577–587. doi:10.1007/s00442-011-2092-z

Laughlin, D. C., Mommer, L., Sabatini, F. M., Bruelheide, H., Kuyper, T. W., McCormack, M. L., et al. (2021). Root traits explain plant species distributions along climatic gradients yet challenge the nature of ecological trade-offs. *Nat. Ecol. Evol.* 5 (8), 1123–1134. doi:10.1038/s41559-021-01471-7

Lavania, D., Siddiqui, M. H., Al-Whaibi, M. H., Singh, A. K., Kumar, R., and Grover, A. (2015). Genetic approaches for breeding heat stress tolerance in faba bean (*Vicia faba* L.). *Acta Physiol. Plant.* 37 (1), 1737. doi:10.1007/s11738-014-1737-z

Lee, D. G., Ahsan, N., Lee, S. H., Lee, J. J., Bahk, J. D., Kang, K. Y., et al. (2009). Chilling stress-induced proteomic changes in rice roots. *J. Plant Physiol.* 166 (1), 1–11. doi:10.1016/j.jplph.2008.02.001

Li, Y., Böck, A., Haseneyer, G., Korzun, V., Wilde, P., Schön, C. C., et al. (2011). Association analysis of frost tolerance in rye using candidate genes and phenotypic data from controlled, semi-controlled, and field phenotyping platforms. *BMC Plant Biol.* 11 (1), 146. doi:10.1186/1471-2229-11-146

Lyu, J. I., Ramekar, R., Kim, J. M., Hung, N. N., Seo, J. S., Kim, J.-B., et al. (2021). Unraveling the complexity of faba bean (*Vicia faba* L.) transcriptome to reveal cold-stress-responsive genes using long-read isoform sequencing technology. *Sci. Rep.* 11 (1), 21094. doi:10.1038/s41598-021-00506-0

Nadeem, M. A. (2021). Türk bakla genetik kaynaklarının tohum mineral İçeriği için fenotipik çeşitliliğin araştırılması. *Uluslar. Tarım ve Yaban Hayatı Bilim. Derg.* 7 (3), 540–550. doi:10.24180/ijaws.949496

O'Sullivan, D. M., and Angra, D. (2016). Advances in faba bean genetics and genomics. *Front. Genet.* 7, 150. doi:10.3389/fgene.2016.00150

Papagiannaki, K., Lagouvardos, K., Kotroni, V., and Papagiannakis, G. (2014). Agricultural losses related to frost events: use of the 850 hPa level temperature as an explanatory variable of the damage cost. *Nat. Hazards Earth Syst. Sci.* 14, 2375–2386. doi:10.5194/nhess-14-2375-2014

Pociecha, E., Kościelniak, J., and Filek, W. (2008). Effects of root flooding and stage of development on the growth and photosynthesis of field bean (*Vicia faba L. minor*). *Acta Physiol. Plant.* 30 (4), 529–535. doi:10.1007/s11738-008-0151-9

Pritchard, J. K., and Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 69 (1), 1–14. doi:10.1086/321275

Rasmusson, D. C., and Lambert, J. W. (1960). "Variety X environment interactions in barley variety tests," in 5-8, December 1960, Abstracts of the annual meeting of the american society of agronomy, Chicago.

Ratinam, M., Abd El Moneim, A. M., and Saxena, M. C. (1994). Variations in sugar content and dry matter distribution in roots and their associations with frost tolerance in certain forage legume species. *J. Agron. Crop Sci.* 173 (5), 345–353. doi:10.1111/j.1439-037X.1994.tb00582.x

Rebaa, F., Abid, G., Aouida, M., Abdelkarim, S., Aroua, I., Muhovski, Y., et al. (2017). Genetic variability in Tunisian populations of faba bean (Vicia faba L. var. major) assessed by morphological and SSR markers. *Physiol. Mol. Biol. Plants* 23 (2), 397–409. doi:10.1007/s12298-017-0419-x

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., et al. (2001). Linkage disequilibrium in the human genome. *Nature* 411 (6834), 199–204. doi:10.1038/35075590

Rezaei, E. E., Siebert, S., and Ewert, F. (2015). Intensity of heat stress in winter wheat - phenology compensates for the adverse effect of global warming. *Environ. Res. Lett.* 10 (2), 024012. doi:10.1088/1748-9326/10/2/024012

Roy, F., Boye, J. I., and Simpson, B. K. (2010). Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. *Food Res. Int.* 43 (2), 432–442. doi:10.1016/j.foodres.2009.09.002

Ryyppö, A., Repo, T., and Vapaavuori, E. (1998). Development of freezing tolerance in roots and shoots of Scots pine seedlings at nonfreezing temperatures. *Can. J. For. Res.* 28 (4), 557–565. doi:10.1139/x98-022

Sallam, A., Arbaoui, M., El-Esawi, M., Abshire, N., and Martsch, R. (2016b). Identification and verification of QTL associated with frost tolerance using linkage mapping and GWAS in winter faba bean. *Front. Plant Sci.* 7, 1098. doi:10.3389/fpls.2016.01098

Sallam, A., Dhanapal, A. P., and Liu, S. (2016a). Association mapping of winter hardiness and yield traits in faba bean (*Vicia faba* L.). *Crop Pasture Sci.* 67, 55–68. doi:10.1071/cp15200

Sallam, A., Ghanbari, M., and Martsch, R. (2017). Genetic analysis of winter hardiness and effect of sowing date on yield traits in winter faba bean. *Sci. Hortic.* 224, 296–301. doi:10.1016/j.scienta.2017.04.015

Sallam, A., and Martsch, R. (2015). Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica* 143 (4), 501–514. doi:10.1007/s10709-015-9848-z

Sallam, A., Martsch, R., and Moursi, Y. S. (2015). Genetic variation in morpho-physiological traits associated with frost tolerance in faba bean (*Vicia faba* L.). *Euphytica* 205 (2), 395–408. doi:10.1007/s10681-015-1395-2

Sallam, A., and Ul-Allah, S. (2019). "Genomics-aided breeding for climate-smart traits in faba bean," in *Genomic designing of climate-smart pulse crops* (Cham: Springer), 359–395.

Sandhu, K. S., Aoun, M., Morris, C. F., and Carter, A. H. (2021). Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. *Biology (Basel)* 10. doi:10.3390/biology10070689

Satovic, Z., Avila, C. M., Cruz-Izquierdo, S., Díaz-Ruíz, R., García-Ruíz, G. M., Palomino, C., et al. (2013). A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). *BMC Genomics* 14 (1), 932. doi:10.1186/1471-2164-14-932

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33 (1), 1–14. doi:10.1007/s11032-013-9917-x

Smékalová, V., Doskočilová, A., Komis, G., and Šamaj, J. (2014). Crosstalk between secondary messengers, hormones and MAPK modules during abiotic stress signalling in plants. *Biotechnol. Adv.* 32 (1), 2–11. doi:10.1016/j.biotechadv.2013.07.009

Snyder, R. L., and Melo-abreu, J. P. D. (2005). *Protection: fundamentals , practice and economics*. Rome: FAO.

Stoddard, F. L., Balko, C., Erskine, W., Khan, H. R., Link, W., and Sarker, A. (2006). Screening techniques and sources of resistance to abiotic stresses in cool-season food legumes. *Euphytica* 147 (1–2), 167–186. doi:10.1007/s10681-006-4723-8

Stelling, D. (1989). Problems of breeding for improved standing ability in dried peas, *Pisum sativum* L. *J. Agr. Crop Sci.* 163.1 (1–2), 21–32. doi:10.1111/j.1439-037X.1989.tb00733.x

Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotechnol.* 2 (3), 195–212. doi:10.9787/pbb.2014.2.3.195

Torres, A. M., Román, B., Avila, C. M., Satovic, Z., Rubiales, D., Sillero, J. C., et al. (2006). Faba bean breeding for resistance against biotic stresses: towards application of marker technology. *Euphytica* 147 (1–2), 67–80. doi:10.1007/s10681-006-4057-6

Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* 3, 731. doi:10.21105/joss.00731

Utz, H. F. (1997). "A computer program for statistical analysis of plant breeding experiments," in *Version 2N. Institute of plant breeding, seed science and population genetics* (Stuttgart, Germany: University of Hohenheim).

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93 (1), 77–78. doi:10.1093/jhered/93.1.77

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., et al. (2016). A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnol. J.* 14 (1), 177–185. doi:10.1111/pbi.12371

Zhou, R., Hyldgaard, B., Yu, X., Rosenqvist, E., Ugarte, R. M., Yu, S., et al. (2018). Phenotyping of faba beans (*Vicia faba* L.) under cold and heat stresses using chlorophyll fluorescence. *Euphytica* 214 (4), 68. doi:10.1007/s10681-018-2154-y

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 24 August 2022
doi: 10.3389/fgene.2022.897696

Innovation in the Breeding of Common Bean Through a Combined Approach of *in vitro* Regeneration and Machine Learning Algorithms

Muhammad Aasim[1], Ramazan Katirci[2], Faheem Shehzad Baloch[1], Zemran Mustafa[3], Allah Bakhsh[4], Muhammad Azhar Nadeem[1], Seyid Amjad Ali[5], Rüştü Hatipoğlu[6], Vahdettin Çiftçi[7], Ephrem Habyarimana[8], Tolga Karaköy[1] and Yong Suk Chung[9]**

[1]Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas, Turkey, [2]Department of Metallurgical and Materials Engineering, Faculty of Engineering and Natural Sciences, Sivas University of Science and Technology, Sivas, Turkey, [3]Department of Plant Production and Technologies, Faculty of Agricultural Science and Technologies, Sivas University of Science and Technology, Sivas, Turkey, [4]Center of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan, [5]Department of Information Systems and Technologies, Bilkent University, Ankara, Turkey, [6]Department of Field Crops, Faculty of Agriculture, University of Çukurova, Adana, Turkey, [7]Department of Field Crops, Faculty of Agriculture, Bolu Abant Izzet Baysal University, Bolu, Turkey, [8]International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India, [9]Department of Plant Resources and Environment, Jeju National University, Jeju, South Korea

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Mohsen Hesami,
University of Guelph, Canada
Muhammad Shoaib,
Bahauddin Zakariya University, Pakistan
Adil Koray Yildiz,
Bozok University, Turkey

***Correspondence:**
Faheem Shehzad Baloch
balochfaheem13@gmail.com
Yong Suk Chung
yschung@jejunu.ac.kr

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

Received: *16 March 2022*
Accepted: *03 June 2022*
Published: *24 August 2022*

Citation:
Aasim M, Katirci R, Baloch FS, Mustafa Z, Bakhsh A, Nadeem MA, Ali SA, Hatipoğlu R, Çiftçi V, Habyarimana E, Karaköy T and Chung YS (2022) Innovation in the Breeding of Common Bean Through a Combined Approach of in vitro Regeneration and Machine Learning Algorithms. Front. Genet. 13:897696. doi: 10.3389/fgene.2022.897696

Common bean is considered a recalcitrant crop for *in vitro* regeneration and needs a repeatable and efficient *in vitro* regeneration protocol for its improvement through biotechnological approaches. In this study, the establishment of efficient and reproducible *in vitro* regeneration followed by predicting and optimizing through machine learning (ML) models, such as artificial neural network algorithms, was performed. Mature embryos of common bean were pretreated with 5, 10, and 20 mg/L benzylaminopurine (BAP) for 20 days followed by isolation of plumular apice for *in vitro* regeneration and cultured on a post-treatment medium containing 0.25, 0.50, 1.0, and 1.50 mg/L BAP for 8 weeks. Plumular apice explants pretreated with 20 mg/L BAP exerted a negative impact and resulted in minimum shoot regeneration frequency and shoot count, but produced longer shoots. All output variables (shoot regeneration frequency, shoot counts, and shoot length) increased significantly with the enhancement of BAP concentration in the post-treatment medium. Interaction of the pretreatment × post-treatment medium revealed the need for a specific combination for inducing a high shoot regeneration frequency. Higher shoot count and shoot length were achieved from the interaction of 5 mg/L BAP × 1.00 mg/L BAP followed by 10 mg/L BAP × 1.50 mg/L BAP and 20 mg/L BAP × 1.50 mg/L BAP. The evaluation of data through ML models revealed that R^2 values ranged from 0.32 to 0.58 (regeneration), 0.01 to 0.22 (shoot counts), and 0.18 to 0.48 (shoot length). On the other hand, the mean squared error values ranged from 0.0596 to 0.0965 for shoot regeneration, 0.0327 to 0.0412 for shoot count, and 0.0258 to 0.0404 for shoot length from all ML models. Among the utilized models, the multilayer perceptron model provided a better prediction and optimization for all output variables, compared to other models. The achieved results can be employed for the prediction and

optimization of plant tissue culture protocols used for biotechnological approaches in a breeding program of common beans.

Keywords: machine learning algorithms, artificial neural network, *in vitro* regeneration, plumular apices, coefficient of determination, mean squared error

INTRODUCTION

Grain legumes are an important pillar of the agricultural system, are considered a vital source of high-quality protein for food and fodder, and play a significant role in sustainable cropping systems (Vanlauwe et al., 2019). Common bean (*Phaseolus vulgaris L.*) is an important grain legume crop and is mostly used worldwide for its pods and palatable seeds (Nadeem et al., 2020a). Common bean contains good concentrations of high-quality protein, minerals particularly zinc and iron, vitamins, and antioxidants and is considered a "grain of hope" for the impoverished community. Common bean was originated in Mesoamerica (Bitocchi et al., 2012) and its domestication in Andean and Mesoamerican regions resulted in the formation of two unique gene pools: Andean gene pool and Mesoamerican gene pool (Kami et al., 1995; Mamidi et al., 2013; Asfaw et al., 2017; Blair et al., 2018; Campa et al., 2018). Common bean is considered one of the most varied legume crops by reflecting variations in its growth habit, plant height, pods, maturity, seed weight and size (Yeken et al., 2019; Nadeem et al., 2020b).

Climate change is becoming a serious threat to agriculture, and various biotic (pathogens and insects) or abiotic (drought and edaphic) factors are contributing significantly to the global common bean production loss (Castillo et al., 2015). Keeping these in view, scientists are trying to develop climate-resilient common bean cultivars having improved agronomic and nutritional traits. The mentioned target can be achieved by the application of modern biotechnological techniques and for that reason, optimization of the *in vitro* plant tissue culture technique for whole plant regeneration is highly demanding. To date, a reasonable number of *in vitro* regeneration protocols have been established and documented. *In vitro* regeneration of common bean is an arduous task due to its recalcitrant nature, genotype dependence, lack of reproducibility, low shoot counts with stunted growth, rooting, and acclimatization. Hence, there is always a need to develop a new, efficient, and repeatable protocol for the application of biotechnological techniques to produce elite cultivars, especially for recalcitrant crops (Aasim et al., 2013). To achieve the objective, selection of potent explants with a high regeneration protocol is highly significant. Considering this, a novel explant "plumular apices" and an *in vitro* regeneration protocol of pretreatment of explants with high benzylaminopurine (BAP) concentration was employed for common bean. Pretreatment is the process of treating seeds or explants with variable stimulants like cytokinins at low to high doses for a certain period, followed by culturing the explants on a post-treatment medium, supplemented with low plant growth regulators (PGRs) or without any PGRs (Özkan and Aasim, 2019).

Conventional plant breeding methodologies include the assessment and classification of genetic diversity, yield component analysis, yield stability analysis, enhanced tolerance to stresses, and hybrid breeding programs. On the other hand, *in vitro* micropropagation, doubled haploid production, artificial polyploidy induction and *Agrobacterium*-mediated gene transformation techniques are considered *in vitro*-based biotechnological breeding methodologies (Niazian and Niedbala, 2020). In plant tissue culture studies, the impacts of input (uni or multi) factors on the regeneration potential (outputs) of desired plants are studied. In general, classical statistical techniques have been employed for analyzing and interpreting the output variables. These techniques are generally based on variance analysis and linear regression models for estimating the correlation between input (independent) and output (dependent) variables. Although these approaches are highly effective, lack of efficacy of complex and nonlinear inputs (Hesami and Jones, 2021; Earl et al., 2021) and high probability (Abbasi et al., 2016; Jamshidi et al., 2016; Farhadi et al., 2020) are the major concerns in plant tissue culture studies due to the sensitivity. These types of issues can be overcome by modern high throughput technologies like machine learning (ML) and artificial neural network (ANN) models for testing and optimizing the output variables concerning the input parameters. Although the application of ML and ANN models in plant sciences specifically in the area of plant tissue culture is in its early stages, it is successfully documented for different aspects of plant tissue culture ranging from *in vitro* sterilization to *in vitro* regeneration and from *in vitro* callogenesis to secondary metabolite production (Hesami et al., 2020a; García-Pérez et al., 2020; Hameg et al., 2020; Hesami & Jones, 2020; Niazian and Niedbala, 2020; Pepe et al., 2021; Salehi et al., 2021; Aasim et al., 2022). In these studies, researchers employed different ML algorithms, and the selection of specific ML models is generally based on the expertise and target set in the study. These data-driven models are highly efficient to parse and interpret different types of datasets (non-normal, nonlinear, and nondeterministic unpredictable data) by using all spectral data along with avoiding irrelevant spectral bands and multicollinearity (Salehi et al., 2020). In this study, an *in vitro* regeneration protocol of common bean was established using novel plumular apice explants. The results regarding output variables were analyzed and interpreted, and input variables were predicted by response surface methodology (RSM). In addition, the results for the output variables were validated using different ML algorithms (support vector regression—SVR, Gaussian process regression—GPR, XGBoost regression—XGBoost, and random forest regression—RF) and an ANN-based multilayer perceptron (MLP) regression model. The performance was evaluated by tabulating the R^2 and the mean squared error

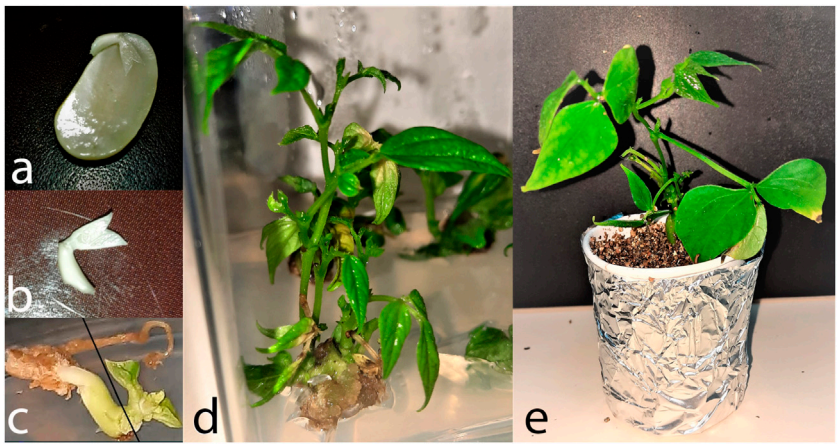


FIGURE 1 | *In vitro* regeneration and rooting of common bean Cv. Karacaşehir 90 **(A)** sterilized seed with the intact embryo, **(B)** isolated embryo ready for inoculating on the pretreatment medium, **(C)** pretreated mature embryo used for isolating the plumular apice explant, **(D)** multiple shoot induction from the plumular apice explant, and **(E)** acclimatized plant in a pot containing vermiculite.

(MSE) metric values for each model (Hesami et al., 2019; Kirtis et al., 2022). The results achieved in this study will open a new window to evaluate the efficiency of the plant tissue culture protocols that are predominantly developed for breeding purposes.

MATERIALS AND METHODS

In vitro Regeneration

The commercial common bean cultivar "Karacaşehir-90" was selected for this study as the plant material. Manually selected uniform seeds were surface sterilized with 3.5% (w/v) NaOCl for 15 min. Thereafter, seeds were continuously rinsed with sterilized dH_2O water for 5–7 min and this process was repeated thrice to remove the traces of NaOCl. Seeds (**Figure 1A**) were awaited for 24 h in dH_2O, followed by isolation of mature embryos (**Figure 1B**) under aseptic conditions. A two-step experiment was designed for this research. At first, mature embryos isolated from sterilized seeds were inoculated on MS (Murashige and Skoog, 1962) media supplemented with 5, 10, and 20 mg/L BAP (pretreatment medium) for 20 days. In the second step, plumular apice explants (**Figure 1C**) were carefully isolated from pretreated mature explants, followed by inoculation on MS media supplemented with low BAP (0.25, 0.50, 1.00, and 1.50 mg L^{-1}) concentrations (post-treatment medium). The explants were cultured for 8 weeks on the post-treatment medium. Four different concentrations (0.25, 0.50, 1.0, and 1.50 mg L^{-1}) of indole-3-butyric acid (IBA) were used for *in vitro* rooting. For acclimatization, rooted plantlets were transferred to pots filled with vermiculite, wrapped in a polyethylene bag, and placed in the growth room.

The basal media used for pretreatment, post-treatment, and rooting were prepared by adding MS (4.4 g/L), commercial sugar (30 g/L), and polyvinyl proline (25 mg L^{-1}). The pH of all media was adjusted to ~5.8 with the aid of 1N HCl or 1N NaOH. The medium was gelled with agar (6.5 g/L) and autoclaved at 121°C for 20 min. All experiments were carried out in the growth room at 24 ± 2°C and 16-h light photoperiod, equipped with white light-emitting diodes at approximately 2000 LUX. All chemicals used in this study were procured from Duchefa (MS, BAP, IBA, and agar) and Sigma-Aldrich (polyvinyl proline).

Response Surface Methodology

The RSM approach was used to model and optimize the selected responses to changing variables and graphical representation of the results. RSM generates continuous multivariable predictions represented as quadratic surfaces, allowing the prediction of optimal values in three-dimensional space. Pretreatment, post-treatment, and their interactive effect values were used as input variables. On the other hand, regeneration frequency (%), shoot count, and shoot length (cm) were used for the response surface calculations. The degree of predicted mathematical model compliance to obtained values was expressed as R^2 fit values. All RSM data analyses, such as analysis of variance, regression, and generation of quadratic polynomial surface equations, graphics and optimal value predictions, were conducted using Minitab v20.4 statistical software.

Modeling Procedures

In this study, interactions of pretreatment (5, 10, and 20 mg/L BAP) × post-treatment BAP doses (0.25, 0.50, 1.0, and 1.50 mg/L) were used as input variables, whereas, *in vitro* regeneration frequency, shoot count, and shoot length were measured as the output variables. ML algorithms of SVR (Hesami et al., 2020b; Katirci and Takci, 2021), GPR (Hu et al., 2019), XGBoost (Chen and Guestrin, 2016), RF (Aggarwal, 2018), and MLP neural network (Silva et al., 2019) were utilized to train and test the model. The performance of the model was assessed using leave-one-out cross-validation (Sammut and Webb, 2011). The hyperparameters of the ML models were optimized using the grid search technique to find the best

model. The open-source Python language (Van Rossum and Drake, 1994) was used to code algorithms using the sklearn library (Pedregosa et al., 2011). MLP, SVR, GP, XGBoost, and RF algorithms were used to predict the outputs. The model performance was evaluated by calculating R^2 (coefficient of determination) and MSE values (Hesami et al., 2019), which are presented in **Eqs. 1 and 2**.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \tilde{Y}\right)^2} \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (2)$$

Y_i represents the measured values, $\hat{Y}_i$ indicates the predicted values, $\tilde{Y}$ denotes the mean of the measured values, and n is the count of samples.

The dataset that is used in plant tissue culture studies may not be linear; hence, nonlinear regression models are essential like the SVR model (Drucker, 1997) as expressed in **Eq. 3**. In SVR, the output y is a real number and it can be used for nonlinear variables.

$$y = w\varphi(x) + b \quad (3)$$

In the above equation, b depicts the bias, w represents the weight, and the elevated features space is presented as $\varphi(x)$, which defines the nonlinearity of input x. In the SVR model, the predicted variable is placed between the upper and lower limit values to minimize the risk. In case the data exceeds these limits, it is set between these values (Smola & Schölkopf, 2004). The kernels of "linear", "poly", "radial basis function (rbf)", "sigmoid", and "precomputed" are present in the SVR model. Among these, the RBF kernel is the most widely used.

The GPR model is another nonparametric supervised learning method that is used mainly to perform Bayesian nonlinear regression and classification tasks. It is a powerful ML algorithm that uses the Gaussian probability density function. The GPR works efficiently with a small dataset, with more accuracy, ease of calculation, and consistency.

The approach is presented in **Eq. 4** for each input x and output y produced by this function.

$$y_i = f(x_i) + \varepsilon \quad (4)$$

Extreme gradient boosting (XGBoost) is a decision-tree-based ensemble ML algorithm that uses a gradient boosting framework that can be used for both regression and classification problems (Chen and Guestrin, 2016). In ML, ensemble learning algorithms combine multiple ML algorithms to obtain a better model. The XGBoost model generates the regression or classification trees by taking previous trees and factoring in their predictions to create a new tree to decrease prediction error. **Eq. 5** indicates the XGBoost objective function and **Eq. 6** shows the model of XGBoost at iteration j that needs to be minimized.

$$Y_i = F(x_i) = \sum_{d=1}^{D} f_d(x_i), f_d \in F, i = 1, \ldots, n \quad (5)$$

$$L_j = \sum_{(i=1)}^{n} l\left(y_i, \hat{y}_i^{(j-1)} + f_j(x_i)\right) + \Omega\left(f_j\right) \quad (6)$$

where l is a differentiable convex loss function that measures the dissimilarity between the prediction $\hat{y}_i$ and the target y_i. The term Ω penalizes the complexity of the model and it also helps to smooth out the final learned weights to avoid overfitting.

The RF model is an alternative supervised ensemble learning method based on the decision trees (Breiman, 2001), which can also be implemented for regression and classification problems. It is one of the most widely used ML models due to its simplicity in design, high efficiency, less susceptibility to overfitting, handling the noise, and ability to manage a large number of features. The forest is generated by multiple decision trees and each tree possesses the same distribution. The MSE metric is used to solve the regression models. It determines the distance between the nodes to define which branch is better for the forest. The following **Eq. 7** describes this concept (Pavlov et al., 2019).

$$y = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)k(x, x_i) + b \quad (7)$$

where y is the value of the data point and n is the number of samples.

The MLP is the most well-known ANN model that consists of more than one perceptron, which includes a nonlinear activation function. MLP is a supervised learning method containing one or more hidden layers. The training continues until the following equation is minimized.

$$E = \frac{1}{K}\sum_{k=1}^{K}\left(y_k - \hat{y}_k\right)^2 \quad (8)$$

where y_k and $\hat{y}_k$ are observed and predicted data points, respectively.

Generating an MLP structure is the most important part that significantly influences the performance of the model. It is a prerequisite to defining the number of neurons in each layer and the number of hidden layers during the construction of the model. MLP is often applied to supervised learning problems. The backpropagation method is implemented to tune the weights and biases of the layers (Hesami et al., 2019).

RESULTS

In vitro Regeneration of Common Bean

Mature embryo explants exposed to 5, 10, and 20 mg L^{-1} BAP for 20 days resulted in enhanced embryo size of approximately 60–70% explants, which allowed to isolate plumular apice explants easily under sterile conditions (**Figure 1C**). Thereafter, explants were inoculated on a post-treatment medium, which resulted in multiple shoot induction within 2–3 weeks along with callus induction from the basal end of some explants. The explants were cultured in the growth room for 8 weeks to induce multiple shoots (**Figure 1D**). The analysis of variance exhibited the variable response of input variables (pretreatment, post-treatment, and pretreatment × post-treatment) on *in vitro* regeneration of common bean. Results

TABLE 1 | Analysis of variance of output variables of common bean.

Treatment	Output variables	*p*-value
Pretreatment	Regeneration (%)	0.000[a]
	Shoot counts	0.234
	Shoot length (cm)	0.013*
Post-treatment	Regeneration (%)	0.421
	Shoot counts	0.329
	Shoot length (cm)	0.000[a]
Pretreatment × post-treatment	Regeneration (%)	0.562
	Shoot counts	0.682
	Shoot length (cm)	0.021*

[a]p< 0.01 *and* *p< 0.05.

revealed the significant impact of pretreatment on the regeneration frequency ($p < 0.01$) and shoot length ($p < 0.05$). Results on post-treatment ($p < 0.01$) application of BAP and combination of pretreatment × post-treatment × post-treatment ($p < 0.01$) revealed a significant impact only on the shoot length. On the other hand, shoot counts have remained insignificant to all input variables (**Table 1**).

Results revealed that elevated pretreatment concentrations negatively affected the regeneration frequency and shoot counts, which ranged from 47.91 to 96.51% and 2.99 to 3.60, respectively. In contrast, the mean shoot length increased with elevated pretreatment concentration and ranged from 1.07 to 1.57 cm (**Supplementary Table S1**). Results of post-treatment revealed a better regeneration frequency at high BAP concentrations ranging from 66.67 to 72.22%. In a similar manner, shoot length also exhibited an increase with a respective increase in BAP concentration. On the other hand, the variable impact of the post-treatment medium (BAP) was observed on shoot counts that ranged from 3.06 to 3.92 (**Supplementary Table S1**). The results on pretreatment × post-treatment exhibited the negative impact of elevated BAP concentration (pretreatment) on shoot regeneration frequency (%) that ranged from 58.33 to 66.67% (10 m g/L BAP) and 41.67 to 58.33% (20 m g/L BAP). However, exposing explants to 5 mg/L BAP resulted in up to 100% regeneration (**Figure 2A**). The results on shoot counts and shoot length showed the variable impact of pretreatment × post-treatment concentrations (**Figures 2B and C**). Outcomes revealed that maximum shoot counts were linked with pretreatment × post-treatment concentrations. The maximum shoot counts were obtained for 10 mg/L × 1.50 mg/L BAP (5.0 shoots), 5 mg/L × 0.50 mg/L BAP (4.67 shoots), and 20 mg/L × 1.50 mg/L BAP (3.33 shoots) (**Figure 2B**). A similar pattern was also observed with the shoot length. The longest shoots were recorded as 1.15 cm ($5\ mg\ L^{-1} \times 1.00\ mg\ L^{-1}$), 1.40 cm ($10\ mg\ L^{-1} \times 1.50\ mg\ L^{-1}$), and 1.79 cm ($20\ mg\ L^{-1} \times 1.50\ mg\ L^{-1}$) (**Supplementary Table S2**). Results revealed that exposing explants to a high BAP concentration (both pretreatment and post-treatment) medium yielded relatively longer shoots (**Figure 2C**) compared to other combinations. In this study, contour plots were also constructed for a better presentation and understanding of the data. In the contour plots, the data were distributed into different subgroups, emphasized with different colors. Contour plots help to find out the best combination for the desired output value. Results of contour plots revealed the optimization of <90% regeneration frequency (**Figure 3A**), 4.0 shoots per explant (**Figure 3B**), and 2.4 cm longer shoots (**Figure 3C**) and presented the doses of the pretreatment and post-treatment medium.

In vitro regenerated shoots inoculated on the rooting medium yielded a relatively high rooting frequency. Although most of the plants were rooted within the first 3–4 weeks, they were kept in the rooting medium for a total of 6 weeks before shifting to pots for acclimatization. The survival rate of rooted plantlets in pots was relatively less than expected (**Figure 1E**). The results revealed that the protocol can be used for *in vitro* regeneration of common bean.

Response Surface Regression Models

The experiment design of the study was based on pretreatment doses and post-treatment doses, followed by selecting the best mathematical model. The regression equations (**Eq. 9–11**) for the response variables [R^2 (measured), R^2 (Adj.), and R^2 (pred.)] were used and their respective values are presented in **Table 2**.

$$\begin{aligned}\textbf{Regeneration frequency} = {} & 152.2 - 13.25\ \text{pre} + 3.9\ \text{post} \\ & + 0.375\ \text{pre*pre} - 3.7\ \text{post*post} \\ & + 0.67\ \text{pre*post}\end{aligned} \tag{9}$$

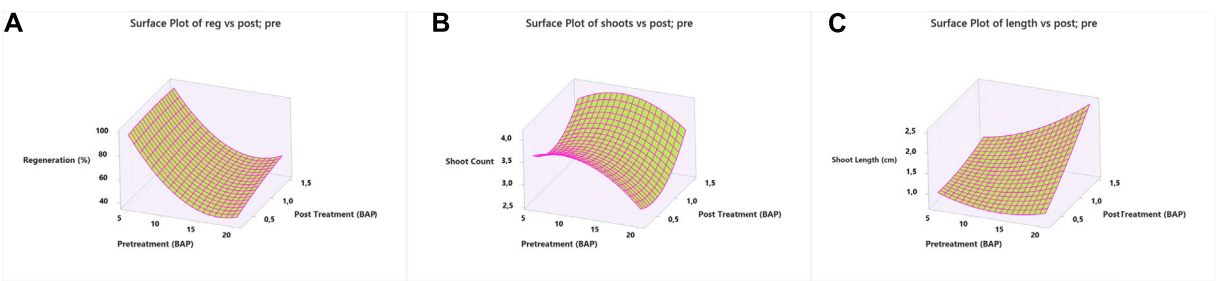


FIGURE 2 | 3D response surface plots of *in vitro* regeneration of common bean **(A)** regeneration, **(B)** shoot count, and **(C)** shoot length.

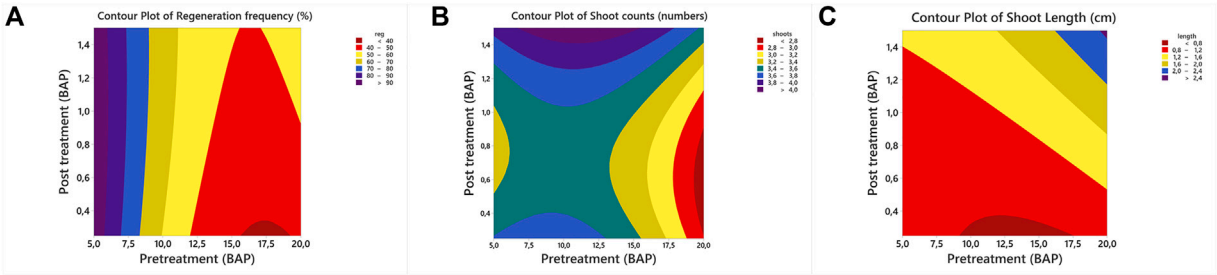


FIGURE 3 | Contour plots of *in vitro* regeneration of common bean **(A)** regeneration, **(B)** shoot count, and **(C)** shoot length.

TABLE 2 | Response surface regression models for *in vitro* regeneration of common bean.

Output variables	R^2	R^2 adj	R^2 pred
Regeneration (%)	56.27	48.98	38.68
Shoot counts	13.74	0.00	0.00
Shoot Length (cm)	47.84	39.14	24.34

$$\text{Shoot Counts} = 3.52 + 0.124\,\text{pre} - 1.74\,\text{post} - 0.00737\,\text{pre*pre} + 1.04\,\text{post*post} + 0.0263\,\text{pre*post} \quad (10)$$

$$\text{Shoot Length} = 1.511 - 0.117\,\text{pre} - 0.50\,\text{post} + 0.00372\,\text{pre*pre} + 0.195\,\text{post*post} + 0.0710\,\text{pre*post} \quad (11)$$

The results of the regression model depicted more R^2 (measured) compared to R^2 (Adj.) and R^2 (pred.) for all tested output variables used in this study. The R^2 values for regeneration were recorded as 56.27 (R^2 measured), 48.98 (R^2 adj.), and 38.68 (R^2 pred.). R^2 for shoot length was recorded as 47.84 (R^2 measured), 39.14 (R^2 adj), and 24.34 (R^2 pred.). The results illustrated that the regression models efficiently presented the data. In contrast, a low R^2 (measured) value of 13.74 with zero values for both R^2 (adj) and R^2 (pred.) was attributed to the shoot counts. The comparatively low R^2 values for shoot counts reflect that the regression model did not find the association between input and output variables. Overall results illustrated a better impact of pretreatment and post-treatment systems on shoot regeneration and shoot length as compared to shoot counts.

Response Prediction of Output Variables

The computations of predicted values for all output variables of *in vitro* regeneration of common beans were also performed by solving the reference equations (**Eq. 7–9**) to predict the impact of input variables on output variables. Results indicated a variable combination of pretreatment and post-treatment dose (BAP) for inducing maximum output values. Moreover, results also showed that a combination of 5 mg/L BAP × 0.995 mg/L BAP may yield 98.88% regeneration frequency (%) (**Figure 4A**). The maximum predicted shoot count (4.142 shoots) was attributed to the 11.061 mg/L BAP × 1.50 mg/L BAP combination (**Figure 4B**). However, the maximum predicted shoot length of 2.447 cm was

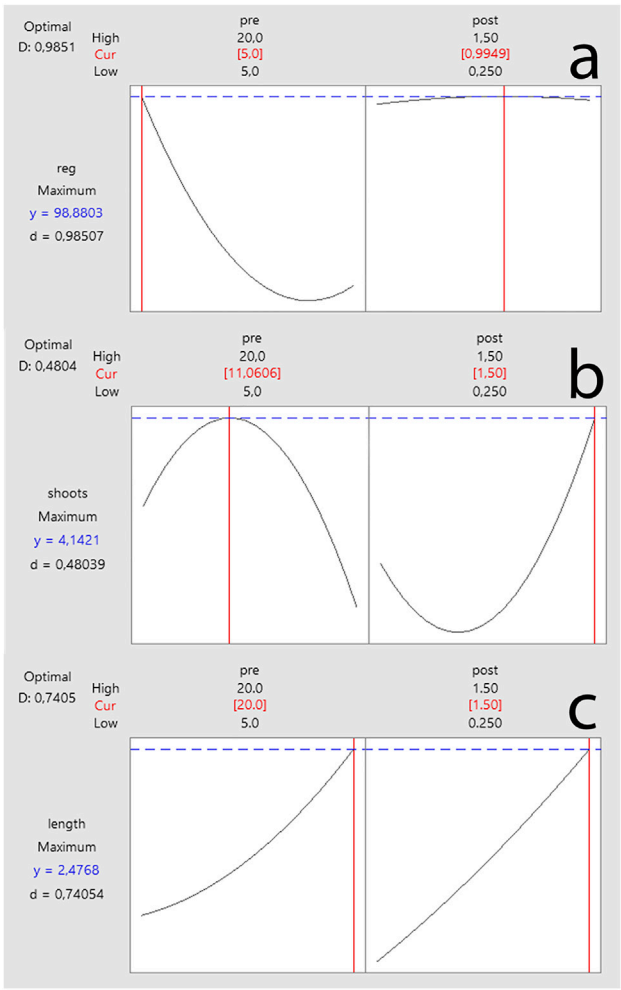


FIGURE 4 | Response prediction of individual output variables on *in vitro* regeneration of common bean **(A)** regeneration, **(B)** shoot count, and **(C)** shoot length.

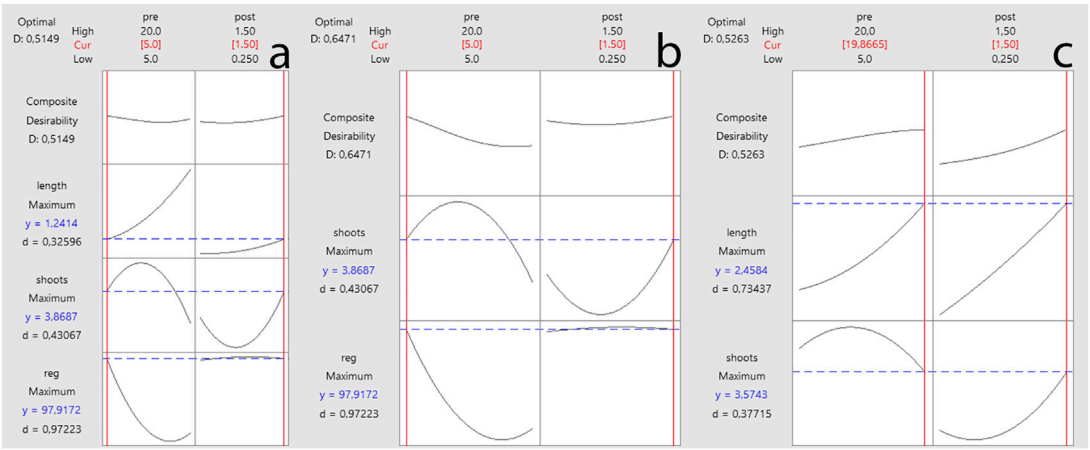


FIGURE 5 | Multiple response prediction of output variables on *in vitro* regeneration of common bean **(A)** regeneration × shoots × length, **(B)** regeneration × shoots, and **(C)** shoots × length.

TABLE 3 | Validity of the models.

	Shoot count		Shoot length		Regeneration	
	R^2	MSE	R^2	MSE	R^2	MSE
MLP	0.22	0.0327	0.48	0.0258	0.58	0.0596
SVR	0.11	0.0371	0.23	0.0377	0.44	0.0803
GP	0.06	0.0392	0.35	0.0318	0.49	0.0724
XGB	0.01	0.0412	0.18	0.0404	0.32	0.0965
RF	0.09	0.0380	0.25	0.0367	0.44	0.0803

attributed to the 20 mg/L BAP × 1.50 mg/L BAP combination (**Figure 4C**). The predicted values by response prediction models were close to the results attained in this study. The achieved results can be confirmed by checking the contour plots for shoot regeneration frequency (**Figure 3A**), shoot counts (**Figure 3B**), and shoot length (**Figure 3C**).

In addition to response prediction of individual output variables, the prediction response of multiple output variables was also constructed to optimize the dose concentration by considering two (reg × shoots and shoots × length) or considering all output variables (reg × shoot counts × shoot length). Results revealed that the combination of 5 mg/L × 1.5 mg/L BAP can be used for reg × shoot counts × shoot length (**Figure 5A**) and regeneration × shoot counts (**Figure 5B**). In contrast, 19.867 mg/L BAP × 1.5 mg/L BAP combination was predicted for shoot counts × shoot length variables (**Figure 5C**).

Machine Learning Algorithms

The R^2 and MSE performance metrics were used to predict the shoot count, shoot length, and regeneration (**Table 3**). Results exhibited variable R^2 values of all output variables for all models. The MSE values indicate the error between the measured and predicted values and varied with each model. Comparison between the five models revealed the better performance of MLP with high R^2 values for all outputs as compared to other models. However, the values differed for each output variable. Results further illustrated the clear relationship between the R^2 and the MSE values. In general, high R^2 values with low MSE values were recorded for all models. Results on shoot regeneration revealed the order of MLP (0.58 R^2; 0.0596 MSE) > GP (0.49 R^2; 0.0724 MSE) > SVR (0.44 R^2; 0.0802 MSE) > RF (0.44 R^2; 0.0803 MSE) > XGBoost (0.32 R^2; 0.0966 MSE). Results on shoot counts were computed in order of MLP (0.22 R^2; 0.0327 MSE) > SVR (0.11 R^2; 0.0371 MSE) > RF (0.09 R^2; 0.0380 MSE) > GP (0.06 R^2; 0.0392 MSE) > XGBoost (0.01 R^2; 0.0412 MSE). The performance of models on shoot count revealed the order of MLP (0.48 R^2; 0.0258 MSE) > GP (0.35 R^2; 0.0318 MSE) > RF (0.25 R^2; 0.0367 MSE) > SVR (0.23 R^2; 0.0377 MSE) > XGBoost (0.18 R^2; 0.0404 MSE). **Figure 6** presents the difference between the predicted and measured values. The horizontal axis refers to the samples while the vertical axis specifies the data collected from the models and the experimental study. The compatibility of the experimental results revealed the better performance of the MLP model for shoot regeneration, shoot counts, and shoot length (**Figure 6**). On the contrary, the XGBoost model exhibited the least compatibility between actual and predicted values.

A data visualization method was used with colors to indicate the relationship between two variables. In the heatmap, it was detected that there is a strong correlation between BAP and shoot length. The overall results displayed a negative correlation for regeneration (−0.67) and shoot counts (−0.25) with BAP (pretreatment) and a positive correlation between BAP (pretreatment) and shoot length (0.34). Results on BAP (post-treatment) revealed a positive correlation with all output variables. On the other hand, a negative correlation between regeneration and shoot length (−0.21) and a positive correlation between regeneration and shoot counts (0.49) were also observed (**Figure 7**). These results indicated the dependence of input factors on output variables.

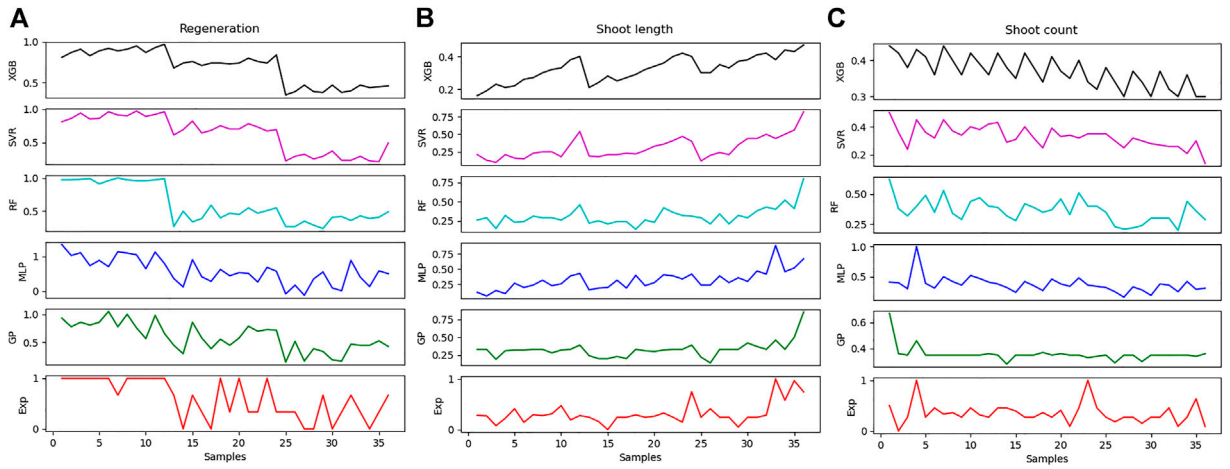


FIGURE 6 | The relationship between the prediction and actual values for **(A)** regeneration, **(B)** shoot length, and **(C)** shoot count.

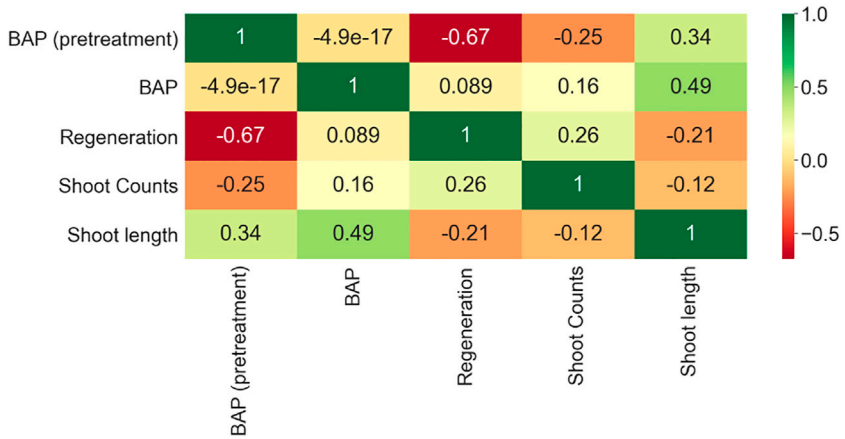


FIGURE 7 | Correlation matrix of inputs and outputs for common bean.

DISCUSSION

The selection of proper explants is a prerequisite for establishing an *in vitro* regeneration protocol, especially for recalcitrant plants like edible legumes. The selection of nontraditional and novel explants is one of the possible and potential solutions to overcome the recalcitrant issue in plant tissue cultures (Wang et al., 2011). Plumular apice are a potent and highly efficient explant due to the presence of meristem. To date, researchers tested plumule or plumular apice explant for *in vitro* shoot regeneration of different edible legumes like pea (Molnár et al., 1999), chickpea (Aasim et al., 2013), peanut (Singh and Hazra, 2009; Day and Aasim, 2017), cowpea (Aasim et al., 2009), pigeon pea (Surekha et al., 2005), and lentils (Aasim, 2012). In all these studies, plumular apices were proved to be efficient for inducing high regeneration frequency with high shoot counts per explant. However, the major problem associated with the use of this potent explant is the isolation from the embryo without any damage due to its smaller size. Pretreatment or pulse treatment of mature (Aasim et al., 2009; Day and Aasim, 2017) or immature embryos (Aasim, 2012) with high cytokinin concentration significantly enhances the embryo size which in turn allows isolation of plumular apice explant properly without any damage (Özkan and Aasim, 2019).

Pretreatment of explants with a high dose of cytokinins or auxins for a certain period is more effective to induce more shoots and more rapid regeneration (Brar et al., 1999; Barpete et al., 2014; Kumari et al., 2017; Özkan and Aasim, 2019) due to the more active division of cells (meristematic cells) found in the explants. However, stunted shoots, heavy callus induction, and deformed shoots (vitrified or hyperhidric shoots) are some of the common and negative features associated with the pretreatment (Aasim et al., 2011a). The pretreatment and post-treatment medium, treatment time, plant, and explants (Day and Aasim, 2017) are some of the factors that regulate the whole regeneration process. The manipulation of triggers (inputs), epigenetic and transcriptional cellular responses to the triggers, and molecules

stem cell niche (Sugimoto et al., 2019) lead to nondeterministic and nonlinear developmental patterns in the plant's cells and tissues (Prasad and Gupta, 2008).

In vitro regeneration is the mainstay of *in vitro*-based breeding methods, and optimization of input variables for the application of modern biotechnological techniques is highly demanding in the modern era of genome editing. The final output is generally analyzed and interpreted by traditional statistical software programs with the aid of tests like least significant difference test, Duncan's multiple range test, Tukey's honestly significant difference test, etc. (Ayuso et al., 2019). These models are not sufficient for the exact prediction of input combinations for the desired output variables. In recent years, modern computer-based software and models have been documented for the exact prediction and validation of the results. The prediction methodologies are divided into three major groups: regression equations, mathematical equations, and computer-based software (Askari et al., 2021). Among these, computer-based software models and simulation programs are gaining popularity with high acceptability by researchers to predict data with more accuracy (Kirtis et al., 2022).

RSM is a computer-based model, used for optimizing and predicting output variables using more than two input variables (Abbasi et al., 2016; Managamuri et al., 2019; Askari et al., 2021; Slimani et al., 2021). The advantage of using contour plots is the distribution of attained results into different subunits, which enables to specify the input variables for the desired output variable (Aasim et al., 2022). RSM predicted the optimal pretreatment and post-treatment BAP concentrations for inducing maximum shoot regeneration frequency, shoot counts and shoot length by estimating the R^2 (measured), R^2 (Adj.), and R^2 (pred.) values of output variables. Furthermore, RSM successfully predicted the input variables by considering individual or multiple output variables. The use of surface plots and contour plots by RSM also clearly illustrated the impact of pretreatment and post-treatment doses of BAP on *in vitro* regeneration output variables of common bean. The use of RSM in plant or agricultural sciences is limited. However, RSM has been employed successfully for predicting the optimal conditions for *in vitro* regeneration and secondary metabolite production of different plants (Bansal et al., 2017; Premkumar et al., 2020; Slimani et al., 2021).

Results on pretreatment dose revealed a negative impact on shoot regeneration frequency and shoot count. Investigation of previous studies on pretreatment with cytokinin discerned the variable impact on shoot regeneration with both positive and negative impacts depending on the genotype, cytokinin type, and concentration. The study on peanuts revealed 100% shoot regeneration with more shoot counts from plumular apices preconditioned with 20 mg L^{-1} BAP as compared to 10 mg L^{-1} BAP (Day and Aasim, 2017). In a similar manner, other studies on cowpea (Brar et al., 1999), *Pongamia pinnata* (Belide et al., 2010), and *Sophora tonkinensis* (Jana et al., 2013) also illustrated the positive impact of pretreatment with cytokinin on shoot regeneration. On the other hand, relatively low regeneration from preconditioned explants has also been documented in peanuts (Akasaka et al., 2000; Matand et al., 2013). The results revealed a decreased shoot count pattern with enhanced pretreatment concentration of BAP. On the contrary, a high concentration of pretreatment dose yielded longer shoots. Results illustrated that shoot counts and shoot length are associated with BAP concentration and other factors like genotype. A previous study on lentils using preconditioned plumular apices explants yielded relatively more shoot counts and shoot length compared to nonconditioned plumular apices explants (Aasim, 2012).

A post-treatment medium enriched with low cytokinin concentration is highly significant and regulates the *in vitro* regeneration from pretreated explants. Results revealed high shoot regeneration frequency, shoot counts, and shoot length from pretreatment and post-treatment of BAP and confirmed the results achieved in chickpea (Aasim et al., 2011b) and lentils (Aasim, 2012). However, the investigation of previous studies clearly illustrated the significance of the correlation between PGRs type and concentration of both pretreatment and post-treatment medium, explant, and genotype (Tang et al., 2012; Kumari et al., 2017) on *in vitro* shoot regeneration. The results confirmed the significance of BAP concentration in the pretreatment and post-treatment medium on *in vitro* shoot count and shoot length of common bean. However, both parameters generated maximum output at a different combination of pretreatment × post-treatment BAP concentration. Previous studies on chickpea and lentils also exhibited a different combination of pretreatment × post-treatment BAP concentration. In chickpea, maximum shoot counts with shorter shoots were associated with high BAP in the post-treatment medium (Aasim et al., 2013). Vice versa, minimum shoot counts with longer shoots of lentils were documented from low BAP concentration in the post-treatment medium (Aasim, 2012). Shoot length is another important factor and maximum shoot length was documented at the high pretreatment × post-treatment combination used in this study. The results are contrary to the findings in peanuts, where shoot length gradually decreased with elevated BAP concentration in the post-treatment medium (Day and Aasim, 2017). Overall results revealed that pretreatment and post-treatment doses of BAP exerted a clear impact on *in vitro* shoot regeneration of common beans as confirmed in other studies (Jahan et al., 2011; Özkan and Aasim, 2019).

In vitro rooting of *in vitro* regenerated shoot is an important step to establishing a successful *in vitro* regeneration protocol for recalcitrant plants. The availability of higher cytokinin concentration in the culture medium is generally supposed to be inhibitive for inducing *in vitro* rooting. Previous studies on the use of pretreatment or post-treatment medium in other crops revealed no negative impact on *in vitro* rooting (Aasim et al., 2009; Day and Aasim, 2017; Özkan and Aasim, 2020), and this study also support their findings and achieved 100% rooting. After successful rooting, rooted plants transferred to pots failed to adopt and a very low frequency of plants survived possibly due to awaiting plants in the rooting medium for a long time, which resulted in damaged roots and ultimately affected the survival percentage.

In recent years, ML and ANN models have been successfully employed in plant tissue culture studies for optimizing different

input variables like a basal medium (Alanagh et al., 2014; Arab et al., 2016; Arab et al., 2018), PGR types, concentration for *in vitro* regeneration (Kirtis et al., 2022), somatic embryogenesis (Niazian et al., 2017), callogenesis (Niazian et al., 2018), *in vitro* sterilization (Hesami et al., 2019; Aasim et al., 2022), and *in vitro* induced double haploid production (Niazian and Shariatpanahi, 2020). The detailed investigation of these studies revealed the use of different performance metrics like R^2, MSE, RMSE, MAE, etc. to validate different ML and ANN models. In this study, five different ML models including the ANN model were used for optimizing and predicting the results. Results divulged the variable response of all tested models to the target output variable. The best model for all parameters was found to be MLP. However, the RF model ranked second for shoot regeneration and shoot count, and the GP model for shoot length. The results confirmed the previous findings by researchers in plant tissue culture studies. An investigation of ML models revealed that the prediction of the model is dependent on inputs, target outputs, and the type of model used (Hesami et al., 2019; Salehi et al., 2020, 2021; Kirtis et al., 2022).

The performance of all the tested models was validated by computing R^2 and MSE scores. Relatively high R^2 values for shoot regeneration and low R^2 values were recorded for the shoot length. A detailed investigation of ML models in plant tissue culture studies revealed the variable R^2 values for different output variables like R^2 = 0.94 (Hesami et al., 2019), R^2 = 0.56–0.85 (Salehi et al., 2020), and R^2 = 0.70 (Hesami and Jones, 2021) and 0.98–1.0 (Kirtis et al., 2022). The R^2 values obtained in this study are relatively less than those of the previous findings but still validated the results in an efficient way. A low R^2 does not reflect the poor performance of the experiment, but rather reflects the low compatibility between input and output variables. High R^2 values reflect the high compatibility between the input and output variables, and they are obtained when the difference between the mean of the measured values and the predicted values is bigger than the difference between the actual and predicted values. The single performance metric does not predict or validate the results accurately, and therefore more than two performance metrics are generally considered for ML modeling. MSE is another powerful performance metric that reflects the error between the actual and predicted values. High MSE values depict the high error and vice versa. The results on MSE values for all output variables exhibited very low values for all the tested models, which reflects the low error between the actual and predicted values (Kirtis et al., 2022).

CONCLUSION

The development of a successful *in vitro* regeneration protocol for the common bean is extremely crucial for the application of modern biotechnological techniques for its improvement. The developed protocol can be employed for the application of *in vitro* biotechnological techniques like genetic transformation and *in vitro* polyploidy induction for its enhancement. Application of ML and ANN models depicted better performance of the MLP model as compared to other models for better prediction and optimization of all output variables. The results achieved in this study proved that ML models are powerful tools to analyze the data and optimize the complex conditions irrespective of the variable inputs, outputs, and responses of models. The accomplished results can be effectively employed for the prediction and optimization of plant tissue culture protocols used for breeding purposes in the future.

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.

AUTHOR CONTRIBUTIONS

MA (Lab Research Coordination, Response Surface Methodology Analysis, and Article Writing), RK (ML and ANN Analysis and Article Writing), ZM (Response Surface Methodology), SA (Statistical Analysis and Graphical Presentation), MN and AB (Conducting Research and Article Writing), FB (Research Coordination, Article Writing, and Control), TK (Research Coordination), YC (Editing, Reviewing the Article, and Providing Funds), VC (Research Coordination), EH (Article Writing and Editing), and RH (Research coordination, Reviewing, and Editing).

FUNDING

Basic Science Research Program supported this research through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2019R1A6A1A11052070).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.897696/full#supplementary-material

REFERENCES

Aasim, M., Day, S., Rezaei, F., and Hajyzadeh, M. (2013). Multiple Shoot Regeneration of Plumular Apices of Chickpea. *Turk. J. Agric. For.* 37, 33–39. doi:10.3906/tar-1204-3810.3906/biy-1207-44

Aasim, M., Day, S., Rezael, F., Hajyzadeh, M., Mahmud, S. T., and Ozcan, S. (2011b). *In Vitro* shoot Regeneration from Preconditioned Explants of Chickpea (Cicer Arietinum L.) Cv. *Gokce. Afr. J. Biotechnol.* 10, 2020–2023. doi:10.5897/AJB11.3214

Aasim, M., Katırcı, R., Akgur, O., Yildirim, B., Mustafa, Z., Nadeem, M. A., et al. (2022). Machine Learning (ML) Algorithms and Artificial Neural Network for

Optimizing *In Vitro* Germination and Growth Indices of Industrial Hemp (*Cannabis Sativa* L.). *Industrial Crops Prod.* 181, 114801. doi:10.1016/j.indcrop.2022.114801

Aasim, M., Khawar, K. M., and Özcan, S. (2009). *In Vitro* micropropagation from Plumular Apices of Turkish Cowpea (Vigna Unguiculata L.) Cultivar Akkiz. *Sci. Hortic.* 122, 468–471. doi:10.1016/j.scienta.2009.05.023

Aasim, M. (2012). Micropropagation of Lentil (Lens Culinaris Medik.) Using Pulse Treatment of Immature Plumular Apices. *Pak. J. Agri. Sci.* 49, 149–154.

Aasim, M., Sahin-Demirbag, N., Khawar, M., Kendir, H., and Özcan, S. (2011a). Direct Axillary Shoot Regeneration from the Mature Seed Explant of the Hairy Vetch (Vicia Villosa Roth). *Arch. Biol. Sci. Belgra* 63, 757–762. doi:10.2298/abs1103757a

Abbasi, Z., Hooshyar, S., and Bagherieh-Najjar, M. B. (2016). Improvement of Callus Production and Shoot Regeneration Using Various Organs of Soybean (Glycine Max L. Merr) by Response Surface Methodology. *Vitro Cell.Dev.Biol.-Plant* 52, 537–545. doi:10.1007/s11627-016-9778-1

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*, 10. Switzerland AG: Springer, 978–983. doi:10.1007/978-3-319-94463-0

Akasaka, Y., Daimon, H., and Mii, M. (2000). Improved Plant Regeneration from Cultured Leaf Segments in Peanut (*Arachis hypogaea* L.) by Limited Exposure to Thidiazuron. *Plant Sci.* 156, 169–175. doi:10.1016/s0168-9452(00)00251-x

Alanagh, E. N., Garoosi, G.-a., Haddad, R., Maleki, S., Landín, M., and Gallego, P. P. (2014). Design of Tissue Culture Media for Efficient Prunus Rootstock Micropropagation Using Artificial Intelligence Models. *Plant Cell Tiss. Organ Cult.* 117, 349–359. doi:10.1007/s11240-014-0444-1

Arab, M. M., Yadollahi, A., Eftekhari, M., Ahmadi, H., Akbari, M., and Khorami, S. S. (2018). Modeling and Optimizing a New Culture Medium for *In Vitro* Rooting of G×N15 Prunus Rootstock Using Artificial Neural Network-Genetic Algorithm. *Sci. Rep.* 8, 9977–10018. doi:10.1038/s41598-018-27858-4

Arab, M. M., Yadollahi, A., Shojaeiyan, A., and Ahmadi, H. (2016). Artificial Neural Network Genetic Algorithm as Powerful Tool to Predict and Optimize *In Vitro* Proliferation Mineral Medium for G × N15 Rootstock. *Front. Plant Sci.* 7, 1526. doi:10.3389/fpls.2016.01526

Asfaw, A., Ambachew, D., Shah, T., and Blair, M. W. (2017). Trait Associations in Diversity Panels of the Two Common Bean (Phaseolus vulgaris L.) Gene Pools Grown under Well-Watered and Water-Stress Conditions. *Front. Plant Sci.* 8, 733. doi:10.3389/fpls.2017.00733

Askari, M., Abbaspour-Gilandeh, Y., Taghinezhad, E., El Shal, A. M., Hegazy, R., and Okasha, M. (2021). Applying the Response Surface Methodology (RSM) Approach to Predict the Tractive Performance of an Agricultural Tractor during Semi-deep Tillage. *Agriculture* 11, 1043. doi:10.3390/agriculture11111043

Ayuso, M., García-Pérez, P., Ramil-Rego, P., Gallego, P. P., and Barreal, M. E. (2019). *In Vitro* culture of the Endangered Plant Eryngium Viviparum as Dual Strategy for its *Ex Situ* Conservation and Source of Bioactive Compounds. *Plant Cell Tiss. Organ Cult.* 138, 427–435. doi:10.1007/s11240-019-01638-y

Bansal, M., Sudhakara Reddy, M., and Kumar, A. (2017). Optimization of Cell Growth and Bacoside-A Production in Suspension Cultures of Bacopa Monnieri (L.) Wettst. Using Response Surface Methodology. *Vitro Cell.Dev.Biol.-Plant* 53, 527–537. doi:10.1007/s11627-017-9847-0

Barpete, S., Aasim, M., Khawar, K. M., Özcan, S. F., and Özcan, S. (2014). Preconditioning Effect of Cytokinins on *In Vitro* Multiplication of Embryonic Node of Grass Pea (Lathyrus Sativus L.) Cultivar Gürbüz. *Turk. J. Biol.* 38, 485–492. doi:10.3906/biy-1312-94

Belide, S., Rao Sajjalaguddam, R., and Paladugu, A. (2010). Cytokinin Preconditioning Enhances Multiple Shoot Regeneration in Pongamia Pinnata (L.) Pierre - a Potential, Non-edible Tree Seed Oil Source for Biodiesel. *Electron Journ Biotechnol* 13, 3–4. doi:10.2225/vol13-issue6-fulltext-5

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., et al. (2012). Mesoamerican Origin of the Common Bean (Phaseolus vulgaris L.) Is Revealed by Sequence Data. *Proc. Natl. Acad. Sci. U.S.A.* 109, E788–E796. doi:10.1073/pnas.1108973109

Blair, M. W., Cortés, A. J., Farmer, A. D., Huang, W., Ambachew, D., Penmetsa, R. V., et al. (2018). Uneven Recombination Rate and Linkage Disequilibrium across a Reference SNP Map for Common Bean (Phaseolus vulgaris L.). *PloS one* 13, e0189597. doi:10.1371/journal.pone.0189597

Brar, M. S., Al-Khayri, J. M., Morelock, T. E., and Anderson, E. J. (1999). Genotypic Response of Cowpea Vigna Unguiculata (L.) to *In Vitro* Regeneration from Cotyledon Explants. *Vitro Cell.Dev.Biol.-Plant* 35, 8–12. doi:10.1007/s11627-999-0002-4

Breiman, L. (2001). Random Forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324

Campa, A., Murube, E., and Ferreira, J. J. (2018). Genetic Diversity, Population Structure, and Linkage Disequilibrium in a Spanish Common Bean Diversity Panel Revealed through Genotyping-By-Sequencing. *Genes* 9, 518. doi:10.3390/genes9110518

Castillo, B. M., Rodriguez de la O, J. L., Mascorro Gallardo, J. O., and Iturriaga, G. (2015). *In Vitro* plants of Common Bean (Phaseolus vulgaris L.) Obtained by Direct Organogenesis. *J. Agric. Sci.* 7, 169–179. doi:10.5539/jas.v7n11p169

Chen, T., and Guestrin, C. (2016). "Xgboost: A Scalable Tree Boosting System," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, San Francisco, CA, August 13–17, 2016, 785–794. doi:10.1145/2939672.2939785

Day, S., and Aasim, M. (2017). *In Vitro* Screening of Preconditioned Plumular Apices Explants of Peanut (Arachis Hypogeae) to Different Salts Concentration. *Fresenius. Environ. Bull.* 4671.

Drucker, H. (1997). Improving Regressors Using Boosting Techniques. *ICML* 97, 107–115.

Earl, C., Belitz, M. W., Laffan, S. W., Barve, V., Barve, N., Soltis, D. E., et al. (2021). Spatial Phylogenetics of Butterflies in Relation to Environmental Drivers and Angiosperm Diversity Across North America. *Iscience* 24 (4), 102239. doi:10.1016/j.isci.2021.102239

Farhadi, S., Salehi, M., Moieni, A., Safaie, N., and Sabet, M. S. (2020). Modeling of Paclitaxel Biosynthesis Elicitation in Corylus Avellana Cell Culture Using Adaptive Neuro-Fuzzy Inference System-Genetic Algorithm (ANFIS-GA) and Multiple Regression Methods. *PLoS ONE* 15 (8), e0237478. doi:10.1371/journal.pone.0237478

García-Pérez, P., Lozano-Milo, E., Landín, M., and Gallego, P. P. (2020). Machine Learning Technology Reveals the Concealed Interactions of Phytohormones on Medicinal Plant *In Vitro* Organogenesis. *Biomolecules* 10 (5), 746. doi:10.3390/biom10050746

Hameg, R., Arteta, T. A., Landin, M., Gallego, P. P., and Barreal, M. E. (2020). Modeling and Optimizing Culture Medium Mineral Composition for *In Vitro* Propagation of Actinidia Arguta. *Front. Plant Sci.* 11, 2088. doi:10.3389/fpls.2020.554905

Hesami, M., and Jones, A. M. P. (2020). Application of Artificial Intelligence Models and Optimization Algorithms in Plant Cell and Tissue Culture. *Appl. Microbiol. Biotechnol.* 104, 9449–9485. doi:10.1007/s00253-020-10888-2

Hesami, M., and Jones, A. M. P. (2021). Modeling and Optimizing Callus Growth and Development in Cannabis Sativa Using Random Forest and Support Vector Machine in Combination with a Genetic Algorithm. *Appl. Microbiol. Biotechnol.* 105, 5201–5212. doi:10.1007/s00253-021-11375-y

Hesami, M., Naderi, R., and Tohidfar, M. (2020a). Introducing a Hybrid Artificial Intelligence Method for High-Throughput Modeling and Optimizing Plant Tissue Culture Processes: the Establishment of a New Embryogenesis Medium for chrysanthemum, as a Case Study. *Appl. Microbiol. Biotechnol.* 104 (23), 10249–10263. doi:10.1007/s00253-020-10978-1

Hesami, M., Naderi, R., Tohidfar, M., and Yoosefzadeh-Najafabadi, M. (2019). Application of Adaptive Neuro-Fuzzy Inference System-Non-Dominated Sorting Genetic Algorithm-II (ANFIS-NSGAII) for Modeling and Optimizing Somatic Embryogenesis of Chrysanthemum. *Front. Plant Sci.* 10, 869. doi:10.3389/fpls.2019.00869

Hesami, M., Naderi, R., Tohidfar, M., and Yoosefzadeh-Najafabadi, M. (2020b). Development of Support Vector Machine-Based Model and Comparative Analysis with Artificial Neural Network for Modeling the Plant Tissue Culture Procedures: Effect of Plant Growth Regulators on Somatic Embryogenesis of chrysanthemum, as a Case Study. *Plant Methods* 16 (1), 1–15. doi:10.1186/s13007-020-00655-9

Hu, Z., Jin, Y., Hu, Q., Sen, S., Zhou, T., and Osman, M. T. (2019). Prediction of Fuel Consumption for Enroute Ship Based on Machine Learning. *IEEE Access* 7, 119497–119505. doi:10.1109/ACCESS.2019.2933630

Jahan, A. A., Anis, M., and Aref, I. M. (2011). Preconditioning of Axillary Buds in Thidiazuron-Supplemented Liquid Media Improves In Vitro Shoot

Multiplication in Nyctanthes arbor-tristis L. *Appl. Biochem. Biotechnol.* 163 (7), 851–859. doi:10.1007/s12010-010-9089-7

Jamshidi, S., Yadollahi, A., Ahmadi, H., Arab, M. M., and Eftekhari, M. (2016). Predicting *In Vitro* Culture Medium Macro-Nutrients Composition for Pear Rootstocks Using Regression Analysis and Neural Network Models. *Front. Plant Sci.* 7, 274. doi:10.3389/fpls.2016.00274

Jana, S., Sivanesan, I., and Jeong, B. R. (2013). Effect of Cytokinins on *In Vitro* Multiplication of Sophora Tonkinensis. *Asian Pac. J. Trop. Biomed.* 3, 549–553. doi:10.1016/S2221-1691(13)60111-2

Kami, J., Velásquez, V. B., Debouck, D. G., and Gepts, P. (1995). Identification of Presumed Ancestral DNA Sequences of Phaseolin in *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1101–1104. doi:10.1073/pnas.92.4.1101

Katirci, R., and Takci, H. (2021). Makine Öğrenmesi Kullanarak Krom Kaplama Örtme Gücünün Tahmin Edilmesi. *Firat Univ. J. Eng.* 33, 709–719. doi:10.35234/fumbd.950667

Kirtis, A., Aasim, M., and Katırcı, R. (2022). Application of Artificial Neural Network and Machine Learning Algorithms for Modeling the *In Vitro* Regeneration of Chickpea (*Cicer Arietinum* L.). *PCTOC* 2, 1–2. doi:10.1007/s11240-022-02255-y

Kumari, K., Lal, M., and Saxena, S. (2017). Enhanced Micropropagation and Tiller Formation in Sugarcane through Pretreatment of Explants with Thidiazuron (TDZ). *3 Biotech.* 7, 1–4. doi:10.1007/s13205-017-0910-7

Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., et al. (2013). Demographic Factors Shaped Diversity in the Two Gene Pools of Wild Common Bean Phaseolus vulgaris L. *Heredity* 110 (3), 267–276. doi:10.1038/hdy.2012.82

Managamuri, U., Vijayalakshmi, M., Krishna Ganduri, V. S. R., Rajulapati, S. B., and Poda, S. (2019). Assessment of Optimized Process Parameters for Superior Bioactive Metabolite Production by Nonomuraea Longicatena VSM-16 Using Response Surface Methodology. *Jyp* 11, 377–381. doi:10.5530/jyp.2019.11.77

Matand, K., Wu, N., Wu, H., Tucker, E., and Love, K. (2013). More Improved Peanut (*Arachis hypogaea* L.) Protocol for Direct Shoot Organogenesis in Mature Dry-Cotyledonary and Root Tissues. *J. Biotech. Res.* 5, 24. doi:10.5897/AJB10.823

Molnár, Z., Jenes, B., and Ördog, V. (1999). *Genetic Transformation of Pea (Pisum Sativum L.) via Particle Bombardment*. Stara Lesna, Slovak Republic: Recent Advances in Plant Biotechnology. September, 4-11.

Murashige, T., and Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant* 15, 473–497. doi:10.1111/j.1399-3054.1962.tb08052.x

Nadeem, M. A., Gündoğdu, M., Ercişli, S., Karaköy, T., Saracoğlu, O., HabyarimanaBaloch, E. F. S., et al. (2020b). Uncovering Phenotypic Diversity and DArTseq Marker Loci Associated with Antioxidant Activity in Common Bean. *Genes* 11, 36. doi:10.3390/genes11010036

Nadeem, M. A., Karaköy, T., Yeken, M. Z., Habyarimana, E., Hatipoğlu, R., Çiftçi, V., et al. (2020a). Phenotypic Characterization of 183 Turkish Common Bean Accessions for Agronomic, Trading, and Consumer-Preferred Plant Characteristics for Breeding Purposes. *Agronomy* 10, 272. doi:10.3390/agronomy10020272

Niazian, M., and Niedbała, G. (2020). Machine Learning for Plant Breeding and Biotechnology. *Agriculture* 10, 436. doi:10.3390/agriculture10100436

Niazian, M., Noori, S. A. S., Galuszka, P., Tohidfar, M., and Mortazavian, S. M. M. (2017). Genetic Stability of Regenerated Plants via Indirect Somatic Embryogenesis and Indirect Shoot Regeneration of Carum Copticum L. *Industrial Crops Prod.* 97, 330–337. doi:10.1016/j.indcrop.2016.12.044

Niazian, M., Sadat-Noori, S. A., Abdipour, M., Tohidfar, M., and Mortazavian, S. M. M. (2018). Image Processing and Artificial Neural Network-Based Models to Measure and Predict Physical Properties of Embryogenic Callus and Number of Somatic Embryos in Ajowan (Trachyspermum Ammi (L.) Sprague). *Vitro Cell.Dev.Biol.-Plant* 54, 54–68. doi:10.1007/s11627-017-9877-7

Niazian, M., and Shariatpanahi, M. E. (2020). In Vitro-based Doubled Haploid Production: Recent Improvements. *Euphytica* 216, 1–21. doi:10.1007/s10681-020-02609-7

Özkan, H., and Aasim, M. (2020). Comparison of *In Vitro* Regeneration Potential of Different Preconditioned and Nonconditioned Explants of Peanut (*Arachis hypogaea* L). *J. Environ. Agric. Sci.* 22, 11–19.

Özkan, H., and Aasim, M. (2019). Potential of Pretreated Explants of Peanut (Arachis Hypogeae Linn.) to Micropropagation under *In Vitro* Conditions. *Pak. J. Agric. Sci.* 56. doi:10.21162/PAKJAS/19.8262

Pavlov, A. V., Serdyuk, J. A., and Ustinov, A. B. (2019). Machine Learning and the Schrödinger Equation. *J. Phys. Conf. Ser.* 1236 (No. 1), 012050. doi:10.1088/1742-6596/1236/1/012050

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J Mach Learn Res*12, 2825–2830.

Pepe, M., Hesami, M., Small, F., and Jones, A. M. P. (2021). Comparative Analysis of Machine Learning and Evolutionary Optimization Algorithms for Precision Micropropagation of Cannabis Sativa: Prediction and Validation of *In Vitro* Shoot Growth and Development Based on the Optimization of Light and Carbohydrate Sources. *Front. Plant Sci.* 12, 757869. doi:10.3389/fpls.2021.757869

Prasad, V. S. S., and Gupta, S. D. (2008). "Applications and Potentials of Artificial Neural Networks in Plant Tissue Culture," in *Plant Tissue Cult. Lett.* (Dordrecht: Springer), 47–67. doi:10.1007/978-1-4020-3694-1_3

Premkumar, G., Karuppanapandian, T., Sureshpandian, C., Arumugam, N., Selvam, A., and Rajarathinam, K. (2020). Optimization of a Liquid Culture System for Shoot Regeneration and Achieving an Enriched Level of Scopadulcic Acid B in the Leaf Organ Cultures of Scoparia Dulcis L. By Response Surface Methodology. *Vitro Cell.Dev.Biol.-Plant* 56, 60–71. doi:10.1007/s11627-019-10037-z

Salehi, A., Mobarhan, M. A., Mohammadi, J., Shahsavarani, H., Shokrgozar, M. A., and Alipour, A. (2020). Efficient Mineralization and Osteogenic Gene Overexpression of Mesenchymal Stem Cells on Decellularized Spinach Leaf Scaffold. *Gene* 757, 144852. doi:10.1016/j.gene.2020.144852

Salehi, M., Farhadi, S., Moieni, A., Safaie, N., and Hesami, M. (2021). A Hybrid Model Based on General Regression Neural Network and Fruit Fly Optimization Algorithm for Forecasting and Optimizing Paclitaxel Biosynthesis in Corylus Avellana Cell Culture. *Plant Methods* 17, 1–13. doi:10.1186/s13007-021-00714-9

Sammut C. and Webb, G.I. (Editors) (2011). *Encyclopedia of Machine Learning* (New York, United States: Springer Science and Business Media). doi:10.1007/1978-0387-30164-8

Silva, T. S., de Freitas Souza, M., Maria da Silva Teófilo, T., Silva dos Santos, M., Formiga Porto, M. A., Martins Souza, C. M., Barbosa dos Santos, J., and Silva, D. V. (2019). Use of Neural Networks to Estimate the Sorption and Desorption Coefficients of Herbicides: A Case Study of Diuron, Hexazinone, and Sulfometuron-Methyl in Brazil. *Chemosphere* 236, 124333. doi:10.1016/j.chemosphere.2019.07.064

Singh, S., and Hazra, S. (2009). Somatic Embryogenesis from the Axillary Meristems of Peanut (*Arachis hypogaea* L.). *Plant Biotechnol. Rep.* . 3, 333-3–340. doi:10.1007/s11816-009-0107-6

Slimani, C., El Goumi, Y., Rais, C., El Ghadraoui, L., Benjelloun, M., and Lazraq, A. (2021). Optimization of Callogenesis/Caulogenesis Induction Protocol in Saffron Plant (Crocus Sativus L.) Using Response Surface Methodology. *Biointerface Res. Appl. Chem.* 12, 4731–4746. doi:10.33263/BRIAC124.47314746

Smola, A. J., and Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics Comput.* 14, 199–222. doi:10.1023/B:STCO.0000035301.49549.88

Sugimoto, K., Temman, H., Kadokura, S., and Matsunaga, S. (2019). To Regenerate or Not to Regenerate: Factors that Drive Plant Regeneration. *Curr. Opin. Plant Biol.* 47, 138–150. doi:10.1016/j.pbi.2018.12.002

Surekha, C., Beena, M. R., Arundhati, A., Singh, P. K., Tuli, R., Dutta-Gupta, A., et al. (2005). Agrobacterium-mediated Genetic Transformation of Pigeon Pea (Cajanus Cajan (L.) Millsp.) Using Embryonal Segments and Development of Transgenic Plants for Resistance against Spodoptera. *Plant Sci.* 169, 1074–1080. doi:10.1016/j.plantsci.2005.07.011

Tang, Y., Chen, L., Li, X. M., Li, J., Luo, Q., Lai, J., et al. (2012). Effect of Culture Conditions on the Plant Regeneration via Organogenesis from Cotyledonary Node of Cowpea (Vigna Unguiculata L. Walp). *Afr. J. Biotechnol.* 11, 3270–3275. doi:10.5897/ajb11.3214

Van Rossum, G., and Drake, F. L., Jr (1994). *Python Reference Manual*. Surrey, England: Network Theory LTD. doi:10.5555/1610526

Vanlauwe, B., Hungria, M., Kanampiu, F., and Giller, K. E. (2019). The Role of Legumes in the Sustainable Intensification of African Smallholder Agriculture: Lessons Learnt and Challenges for the Future. *Agric. Ecosyst. Environ.* 284, 106583. doi:10.1016/j.agee.2019.106583

Wang, X. J., Liu, S. T., Xia, H., Wan, S. B., Zhao, C. Z., and Li, A. Q. (2011). Peanut ('*Arachis hypogaea* L.') Omics and Biotechnology in China. *Plant Omics* 4, 339–349. doi:10.3316/informit.684496127565121

Yeken, M. Z., Nadeem, M. A., Karaköy, T., Baloch, F. S., and Çiftçi, V. (2019). Determination of Turkish Common Bean Germplasm for Morpho-Agronomic

and Mineral Variations for Breeding Perspectives in Turkey. Kahramanmaraş Sütçü İmam Üniv. *Tarım Doğa Derg.* 22, 38–50. doi:10.18016/ksutarimdoga.vi.563740

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 © 2022 Aasim, Katirci, Baloch, Mustafa, Bakhsh, Nadeem, Ali, Hatipoğlu, Çiftçi, Habyarimana, Karaköy and Chung. 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 September 2022
doi: 10.3389/fgene.2022.905824

Incorporation of Soil-Derived Covariates in Progeny Testing and Line Selection to Enhance Genomic Prediction Accuracy in Soybean Breeding

Caio Canella Vieira[1†], *Reyna Persa*[2†], *Pengyin Chen*[1] *and Diego Jarquin*[2*]

[1]Division of Plant Science and Technology, Fisher Delta Research, Extension and Education Center, University of Missouri, Portageville, MO, United States, [2]Agronomy Department, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:
Aamir Raina,
Aligarh Muslim University, India

Reviewed by:
Yang Xu,
Yangzhou University, China
Chengsong Zhu,
University of Texas Southwestern Medical Center, United States
Babu Valliyodan,
Lincoln University, United States

***Correspondence:**
Diego Jarquin
jhernandezjarqui@ufl.edu

[†]*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: *27 March 2022*
Accepted: *08 June 2022*
Published: *08 September 2022*

Citation:
Canella Vieira C, Persa R, Chen P and Jarquin D (2022) Incorporation of Soil-Derived Covariates in Progeny Testing and Line Selection to Enhance Genomic Prediction Accuracy in Soybean Breeding. Front. Genet. 13:905824. doi: 10.3389/fgene.2022.905824

The availability of high-dimensional molecular markers has allowed plant breeding programs to maximize their efficiency through the genomic prediction of a phenotype of interest. Yield is a complex quantitative trait whose expression is sensitive to environmental stimuli. In this research, we investigated the potential of incorporating soil texture information and its interaction with molecular markers *via* covariance structures for enhancing predictive ability across breeding scenarios. A total of 797 soybean lines derived from 367 unique bi-parental populations were genotyped using the Illumina BARCSoySNP6K and tested for yield during 5 years in Tiptonville silt loam, Sharkey clay, and Malden fine sand environments. Four statistical models were considered, including the GBLUP model (M1), the reaction norm model (M2) including the interaction between molecular markers and the environment (G×E), an extended version of M2 that also includes soil type (S), and the interaction between soil type and molecular markers (G×S) (M3), and a parsimonious version of M3 which discards the G×E term (M4). Four cross-validation scenarios simulating progeny testing and line selection of tested–untested genotypes (TG, UG) in observed–unobserved environments [OE, UE] were implemented (CV2 [TG, OE], CV1 [UG, OE], CV0 [TG, UE], and CV00 [UG, UE]). Across environments, the addition of G×S interaction in M3 decreased the amount of variability captured by the environment (–30.4%) and residual (–39.2%) terms as compared to M1. Within environments, the G×S term in M3 reduced the variability captured by the residual term by 60 and 30% when compared to M1 and M2, respectively. M3 outperformed all the other models in CV2 (0.577), CV1 (0.480), and CV0 (0.488). In addition to the Pearson correlation, other measures were considered to assess predictive ability and these showed that the addition of soil texture seems to structure/dissect the environmental term revealing its components that could enhance or hinder the predictability of a model, especially in the most complex prediction scenario (CV00). Hence, the availability of soil texture information before the growing season could be used to optimize the efficiency of a breeding program by allowing the reconsideration of

field experimental design, allocation of resources, reduction of preliminary trials, and shortening of the breeding cycle.

Keywords: genomic prediction/selection, genotype × environment G×E interaction, soil covariates, genetic gain, soybean breeding

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] represents the largest and most concentrated segment of global agricultural trade (Gale et al., 2019). It is the crop that delivers the highest amount of protein per hectare and accounts for over 60% of total global oilseed production (United States Department of Agriculture, 2022). Worldwide, Brazil (37%, 139,000 MT), United States (32%, 120,700 MT), and Argentina (12%, 46,500 MT) account for over 80% of the soybean production (United States Department of Agriculture, 2022). Over the last two decades (2001/2002 to 2021/2022), soybean production has nearly doubled from 182,830 to 363,860 MT (United States Department of Agriculture, 2002; United States Department of Agriculture, 2022). The substantial increase in soybean production can be attributed to advances in agronomical practices (Specht et al., 1999; Mourtzinis et al., 2018), faster implementation of novel technologies in farming operations (Liu et al., 2008; Ainsworth et al., 2012; Vieira and Chen, 2021), and the development of improved soybean cultivars (Salado-Navarro et al., 1993; Voldeng et al., 1997; Specht et al., 1999; Specht and Williams, 2015; Vieira and Chen, 2021), of which the availability of high dimensional genomic (Song et al., 2013, 2020) and phenomic data (Moreira et al., 2019, 2020; Parmley et al., 2019; Zhou et al., 2022), as well as the integration of environmental covariates (ECs) through predictive analytics, have contributed to accelerated genetic gains (Jarquin et al., 2014a; Jarquin et al., 2014b; Persa et al., 2020; Widener et al., 2021).

Marker-assisted selection (MAS) has greatly contributed to the improvement and selection of soybean traits regulated by major effect genes, including biotic (Pham et al., 2013; Shi et al., 2015) and abiotic tolerance (Pathan et al., 2007; Wu et al., 2020), as well as seed composition–related traits (Pham et al., 2010; Patil et al., 2017). On the other hand, yield is a highly complex quantitative trait regulated by many genes with small effects, thus limited success has been reported in applying MAS (Concibido et al., 2003; Jarquin et al., 2014a). Bernardo (1994) was the first one who proposed the use of genomic variants (RLFPs) for predicting trait performance for selecting genotypes (genomic selection, GS), back then, the number of these covariates was limited/reduced. Later, Meuwissen et al. (2001) proposed a new methodology to deal with the curse of the dimensionality problem ($n < p$; n is the number of data points available for model fitting and p is the number of genomic variants) and it is considered a landmark in genomic research. The concept of GS revolves around using the information of all molecular markers—large and small effects—to develop prediction models for the phenotype of interest. The major advantage of GS relies on the ability to predict the yield of genotypes to allow the identification and selection of the most promising individuals earlier in the breeding pipeline, which not only reduce costs, time, and space but enhance the genetic gain by reducing the length of the breeding cycle, increasing selection intensity, as well as allowing the breeders to have a clear knowledge of the genetics of the materials early in the pipeline (Jarquin et al., 2014b; Crossa et al., 2017; Vieira and Chen, 2021; Wartha and Lorenz, 2021).

In soybean, the first application of GS was reported by Jarquin et al. (2014a). By using a standard G-BLUP model including only additive effects and an extended version of the G-BLUP model including additive-by-additive effects, a prediction accuracy of 0.64 for grain yield and roughly 41% of the phenotypic variance explained by the genotypic component were reported using 301 lines of the University of Nebraska soybean breeding program. Usually, different response patterns in a set of genotypes are observed when these are tested under different environmental conditions complicating the selection of the most promising candidates (Crossa et al., 2004). The presence of these changes in the response pattern of the ranking of the genotypes is also known as the genotype-by-environment interaction effect. To allow the consideration of this interaction effect in prediction models using weather data, Jarquin et al. (2014b) proposed a reaction norm model that allows the incorporation of the main and interaction effects of both high-dimensional molecular markers and EC through covariance structures using data from wheat cultivars tested in 340 environments. In the cross-validation scenario that considers the prediction of the performance of genotypes that have never been evaluated in field trials (CV1), in comparison with the conventional main effect genomic selection model, the reaction norm model enhanced prediction accuracy by 35%, whereas in the cross-validation scenario where all genotypes had at least one field evaluation available (CV2), a 17% increase in predictive ability was observed (Jarquin et al., 2014b). Using the soybean nested association panel (SoyNAM), Xavier et al. (2016) investigated the impacts of training population size, genotyping density, and 14 prediction models on the accuracy of genomic prediction. These authors showed that the training population size was the most impactful factor in the accuracy improvement. Ma et al. (2016) used ridge regression best linear unbiased prediction (rrBLUP) (Endelman, 2011) with fivefold cross-validation to explore strategies of marker preselection. The prediction accuracy based on markers selected with a haplotype block analysis–based approach increased by approximately 4% compared with random or equidistant marker sampling. Stewart-Brown et al. (2019) investigated the effects of two relatedness strategies among genotypes in overall prediction accuracies and found both methods returned similar accuracies. The first method was based on each bi-parental population and

utilized a training set of full-sibs of the validation set. The second method utilized a training set of all remaining breeding lines except for full-sibs of the validation set to predict across populations. Persa et al. (2020) expanded the reaction norm model proposed by Jarquin et al. (2014b) by incorporating the interaction between genotypes' families and the environment under the premise that the differential responses of families to environmental stimuli could be used for enhancing the selection process in target environments. The most comprehensive model improved the predictive ability by 41% (CV1) and 49% (CV2) compared to the standard GBLUP, and roughly 17% as compared to the conventional reaction norm model. Widener et al. (2021) included three EC (mean minimum daily temperature, mean maximum daily temperature, and mean daily precipitation) interactions with molecular markers into the reaction norm model and no substantial increase in prediction accuracy was observed and resulted in more often negative predictions although these authors were only interested in assessing strategies to selecting sets of environments for model training. These authors found that in predicting the most dissimilar environment (based on phenotypes and environmental covariates) a reduced set of environments is adequate to optimize predictive ability while for the most similar environment, as the number of environments in the training set increased the predictive ability was improved too.

The covariance structure proposed in the reaction norm allows the borrowing of information between genotypes based on environmental and genomic similarities. For instance, in Jarquin et al. (2014b), the covariance matrices describing the similarities between environmental conditions and genetic information permit the borrowing of information between environments and molecular markers, respectively. The cross-validation scenario where untested genotypes are being predicted in untested environments (CV00) is often the challenge faced in the early stages of a breeding pipeline also known as the progeny stage. In this situation, the environmental conditions in upcoming growing seasons are often unpredictable and distinct from what was used in the model's training dataset limiting the main advantage of the approach based on conventional covariance structures only. Soil-related information such as soil texture is generally constant across years and readily available before the growing season. Consequently, leveraging the information of soil texture as the main effect as well as its interaction with molecular markers could potentially increase predictive ability, particularly in scenarios considering untested genotypes in untested environments. Therefore, the objective of this study was to investigate the potential of including soil-derived covariates in the reaction norm model to enhance the predictive ability under common plant breeding scenarios, including the prediction of untested genotypes in untested environments (progeny testing) as well as multiple combinations of tested genotypes in tested and untested environment simulating line selection. A set of 797 advanced soybean breeding lines derived from unique 367 bi-parental populations was used in this study. Lines were evaluated for grain yield between 2017 and 2021 and genotyped using the Illumina Infinium BARCSoySNP6K BeadChip.

MATERIALS AND METHODS

Plant Materials and Field Trials

A set of 797 advanced soybean breeding lines derived from 367 unique bi-parental populations developed by the University of Missouri–Fisher Delta Research, Extension, and Education Center (MU-FDREEC), soybean breeding program was used in this study. The lines comprised 5 years (2017–2021) of internal advanced yield trials at the MU-FDREEC. Five seeds of each line were germinated in paper pouches for 3–4 days at room temperature and seedlings were transplanted into micropots filled with sterilized sandy loam soil. Genomic DNA was extracted from lyophilized young trifoliate leaf tissue (V3) (Fehr et al., 1971) using the Qiagen DNeasy Plant 96 kit (QIAGEN, Valencia, CA, United States) and respective protocol. DNA concentration was quantified using a spectrophotometer (NanoDrop Technologies Inc., Centerville, DE, United States) and normalized at 50 ng/μl. DNA samples were genotyped in the USDA-ARS Soybean Genomics and Improvement Laboratory using the Illumina Infinium BARCSoySNP6K BeadChip (Song et al., 2020). The single nucleotide polymorphism (SNP) alleles were called using the Illumina Genome Studio Genotyping Module (Illumina, Inc., San Diego, CA, United States).

Field trials were conducted for 5 years (2017–2021) at the Lee Farm in Portageville, MO (36°23′44.2″N latitude and 89°36′52.3″W longitude) and the Rhodes Farm in Clarkton, MO (36°29′14.8″N latitude and 89°57′39.0″W longitude) using a three-replicate randomized complete block design. At the Lee Farm, trials were conducted each year in four environments consisting of two Tiptonville silt loam and two Sharkey clays. Tiptonville silt loam consists of very deep, nearly level, moderately well-drained soils formed in silty alluvium (United States Department of Agriculture, 2018a), whereas Sharkey clay is very deep, poorly drained, and very slowly permeable in soils that is formed in clayey alluvium (United States Department of Agriculture, 2013). At the Rhodes farm, trials were conducted in one Malden fine sand environment each year. This consists of very deep, excessively drained soils formed in sandy alluvium (United States Department of Agriculture, 2018b). Each plot consisted of four rows 3.66 m long spaced 0.76 m apart. The two center rows of each plot were harvested with a plot combined for seed yield adjusted to 13% seed moisture.

Statistical Models

For assessing the effects of the soil type–derived covariates and their interactions with environmental factors in genomic prediction, four models were considered.

M1: E+L+G

This model allows the inclusion of the main effect of the molecular markers *via* covariance structures. Suppose that the

genomic effect g_i of the ith line can be characterized by a linear combination between p molecular markers x_m $(m = 1, 2, \ldots, p)$ and their corresponding effects b_m such that $g_i = \sum_{m=1}^{p} x_m b_m$, with $b_m \sim N(0, \sigma_b^2)$. If we include all the genomic effects into a single vector, we have $\mathbf{g} = \mathbf{Xb}$. From results of the multivariate normal density, the vector of genomic effects $\mathbf{g} = \{g_i\} \sim N(\mathbf{0}, \mathbf{G}\sigma_{\mathbf{g}}^2)$ with $\mathbf{G} = \frac{\mathbf{XX}'}{p}$, and $\sigma_g^2 = p\sigma_b^2$ as the corresponding variance component. In this way, the linear predictor becomes.

$$y_{ij} = \mu + E_j + L_i + g_i + \varepsilon_{ij} \quad (1)$$

where the yield response y_{ij} of the ith genotype observed at the jth environment can be modeled as the sum of a mean effect μ common to all genotypes across environments, a random effect of the ith line L_i following an independent and identically distributed (IID) normal density centered on zero and variance σ_L^2 such that $L_i \sim N(0, \sigma_L^2)$, a random environmental effect of the jth environment E_j following IID normal densities centered on zero and variance σ_E^2 such that $E_j \sim N(0, \sigma_E^2)$, and a random effect ϵ_{ij} addressing the unexplained variability by these model terms such that $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.

M2: E+L+G+G×E

To consider the effect of the environmental stimuli on the genomic responses, Jarquin et al. (2014b) proposed the reaction norm model. Briefly, this model indirectly allows the inclusion of the interaction between each molecular marker and each environment or environmental covariate in prediction models *via* covariance structures. Consider gE_{ij} as the random effect explaining the genomic interaction between the ith genotype and the jth environment such that the vector of interaction effects $\mathbf{gE} = \{gE_{ij}\} \sim N(\mathbf{0}, Z_E Z_E' \# Z_L \mathbf{G} Z_L' \sigma_{gE}^2)$, where Z_L and Z_E are the incidence matrices that connect phenotypes with genotypes and environments, respectively, σ_{gE}^2 is the corresponding variance component, and "#" represents the Hamadard product (cell-by-cell product) between two matrices of the same dimensions. Adding this model term to M1 results in the following linear predictor:

$$y_{ij} = \mu + E_j + L_i + g_i + gE_{ij} + \varepsilon_{ij} \quad (2)$$

This model has shown significant improvements in predictive ability compared with the conventional GS model (M1) when predicting the yield of genotypes in already observed environments. However, in more challenging scenarios like those where no phenotypic records from the target environment are available for any genotype, the advantage becomes less pronounced likely due to the environmental stimuli not being properly accounted for. Also, predicting future environments poses an extra challenge since it is not feasible to forecast the expected weather conditions in a precise manner limiting the usefulness of M2 in these cases.

M3: E+L+S+G+G×E+G×S

An important component of the environmental stimuli that genotypes are exposed to is the multiple soil conditions, of which soil structures are factors that can be easily obtained in advance during the planning stage of the experiments. The current model attempts to leverage the information on the soil structure in the prediction context. Consider S_k as the random effect that represents the soil type where the soybean cultivars were planted (k = 1, 2,. . ., K). Furthermore, if we assume these effects as IID outcomes from a normal distribution centered on zero and with a common variance σ_S^2 we have $S_k \sim \mathrm{N}(0, \sigma_S^2)$. This model term allows the inclusion of the main effect of the soil type in the prediction model. In principle, it is assumed that the effect of soil type is the same for all genotypes planted in a given experiment. Thus, this model term will not help to improve the predictive ability because their effects are common to all genotypes within the same experiment. For this reason, we also considered the interaction between the molecular markers and the soil type to permitting specific responses within environments also allows the borrowing of information between genotypes planted at different soil types. For this, we used the same principles as in M2 such that gS_{ik} represents the interaction effect of the ith genotyped at the kth soil type. If we include these interaction effects in a vector we have $\mathbf{gS} = \{gS_{ik}\} \sim N(\mathbf{0}, Z_L \mathbf{G} Z_L' \# Z_S Z_S' \sigma_{gS}^2)$, where Z_S is the incidence matrix that connects phenotypes with the soil type where the genotypes are observed, and σ_{gS}^2 represents the associated variance component. Combining this model term with M3, we have the resulting linear predictor.

$$y_{ij} = \mu + E_j + S_k + L_i + g_i + gE_{ij} + gS_{ik} + \varepsilon_{ij} \quad (3)$$

where all of the remaining terms remain as previously defined.

M4: E+L+S+G+G×S

Finally, a fourth model (M4) results from dropping the G×E term from M3. It is an attempt to have an intermediate implementation between models M2 and M3. The resulting model is as follows:

$$y_{ij} = \mu + E_j + S_k + L_i + g_i + gS_{ik} + \varepsilon_{ij} \quad (4)$$

where all of the remaining terms remain as previously defined.

Cross-Validation Schemes

In this study, four cross-validation schemes that simulate realistic prediction scenarios of interest for breeders for screening, selecting, and advancing genotypes through the breeding pipeline were implemented. The goal of considering these four prediction scenarios is to evaluate if in any of these the integration of soil-derived covariates accomplishes significant improvements in predictive ability. Persa et al. (2020) provide a comprehensive review of these four cross-validation scenarios and an extension to balancing the sample sizes in training and testing sets across cross-validation schemes.

The first prediction scenario is called CV2 (tested genotypes in observed environments), and it refers to the problem of predicting already tested genotypes in already observed environments. The main purpose of this scheme is to assess the predictability of partial field trials. Few genotypes have already been observed in some environments but not in others and the interest is to predict their performance in those environments where these genotypes were not

observed. In this study, a fivefold cross-validation was considered such that around 20% of the phenotypic values were assigned to the testing set and the remaining 80% (or four folds) to the training set which is used for model calibration. The model evaluation was conducted by predicting each fold (one at a time) using the remaining four folds for calibration, and this procedure was repeated until all the five folds were completed. This previous procedure was repeated 10 times.

The second prediction scenario is CV1 (untested genotypes in observed environments), and it refers to the problem of predicting untested genotypes in already observed environments where other genotypes were already tested. This prediction scenario mimics the problem of predicting (novel or newly developed) genotypes that were not observed in any of the environments; however, in these environments there is available phenotypic information for other genotypes. Even though the phenotypic information for these target genotypes of interest is not available, it is possible to borrow information from other genotypes *via* genomic data to allow the prediction of the unobserved genotypes. Also, a fivefold cross-validation was considered. In this CV, genotypes were assigned to folds instead of phenotypes such that all phenotypic records from the same genotype are encountered in the same fold. Under this scenario, around 20% of the genotypes were used as validation or testing set and the rest (~80% of the genotypes) were considered for the model's calibration. Similarly, to CV2, each fold was predicted (one at a time) using the remaining four folds and this procedure was repeated 10 times.

The CV0 (tested genotypes in unobserved environments) cross-validation scheme considers the scenario of predicting the performance of already observed genotypes in other environments and the interest is to predict their performance in an unobserved/novel environment. Under this scheme, the genotypes' mean performance is predicted in a hypothetical unobserved environment. The training set includes phenotypic records from all the genotypes in these already observed environments. The validation is conducted by predicting the performance of all the lines in one unobserved environment (one at a time) using the information of the remaining environments (training set). These steps are repeated for every environment.

CV00 (untested genotypes in unobserved environments) is perhaps the most interesting cross-validation scenario for breeders but also it is the most challenging. It considers the prediction of novel genotypes that have not been tested in any environment yet, and breeders are interested in their performance in an unobserved/novel environment. The strategy for estimating untested genotypes in new environments consists of removing all the phenotypic information from the target environment as well as all the phenotypic information from the training set but corresponding to only to those genotypes in the testing set (unobserved environment).

Model Assessment

The predictive ability of the different models for the different cross-validation schemes was calculated as the within-environment correlation between the predicted and observed values. These correlations provide an assessment of the model's predictive ability at the environment level which may vary substantially across environments due to a large number of unaccounted environmental conditions and sample sizes of the environments.

A general assessment across environments predictability is obtained by computing the weighted average correlation to account for uncertainty and the sample size of the environments as proposed by Tiezzi et al. (2017).

$$r_w = \frac{\sum_{j=1}^{J} \frac{r_j}{V(r_j)}}{\sum_{j=1}^{J} \frac{1}{V(r_j)}}$$

where $V(r_j) = \frac{1-r_j^2}{n_j-2}$, r_j represents the Pearson correlation between predicted and observed records at the $j^{th}(w = 1, \ldots, 50)$ environment; $V(r_j)$ and n_j corresponds to the sampling variance and number observations, respectively.

Variance Components

In general, the addition of model terms would result in a change in the predictive ability. To assess the importance/contribution of these terms, a full data analysis (i.e., non-missing values) was conducted to compute the variance components and examine the relative contribution of the different model components for each model. For this, the proportion of explained variability from each model term z is calculated as the ratio of the associated variance component to the sum of all t variance components ($z = 1, 2,..,t$) in the model multiplied by 100

$$\left(\frac{\sigma_z^2}{\sum_{z=1}^{t}\sigma_z^2} \times 100\right)$$

RESULTS

Variance Components

The relative amount of phenotypic variability (percentage) explained by the different model terms (across and within environments) for the four models (M1–M4) is provided in **Table 1**. Across environments, in M1; the environment component (E) captured the largest amount of phenotypic variability (65.7%) while the lines (L) and the main effect of the markers (G) explained 1.2 and 7.3% respectively, and the remaining non-explained variability addressed by the error term (R) was 25.8%. The addition of model terms significantly reduced the amount of variability captured by the E and R terms. Under the most complex model (M3), the corresponding values were reduced by 30.4% (45.7%) and 39.2% (15.7%), respectively. Concerning the percentage of within environment variability (after discarding the E term), the residual term R captured 75.2% with M1 while with M3, it was reduced by almost threefold to 28.9%. The interaction between molecular markers and environments (G×E) and between molecular markers and soil type (G×S) explained 20 and 17.5% of the phenotypic

TABLE 1 | Percentage of phenotypic variability explained by the different model components across and within environments for the four models (M1–M4).

Model	% Of across environment variability							% Of within environment variability					
	E[a]	L	S	G	G×E	G×S	R	L	S	G	G×E	G×S	R
M1: E+L+G	65.7	1.2		7.3			25.8	3.6		21.3			75.2
M2: E+L+G+G×E	65.7	1.6		6.0	12.7		14.0	4.5		17.5	37.1		40.9
M3: E+L+S+G+G×E+G×S	45.7	2.2	12.5	3.5	10.9	9.5	15.7	4.0	23.1	6.5	20.0	17.5	28.9
M4: E+L+S+G+G×S	64.0	1.5	0.0	4.2		9.5	20.8	4.2		11.6		26.4	57.8

[a]*The letters E, L, S, and G denote the mean effects of environments, genotypes, soil type, and molecular markers, respectively, whereas G×E and G×S reflect the interaction of each molecular marker with environments and soil type, respectively. The residual variance is denoted by R.*

TABLE 2 | Weighted mean average correlation across environments for four cross-validation schemes and four models.

Model[a]	CV2[b]	CV1	CV0	CV00
M1: E+L+G	0.461	0.359	0.461	**0.240**[c]
M2: E+L+G+G×E	0.558	**0.480**	0.459	0.192
M3: E+L+S+G+G×E+G×S	**0.577**	**0.480**	**0.488**	0.227
M4: E+L+S+G+G×S	0.515	0.405	0.484	0.231

[a]*E, L, S, and G constitute the main effect of the environments, genotypes, soil type, and molecular markers; and G×E and G×S evoke the interaction between each molecular marker with environments and soil type, respectively.*

[b]*CV2 considers the case of predicting incomplete field trials (i.e., some genotypes tested in some environments but not others), whereas CV1 assessed the accuracy of predicting newly developed genotypes. CV0 represents plant performance in novel environments of previously studied genotypes. CV00 assesses new genotypes in novel environments. For CV2 and CV1, 10 replicates of fivefold cross-validation were considered while for CV0 and CV00 the leave one environment out scheme was implemented.*

[c]*Bolded numbers indicate the best model performance for each cross-validation scheme.*

variability, respectively. These results highlight the importance of considering the interaction between molecular markers and environmental descriptors (environments and soil type) with the potential for improving predictive ability.

Predictive Ability

A very quick assessment of the ability of the different models for performing predictions can be achieved by revising the within environment mean average correlation between predicted and observed values. **Table 2** displays the mean average correlations for the four cross-validation schemes (CV2, CV1, CV0, and CV00) and to the four prediction models (M1–M4), and the results of the best model are highlighted in boldface by columns. Under the incomplete field trial scenario (CV2), the best model was M3 (0.577) which improved the conventional genomic selection model (M1) by 25.1% and was approximately 4% superior to the reaction norm model including G×E (M2). For the scenario of predicting newly developed lines in observed environments (CV1), models M2 and M3 performed similarly (~0.48), outperforming M1 by 34%. When predicting the yield of already tested genotypes in novel environments (CV0), the inclusion of G×E and G×S did not provide substantial improvement in overall accuracy as observed in the other cross-validation scenarios. In this case, the best model was M3 (0.488) which slightly outperformed M1 (0.461), M2 (0.459), and M4 (0.484). Thus, an improvement of 6% in the predictive ability was observed in M3 as compared to M1. In the most challenging and interesting prediction scenario consisting of predicting new genotypes in novel environments (CV00), the main effect model M1 returned the highest average correlation (0.240), followed by M4 (0.231), M3 (0.227), and M2 (0.192). In general, when considering only the mean average correlation as the unique criteria for selecting the best prediction model, M3 outperformed the other models in CV2, CV1, and CV0, while under CV00 the conventional main effect model M1 yielded the highest predictive accuracy.

Within Environment Predictive Ability as a Function of the Sample Size

Supplementary Figures S1, S2 in the Supplemental Section display the within-environment average correlation (*y*-axis) between predicted and observed values (10 replicates of fivefold cross-validation) as a function of the sample size of the environments (*x*-axis) under CV2 and CV1 prediction scenarios for the four models (M1–M4). For CV0 and CV00, since these do not involve a randomization process because each environment is left out at a time, the correlation between predicted and observed values are computed only once within environments and their corresponding results are displayed in **Supplementary Figures S3, S4**, respectively. For the four cross-validation schemes, the correlations for each model-environment are provided in Supplemental Section in **Supplementary Tables S1–S4**.

Under the CV2 scenario, in **Supplementary Figure S1** (**Supplementary Table S1**); we observed that as the number of genotypes in the target environment increased the mean average correlation also increased. Negative correlations were observed with the M1 (panel A) model in 11 of the 50 environments, while these negative values were observed with the M2, M3, and M4 models in only six, four, and four environments, respectively. For the CV1 scenario, a similar trend to the previous prediction scheme was observed. The main effect model M1 returned negative values in eight environments, M2 returned negative values in only five environments, M3 returned the lowest number of environments with adverse outcomes (3), and the intermediate model M4 resulted in five environments with negative correlations (**Supplementary Figure S2** and **Supplementary Table S2**). In the CV0 scheme, the model M1 returned nine out of the 50 environments with negative correlations while with M2 10 out of the 50 environments resulted in negative correlations (**Supplementary Figure S3**

and **Supplementary Table S3**). The interaction models that consider the soil type (M3 and M4) resulted in only five environments with negative results. Regarding the most complex prediction scenario CV00, the main effect model M1 returned negative results in nine out of the 50 environments, M2 resulted in 10 environments with adverse outcomes while M3 and M4 returned only six and five environments with negative correlations, respectively (**Supplementary Figure S4** and **Supplementary Table S4**).

Predictive Ability of Genotypes in Environments

Another way to assess model predictive ability was introduced by Jarquin et al. (2014b). These authors superimposed a grid on the scatter plot between predicted and observed values with the grid's vertical and horizontal lines represent the empirical percentiles (20, 50, and 80%) of the predicted and observed values, respectively. Also, within each rectangle of the grid, the proportion of genotypes at each category in the *y*-axis (observed values) conditional on the groups/categories defined by the *x*-axis (predicted values) is displayed. **Supplementary Figures S5–S8** in Supplemental Section contain the corresponding conditional plots for the four cross-validations and the four models.

For CV2, among the top 20% (i.e., to the right from the vertical line in the 80% mark on the *x*-axis) of the predicted genotypes in environments with model M1 (top right panel A), 68% of these showed an observed performance among the top 20% (i.e., above the 80% of the horizontal line in the *y*-axis) phenotypes in fields (**Supplementary Figure S5**). On the other hand, out of the bottom 20% (i.e., to the left from the vertical line in the 20% mark on the *x*-axis) of the genotypes predicted to have the lowest performance in fields, 71% were among the observed genotypes with the poorest performance. In addition, a linear regression between the predicted and observed values was performed, as well the mean squared error (MSE) and the weighted average correlation across environments (Cor) were added to the plot. An R-square (R^2) of 0.66 resulted from regressing the observed values on the predicted values, MSE = 94.1 and a Cor = 0.461.

Using the M2 model, 71% of the genotypes projected to have superior performance in fields (i.e., among the top 20%) were classified in the right category while 74% of those predicted with the poorest performance were among the phenotypes with the lowest performance. The resulting R^2 was 0.72 for a MSE = 77.6 and a Cor = 0.558. The most complex model M3, returned classification successes of the top and the worse genotypes in fields of 71 and 76%, respectively, for an R^2 = 0.73, MSE = 75.7, and a Cor = 0.577. For the intermediate model M4, the corresponding classification successes were 69% (top 20%) and 74% (bottom 20%) with R^2 = 0.69, MSE = 86.2, and a Cor = 0.515.

For the CV1 cross-validation scheme, M1 (**Supplementary Figure S6**) returned a classification success of 67% for the top 20% of the genotypes in fields and 70% for those with the poorest performance, with an R^2 = 0.64, MSE = 100.8, and a Cor = 0.359. With M2 the corresponding classification successes were 71 and 73%, with an R^2 = 0.7, MSE = 84.7, and Cor = 0.48. Similar values to those obtained with M2 were obtained with M3 for all the mentioned criteria. Finally, with M4 a slight reduction in the classification success was observed for the top 20% (67%) and the lowest 20% (71%) as compared to M2 and M3, with an R^2 = 0.66, MSE = 95.1, and Cor = 0.405.

When predicting already tested genotypes in untested environments (CV0), M1 returned a low classification success of the top and bottom 20% genotypes (25 and 29%, respectively), with an R^2 = 0.03, MSE = 291.8, and a Cor = 0.461. There, M2 and M3 returned similar results to those from M1 with a slight decrease in the MSE and a slight improvement of the weighted average correlation with M3 (0.488) (**Supplementary Figure S7**). The most promising model in this scenario was M4 which returned a classification success of the top and bottom 20% of 34 and 36%, respectively. It also returned the highest R^2 (0.11) and the smallest MSE (251.9) among all models leveraging the advantage of including soil type in interaction with molecular markers in the prediction models.

For the most complex prediction scenario CV00, the classification success rate, R^2, and Cor values were reduced across all models while the MSE increased simultaneously. M1 returned a classification success rate of the top and bottom 20% performing lines of 17 and 25%, respectively, with an R^2 = 0, MSE = 305, and a Cor = 0.24. In M2, M3, and M4, the average weighted correlation was reduced to 0.192, 0.227, and 0.231, respectively (**Supplementary Figure S8**). However, the classification success of the top and bottom performing lines was improved with M3, especially on the ability to detect the top 20% genotypes. For this model, the classification success was 29% for identifying the top 20% genotypes while it was 26% for screening out the worst-performing genotypes.

Overall Performance of Genotypes

Another approach used to assess the model performance was the overall performance of the genotypes. For this, within each environment, the phenotypic and predicted values of all genotypes were adjusted by their corresponding environmental mean (centered on zero) followed by the computation of the across environment mean for all lines. **Figures 1–4** display the classification success of the adjusted genotypes (observed and predicted values) marked by the advancement fate of each genotype including the advanced yield trial (AYT, gray), USDA Preliminary trials (USDA-UP, yellow), USDA Uniform trials (USDA-UT, orange), and Commercial Release (Release, blue). Detailed information on each stage of the breeding pipeline and selection criteria for line advancement were reported in Vieira and Chen (2021).

Figure 1 displays the results corresponding to the CV2 scenario. M1 returned a classification success of 76% for the top 20% of the predicted (adjusted) genotypes, and 79% success for the bottom 20% of the genotypes. In addition, the means of the predictions corresponding to the different advancement fate

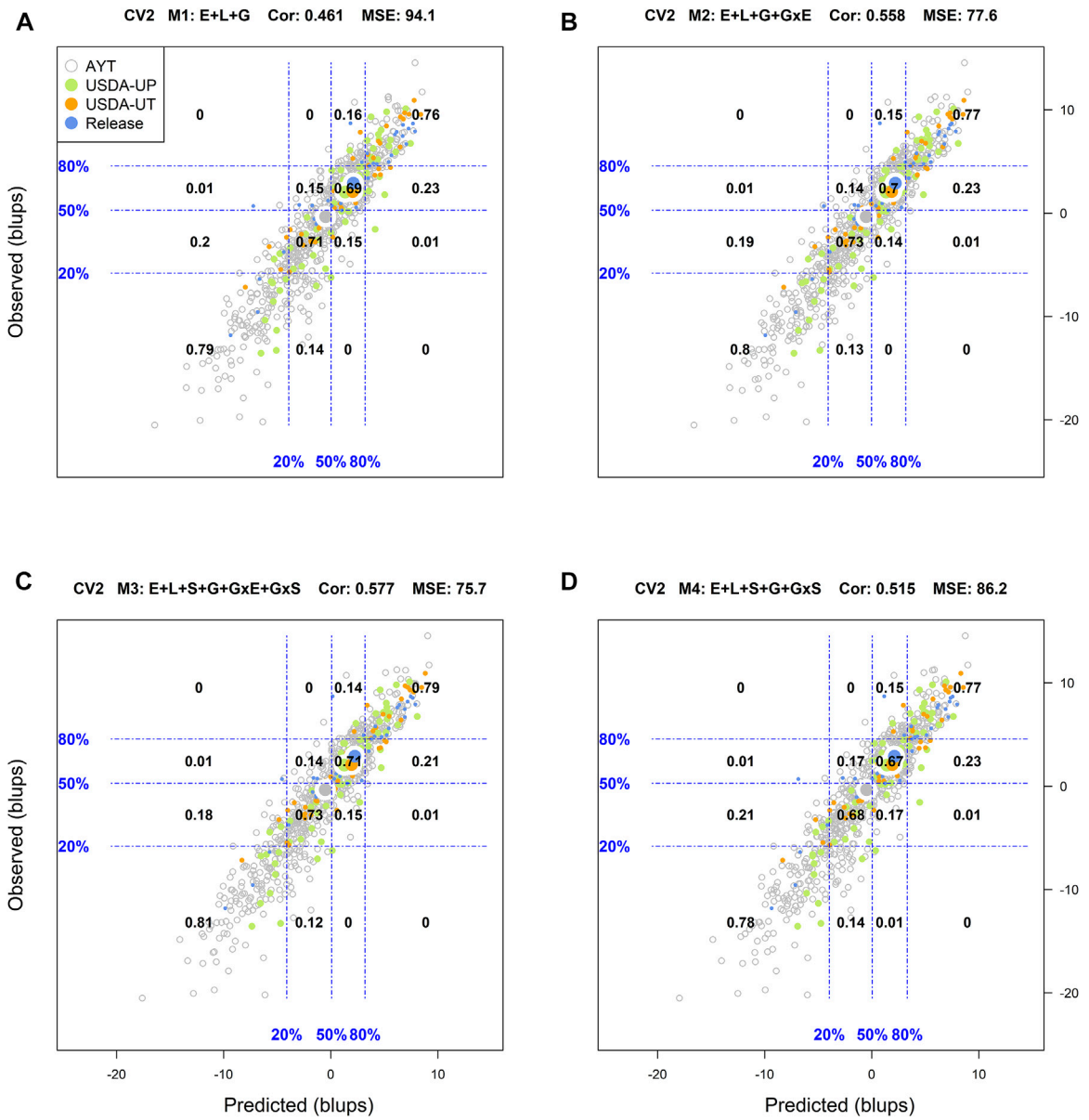


FIGURE 1 | Genotypic means (BLUP-centered on zero within environments) of observed versus predicted cross-validation predictions of four models (M1–M4) under the cross-validation scheme CV2, which mimics the incomplete field trial prediction scenario (predicting tested genotypes in observed environments). Models and terms are described in detail in the *Materials and Methods* section (**Eqs 1**–**4**). Horizontal and vertical dashed lines indicate the 20, 50, and 80% empirical percentiles corresponding to the genotypic means of observed and predicted values; the numbers inside the grid provide the conditional proportions observed on the *y*-axis for the different percentiles on the *x*-axis (e.g., out of the top 20% of the predicted values with M3 (panel C), 79% [top right] of these correspond to genotypes that showed a performance among the 20% across fields).

aligns with their counterpart based on phenotypes. There, the mean of the adjusted genotypes of the release (blue) group was superior followed by USDA-UT (orange), USDA-UP (yellow), and AYT (gray). Regarding M2 and M3, improvements in the classification accuracy were observed as compared to M1. With M3, a classification success of 76% was obtained for those genotypes in the top 20 and 79% for those in the lowest 20%. M4 returned intermediate results between M1 and M3 (**Figure 1**).

Similar to CV2, the corresponding results of CV1 are displayed in **Figure 2**. As expected, predicting new genotypes resulted in a significant reduction of the predictive ability of the models. With M1, the classification success of the top and bottom 20% of the predicted genotypes was 0.45, and 0.48, respectively. In this cross-validation scheme, the best results were shown in M2 with a classification success of 48% of the genotypes in the top 20 and 52% in the bottom 20%. Model M3 was the second-best model

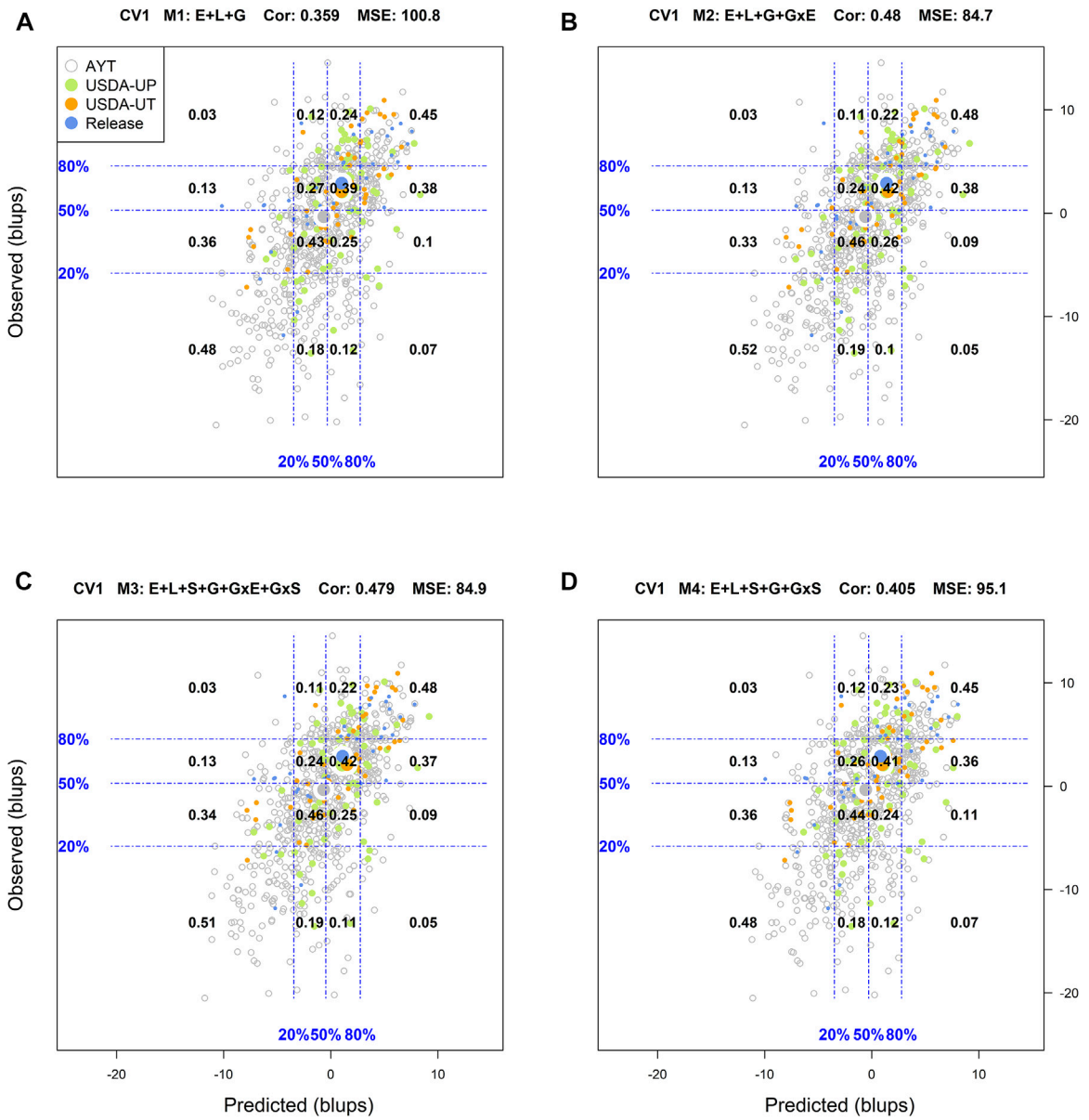


FIGURE 2 | Genotypic means (BLUP-centered on zero within environments) of observed versus predicted cross-validation predictions of four models (M1–M4) under the cross-validation scheme CV1, which mimics the prediction scenario of newly developed lines (predicting untested genotypes in observed environments). Models and terms are described in detail in the *Materials and Methods* section (**Eqs 1**–**4**). Horizontal and vertical dashed lines indicate the 20, 50, and 80% empirical percentiles corresponding to the observed and predicted values; the numbers inside the grid provide the conditional proportions observed on the *y*-axis for the different percentiles on the *x*-axis (e.g., out of the top 20% of the predicted values with M3 (panel C), 48% [top right] of these correspond to phenotypes that showed a performance among the 20% in fields).

with the corresponding values for top and bottom 20% of 48 and 51%, respectively.

Regarding the prediction of the overall performance of tested genotypes in untested environments (CV0), M1 returned a classification success of 51 and 52% for the top and bottom 20%, respectively. The best results predicting the top 20% of the genotypes were obtained with M3 (55%), while M1 was the best (52%) for the bottom 20%. Model 4 produced intermediated results and it was the most stable across the diagonal in the grid (i.e., including the other percentiles), whereas M2 returned the poorest performance.

Finally, for the most complex prediction scenario CV00, M1 returned a classification success of 35% for the top 20 and 36% for the bottom 20%. M1 was the most accurate model in classifying genotypes with the poorest performance. M4 outperformed this model in the identification of the superior genotypes with a success rate of 39%. The remaining models underperformed M1 in identifying

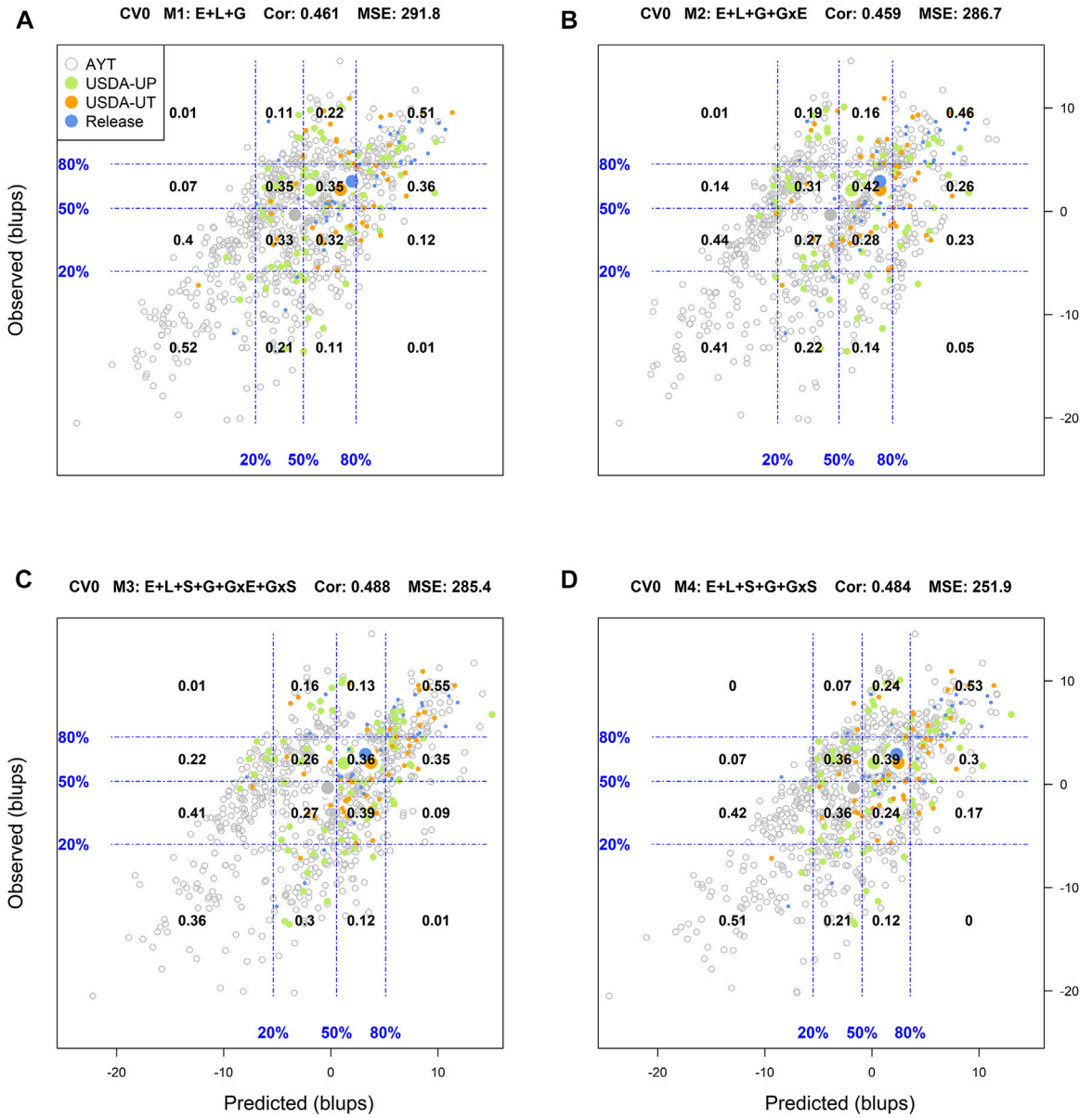


FIGURE 3 | Genotypic means (BLUP-centered on zero within environments) of observed versus predicted cross-validation predictions of four models (M1–M4) under the cross-validation scheme CV0, which mimics the prediction scenario of predicting in novel environments (predicting tested genotypes in unobserved environments). Models and terms are described in detail in the *Materials and Methods* section (**Eqs 1**–**4**). Horizontal and vertical dashed lines indicate the 20, 50, and 80% empirical percentiles corresponding to the observed and predicted values; the numbers inside the grid provide the conditional proportions observed on the *y*-axis for the different percentiles in the *x*-axis (e.g., out of the top 20% of the predicted values with M3 (panel C), 55% [top right] of these correspond to phenotypes that showed a performance among the 20% in fields).

genotypes in both extremes, where M3 was slightly superior to M4 in the bottom 20% (0.32 vs. 0.30).

DISCUSSION

As the fields of genomics and data analytics substantially evolved over the past decade, the concept of genomic selection applied to phenotypic prediction revolutionized commercial and public breeding programs by allowing plant breeders to predict the phenotype of interest in untested genotypes (Crossa et al., 2017; Vieira and Chen, 2021; Wartha and Lorenz, 2021). Genomic selection has covered multiple fronts of the breeder's equation maximizing the genetic gain in a given breeding cycle. For instance, a large component of a breeding cycle is allocated to

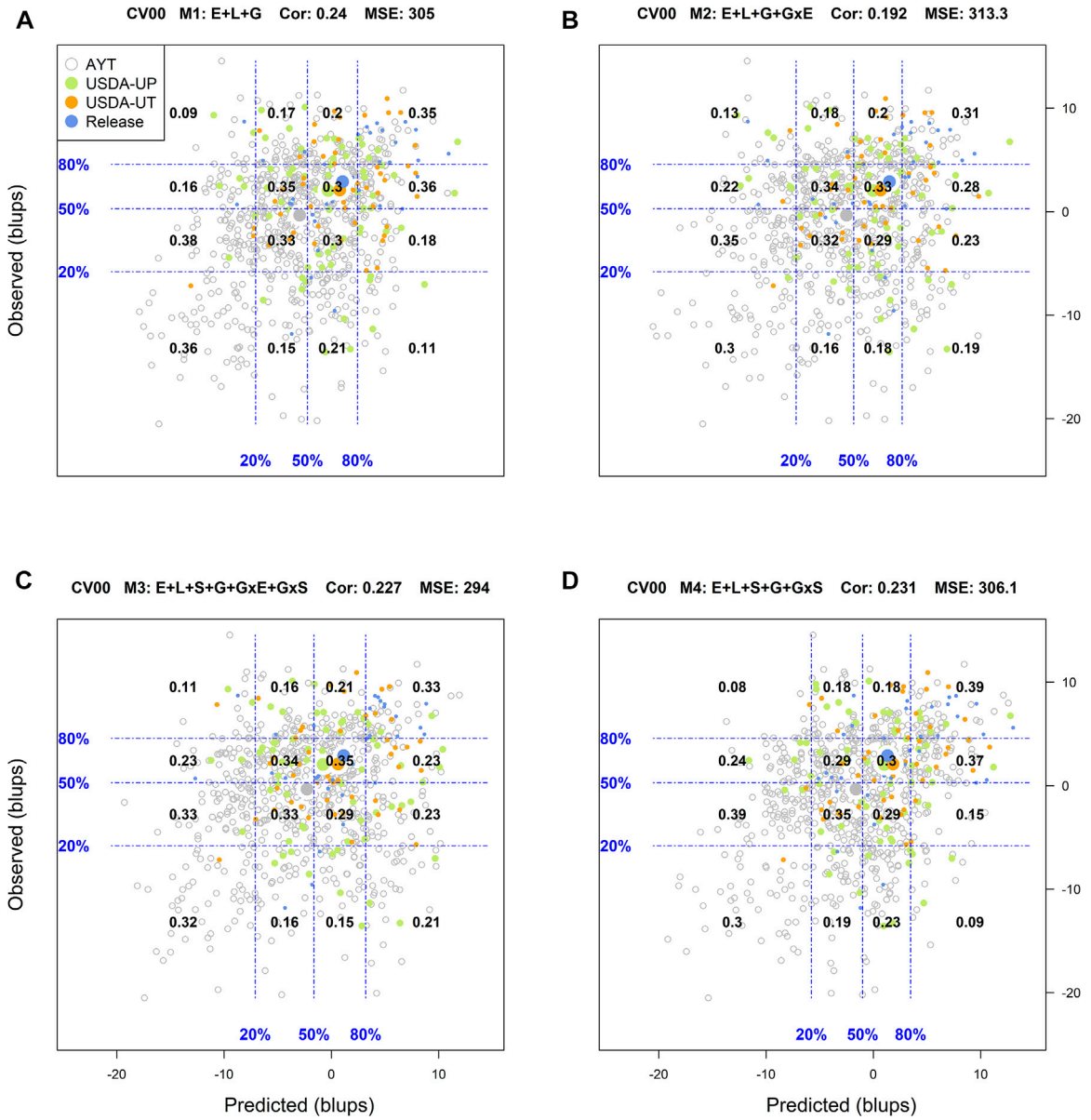


FIGURE 4 | Genotypic means (BLUP-centered on zero within environments) of observed versus predicted cross-validation predictions of four models (M1–M4) under the cross-validation scheme CV00, which mimics the prediction scenario of predicting newly developed lines in novel environments (predicting untested genotypes in unobserved environments). Models and terms are described in detail in the *Materials and Methods* section (**Eqs 1**–**4**). Horizontal and vertical dashed lines indicate the 20, 50, and 80% empirical percentiles corresponding to the observed and predicted values; the numbers inside the grid provide the conditional proportions observed on the *y*-axis for the different percentiles on the *x*-axis (e.g., out of the top 20% of the predicted values with M4 (panel D), 39% [top right] of these correspond to phenotypes that showed a performance among the 20% in fields).

progeny selection and preliminary yield trials of which the main objective is to characterize the genetic diversity present in a population of interest by evaluating a large number of genotypes for yield and overall agronomic traits. Genomic selection rises as a statistically powerful solution generating predicted values for unobserved genotypes, allowing plant breeders to shorten the breeding cycle and significantly minimize the costs associated with extensive field trials (Vieira and Chen, 2021; Wartha and Lorenz, 2021). Up to this date, however, the wide and large-scale implementation of genomic selection across plant breeding programs still faces challenges and drawbacks.

It is well-known that the expression of a phenotype is a function of the genotype, the environment, and the interaction between the genotype and environment (G×E) providing the relative performance of genotypes across different environments

(Kang, 1997; de Leon et al., 2016). The differential response of genotypes across environments for a given phenotype of interest guide critical decisions in a plant breeding program, including the selection and advancement of genotypes as well as overall logistics and allocation of resources for multi-environment trials (Hill, 1975; Cooper and DeLacy, 1994; Kang, 1997; de Leon et al., 2016). Yield is a highly complex and quantitative trait regulated by numerous large and small-effect genes, of which its expression is immensely dependable on the genotype interaction with various components of the environment including pathogens (Rincker et al., 2017; Vieira et al., 2021), pests (Haile et al., 1998; Rocha et al., 2015), weeds (Oerke, 2006; Soltani et al., 2017), temperature, light, and precipitation (Runge and Odell, 1960; Goldblum, 2009; Alsajri et al., 2020), and soil-derived factors (Cox et al., 2003; Kaspar et al., 2004; Anthony et al., 2012). Thus, a practical and accurate implementation of genomic selection for yield relies on understanding and accounting for the interaction of molecular markers with the environment and/or its multiple components.

In this research, we aimed to expand the reaction norm model initially proposed by Jarquin et al. (2014b) which accounts for the interaction between molecular markers and the environment through covariance structures. Here, we investigated the potential of incorporating soil-derived covariates to enhance the predictive ability of yield across multiple cross-validation scenarios simulating progeny testing and line selection. A straightforward approach to examine the relative contribution of each model term is through the computation of variance components. Across environments, we observed that the addition of the G×S interaction in M3 substantially decreased the amount of variability captured by both the environment (−30.4%) and residual (−39.2%) terms as compared to the conventional GBLUP model (M1). When compared to the reaction norm model (M2), the addition of G×S equally reduced the amount of variability captured by the environment (−30.4%). Within environments, a larger reduction in variability captured by the residual term was observed in M3. Interestingly, the addition of the G×S term in M3 reduced the variability captured by the residual term by roughly 60 and 30% when compared to M1 and M2, respectively. The addition of soil-derived covariates seems to structure/dissect the environment term revealing components of the environment that could potentially enhance or hinder the performance of a model. This creates opportunities to explore more complex and readily available environmental components, which through covariance structures, could allow the borrowing of information across environmental components enhancing the predictive ability in challenging cross-validation scenarios. For instance, the amount of variability explained by both G×E (20.0%) and G×S (17.5%) in M3 shows that the inclusion of these terms increases the proportion of variance accounted for by the model, and therefore, it can enhance its predictive ability.

In regards to the predictive ability of each model across the proposed cross-validation scenarios, M3 outperformed the other models in CV2, CV1, and CV0. The conventional genomic selection model (GBLUP, M1) was the best in CV00. In the incomplete field trial scenario (CV2), M3 substantially outperformed M1 (25.1%). The ability of the covariance structures to borrow information from already observed genotypes in tested environments increased the model's performance. In this case, the addition of G×S provides a slight edge over the reaction norm model (M2, 4%), highlighting the benefit of accounting for possible interactions between markers and soil types in overall prediction accuracy. An alternative methodology to assess the practical accuracy of the model consisting of empirical percentiles of the predicted and observed values was proposed by Jarquin et al. (2014b). Here, we observed that with M3 the classification accuracy for the top and bottom 20% percentile was 0.79 and 0.81, respectively. This represents approximately a 4% increment in classification accuracy as compared to M1. All four models flawlessly avoided misclassifying a top 20% percentile genotype as a bottom 20% percentile and *vice versa*, encouraging the practical applications of genomic prediction for line selection throughout the breeding pipeline. These results provide an opportunity to reconsider the experimental design in field trials, including the number of replications as well as overall resource allocation in multi-environment field trials. The prediction models can precisely discard inferior genotypes with nearly full confidence reducing the need for extensive preliminary field trials.

In CV1, M2 and M3 performed approximately 34% better than M1. In this cross-validation scenario, the genotypes are untested but the environment has been already observed with a different set of genotypes. The covariance structures allow the borrowing of information from previously observed genotypes, especially the main effects of molecular markers and the interaction between the markers and the environment. However, the structuring of the environment through the addition of G×S did not yield any advantages in prediction accuracy as compared to M2. Jarquin et al. (2021) observed similar results in CV1 when including the interactions using only weather data. This was attributed to G×E sufficiently capturing the similarities among pairs of environments leaving limited variance left to be explained by G×S. In the cross-validation scenario aiming to predict the yield of already tested genotypes in unobserved environments (CV0), M3 outperformed M1 and M2 by roughly 6%. These results provide an opportunity to explore alternative multi-environment testing and resource allocation throughout line selection in a breeding program. For instance, by leveraging the information of molecular markers of a different set of observed genotypes and known environments, plant breeders may be able to simulate multiple yield trials in a given growing season substantially increasing statistical power and confidence in line selection and advancement without necessarily increasing the investment in field operations. Similarly, the results from CV0 support both the reduction in the number of physical locations and the simulation of yield trials across diverse untested environments. This can substantially reduce the overall cost of a breeding pipeline while simultaneously enhancing statistical power and confidence in identifying genotypes with superior yield and overall adaptability.

In the most challenging cross-validation scenario considering untested genotypes in unobserved environments (CV00), M3 substantially outperformed M2 (19%), whereas M4 slightly outperformed M3 by 2%. Here, it highlighted the main advantage of incorporating G×S and S in the model. As previously discussed, the soil texture is generally constant across years and readily available before the growing season whereas the environment is often unfeasible to be accurately predicted prior to the growing season. Therefore, the borrowing of information from both soil covariate and molecular markers (in interaction) resulted in higher prediction accuracies as compared to M2. Although M1 yielded the highest prediction accuracy among the four models, M4 showed superior classification accuracy in the top 20% empirical percentiles (12% advantage over M1). By considering the advancement fate of the genotypes included in this analysis (AYT, USDA-UP, USDA-UT, and Release), nearly all the genotypes commercially released are concentrated in the top 20% and 50% observed and predicted empirical percentiles. This shows that, although the model may misclassify the empirical percentiles and/or show relatively low prediction accuracy, it does not negatively affect the identification and selection of the very best genotypes that will eventually be commercially released. These results support the modernization of a conventional breeding pipeline by precisely eliminating inferior genotypes prior to any field testing. Nearly 2 years of a conventional breeding pipeline is devoted to the assessment of the entire pool of genotypes representing a breeding cycle (Vieira and Chen, 2021). After reaching desired homozygosity ($F_{4:5}$), a large number of genotypes are tested in progeny rows to visually evaluate their yield potential and overall agronomic traits. Selected genotypes, often consisting of many inferior genotypes mistakenly selected by subjective standards, are then tested in preliminary multi-environment yield trials. As seemed in CV00, the implementation of genomic selection has the potential for eliminating 2 years of extensive field testing by predicting the breeding values of untested genotypes. Thus, the wide implementation of genomic selection throughout a breeding pipeline holds promising improvements in cost efficiency, shortening the duration of the cycle, and overall genetic gain.

CONCLUSION

The increasing availability of high-dimensional genomic data has allowed breeding programs to implement genomic selection to optimize the efficiency of a given breeding pipeline. Although widely adopted in commercial programs, the application of genomic selection in the public sector still faces limitations associated with costs, data availability, and technical support. In this research, we investigated the potential of incorporating soil texture and its interaction with molecular markers through covariance structures to increase prediction accuracy. As an approach to structuring the environmental term, the inclusion of G×S was shown to benefit the predictive ability of the models across multiple cross-validation scenarios. It is hypothesized that the availability of the soil texture prior to the growing season may have been essential to maximizing the functionality of covariance structures, particularly in scenarios with untested genotypes in untested environments. In addition, we demonstrated the applications of genomic selection across multiple stages of a breeding pipeline through four different cross-validation scenarios. In both progeny testing and line selection, we highlight the potential of genomic selection to optimize the efficiency of a soybean breeding program and discuss the opportunities to reconsider field experimental designs, allocation of resources, and reduction of preliminary field trials. Further studies considering covariates that are readily available before the growing season are encouraged to better understand the effect of the environment and enhance predictive ability. In addition, alternative metrics to assess the true potential and applicability of a model should be investigated to embolden the wide implementation of genomic selection in the public sector.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and link to the data can be found as follows: Dryad; https://doi.org/10.5061/dryad.z8w9ghxf9.

AUTHOR CONTRIBUTIONS

PC and DJ contributed to conception and design of the study. CV collected the data and organized the database. RP and DJ performed the statistical analysis. CV, RP, and DJ wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank the University of Missouri–Fisher Delta Research, Extension, and Education Center soybean breeding team for their hard work and support in preparing and 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.2022.905824/full#supplementary-material

REFERENCES

Ainsworth, E. A., Yendrek, C. R., Skoneczka, J. A., and Long, S. P. (2012). Accelerating Yield Potential in Soybean: Potential Targets for Biotechnological Improvement. *Plant Cell Environ.* 35, 38–52. doi:10.1111/j.1365-3040.2011.02378.x

Alsajri, F. A., Wijewardana, C., Irby, J. T., Bellaloui, N., Krutz, L. J., Golden, B., et al. (2020). Developing Functional Relationships between Temperature and Soybean Yield and Seed Quality. *Agron. J.* 112, 194–204. doi:10.1002/agj2.20034

Anthony, P., Malzer, G., Sparrow, S., and Zhang, M. (2012). Soybean Yield and Quality in Relation to Soil Properties. *Agron. J.* 104, 1443–1458. doi:10.2134/agronj2012.0095

Bernardo, R. (1994). Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. *Crop Sci.* 34, 20–25. doi:10.2135/cropsci1994.0011183X003400010003x

Concibido, V., la Vallee, B., Mclaird, P., Pineda, N., Meyer, J., Hummel, L., et al. (2003). Introgression of a Quantitative Trait Locus for Yield from *Glycine Soja* into Commercial Soybean Cultivars. *Theor. Appl. Genet.* 106, 575–582. doi:10.1007/s00122-002-1071-5

Cooper, M., and DeLacy, I. H. (1994). Relationships Among Analytical Methods Used to Study Genotypic Variation and Genotype-By-Environment Interaction in Plant Breeding Multi-Environment Experiments. *Theor. Appl. Genet.* 88, 561–572. doi:10.1007/BF01240919

Cox, M. S., Gerard, P. D., Wardlaw, M. C., and Abshire, M. J. (2003). Variability of Selected Soil Properties and Their Relationships with Soybean Yield. *Soil Sci. Soc. Am. J.* 67, 1296–1302. doi:10.2136/sssaj2003.1296

Crossa, J., Yang, R.-C., and Cornelius, P. L. (2004). Studying Crossover Genotype × Environment Interaction Using Linear-Bilinear Models and Mixed Models. *J. Agric. Biol. Environ. Statistics* 9 (3), 362–380. doi:10.1198/108571104X4423

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 Leon, N., Jannink, J.-L., Edwards, J. W., and Kaeppler, S. M. (2016). Introduction to a Special Issue on Genotype by Environment Interaction. *Crop Sci.* 56, 2081–2089. doi:10.2135/cropsci2016.07.0002in

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

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of Development Descriptions for Soybeans, Glycine Max (L.) Merrill. *Crop Sci.* 11, 929–931. doi:10.2135/cropsci1971.0011183X001100060051x

Gale, F., Valdes, C., and Ash, M. (2019). *Interdependence of China, United States, and Brazil in Soybean Trade*. Washington, D.C: Economic Research Service - USDA, 1–48.

Goldblum, D. (2009). Sensitivity of Corn and Soybean Yield in Illinois to Air Temperature and Precipitation: The Potential Impact of Future Climate Change. *Phys. Geogr.* 30, 27–42. doi:10.2747/0272-3646.30.1.27

Haile, F. J., Higley, L. G., and Specht, J. E. (1998). Soybean Cultivars and Insect Defoliation: Yield Loss and Economic Injury Levels. *Agron. J.* 90, 344–352. doi:10.2134/agronj1998.00021962009000030006x

Hill, J. (1975). Genotype-environment Interaction - A Challenge for Plant Breeding. *J. Agric. Sci.* 85, 477–493. doi:10.1017/S0021859600062365

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014a). Genotyping by Sequencing for Genomic Prediction in a Soybean Breeding Population. *BMC Genomics* 15, 740–810. doi:10.1186/1471-2164-15-740

Jarquín, D., Crossa, J., Lacaze, X., du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014b). 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

Jarquin, D., de Leon, N., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I., et al. (2021). Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield within the Genomes to Fields Maize Project. *Front. Genet.* 11, 1–11. doi:10.3389/fgene.2020.592769

Kang, M. S. (1997). "Using Genotype-By-Environment Interaction for Crop Cultivar Development," in *Advances in Agronomy*. Editor D. Sparks (Cambridge, MA: Academic Press), 199–252. doi:10.1016/S0065-2113(08)60569-6

Kaspar, T. C., Pulido, D. J., Fenton, T. E., Colvin, T. S., Karlen, D. L., Jaynes, D. B., et al. (2004). Relationship of Corn and Soybean Yield to Soil and Terrain Properties. *Agron. J.* 96, 700–709. doi:10.2134/agronj2004.0700

Liu, X., Jin, J., Wang, G., and Herbert, S. J. (2008). Soybean Yield Physiology and Development of High-Yielding Practices in Northeast China. *Field Crops Res.* 105, 157–171. doi:10.1016/j.fcr.2007.09.003

Ma, Y., Reif, J. C., Jiang, Y., Wen, Z., Wang, D., Liu, Z., et al. (2016). Potential of Marker Selection to Increase Prediction Accuracy of Genomic Selection in Soybean (Glycine Max L.). *Mol. Breed.* 36 (8), 113. doi:10.1007/s11032-016-0504-9

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. doi:10.1093/genetics/157.4.1819

Moreira, F. F., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2019). Improving the Efficiency of Soybean Breeding with High-Throughput Canopy Phenotyping. *Plant Methods* 15, 1–9. doi:10.1186/s13007-019-0519-4

Moreira, F. F., Oliveira, H. R., Volenec, J. J., Rainey, K. M., and Brito, L. F. (2020). Integrating High-Throughput Phenotyping and Statistical Genomic Methods to Genetically Improve Longitudinal Traits in Crops. *Front. Plant Sci.* 11, 1–18. doi:10.3389/fpls.2020.00681

Mourtzinis, S., Rattalino Edreira, J. I., Grassini, P., Roth, A. C., Casteel, S. N., Ciampitti, I. A., et al. (2018). Sifting and Winnowing: Analysis of Farmer Field Data for Soybean in the US North-Central Region. *Field Crops Res.* 221, 130–141. doi:10.1016/j.fcr.2018.02.024

Oerke, E.-C. (2006). Crop Losses to Pests. *J. Agric. Sci.* 144, 31–43. doi:10.1017/S0021859605005708

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2019). Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. *Plant Phenomics* 2019, 1–15. doi:10.34133/2019/5809404

Pathan, M. S., Lee, J.-D., Shannon, J. G., and Nguyen, H. T. (2007). "Recent Advances in Breeding for Drought and Salt Stress Tolerance in Soybean," in *Advances in Molecular Breeding toward Drought and Salt Tolerant Crops*. Editors M. A. Jenks, P. M. Hasegawa, and S. M. Jain (Dordrecht: Springer Netherlands), 739–773. doi:10.1007/978-1-4020-5578-2_30

Patil, G., Chaudhary, J., Vuong, T. D., Jenkins, B., Qiu, D., Kadam, S., et al. (2017). Development of SNP Genotyping Assays for Seed Composition Traits in Soybean. *Int. J. Plant Genomics* 2017, 1–12. doi:10.1155/2017/6572969

Persa, R., Iwata, H., and Jarquin, D. (2020). Use of Family Structure Information in Interaction with Environments for Leveraging Genomic Prediction Models. *Crop J.* 8, 843–854. doi:10.1016/j.cj.2020.06.004

Pham, A.-T., Lee, J.-D., Shannon, J. G., and Bilyeu, K. D. (2010). Mutant Alleles of FAD2-1A and FAD2-1Bcombine to Produce Soybeans with the High Oleic Acid Seed Oil Trait. *BMC Plant Biol.* 10, 195. doi:10.1186/1471-2229-10-195

Pham, A.-T., McNally, K., Abdel-Haleem, H., Roger Boerma, H., and Li, Z. (2013). Fine Mapping and Identification of Candidate Genes Controlling the Resistance to Southern Root-Knot Nematode in PI 96354. *Theor. Appl. Genet.* 126, 1825–1838. doi:10.1007/s00122-013-2095-8

Rincker, K., Cary, T., and Diers, B. W. (2017). Impact of Soybean Cyst Nematode Resistance on Soybean Yield. *Crop Sci.* 57, 1373–1382. doi:10.2135/cropsci2016.07.0628

Rocha, F., Vieira, C. C., Ferreira, M. C., Oliveira, K. C., Moreira, F. F., and Pinheiro, J. B. (2015). Selection of Soybean Lines Exhibiting Resistance to Stink Bug Complex in Distinct Environments. *Food Energy Secur* 4, 133–143. doi:10.1002/fes3.57

Runge, E. C. A., and Odell, R. T. (1960). The Relation between Precipitation, Temperature, and the Yield of Soybeans on the Agronomy South Farm, Urbana, Illinois. *Agron. J.* 52, 245–247. doi:10.2134/agronj1960.00021962005200050001x

Salado-Navarro, L. R., Sinclair, T. R., and Hinson, K. (1993). Changes in Yield and Seed Growth Traits in Soybean Cultivars Released in the Southern USA from 1945 to 1983. *Crop Sci.* 33, 1204–1209. doi:10.2135/cropsci1993.0011183X003300060019x

Shi, Z., Liu, S., Noe, J., Arelli, P., Meksem, K., and Li, Z. (2015). SNP Identification and Marker Assay Development for High-Throughput Selection of Soybean Cyst Nematode Resistance. *BMC Genomics* 16, 314. doi:10.1186/s12864-015-1531-3

Soltani, N., DIlle, J. A., Burke, I. C., Everman, W. J., Vangessel, M. J., Davis, V. M., et al. (2017). Perspectives on Potential Soybean Yield Losses from Weeds in North America. *Weed Technol.* 31, 148–154. doi:10.1017/wet.2016.2

Song, Q., Hyten, D. L., Jia, G., Quigley, C. v., Fickus, E. W., Nelson, R. L., et al. (2013). Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE* 8, e54985–12. doi:10.1371/journal.pone.0054985

Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., et al. (2020). Soybean BARCSoySNP6K: An Assay for Soybean Genetics and Breeding Research. *Plant J.* 104, 800–811. doi:10.1111/tpj.14960

Specht, J. E., and Williams, J. H. (2015). *Contribution of Genetic Technology to Soybean Productivity - Retrospect and Prospect*. Madison, WI: American Society of Agronomy and Crop Science Society of America, 49–74. doi:10.2135/cssaspecpub7.c3

Specht, J. E., Hume, D. J., and Kumudini, S. v. (1999). Soybean Yield Potential-A Genetic and Physiological Perspective. *Crop Sci.* 39, 1560–1570. doi:10.2135/cropsci1999.3961560x

Stewart-Brown, B. B., Song, Q., Vaughn, J. N., and Li, Z. (2019). Genomic Selection for Yield and Seed Composition Traits within an Applied Soybean Breeding Program. *G3 Genes Genomes Genetics* 9 (7), 2253–2265. doi:10.1534/g3.118.200917

Tiezzi, F., de los Campos, G., Parker Gaddis, K. L., and Maltecca, C. (2017). Genotype by Environment (Climate) Interaction Improves Genomic Prediction for Production Traits in US Holstein Cattle. *J. Dairy Sci.* 100 (3), 2042–2056. doi:10.3168/jds.2016-11543

United States Department of Agriculture (2002). *Major Oilseeds: World Supply and Distribution*. Washington, DC. Available at: https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/pk02cb11m/6395w746k/oilseed-trade-01-01-2002.pdf (Accessed February 25, 2022).

United States Department of Agriculture (2013). Official Soil Series Descriptions and Series Classification: Tiptonville Series. Available at: https://soilseries.sc.egov.usda.gov/OSD_Docs/T/TIPTONVILLE.html (Accessed May 5, 2022).

United States Department of Agriculture (2018a). Official Soil Series Descriptions and Series Classification: Sharkey Series. Available at: https://soilseries.sc.egov.usda.gov/OSD_Docs/S/SHARKEY.html (Accessed May 5, 2022).

United States Department of Agriculture (2018b). Official Soil Series Descriptions and Series Classification: Malden Series. Available at: https://soilseries.sc.egov.usda.gov/OSD_Docs/M/MALDEN.html (Accessed May 5, 2022).

United States Department of Agriculture (2022). *Oilseeds: World Markets and Trade*. Washington, DC. Available at: https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf (Accessed February 25, 2022).

Vieira, C. C., and Chen, P. (2021). The Numbers Game of Soybean Breeding in the United States. *Crop Breed. Appl. Biotechnol.* 21, 387521–387531. doi:10.1590/1984-70332021v21sa23

Canella Vieira, C., Chen, P., Usovsky, M., Vuong, T., Howland, A. D., Nguyen, H. T., et al. (2021). A Major Quantitative Trait Locus Resistant to Southern Root-knot Nematode Sustains Soybean Yield under Nematode Pressure. *Crop Sci.* 61, 1773–1782. doi:10.1002/csc2.20443

Voldeng, H. D., Cober, E. R., Hume, D. J., Gillard, C., and Morrison, M. J. (1997). Fifty-Eight Years of Genetic Improvement of Short-Season Soybean Cultivars in Canada. *Crop Sci.* 37, 428–431. doi:10.2135/cropsci1997.0011183X003700020020x

Wartha, C. A., and Lorenz, A. J. (2021). Implementation of Genomic Selection in Public-Sector Plant Breeding Programs: Current Status and Opportunities. *Crop Breed. Appl. Biotechnol.* 21, 1–19. doi:10.1590/1984-70332021v21sa28

Widener, S., Graef, G., Lipka, A. E., and Jarquin, D. (2021). An Assessment of the Factors Influencing the Prediction Accuracy of Genomic Prediction Models across Multiple Environments. *Front. Genet.* 12, 689319. doi:10.3389/fgene.2021.689319

Wu, C., Mozzoni, L. A., Moseley, D., Hummer, W., Ye, H., Chen, P., et al. (2020). Genome-wide Association Mapping of Flooding Tolerance in Soybean. *Mol. Breed.* 40, 4. doi:10.1007/s11032-019-1086-0

Xavier, A., Muir, W. M., and Rainey, K. M. (2016). Assessing Predictive Properties of Genome-wide Selection in Soybeans. *G3 Genes Genomes Genetics* 6 (8), 2611–2616. doi:10.1534/g3.116.032268

Zhou, J., Beche, E., Vieira, C. C., Yungbluth, D., Zhou, J., Scaboo, A., et al. (2022). Improve Soybean Variety Selection Accuracy Using UAV-Based High-Throughput Phenotyping Technology. *Front. Plant Sci.* 12, 768742. doi:10.3389/fpls.2021.768742

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 © 2022 Canella Vieira, Persa, Chen and Jarquin. 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 | Frontiers in Genetics

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fgene.2022.1000440

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Abe Shegro Gerrano,
Agricultural Research Council of South Africa, South Africa
Sheikh Mansoor,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, India

*CORRESPONDENCE
Aditya Pratap,
Aditya.Pratap@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 22 July 2022
ACCEPTED 06 October 2022
PUBLISHED 03 November 2022

CITATION
Kumari G, Shanmugavadivel PS, Lavanya GR, Tiwari P, Singh D, Gore PG, Tripathi K, Madhavan Nair R, Gupta S and Pratap A (2022), Association mapping for important agronomic traits in wild and cultivated Vigna species using cross-species and cross-genera simple sequence repeat markers. *Front. Genet.* 13:1000440. doi: 10.3389/fgene.2022.1000440

COPYRIGHT
© 2022 Kumari, Shanmugavadivel, Lavanya, Tiwari, Singh, Gore, Tripathi, Madhavan Nair, Gupta and Pratap. 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.

Association mapping for important agronomic traits in wild and cultivated *Vigna* species using cross-species and cross-genera simple sequence repeat markers

Gita Kumari[1†], P. S. Shanmugavadivel[1], G. Roopa Lavanya[2], Pravin Tiwari[1], Dharmpal Singh[1], P. G. Gore[3], Kuldeep Tripathi[3], Ramakrishnan Madhavan Nair[4], Sanjeev Gupta[1] and Aditya Pratap[1*†]

[1]ICAR-Indian Institute of Pulses Research, Kanpur, India, [2]Sam Higginbottom University of Agricultural Technology and Sciences, Prayagraj, India, [3]ICAR-National Bureau of Plant Genetic Resources, New Delhi, India, [4]World Vegetable Centre, ICRISAT Campus, Patancheru, India

The genus *Vigna* is an agronomically important taxon, with many of its species inhabiting a wide range of environments and offering numerous useful genes for the improvement of the cultivated types. The present study aimed to detect the genomic regions associated with yield-attributing traits by genome-wide association mapping. A diverse panel of 98 wild and cultivated *Vigna* accessions (acc.) belonging to 13 different species was evaluated for yield and related traits during the *kharif* season of 2017 and 2018. The panel was also genotyped using 92 cross-genera and cross-species simple sequence repeat markers to study the population genetic structure and useful market-trait associations. The PCA and trait correlation established relationships amongst the traits during both seasons while 100-seed weight (HSW) had a positive correlation with pod length (PL), and days to first flowering (DFF) with days to maturity (DM). The population genetic structure analysis grouped different acc. into three genetically distinct sub-populations with SP-1 comprising 34 acc., SP-2 (24 acc.), and SP-3 (33 acc.) and one admixture group (7 acc.). Mixed linear model analysis revealed an association of 13 markers, namely, VR018, VR039, VR022, CEDG033, GMES0337, MBSSR008, CEDG220, VM27, CP1225, CP08695, CEDG100, CEDG008, and CEDG096A with nine traits. Seven of the aforementioned markers, namely, VR018 for plant height (PH) and terminal leaflet length (TLL), VR022 for HSW and pod length (PL), CEDG033 for DFF and DM, MBSSR008 for DFF and DM, CP1225 for CC at 30 days (CC30), DFF and DM, CEDG100 for PH and terminal leaflet length (TLL), and CEDG096A for CC30 and chlorophyll content at 45 days were associated with multiple traits. The marker CEDG100, associated with HSW, PH, and TLL, is co-localized in gene-encoding histone–lysine N-methyltransferase *ATX5*. Similarly, VR22, associated with PL and HSW, is co-located in gene-

encoding *SHOOT GRAVITROPISM 5* in mungbean. These associations may be highly useful for marker-assisted genetic improvement of mungbean and other related *Vigna* species.

KEYWORDS

***Vigna*, *V. radiata*, cross-species SSR, genetic diversity, association study**

1 Introduction

The genus *Vigna* is an important taxon playing a prominent role in food and nutritional security and environmental sustainability. Several members of this genus are known to inhabit a wide range of agro-ecological regions across the globe including farmlands in tropical and subtropical regions, pastures, sandy beaches, mountain cliffs, roadsides, and field bunds. A few species are also found inhabiting extreme climates, thereby possessing special adaptive genes (Takahashi et al., 2016); therefore, these can serve as a potential source for yield- and adaptation-related traits (Pratap et al., 2015; Pratap, 2021). Most members of this genus are diploid with 2n = 2x = 22 having an exception of *V. glabrescens* which is a tetraploid (2n = 4x = 44). This genus is characterized by five sub-genera, namely, *Haydonia*, *Lasiospron*, *Ceratropis*, *Plectotropis*, and *Vigna*. Among these, the crop species have been developed from the later three sub-genera. *Ceratotropis*, also known as the Asiatic *Vigna*, is the most important in terms of most of the agronomic species categorized in this group. The agronomic species in the genus *Vigna* are autogamous and cross-pollination does not occur in nature.

India has been bestowed with an abundance of genetic diversity of *Vigna* including its cultivated, wild, and weedy types (Bisht et al., 2005; Pratap et al., 2015). It is the largest producer, processor, consumer, and importer of the two most important members of this group, the mungbean (*V. radiata* L. emWilczek) and the urdbean (*V. mungo* L. Hepper). The country is also considered as the region of the first domestication of important cultivated *Vigna* species including mungbean, urdbean (Baudoin and Marechal, 1988), ricebean (de Candolle, 1886), and mothbean (Smartt, 1985), while the progenitors of mungbean and urdbean, namely, *V. sublobata* and *V. sylvestris*, are found in abundance in cultivated as well as waste lands (Chandel et al., 1984; Lawn and Cottrell, 1988). Fortunately, the large variability of *Vigna* species spread across the Indian subcontinent has been collected extensively and conserved in the National Gene Banks at the ICAR-National Bureau of Plant Genetic Resources, New Delhi (ICAR-NBPGR), and ICAR-Indian Institute of Pulses Research, Kanpur (ICAR-IIPR). Most of these accessions have been studied and characterized at the morphological and molecular levels (Pratap et al., 2015; Pratap et al., 2017; Gore et al., 2019; Pratap et al., 2021b; Gore et al., 2021; Kumari et al., 2021), and many of them including *V. radiata*, *V. mungo*, *V. umbellata*, *V. sylvestris*, and *V. trilobata* have been deployed in hybridization programs for the genetic improvement of the cultivated types. A few of the wild relatives of *Vigna* are also available for neo-domestication (Gore et al., 2019). Some of them are valued as forage, cover, and green manure crops (Kumari et al., 2021). Many species also confer numerous valuable genes for agronomic, stress resistance, and seed quality traits (Pratap et al., 2015; Douglas et al., 2020; Pratap et al., 2020), namely, resistance/tolerance to drought and waterlogging (Bisht et al., 2005; Tomooka et al., 2006), soil salinity (Chankaew et al., 2014; Yoshida et al., 2016), heat and cold stress (HanumanthaRao et al., 2016), acidic or alkaline soils (Soares et al., 2014), bruchids (Aidbhavi et al., 2021), Cercospora leaf spot (Chankaew et al., 2011; Singh et al., 2017), and yellow mosaic disease [for a review, please see (Singh et al., 2020)] in addition to superior agronomic traits and photo-thermo period insensitivity (Pratap et al., 2014b; Basu et al., 2019). Extensive efforts have been undertaken for their genetic improvement and methods such as hybridization, selection, and mutation have been abundantly deployed in addition to pre-breeding and distant hybridization (Singh et al., 2017; Pratap et al., 2021b).

Information on the diversity at the genetic level, population genetic structure, and marker–trait association for useful traits provides useful information for deploying marker-assisted breeding for targeted and time-bound genetic improvement of crop plants. Information on the association of these traits with seed yield is of utmost importance to undertake a selective breeding program toward developing a desirable combination of yield-contributing traits. In this direction, the genome-wide association study (GWAS) is a high-resolution method for genetic mapping of traits using existing germplasm and their phenotypic information for the trait concerned (Flint-Garcia et al., 2003). It also helps us understand the genetic basis of complex traits and allows studying a wide range of alleles at each locus and the identification of useful marker–trait associations at the whole-genome level in addition to the identification of elite alleles for significantly associated loci (Rohilla et al., 2021). The higher mapping resolution of traits from association mapping provides an opportunity for the adoption of MAS in crop breeding programs (Mackay and Powell, 2007). Nonetheless, very few reports are available which document genetic diversity in *Vigna* species at the molecular level, while no report is available on useful marker–trait associations in wild and diverse *Vigna* accessions. The present study aimed to evaluate the genetic diversity and marker–trait associations in

an extensive panel of wild and cultivated *Vigna* species genotyped with cross-genera and cross-species SSR markers so as to identify useful marker–trait associations which could be effectively deployed not only in mungbean improvement programs but also in other related *Vigna* crops such as urdbean which lack genomic information.

2 Materials and methods

2.1 Plant materials

The plant materials for this study comprised 98 genotypes belonging to cultivated (13 accessions) and wild (85 accessions) *Vigna* species (Supplementary Table S1). Species-wise, the accessions (acc.) belonged to *V. umbellata* (16 acc.), *V. trilobata* (20 acc.), *V. mungo* var. *mungo* (9 acc.), *V. radiata* (6 acc.), *V. radiata* var. *radiata* (6 acc.), *V. radiata* var. *sublobata* (6 acc.), *V. silvestris* (4 acc.), *V. unguiculata* (4 acc.), *V. dalzelliana* and *V. trinervia* var. *bourneae* (3 acc. each), *V. radiata* var. *setulosa*, *V. stipulaceae* and *V. vexiliata* (2 acc. each), and *V. trinervia* and *V. glabrescence* (1 acc. each). All the accessions were collected under the "National Exploration Plan" coordinated by the ICAR-NBPGR, New Delhi, which is the nodal institute at the national level for the collection and conservation of plant genetic resources in India, supported by the ICAR-IIPR, Kanpur, and other institutes of the Indian Council of Agricultural Research (ICAR). A team of experts comprising taxonomists/botanists and crop breeders undertook different exploration missions to collect these wild accessions over the years and identified each accession. Subsequently, seed samples of all collections were multiplied and deposited in the National Genebank housed at the ICAR-NBPGR and are available freely to all researchers nationally and globally as per national legislation. As evidenced from previous morphological analysis as well as the literature available, this panel of genotypes represented a wide range of genetic diversity to represent important recombination events as well as high genetic diversity.

2.2 Phenotyping of the selected panel

All the *Vigna* genotypes were grown in natural field conditions in the wide hybridization garden in an augmented design during *Kharif* (rainy season) 2017 and 2018 at the Main Research Farm, ICAR-IIPR, Kanpur, India. Kanpur is located at 26°27′N latitude, 80°14′E longitude, 152.4 m above mean sea level (amsl) and experiences tropical climate with a long-term mean annual rainfall of 820 mm. (Annual report, 2019). About 80% of the total rainfall is received during the southwest monsoon season (July–September). The recommended package of agronomic practices for growing mungbean in the region was followed to raise healthy plants. The phenotypic data were recorded on five random plants of each genotype at the specified stage of plant growth, and these were averaged for use in the downstream analysis. The traits included chlorophyll content at 30 days (CC30) and 45 days (CC45), measured using a SPAD photometer from five top leaves in five random plants of each accession, averaged over all samples; days to first flowering (DFF); days to first pod maturity (DM); plant height in cm (PH); peduncle length in cm (PEDLTH); pod length in cm (PL); terminal leaflet length in cm (TLL); number of seeds per plant (NSPP); and 100-seed weight in g (HSW).

2.3 Phenotypic data analysis

The phenotypic data were analyzed for the descriptive statistics (R package = "pastecs"), and the linear mixed model approach was followed to calculate BLUP values for each genotype using the R package ("lme4"). The BLUP values were considered for determining the strength of a relationship between variables following trait correlation (R packages = "Corrplot" and "PerformanceAnalytics") and principal component analysis (PCA) of the traits (R packages = "factoxtra" and "FactoMineR").

2.4 Genotyping the panel with cross-genera and cross-species simple sequence repeats

The total genomic DNA was extracted from fresh young leaves of each accession at the early vegetative stage (within 10–12 days of sowing) following the CTAB method (Doyle and Doyle, 1990) with slight modifications (Pratap et al., 2015). The quality of the extracted DNA was analyzed on 0.8% agarose gel and the quantity was determined using a Nanodrop spectrophotometer ND 1000 (Nanodrop Technologies, DE, United States). The DNA of each sample was normalized to a concentration of 20–30 ng/μL. The panel was genotyped with 92 polymorphic SSRs belonging to different *Vigna* species namely, cowpea (Li et al., 2001), adzuki bean (Wang et al., 2004), mungbean (Kumar et al., 2002; Somta et al., 2009), and common bean (Gaitan-Solis et al., 2002; Blair et al., 2003) (Supplementary Table S2). PCR amplification was carried out on a tetrad thermal cycler (G-strom, Somerset, United Kingdom) in a reaction volume of 20 μL containing 50–60 ng template DNA, 0.4 μL 10 mM dNTPs, 0.6 U of *Taq* DNA polymerase (Fermentas, Mumbai), 2 μL of 10X *Taq* buffer A (Fermentas, Mumbai) with $MgCl_2$, and 5 pmole each of forward and reverse primers (ILS, India). PCR amplifications were performed at an initial denaturation of 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 15 s, primer-specific annealing for 15 s at 45–55°C, and extension at 72°C for 1 min followed by the final extension at 72°C for 5 min. The PCR products were resolved on

3.5% metaphor®agarose gel in 1X TAE buffer for 3–4 h at 80–100 V and stained with ethidium bromide. The gels were documented using a gel documentation system (Uvitech, Cambridge), and alleles were recorded on all genotypes according to their fragment sizes (in base pairs).

2.5 Population structure analysis

The Bayesian model-based STRUCTURE v2.3.4 tool was used to find the number of sub-populations (Q-matrix) in the selected *Vigna* association panel using 92 SSR markers (Pritchard et al., 2000). The genotypic data were analyzed with 10 independent runs for each cluster (*K*), ranging from 1 to 10 by setting the burn-in period of 30,000 and the number of Markov chain Monte Carlo iterations of 60,000 along with the admixture model and correlated allele frequencies. The optimum number of sub-populations (k) was determined using Structure Harvester Web v0.6.94 (Earl and Vonholdt, 2012) based on the ad-hoc criterion (Delta *K*) (Evanno et al., 2005).

2.6 Analysis of linkage disequilibrium between markers and marker–trait association

The software program TASSEL v2.0.1 (Bradbury et al., 2007) was used to evaluate linkage disequilibrium between the studied markers and marker–trait associations (MTAs) using the genotypic and 2 years' phenotypic data following the general linear model (GLM with Q-matrix) and mixed linear model (MLM with Q + K matrix). The Q-matrix derived from the STRUCTURE program and the relative kinship matrix calculated by TASSEL software was used for GLM and MLM analyses. The number of permutation runs in GLM was set to 10,000 to obtain a marker significance value of corrected $p < 0.00001$ (at alpha = 0.001 and $n = 92$) to declare MTAs. MLM with Q + K matrix was analyzed following default run parameters, namely, convergence criterion of 1.0×10^{-4}, and the maximum number of iterations was set to 200. Significant MTAs were declared at alpha 0.05, and corrected *p*-value of ≤0.00054 with relative magnitude was represented by the R^2 value as the portion of variation explained by the marker.

2.7 Localization of markers associated with traits on the *V. radiata* genome

Primer sequences of all associated SSRs were BLAST searched against the *V. radiata* var. *radiata* (tax id: 3,916) genome and searched for their 100% identity to find their exact physical position in a genome. The corresponding physical position of each SSR was used to extract the nucleotide sequences to search and verify the presence of the SSR motif flanking the primer region following the SSRIT tool (Temnykh et al., 2001) and searched for the co-localized or flanking genes.

3 Results

3.1 Descriptive statistics

The selected *Vigna* panel revealed wide phenotypic variations for all the 10 traits in both growing seasons (Table 1). The highest coefficient of variation (CV) was observed for HSW (83%) followed by PH (65%) and DFF (52%). The lowest CV was observed for CC45 (19%) and CC30 (15%). Data from both seasons were comparable to each other, and not much deviation was observed.

3.2 Correlation of traits with genotypic values obtained from the mixed model

Correlation of traits with predicted genotype values of each genotype from the mixed model approach revealed a high positive correlation between NSPP and PL (1.0), DFF and DM (0.97), CC30 and CC45 (0.82), HSW and NSPP (0.72), and HSW and PL (0.72). HSW exhibited a significant positive correlation with most of the traits except CC30, CC45, and peduncle length. NSPP recorded a significant positive correlation with peduncle length, PH, TLL, and PL. PH, DFF, and DM traits revealed a negative correlation with CC45 and CC30. Likewise, PL recorded a significant negative correlation with TLL (−0.38) (Figure 1).

The predicted genotype values of each genotype with 2018 data revealed highly significant positive correlations between DFF and DM (0.97), CC30 and CC45 (0.7), and HSW and PL (0.57). On the contrary, PH, DFF, and DM revealed a negative correlation with CC45 and CC30 as in the 2017 data. Similarly, DFF and DM had a significant negative correlation with NSPP. PH, PL, and TLL revealed a positive correlation with each of these three traits (Figure 1).

3.3 Principal component analysis with BLUP

The principal component analysis (PCA) established relationships amongst the studied traits during 2017 and 2018. The PCA of 2017 data revealed that dimension 1 (37%), dimension 2 (27%), dimension 3 (12.5%), and dimension 4 (10.12%) altogether contributed around 86.62% of the explained phenotypic variances of the studied traits (Figure 2). Dimension 1 had more than 10% contribution from HSW, NSPP, PH, and PL. Dimensions 2 and 4 had >10% contribution from

TABLE 1 Descriptive statistics of quantitative traits.

Descriptive statistics	PH		CC30		CC45		TLL		PEDLTH		DFF		DM		PL		NSPP		HSW	
	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018	2017	2018
Minimum	3.2	5.8	3.86	25.59	26.84	28.65	3.2	2.6	5.8	5.9	26	25	35	37	2.64	2.62	3.58	3.56	0.596	0.65
Maximum	144.1	197.2	52.84	60.3	63.32	67.16	12	11.6	34.6	34.6	132	129	56	151	16.5	11.56	13.8	13.4	19.927	20.7
Range	140.9	191.4	28.98	34.71	36.48	38.51	8.8	9	28.8	28.7	106	104	121	114	13.86	8.94	10.22	9.84	19.331	20.05
Sum	4,768.31	5,714.6	3,756.28	4,008.94	4,079.7	4,081.22	800	744.9	1,486.94	1,563.6	4,478	4,336	6,401	6,303	549.21	554.165	855.88	855.93	287.807	291.42
Mean	48.65	58.31	38.33	40.91	41.63	41.65	8.16	7.6	15.17	15.96	45.69	44.25	65.32	64.31	5.6	5.65	8.73	8.73	2.94	2.97
SE mean	3.18	3.56	0.61	0.61	0.8	0.72	0.23	0.22	0.67	0.68	2.35	2.34	2.65	2.63	0.197	0.17	0.26	0.26	0.25	0.25
Var.	993.66	1,241.53	36.84	36.53	62.31	50.8	5.04	4.79	43.83	45.61	541.39	536.66	689.89	677.02	3.47	2.89	6.49	6.63	5.97	6.15
St. deviation	31.52	35.24	6.07	6.04	7.89	7.13	2.25	2.19	6.62	6.75	23.27	23.17	26.27	26.02	1.86	1.7	2.555	2.58	2.44	2.48
CV	0.65	0.6	0.16	0.15	0.19	0.17	0.28	0.29	0.44	0.42	0.51	0.52	0.4	0.41	0.33	0.3	0.29	0.29	0.83	0.83

Where PH: plant height in cm; CC30: chlorophyll content at 30 days; CC45: chlorophyll content at 45 days; DFF: days to first flowering; DM: days to first pod maturity; PH: plant height in cm; PEDLTH: peduncle length in cm; PL: pod length in cm; TLL: terminal leaflet length in cm; NSPP: number of seeds per plant; and HSW: 100-seed weight in g.

10.3389/fgene.2022.1000440

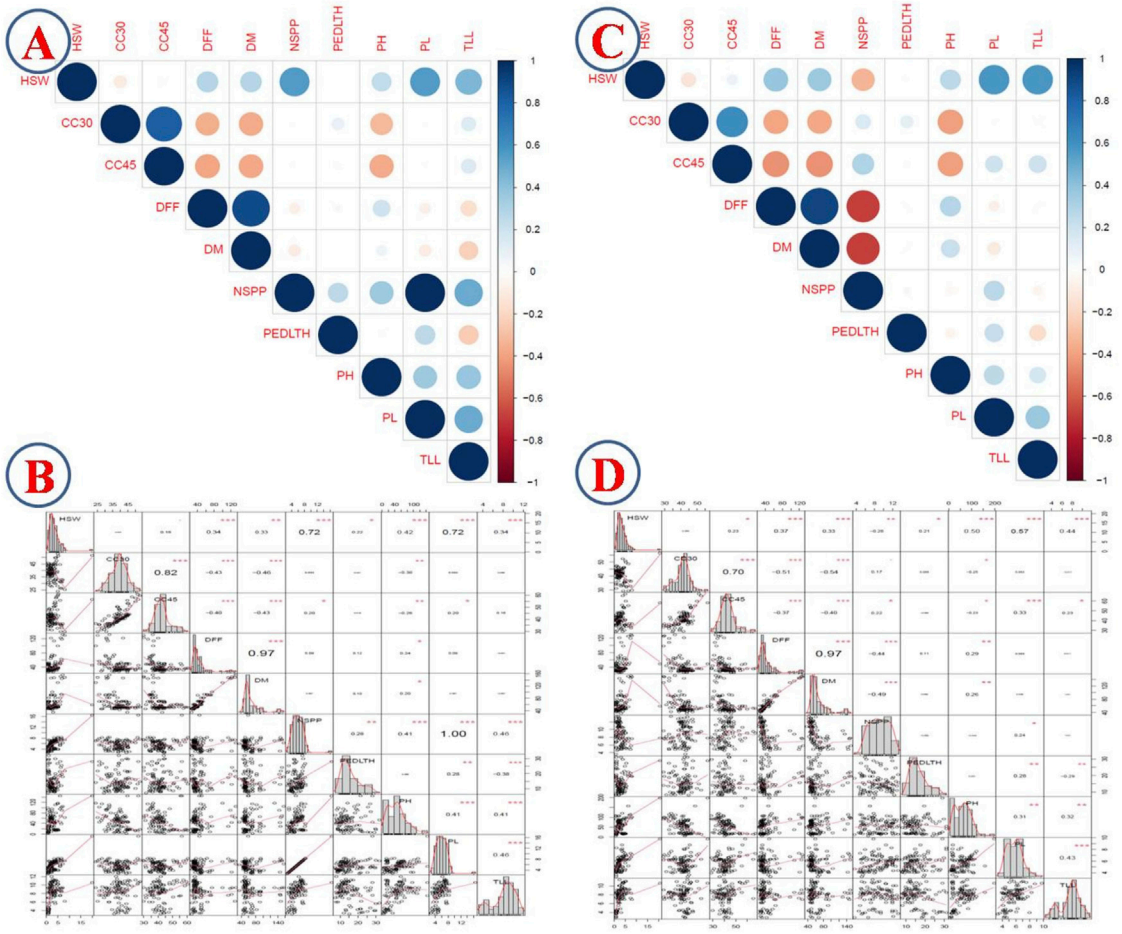


FIGURE 1
Correlation analysis of quantitative traits with BLUP values; **(A)** correlogram of 2017 data; **(B)** correlation table of 2017 data; **(C)** correlogram of 2018 data; **(D)** correlation table of 2018.

DFF, DM, CC30, and CC45. Dimension 3 had >10% variation from PEDLTH and TLL. The PCA also revealed that DFF and DM had a negative correlation with CC30 and CC45. HSW showed a positive correlation with the pod length and number of pods per plant (Figure 2).

The PCA of 2018 also revealed that dimension 1 (33.2%), dimension 2 (23.13%), dimension 3 (12.99%), and dimension 4 (10.81%) altogether contributed around 80.13% of the explained phenotypic variances of the studied traits (Figure 2). The phenotypic variance of >10% in dimension 1 was contributed by PH, CC30, DM, DFF, and HSW. Dimension 2 had >10% phenotypic variance contribution from PH, PL, TLL, HSW, and CC45. Dimension 3 had >10% phenotypic variances contribution from PEDLTH, TLL, and CC45, and dimension 4 had >10% phenotypic variances contribution from NSPP, PEDLTH, TLL, and CC30. The PCA also revealed a negative correlation of DFF and DM with NSPP, CC30, and CC45 (Figure 2).

3.4 Population genetic structure analysis

Population genetic structure analysis revealed three sub-populations (K = 3) in the selected panel of 13 cultivated mungbean varieties and 85 accessions belonging to 15 different *Vigna* species (Figure 3). The major sub-population 1 (Q1 in red color) represented 34 genotypes (35%) accommodating *V. trilobata* (17 acc.), *V. umbellata* (13 acc.), *V. vexillata* (2 acc.), and *V. dalzelliana* (2 acc.). Sub-population 2 (Q2 in green color) comprised 24 (24.5%) genotypes including cultivated *Vigna* (13 acc.); *V. trilobata* and *V. umbellata* (3 acc. each); *V. stipulaceae* (2 acc.); and 1 accession each of *V. unguiculata*, *V. glabrescence*, and *V. radiata*. Sub-population 3 (Q3 in blue color) consisted of 33 genotypes, and these belonged to *V. mungo* (9 acc.), *V. radiata* (11 acc.), *V. radiata* var. *sublobata* (6 acc.), *V. trinervia* var. *bourneae* (3 acc.), and *V. silvestris* and *V. radiata* var. *setulosa* (2 acc. each). A total of seven genotypes,

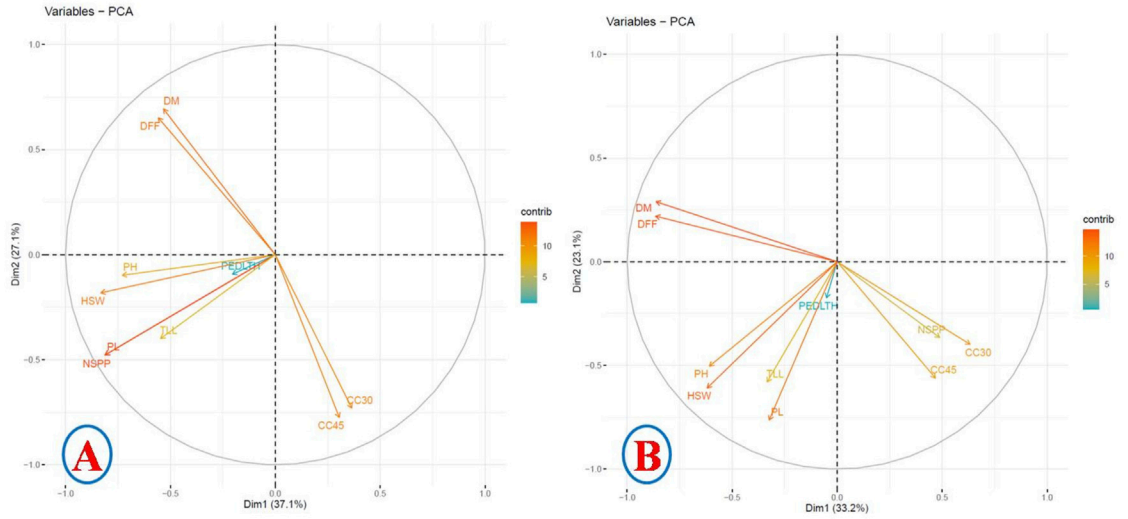


FIGURE 2
Principal component analysis with correlation. **(A)** 2017 data. **(B)** 2018 data.

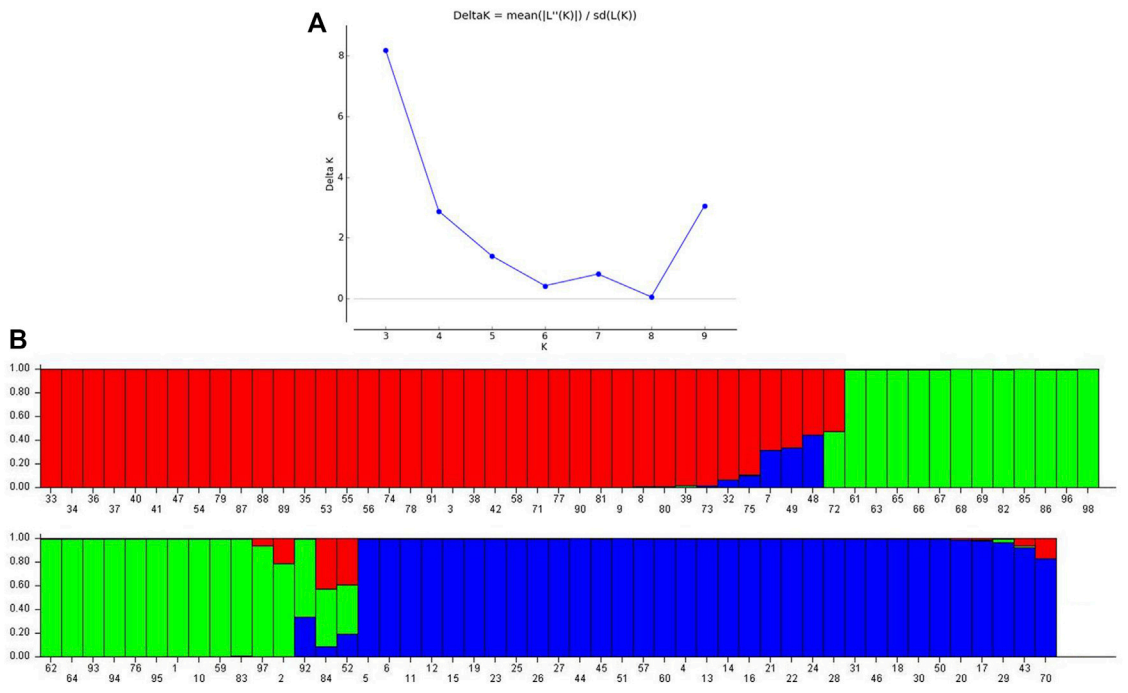


FIGURE 3
Population genetic structure analysis: **(A)** Evanno Delta K-based sub-population prediction. **(B)** Three distinct sub-populations in a chosen *Vigna* panel.

namely, IC247408 (*V. dalzelliana*), IC277014 (*V. silvestris*), IC277021 (*V. silvestris*), JAP/10–51 (*V. trinervia*), TCR279, IC298665, and NSB007 (*V. unguiculata*), were considered as the admixture class since these shared genomic content from Q1, Q2, and Q3 sub-populations. The genotype IC247408 considered as the admixture class had a major genome frequency belonging to Q1 (0.684), whereas two other genotypes belonging to *V. dalzelliana* clustered with *V. trilobata* and *V. umbellata*.

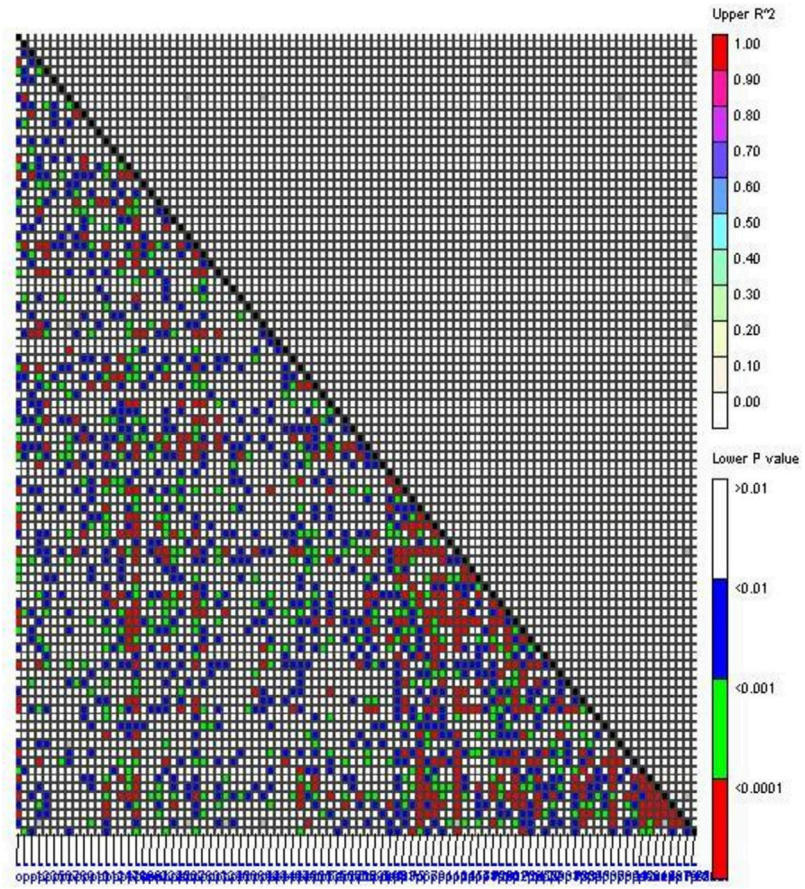


FIGURE 4
Linkage disequilibrium between marker pairs. The lower diagonal part represents *p*-values between the marker pairs; the upper diagonal part represents the R^2 value between marker pairs.

Similarly, TCR279 (0.65), IC298665 (0.42), and NSB007 (0.49) belonging to *V. unguiculata* had major genome frequency related to the Q2 sub-population, whereas Goa cowpea (*V. unguiculata*) grouped with the genotypes of cultivated mungbean varieties and *V. trilobata*, *V. umbellata*, *V. stipulaceae*, *V. glabrescence*, and *V. radiata*.

3.5 LD between studied markers

The r^2 value between alleles of two loci varied between 0.007 (CEDG071 vs. BMD-35) and 0.048 (CEDG256 vs. CEDG185) (Figure 4). Most of the marker pairs had r^2 values less than 0.02 (Supplementary Table S4). The observed r^2 values were near zero which indicated that the markers were in perfect equilibrium, and therefore, two markers will not provide identical information. Similarly, the D′value between two loci varied from 0.83 (CEDG150 vs. CEDG100) to 0.28 (CEDG060 vs. BMD-6). Most marker pairs have a D'value of more than 0.55 (3,250 pairs), thus indicating the existence of optimal recombination in the present population.

3.6 GLM (Q)-based association analysis

GLM with Q and phenotypic data of the experiments conducted in 2017 identified 34 MTAs for 9 out of 10 studied traits except for NSPP at a corrected *p*-value of <0.00001 (Supplementary Table S3). A maximum of 8 MTAs were identified for HSW followed by 6 for DFF; 5 for DM; 4 for PL; 3 each for CC30, PEDLTH, and PH; and 1 each for CC45 and TLL. The marker CEDG225 was observed to have an association with CC30, CC45, DFF, DM, and PH. Similarly, the markers, VR022, CEDG118, CP1225, and DMBSSR016, were also associated with HSW and PL. The markers CEDG096A, CEDG033, and JMES1424 were found associated with highly correlated traits such as DFF and DM.

TABLE 2 Significant marker–trait associations identified from the MLM (Q + K) approach in different environments.

Trait	Locus	Allele	2017 data			2018 data		
			F-value	*p*-value	R^2	F-value	*p*-value	R^2
DFF	CEDC033	150	4.34	5.54E-06	14.49	4.25	7.54E-06	15.14
	MBSSR008	200	6.52	8.34E-09	2.34	6.50	9.81E-09	3.13
	CP1225	205	-			2.92	5.35E-04	1.52
DM	CEDC033	150	3.2	2.93E-04	15.51	3.25	2.55E-04	15.54
	MBSSR008	220	5.60	1.35E-07	2.86	5.38	2.72E-07	2.80
	CP1225	205	3.08	2.91E-04	2.61	-		
PH	CP08695	260	3.86	7.83E-05	22.49	-		
	CEDG100	220	3.10	2.11E-04	34.87	4.00	6.69E-06	19.87
	VR018	220	-			3.07	2.44E-04	17.93
HSW	VR022	290	9.80	2.94E-13	2.48	11.28	9.62E-15	2.12
	GMES0337	195	4.06	9.08E-05	2.19	4.90	8.11E-06	2.52
	CEDG008	115	2.89	4.49E-04	1.63	3.32	8.11E-05	1.79
PL	VR022	290	5.18	1.52E-07	2.06	-		
TLL	CEDG100	185	3.02	2.91E-04	1.96	-		
	VR018	235	-			4.04	5.80E-06	1.96
PEDLTH	CEDG220	145	4.27	2.13E-05	7.27	4.96	2.52E-06	6.98
	VM27	140	2.85	4.96E-04	11.82	-		
CC30	VR039	140	3.78	9.97E-05	6.16	-		
	CP1225	190	-			3.65	3.62E-05	3.10
	CEDG096A	195	-			3.17	1.26E-04	11.59
CC45	CEDG096A	195	-			2.81	5.39E-04	13.06

$p \leq 0.00054$; R^2 = phenotypic variance; PH: plant height in cm; CC30: chlorophyll content at 30 days; CC45: chlorophyll content at 45 days; DFF: days to first flowering; DM: days to first pod maturity; PH: plant height in cm; PEDLTH: peduncle length in cm; PL: pod length in cm; TLL: terminal leaflet length in cm; and HSW: 100-seed weight in g.

GLM with Q and the phenotypic data of the experiments conducted in 2018 identified 30 MTAs for 8 out of the 10 studied traits, except NSPP and PL, at a corrected *p*-value < 0.00001 (Supplementary Table S3). A maximum of 9 MTAs for HSW and the lowest 1 MTA for CC45 were identified. A total of 6 MTAs each for DFF and DM and 2 each for CC30, PEDLTH, PH, and TLL were also identified. Of these MTAs, the loci CEDG225, CEDG096A, CEDG033, MBSSR008, and CEDG100 were observed to be associated with more than one trait. The marker CEDG225 was associated with CC30, DFF, DM, and TLL. Similarly, CEDG096A was associated with CC30, CC45, DFF, and DM. The markers CEDG033 and MBSSR008 were also associated with DFF and DM. Marker CEDG100 was associated with HSW and PH. A total of 24 MTAs associated with HSW (8), DFF (6), DM (5), PEDLTH (2), PH (1), TLL (1), and CC30 (1) were consistently expressed in both seasons.

3.7 Mixed linear model (Q+K)-based association analysis

MLM analysis revealed an association of 13 markers, namely, VR018, VR039, VR022, CEDG033, GMES0337, MBSSR008, CEDG220, VM27, CP1225, CP08695, CEDG100, CEDG008, and CEDG096A, with 9 out of 10 studied traits except for NSPP from the data of both the years at value < 0.00054 with the Bonferroni correction (alpha 0.05; n:92) (Table 2; Figure 5). Of these 13 markers, seven markers, namely, VR018 (PH and TLL), VR022 (HSW and PL), CEDG033 (DFF and DM), MBSSR008 (DFF and DM), CP1225 (CC30, DFF, and DM), CEDG100 (PH and TLL), and CEDG096A (CC30 and CC45) were observed to be associated with multiple traits. The phenotypic variation explained by these MTAs varied from 1.52 (DFF-CP1225) to 34.87 (PH-CEDG100). The maximum number of 3 MTAs was identified for traits like CC30, DFF, DM, PH, TL, and HSW, and a minimum of 1 MTA was identified for PL (VR022) and CC45 (CEDG096A). The MTAs identified for DFF and DM were also similar. The association of markers CEDG033 and MBSSR008 linked with DFF and DM were consistent across the 2 years of experimentation, explaining up to 15.54% phenotypic variation. The markers CEDG100, CP08695, and VR018 were linked with PH expression, and among these, VR018 identified from 2 years of data explained 34.86% and 19.87% phenotypic variation in 2017 and 2018, respectively. Importantly, markers such as VR022, GMES0337, and CEDG008 have been consistently identified as linked with

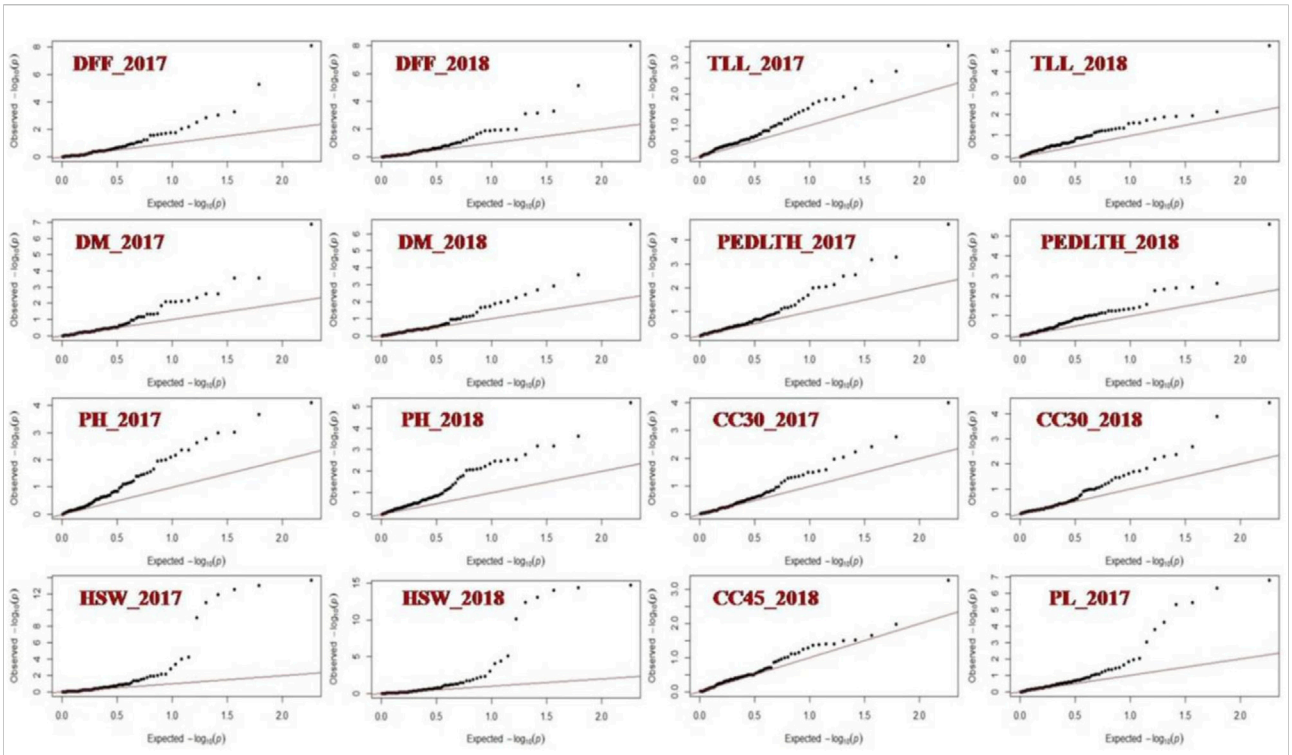


FIGURE 5
QQ bi-plot showing the association of markers with yield-related traits in *Vigna* species.

HSW expression in the 2 years' data, and the phenotypic variation explained by them varied from 1.62 to 2.5%. The phenotypic variation of MTA linked with CC30 (CEDG096A, CP1225, and VR039) varied from 3.09 to 11.58. The marker CEDG096A linked with CC30 and CC45 was identified only in 2018 which explained 11.58% and 13% phenotypic variation for these traits, respectively. A total of 15 MTAs (CEDG096A and VR039 for CC30; CEDG096A for CC45; CEDG033 and MBSSR008 for DFF and DM; CEDG100 and CP08695 for PH; VR022 for PL; VR018 for TLL; CEDG220 for PEDLTH; and VR022, GMES0337, and CEDG008 for HSW) identified for 9 traits from both seasons following the MLM (Q + K) approach were also identified through the GLM approach.

3.8 Genes co-localized with trait-linked simple sequence repeat markers

The primer pairs of five markers were exactly (100%) aligned on the *Vigna* reference genome (*V. radiata* var. *radiata* (taxid: 3,916)) (Table 3) and also possessed the same SSR motif as reported earlier in related *Vigna* species. Interestingly, four markers, namely, CP00361, CEDG100, VR022, and VR039, linked with various traits in this study are co-localized in protein-coding genes of the mungbean genome such as uncharacterized protein (LOC106765295), histone–lysine N-methyltransferase *ATX5* (LOC106775878), *SHOOT GRAVITROPISM 5* (LOC106774869), and putative UPF0481 protein *At3g02645* (LOC106767008), respectively. Another primer pair of the marker CEDG220 associated with peduncle length is found to be intergenic between the genes coding for GTP-binding nuclear protein Ran-3 (LOC106768487) and cytochrome P450 734A1 (LOC106767283).

4 Discussion

The test panel of 98 highly diverse *Vigna* genotypes belonging to cultivated (13 acc.) and wild (85 acc.) species was genotyped with cross-species and cross-genera SSR primers from adzuki bean-, common bean-, and scarlet runner bean (Pratap et al., 2015; Kumari et al., 2021)-identified MTAs for 10 quantitative traits. The SSRs can be effectively utilized for linkage studies and gene/QTL/association/comparative mapping studies in crops such as *Vigna* sp. In recent years, efforts have been made to develop various SSR markers such as genic/EST- and genomic-SSRs in many *Vigna* species (Wang et al., 2004; Gwag et al., 2006; Somta et al., 2008; Chen et al., 2015). Successful transferability of SSRs from adzuki bean, common bean, and scarlet runner bean was

TABLE 3 Trait-linked SSRs flanking candidate genes in *V. radiata*.

S. No.	Marker name	Genes of *V. radiata* var. *radiata* nesting SSR markers	Start and end of forward primer	Start and end of reverse primer	Amplicon length in *V. radiata*	SSR motifs in *V. radiata var. radiata*
1	CP00361	Uncharacterized LOC106765295	322–345	106–129	240	(CT)5N(CT)5N (TC)8
2	CEDG100	Histone–lysine N-methyltransferase ATX5 (LOC106775878)	283–304	465–444	183	(CT)6N (TC)9
3	CEDG220	CHR7; intergenic between LOC106768487 (42,176,444. 42,178,582) and LOC106767283 (42,196,013.42,199,609)	152–172	23–44	150	(CT)12
4	VR022	SHOOT GRAVITROPISM 5 (LOC106774869)	66–92	188–209	144	(TCTA)5N (TC)5
5	VR039	Putative UPF0481 protein *At3g02645* (LOC106767008)	1,117–1,138	1,226–1,246	130	(AGA)5

reported earlier in mungbean (Wang et al., 2009; Dikshit et al., 2012; Gupta et al., 2013a; Pratap et al., 2015; Pratap et al., 2016; Pratap et al., 2020; Kumari et al., 2021). Earlier, cross-genera- and cross-species-specific SSRs have been very well utilized in linkage and association mapping for identifying genes/QTLs responsible for various traits in mungbean, blackgram, *V. marina*, and cowpea in addition to their use in studying genetic diversity (Isemura et al., 2012; Gupta et al., 2013b; Singh et al., 2020; Rohilla et al., 2021; Vadivel et al., 2021).

Agronomically important yield-attributing traits, namely, PH, NSP, HSW, DF, DM, and resistance to various biotic and abiotic stresses, in most of the crop plants are likely to be controlled by many genes/QTLs with additive gene actions (Bharadwaj et al., 2021). Therefore, effective selection could be practiced along with modern genomic tools even in the early generations of crops. A large amount of phenotypic variation and correlations among traits observed in the panel of cultivated and wild *Vigna* accessions in the current study are in agreement with the earlier reports (Pratap et al., 2017; Azam et al., 2018; Nair et al., 2019; Kumari et al., 2021). The variability at the genotypic and phenotypic levels in the present study could be attributed to a large number of accessions belonging to 19 *Vigna* species which represented different eco-geographical regions of India. The positive correlation of HSW and PL, DFF, and DM was observed, which will greatly favor the development of short-duration mungbean cultivars with higher yield potential. Simultaneously, a focus will be required on the number of pods in a bunch in order to maintain more yield per plant.

Assigning different individuals to defined populations is highly beneficial in population genetics studies (Pritchard et al., 2000), whereas making a population classification can provide an inference of individual ancestry that might not have been adequately defined beforehand. In this study, 92.8% of the studied genotypes were distinctly grouped into three sub-populations and the remaining 7.2% of the studied lines were grouped as the admixture class. Majority of the cultivated genotypes of mungbean and the genotypes belonging to the primary gene pool such as *V. mungo* (9 acc.), *V. radiata* (11 acc.), *V. radiata* var. *sublobata* (6 acc.), *V. trinervia* var. *bourneae* (3 acc.), *V. silvestris* (2 acc.), and *V. radiata* var. *setulosa* (2 acc.) were clustered in sub-populations 2 and 3, respectively. However, most of the genotypes (88%) in sub-population 1 belonged to the secondary gene pool, and these included *V. trilobata* (17 acc.) and *V. umbellata* (13 acc.). *V. trilobata* have a higher number of pods with small seeds and varying capacities of plant growth habit (Pandiyan et al., 2012). Rice bean (*V. umbellata*) is mainly grown in northern India and Southeast Asia, and is considered as a donor for resistance to bruchids, yellow mosaic virus, Cercospora leaf spot, and bacterial leaf spot (Bhanu et al., 2018; Aidbhavi et al., 2021). The accession IC 251342 of *V. umbellata* was reported as photo- and thermo-period tolerant (Pratap et al., 2014a). Therefore, *V. umbellata* can be effectively utilized as a donor in mungbean and urdbean improvement programs to impart biotic and abiotic stress tolerance with due consideration of addressing pre- and post-fertilization barriers (Kumar et al., 2007; Chen et al., 2017; Pratap et al., 2018; Pratap et al., 2021b). So far, limited studies have been undertaken on the analysis of the population genetic structure in mungbean and other *Vigna* species; in most of them, 3–4 sub-populations in various germplasm panels and 6 sub-populations in a panel consisting of released varieties, advanced breeding materials, and elite lines were identified (Noble et al., 2012; Breria et al., 2020; Sokolkova et al., 2020; Kumari et al., 2021). The grouping of genotypes in our study is also in conformity with the earlier studies. Furthermore, the genotypes belonging to specific gene pools were grouped together as per the gene-pool classification of mungbean (Fatokun et al., 1993). Nonetheless, since the taxonomy of *Vigna* crops to date is primarily based on morphological attributes, the chance of misclassification of few *Vigna* species could have been encountered (Tomooka et al., 2006). Therefore, correct molecular taxonomy of the wild relatives of *Vigna* accessions is of prime importance in order to decipher their relationship and diversity among the various accessions of *Vigna* species for their further effective usage in

various *Vigna* improvement programs (Takahashi et al., 2018; Kumari et al., 2021).

Mapping gene(s) or genomic region(s) which regulate important traits helps us to effectively utilize modern breeding technologies to expedite the varietal development process. Few studies on mapping genes/QTLs following QTL and association mapping approaches were reported in mungbean which included salinity stress tolerance (Breria et al., 2020), MYMIV resistance (Chankaew et al., 2011; Singh et al., 2017; Singh et al., 2020), maturity and hypocotyl color (Sokolkova et al., 2020), seed mineral concentration (Wu et al., 2020), and domestication-related traits (Isemura et al., 2012). QTL mapping for HSW, NPP, NSP, and MYMIV tolerance was reported using a limited number (100) of segregating RILs and few (46) polymorphic SSR markers (Singh et al., 2013). Hence, mapping genes/QTLs following a diverse panel consisting of an extensive collection of different *Vigna* species will really benefit not only the mungbean development program but also other closely related *Vigna* species. Out of the identified MTAs from GLM and MLM approaches in this study, 11 MTAs were previously reported in various mapping populations of *Vigna* species following the QTL mapping approach for traits such as MYMIV (CEDG100) in mungbean (Kitsanachandee et al., 2013), MYMV (CEDG225) in blackgram (Gupta et al., 2013b; Vadivel et al., 2021) and mungbean (CEDG225, MBSSR008, and VM27) (Singh et al., 2020), Cercosporaleaf spot (CEDG008) (Tantasawat et al., 2020), HSW (VM27) in *V. marina* (Chankaew et al., 2014), and HSW (CP1225) in mungbean (Alam et al., 2014). In addition to these, QTLs related to domestication-related traits such as seed length and number of seeds per pod linked/flanking the SSR marker CEDG220; pod width with CEDG096A; total number of pods with CEDG096A; stem internode length with CEDG271 and CP00361; HSW with JMES1424; and percent of shattered pods with MBSSR008 in mungbean using BC_1F_1 biparental a mapping population derived from a cross between a wild mungbean accession (JP211874) and a cultivated mungbean landrace (JP229096 cv. Sukhothai) (Isemura et al., 2012) were also identified for different traits in our study. The marker CP1225 identified for HSW in this study is consistent with an earlier report (Alam et al., 2014) where it was identified for the same trait in the F_2 mungbean mapping population derived from BM1 × BM6 by single regression analysis.

The marker CEDG100 was associated with HSW, PH, and TLL, and this MTA was consistently expressed for PH expression in both the seasons' data explaining >30% of phenotypic variation, and hence, was considered as a major and consistent MTA for PH. This genic/EST SSR marker co-localized gene encoding histone–lysine N-methyltransferase ATX5 is implicated in epigenetics, specifically in the methylation of histones. Histone methylation is an important epigenetic modification which plays a crucial role in regulating the gene expression by either increasing or decreasing the target gene expression and genome stability (Kumar and Mohapatra, 2021). The histone–lysine methyltransferase (HKMTase), specifically ATX5, is involved in the methylation of lysine residue (H3K4 di- and trimethylation) present on the tail of the histone protein and *ARABIDOPSIS TRITHORAX 5* (*ATX5*) reported to function in abscisic acid and dehydration stress responses (Chen et al., 2017; Liu et al., 2018a) and glucose signaling in *Arabidopsis* (Liu et al., 2018b). Importantly, *ATX5* along with *ATX3 and ATX4* in *Arabidopsis* has a role in determining plant height (Chen et al., 2017). The triple mutant of *atx3-1*, *atx4-1*, and *atx5-1* exhibits dwarf and small rosette leaf phenotypes without altering the flowering time in *Arabidopsis* (Chen et al., 2017). Similarly, the VR22 marker associated with PL and HSW is co-located in gene-encoding *SHOOT GRAVITROPISM 5* (*SGR5*) (LOC106774869) in mungbean. Gravity affects many biological processes including negative (directing shoots upwards) and positive (downwards) gravitropism. *SGRs* have been shown to mediate the gravitropic responses of different plant organs. SRG5 is a C2H2-type ZF protein and functions primarily in the early steps of gravity perception in inflorescence stems (Morita et al., 2006). The mutant of *SGR5* exhibits an altered gravitropic response of the inflorescence stems by altering the patterns of starch accumulation or deposition in the endodermal cells of inflorescence stems (Tanimoto et al., 2008; Kim et al., 2016). Alternative splicing of *SGR5* into *SGR5a* and *SGR5b* modulates the gravitropic response of inflorescence stems at high temperatures in *Arabidopsis* (Kim et al., 2016). In addition to *SGR5*, other genes determining shoot gravitropism are *BIG GRAIN 1* (*BG1*) in rice (Liu et al., 2015); *LAZY1* (*LA1*) in rice, maize, and tea; and *Arabidopsis* (Overbeek, 1936; Jones and Adair, 1938; Li et al., 2007; Yoshihara et al., 2013), which determine spreading of tillers in rice, prostate culms in maize, and outward orientation branching and wider angles in *Arabidopsis* by primarily modulating polar auxin transport (PAT). Plant lateral organs such as primary and secondary branches are generated at a defined angle termed the gravitropic set-point angle and thus, determine the overall plant architecture which is mainly maintained by gravitropism (Digby and Firn, 1995). Hence, developing a variety with ideal plant architecture (IPA) in *Vigna* is necessary to enhance crop yield by harvesting light sources and converting sources efficiently to sink in plants. The important characteristics of IPA in *Vigna* include increased number of pods per plant as well as bunch, number of seeds per pod, high HSW, and short plant-type with erect growth habit.

Days to flowering and maturity are important in the mungbean improvement program to breed shorter-duration crop varieties with higher yields. The markers CEDC033 and MBSSR008 associated with the expression of DFF and DM have been observed to be consistent in both seasons. However, we could not identify these markers associated with HSW as identified in correlation studies where HSW had a positive correlation with DFF and DM traits. This could be ascribed to genes controlling HSW in *Vigna*. The MTAs identified in this

study could be effectively utilized in improving other closely related *Vigna* species which lack genomic information such as trait-linked markers for deploying in marker-assisted crop improvement programs. The MTAs identified in this study for the traits DFF, DM, PH, and HSW can be further explored to utilize in the *Vigna* improvement program following marker-assisted breeding and cloning of the genes to help us understand the molecular mechanism controlling the expression of these traits in various *Vigna* species.

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.

Author contributions

GK, AP, and SG conceptualized, planned, and executed the experiment. GK, PS, and AP analyzed the data and interpreted the results. GK, PT, DS, PG, and KT undertook field and laboratory experiments. PG provided seeds of some of the accessions. SG and RN provided suggestions for the improvement of the experiments and the manuscript. All the authors made a direct intellectual contribution to the manuscript, and all read and approved the manuscript.

Funding

The authors acknowledge the funding support received from the Department of Biotechnology, Government of India (Grant Nos. BT/Ag/Network/Pulses-I/2017–18) and ACIAR, Australia (Grant Nos. CIM/2014/79), in the form of funded research projects.

Acknowledgments

The authors would like to thank the Indian Council of Agricultural Research for providing all the necessary laboratory and field facilities for conducting this research.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1000440/full#supplementary-material

References

Aidbhavi, R., Pratap, A., Verma, P., Lamichaney, A., Bandi, S. M., Nitesh, S. D., et al. (2021). Screening of endemic wild *Vigna* accessions for resistance to three bruchid species. *J. Stored Prod. Res.* 93, 101864. doi:10.1016/j.jspr.2021.101864

Alam, A. K. M., Somta, P., and Srinives, P. (2014). Identification and confirmation of quantitative trait loci controlling resistance to mungbean yellow mosaic disease in mungbean (*Vigna radiata* L. Wilczek). *Mol. Breed.* 34, 1497–1506. doi:10.1007/s11032-014-0133-0

Annual Report (2019). *Annual report*. Kanpur – 208024: ICAR-Indian Institute of Pulses Research, 146.

Azam, M. G., Hossain, M. A., Alam, M. S., Rahman, K. S., and Hossain, M. (2018). Genetic variability, heritability and correlation path analysis in mungbean (*Vigna radiata* L.Wilczek). *Bangladesh J. Agric. Res.* 43, 407–416. doi:10.3329/bjar.v43i3.38389

Basu, P. S., Pratap, A., Gupta, S., Sharma, K., Tomar, R., and Singh, N. P. (2019). Physiological traits for shortening crop duration and improving productivity of greengram (*V. radiata* L. Wilczek) under high temperature. *Front. Plant Sci.* 10, 1508. doi:10.3389/fpls.2019.01508

Baudoin, J. P., and Marechal, R. (1988). "Taxonomy and evolution of the genus (*Vigna* mungbean)," in Proceedings of second international symposium (Bangkok, Shanhua, Taiwan: AVRDC), 2–12.

Bhanu, A. N., Singh, M. N., and Srivastava, K. (2018). Efficient hybridization procedure for better pod setting in inter-specific crosses involving (*Vigna)* species. *Adv. Plants Agric. Res.* 8, 101–105. doi:10.15406/apar.2018.08.00298

Bharadwaj, C., Tripathi, S., Soren, K. R., Thudi, M., Singh, R. K., Sheoran, S., et al. (2021). Introgression of "QTL-hotspot" region enhances drought tolerance and grain yield in three elite chickpea cultivars. *Plant Genome* 14, e20076. doi:10.1002/tpg2.20076

Bisht, I. S., Bhat, K. V., Lakhanpaul, S., Latha, M., Jayan, P. K., Biswas, B. K., et al. (2005). Diversity and genetic resources of wild *Vigna* species in India. *Genet. Resour. Crop Evol.* 52, 53–68. doi:10.1007/s10722-005-0286-0

Blair, M. W., Giraldo, M. C., Buendía, H. F., Tovar, E., Duque, M. C., and Beebe, S. E. (2003). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.) microsatellite marker diversity in common bean *(Phaseolus vulgaris L.). Theor. Appl. Genet.* 113, 100–109. doi:10.1007/s00122-006-0276-4

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 (19), 2633–2635. doi:10.1093/bioinformatics/btm308

Breria, C. M., Hsieh, C. H., Yen, J. Y., Nair, R., Lin, C. Y., Huang, S. M., et al. (2020). Population structure of the world vegetable center mungbean mini core

collection and genome-wide association mapping of loci associated with variation of seed coat luster. *Trop. Plant Biol.* 13, 1–12. doi:10.1007/s12042-019-09236-0

Chandel, K. P. S., Lester, R. N., and Starling, R. J. (1984). The wild ancestors of urd and mungbeans (*Vigna mungo* L. Hepper) and (*V. radiata* L. Wilczek). *Bot. J. Linn. Soc.* 89, 85–96. doi:10.1111/j.1095-8339.1984.tb01002.x

Chankaew, S., Somta, P., Sorajjapinun, W., and Srinives, P. (2011). Quantitative trait loci mapping of Cercospora leaf spot resistance in mungbean, (*Vigna radiata* L. Wilczek). *Mol. Breed.* 28, 255–264. doi:10.1007/s11032-010-9478-1

Chankaew, S., Isemura, T., Naito, K., Ogiso-Tanaka, E., Tomooka, N., Somta, P., et al. (2014). QTL mapping for salt tolerance and domestication-related traits in (*Vigna marina* subsp. Oblonga), a halophytic species. *Theor. Appl. Genet.* 127, 691–702. doi:10.1007/s00122-013-2251-1

Chen, H., Liu, L., Wang, L., Wang, S., Somta, P., and Cheng, X. (2015). Development and validation of EST-SSR markers from the transcriptome of adzuki bean (*Vigna angularis*). *PLoS ONE* 10, e0131939. doi:10.1371/journal.pone.0131939

Chen, L. Q., Luo, J. H., Cui, Z. H., Xue, M., Wang, L., Zhang, X. Y., et al. (2017). *atx3, atx4*, and *atx5* encode putative h3k4 methyltransferases and are critical for plant development. *Plant Physiol.* 174, 1795–1806. doi:10.1104/pp.16.01944

de Candolle, A. (1886). "Plants cultivated for their seeds," in *Origin of cultivated plants*. 2nd Edition (London: Kegan Paul Trench & Co.), XLIX, 376–384. International Scientific Series.

Digby, J., and Firn, R. D. (1995). The gravitropic set-point angle (GSA): The identification of an important developmentally controlled variable governing plant architecture. *Plant Cell Environ.* 18, 1434–1440. doi:10.1111/j.1365-3040.1995.tb00205.x

Dikshit, H. K., Singh, D., Singh, A., Jain, N., Kumari, J., and Sharma, T. (2012). Utility of adzuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi] simple sequence repeat (SSR) markers in genetic analysis of mungbean and related *Vigna* spp. *Afr. J. Biotechnol.* 11, 13261–13268. doi:10.5897/AJB12.1720

Douglas, C., Pratap, A., Rao, B. H., Manu, B., Dubey, S., Singh, P., et al. (2020). "Breeding progress and future challenges: Abiotic stresses," in *The mungbean genome*. Editors R. M. Nair, R. Schafleitner, and S. H. Lee (Cham: Springer), 81–96. Compendium of Plant Genomes.

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

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

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

Fatokun, C. A., Danesh, D., Young, N. D., and Stewart, E. L. (1993). Molecular taxonomic relationships in the genus *Vigna* based on RFLP analysis. *Theor. Appl. Genet.* 86, 97–104. doi:10.1007/BF00223813

Flint-Garcia, S. A., Thornsberry, J. M., and Iv, B. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi:10.1146/annurev.arplant.54.031902.134907

Gaitan-Solis, E., Duque, M. C., Edwards, K. J., and Tohme, J. (2002). Microsatellite repeats in common bean (*Phaseolus vulgaris*). Isolation, characterization and cross-species amplification in *Phaseolus* ssp. *Crop Sci.* 42, 2128–2136. doi:10.2135/cropsci2002.2128

Gore, P. G., Tripathi, K., Pratap, A., Bhat, K., Umdale, S. D., Gupta, V., et al. (2019). Delineating taxonomic identity of two closely related *Vigna* species of section aconitifoliae: (*V. trilobata* L. Verdc.) and (*V. stipulacea* lam. Kuntz) in India. *Genet. Resour. Crop Evol.* 66, 1155–1165. doi:10.1007/s10722-019-00767-9

Gore, P. G., Das, A., Bhardwaj, R., Tripathi, K., Pratap, A., Dikshit, H. K., et al. (2021). Understanding G × E interaction for nutritional and antinutritional factors in a diverse panel of (*Vigna stipulacea* Lam. Kuntz) germplasm tested over the locations. *Front. Plant Sci.* 12, 766645. doi:10.3389/fpls.2021.766645

Gupta, S., Gupta, D. S., Anjum, K. T., Pratap, A., and Kumar, J. (2013a). Transferability of simple sequence repeat markers in blackgram. *Aust. J. Crop Sci.* 7, 345–353.

Gupta, S., Gupta, D. S., Anjum, T. K., Pratap, A., and Kumar, J. (2013b). Inheritance and molecular tagging of MYMIV resistance gene in blackgram (*Vigna mungo* L. Hepper). *Euphytica* 193, 27–37. doi:10.1007/s10681-013-0884-4

Gwag, J. G., Chung, J. W., Chung, H. K., Lee, J. H., Ma, K. H., Dixit, A., et al. (2006). Characterization of new microsatellite markers in mungbean, (*Vigna radiata l.). Mol. Ecol. Notes* 6, 1132–1134. doi:10.1111/j.1471-8286.2006.01461.x

HanumanthaRao, B., Nair, R. M., and Nayyar, H. (2016). Salinity and high temperature tolerance in mungbean (*Vigna radiata* L. Wilczek) from a physiological perspective. *Front. Plant Sci.* 7, 957. doi:10.3389/fpls.2016.00957

Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE* 7, 41304. doi:10.1371/journal.pone.0041304

Jones, J. W., and Adair, C. R. (1938). A "lazy" mutation in rice. *J. Hered.* 2, 315–318. doi:10.1093/oxfordjournals.jhered.a104527

Kim, J. Y., Ryu, J. Y., Baek, K., and Park, C. M. (2016). High temperature attenuates the gravitropism of inflorescence stems by inducing SHOOT GRAVITROPISM 5 alternative splicing in Arabidopsis. *New Phytol.* 209, 265–279. doi:10.1111/nph.13602

Kitsanachandee, R., Somta, P., Chatchawankanphanich, O., Akhtar, K. P., Shah, T. M., Nair, R. M., et al. (2013). Detection of quantitative trait loci for mungbean yellow mosaic India virus (MYMIV) resistance in mungbean (*Vigna radiata* L. Wilczek) in India and Pakistan. *Breed. Sci.* 63, 367–373. doi:10.1270/jsbbs.63.367

Kumar, S., and Mohapatra, T. (2021). Dynamics of DNA methylation and its functions in plant growth and development. *Front. Plant Sci.* 12, 596236. doi:10.3389/fpls.2021.596236

Kumar, S. V., Tan, S. G., Quah, S. C., and Yusoff, K. (2002). Isolation of microsatellite markers in mungbean, *Vigna radiata*. *Mol. Ecol. Notes* 2, 96–98. doi:10.1046/j.1471-8286.2002.00158.x

Kumar, N. P., Pandiyan, M., and Veerabadhiran, P. (2007). Pre-fertilization barriers in *Vigna radiata* × *Vigna umbellata*. *Plant Arch.* 7, 377–380.

Kumari, G., Lavanya, G. R., Shanmugavadivel, P. S., Singh, Y., Singh, P., Patidar, B., et al. (2021). Genetic diversity and population genetic structure analysis of an extensive collection of wild and cultivated *Vigna* accessions. *Mol. Genet. Genomics.* 296, 1337–1353. doi:10.1007/s00438-021-01825-7

Lawn, R. J., and Cottrell, A. (1988). Wild mungbean and its relatives in Australia. *Biol. Columb. Ohio)* 35, 267–273.

Li, C. D., Fatokun, C. A., Ubi, B., Singh, B. B., and Scoles, G. J. (2001). Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. *Crop Sci.* 41, 189–197. doi:10.2135/cropsci2001.411189x

Li, P., Wang, Y., Qian, Q., Fu, Z., Wang, M., Zeng, D., et al. (2007). *LAZY1* controls rice shoot gravitropism through regulating polar auxin transport. *Cell Res.* 17, 402–410. doi:10.1038/cr.2007.38

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, Y., Zhan, A., Yin, H., Meng, Q., Yu, X., Huang, S., et al. (2018a). Trithorax-group proteins Arabidopsis TRITHORAX4 (ATX4) and ATX5 function in abscisic acid and dehydration stress responses. *New Phytol.* 217, 1582–1597. doi:10.1111/nph.14933

Liu, Y., Wang, J., Yin, H., Zhang, A., Huang, S., Wang, T. J., et al. (2018b). Trithorax-group protein ATX5 mediates the glucose response via impacting the HY1-ABI4 signaling module. *Plant Mol. Biol.* 98, 495–506. doi:10.1007/s11103-018-0791-0

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12, 57–63. doi:10.1016/j.tplants.2006.12.001

Morita, M. T., Sakaguchi, K., Kiyose, S. I., Taira, K., Kato, T., Nakamura, M., et al. (2006). A C2H2-type zinc finger protein, SGR5, is involved in early events of gravitropism in Arabidopsis inflorescence stems. *Plant J.* 47, 619–628. doi:10.1111/j.1365-313X.2006.02807.x

Nair, R. M., Pandey, A. K., War, A. R., Hanumantharao, B., Shwe, T., Alamn, A., et al. (2019). Biotic and abiotic constraints in mungbean production - progress in genetic improvement. *Front. Plant Sci.* 10, 1340. doi:10.3389/fpls.2019.01340

Noble, T. J., Tao, Y., Mace, E. S., Williams, B., Jordan, D. R., Douglas, C. A., et al. (2012). Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. *Front. Plant Sci.* 8, 2102. doi:10.3389/fpls.2017.02102

Overbeek, J. V. (1936). LAZY, " an a-geotropic form of maize: "gravitational indifference" rather than structural weakness accounts for prostrate growth-habit of this form. *J. Hered.* 27, 93–96. doi:10.1093/oxfordjournals.jhered.a104191

Pandiyan, M., Senthil, N., Suresh, R., Chakravarthy, N., and Packiaraj, D. (2012). Interspecific hybridization of *Vigna radiata* x *Vigna trilobata*. *Wudpecker J. Agric. Res.* 1, 233–234.

Pratap, A., Basu, P. S., Gupta, S., Malviya, N., Rajan, N., Tomar, R., et al. (2014a). Identification and characterization of sources for photo- and thermo-insensitivity in *Vigna* species. *Plant Breed.* 133, 756–764. doi:10.1111/pbr.12215

Pratap, A., Malviya, N., Tomar, R., Gupta, D. S., and Kumar, J. (2014b). "Vigna," in *Alien gene transfer in crop plants*. Editors A. Pratap and J. Kumar (New york,

USA: Springer Business+ Science Media), Vol. 2, 163–190. Achievements and Impacts.

Pratap, A., Gupta, S., Malviya, N., Tomar, R., Maurya, R., John, K. J., et al. (2015). Genome scanning of asiatic *vigna* species for discerning population genetic structure based on microsatellite variation. *Mol. Breed.* 35, 178. doi:10.1007/s11032-015-0355-9

Pratap, A., Gupta, S., Tomar, R., Malviya, N., Maurya, R., Pandey, V. R., et al. (2016). Singh Cross-genera amplification of informative microsatellite markers from common bean and scarlet runner bean for assessment of genetic diversity in mungbean (*Vigna radiata*). *Plant Breed.* 135, 499–505. doi:10.1111/pbr.12376

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. Genet. Genomics* 292, 1237–1245. doi:10.1007/s00438-017-1343-z

Pratap, A., Prajapati, U., Singh, C. M., Gupta, S., Rathore, M., Malviya, N., et al. (2018). Potential, constraints and applications of *in vitro* methods in improving grain legumes. *Plant Breed.* 137, 235–249. doi:10.1111/pbr.12590

Pratap, A., Douglas, C., Prajapati, U., Kumari, G., War, A. R., Tomar, R., et al. (2020). "Breeding progress and future challenges: Biotic stresses," in *The mungbean genome*. Editors R. M. Nair, R. Schafleitner, and S. H. Lee (Cham: Springer), 55–80. Compendium of Plant Genomes.

Pratap, A., Das, A., Kumar, S., and Gupta, S. (2021b). Current perspectives on introgression breeding in food legumes. *Front. Plant Sci.* 11, 589189. doi:10.3389/fpls.2020.589189

Pratap, A. (2021). "Mungbean," in *The beans and the peas: From orphan to mainstream crops*. Editors A. Pratap and S. Gupta (Duxdorf, U.K.: Elsevier), 1–32.

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

Rohilla, V., Yadav, R. K., Poonia, A., Sheoran, R., Kumari, G., Shanmugavadivel, P. S., et al. (2021). Association mapping for yield attributing traits and yellow mosaic disease resistance in mungbean (*Vigna radiata* (l.) Wilczek). *Front. Plant Sci.* 12, 749439. doi:10.3389/fpls.2021.749439

Singh, D. P., Singh, B. B., and Pratap, A. (2017). Genetic improvement of mungbean and urdbean and their role in enhancing pulse production in India. *Indian J. Genet. Plant Breed.* 76, 550–567. doi:10.5958/0975-6906.2016.00072.9

Singh, C. M., Pratap, A., Gupta, S., Biradar, R. S., and Singh, N. P. (2020). Association mapping for mungbean yellow mosaic India virus resistance in mungbean (*Vigna radiata* L. Wilczek). *3 Biotech.* 10 (2), 33. doi:10.1007/s13205-019-2035-7

Singh, I., Sandhu, J. S., Gupta, S. K., and Singh, S. (2013). Introgression of productivity and other desirable traitsfrom ricebean (*Vigna umbellata*) into black gram (*Vigna mungo*). *Plant Breed* 132, 401–406. doi:10.1111/pbr.12068

Smartt, J. (1985). Evolution of grain legumes. III. Pulses in the genus *Vigna*. *Exp. Agric.* 21, 87–100. doi:10.1017/s0014479700012370

Soares, B. L., Ferreira, P. A. A., de Oliveira-Longatti, S. M., Marra, L. M., Rufini, M., Moreira, M. J. F. M. S., et al. (2014). Cowpea symbiotic efficiency, pH and aluminum tolerance in nitrogen-fixing bacteria. *Sci. Agric.* 71, 171–180. doi:10.1590/s0103-90162014000300001

Sokolkova, A., Burlyaeva, M., Valiannikova, T., Vishnyakova, M., Schafleitner, R., Lee, C. R., et al. (2020). Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan). *BMC Plant Biol.* 20, 363. doi:10.1186/s12870-020-02579-x

Somta, P., Musch, W., Kongsamai, B., Chanprame, S., Nakasathien, S., Toojinda, T., et al. (2008). New microsatellite markers isolated from mungbean (*Vigna radiata* L. Wilczek). *Mol. Ecol. Resour.* 8, 1155–1157. doi:10.1111/j.1755-0998.2008.02219.x

Somta, P., Seehalak, W., and Srinives, P. (2009). Development, characterization and cross-species amplification of mungbean (*Vigna radiata*) genic microsatellite markers. *Conserv. Genet.* 10, 1939–1943. doi:10.1007/s10592-009-9860-x

Takahashi, Y., Somta, P., Muto, C., Iseki, K., Naito, K., Pandiyan, M., et al. (2016). Novel genetic resources in the genus *Vigna* unveiled from gene bank accessions. *PLoS ONE* 11, e0147568. doi:10.1371/journal.pone.0147568

Takahashi, Y., Muto, C., Iseki, K., Naito, K., Somta, P., Pandiyan, M., et al. (2018). A new taxonomic treatment for some wild relatives of mungbean (*Vigna radiata* L. Wilczek) based on their molecular phylogenetic relationships and morphological variations. *Genet. Resour. Crop Evol.* 65, 1109–1121. doi:10.1007/s10722-017-0599-9

Tanimoto, M., Tremblay, R., and Colasanti, J. (2008). Altered gravitropic response, amyloplast sedimentation and circumnutation in the Arabidopsis shoot gravitropism 5 mutant are associated with reduced starch levels. *Plant Mol. Biol.* 67, 57–69. doi:10.1007/s11103-008-9301-0

Tantasawat, P. A., Poolsawat, O., Arsakit, K., and Papan, P. (2020). Identification of ISSR, ISSR-RGA and SSR markers associated with cercospora leaf spot resistance gene in mungbean. *Int. J. Agric. Biol.* 23, 447–453. doi:10.17957/IJAB/15.1308

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11, 1441–1452. doi:10.1101/gr.184001

Tomooka, N., Kaga, A., and Vaughan, D. A. (2006). "The asian *vigna* (*vigna* subgenus ceratotropis) biodiversity and evolution," in *Plant genome: biodiversity and evolution. Part C: phanerogams (angiosperms– dicotyledons)*. Editors A. K. Sharma and A. Sharma (Science Enfield), Vol. 1, 87–126.

Vadivel, K., Manivannan, N., Mahalingam, A., Satya, V. K., Vanniarajan, C., and Ragul, S. (2021). Identification and validation of quantitative trait loci of mungbean yellow mosaic virus disease resistance in blackgram (*Vigna mungo* L. Hepper). *Legume Res. LR* 4459. doi:10.18805/LR-4459

Wang, X. W., Kaga, A., Tomooka, N., and Vaughan, D. A. (2004). The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean (*Vigna* angularis Willd. Ohwi & Ohashi). *Theor. Appl. Genet.* 109, 352–360. doi:10.1007/s00122-004-1634-8

Wang, L. X., Cheng, X. Z., Wang, S. H., Liu, C. Y., and Liang, H. (2009). Transferability of SSR from adzuki bean to mungbean. *A. A. S.* 35, 816–820. doi:10.3724/sp.j.1006.2009.00816

Wu, X., Islam, A. S. M. F., Limpot, N., Mackasmiel, L., Mierzwa, J., Cortés, A. J., et al. (2020). Genome-wide SNP identification and association mapping for seed mineral concentration in mungbean (*Vigna radiata* L.). *Front. Genet.* 11, 656. doi:10.3389/fgene.2020.00656

Yoshida, Y., Marubodee, R., Ogiso-Tanaka, E., Iseki, K., Isemura, T., Takahashi, Y., et al. (2016). Salt tolerance in wild relatives of adzuki bean, *Vigna angularis* (Willd.) Ohwi et Ohashi. *Genet. Resour. Crop Evol.* 63, 627–637. doi:10.1007/s10722-015-0272-0

Yoshihara, T., Spalding, E. P., and Iino, M. (2013). AtLAZY1 is a signaling component required for gravitropism of the *Arabidopsis thaliana* inflorescence. *Plant J.* 74, 267–279. doi:10.1111/tpj.12118

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 14 November 2022
DOI 10.3389/fgene.2022.939182

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Yingpeng Han,
Northeast Agricultural University, China
Asela Wijeratne,
University of Memphis, United States

*CORRESPONDENCE
Milind B. Ratnaparkhe,
milind.ratnaparkhe@icar.gov.in

†These authors share first authorship

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

RECEIVED 08 May 2022
ACCEPTED 21 October 2022
PUBLISHED 14 November 2022

CITATION
Chandra S, Choudhary M, Bagaria PK, Nataraj V, Kumawat G, Choudhary JR, Sonah H, Gupta S, Wani SH and Ratnaparkhe MB (2022), Progress and prospectus in genetics and genomics of *Phytophthora* root and stem rot resistance in soybean (*Glycine max* L.). *Front. Genet.* 13:939182.
doi: 10.3389/fgene.2022.939182

COPYRIGHT
© 2022 Chandra, Choudhary, Bagaria, Nataraj, Kumawat, Choudhary, Sonah, Gupta, Wani and Ratnaparkhe. 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.

Progress and prospectus in genetics and genomics of *Phytophthora* root and stem rot resistance in soybean (*Glycine max* L.)

Subhash Chandra[1†], Mukesh Choudhary[2†], Pravin K. Bagaria[3], Vennampally Nataraj[1], Giriraj Kumawat[1], Jeet Ram Choudhary[4], Humira Sonah[5], Sanjay Gupta[1], Shabir Hussain Wani[6] and Milind B. Ratnaparkhe[1*]

[1]ICAR-Indian Institute of Soybean Research, Indore, India, [2]ICAR-Indian Institute of Maize Research, Ludhiana, India, [3]Department of Plant Pathology, Punjab Agricultural University, Ludhiana, India, [4]ICAR-Indian Agricultural Research Institute, New Delhi, India, [5]National Agri-Food Biotechnology Institute, Mohali, India, [6]Mountain Research Centre for Field Crops, Sher-e-Kashmir University of Agricultural Sciences and Technology, Srinagar, Jammu and Kashmir, India

Soybean is one of the largest sources of protein and oil in the world and is also considered a "super crop" due to several industrial advantages. However, enhanced acreage and adoption of monoculture practices rendered the crop vulnerable to several diseases. Phytophthora root and stem rot (PRSR) caused by *Phytophthora sojae* is one of the most prevalent diseases adversely affecting soybean production globally. Deployment of genetic resistance is the most sustainable approach for avoiding yield losses due to this disease. PRSR resistance is complex in nature and difficult to address by conventional breeding alone. Genetic mapping through a cost-effective sequencing platform facilitates identification of candidate genes and associated molecular markers for genetic improvement against PRSR. Furthermore, with the help of novel genomic approaches, identification and functional characterization of *Rps* (resistance to *Phytophthora sojae*) have also progressed in the recent past, and more than 30 *Rps* genes imparting complete resistance to different PRSR pathotypes have been reported. In addition, many genomic regions imparting partial resistance have also been identified. Furthermore, the adoption of emerging approaches like genome editing, genomic-assisted breeding, and genomic selection can assist in the functional characterization of novel genes and their rapid introgression for PRSR resistance. Hence, in the near future, soybean growers will likely witness an increase in production by adopting PRSR-resistant cultivars. This review highlights the progress made in deciphering the genetic architecture of PRSR resistance, genomic advances, and future perspectives for the deployment of PRSR resistance in soybean for the sustainable management of PRSR disease.

KEYWORDS

soybean, *Phytophthora*, disease resistance, gene stacking, sustainable management, genomic approaches

Introduction

Soybean (*Glycine max* (L.) Merrill) is an important legume crop that fulfills the substantial demand for food and feed globally. It is high in protein and oil content and also serves as a source of nutraceuticals such as bioflavonoids, lecithins, phytosterols, saponins, and tocopherols. Its oil is mainly used for domestic purposes; however, recent trends move toward the use of soybean oil as biodiesel to decrease reliance on fossil fuels (Mofijur et al., 2013). Approximately 70% of soybean's economic value is for its meal, of which 97% is consumed as livestock and poultry feed (Raghuvanshi and Bisht 2010). The demand for soybean in the international market is increasing due to interest in functional food and the use of various soybean seed constituents and by-products in a wide array of specific industrial products (Kumawat et al., 2016). Globally, soybean is grown in an area of 122.6 million hectares (mha) with an annual average production of 336.6 million tons (mt) (USDA, 2020). The leading producers of soybean are Brazil, the United States, China, Argentina, and India (USDA, 2020). Like other food crops, soybean production is being challenged by various forms of abiotic and biotic stresses. The remarkable growth in the number of major diseases and their area has been observed in the past 50 years; subsequently, it negatively affects soybean production worldwide. The impact of diseases on soybean may be cited to the fact that the average annual economic loss due to soybean diseases in the US reached nearly $4.55 billion based on an investigation from 1996 to 2016 (Bandara et al., 2020). This increase in the number and spread of diseases can be attributed to enhanced acreage in new un-adapted regions and monoculture practices resulting in high pathogen density. Various factors governing the disease severity and economic losses include the pathogen type, plant tissue under attack, affected number of plants, the severity of an attack, pathogen-favoring environment, host plant vulnerability, plant stress level, and crop development stages (Hartman and Hill, 2010).

Among the various soybean diseases, Phytophthora root and stem rot (PRSR), caused by the soil-borne pathogen *Phytophthora sojae* (Kaufmann and Gerdemann) (oomycete pathogen), is the second most important economic disease after soybean cyst nematodes in the world. Earlier, *P. sojae* was part of the *Phytophthora megasperma* species complex which causes rot diseases in plants (Lin et al., 2021). *Phytophthora sansomeana* is identified as another causal agent for root rot in soybean, and partial resistance to *P. sansomeana* in soybean has been observed (Lin et al., 2021). Like *P. sojae, P. sansomeana* is also part of the *P. megasperma* complex. However, stem and root rot caused by *P. sansomeana* is not included in this review to keep the article length in check. PRSR drastically limits the yields of soybean globally as losses caused by it range between 10 and 40% or complete yield loss in some scenarios (Xiao et al., 2011; Zhang et al., 2013a). In the last few decades, *Phytophthora* root and stem rot resistance has been characterized by many researchers (Ryley et al., 1998; Dorrance et al., 2003; Grau et al., 2004; Sugimoto et al., 2006; Dorrance et al., 2008). In the United States, a loss of nearly 20.5 million tons was reported from 1996 to 2014, with an average annual loss of over 1.1 million tons due to this pathogen (Allen et al., 2017). *P. sojae* was first reported in Indiana state of the United States in 1948 (Kaufmann, 1957). Later, it spread to the major soybean growing areas of the United States, particularly in the pathogen-favoring environment of the Northern United States (Dorrance and Schmitthenner 2000). In addition to the United States, PRSR has been reported in other soybean-producing continents, namely, Asia, Africa, Australia, and Europe. The occurrence and development of PRSR are facilitated by poorly drained clay soils, low temperatures, and high rains (Kaufmann, 1957; Han et al., 2008). PRSR is generally characterized by the damping-off of seedlings and rotting of roots in adult plants (Tyler, 2007), and affected plants exhibit red–brown water-soaked lesions, wilting, and chlorosis, which in the case of extreme severity leads to mortality (Schmitthenner 1985; Dorrance et al., 2003).

P. sojae has abundant pathogenic diversity, and complete and partial resistance reactions have been reported for this pathogen. PRSR is being managed by cultivars with one or two dominant resistance genes for *Phytophthora sojae* named "*Rps*" (Jang and Lee 2020). However, the *Rps* genes are race-specific and useful as introgression of such genes is easy; but partial resistance has its own advantage for long-term protection. The first resistance gene against soybean *P. sojae* (named *Rps1a*) was identified in the 1950s (Bernard et al., 1957). Later, with the advent of sequencing technology and development of abundant simple sequence repeat (SSR) markers (Song et al., 2010), molecular linkage mapping gained pace, and nearly 30 *Rps* genes have been identified to date (Lin et al., 2013; Sun et al., 2014a; Ping et al., 2016; Niu et al., 2017; Zhong et al., 2019; Zhong et al., 2019; Jiang et al., 2020). In addition to SSR markers, a large number of single nucleotide polymorphisms (SNPs) and insertion/deletion markers for fine genetic mapping and molecular breeding have also been studied in different mapping populations (Li et al., 2016a; Li et al., 2017a). Functional characterization of identified genes has also gained progress in the recent past (Fan et al., 2015; Fan et al., 2017; Jang and Lee, 2020; Zhou et al., 2022). Few transcriptomic studies also uncovered molecular pathways in response to *P. sojae* infection (Guo et al., 2011; Lin et al., 2014). Newly identified genes for PRSR resistance serve as a good source for modern breeding programs to improve the resistance of cultivars to PRSR disease. Furthermore, the identified quantitative disease resistance loci (QDRL) can be employed

in gene stock mining for the identification of novel alleles. This review aims to provide the current progress and future perspectives on genetics and genomics-assisted studies of *P. sojae* *R*-genes/QDRL and their utilization in soybean improvement.

Disease management, pathogenic diversity, and potential genetic resources

PRSR is a serious concern today as it causes a significant yield loss in soybean production. Current PRSR control strategies include applications of various fungicides (Anderson and Buzzell, 1992), improving soil drainage systems (Schmitthenner, 1985), tillage systems (Workneh et al., 1998), application of calcium-containing fertilizers (Sugimoto et al., 2010), and the use of resistant varieties (Schmitthenner, 1999; Dorrance et al., 2003). Germplasm screening-based identification of resistant genotypes and development of PRSR-resistant soybean cultivars is the most effective and sustainable approach for minimizing yield losses (Burnham et al., 2003).

Currently, the management of PRSR is largely dependent on resistant cultivars, having one or more resistance genes. For understanding, pathogen race refers to a pathogen's ability to cause disease in its host (Anderson et al., 2010); in other words, the pathogen race attacks certain resistance genes (Dorrance et al., 2016), and this kind of resistance is accompanied by several mechanisms including effector-triggered immunity (ETI), where R gene products in the host is recognized (directly/indirectly) by specific pathogen effectors termed avirulence (Avr) proteins (Li et al., 2021). Till now, nine *Avr* genes of *P. sojae* have been cloned (Yang et al., 2019). Soybean R genes whose products recognize *P. sojae* Avr effectors and trigger *Phytophthora* resistance are known as *Rps* (resistance to *P. sojae*) genes (Tyler and Gijzen, 2014). For understanding, avirulence 1c (*Avr1c*) gene in *P. sojae* confers the resistance by *Rps 1c* gene in soybean populations; the K105 amino-acid residue in *Avr* gene is the main determinant of the avirulence of *Avr1c* that interacts with *Rps* gene (Yang et al., 2019). So *Rps* genes have the potential to combat PRSR, but they are race-specific; therefore, they would be operational against limited *P. sojae* isolates, and each *Rps* gene often remains effective for about 8–15 years, which leads to the emergence of new isolates after a certain period (Schmitthenner 1985; Dorrance et al., 2003).

The diversity of the *P. sojae* population has been investigated in the United States and Canada since the 1960s, and most of the *P. sojae* isolates were determined based on studies with 15 host-differentials (Dorrance et al., 2004). During the 1980s, *Rps* genes *1a*, *1d*, and *1k* have been widely exploited to combat PRSR losses; however, the emergence of new isolates and enhancement in virulence lead to the evolution of more than 55 races reported against eight soybean differentials (*Rps1a*, *1b*, *1c*, *1d*, *1k*, *2*, *3a*, *6*, and *7* genes) (Abney et al., 1997; Grau et al., 2004). A total of 213 virulent pathotypes were identified from 873 isolates of the North Central United States (Dorrance et al., 2016).

In China, after the identification of *P. sojae* in the Heilongjiang region in 1991 (Su and Shen 1993), the incidence of the pathogen was reported mainly in the Inner Mongolia Autonomous Region, Xinjiang Uygur Autonomous Region, and Fujian Province till 2015 (Wen and Chen 2002; Chen et al., 2004; Liu et al., 2006; Xiao et al., 2011). During 2005–2007 in Heilongjiang Province, a total of 96 isolates were collected and investigated, which revealed that four out of the eight races had new pathotypes (Zhang et al., 2010).

PRSR was first reported in Hokkaido, Japan, in 1977 (Tsuchiya, 1990). Tsuchiya (1990) found the genetic differences between American and Japanese isolates during the investigation of 49 Japanese isolates and 55 known American *P. sojae* races. Sugimoto et al. (2006) collected 51 isolates from Hyogo in Japan and identified four new races. More than 100 *P. sojae* isolates were reported from 14 different regions for 14 *Rps* genes including *Rps1a*, *1b*, *1c*, *1d, 1k*, *3b*, *7*, and *8*; among them, *Rps1d* and *1k* were determined as the most promising resistance genes (Moriwaki 2010). Similarly, in Brazil, *P. sojae* was found to be different from other regions; 17 pathotypes were determined based on 37 Brazilian isolates, which were genetically different from the previously reported ones (Costamilan et al., 2013). Subsequently, *Rps1a*, *1c*, and *1k* were highly utilized in Brazilian soybean breeding programs, whereas *Rps1a* and *1c* were not effective in the United States (Costamilan et al., 2013). In South Korea, PRSR was first reported 2 decades ago (Jee et al., 1998). Kang et al. (2019) reported genetic differences among the pathotypes of Korean *P. sojae* isolates. Thus, based on these facts, it is a prerequisite to collect *P. sojae* isolates from several regions/fields and to assess them with differential varieties.

A number of genetic sources for *P. sojae* resistance have been identified and utilized to map *Rps* genes and to develop resistant cultivars through different breeding strategies (Table 1). Similarly, numerous genetic resources for incomplete or partial resistance for *P. sojae* have been utilized for genetic studies in the form of breeding lines and introgression (Table 2). Dorrance and Schmitthenner (2000) evaluated over 1,000 accessions from USDA germplasm accessions and found 162 accessions to be resistant to three races (7, 17, and 25). In addition to this, they also reported partial resistance in 55.5% of the 887 accessions for *P. sojae*. Kang et al. (2019) evaluated the *Rps* resistance against four isolates in 20 popular varieties of South Korea, while Daewon was identified as a resistant cultivar.

Genetics of complete resistance versus partial resistance

There are two types of resistance to *P. sojae* reported in soybean, namely, complete resistance and partial resistance

TABLE 1 Details of *P. sojae*-resistant genes (*Rps*), their source, chromosomal positions, and associated markers.

S. No.	Name of *Rps* gene	Chr no. (LG)	Source	[a]Position 1 (Mbp)	[a]Position 2 (Mbp)	Flanking marker 1	Flanking marker 2	References
1	*Rps1a*	3 (N)	L88-8470, Mukden, and Harlon	3.2	3.9	Satt159 (BARCSOYSSR_03_0180)	Satt009 (BARCSOYSSR_03_0226)	Bernard et al. (1957), Weng et al. (2001)
2	*Rps1b*	3 (N)	L77-1863	3.4	5.7	Satt152 (BARCSOYSSR_03_0192)	Satt530	Mueller et al. (1978), Demirbas et al. (2001)
3	*Rps1c*	3 (N)	L75-3735	3.4	9.2	Satt152 (BARCSOYSSR_03_0192)	Satt584 (BARCSOYSSR_03_0442)	Mueller et al. (1978), Demirbas et al. (2001)
4	*Rps1d*	3 (N)	L93-3312 and PI 103091	3.4	3.5	Satt152 (BARCSOYSSR_03_0192)	Sat_186 (BARCSOYSSR_03_0204)	Buzzell and Anderson (1992), Sugimoto et al. (2008)
5	*Rps1k*	3 (N)	L77-1794, Williams82, and E00003	-	-	CG1 (AFLP)	-	Bernard and Cremeens (1981), Kasuga et al. (1997), Gao and Bhattacharyya (2008)
6	*Rps2*	16 (J)	L76-1988, P I398440, and P I398694	1.64	34.03	Satt287 (BARCSOYSSR_16_0090)	Satt547 (BARCSOYSSR_16_1165)	Kilen et al. (1974), Demirbas et al. (2001), Gordon et al. (2007)
7	*Rps3a*[b]	13 (F)	L83-570, P I3 99036, PI408097, and PI424354	23.68	37.6	Satt374 (Sat_309)	Satt144	Mueller et al. (1978), Demirbas et al. (2001), Gordon et al. (2007)
8	*Rps3b*	13 (F)	L91-8347	-	-	-	-	Ploper et al. (1985), Demirbas et al. (2001)
9	*Rps3c*	13 (F)	L92-7857	-	-	-	-	Sugimoto et al. (2012), Demirbas et al. (2001)
10	*Rps4*	18 (G)	L85-2,352, and PI399036	53.8	56.3	Satt191 (BARCSOYSSR_18_1750)	Sat_064 (BARCSOYSSR_18_1858)	Athow et al. (1980), Demirbas et al. (2001), Sandhu et al. (2004), Gordon et al. (2007)
11	*Rps5*	18 (G)	L85-3059 and PI399036	-	53.9	-	Satt472 (BARCSOYSSR_18_1708)	Buzzell and Anderson, 1981; Sahoo et al., 2017
12	*Rps6*	18 (G)	L89-1,581, PI399079, and PI399036	54.5	-	Satt191 (BARCSOYSSR_18_1750)	Sat_372	Athow and Laviolette (1982), Gordon et al. (2007)
13	*Rps7*	3 (N)	L93-3258, OX281, and PI408097	3.9	18.4	Satt009 (BARCSOYSSR_03_0226)	Satt125 (BARCSOYSSR_03_0564)	Anderson and Buzzel (1992), Weng et al. (2001), Gordon et al. (2007)
14	*Rps8*	13 (F)	PI 399073	24.3	28.9	Satt425 (BARCSOYSSR_13_0784)	Satt114 (BARCSOYSSR_13_1055)	Gordon et al. (2004), Gordon et al. (2006), Sandhu et al. (2005)
15	*Rps9*	3 (N)	Ludou 4 and Cangdou 5	2.94	3.15	Satt631 (BARCSOYSSR_03_0162)	Sat_186 (BARCSOYSSR_03_0204)	Wu et al. (2011a)
16	*Rps10*	17 (D2)	Wandou 15	30.8	31.1	Sattwd15-24	Sattwd15-47	Zhang et al. (2013a)
17	*Rps11*	7 (M)	PI 594527	5.42	5.77	BARCSOYSSR_07_0266	BARCSOYSSR_07_0300	Ping et al. (2016)
18	*Rps12*	18 (G)	PI 399036	56	56.3	BARCSOYSSR_18_1840	Sat_064	Sahoo et al. (2017)
19	*Rps13*	18 (G)	PI 399036	-	-	Sat_064	BARCSOYSSR_18_1859	Sahoo et al. (2021)
20	*RpsUN1*	3 (N)	PI 567139B	3.2	4.3	Satt159 (BARCSOYSSR_03_0180)	BARCSOYSSR_03_0250	Lin et al. (2013), Li et al. (2016a)
21	*RpsUN2*	16 (J)	PI 567139B	36.9	37.3	BARCSOYSSR_16_1275	Sat_144 (BARCSOYSSR_16_1294	Lin et al. (2013), Li et al. (2016a)
22	*Rps Yu25*	3 (N)	Zheng 92116	3.19	3.33	Sat_186 (BARCSOYSSR_03_0204)	Satt_152	Sun et al. (2011)
23	*RpsYD29*	3 (N)	Yudou 29	3.9	4.1	SattWM82-50	Satt1k4b	Zhang et al. (2013b)
24	*RpsYD25*	3 (N)	Yudou 25	2.2	4.5	Satt1k3	BARCSOYSSR_03_0253	Fan et al. (2009), Zhong et al. (2020)

(Continued on following page)

TABLE 1 (*Continued*) Details of *P. sojae*-resistant genes (*Rps*), their source, chromosomal positions, and associated markers.

S. No.	Name of *Rps* gene	Chr no. (LG)	Source	[a]Position 1 (Mbp)	[a]Position 2 (Mbp)	Flanking marker 1	Flanking marker 2	References
25	*RpsYB30*	19 (L)	Youbian 30	33.9	34.8	Satt497 (BARCSOYSSR_19_0760)	Satt313 (BARCSOYSSR_19_0788)	Zhendong et al. (2010)
26	*RpsSu*	10 (O)	Su88-M21	1	39.4	Satt358	Sat_242 (BARCSOYSSR_10_1104)	Wu et al. (2011b)
27	*RpsZS18*	2 (D1b)	Zaoshu18	43.37	44.3	ZCSSR33	ZCSSR46	Yao et al. (2010), Zhong et al. (2018a)
28	*RpsSN10*	13 (F)	Suinong 10	16.6	16.9	Satt423 (BARCSOYSSR_13_0264)	Satt149 (BARCSOYSSR_13_0245)	Yu et al. (2010)
29	*Rps1?*	3 (N)	Waseshiroge	3.9	4.5	Satt009 (BARCSOYSSR_03_0226)	T0003044871	Sugimoto et al. (2011)
30	*RpsJS*	18 (G)	Nannong 10–1	56.3	56.6	BARCSOYSSR_18_1859	BARCSOYSSR_18_1864	Sun et al. (2014a)
31	*RpsWY*	3 (N)	Wayao	2.9	3.4	Satt631 (BARCSOYSSR_03_0162)	Satt152 (BARCSOYSSR_03_0192)	Cheng et al. (2017)
32	*RpsQ*	3 (N)	Qichadou 1	3	3.1	BARCSOYSSR_03_0165	InDel281	Li et al. (2017a)
33	*RpsHN*	3 (N)	Meng8206	4.2	4.5	SSRSOYN-25	SSRSOYN-44	Niu et al. (2017)
34	*Rps HC18*	3 (N)	Huachun 18	4.5	4.6	BARCSOYSSR_03_0269	BARCSOYSSR_03_0272	Zhong et al. (2018b)
35	*RpsX*	3 (N)	Xiu94-11	2.9	3.2	InDelxz6	BARCSOYSSR_03_0175	Zhong et al. (2019)
36	*RpsGZ*	3 (N)	Guizao1	32.3	-	Gm_03_bin31	-	Jiang et al. (2020)
37	-	16 (J)	-		4.0*	-	BARC-014467–01559 -	Huang et al. (2016)
38	-	20 (I)	-	-	46.6*	BARC-013645–01207	-	Huang et al. (2016)

[a]Physical position of the left marker and right flanking markers is based on the genome assembly Wms.82. v1. a2 and approximate physical positions with an asterisk (*) are based on the genome assembly Glyma. Wm82. a1.
[b]Physical positions and associated markers on *Rps3a* are based on Gordon et al. (2007).

(Sugimoto et al., 2012). Complete resistance is race-specific and exhibits a single dominant resistance gene (*Rps*) that provides immunity or near immunity, whereas partial resistance is controlled by major and minor genes, and it restricts pathogen colonization and spread (Dorrance et al., 2003, 2004; Sugimoto et al., 2012). Previous studies over the last 2–3 decades identified both complete and partial resistance to *P. sojae* (Burnham et al., 2003; Dorrance et al., 2004; Sugimoto et al., 2012; Jang and Lee 2020).

During the single dominant gene resistance mechanism against *P. sojae*, expressed products of *Rps* genes interact with those of *P. sojae* through a gene-for-gene interaction and prevent disease development in plants (Hartwig et al., 1968; Schmitthenner, 1999; Burnham et al., 2003; Fan et al., 2009). There are very few reports explaining the detailed expressed products of *Rps* genes; Gao et al. (2005) mentioned the role of coiled-coil–nucleotide-binding–leucine-rich repeat (CC-NB-LRR)-type proteins in the case of *Rps1*-k locus ,and Li et al. (2021) demonstrated that *E3 ligase GmPUB1* protein is required for the interaction of *P. sojae* effector protein *Avr1b* with the resistance of *Rps1b* and *Rps1k* in soybean. As an example, during an investigation of the inheritance pattern of *Rps* genes, Li et al. (2017b) used detached-petiole and hypocotyl inoculation methods in F_2 and $F_{2:3}$ populations derived from a cross "Zhonghuang47" × "Xiu94-11." A segregation ratio of 3:1 for the resistance and the susceptible reaction indicated a single dominant gene for *P. sojae* resistance in their study. All the *Rps* genes provide race-specific and complete resistance with the exception of *Rps2*, which provides incomplete resistance (Mideros et al., 2007).

Genes for complete resistance

To the best of our knowledge, more than 30 *Rps* genes/alleles have been reported and are present on 10 different chromosomes in soybean (Table 1). Most of the *Rps* loci are located on chromosome 3 (14 genes), followed by chromosome 18 (6 genes) and chromosome 13 (5 genes). The *Rps* genes on these three chromosomes constitute approximately 70% of the total *Rps* genes reported (Figure 1). *Rps1* (with five different alleles, *Rps1a*, *Rps1b*, *Rps1c*, *Rps1d*, and *Rps1k*), *Rps7*, *Rps9*, *RpsYu25*, *RpsYD29*, *RpsYD25*, *RpsUN1*, *RpsWY*, *RpsQ*, *RpsHC18*, *RpsX*, *RpsHN*, and *RpsGZ* and an unnamed *Rps* gene (*Rps1?*) were mapped on the short arm of chromosome 3 (Figure 1; Table 1). Similarly, *Rps4*, *Rps5*, *Rps6*, *Rps12*, *Rps13*, and *RpsJS* are located on chromosome 18; *Rps2*, *RpsUN2*, and one

TABLE 2 Details of QDRL for *P. sojae* resistance identified through the bi-parental mapping approach in soybean.

[b]Plant material/ population	QDRL/ genomic region	[a]Potential linked marker	Chromosome	PVE/ R^2 (%)	Marker	Environment	Reference
Conrad × Sloan (RILs)	2	Satt579 and Satt600; Satt252 and Satt149	2 and 13	10.6 and 32.4	SSR	Growth chamber	Burnham et al. (2003)
Conrad × Harosoy (RILs)	2	Satt266 and Satt579; Satt252 and Satt423	2 and 13	15.9–35.0	SSR	Growth chamber	Burnham et al. (2003)
Conrad × Williams (RILs)	2	Satt579 and Satt600; Satt252 and Satt149	2 and 13	20.7–21.4	SSR	Growth chamber	Burnham et al. (2003)
Conrad × OX760-6-1 (RILs)	1	Satt414 and Satt596	16	13.7–21.5	SSR	Field	Weng et al. (2007)
Conrad × OX760-6-1 (RILs)	3	OPL18 and Satt274; Satt509 and Satt030; Satt343 and OPG16600	2 and 13	2.4–21.6%	RAPDs and SSR	Greenhouse	Han et al. (2008)
Hefeng 25 × Conrad (RILs)	8	Satt579 and Sat_089; Satt325 and Satt343; Satt277 and Satt365	2.6, 8, 11, and 13	4.24–27.98	SSR	Greenhouse and Field	Li et al. (2010)
V71-370 × PI407162 (RILs)	3	Satt414, Satt529, Sat_163, and SLP142	16, 18, and 20	7–32	SSR	Greenhouse	Tucker et al. (2010)
Conrad × Sloan (RILs)	5	Satt353, Sct_033, Satt574, GMH_OSU31, GML_OSU10, and F424_294	12, 13, 14, 17, and 19	4–7	SSR; SNP	Greenhouse	Wang et al. (2010)
Su88-M21 × Xinyixiaoheidou (RILs)	3	Satt520, Satt557,Satt598, Satt651, Satt420, and Sat_274	6, 10, and 15	4.3–15.9	SSR	Greenhouse	Wu et al. (2011c)
Conrad × Sloan (RILs)	5	Satt527, BARCSOYSSR_19_1473, BARC-060037–16311, and BARCSOYSSR_18_1777	1, 18, and 19	4.8–19.6	SNP	Greenhouse	Wang et al. (2012)
S99–2,281 × PI 408105A (RILs)	2	Sat_154 Sat_375, Sat_300, and BARC-023721–03465	13 and 17	7.5–35.8	SSR; SNP	Greenhouse	Nguyen et al. (2012)
OX20–8 × PI 398841, (RILs)	3	BARC-044479–08708, BARCSOYSSR_13_1103, BARC-031343–07057, (BARCSOYSSR_13_0981), and BARCSOYSSR_13_1131	1, 13, and 18	4–16	SNP	Field	Lee et al. (2013a)
OX20–8 × PI 407861A (RILs)	9	BARC-051883–11286, Sat_234, and BARCSOYSSR_15_0160	3, 4, 8, 10, 13, 15, and 18	2.4–8.6	SNP	Greenhouse	Lee et al. (2013b)
Combined populations (6 NAM)	16	BARC-025777–05064, BARC-047665–10370, BARCSOYSSR13_1106 and BARCSOYSSR13_1103	1, 3, 12, 13, 16, and 18	4–45	SNP	Greenhouse amd field	Lee et al. (2014)
Conrad × Sloan, (RILs)	10	BARC_2.0_Gm18_56710850, BARC_2.0_Gm18_56876857, BARCSOYSSR_19_1286 and BARC_2.0_Gm19_46116996	1, 4, 9, 15, 16, 18, and 19	2–13.6	SNP	Greenhouse	Stasko et al. (2016)
PI 399036 × AR2 (AX20925) (RILs)	6	BARC-064609–18739, BARC-039977–07624, BARC-042881–08448 and BARC-019805–04379	2, 3, 6, 12, 15, and 19	5–14	SSR	Growth chamber	Abeysekara et al. (2016)
PI 399036 × AR3 (AX20931) (RILs)	7	BARC-065787–19749, BARC-056237–14178, BARC-017625–02635 and BARC-055533–13402	2, 7, 5, 8, 9, 13, 14, 15, 17, and 20	5–30	SSR	Growth chamber	Abeysekara et al. (2016)
Combined populations (2 NAM)	4	Gm13_29043806_T_C, Gm13_39560450_G_A, and Gm06_11776489_C_A	6, 13, and 18	7–42.2	SNP	Growth chamber	Scott et al. (2019)
PI 449459 × Misty	2	Chr13:28842184, Chr13:30776191, Chr19:50040258, and Chr19:50556102	13 and 19	13.1–17.6	SNP	Growth chamber	de Ronne et al. (2019)
Hefeng 25 ×DongongL-28	2	Chr03-41803925, Chr03-41822143, Chr03-3904775, and Chr03-4404630	3	5.8–56.0	SNP and SLAF	Growth chamber	Zhao et al. (2020)
Williams × PI 407974B and Williams × PI 424487B	3	ss715586321, ss715632438, and ss715632427	3 and 18	56–89	SSR and SNP	-	Bolanos-Carri et al. (2021)

[a]Markers which explained maximum phenotypic variations and near identified genomic regions.
[b]PRSR, resistant parent depicted in bold letters.

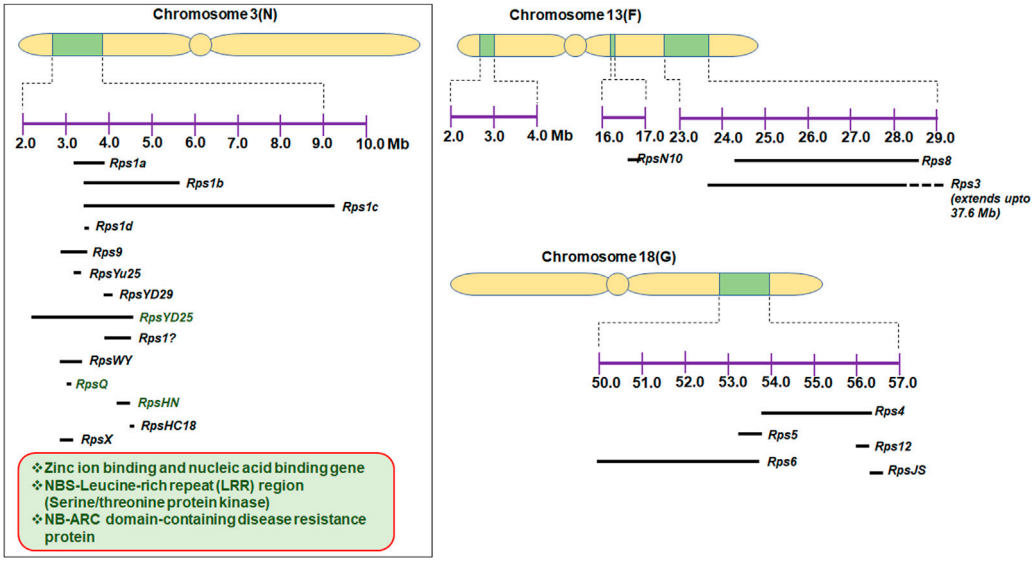


FIGURE 1
Genomic regions of chromosomes 3, 13, and 18, where more than 20 *Rps* genes were mapped; some potential characterized candidate genes are also depicted on chromosome 3.

unknown *Rps* are located on chromosome 16; *Rps3* (three alleles, *Rps3a*, *Rps3b*, and *Rps3c*) and *RpsSN10* which was linked with *Rps8* were mapped on chromosome 13. Furthermore, the remaining genes, namely, *RpsZS18*, *Rps11*, *RpsSu*, *Rps10*, and *RpsYB30*, and an unnamed *Rps* were identified on chromosomes 2, 7, 10, 17, 19, and 20, respectively (Sandhu et al., 2004; Sandhu et al., 2005; Gordon et al., 2007; Yao et al., 2010; Yu et al., 2010; Wu et al., 2011b; Zhang et al., 2013a; Lin et al., 2013; Sun et al., 2014a; Li et al., 2016a; Huang et al., 2016; Ping et al., 2016; Sahoo et al., 2017; Sahoo et al., 2021). A few genomic regions were repeatedly found in many mapping studies using bi-parent populations. For example, on chromosome 3, a genomic region of ~2 Mb was found to be a hot spot, where major resistance was identified in over 10 investigations using different resistance sources (Figure 1). Zhong et al. (2019) identified *RpsX* in soybean cultivar Xiu94-11; subsequently, it was revealed that *RpsX* was located in the 242-kb genomic region spanning the *RpsQ* locus on chromosome 3. Zhong et al. (2020) fine-mapped *RpsYD25* in 1127 $F_{3:4}$ families derived from "Zaoshu18" and "Yudou25;" subsequently, 7 out of 178 soybean genotypes containing *RpsYD25* were identified using five co-segregated SSR markers. Recently, Jiang et al. (2020) have fine-mapped *RpsGZ* to a 367.371-kb genomic region on chromosome 3 in recombinant inbred lines (RILs) derived from a cross of the resistant cultivar "Guizao1" and the susceptible cultivar "BRSMG68." Sahoo et al. (2017) identified *Rps12* on chromosome 18 in an RIL population developed by crossing the *P. sojae* resistant cultivar "PI399036" with the susceptible "AR2" line, and this gene was mapped at 2.2 cM proximal to the *NBSRps4/6*-like sequence that co-segregated with the *Phytophthora* resistance genes *Rps4* and *Rps6*.

In general, *Rps* gene efficacy is limited to 8–15 years (Grau et al., 2004; Sugimoto et al., 2012). Therefore, continuous efforts are required for the identification of new *Rps* genes and for the development of PRSR-tolerant cultivars. Under the conditions of high disease pressure, cultivars with complete resistance are far more effective than those having partial resistance to *P. sojae* (Schmitthenner, 1999; Dorrance et al., 2003). Contradictorily, partial resistance conferred by many QDRL has been found to be durable compared to complete resistance (single *Rps* gene) in the United States where *P. sojae* races evolve at a much faster rate to knock down even the most effective *Rps* genes (Dorrance et al., 2003). This indicates the significance of both complete resistance and partial resistance to *P. sojae* in different situations.

Quantitative disease resistance loci for partial resistance

Partial resistance to *P. sojae* is a quantitative trait which is usually race non-specific and provides long-term resistance stability against the pathogen (Schmitthenner 1985; Dorrance et al., 2003; Dorrance et al., 2008; Wang et al., 2012). Partial resistance has moderate to high heritability and thus can be improved through selection pressure. For stable and durable

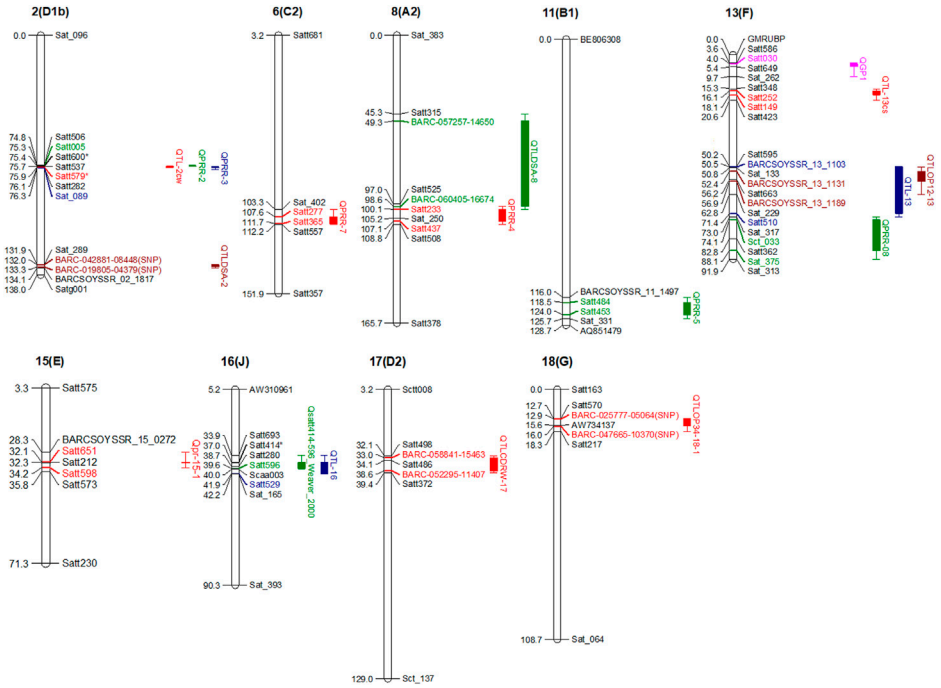


FIGURE 2
Major QDRL (phenotypic variations explaining (PVE) more than 10%) identified for *P. sojae* resistance along with their flanking makers and chromosomal locations.

management of PRSR, partial resistance along with complete resistance (*Rps* genes) may be used, as both types of resistance have different mechanisms to respond to PRSR.

Usually, the levels of partial resistance are evaluated using lesion length measurement, root rot score, tray test, inoculum layer test, or field evaluation (Tooley and Grau 1984; Dorrance et al., 2008). The development of cultivars with increased levels of partial or incomplete resistance needs the identification and characterization of novel resources of partial resistance.

Partial resistance or field resistance to *P. sojae* is governed by several genomic regions called quantitative trait loci (QTL or alternatively termed QDRL), each contributing a certain magnitude of resistance (Scott et al., 2019). There are a number of resources that have been utilized for mapping QDRL for partial resistance to *P. sojae* (Table 2). Extensive mapping studies using two contrasting parents in soybean reported about more than 90 QDRL for partial resistance to *P. sojae* (Table 2; Figure 2). Later on, the large confidence interval spanning genomic regions against *P. sojae* was further narrowed down through fine-mapping to pinpoint the exact position of QDRL (Huang et al., 2016; Karhoff et al., 2019). The cultivar "Conrad" that does not exhibit *Rps* genes but shows high partial resistance has been extensively used in QDRL mapping, identifying over 35 QDRL using different bi-parental populations (Burnham et al., 2003; Weng et al., 2007; Han et al., 2008; Li et al., 2010; Wang et al., 2010; Wang et al., 2012; Stasko et al., 2016). Some common QDRL were identified in a "Conrad" × "Sloan" population against three isolates of *P. sojae*, showing that a common resistance mechanism may occur in response to the individual inoculated isolates (Stasko et al., 2016). The detailed list of recent mapping studies using bi-parental populations leading to the identification of major QDRL along with significant markers imparting partial resistance to *P. sojae* is given in Table 2. Although over 15 QDRL explained more than 10% phenotypic variance (PV), the majority of QDRL explained <10% of the PV for partial resistance toward PRSR (Table 2) (Burnham et al., 2003; Weng et al., 2007; Han et al., 2008; Li et al., 2010; Nguyen et al., 2012; Lee et al., 2013a; Lee et al., 2014; Abeysekara et al., 2016; de Ronne et al., 2019; Scott et al., 2019). Apart from RILs, nested association mapping (NAM) populations have also been used to map QDRL associated with PRSR (Lee et al., 2014). Recently, Scott et al. (2019) have carried out inoculation of two Soy-NAM populations with *P. sojae* isolate Win371 for the identification of major QDRL (Figure 2). Four major QDRL were identified by Abeysekara et al. (2016) using RILs derived from "AX20925" (PI 399036 × AR2) and "AX20931" (PI 399036 × AR3). In the latest study, Zhao

TABLE 3 Details of genomic regions associated with *P. sojae* resistance identified through the association mapping/GWAS approach in soybean.

No. of genotypes	GWAS loci	Chromosome	PVE (%)	Markers used	Method	References
214	4	2.17	5.24–8.14	138 SSRs	GLM and MLM	Sun et al. (2014b)
797	16	3, 13, and 19	2.5–3.8	19,303 SNPs	MLM	Schneider et al. (2016)
224	14	3, 6, 4, 9, 11, 15, 16, and 20	-	1,645 SNPs	GLM and MLM	Huang et al. (2016)
44–7431	28	3, 13, and 18	-	42,449	MLM	Chang et al. (2016)
189	32	3, 4, 5, 7, 10, 13, 14, and 18	-	33,625 SNPs	GML and MLM-Q + K	Qin et al. (2017)
337	26	1	6.14–11.18	60,862 SNPs	GML and MLM-Q + K	Niu et al. (2018)
279	3	13	-	59,845 SNPs	GLM and MLM (Q + K)	Li et al. (2016b)
169	8	3, 15	13.9–21.1	3,807 SNPs	MLM	Ludke et al. (2019)
478	24	2, 3, 5, 6, 10, 11, 12, 13, 18, and 20	2.28–12.1	34,248 SNPs	MLMM	Rolling et al., 2020
495	24	2, 3, 4, 5, 8, 11, 13, 14, 15, 16, 17, and 18	0.21–13.11	33,234 SNPs	MLMM	Rolling et al., 2020
225	8	3, 7, 14, 15, and 17	25.3–33.6	28,722 SNPs	CMLM and FARMCPU	Zhao et al. (2020)
2,233 (16 panels)	75	All chromosomes	-	~33,641–40,954 SNPs	CMLM, MLMM_cof, and FARMCPU	Van et al. (2020)

[a]Markers which explained maximum phenotypic variations and near identified genomic regions.

et al. (2020) identified quantitative trait nucleotides (QTNs) explaining up to 56% PV on chromosome 3 using RILs derived from crossing "DongnongL-28" and "Hefeng 25."

Alternatively, another approach, genome-wide association studies (GWAS), provide high-resolution mapping than the traditional bi-parental mapping strategy. In a SSR-based association mapping study among 214 soybean accessions, four SSR alleles, *viz.*, *Satt634-133*, *Satt634-149*, *Sat_222-168*, and *Satt301-190*, were found to be significantly associated with *P. sojae* partial resistance (Sun et al., 2014b). Similarly, in another GWAS on resistance to 11 *P. sojae* isolates involving 224 germplasm accessions, Huang et al. (2016) identified 14 marker–trait associations for PRSR resistance including five novel loci. In USDA soybean germplasm, significant associations were detected for 28 SNPs located on chromosomes 3, 13, and 18 (Chang et al., 2016). The updated information on all GWAS conducted on soybean against PRSR is given in Table 3. The majority of association studies identified SNPs explaining small variations (minor QDRL); however, some of the studies (Ludke et al., 2019; Rolling et al., 2020; Zhao et al., 2020) identified major QDRL explaining variations for PRSR. Ludke et al. (2019) conducted a SNP-based GWAS on 169 soybean cultivars for *P. sojae* resistance and identified four QDRL on two chromosomes (two each on chromosomes 3 and 15). Interestingly, the identified genomic regions were found to be co-localized with already known and annotated resistance genes. Recently, Rolling et al. (2020) analyzed QDRL in 478 and 495 plant introductions (PIs) against *P. sojae* isolates OH.121 and C2.S1, respectively, and 24 significant associated SNPs were identified. Five QDRL identified in this study were found to be co-localized with *P. sojae* meta-QDRL identified from previous bi-parental mapping studies (Rolling et al., 2020). Using available disease phenotypic information, Van et al. (2020) identified 75 novel QTNs using 16 panels consisting of 2,233 soybean accessions. The identified SNPs linked to QDRL can be used in marker-assisted selection for introgression and stacking of partial PRSR resistance loci for imparting durable resistance.

Candidate genes for *Phytophthora* root and stem rot resistance

Characterization of putative genes imparting resistance to *P. sojae* has also been progressed. Graham et al. (2002) characterized the sequence of the *Rps2* genomic region. *Rps2* locus sequences included 16 resistance gene homologs with similarities to the *TIR/NBD/LRR* family of disease resistance genes, a leucine zipper protein, four gene sequences with similarities to Ca^{2+}-binding domains of a *calmodulin* gene, and three genes with homology to an NtPRp27-like protein (Graham et al., 2002). Sequencing of the *Rps1k* locus identified a coiled-coil–nucleotide-binding site–leucine-rich repeat (*CC-NBS-LRR*)-type gene (Bhattacharyya et al., 2005). Further characterization of *Rps1k* by bacterial artificial chromosome (BAC) sequencing revealed the presence of two nucleotide-binding site–leucine-rich repeat (NBS-LRR)-encoding genes (*Rps1k-1* and *Rps1k-2*) (Gao et al., 2005; Gao and Bhattacharyya, 2008; Sandhu et al., 2009). Sandhu et al. (2004) demonstrated that the deletion of *NBSRps4/6* in mutant M1 is correlated with the loss of *Rps4* function. With the availability of a

TABLE 4 Putative candidate genes identified for *P. sojae* resistance.

Parents	Rps genes	Gene model number	Gene function	References
Conrad × Sloan	-	*Glyma19g40800*	Transducin/WD40 domain-containing	Wang et al. (2012)
		Glyma19g40840	Pectinesterase	
		Glyma19g40940	Glycoside hydrolase family 28 protein	
		Glyma19g41590	2-Deoxyglucose-6-phosphate phosphatase	
		Glyma19g41900	Phloem-specific lectin PP2-like protein	
		Glyma19g42120	Heparan-alpha-glucosaminide N-acetyltransferase	
		Glyma19g42200	Rapid alkalinization factor	
		Glyma19g42210	Rad9	
		Glyma19g42220	Respiratory burst oxidase 2	
		Glyma19g42240	Histone H2A 7	
		Glyma19g42390	Cyclin-dependent protein kinase	
Wandou 15 and Williams	*Rps10*	*Glyma17g28950.1*	Serine/threonine (Ser/Thr) protein kinases	Zhang et al. (2013a)
		Glyma17g28970.1	Serine/threonine (Ser/Thr) protein kinases	
Jikedou 2 × Yudou 29	*RpsYD29*	*Glyma03g04030.1*	NBS-LRR	Zhang et al. (2013b), Gao and Bhattacharyya (2008)
		Glyma03g04080.1	NBS-LRR	
Nannong 10–1 (P1) × 06–070583 (P2)	*Rps JS*	*Glyma18g51930*	NBS-LRR	Sun et al. (2014a)
		Glyma18g51950	NBS-LRR	
		Glyma18g51960	NBS-LRR	
Germplasm panel (797)	-	*Glyma.03G034400*	NBR-gene	Schneider et al. (2016)
		Glyma.03G034200	Plant defense	
		Glyma.03G035700	Abscisic acid responsive stress	
		Glyma.13G194100	NB-LRR-encoding genes	
		Glyma.19G245400	PR4-related chitin-binding proteins	
		Glyma.19G248900	Ethylene/JA responsive transcription factor	
Germplasm panel (279)	-	*Glyma13g32980*	*Coat protein I* (COPI)-related gene	Li et al. (2016b)
		Glyma13g33900	2OGFE (II) oxygenase family protein	
		Glyma13g33512	Pentatricopeptide (PPR) repeat	
		Glyma13g33536	Leucine-rich repeat domain protein	
		Glyma13g33740	Leucine-rich repeat domain protein	
		Glyma13g33243	Gpi16 subunit	
		Glyma13g33260	Zn-finger protein	
Germplasm panel (224)	-	*Glyma15g41680*	LEM3 (ligand-effect modulator 3) family/CDC50-related	Huang et al. (2016)
		Glyma03g28660	ARF-related/ADP-ribosylation factor	
		Glyma16g30140	Predicted lipase class 3 gene	
		Glyma16g04700	Thioredoxin	
		Glyma20g39240	DEAD/DEAH box helicase	
		Glyma06g01080	2OG-Fe (II) oxygenase superfamily	
		Glyma16g14080	Serine/threonine protein kinase	
		Glyma11g11100	Phototropic-responsive NPH3 family protein	
		Glyma16g31930	Zinc finger domain	
		Glyma03g04960	Lipid transport protein	

(Continued on following page)

TABLE 4 (*Continued*) Putative candidate genes identified for *P. sojae* resistance.

Parents	Rps genes	Gene model number	Gene function	References
		Glyma04g40800	Serine/threonine protein kinase	
		Glyma09g04310	Ankyrin repeat and calmodulin-binding motif	
Germplasm panel (189)	-	*Glyma.03g034400*	LRR and NB-ARC domains containing disease resistance protein	Qin et al. (2017)
		Glyma.05g209300	Disease resistance protein (TIR-NBS class)	
		Glyma.05g213400	Disease resistance responsive (dirigent-like protein) family protein	
		Glyma.13g184800	LRR and NB-ARC domains containing disease resistance protein	
		Glyma.07g007800	Disease resistance protein RPS4-RELATED	
		Glyma.03g037000	LRR and NB-ARC domains containing disease resistance protein	
		Glyma.04g205200	Defense response	
		Glyma.13g028100	RPS4-related disease resistance protein	
		Glyma.03g149600	Resistance to Phytophthora 1	
		Glyma.10g127500	Disease resistance responsive (dirigent-like protein) family protein	
		Glyma.10g129400	Disease resistance family protein/LRR family protein	
		Glyma.10g184300	RPS4-related disease resistance protein	
		Glyma.10g196700	Disease resistance protein (CC-NBS-LRR class) family	
		Glyma.14g079500	Arabidopsis broad-spectrum mildew resistance protein RPW8	
		Glyma.14g079600	Arabidopsis broad-spectrum mildew resistance protein RPW8	
Jikedou 2 × Qichadou 1	*RpsQ*	*Glyma.03g27200*	Protein with a typical serine/threonine protein kinase structure	Li et al. (2017a)
Meng8206 × Linhedafenqing and Meng8206 Zhengyang	*RpsHN*	*Glyma.03g04260*	NB-ARC domain-containing disease resistance protein	Niu et al. (2017)
		Glyma.03g04300	NB-ARC domain-containing disease resistance protein	
		Glyma.03g04340	Serine/threonine protein kinase	
Huachun 2 ×Wayao	*RpsWY*	*Glyma03g04350*	Pentatricopeptide repeat-containing protein	Cheng et al. (2017)
		Glyma03g04360	Transposase/serine/threonine protein	
		Glyma03g04370	Non-specific lipid-transfer protein 3-like protein	
Germplasm panel (337)	-	*Glyma01g32800*	Serine/threonine protein kinase	Niu et al. (2018)
		Glyma01g32855	Leucine-rich repeat protein kinase family proteins	
Williams×Zaoshu18	*RpsZS18*	*Glyma.02g245700*	EF-hand calcium-binding domain	Zhong et al. (2018a)
		Glyma.02g245800	pfkB carbohydrate kinase	
Germplasm panel (169)		*Glyma03g03480*	Auxin-responsive family protein	Ludke et al. (2019)
		Glyma03g04990	Aalanine-glyoxylate aminotransferase/beta-Alanine-pyruvate aminotransferase	

(Continued on following page)

TABLE 4 (*Continued*) Putative candidate genes identified for *P. sojae* resistance.

Parents	Rps genes	Gene model number	Gene function	References
		Glyma03g05070	Short-chain dehydrogenase/reductase (SDR) family protein	
		Glyma15g20550	Pectinesterase family protein	
		Glyma15g21130	Expansin-like B3 precursor (EXLB3)	
Zhonghuang47 × Xiu94-11	*RpsX*	*Glyma.03g027200*	Leucine-rich repeat (LRR) region	Zhong et al. (2019)
Hefeng 25 × DongongL-28; Germplasm (225)	-	*Glyma.03G033700*	C2H2-like zinc finger protein	Zhao et al. (2020)
		Glyma.03G033800	Cell wall β-expansin protein	
Germplasm panel (376)	-	*Glyma05g146400*	Mannosyl oligosaccharide glucosidases	Van et al. (2020)
		Glym05g146500	Mannosyl oligosaccharide glucosidases	
		Glym.05g146600	ER metallopeptidase	
		Glyma05g146900	Heparan sulfate glycosyltransferase-related	
PI 449459 × Misty	-	*Glyma.13G190400*	NBS-LRR	de Ronee et al. (2019)
		Glyma.19G262700	AP2/ERF-type transcription factor	
Zaoshu18 × Yudou25	*RpsYD25*	*Glyma.03g034700*	Zinc ion binding- and nucleic acid-binding gene	Zhong et al. (2020)
		Glyma.03g034800	NBS-LRR	
		Glyma.03g034900	NBS-LRR	
Guizao1 × BRSMG68	*RpsGZ*	*Glyma.03G034400*	Disease resistance protein (NBS-LRR class), putative	Jiang et al. (2020)
		Glyma.03G034500	Disease resistance protein (NBS-LRR class), putative	
		Glyma.03G034800	Disease resistance protein (NBS-LRR class), putative	
		Glyma.03G034900	Disease resistance protein (NBS-LRR class), putative	
		Glyma.03G035000	Domain of unknown function DUF223	
		Glyma.03G035100	PIF1-like helicase	
		Glyma.03G035200	CW-type zinc finger; B3 DNA-binding domain	
		Glyma.03G035300	Disease resistance protein (NBS-LRR class)	
		Glyma.03G035400	PPR repeat	
		Glyma.03G035500	Plant mobile domain	
		Glyma.03G035600	Protease inhibitor/seed storage/LTP family	
		Glyma.03G035800	Pollen allergen; rare lipoprotein A (RlpA)-like double-psi beta-barrel	
		Glyma.03G035900	Membrane attack complex/perforin domain	
		Glyma.03G036000	Protein tyrosine kinase; serine–threonine protein kinase	
		Glyma.03G036200	Multidrug resistance protein	

complete reference genome sequence, genomic regions of different *Rps* regions were analyzed for the identification of candidate resistance genes. A list of putative candidate genes for *P. sojae* resistance is given in Table 4 along with their gene annotations. The maximum number of candidate genes reported are from chromosome 3 (Zhang et al., 2013a; Zhang et al., 2013b; Sun et al., 2014a; Li et al., 2017a; Cheng et al., 2017; Niu et al., 2017; Zhong et al., 2018a; Zhong et al., 2019; Jiang et al., 2020; Zhong et al., 2020). Some of these genes, *viz.*, zinc ion binding- and nucleic acid-binding genes, NB-ARC domain-containing disease resistance

proteins, and *NBS-LRR* genes, were functionally analyzed (Li et al., 2017a; Niu et al., 2017; Zhong et al., 2020) (Figure 1). Li et al. (2016a) reported multiple copies of *R-gene*-type annotation in *RpsUN1* and *UN2*. Li et al. (2016b) conducted GWAS in an association panel of 279 accessions and identified seven candidate genes on chromosome 13 that are reported to govern natural variations for partial resistance to *P. sojae*. Unlike Li et al. (2016a), *non-NBS-LRR* types of genes have also been proposed as candidates for another *Rps* allele on chromosome 3 (Cheng et al., 2017). Cheng et al. (2017) identified candidate genes against *P. sojae* using the high-throughput genome-wide sequencing approach by mapping 3,469 recombination bins in RILs. This study revealed the localization of *RpsWY* gene in bin 401 (on chromosome 3). Bin 401 was found to contain three genes, namely, *pentatricopeptide repeat-containing protein, transposase/serine/threonine protein*, and *non-specific lipid-transfer protein 3-like protein*. Sahoo et al. (2017) also reported several *NBS-LRR-like* genes in genetic investigations of *Rps12*. Jiang et al. (2020) and Zhong et al. (2020) also reported *NBS-LRR* and *zinc ion-binding* genes as candidate genes by fine mapping of *RpsYD25* and *RpsGZ*. Though reference genome sequencing can provide information on the majority of genes present in the identified genomic region, *de novo* sequencing of the haplotype carrying the target *Rps* gene is important to identify candidate genes.

Rps 11 showed resistance to 12 races of *P. sojae*; therefore, it is a broad-spectrum resistance gene (Wang et al., 2021). Wang et al. (2021) sequenced the genome of "PI 594527" by long-read sequencing, and the assembled genome sequence identified that the *Rps11* locus was present in a genomic region harboring a cluster of 12 *NLR* genes of a single origin in soybean. Fine mapping and gene expression analysis pinpointed a 27.7-kb *NLR* gene (Wang et al., 2021). Genetic transformation of an *Rps11*-coding DNA sequence in a susceptible soybean genotype conferred a resistant phenotype. Pan-genome analysis revealed that *Rps11* is a unique gene in "PI 594527" and does not have any other allelic copy in the other genotypes. The isolation of *Rps11* will help soybean breeders accelerate the improvement of broad-spectrum resistance to *P. sojae* in soybean. The unique structural features of *Rps11* make it a suitable model to investigate the resistance mechanism to further improve high-yielding cultivars.

Transcriptomic studies on PRSR

Recent developments in the genomics of *P. sojae* and soybean have made this pathosystem a model to understand molecular bases underpinning plant–oomycete interactions (Guo et al., 2011). Furthermore, transcriptomics of PRSR resistance in soybean has been extensively carried out to study the candidate genes and the role of biochemical pathways involved in conferring resistance. Through microarray analysis, genes governing pathogenesis-related proteins and enzymes involved in phytoalexin biosynthesis were found to be upregulated and reached a peak at 24 dpi. On the other hand, genes encoding lipoxygenases and peroxidases were found to be downregulated during the infection process (Moy et al., 2004).

To gain deep insights into the molecular basis of resistance to *P. sojae*, differential gene expression in response to *P. sojae* infection in the cultivar "Suinong 10" was studied by Xu et al. (2012). A total of eight transcripts were found to be upregulated in the treated plants as compared to those of the control. These transcripts are responsible for enzymes involved in the phytoalexin biosynthesis pathway and pathogenesis-related proteins and some defense response-related proteins such as phenylalanine ammonia-lyase, WRKY transcription factor 31, isoflavone reductase, pleiotropic drug resistance protein 12, and major allergen *Pru ar 1* (Xu et al., 2012). Molecular responses induced by different *Rps* genes and the association of phytohormone signaling pathways with disease reactions to *P. sojae* infection were studied by Lin et al. (2014). Transcriptome analysis on 10 near-isogenic lines (NILs) (*Rps1-a*, *1-b*, *1-c*, *1-k*, *Rps3-a*, *3-b*, *3-c*, *Rps4*, *5*, and *6*, each in the genetic background of "Williams") and the susceptible genotype, "Williams" during pre- and post-inoculation was carried out to identify differentially expressed genes (DEGs) across different treatments (Lin et al., 2014). A total of 5,806 incompatible interaction genes (IIGs) were identified by comparing DEGs between "Williams" and NILs, and 1,139 compatible interaction genes (CIGs) were identified in "Williams." Of these 5,806 IIGs, 23 were found to be common across 10 NILs and are mostly associated with biotic and abiotic stress responses, suggesting the overlap of molecular responses induced by different *Rps* genes. Two *NPR-1*-like IIGs, *Glyma02g45260* and *Glyma14g03510*, were involved in mediating the SA signaling pathway during an incompatible reaction, suggesting the role of the SA pathway in genetic resistance. Several *JAZ*-like proteins that repress the jasmonic acid (JA) pathway were found, such as IIGs and/or CIGs. These proteins were downregulated in NILs and were upregulated in "Williams." Also, a *JAR1* homolog, *Glyma07g06370* that activates the JA signaling pathway, was upregulated during the susceptible reaction in Williams. Genes that repress the ethylene (ET) pathway were found to be downregulated in NILs and upregulated in "Williams," suggesting that the ET pathway was repressed during the susceptible host reaction in "Williams" and activated in NILs during the incompatible reaction. In addition, three *BAK1* homolog IIGs that activate brassinosteroid (BR) signaling were found to be upregulated in NILs, suggesting the role of the BR signaling pathway during defense against *P. sojae*.

Role of transcription factors

Transcription factors (TFs) are master switches for regulating the expression of genes and controlling several signaling pathways (Chattopadhyay et al., 2019) and also play a vital role in different defense mechanisms in different plant species

against different phytopathogens. In soybean, several TFs have been identified for their role in regulating genes and pathways involved in resistance to *P. sojae*. A *bHLH* (basic helix–loop-helix) transcription factor associated with resistance to *P. sojae* was functionally characterized through its hypo- and hyper-expression in a resistant soybean genotype, "L77-1863," and was designated as *GmPIB1*. *GmPIB1* represses the expression of the *GmSPOD1* gene by directly binding to its promoter. Through RNAi assay, it was found that *GmSPOD1* is involved in the production of reactive oxygen species (ROS) during *P. sojae* infection. Hence, the role of *GmPIB1* TF in *P. sojae* resistance through reduced ROS production has been established (Cheng et al., 2018). Several *ethylene-responsive element binding factor* (*ERF*) transcription factors are linked with disease resistance in different plants (Gu et al., 2000; Song et al., 2005). An ERF-associated amphiphilic repression (EAR) motif-containing ERF TF, *GmERF5*, conferring resistance to *P. sojae* through the positive regulation of pathogenesis-related (PR) genes, *PR10*, *PR1-1*, and *PR10-1*, has been identified (Dong et al., 2015). A TF gene, *GmWRKY40*, was found to impart resistance in soybean to *P. sojae* and acts as a positive regulator of ROS accumulation and the JA signaling pathway (Cui et al., 2019). A transcription factor, *GmMYB29A2*, was found to impart resistance to *P. sojae* infection in soybean through the regulation of *glyceollin I* accumulation (Jahan et al., 2020). *WRKY transcription factor 31* identified in response to *P. sojae* infection (Xu et al., 2012) was functionally characterized through overexpression and RNAi silencing (Fan et al., 2017). Gene *GmWRKY31* interacts with *GmHDL56* and jointly engages in the activation of *GmNPR1*, which in turn manifests resistance during the Suinong 10–*P. sojae* interaction. Another TF, *GMERF113*, was isolated from "Suinong 10" and characterized for its response to *P. sojae* infection in a susceptible genotype "Dongnog 5." The overexpression of *GMERF113* in this genotype resulted in an enhanced resistance level and expression of pathogenesis-related genes, *PR1* and *PR10-1*. Thus, the role of *GMERF113* in the defense mechanism through positive regulation of these two pathogenesis-related genes has been well-demonstrated (Zhao et al., 2017).

Role of enzymes and proteins

Fan et al. (2015) studied the expression of class 10 protein *Gly m 4l* and found its role in the resistance to *P. sojae*. Zhang et al. (2017) identified a *phenylalanine ammonia-lyase* (*PAL*) gene family member, *GmPAL2.1*, to be linked with resistance to *P. sojae* through reverse genetics. The role of enzyme class 4-coumarate: CoA ligase (4CL) in plant defense against pathogens has been investigated extensively (Ehlting et al., 1999). A member of the 4CL (enzyme 4-coumarate: CoA ligase) gene family, *GmPI4L*, identified in soybean is associated with resistance to *P. sojae* infection through the enhanced production of glyceollins, genistein, and daidzein in soybean, laying the foundation for the enzymatic basis for resistance to this pathogen (Chen et al., 2019). The mediator complex is a part of RNA polymerase II, which acts as a regulatory element of the transcription process. A mediator subunit, *GmMED16-1* in soybean, through its silencing, was found to govern *P. sojae* by modulating the transcription of *NPR1*, *PR1a*, and *PR5* genes (Xue et al., 2019).

Role of miRNAs

MicroRNAs (miRNAs) are also known to be regulated under defense mechanisms in several plant species. Guo et al. (2011) revealed the role of miRNA in *P. sojae* resistance. Wong et al. (2014) identified miR393 and miR166, as triggered by heat-inactivated *P. sojae* hyphae, suggesting their roles in soybean basal defense. Knockdown of miR393 led to the increased susceptibility of soybean to *P. sojae*. The expression of isoflavonoid synthesis genes was drastically reduced in miR393 knockdown roots, suggesting that miR393 promotes soybean defense against *P. sojae*.

Molecular breeding for resistance to *P. sojae*

Soybean witnessed a significant improvement in yields over the past 60 years through conventional breeding approaches. Soybean yields were estimated to improve at the rate of 23 kg/ha/annum (Specht et al., 1999), and Wilcox (2001) reported an increase of 60% in seed yields over the past 60 years in the United States of America. The significant increase in yields has been witnessed mainly due to the toiling efforts of conventional breeding-based public sector soybean breeding programs. But considering the limitations of conventional breeding methods for *P. sojae* resistance improvement, further progress for yield enhancement is stagnated at the global level. The stagnated progress due to *P. sojae* infection can be further brought back to an accelerated track by the adoption of MAS and genomics-aided approaches in the PRSR resistance soybean breeding programs.

Molecular markers ranging from hybridization (RFLP and AFLP) and polymerase chain reaction-based markers (SSRs) to sequencing-based markers (SNPs) have been used to a greater extent for high-resolution mapping as well as for fine mapping of genomic regions governing the resistance to *P. sojae* (Table 2).

The identified major genomic regions for *P. sojae* resistance can be introgressed into elite soybean cultivars through the use of genomics-assisted breeding techniques, *viz.*, marker-assisted backcross breeding (MABB), marker-assisted recurrent selection (MARS), marker-assisted gene pyramiding (MAGP), and genomic selection (GS). The identified major QDRL can be

targeted for introgression into elite cultivars using the MABB approach (Ribaut and Ragot, 2007; Choudhary et al., 2019). Selection of *Rps* gene for introgression is also very important as it depends on particular regions of cultivation. Dorrance et al. (2016) estimated pathotype variability in 11 different states of the US with 873 isolates and concluded that *Rps 6* and *Rps 8* are more effective against the majority of isolates collected from northern regions. Several efforts have been made for the introgression of single-gene (*Rps*)-mediated resistance into soybean cultivars for controlling PRSR (Roth et al., 2020). Six of these genes (*Rps1a*, *Rps1b*, *Rps1c*, *Rps1k*, *Rps3a*, *Rps 6*, and *Rps3a*) already exist in commercial varieties and provide disease management against *Phytophthora* root and stem rot (Roth et al., 2020), which were transferred with the help of conventional approaches. In Japan, "Hyogo Prefecture," the black-seeded PRSR-resistant line, was used as the donor for introgression and for the development of resistant cultivars (Sugimoto et al., 2010). Although plant breeders use MAS-based approaches mainly for transferring *Rps* genes in soybean (Li et al., 2010; Ramalingam et al., 2020), due to high disease pressure, rapid evolution in the pathotypes of *P. sojae* has been witnessed over the past 3 decades, hence making vertical resistance ineffective. This forced the plant breeders to target partial resistance for the effective and sustainable management of PRSR (Schmitthenner, 1985). Studies on mapping QDRL dissected the genetic basis of partial resistance to *P. sojae* and revealed small-to-moderate effect QDRL, many of which individually explained less than 10% of phenotypic variance for PRSR in a population (Table 2). The difficulty of identifying small-effect QDRL in small mapping populations can be resolved by deploying joint linkage QDRL analysis of multiple populations (Lander and Kruglyak, 1995; Beavis 1998). Although relatively less, a good number of major QDRL have been mapped for PRSR partial resistance in soybean (Figure 2).

The utilization of MABB is restricted to the introgression of major QDRL only, which have more PVE (phenotypic variance explanation) percentages and limited localization in the genome, as it is very difficult to follow a large number of QDRL during introgression programs. Hence, other molecular breeding approaches such as MAGP, MARS, and genomic selection can serve as a good alternative for accumulating favorable QDRL (minor and major effects) for PRSR resistance. Pyramiding of PRSR-resistant QDRL was demonstrated by Li et al. (2010) by targeting seven consistent QDRL (detected across multiple environments) from two different donors ("Conrad" and "Hefeng 25"). Limited efforts of QDRL stacking for PRSR resistance revealed a significant increase in the tolerance level of introgressed lines, and the tolerance level against PRSR was found to be positively correlated with the number of QDRL stacked (Li et al., 2010). Recently, Karhoff et al. (2019) demonstrated the genetic gains from selections of a major QTL for partial resistance to *P. sojae*. The introgression of a resistance allele from the respective "PI 427105B" and "PI 427106" improved the genetic levels of resistance to *P. sojae* by ~20% and ~40%, respectively, and the yield by 13%–29% under diseased conditions (Karhoff et al., 2019). These are a few examples of PRSR resistance introgression through molecular breeding, demonstrating the fruitful results of genetic and genomic mapping for PRSR resistance. Dorrance et al. (2016) emphasized on stacking of *Rps* genes with strong partial resistance for limiting the loss caused by PRSR. With the new genomics-assisted breeding approaches, it will be practically more feasible and applicable in stacking of major genes for complete resistance and multiple QDRL of partial resistance for imparting sustainable PRSR resistance in soybean cultivars.

Genome editing for understanding PRSR resistance

Not only naturally available and induced mutations are the source for introducing new resistance genes in crop improvement programs, but also genetic engineering and gene editing (genome editing) are techniques that enable precise and targeted modifications. Now, gene-editing technologies are gaining momentum for crop improvement as they are more similar to the widely accepted "mutation breeding" technology.

CRISPR/Cas9 gene editing is particularly useful in deciphering the plant–pathogen interaction and understanding effector-triggered immunity. Pathogen avirulence (*Avr*) effectors interplay with corresponding plant resistance (*R*) proteins and activate robust immune responses in the host plant. *Avr4/6*, an RxLR effector gene of *P. sojae*, which is recognized by soybean *R*-genes (*Rps6* and *Rps4*), was edited using CRISPR/Cas9 technology to study its possible role in pathogenicity (Fang and Tyler 2016). This study validated the contribution of *Avr4/6* in pathogen recognition by soybean *R*-gene loci, *Rps4* and *Rps6*. Ochola et al. (2020) engineered the promoter region of *PsAvr3b* gene which is recognized by *Rps3b*, and mutants with low *PsAvr3b* expression successfully colonized soybean plants carrying the cognate *R*-gene *Rps3b*. Wang L et al. (2020) edited *PsSu(z)12* gene associated with effector locus *Avr1b*. *PsSu(z)12* is epigenetically governed and encodes a core subunit of the *H3K27me3 methyltransferase complex*. CRISPR/Cas9-mediated H3K27me3 depletion within the *Avr1b* genomic region was correlated with impaired *Avr1b* gene silencing, and the mutants lost their ability to evade immune recognition by soybeans carrying *Rps1b* (Wang P et al., 2020). Tan et al. (2020) studied knockout mutants of *P. sojae* generated *via* the CRISPR/Cas9 system for the *PsGH7a* (*GH7 family cellobiohydrolase*) gene, and the mutants were found to have reduced virulence on susceptible soybean as compared to wild-type strain "P6497." It is expected that in the future, the CRISPR/Cas9 system coupled with other genomic techniques will be an

important approach to create disease-resistant cultivars that can withstand biotic stresses (Kumar et al., 2020).

Challenges and future perspectives

The urgency and significance of *P. sojae*-resistant cultivar development can be realized from its vast spread and rapid occurrence of the disease across soybean-growing areas. This demands a strong emphasis on strengthening *P. sojae* resistance soybean breeding programs globally. Although significant progress has been made through the utilization of race-specific resistance genes (*Rps* genes), the rapid evolution of pathotypes in *P. sojae* resulted in resistance breakdown. This problem was quickly assessed by soybean breeders and, hence, shifted the focus to partial resistance (horizontal resistance) which provides relatively broad and highly durable resistance. Extensive genetic and genomics studies identified several major genes and QDRL for *P. sojae* resistance. The *Rps*-linked markers can be utilized in the selection of genotypes having PRSR resistance genes in early stages, and subsequent backcrossing will enable the rapid development of PRSR-resistant soybean cultivars. Marker-assisted breeding approaches such as MAGP can help in pyramiding vertical and horizontal resistance by the utilization of major resistance genes and QDRL identified in different genetic backgrounds. This strategy of combining complete and partial resistance in the same cultivars will prove to be the most effective approach in the near future. Soybean breeders need to continuously identify novel and unique resistance genes to cope with the emerging new pathotypes (Sugimoto et al., 2011). Though it is challenging to incorporate a large number of genes and QDRL from multiple genetic backgrounds into a single background using MABB, MARS and genomic selection can be used in resistance breeding programs to incorporate all PRSR resistance loci for durable resistance. It will be useful to mine the germplasm and geographical regions with enormous diversity for the presence of resistance to prevailing *P. sojae* pathotypes. For example, soybean germplasm collections in the Republic of Korea have greater variability for resistance to *P. sojae* for specific *Rps* loci, as well as partial resistance (Dorrance and Schmitthenner, 2000), and can be used for incorporating durable resistance through large-scale breeding programs. Emerging approaches such as gene discovery through re-sequencing, proteomics, metabolomics, RNA-seq, and exome sequencing of soybean and its wild relatives need to be exploited at a broader level. Furthermore, the QTL-seq approach will likely augment the rapid identification of novel QDRL and advancement of selected progenies for cultivar improvement (Zhang et al., 2018). The CRISPR/Cas9-mediated identification of effector-triggered immunity and *R*-gene editing is a highly targeted approach for the understanding and rapid development of PRSR resistance. Thus, different "Omics" approaches may be employed to explore the plant defense mechanisms in plant–pathogen interactions along with a gene-editing approach. In addition to the genetic improvement of cultivars for PRSR resistance, other alternative approaches need to be adopted and integrated to achieve prolonged resistance. Such approaches include the identification of effective compounds such as calcium that could help control the PRSR to certain levels (Sugimoto et al., 2010). Since the roots are primary targets for PRSR infection, the extensive comparative study of root traits in wild relatives or resistant cultivars to those of susceptible cultivars will help in the identification of certain target traits for phenotyping and resistance management. For such studies, phenotyping platforms that help in better visualization of root system architecture should be given high priority. The combined approach of genetic resistance, integrated disease management, and climate-smart agronomic practices can pave the path for the sustainable management of PRSR in soybean.

Author contributions

SC and MC initiated the idea and drafted the manuscript. PB, VN, GK, and JC performed the literature search and conducted data analysis. SW, HS, MR, and SG critically revised the manuscript.

Acknowledgments

The authors thank the Indian Council of Agricultural Research and the Department of Biotechnology, Govt. of India, for encouragement and support.

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.

References

Abeysekara, N. S., Matthiesen, R. L., Cianzio, S. R., Bhattacharyya, M. K., and Robertson, A. E. (2016). Novel sources of partial resistance against *Phytophthora sojae* in soybean PI 399036. *Crop Sci.* 56, 2322–2335. doi:10.2135/cropsci2015.09.0578

Abney, T., Melgar, J., Richards, T., Scott, D., Grogan, J., and Young, J. (1997). New races of *Phytophthora sojae* with Rps 1-d virulence. *Plant Dis.* 81, 653–655. doi:10.1094/PDIS.1997.81.6.653

Allen, T. W., Bradley, C. A., Sisson, A. J., Byamukama, E., Chilvers, M. I., Coker, C. M., et al. (2017). Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Prog.* 18, 19–27. doi:10.1094/PHP-RS-16-0066

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., and Singh, K. B. (2010). Plants versus pathogens: An evolutionary arms race. *Funct. Plant Biol.* 37 (6), 499–512. doi:10.1071/FP09304

Anderson, T. R., and Buzzell, R. I. (1992). Diversity and frequency of races of *Phytophthora megasperma* f. sp. *glycinea* in soybean fields in Essex County, Ontario, 1980-1989. *Plant Dis.* 76, 587–589. doi:10.1094/PD-76-0587

Athow, K. L., Laviolette, F. A., and Mueller, E. H. (1980). A new major gene for resistance to *Phytophthora megasperma* var *sojae* in soybean. *Phytopathology* 70, 977–980. doi:10.1094/phyto-70-977

Athow, K. L., and Laviolette, F. A. (1982). Rps 6, a major gene for resistance to *Phytophthora megasperma* f. sp. *glycinea* in soybean. *Phytopathology* 72, 1564–1567. doi:10.1094/phyto-72-1564

Bandara, A. Y., Weerasooriya, D. K., Bradley, C. A., Allen, T. W., and Esker, P. D. (2020). Dissecting the economic impact of soybean diseases in the United States over two decades. *PLOS one* 15 (4), e0231141. doi:10.1371/journal.pone.0231141

Beavis, W. D. (1998). "QTL analyses: Power, precision, and accuracy," in *Molecular dissection of complex traits*. Editor A. H. Paterson (Boca Raton, US App: CRC Press), 145–162.

Bernard, R. L., and Cremeens, C. R. (1981). An allele at the rps1 locus from the variety 'Kingwa. *Soybean Genet. Newsl.* 8, 14. Available at: https://lib.dr.iastate.edu/soybean.genetics/vol8/iss1/14.

Bernard, R. L., Smith, P. E., Kaufmann, M. J., and Schmitthenner, A. F. (1957). Inheritance of resistance to Phytophthora root and stem rot in the soybean 1. *Agron. J.* 49, 391. doi:10.2134/agronj1957.00021962004900070016x

Bhattacharyya, M. K., Narayanan, N. N., Gao, H. O., Santra, D. K., Salimath, S. S., Kasuga, T., et al. (2005). Identification of a large cluster of coiled coil-nucleotide binding site–leucine rich repeat-type genes from the *Rps1* region containing Phytophthora resistance genes in soybean. *Theor. Appl. Genet.* 111, 75–86. doi:10.1007/s00122-005-1993-9

Bolanos-Carriel, C., Batnini, A., McHale, L. K., and Dorrance, A. (2021). Identification of resistance loci towards *Phytophthora sojae* (*Rps*) in South Korean soybean plant introductions 407974B and 424487B. *Crop Sci* 62, 275–285. doi:10.1002/csc2.20596

Burnham, K. D., Dorrance, A. E., VanToai, T. T., and St Martin, S. K. (2003). Quantitative trait loci for partial resistance to *Phytophthora sojae* in soybean. *Crop Sci.* 43, 1610–1617. doi:10.2135/cropsci2003.1610

Buzzell, R. I., and Anderson, T. R. (1981). Another major gene for resistance to *Phytophthora megasperma var. sojae* in soybean. *Soybean Genet. Newsl.* 18, 30–33. Available at: http://lib.dr.iastate.edu/soybeangenetics/vol8/iss1/10.

Buzzell, R. I., and Anderson, T. R. (1992). Inheritance and race reaction of a new soybean *Rps1* allele. *Plant Dis.* 76, 600–601. doi:10.1094/PD-76-0600

Chang, H. X., Lipka, A. E., Domier, L. L., and Hartman, G. L. (2016). Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Phytopathology* 106 (10), 1139–1151. doi:10.1094/PHYTO-01-16-0042-FI

Chattopadhyay, A., Purohit, J., Tiwari, K. K., and Deshmukh, R. (2019). Targeting transcription factors for plant disease resistance: Shifting paradigm. *Curr. Sci.* 117 (10), 1598–1607. doi:10.18520/cs/v117/i10/1598-1607

Chen, Q., Weng, Q., Wang, Y., and Zheng, X. (2004). Identification and sequencing of ribosomal DNA-ITS of *Phytophthora sojae* in Fujian. *Acta Phytopathol. Sin.* 34, 112–116.

Chen, X., Fang, X., Zhang, Y., Wang, X., Zhang, C., Yan, X., et al. (2019). Overexpression of a soybean 4-coumaric acid: coenzyme A ligase (*GmPI4L*) enhances resistance to *Phytophthora sojae* in soybean. *Funct. Plant Biol.* 46, 304–313. doi:10.1071/FP18111

Cheng, Q., Dong, L., Gao, T., Liu, T., Li, N., Wang, L., et al.(2018). The bHLH transcription factor GmPIB1 facilitates resistance to *Phytophthora sojae* in *Glycine max*. *J. Exp. Bot.* 69 (10), 2527–2541. doi:10.1093/jxb/ery103

Cheng, Y., Ma, Q., Ren, H., Xia, Q., Song, E., Tan, Z., et al. (2017). Fine mapping of a Phytophthora-resistance gene RpsWY in soybean (*Glycine max* L.) by high-throughput genome-wide sequencing. *Theor. Appl. Genet.* 130, 1041–1051. doi:10.1007/s00122-017-2869-5

Choudhary, M., Wani, S. H., Kumar, P., Bagaria, P. K., Rakshit, S., Roorkiwal, M., et al. (2019). QTLian breeding for climate resilience in cereals: Progress and prospects. *Funct. Integr. Genomics* 19, 685–701. doi:10.1007/s10142-019-00684-1

Costamilan, L. M., Clebsch, C. C., Soares, R. M., Seixas, C. D. S., Godoy, C. V., and Dorrance, A. E. (2013). Pathogenic diversity of *Phytophthora sojae* pathotypes from Brazil. *Eur. J. Plant Pathol.* 135, 845–853. doi:10.1007/s10658-012-0128-9

Cui, X., Yan, Q., Gan, S., Xue, D., Wang, H., Xing, H., et al. (2019). GmWRKY40, a member of the WRKY transcription factor genes identified from *Glycine max* L, enhanced the resistance to. *BMC Plant Biol.* 19, 598. doi:10.1186/s12870-019-2132-0

de Ronne, M., Labbé, C., Lebreton, A., Sonah, H., Deshmukh, R., Jean, M., et al. (2019). Integrated QTL mapping, gene expression and nucleotide variation analyses to investigate complex quantitative traits: A case study with the soybean– *Phytophthora sojae* interaction. *Plant Biotechnol. J.* 18, 1492–1494. doi:10.1111/pbi.13301

Demirbas, A., Rector, B. G., Lohnes, D. G., Fioritto, R. J., Graef, G. L., Cregan, P. B., et al. (2001). Simple sequence repeat markers linked to the soybean Rps genes for Phytophthora resistance. *Crop Sci.* 41, 1220–1227. doi:10.2135/cropsci20014141220x

Dong, L., Cheng, Y., Wu, J., Cheng, Q., Li, W., Fan, S., et al. (2015). Overexpression of *GmERF5*, a new member of the soybean EAR motif-containing ERF transcription factor, enhances resistance to *Phytophthora sojae* in soybean. *J. Exp. Bot.* 66 (9), 2635–2647. doi:10.1093/jxb/erv078

Dorrance, A. E., Berry, S. A., Anderson, T. R., and Meharg, C. (2008). Isolation, storage, pathotype characterization, and evaluation of resistance for *Phytophthora sojae* in soybean. *Plant Health Prog.* 10, 1094–1101. doi:10.1094/PHP-2008-0118-01-DG

Dorrance, A. E., Jia, H., and Abney, T. S. (2004). Evaluation of soybean differentials for their interaction with *Phytophthora sojae*. *Plant Health Prog.* 5, 9. doi:10.1094/PHP-2004-0309-01-RS

Dorrance, A. E., McClure, S. A., and St Martin, S. K. (2003). Effect of partial resistance on Phytophthora stem rot incidence and yield of soybean in Ohio. *Plant Dis.* 87, 308–312. doi:10.1094/PDIS2003873308

Dorrance, A. E., and Schmitthenner, A. F. (2000). New sources of resistance to *Phytophthora sojae* in the soybean plant introductions. *Plant Dis.* 84, 1303–1308. doi:10.1094/PDIS200084121303

Dorrance, A., Kurle, J., Robertson, A., Bradley, C., Giesler, L., Wise, K., et al. (2016). Pathotype diversity of *Phytophthora sojae* in eleven states in the United States. *Plant Dis.* 100, 1429–1437. doi:10.1094/PDIS-08-15-0879-RE

Ehlting, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E., and Kombrink, E. (1999). Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionary classes in angiosperms. *Plant J.* 19, 9–20. doi:10.1046/j1365-313X199900491x

Fan, A., Wang, X., Fang, X., Wu, X., and Zhu, Z. (2009). Molecular identification of Phytophthora resistance gene in soybean cultivar yudou 25. *A. A. S.* 35, 1844–1850. doi:10.3724/SPJ1006200901844

Fan, S., Dong, L., Han, D., Zhang, F., Wu, J., Jiang, L., et al. (2017). *GmWRKY31* and *GmHDL56* enhances resistance to *Phytophthora sojae* by regulating defense-related gene expression in soybean. *Front. Plant Sci.* 8, 781. doi:10.3389/fpls.2017.00781

Fan, S., Jiang, L., Wu, J., Dong, L., Cheng, Q., Xu, P., et al. (2015). A novel pathogenesis related class 10 protein glym 4l, increases resistance upon *Phytophthora sojae* infection inSoybean (*Glycine max* [L] merr). *PLoS ONE* 10 (10), e0140364. doi:10.1371/journal.pone.0140364

Fang, Y., and Tyler, B. M. (2016). Efficient disruption and replacement of an effector gene in theoomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* 17, 127–139. doi:10.1111/mpp12318

FAOSTAT (2019). Available at: https://www.fao.org/faostat/en/#data/QCL.

Gao, H., Narayanan, N. N., Ellison, L., and Hattacharyya, M. K. (2005). Two classes of highly similar coiled coil-nucleotide binding-leucine rich repeat genes isolated from the *Rps1-k* locus encode *Phytophthora* resistance in soybean. *Mol. Plant. Microbe. Interact.* 18, 1035–1045. doi:10.1094/MPMI-18-1035

Gao, H. Y., and Bhattacharyya, M. K. (2008). The soybean-*Phytophthora* resistance locus *Rps1-k* encompasses coiled coil-nucleotide binding-leucine rich

repeat-like genes and repetitive sequences. *BMC Plant Biol.* 8, 29. doi:10.1186/1471-2229-8-29-2229-8-29

Gordon, S. G., St Martin, S. K., and Dorrance, A. E. (2004). *Mapping* Rps8, *a gene for resistance to Phytophthora root and stem rot in soybean*. Madison, WI: Crop Sci Soc America Annu Mtg, ASA-CSA-SS, 320.

Gordon, S. G., St Martin, S. K., and Dorrance, A. E. (2006). *Rps 8* maps to a resistance gene rich region on soybean molecular linkage group F. *Crop Sci.* 46, 168–173. doi:10.2135/cropsci200404-0024

Gordon, S., Kowitwanich, K., Pipatpongpinyo, W., St Martin, S., and Dorrance, A. (2007). Molecular marker analysis of soybean plant introductions with resistance to *Phytophthora sojae*. *Phytopathology* 97, 113–118. doi:10.1094/PHYTO-97-0113

Graham, M. A., Marek, L. F., and Shoemaker, R. C. (2002). Organization, expression and evolution of a disease resistance gene cluster in soybean. *Genetics* 162, 1961–1977. doi:10.1093/genetics/162.4.1961

Grau, C. R., Dorrance, A. E., Bond, J., and Russin, J. S. (2004). "Fungal diseases," in *Soybeans: Improvement, production, and uses*. Editors H. R. Boerma and J. E. Specht. 3rd edn. (American Soc Agron MA), 679–763.

Gu, Y. Q., Yang, C., Thara, V. K., Zhou, J., and Martin, G. B. (2000). *Pti4* is induced by ethylene and salicylic acid, and its product is phosphorylated by the *Pto kinase*. *Plant Cell* 12, 771–786. doi:10.1105/tpc.12.5.771

Guo, N., Ye, W. W., Wu, X. L., Shen, D. Y., Wang, Y. C., Xing, H., et al. (2011). Microarray profiling reveals microRNAs involving soybean resistance to Phytophthora sojae. *Genome* 54 (11), 954–958. doi:10.1139/g11-050

Han, Y., Teng, W., Yu, K., Poysa, V., Anderson, T., Qiu, L., et al. (2008). Mapping QTL tolerance to Phytophthora root rot in soybean using microsatellite and RAPD/SCAR derived markers. *Euphytica* 162, 231–239. doi:10.1007/s10681-007-9558-4

Hartman, G. L., and Hill, C. B. (2010). "Diseases of soybean and their management," in *The soybean: Botany, production, and uses*. Editor G. Singh (Cambridge, MA: CAB International), 276–299.

Hartwig, E. E., Keeling, B. L., and Edwards, C. J., Jr (1968). Inheritance of reaction to Phytophthora rot in the soybean. *Crop Sci.* 8, 634–636. doi:10.2135/cropsci1968.0011183x000800050042x

Huang, J., Guo, N., Li, Y., Sun, J., Hu, G., Zhang, H., et al. (2016). Phenotypic evaluation and genetic dissection of resistance to *Phytophthora sojae* in the Chinese soybean mini core collection. *BMC Genet.* 17, 85. doi:10.1186/s12863-016-0383-4

Jahan, M. A., Harris, B., Lowery, M., Infante, A. M., Percifield, R. J., and Kovinicha, N. (2020). Glyceollin transcription factor *GmMYB29A2* regulates soybean resistance to *Phytophthora sojae*. *Plant Physiol.* 183, 530–546. doi:10.1104/pp1901293Available at: wwwplantphysiolorg/cgi/doi/101104/pp1901293.

Jang, I. H., and Lee, S. (2020). A review and perspective on soybean (*Glycine max* L) breeding for the resistance to *Phytophthora sojae* in Korea. *Plant Breed. Biotechnol.* 8 (2), 114–130. doi:10.9787/pbb.2020.8.2.114

Jee, H., Kim, W., and Cho, W. (1998). Occurrence of Phytophthora root rot on soybean (*Glycine max*) and identification of the causal fungus RDA. *J. Crop Prot.* 40, 16–22.

Jiang, B., Cheng, Y., Cai, Z., Li, M., Jiang, Z., Ma, R., et al. (2020). Fine mapping of a Phytophthora-resistance locus *RpsGZ* in soybean using genotyping-by-sequencing. *BMC Genomics* 21 (1), 280. doi:10.1186/s12864-020-6668-z

Kang, I. J., Kang, S., Jang, I. H., Jang, Y. W., Shim, H. K., Heu, S., et al. (2019). Identification of new isolates of *Phytophthora sojae* and the reactions of Korean soybean cultivars following hypocotyl inoculation. *Plant Pathol. J.* 35, 698–704. doi:10.5423/PPJ.NT.09.2019.0249

Karhoff, S., Lee, S., Mian, R., Ralston, T. I., Niblack, T. L., Dorrance, A. E., et al. (2019). Phenotypic characterization of a major quantitative disease resistance locus for partial resistance to Phytophthora sojae. *Crop Sci.* 59, 968–980. doi:10.2135/cropsci2018.08.0514

Kasuga, T., Salimath, S. S., Shi, J., Gijzen, M., Buzzell, R. I., and Bhattacharyya, M. K. (1997). High resolution genetic and physical mapping of molecular markers linked to the Phytophthora resistance gene Rps1-k in soybean. *Mol. Plant. Microbe. Interact.* 10 (9), 1035–1044. doi:10.1094/mpmi.1997.10.9.1035

Kaufmann, M. J. (1957). *Root and stem rot of soybean caused by Phytophthora sojae n. sp*. Champaign: University of Illinois at Urbana-Champaign.

Kilen, T. C., Hartwig, E. E., and Keeling, B. L. (1974). Inheritance of a second major gene for resistance to Phytophthora rot in soybeans 1. *Crop Sci.* 14 (2), 260–262. doi:10.2135/cropsci1974.0011183x001400020027x

Kumar, K., Gambhir, G., Dass, A., Tripathi, A. K., Singh, A., Jha, A. K., et al. (2020). Genetically modified crops: Current status and future prospects. *Planta* 251 (4), 91–27. doi:10.1007/s00425-020-03372-8

Kumawat, G., Gupta, S., Ratnaparkhe, M. B., Maranna, S., and Satpute, G. K. (2016). QTLomics in Soybean: a way forward for translational genomics and breeding. *Frontiers in Plant Science*, 7–1852. doi:10.3389/fpls.2016.01852

Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11 (3), 241–247. doi:10.1038/ng1195-241

Lee, S., Mian, M. A., McHale, L. K., Wang, H., Wijeratne, A. J., Sneller, C. H., et al. (2013a). Novel quantitative trait loci for partial resistance to Phytophthora sojae in soybean PI 398841. *Theor. Appl. Genet.* 126 (4), 1121–1132. doi:10.1007/s00122-013-2040-x

Lee, S., Mian, M. A., Sneller, C. H., Wang, H., Dorrance, A. E., and McHale, L. K. (2014). Joint linkage QTL analyses for partial resistance to Phytophthora sojae in soybean using six nested inbred populations with heterogeneous conditions. *Theor. Appl. Genet.* 127, 429–444. doi:10.1007/s00122-013-2229-z

Lee, S., Rouf Mian, M. A., McHale, L. K., Sneller, C. H., and Dorrance, A. E. (2013b). Identification of quantitative trait loci conditioning partial resistance to Phytophthora sojae in soybean PI 407861A. *Crop Sci.* 53 (3), 1022–1031. doi:10.2135/cropsci2012.10.0578

Li, L., Guo, N., Niu, J., Wang, Z., Cui, X., Sun, J., et al. (2016b). Loci and candidate gene identification for resistance to Phytophthora sojae via association analysis in soybean [*Glycine max* (L.) Merr.]. *Mol. Genet. Genomics* 291 (3), 1095–1103. doi:10.1007/s00438-015-1164-x

Li, L., Lin, F., Wang, W., Ping, J., Fitzgerald, J. C., Zhao, M., et al. (2016a). Fine mapping and candidate gene analysis of two loci conferring resistance to Phytophthora sojae in soybean. *Theor. Appl. Genet.* 129 (12), 2379–2386. doi:10.1007/s00122-016-2777-0

Li, S., Hanlon, R., Wise, H., Pal, N., Brar, H., Liao, C., et al. (2021). Interaction of Phytophthora sojae effector Avr1b with E3 ubiquitin ligase GmPUB1 is required for recognition by soybeans carrying Phytophthora resistance rps1-b and rps1-k genes. *Front. Plant Sci.* 12, 725571. doi:10.3389/fpls.2021.725571

Li, X., Han, Y., Teng, W., Zhang, S., Yu, K., Poysa, V., et al. (2010). Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environments from soybean cultivars 'Conrad'and 'Hefeng 25. *Theor. Appl. Genet.* 121 (4), 651–658. doi:10.1007/s00122-010-1337-2

Li, Y., Sun, S., Zhong, C., Wang, X., Wu, X., and Zhu, Z. (2017a). Genetic mapping and development of co-segregating markers of RpsQ, which provides resistance to Phytophthora sojae in soybean. *Theor. Appl. Genet.* 130 (6), 1223–1233. doi:10.1007/s00122-017-2883-7

Li, Y., Sun, S., Zhong, C., and Zhu, Z. (2017b). Detached-petiole inoculation method to evaluate Phytophthora root rot resistance in soybean plants. *Crop Pasture Sci.* 68 (6), 555–560. doi:10.1071/cp17158

Lin, F., Li, W., McCoy, A. G., Gao, X., Collins, P. J., Zhang, N., et al. (2021). Molecular mapping of quantitative disease resistance loci for soybean partial resistance to Phytophthora sansomeana. *Theor. Appl. Genet.* 134 (7), 1977–1987. doi:10.1007/s00122-021-03799-x

Lin, F., Zhao, M., Baumann, D. D., Ping, J., Sun, L., Liu, Y., et al. (2014). Molecular response to the pathogen Phytophthora sojae among ten soybean near isogenic lines revealed by comparative transcriptomics. *BMC genomics* 15 (1), 18–13. doi:10.1186/1471-2164-15-18

Lin, F., Zhao, M., Ping, J., Johnson, A., Zhang, B., Abney, T. S., et al. (2013). Molecular mapping of two genes conferring resistance to Phytophthora sojae in a soybean landrace PI 567139B. *Theor. Appl. Genet.* 126 (8), 2177–2185. doi:10.1007/s00122-013-2127-4

Liu, F. P., Gao, Z. M., Wang, X. F., Yu, C., and Pan, Y. M. (2006). On identification and biological characteristics of the causal organism of Phytophthora root rot of soybean in Anhui. *J. Guangxi Agric. Biol. Sci.* 25, 48–51.

Ludke, W. H., Schuster, I., Silva, F. L. D., Montecelli, T. D. N., Soares, B. D. A., Oliveira, A. B. D., et al. (2019). SNP markers associated with soybean partial resistance to Phytophthora sojae. *Crop Breed. Appl. Technol.* 19, 31–39. doi:10.1590/1984-70332019v19n1a05

Mideros, S., Nita, M., and Dorrance, A. E. (2007). Characterization of components of partial resistance, *Rps2*, and root resistance to *Phytophthora sojae* in soybean. *Phytopathology* 97, 655–662. doi:10.1094/PHYTO-97-5-0655

Mofijur, M., Atabani, A. E., Masjuki, H. A., Kalam, M. A., and Masum, B. M. (2013). A study on the effects of promising edible and non-edible biodiesel feedstocks on engine performance and emissions production: A comparative evaluation. *Renew. Sustain. energy Rev.* 23, 391–404. doi:10.1016/j.rser.2013.03.009

Moriwaki, J. (2010). Aiming at the construction of the race distinction system of Japanese Phytophthora sojae. *Plant Prot. Jpn.* 64, 508–510.

Moy, P., Qutob, D., Chapman, B. P., Atkinson, I., and Gijzen, M. (2004). Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Mol. Plant. Microbe. Interact.* 17 (10), 1051–1062. doi:10.1094/MPMI.2004.17.10.1051

Mueller, E. H., Athow, K. L., and Laviolette, F. A. (1978). Inheritance of resistence to four physiologic races of Phytophthora, *Phytophthora megasperma Var. sojae Phytopathol* 68, 1318–1322.

Nguyen, V. T., Vuong, T. D., VanToai, T., Lee, J. D., Wu, X., Mian, M. R., et al. (2012). Mapping of quantitative trait loci associated with resistance to Phytophthora sojae and flooding tolerance in soybean. *Crop Sci.* 52 (6), 2481–2493. doi:10.2135/cropsci2011.09.0466

Niu, J., Guo, N., Sun, J., Li, L., Cao, Y., Li, S., et al. (2017). Fine mapping of a resistance gene RpsHN that controls Phytophthora sojae using recombinant inbred lines and secondary populations. *Front. Plant Sci.* 8, 538. doi:10.3389/fpls.2017.00538

Niu, J., Guo, N., Zhang, Z., Wang, Z., Huang, J., Zhao, J., et al. (2018). Genome-wide SNP-based association mapping of resistance to Phytophthora sojae in soybean (Glycine max (L.) Merr.). *Euphytica* 214 (10), 187–189. doi:10.1007/s10681-018-2262-8

Ochola, S., Huang, J., Ali, H., Shu, H., Shen, D., Qiu, M., et al. (2020). Editing of an effector gene promoter sequence impacts plant-Phytophthora interaction. *J. Integr. Plant Biol.* 62 (3), 378–392. doi:10.1111/jipb.12883

Ping, J., Fitzgerald, J. C., Zhang, C., Lin, F., Bai, Y., Wang, D., et al. (2016). Identification and molecular mapping of Rps11, a novel gene conferring resistance to Phytophthora sojae in soybean. *Theor. Appl. Genet.* 129 (2), 445–451. doi:10.1007/s00122-015-2638-2

Ploper, L. D., Athow, K. L., and Laviolette, F. A. (1985). A new allele at the*Rps*$_3$Locus for resistance to*Phytophthora megaspermaf.* sp.*glycinea*in soybean. *Phytopathology* 75 (6), 690–694. doi:10.1094/phyto-75-690

Qin, J., Song, Q., Shi, A., Li, S., Zhang, M., and Zhang, B. (2017). Genome-wide association mapping of resistance to *Phytophthora sojae* in a soybean [Glycine max (L.) Merr.] germplasm panel from maturity groups IV and V. *PLoS One* 12 (9), e0184613. doi:10.1371/journal.pone.0184613

Raghuvanshi, R. S., and Bisht, K. (2010). "Uses of soybean: Products and preparation," in *The soybean: Botany, production, and uses CABI international*. Editor G. Singh. (Wallingford, UK: CABI International, 404–426.

Ramalingam, J., Alagarasan, G., Savitha, P., Lydia, K., Pothiraj, G., Vijayakumar, E., et al. (2020). Improved host-plant resistance to Phytophthora rot and powdery mildew in soybean (*Glycine max* (L.) Merr.). *Sci. Rep.* 10 (1), 13928–14011. doi:10.1038/s41598-020-70702-x

Ribaut, J. M., and Ragot, M. (2007). Marker-assisted selection to improve drought adaptation in maize: The backcross approach, perspectives, limitations, and alternatives. *J. Exp. Bot.* 58 (2), 351–360. doi:10.1093/jxb/erl214

Rolling, W., Lake, R., Dorrance, A. E., and McHale, L. K. (2020). Genome-wide association analyses of quantitative disease resistance in diverse sets of soybean [Glycine max (L.) Merr.] plant introductions. *PLoS ONE.* 15 (3). doi:10.1371/journal.pone.0227710

Roth, M. G., Webster, R. W., Mueller, D. S., Chilvers, M. I., Faske, T. R., Mathew, F. M., et al. (2020). Integrated management of important soybean pathogens of the United States in changing climate. *J. Integr. Pest Manag.* 11 (1), 17. doi:10.1093/jipm/pmaa013

Ryley, M., Obst, N., Irwin, J., and Drenth, A. (1998). Changes in the racial composition of Phytophthora sojae in Australia between1979 and 1996. *Plant Dis.* 82, 1048–1054. doi:10.1094/PDIS.1998.82.9.1048

Sahoo, D. K., Abeysekara, N. S., Cianzio, S. R., Robertson, A. E., and Bhattacharyya, M. K. (2017). A novel Phytophthora sojae resistance Rps12 gene mapped to a genomic region that contains several Rps genes. *PloS one* 12 (1), e0169950. doi:10.1371/journal.pone.0169950

Sahoo, D. K., Das, A., Huang, X., Cianzio, S., and Bhattacharyya, M. K. (2021). Tightly linked Rps12 and Rps13 genes provide broad-spectrum Phytophthora resistance in soybean. *Sci. Rep.* 11 (1), 16907–16913. doi:10.1038/s41598-021-96425-1

Sandhu, D., Gao, H., Cianzio, S., and Bhattacharyya, M. K. (2004). Deletion of a disease resistance nucleotide-binding-site leucine-rich-repeat-like sequence is associated with the loss of the Phytophthora resistance gene Rps4 in soybean. *Genetics* 168 (4), 2157–2167. doi:10.1534/genetics.104.032037

Sandhu, D., Schallock, K. G., Rivera-Velez, N., Lundeen, P., Cianzio, S., and Bhattacharyya, M. K. (2005). Soybean Phytophthora resistance gene Rps8 maps closely to the Rps3 region. *J. Hered.* 96 (5), 536–541. doi:10.1093/jhered/esi081

Sandhu, D., Tasma, I., Frasch, R., and Bhattacharyya, M. K. (2009). Systemic acquired resistance in soybean is regulated by two proteins, orthologous to Arabidopsis NPR1. *BMC Plant Biol.* 9 (1), 105–114. doi:10.1186/1471-2229-9-105

Schmitthenner, A. F. (1999). "Phytophthora rot of soybean," in *Compendium of soybean diseases*. Editors G. L. Hartman, J. B. Sinclair, and J. C. Rupe. 4th edn (USA: The American Phytopathological Society (APS) Press), 39–42.

Schmitthenner, A. F. (1985). Problems and progress in control of Phytophthora root rot of soybean. *Plant Dis.* 69, 362–368doi. doi:10.1094/pd-69-362

Schneider, R., Rolling, W., Song, Q., Cregan, P., Dorrance, A. E., and McHale, L. K. (2016). Genome-wide association mapping of partial resistance to Phytophthora sojae in soybean plant introductions from the Republic of Korea. *BMC genomics* 17 (1), 607–614. doi:10.1186/s12864-016-2918-5

Scott, K., Balk, C., Veney, D., McHale, L. K., and Dorrance, A. E. (2019). Quantitative disease resistance loci towards Phytophthora sojae and three species of Pythium in six soybean nested association mapping populations. *Crop Sci.* 59 (2), 605–623. doi:10.2135/cropsci2018.09.0573

Song, C. P., Agarwal, M., Ohta, M., Guo, Y., Halfter, U., Wang, P., et al. (2005). Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *Plant Cell* 17 (8), 2384–2396. doi:10.1105/tpc.105.033043

Song, Q., Jia, G., Zhu, Y., Grant, D., Nelson, R. T., Hwang, E. Y., et al. (2010). Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1. 0) in soybean. *Crop Sci.* 50 (5), 1950–1960. doi:10.2135/cropsci2009.10.0607

Specht, J. E., Hume, D. J., and Kumudini, S. V. (1999). Soybean yield potential—A genetic and physiological perspective. *Crop Sci.* 39 (6), 1560–1570. doi:10.2135/cropsci1999.3961560x

Stasko, A. K., Wickramasinghe, D., Nauth, B. J., Acharya, B., Ellis, M. L., Taylor, C. G., et al. (2016). High-density mapping of resistance QTL toward Phytophthora sojae, Pythium irregulare, and Fusarium graminearum in the same soybean population. *Crop Sci.* 56 (5), 2476–2492. doi:10.2135/cropsci2015.12.0749

Su, Y. C., and Shen, C. Y. (1993). The discovery and biological characteristic studies of Phytophthora megasperma f. sp. glycinea on soyabean in China. *Acta Phytopathol. Sin.* 23 (4), 341–347.

Sugimoto, T., Kato, M., Yoshida, S., Matsumoto, I., Kobayashi, T., Kaga, A., et al. (2012). Pathogenic diversity of *Phytophthora sojae* and breeding strategies to develop Phytophthora-resistant soybeans. *Breed. Sci.* 61, 511–522. doi:10.1270/jsbbs61511

Sugimoto, T., Watanabe, K., Yoshida, S., Aino, M., Furiki, M., Shiono, M., et al. (2010). Field application of calcium to reduce Phytophthora stem rot of soybean, and calcium distribution in plants. *Plant Dis.* 94, 812–819. doi:10.1094/PDIS-94-7-0812

Sugimoto, T., Yoshida, S., Aino, M., Watanabe, K., Shiwaku, K., and Sugimoto, M. (2006). Race distribution of *Phytophthora sojae* on soybean in Hyogo. *J. Gen. Plant Pathol.* 72, 92–97. doi:10.1007/s10327-005-0255-7

Sugimoto, T., Yoshida, S., Kaga, A., Hajika, M., Watanabe, K., Aino, M., et al. (2011). Genetic analysis and identification of DNA markers linked to a novel Phytophthora sojae resistance gene in the Japanese soybean cultivar. Waseshiroge. *Euphytica* 182 (1), 133–145. doi:10.1007/s10681-011-0525-8

Sugimoto, T., Yoshida, S., Watanabe, K., Aino, M., Kanto, T., Maekawa, K., et al. (2008). Identification of SSR markers linked to the Phytophthora resistance gene Rps1-d in soybean. *Plant Breed.* 127 (2), 154–159. doi:10.1111/j.1439-0523.2007.01440.x

Sun, J., Guo, N., Lei, J., Li, L., Hu, G., and Xing, H. (2014b). Association mapping for partial resistance to *Phytophthora sojae* in soybean (*Glycine max* (L) Merr). *J. Genet.* 93, 355–363. doi:10.1007/s12041-014-0383-y

Sun, J., Li, L., Zhao, J., Huang, J., Yan, Q., Xing, H., et al. (2014a). Genetic analysis and fine mapping of RpsJS, a novel resistance gene to Phytophthora sojae in soybean [*Glycine max* (L.) Merr.]. *Theor. Appl. Genet.* 127 (4), 913–919. doi:10.1007/s00122-014-2266-2

Sun, S., Wu, X. L., Zhao, J. M., Wang, Y. C., Tang, Q. H., Yu, D. Y., et al. (2011). Characterization and mapping of RpsYu25, a novel resistance gene to *Phytophthora sojae*. *Plant Breed.* 130 (2), 139–143. doi:10.1111/j1439-0523201001794x

Tan, X., Hu, Y., Jia, Y., Hou, X., Xu, Q., Han, C., et al. (2020). A conserved glycoside hydrolase family 7 cellobiohydrolase PsGH7a of Phytophthora sojae is required for full virulence on soybean. *Front. Microbiol.* 11, 1285. doi:10.3389/fmicb.2020.01285

Tooley, P. W., and Grau, C. R. (1984). Field characterization of rate-reducing resistance to Phytophthora megasperma f. sp. glycinea in soybean. *Phytopathology* 74 (10), 1201–1208. doi:10.1094/phyto-74-1201

Tsuchiya, S. (1990). Classification of race of Phytophthora sojae and screening of resistant soybean cultivars to Phytophthora stem rot. *Ann. Phytopathol. Soc. Jpn.* 56, 144.

Tucker, D. M., Saghai Maroof, M. A., Mideros, S., Skoneczka, J. A., Nabati, D. A., Buss, G. R., et al. (2010). Mapping quantitative trait loci for partial resistance to Phytophthora sojae in a soybean interspecific cross. *Crop Sci.* 50 (2), 628–635. doi:10.2135/cropsci2009.03.0161

Tyler, B. M., and Gijzen, M. J. (2014). "The Phytophthora sojae genome sequence foundation for a revolution," in *Genomics of plant-associated fungi and oomycetes: Dicot pathogens*. Editors R. A. Dean, A. Lichens-Park, and C. Kole (Heidelberg: Springer), 133–157. doi:10.1007/978-3-662-44056-8_7

Tyler, B. M. (2007). Phytophthora sojae: Root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* 8 (1), 1–8. doi:10.1111/j.1364-3703.2006.00373.x

USDA (2020). *Oilseeds - world markets and trade, a USDA publication*. SW Washington DC: USDA. Available at: https://wwwfasusdagov/data/oilseeds-world-markets-and-trade.

Van, K., Rolling, W., Biyashev, R. M., Matthiesen, R. L., Abeysekara, N. S., Robertson, A. E., et al. (2020). *Mining germplasm panels and phenotypic datasets to identify loci for resistance to Phytophthora sojae in soybean*. doi:10.1002/tpg2.20063

Wang, H., Waller, L., Tripathy, S., St Martin, S. K., Zhou, L., Krampis, K., et al. (2010). Analysis of genes underlying soybean quantitative trait loci conferring partial resistance to Phytophthora sojae. *Plant Genome* 3 (1), 23–40. doi:10.3835/plantgenome2009.12.0029

Wang, H., Wijeratne, A., Wijeratne, S., Lee, S., Taylor, C. G., St Martin, S. K., et al. (2012). Dissection of two soybean QTL conferring partial resistance to Phytophthora sojae through sequence and gene expression analysis. *BMC genomics* 13 (1), 428–523. doi:10.1186/1471-2164-13-428

Wang L, L., Chen, H., Li, J., Shu, H., Zhang, X., Wang, Y., et al. (2020). Effector gene silencing mediated by histone methylation underpins host adaptation in an oomycete plant pathogen. *Nucleic Acids Res.* 48 (4), 1790–1799. doi:10.1093/nar/gkz1160

Wang P, P., Wu, H., Zhao, G., He, Y., Kong, W., Zhang, J., et al. (2020). Transcriptome analysis clarified genes involved in resistance to Phytophthora capsici in melon. *PloS one* 15 (2), e0227284. doi:10.1371/journalpone0227284

Wang, W., Chen, L., Fengler, K., Bolar, J., Llaca, V., Wang, X., et al. (2021). A giant NLR gene confers broad-spectrum resistance to *Phytophthora sojae* in soybean. *Nat. Commun.* 12, 6263. doi:10.1038/s41467-021-26554-8

Wen, J., and Chen, H. (2002). Races and virulence phenotypes of Phytophthora sojae in Heilongjiang province and Mongolia Region of China. *Chinese Journal of Oil Crop Sciences* 24, 63–66.

Weng, C., Yu, K., Anderson, T. R., and Poysa, V. (2007). A quantitative trait locus influencing tolerance to Phytophthora root rot in the soybean cultivar 'Conrad. *Euphytica* 158 (1), 81–86. doi:10.1007/s10681-007-9428-0

Weng, C., Yu, K., Anderson, T. R., and Poysa, V. (2001). Mapping genes conferring resistance to Phytophthora root rot of soybean, Rps1a and Rps7. *J. Hered.* 92 (5), 442–446. doi:10.1093/jhered/925442

Wilcox, J. R. (2001). Sixty years of improvement in publicly developed elite soybean lines. *Crop Sci.* 41 (6), 1711–1716. doi:10.2135/cropsci20011711

Wong, J., Gao, L., Yang, Y., Zhai, J., Arikit, S., Yu, Y., et al. (2014). Roles of small RNA s in soybean defense against *Phytophthora sojae* infection. *Plant J.* 79 (6), 928–940. doi:10.1111/tpj.12590

Workneh, F., Yang, X. B., and Tylka, G. L. (1998). Effect of tillage practices on vertical distribution of Phytophthora sojae. *Plant Dis.* 82 (11), 1258–1263. doi:10.1094/PDIS.1998.82.11.1258

Wu, X., Zhang, B., Shi, S., Zhao, J., Feng, Y., Na, G., et al. (2011a). Identification, genetic analysis and mapping of resistance to Phytophthora sojae of Pm28 in soybean. *Agr. Sci. China* 10, 1506–1511. doi:10.1016/S1671-2927(11)60145-4

Wu, X., Zhou, B., Sun, S., Zhao, J., Chen, S., Gai, J., et al. (2011b). Genetic analysis and mapping of resistance to Phytophthora sojae of Pm14 in soybean. *Scientia Agricultura Sinica* 44, 456–460.

Wu, X., Zhou, B., Zhao, J., Guo, N., Zhang, B., Yang, F., et al. (2011c). Identification of quantitative trait loci for partial resistance to Phytophthora sojae in soybean. *Plant Breed.* 130, 144–149. doi:10.1111/j.1439-0523.2010.01799.x

Xiao, J. T., Miao, M. I. A. O., Gao, K., Xiang, G. L., Yang, S., and Dong, S. M. (2011). Genetic diversity of Phytophthora sojae in China based on RFLP. *Sci. Agric. Sin.* 44 (20), 4190–4198. doi:10.3864/j.issn.0578-1752.2011.20.007

Xu, P. F., Wu, J. J., Xue, A., Li, W. B., Chen, W. Y., Lai, W. E. I., et al. (2012). Differentially expressed genes of soybean during infection by *Phytophthora sojae*. *J. Integr. Agric.* 11 (3), 368–377. doi:10.1016/s2095-3119(12)60021-5

Xue, D., Guo, N., Zhang, X. L., Zhao, J. M., Bu, Y. P., Jiang, D. L., et al. (2019). Genome-wide analysis reveals the role of mediator complex in the soybean—phytophthora sojae interaction. *Int. J. Mol. Sci.* 20 (18), 4570. doi:10.3390/ijms20184570

Yang, J., Wang, X., Guo, B., Huang, J., Ye, W., Dong, S., et al. (2019). Polymorphism in natural alleles of the avirulence gene Avr1c is associated with the host adaptation of Phytophthora sojae. *Phytopathol. Res.* 1 (1), 28–13. doi:10.1186/s42483-019-0035-5

Yao, H. Y., Wang, X. M., Wu, X. F., Xiao, Y. N., and Zhu, Z. D. (2010). Molecular mapping of Phytophthora resistance gene in soybean cultivar Zaoshu18. *J. Plant Genet. Resour.* 11 (2), 213–217.

Yu, A., Xu, P., Wang, J., Zhang, S., Wu, J., Li, W., and Jiang, L. (2010). Genetic analysis and SSR mapping of gene resistance to Phytophthora sojae race 1 in soybean cv Suinong 10. *Chin. J. Oil Crop Sci.* 32 (4), 462–466.

Zhang, C., Wang, X., Zhang, F., Dong, L., Wu, J., Cheng, Q., et al. (2017). Phenylalanine ammonia-lyase2. 1 contributes to the soybean response towards Phytophthora sojae infection. *Sci. Rep.* 7 (1), 7242–7313. doi:10.1038/s41598-017-07832-2

Zhang, J., Xia, C., Duan, C., Sun, S., Wang, X., Wu, X., et al. (2013a). Identification and candidate gene analysis of a novel Phytophthora resistance gene Rps10 in a Chinese soybean cultivar. *PloS one* 8 (7), e69799. doi:10.1371/journal.pone.0069799

Zhang, J., Xia, C., Wang, X., Duan, C., Sun, S., Wu, X., et al. (2013b). Genetic characterization and fine mapping of the novel Phytophthora resistance gene in a Chinese soybean cultivar. *Theor. Appl. Genet.* 126 (6), 1555–1561. doi:10.1007/s00122-013-2073-1

Zhang, S., Xu, P., Wu, J., Xue, A. G., Zhang, J., Li, W., et al. (2010). Races of Phytophthora sojae and their virulences on soybean cultivars in Heilongjiang, China. *Plant Dis.* 94 (1), 87–91. doi:10.1094/PDIS-94-1-0087

Zhang, X., Wang, W., Guo, N., Zhang, Y., Bu, Y., Zhao, J., et al. (2018). Combining QTL-seq and linkage mapping to fine map a wild soybean allele characteristic of greater plant height. *BMC genomics* 19 (1), 226–312. doi:10.1186/s12864-018-4582-4

Zhao, X., Bao, D., Wang, W., Zhang, C., Jing, Y., Jiang, H., et al. (2020). Loci and candidate gene identification for soybean resistance to Phytophthora root rot race 1 in combination with association and linkage mapping. *Mol. Breed.* 40 (10), 100–112. doi:10.1007/s11032-020-01179-9

Zhao, Y., Chang, X., Qi, D., Dong, L., Wang, G., Fan, S., et al. (2017). A novel soybean ERF transcription factor, GmERF113, increases resistance to Phytophthora sojae infection in soybean. *Front. Plant Sci.* 8, 299. doi:10.3389/fpls.2017.00299

Zhendong, Z., Yunlong, H., Xiaoming, W., Junbin, H., and Xiaofei, W. (2010). Molecular identification of a novel Phytophthora resistance gene in soybean. *Eur. PMC* 33, 154–157.

Zhong, C., Li, Y., Sun, S., Duan, C., and Zhu, Z. (2019). Genetic mapping and molecular characterization of a broad-spectrum Phytophthora sojae resistance gene in Chinese soybean. *Int. J. Mol. Sci.* 20 (8), 1809. doi:10.3390/ijms20081809

Zhong, C., Sun, S., Li, Y., Duan, C., and Zhu, Z. (2018b). Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel Phytophthora resistance gene, RpsHC18, in soybean. *Theor. Appl. Genet.* 131 (3), 525–538. doi:10.1007/s00122-017-3016-z

Zhong, C., Sun, S., Yao, L., Ding, J., Duan, C., and Zhu, Z. (2018a). Fine mapping and identification of a novel Phytophthora root rot resistance locus RpsZS18 on chromosome 2 in soybean. *Front. Plant Sci.* 9, 44. doi:10.3389/fpls.2018.00044

Zhong, C., Sun, S., Zhang, X., Duan, C., and Zhu, Z. (2020). Fine mapping, candidate gene identification and co-segregating marker development for the Phytophthora root rot resistance gene RpsYD25. *Front. Genet.* 11, 799. doi:10.3389/fgene.2020.00799

Zhou, L., Deng, S., Xuan, H., Fan, X., Sun, R., Zhao, J., et al. (2022). A novel TIR-NBS-LRR gene regulates immune response to Phytophthora root rot in soybean. *Crop J.* doi:10.1016/j.cj.2022.03.003

frontiers | Frontiers in Genetics

TYPE Data Report
PUBLISHED 10 January 2023
DOI 10.3389/fgene.2022.1048476

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Muhammad Shahid Iqbal,
Institute of Cotton Research (CAS), China
Silvia Pampana,
University of Pisa, Italy

*CORRESPONDENCE
Kishor Gaikwad,
Kish2012@gmail.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 19 September 2022
ACCEPTED 22 December 2022
PUBLISHED 10 January 2023

CITATION
Tyagi A, Sharma S, Srivastava H, Singh A, Kaila T, Ali S, Gaikwad AB, Singh NK and Gaikwad K (2023), Transcriptome profiling of two contrasting pigeon pea (*Cajanus cajan*) genotypes in response to waterlogging stress.
Front. Genet. 13:1048476.
doi: 10.3389/fgene.2022.1048476

COPYRIGHT
© 2023 Tyagi, Sharma, Srivastava, Singh, Kaila, Ali, Gaikwad, Singh and Gaikwad. 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.

Transcriptome profiling of two contrasting pigeon pea (*Cajanus cajan*) genotypes in response to waterlogging stress

Anshika Tyagi[1,2†], Sandhya Sharma[1†], Harsha Srivastava[1], Anuradha Singh[1], Tanvi Kaila[1], Sajad Ali[2†], Ambika B. Gaikwad[3], N. K Singh[1] and Kishor Gaikwad[1*]

[1]National Institute for Plant Biotechnology, New Delhi, India, [2]Department of Biotechnology Yeungnam University, Gyeongsan, South Korea, [3]ICAR-National Bureau of Plant Genetic Resources, New Delhi, India

KEYWORDS

pigeon pea, waterlogging stress, RNAseq, BUSCO, functional annotation, signaling

Introduction

Pigeon pea (*Cajanus cajan* L.) is one of the important legume crop that contributes significantly to the nutritional stability and economy of billions of people in most developing nations (Sharma et al., 2020). Like other legumes, pigeon pea often offers more balanced and nutrient-dense calories and proteins (20%–22%) than cereals, making them essential in terms of food security. It is the sixth most significant legume food crop in the world, with a cultivation area of about 5 million hectares (ha) (Varshney et al., 2012). However, environmental stressors pose a persistent threat to the pigeon pea crop's production, yield, and quality. Among them waterlogging is one of the most harmful stress in pigeon pea which results in huge yield and economic losses around the world (Sultana et al., 2013; Tyagi et al., 2017; Tyagi et al., 2022). Overall, waterlogging has been observed to cause an annual loss of about .28–1.1 million tons per hectare which reduces production by 25%–30% in pigeon pea (Sultana, 2010; ICRISAT, 2011; Bansal and Srivastava, 2012). Waterlogging affects pigeon pea at all growth stages, but is more severe during the seedling and vegetative phases, thereby showing wilting, senescence and chlorosis (Bansal and Srivastava, 2012). Additionally, waterlogging also makes pigeon pea plants more susceptible to fungal diseases like Fusarium wilt and Phytophthora blight which results in significant yield losses (Yohan et al., 2017). Generally, pigeon pea is grown in low-input, risk-prone marginal environments and low-lying places that are more susceptible to waterlogging (Varshney et al., 2012; Duhan and Sheokand 2020). During waterlogging, the inhibition of aerobic respiration hinders growth and a variety of developmental processes, including seed germination, vegetative growth, and subsequent reproductive growth (Pan et al., 2021). Additionally, in waterlogged soils, ethylene and carbon dioxide levels increase dramatically in the root area which in turn alters the functions of soil microbiome that leads to an intense de-nitrification and accumulation of ammonium and polyphenolic compounds (Arduini et al., 2019). Also, it restricts the availability of nutrients like nitrogen (N) and sulphur (S) or changes them into a form that plants cannot absorb. It also changes ion homeostasis zinc (Zn), phosphorous (P), manganese (Mn), and iron (Fe), which can reach lethal levels to plants (Arduini et al., 2019). The most common effect of waterlogging stress is oxygen deficiency (hypoxia) and ethylene accumulation in plants, which can restrict root growth and root permeability, both of which lead to cell death (Sasidharan et al., 2018; Pan et al., 2021). However, plants use their multifaceted defense system in response to waterlogging stress by regulating their morphological, biochemical and molecular traits. For example, the formation of aerenchyma in roots is one of the main traits in plants that confer waterlogging tolerance (Luan et al., 2018). At the physiological and biochemical levels,

plants produce numerous molecules such as osmolytes, calcium (Ca^{2+}), reactive oxygen species (ROS), hormones, antioxidants that confers waterlogging tolerance (Pan et al., 2021). However, the molecular traits that confers waterlogging tolerance is least understood with many knowledge gaps. For example, how plants perceive waterlogging stress and triggers signal transduction pathways that in turn leads the expression of stress responsive genes. Additionally, the role of different sensors or receptors, ion channels, nicotinamide adenine dinucleotide phosphate (NADPH) etc that are involved in waterlogging signal transduction warrants future investigation.

Unlike other crops, pigeon pea genetic advancement has been hampered by scarce genomic resources and a lack of genetic variety in the basic gene pool which pose a significant obstacle to its improvement in terms of stress resistance and yield (Bohra et al., 2010; Varshney et al., 2012). Although there has been significant advancement in understanding the complexity of waterlogging signaling dynamics in model and cereal plants (Eysholdt-Derzsó et al., 2017; Yamauchi et al., 2018), but there is limited information in the most of the legume crops particularly in pigeon pea. For instance, the defense signaling pathways, hormonal crosstalk, regulatory genes, and transcriptional factors involved in waterlogging tolerance in pigeon pea cultivars remains enigmatic despite the availability of high throughput tools. Previous studies have identified many sensitive and tolerant pigeon pea genotypes which were mainly based on the morphological, physiological, biochemical traits and days of survival (Sultana et al., 2013). Although different pigeon pea waterlogging genotypes were found by these investigations, it is still largely unknown how these genotypes control waterlogging tolerance at the molecular level. Therefore, it is necessary to decode the waterlogging tolerance in pigeon pea cultivars and identify potential target genes for developing future climatic smart resilient pigeon pea genotypes in order to maintain productive agriculture and ensuring food security. In this work we first studied the effect of waterlogging stress in two contrasting pigeon pea genotypes viz., JBP-110B (tolerant) and ICP 7035 (sensitive) as well as their transcriptional profiling using *De-novo* transcriptome assembly (unpublished data). This comprehensive transcriptomic study data has led the important findings on the differentially expressed genes regulating waterlogging signaling mechanism in susceptible and tolerant pigeon pea genotypes which can be applied to subsequent research on the improvement of waterlogging resilience in pigeon pea and other legume crops.

Value of the data

- Pigeon pea is a rich source of protein for poor vegetarian people widely grown in Indian, Africa and Southeast Asia subcontinent.
- Waterlogging is the most detrimental abiotic stress in pigeon pea. Despite the availability of high through put tools, genomic resource for waterlogging tolerance trait in pigeon pea remains unknown.
- In this study, we have generated a comprehensive global gene expression profiling dataset for two contrasting pigeon pea genotypes, JBP-110B (tolerant) and ICP 7035 (sensitive) using *De novo* RNA-seq analysis. A total of 39.2 GB of RNA seq data were confirmed by Benchmarking universal Single-Copy Orthologs (BUSCO) and gene ontology (GO) enrichment analysis using Illumina Hiseq paired-end sequencing.

TABLE 1 Statistical summary of RNA-seq data used in this study.

Assembly	Trinity_cdhit
contigs (≥ 0 bp)	90084
contigs (≥ 1,000 bp)	45590
contigs (≥ 5,000 bp)	1549
contigs (≥ 10000 bp)	312
contigs (≥ 25000 bp)	47
contigs (≥ 50000 bp)	5
Total length (≥ 0 bp)	127177704
Total length (≥ 1,000 bp)	104368225
Total length (≥ 5,000 bp)	13665753
Total length (≥ 10000 bp)	5832885
Total length (≥ 25000 bp)	1637564
Total length (≥ 50000 bp)	294277
contigs	65036
Largest contig	62035
Total length	118440920
GC (%)	42.62
N50	2229
N75	1456
L50	16524
L75	32872
N's per 100 kbp	0

- This transcriptomic data can provide novel insights in waterlogging signaling mechanism and also aids in the identification of potential target genes, transcriptional factors and other key molecular players.

Material and methods

Plant material and waterlogging stress treatment

Seeds of two contrasting pigeon pea genotypes JBB-110B (Tolerant) and ICP 7035 (Sensitive) were sown in pots (0.8 m deep and 12 m diameter) containing autoclaved soilrite mixture in three biological replicates under controlled lighting and temperature conditions, with a maximum temperature of 30°C–32°C and a minimum temperature of 22°C–25°C in the Phenomics Facility (PF) at National Institute for Plant Biotechnology (NIPB), New Delhi. For both genotypes, 5 seeds were initially planted in each pot. After 14 days of germination, the plants were then thinned to three healthy plants and allowed to grow for 4 weeks. One-month old plants were exposed to waterlogging stress for 7 days. Briefly, pots were dipped in a plastic tray containing water and water level was maintained at 5 cm level above the soil. Leaf tissue samples from control and

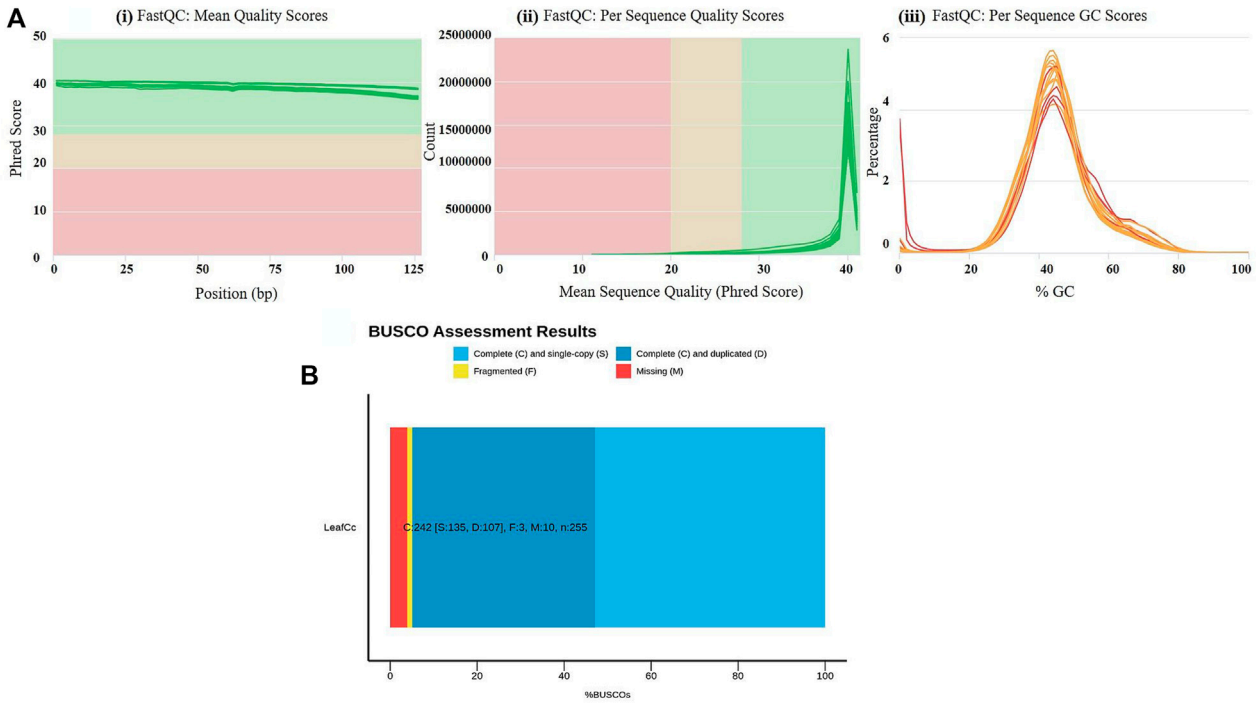


FIGURE 1
(A) Results of raw read preprocessing. (i) Mean quality scores per read. The *x*-axis represents the mean quality scores, and the *y*-axis depicts the read counts. (ii) Per sequence quality scores. The *x*-axis represents the position, and the *y*-axis depicts the Phred score. (iii) GC content of reads. The *x*-axis represents the GC content, and the *y*-axis depicts the ratio of reads. Quality assessment metrics for trimmed and filtered RNA-Seq data used to make the *de novo* transcriptome assembly. **(B)** % BUSCO assessment results of leaf RNAseq data in *C. cajan* for quality check and completeness analysis showing maximum number of unigenes categorized in complete (C), single copy (S), and duplicated (D) genes.

waterlogging stress conditions were collected (at the fourth day/first visible waterlogging induced symptom) in three biological replicates, immediately dip into liquid nitrogen and finally stored at −80°C for further processing.

RNA extraction, transcriptome library preparation, and sequencing

Total RNA was isolated from each sample (control and waterlogging treated) as per the protocol mentioned in Spectrum Plant Total RNA Extraction Kit (SIGMA). Eight RNA samples were taken for library preparation in duplicates. Overall, eight transcriptome libraries were prepared. The RNA samples were quantified using Nanodrop and Qubit. The input concentration of RNA was taken as 1 μg for library preparation. QIAseq® Stranded mRNA Select kit (Qiagen) was used for transcriptome library preparation. PolyA mRNA was enriched from total RNA, and then it was fragmented followed by first strand synthesis. Then second strand synthesis, end repair and A-addition were done. The adapters were ligated and the library was amplified by polymerase chain reaction (PCR). The final library was quantified by qubit and quality check was done using Agilent Bioanalyzer (Agilent 2,100). With the help of Agilent DNA High Sensitivity kit, the library size distribution was assessed, and the library was run on Illumina Hiseq platform.

Quality control and *de novo* transcriptome assembly

Initial quality control (QC) of both the samples in biological replicates was carried out using FastQC version v0.11.9 to check the per base sequence quality of the raw reads of leaf transcriptome. Adapter removal and trimming was performed by TrimGalore version v0.6.1 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and FASTX-Toolkit version v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit), respectively. Thereafter, the results of all samples were run on MultiQC version v1.12 (Ewels et al., 2016). After filtering, the *de novo* assembly was performed in the Trinity software version v2.14.0 with default parameters (Grabherr et al., 2011). We further used BUSCO tool version v4 to evaluate the overall completeness of the final transcriptome assembly (Manni et al., 2021). From an evolutionary perspective, it is fair to predict that these genes will be found in a given genome as single copies, hence BUSCO is excellent for determining assembly completeness. For this work, we utilized the transcriptome evaluation mode with the eukaryote lineage database (eukaryota orthoDB9). The CD-HIT software version v4.6.1 was used to obtain non-redundant unigenes (Li and Godzik, 2006; Fu et al., 2012). OmicsBox version v2.1 (https://www.biobam.com/omicsbox/) was employed for annotation based on the GO terms *viz.* Cellular components, molecular functions, and biological processes. Finally, KEGG (Kyoto encyclopedia of genes and genomes)

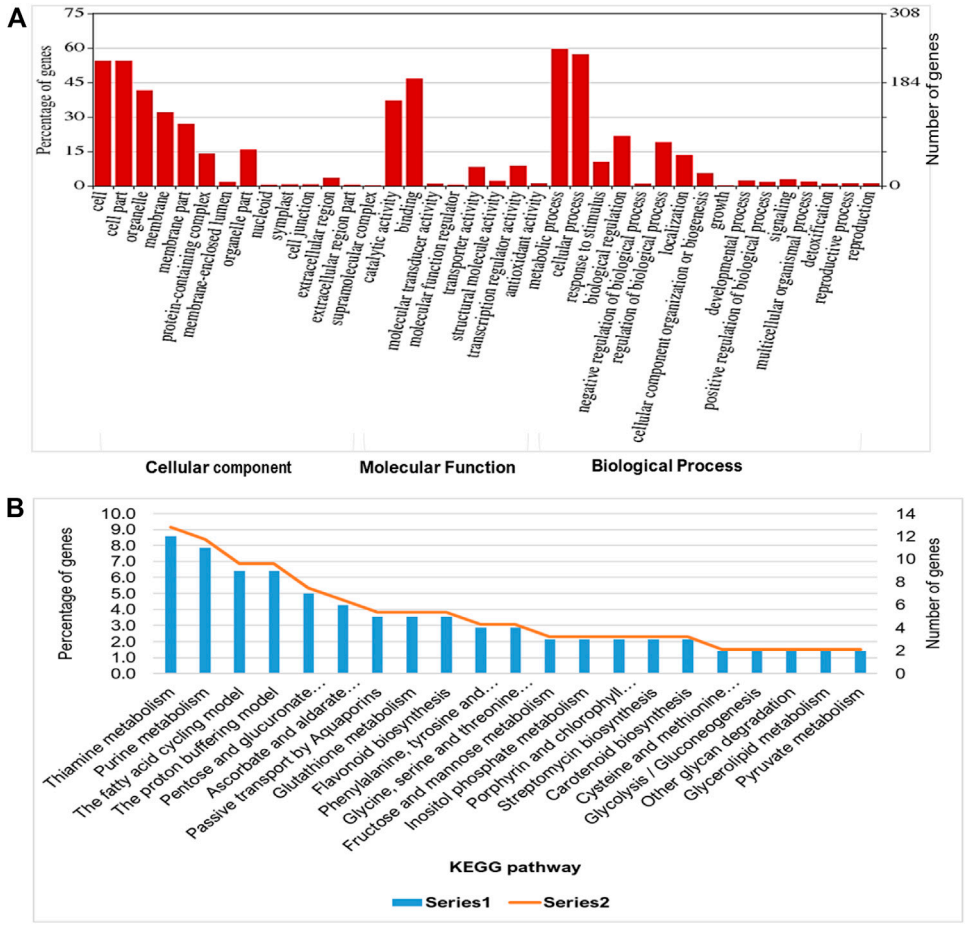


FIGURE 2
Functional enrichment analysis of predicted transcript targets during waterlogging stress in pigeon pea: **(A)** Gene ontology (Cellular components, molecular function, and biological process) and **(B)** KEGG pathway analysis to annotate the unigenes for waterlogging RNAseq data using BLASTX program.

pathway enrichment analysis of all the DEGs was carried out using OmicsBox (Götz et al., 2008).

Results

RNA-seq and *de novo* transcriptome assembly

To evaluate the waterlogging tolerance or susceptibility, we first screened numerous pigeon pea genotypes based on morphological, physiological, and biochemical parameters (unpublished data). Firstly, we checked the quality control (QC) of both the samples using FastQC version v0.11.9 to examine the per base sequence quality of the raw reads of leaf transcriptome (Figure 1A). Based on the findings, we chose two distinct pigeon pea genotypes-JBP-110B (tolerant) and ICP 7035 (sensitive)-for RNA sequencing. Illumina Hiseq sequencing run produced a total of 217,090,406 and 262,705,712 raw paired-end reads of 150 × 2 bp (base pair) from 8 RNA libraries in JBP-110 (tolerant) and ICP 7035 (sensitive) genotypes. A total of 211,324,886 clean reads from the JBP-110 and 255,693,868 from the ICP 7035 genotypes were obtained after the removal of poly-A tails, adapters, primer, short and low-quality sequences using trimming process. Additionally, Trinity software was used for the *de novo* assembly of the pooled reads (467,018,754) from both samples. There were 1,457,155 transcripts in total, with an average length of 545.10 bp and a N50 value of 1984 bp. Our assembly was shown to be relatively complete by BUSCO analysis, with 94.9% (n = 242) of BUSCOs being full sequences, just 1.2% (n = 3) being fragmented sequences, and 3.9% (n = 10) being absent in the assembly with eukaryotic lineage (Figure 1B). Using CD-HIT software, a total of 90,084 unigenes with an average length of 1,411.77 bp and a N50 value of 2,229 bp were obtained after *de novo* assembly. The unigenes had an average guanine-cytosine (GC) content of 42.62%. The *de novo* assembly statistics summary of the RNA seq data is shown in Table 1. Annotation results showed that, in total, 66,106 (73.3%) unigenes annotated from non-redundant (NR), GO and KEGG databases. Among them the maximum number of hits related to transcriptional regulation, integral component of membrane, metal ion binding, and thiamine metabolism were found to be dominant (Figures 2A, B). The dataset generated from all the samples used in current transcriptome analysis (BioProject: PRJNA637701) are deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database with accession number (SRR11940026, SRR11940027, SRR11940028, SRR11940029). This data can also be utilized for comparative studies with data from

other crops to identify similarities and differences in their adaptive responses to waterlogging.

Conclusion

In the years between 2006 and 2016, floods were responsible for about two-thirds of all crop loss and destruction globally, amounting to huge yield losses (Food and Agriculture Organization of the United Nations, 2017). Similarly, waterlogging has been a major concern in legume crops especially in pigeon pea which requires timely improvement in order to maintain crop productivity. Pigeon pea an orphan crop has been neglected for its trait improvement despite being an important crop for under developed countries. In this context, we systematically studied the effect of waterlogging stress in two contrasting pigeon pea genotypes and their transcriptional profiling. To date, this is the first comparative dataset for *De-novo* transcriptome profiling under waterlogging stress in pigeon pea. The candidate unigenes discovered in this study will be extremely important for additional thorough research, such as single nucleotide polymorphisms (SNP) calling and novel non-coding RNAs such as microRNAs (miRNAs), and long non-coding RNAs (lncRNAs), aside from studies of differential gene expression, to elucidate the molecular mechanisms governing waterlogging tolerance in pigeon pea. Additionally, using gene editing or overexpression, we might modulate their expression and functionally validate them to develop waterlogging tolerant pigeon pea cultivars.

Data availability statement

The datasets presented in this study can be found in online repositories (SRA database, http://www.ncbi.nlm.nih.gov/sra) with the following accession information: BioProject ID: PRJNA637701. The names of the repository and accession number can be found in the article.

Author contributions

KG, NKS, AG, and AT designed the experiment and finalized the manuscript. AT performed the experiment, data collection, data interpretation and wrote the manuscript. SS and AS contributed to data collection. HS contributed to bioinformatics analysis. TK contributed to library preparation. KG, AG, and SA helped in writing and editing of draft manuscript. All authors read and approved the manuscript.

Acknowledgments

All the authors are thankful to the Director, ICAR-NIPB for providing the research facility. Financial support for the first author from the DBT-RA Program in Biotechnology and Life Sciences is gratefully acknowledged. The authors also acknowledge the ICAR-CRP on Genomics (ICAR-NBFGR, Lucknow, India) for financial support of this study.

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.

References

Arduini, I., Kokubun, M., and Licausi, F. (2019). Editorial: Crop response to waterlogging. *Front. Plant Sci.* 10, 1578. doi:10.3389/fpls.2019.01578

Bansal, R., and Srivastava, J. P. (2012). Antioxidative defense system in pigeonpea roots under waterlogging stress. *Acta Physiol. Plant.* 34 (2), 515–522. doi:10.1007/s11738-012-0977-z

Bohra, A., Mallikarjuna, N., Saxena, K. B., Upadhyaya, H. D., Vales, I., and Varshney, R. K. (2010). Harnessing the potential of crop wild relatives through genomics tools for pigeonpea improvement. *J. Plant Biol.* 37 (1), 83–98.

Duhan, S., and Sheokand, S. (2020). "Mechanism of waterlogging stress tolerance in pigeonpea plants: Biochemical and anatomical adaptation under waterlogging," in *Plant ecophysiology and adaptation under climate change: Mechanisms and perspectives II*. Editor M. Hasanuzzaman (Singapore: Springer). doi:10.1007/978-981-15-2172-0_5

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32 (19), 3047–3048. doi:10.1093/bioinformatics/btw354

Eysholdt-Derzsó, E., and Sauter, M. (2017). Root bending is antagonistically affected by hypoxia and ERF-mediated transcription via auxin signaling. *Plant Physiol.* 175, 412–423. doi:10.1104/pp.17.00555

Food and Agriculture Organization of the United Nations [FAO] (2017). *The impact of disasters and crises on agriculture and food security*. Rome: FAO, 168.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28 (23), 3150–3152. doi:10.1093/bioinformatics/bts565

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids Res.* 36 (10), 3420–3435. doi:10.1093/nar/gkn176

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 (7), 644–652. doi:10.1038/nbt.1883

ICRISAT (2011). *Half-yearly progress report, june - december 2011. Selection and utilization of water-logging tolerant cultivars in pigeonpea*. Patancheruvu: ICRISAT.

Li, W., and Godzik, A. (2006). Cd-Hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22 (13), 1658–1659. doi:10.1093/bioinformatics/btl158

Luan, H., Guo, B., Pan, Y., Lv, C., Shen, H., and Xu, R. (2018). Morpho-anatomical and physiological responses to waterlogging stress in different barley (Hordeum vulgare L.) genotypes. *Plant Growth Regul.* 85 (3), 399–409. doi:10.1007/s10725-018-0401-9

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38 (10), 4647–4654. doi:10.1093/molbev/msab199

Pan, J., Sharif, R., Xu, X., and Chen, X. (2021). Mechanisms of waterlogging tolerance in plants: Research progress and prospects. *Front. Plant Sci.* 11, 627331. doi:10.3389/fpls.2020.627331

Sasidharan, R., Hartman, S., Liu, Z., Martopawiro, S., Sajeev, N., van Veen, H., et al. (2018). Signal dynamics and interactions during flooding stress. *Plant Physiol.* 176, 1106–1117. doi:10.1104/pp.17.01232

Sharma, S., Paul, P. J., Sameer Kumar, C. V., and Nimje, C. (2020). Utilizing wild *Cajanus platycarpus*, a tertiary genepool species for enriching variability in the primary genepool for pigeonpea improvement. *Front. Plant Sci.* 23 (11), 1055. doi:10.3389/fpls.2020.01055

Sultana, R. (2010). *Can a drowning pigeonpea perform*. SA Trends, 102.

Sultana, R., Vales, M., Saxena, K., Rathore, A., Rao, S., Rao, S., et al. (2013). Waterlogging tolerance in pigeonpea (*Cajanus cajan* (L.) Millsp.): Genotypic variability and identification of tolerant genotypes. *J. Agric. Sci.* 151 (5), 659–671. doi:10.1017/s0021859612000755

Tyagi, A., Raje, R. S., and Chand, S. (2017). Comparative analysis of pigeon pea genotypes for physiological traits under water logging conditions using hydroponics and pot culture. *Trends Biosci.* 10 (43), 8982–8990.

Tyagi, A., Sharma, S., Ali, S., and Gaikwad, K. (2022). Crosstalk between H_2S and NO: An emerging signalling pathway during waterlogging stress in legume crops. *Plant Biol.* 24 (4), 576–586. doi:10.1111/plb.13319

Varshney, R., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (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

Yamauchi, T., Colmer, T. D., Pedersen, O., and Nakazono, M. (2018). Regulation of root traits for internal aeration and tolerance to soil waterlogging-flooding stress. *Plant Physiol.* 176 (2), 1118–1130. doi:10.1104/pp.17.01157

Yohan, Y., Umamahesh, V., Sudhakar, P., Reddiramu, Y., and Sreedhar, D. (2017). Influence of waterlogging on certain biochemical and yield parameters of pigeonpea (Cajanus cajan (L.) millsp). *Int. J. Pure App. Biosci.* 5 (4), 1862–1868. doi:10.18782/2320-7051.2964

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 23 January 2023
DOI 10.3389/fgene.2023.996828

Check for updates

OPEN ACCESS

EDITED BY
Aamir Raina,
Aligarh Muslim University, India

REVIEWED BY
Parvaze Sofi,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India
Momina Hussain,
National Institute for Biotechnology and Genetic Engineering, Pakistan

*CORRESPONDENCE
Jai C. Rana,
j.rana@cgiar.org

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

RECEIVED 18 July 2022
ACCEPTED 05 January 2023
PUBLISHED 23 January 2023

CITATION
Gayacharan, Parida SK, Mondal N, Yadav R, Vishwakarma H and Rana JC (2023), Mining legume germplasm for genetic gains: An Indian perspective. *Front. Genet.* 14:996828. doi: 10.3389/fgene.2023.996828

COPYRIGHT
© 2023 Gayacharan, Parida, Mondal, Yadav, Vishwakarma and Rana. 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.

Mining legume germplasm for genetic gains: An Indian perspective

Gayacharan[1], Swarup K. Parida[2], Nupur Mondal[3], Rashmi Yadav[1], Harinder Vishwakarma[1] and Jai C. Rana[4]*

[1]ICAR-National Bureau of Plant Genetic Resources, New Delhi, India, [2]DBT-National Institute of Plant Genome Research, New Delhi, India, [3]Shivaji College, University of Delhi, New Delhi, India, [4]Alliance of Bioversity International and CIAT, India Office, National Agricultural Science Complex, New Delhi, India

Legumes play a significant role in food and nutritional security and contribute to environmental sustainability. Although legumes are highly beneficial crops, it has not yet been possible to enhance their yield and production to a satisfactory level. Amid a rising population and low yield levels, *per capita* average legume consumption in India has fallen by 71% over the last 50 years, and this has led to protein-related malnutrition in a large segment of the Indian population, especially women and children. Several factors have hindered attempts to achieve yield enhancement in grain legumes, including biotic and abiotic pressures, a lack of good ideotypes, less amenability to mechanization, poorer responsiveness to fertilizer input, and a poor genetic base. Therefore, there is a need to mine the approximately 0.4 million *ex situ* collections of legumes that are being conserved in gene banks globally for identification of ideal donors for various traits. The Indian National Gene Bank conserves over 63,000 accessions of legumes belonging to 61 species. Recent initiatives have been undertaken in consortia mode with the aim of unlocking the genetic potential of *ex situ* collections and conducting large-scale germplasm characterization and evaluation analyses. We assume that large-scale phenotyping integrated with omics-based science will aid the identification of target traits and their use to enhance genetic gains. Additionally, in cases where the genetic base of major legumes is narrow, wild relatives have been evaluated, and these are being exploited through pre-breeding. Thus far, >200 accessions of various legumes have been registered as unique donors for various traits of interest.

KEYWORDS

pulse production, crop domestication, biotic and abiotic stresses, legume genomics, legume collections

1 Introduction

Legumes of the family Fabaceae are among the most important plant groups on planet Earth. While legumes are an important source of food and nutrition, they also play an important role in improving soil health and ecosystem sustainability. Legume grains are often considered to be "the poor man's meat," as the vegetarian human population is highly dependent on legume grains for its protein needs (Roy et al., 2017). The "green revolution" has helped several countries to attain self-sufficiency in food, which can primarily be attributed to a manyfold increase in the production of cereals, particularly rice, wheat, and maize. However, similar advances in grain legume production have not been achieved (Figure 1), probably because legumes are less amenable to the adoption of green revolution technologies. Over 200 species of legumes are cultivated worldwide. Of these, we list the major grain legume crops, with their production and yield status and taxonomic information, in Table 1.

India is the largest producer and consumer of grain legumes globally. India's contribution constitutes around 28.12% of global grain legume production (ca. 23.37 million tonnes), and this is the output of ca. 29 million ha of cultivated land (Department of Economics and Statistics, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, GoI). Of this total, around 34% of the cultivated land (9.89 mha) is covered by cultivation of chickpea alone; this is followed by black gram (4.81 mha), pigeon pea (4.72 mha), green gram (4.61 mha), lentil (1.43 mha), and field pea (0.75 mha). Other minor legumes cultivated in India are green gram, cowpea, moth bean, grass pea, and horse gram. In 2020, global grain legume production was approximately 83.1 million tonnes; this figure is the sum of the production values of grain legumes grown globally, specifically beans (dry), chickpeas, peas (dry), lentils, cowpeas, pigeon peas, Bambara beans, and other minor pulses (www.fao.org/faostat/en). The primary contributors to this total (at 42.12 million tonnes) were the five major grain legume producing countries of India (28.12%), Canada (9.24%), Myanmar (4.84%), Nigeria (4.47%), and the Russian Federation (4.01%). The major producing countries for each crop are given in Table 1.

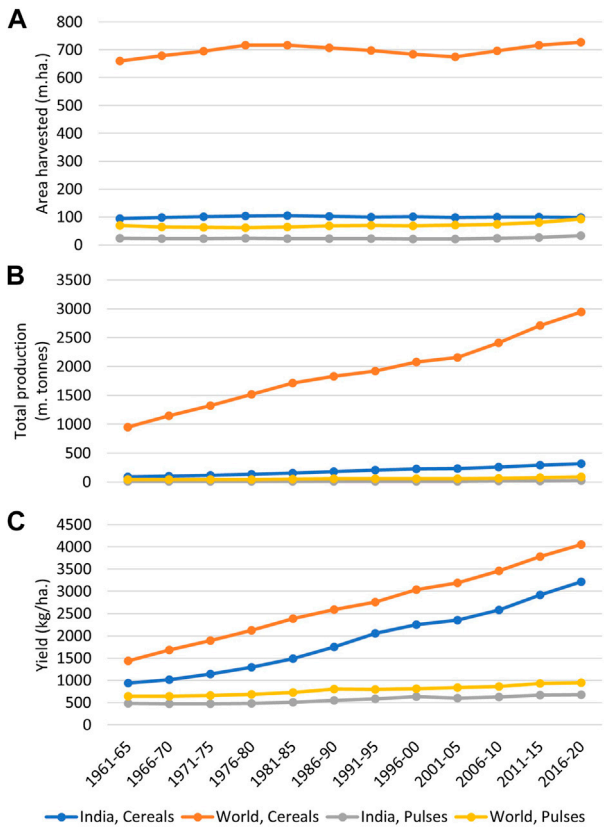


FIGURE 1
A graphical comparison of cereals and pulses in terms of total area harvested **(A)**, total production **(B)**, and yield **(C)** in India and the world. The graph indicates how the onset of the green revolution has tremendously enhanced the production of cereals in India and worldwide, which can be primarily attributed to yield improvement in these crops. By comparison, yield and production improvements in pulses have remained insignificant during this period (Data source: FAOSTAT, 2022).

Globally, over 3,800 improved cultivars of grain legumes have been released, with improvements to traits such as yield, crop duration, and nutritional qualities (Pratap et al., 2022). However, between 1961 and 2020, only a 1.5-fold increase in grain legume productivity was achieved, from 637 kg/ha to 964 kg/ha (Pratap et al., 2022; Figure 1). This can be primarily attributed to various factors, such as a narrow genetic base in cultivated gene pools, poor plant ideotype, high susceptibility to insect pests and diseases, a lack of robust seed systems, and frequent stresses from drought, heat, and flooding.

Over 850 high-yielding varieties of food legumes have been developed in India, and these are now playing a vital role in food legume production (Chauhan et al., 2016). However, the foundation of any crop breeding program is based on only a small number of parental lines, which has led to a narrow genetic base in these cultivated varieties. In a pedigree analysis, it was found that 41% of chickpea varieties had PB 7 as one of its ancestors; in pigeon pea, T 1 and T 190 appeared in 34% of varieties; and T 9 and T 1 appeared in 64% and 35% of varieties of black gram and green gram, respectively (Kumar et al., 2004). Furthermore, in the process of rigorous selection in the development of a variety, alleles conferring defense mechanisms are also lost. This is one of the reasons that the actual yield of most food legume crops is half their potential yield. Recently, drastic climatic change, to which abrupt temperature rises, erratic and heavy rainfall, frequent droughts, episodes of flooding, and rapid pest and pathogen evolution can be attributed, has exceeded the adaptation capability of modern varieties (Guo, 2022). As a result, the breakdown of resistance to biotic stress has become rather common in modern cultivars (Sharma et al., 1999; Burdon et al., 2014; Rex Consortium, 2016; Mbinda and Masaki, 2021; Hu et al., 2022; Van de Wouw et al., 2022). Therefore, *ex situ* collections are now being utilized to increase genetic variability in modern cultivars in order to improve their climate resilience, including *via* genetic gains in breeding programs. Advances in genomics, phenomics, and breeding methods are playing an important role and exerting a significant impact on legume improvement by accelerating genetic gains *via* enhancements to selection efficiency and the advancement of desired genotypes with high precision.

It is well understood that, in terms of enhancing the variability of a crop gene pool, landraces are the primary resource; the desired traits need to be sought out among these, as they are easy to cross and their use significantly reduces the chances of linkage drag as compared to the use of wild species. Additionally, landraces are well adapted to microclimatic niches and have several superior traits in terms of nutritional value. In this study, we have focused on the identification of desired genes and traits and their utilization in the improvement of legume crops. We also propose a comprehensive strategy for the enhancement of genetic gains (Figure 2).

2 Legume germplasm collections in the Indian national gene bank

The collection, conservation, and selection of germplasm are the primary components of the crop domestication process. Wild species were initially brought under cultivation and improved through selection for their agronomic traits, and the practice is still being followed by farmers and breeders. Diverse

TABLE 1 Information on the production and yield status of the major grain legume crops cultivated worldwide, along with their botanical names and chromosome numbers.

Crop	Botanical name	Chromosome number n) and ploidy level x)	Production: major producing countries (million tonnes)			Countries with highest yield (kg/ha)			Total world production (MT)
			1st	2nd	3rd	1st	2nd	3rd	
Chickpea	*Cicer arietinum*	2n = 2x = 16	India (7.06)	Australia (1.32)	Myanmar (0.56)	China (5177)	Israel (4148)	Sudan (4048)	14.6
Green gram*	*Vigna radiata*	2n = 2x = 22	India (2.45)	Myanmar (1.45)	Bangladesh (0.18)	Myanmar (1,239)	Bangladesh (1,030)	Pakistan (730)	ca. 6.0
Black gram*	*Vigna mungo*	2n = 2x = 22	India (3.06)	Myanmar (1.35)	-	Myanmar (1,432)	India (546)	-	ca. 5.0
Lentil	*Lens culinaris*	2n = 2x = 14	Canada (2.64)	India (1.24)	Australia (0.52)	Jordan (3480)	China (2476)	New Zealand (2452)	6.54
Pigeon pea	*Cajanus cajan*	2n = 2x = 22	India (3.78)	Myanmar (0.44)	Malawi (0.42)	Puerto Rico (1858)	Philippines (1821)	Thailand (1701)	5.05
Field pea	*Pisum sativum*	2n = 2x = 14	Canada (4.27)	Russia (2.58)	China (1.46)	Burundi (4809)	Lebanon (4547)	Denmark (3872)	14.65
Cowpea, dry	*Vigna unguiculata*	2n = 2x = 22	Nigeria (3.66)	Niger (2.27)	Burkina Faso (0.62)	Iraq (4083)	North Macedonia (3766)	Egypt (3637)	8.35
Beans, dry	*Phaseolus* and *Vigna* spp.	-	India (5.84)	Myanmar (2.96)	Brazil (2.90)	Mali (10042)	Montenegro (6701)	Tajikistan (6451)	27.46
Broad bean	*Vicia faba*	2n = 2x = 12	China (1.74)	Ethiopia (0.98)	United Kingdom (0.58)	Argentina (8917)	Guyana (8512)	Uzbekistan (5525)	5.47
Total pulses	—	—	—	—	—	964.04	—	—	92.29

Source: FAOSTAT (https://www.fao.org/faostat/en/), as updated on 19 December 2022. Figures represent average yield and production for the period of 2016–2020.
[a]Production and yield [a]ata for green gram and black gram are taken from two other studies (Schreinemachers et al., 2019; Khine et al., 2021).

environments of crop cultivation, including rainfed, dryland, and coastal areas, flood-prone areas, and areas at high altitude, as well as disease hotspots and human preferences in terms of nutritional qualities, aesthetics, and cultural values, have played important roles in the development and deployment of diverse germplasm. Although diversity has been continually developing and has been sustained through traditional practices over the last several thousand years, crop diversity has recently come under threat due to increasing pressure arising from demographic, sociocultural, and technological changes.

India is rich center of diversity for several cultivated crops, including important legumes such as chickpea (*Cicer arietinum*), moth bean (*Vigna aconitifolia*), rice bean (*Vigna umbellata*), cowpea (*Vigna unguiculata*), yard-long bean (*Vigna unguiculata* subsp. *sesquipedalis*), green gram (*V. radiata*), black gram (*V. mungo*), horse gram (*Macrotyloma uniflorum*), and dolichos bean (*Lablab purpureus*) (Zeven and Zhukovsky, 1975; Hawkes, 1983). The development of extensive and organized germplasm collections and conservation activity in India began only after the establishment of the National Bureau of Plant Genetic Resources (NBPGR) in 1976 (Rana et al., 2016). Since then, around 63,000 accessions of legumes have been collected and conserved in *ex situ* conditions (Table 2). The organization is continuously enriching its collections based on gap analysis with respect to earlier collections established within India and also introducing accessions from abroad. Globally, over 0.7 million legume germplasms, including their crop wild relatives (CWRs), are conserved in 276 gene banks distributed worldwide (WIEWS, 2022).

3 Utilization of grain legume germplasm for crop improvement

Crop evolution in early times was based entirely on appearance and performance in terms of agro-morphological traits, and these are still the primary focus of plant breeders and researchers. During the domestication process and subsequent structured breeding programs, genotypes with greater biotic and abiotic stress tolerance are often unintentionally selected, but agronomic traits have always been the prime target for selection. Landraces, which are locally adapted cultivars with a high level of genetic variability developed by farmers over the years, are the primary source of such traits in modern breeding programs. However, in terms of the utilization of germplasm from gene banks, it has become difficult to identify a manageable number of accessions with the desired levels of variability and traits. Recognizing this challenge, Frankel (1984) proposed the concept of a core collection, a minimum number of representative accessions representing maximum variability across the entire collection. Since then, several crop-specific diverse core sets have been developed (Table 3), and this has accelerated the utilization of gene bank collections. A number of significant

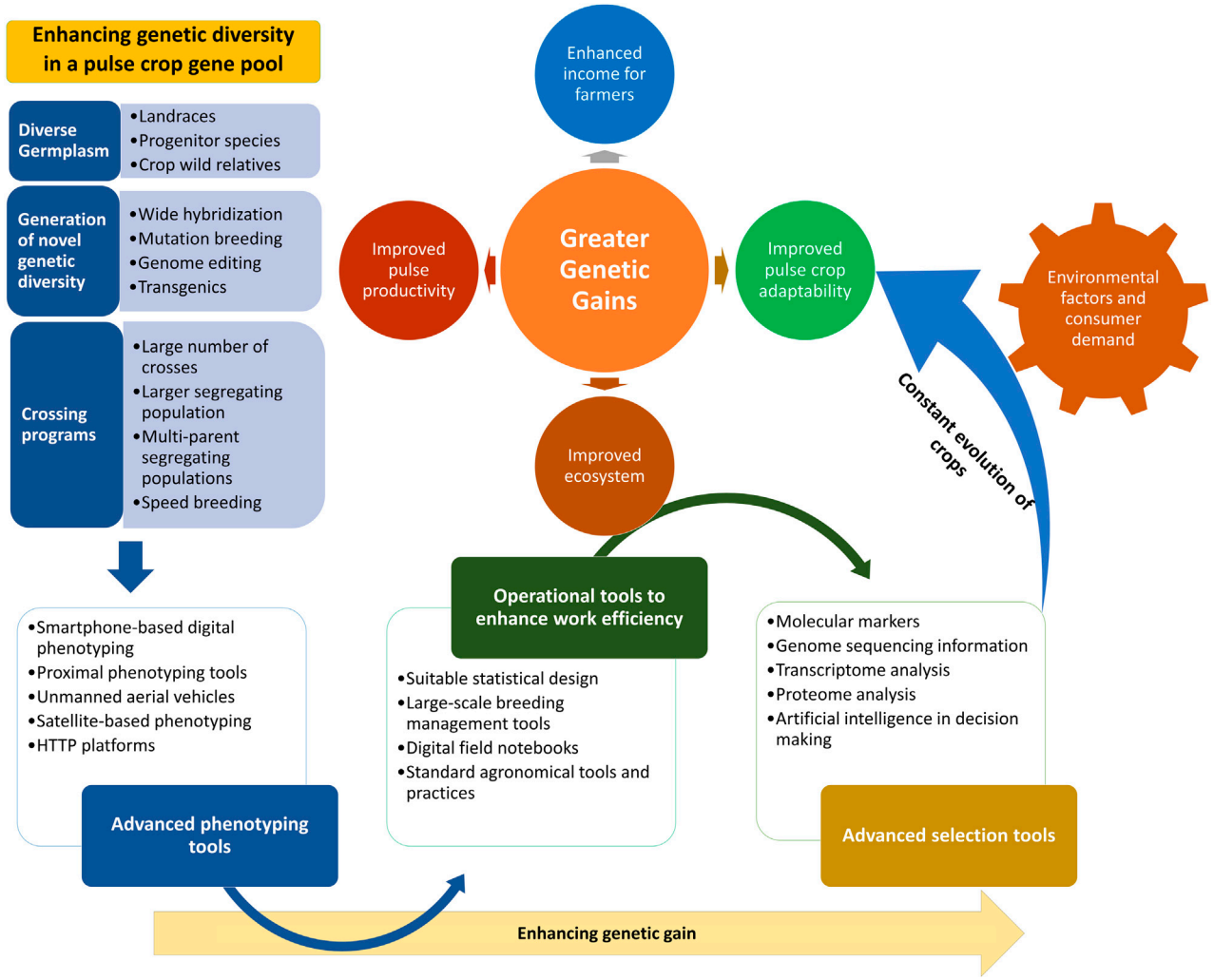


FIGURE 2
Strategy to enhance genetic gains through utilization of advanced phenotyping tools, efficient operational tools, and advanced selection methods and technologies. A strategy to achieve higher genetic gains by broadening the genetic base through the infusion of increasing levels of variability from diverse sources into the target breeding populations is illustrated. The integration of improved crossing program strategies and advanced tools for phenotyping, operations, and desired genotype selection will further enhance the genetic gains made. This strategy will help with the attainment of greater genetic gains along with enhanced crop adaptability to changing climatic conditions.

studies conducted to date in the area of trait identification and utilization are discussed below, presented in crop-wise fashion, and promising trait-specific accessions are summarized in Table 4 (biotic stress resistance) and Table 5 (abiotic stress tolerance). We also find that in the process of breeding modern varieties, the focus on yield *per se* has eventually led to a gradual decrease in the nutritional qualities of new varieties. Comparative studies on the nutritional composition of landraces and traditional cultivars in various crops, such as vegetables and fruits (Davis et al., 2004), wheat (Fan et al., 2008), the potato (White et al., 2009), the common bean (Celmeli et al., 2018), and green gram (Ebert et al., 2017), have indicated that the improved varieties are poorer than the older varieties in terms of nutritional value. Therefore, recognizing the significance of nutritional value and of the availability of nutritional variability in germplasm, we also discuss the important nutritional characteristics of each legume crop in the following sections.

3.1 Chickpea

3.1.1 Agronomic traits

Development of the first core collection in the domain of legumes was reported by Hannan et al. (1994) with the objective of making use of chickpea collections. In this study, a diverse set of 505 chickpea accessions was designated as a core set; this was derived from 7,613 accessions conserved in the Western Regional Plant Introduction Station (WRPIS), USDA. Later, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) developed a core set of 1,956 accessions based on information on the geographic origins and on 13 morphological traits for 16,991 accessions (Upadhyay and Ortiz, 2001). Following this exercise, Upadhyay et al. (2007a) identified 28 early-maturing chickpea germplasm lines having wide geographical distribution. Based on multi-location trials of the core set, ICC 16641, ICC 16644, ICC 11040, ICC 11180, and ICC 12424 were further identified as

TABLE 2 Status of collections of grain legume crops and their wild relatives available in the Indian National Gene Bank.

Crop/species name	Exotic	Indigenous	Total
Chickpea (*Cicer arietinum*)	2961	11452	14413
***Cicer* wild species** *C. bijugum* (31), *C. chorassanicum* (2), *C. cuneatum* (6), *C. echinospermum* (18), *C. judaicum* (54), *C. microphyllum* (35), *C. pinnatifidum* (27), *C. reticulatum* (18), *C. yamashitae* (4), unknown species (9)	148	56	204
Pigeon pea (*Cajanus cajan*)	306	10904	11210
***Cajanus* wild species** *C. cajanifolius* (2), *C. albicans* (3), *C. scarabaeoides* (49), *C. volubilis* (1), *C. sp.* (2), *Atylosia* (18), *Rhynchosia aurea* (1), *R. bracteata* (1), *R. himalensis* (1), *R. minima* (10), *R. sublobata* (4)	0	92	92
Lentil (*Lens culinaris*)	556	1835	2391
Other *Lens* species *L. culinaris* subsp. *odemensis* (29), *L. culinaris* subsp. *orientalis* (63), *L. culinaris* subsp. *tomentosus* (6), *L. esculenta (15), L. lamottei* (3), *L. ervoides* (67), *L. nigricans* (21), *L. odemensis* (6)	202	8	210
Pea (*Pisum sativum*)	1,082	3075	4157
Other *Pisum* species *Pisum sativum* subsp. *hortense* 7), *Pisum sativum* var. *arvense* (260)	25	242	267
Green gram (*Vigna radiata*)	535	3406	3941
Black gram (*Vigna mungo*)	5	2096	2097
Cowpea (*Vigna unguiculata*)	1063	2583	3646
Moth bean (*Vigna aconitifolia*)	37	1472	1509
Rice bean (*Vigna umbellata*)	144	1883	2027
Adzuki bean (*Vigna angularis*)	97	89	186
Yard-long bean (*Vigna unguiculata* subsp. *sesquipedalis*)	1	128	129
***Vigna* wild species** *Vigna radiata* var. *sublobata* (228), *V. radiata* var. *setulosa* (3), *V. mungo* var. *silvestris* (17), *V. angularis* var. *nipponensis* (9), *V. bourneae* (4), *V. dalzelliana* (30), *V. hainiana* (6), *V. khandalensis* (1), *V. membranacea* (1), *V. minima* (1), *V. nepalensis* (3), *V. parkeri* (2), *V. pilosa* (4), *V. racemosa* (2), *V. reticulata* (1), *V. stipulacea* (6), *V. trilobata* (144), *V. trinervia* (2), *V. trinervia* var. *bourneae* (11), *V. vexillata* (109), *V. marina* (2), *V. wightii* (1), *Vigna sp.* (13)	9	591	600
Common bean (*Phaseolus vulgaris*)	1,669	2236	3905
Horse gram (*Macrotyloma uniflorum*)	11	3122	3133
Grass pea (*Lathyrus sativus*)	90	2524	2614
Fava bean (*Vicia faba*)	354	500	854
Total	**9295**	**48294**	**57585**

Source: Indian National Gene Bank database (http://www.nbpgr.ernet.in:8080/PGRPortal).
Bold values in the first column are legume crops followed by their related wild species.

extra-early maturing lines, while ICC 14648, ICC 16641, and ICC 16644 were identified as having higher seed weight. Additionally, in an evaluation of 1,956 accessions of the chickpea core set on 14 agronomic traits, several superior accessions were identified in terms of early flowering, pods/plant, seed yield, and seed weight (Upadhyaya et al., 2007b). Furthermore, in order to reduce the size of the core collections, a mini-core set of 211 accessions was developed based on more extensive phenotypic data and a suitable statistical approach (Upadhyaya and Ortiz, 2001); this has been extensively utilized for the evaluation and identification of important traits (Upadhyaya et al., 2010). Promising accessions for traits such as water use efficiency (ICC 16374, ICC 1422, ICC 4958, ICC 10945, ICC 16374, ICC 16903) and biotic-abiotic stresses were identified (Upadhyaya et al., 2010). Erect type chickpea lines suitable for mechanical harvesting were also identified (Upadhyaya et al., 2017). A similar approach was followed at the Indian National Gene Bank to accelerate the utilization of chickpea germplasm; there, the gene bank's entire chickpea collection (14,651 accessions) was characterized and evaluated for agronomic traits in 2012, and several promising accessions in

TABLE 3 List of core collections developed for grain legume crops.

Crop	Core/mini-core collection size	Base accessions	Traits	References
Chickpea	1956	16,991	13 morphological quantitative traits; passport information	Upadhyaya et al. (2001)
	211[a]	1,956	22 morphological and agronomic traits	Upadhyaya and Ortiz, (2001)
	1,103	14,651	Eight quantitative and 12 qualitative agro-morphological traits	Archak et al. (2016)
Pigeon pea	1,290	12,153	Geographic origin; 14 qualitative morphological traits	Reddy et al. (2005)
	146[a]	1,290	18 qualitative and16 quantitative traits	Upadhyaya et al. (2006)
Lentil	287	2,390	Documented diversity	Simon & Hannan, (1995)
	170	2,324	26 agro-morphological traits	Tripathi et al. (2021a)
Green gram	1,481	5,234	Geographic origin; 8 quantitative traits	Schafleitner et al. (2015)
	152	1,532	Geographical origin; 19 quantitative and 19 qualitative traits	Bisht et al. (1998)
	289*	1,481	Phenotypic and SSR genotypic data	Schafleitner et al. (2015)
Adzuki bean	96	616	13 SSR molecular markers	Xu et al. (2008)
Common bean	171	423	Seed coat traits; geographical information; 46 SSR markers	McClean et al. (2012)
	300	544	Geographical information; morphological traits; phaseolin seed protein	Logozzo et al. (2007)
	52	388	Agro-morphological traits; phaseolin seed protein	Rodiño et al. (2003)
Cowpea	2062	12,000	Geographical information; 28 agro-botanical traits	Mahalakshmi et al. (2007)
Pea	48	731	21 SSR markers	Xu-Xiao et al. (2008)
Sem	46	249	28 agro-morphological traits	Pengelly & Maass, (2001)

[a]Mini-core.

terms of agronomic traits were identified (Archak et al., 2016). The characterization of a large number of accessions also provides the opportunity to identify rare and unique morphotypes, which sometimes turns out to be very useful. For example, in the study carried out by Archak et al. (2016), accession IC486088 was found to have upright podding behavior, which makes it a potential donor that could be used in altering chickpea plant type (Singh et al., 2013) (Figure 3). To enhance the utilization of such unique germplasm of economic or scientific value in crop improvement programs, these are registered with a national germplasm registration facility, i.e., the Germplasm Registration and Information System (GRIS; http://www.nbpgr.ernet.in:8080/registration/AboutUs.aspx). As of December 2022, a total of 28 unique accessions of chickpea have been registered by this facility. These unique traits help in the development of plant types and/or high-yielding cultivars. For example, a unique determinate phenotype was identified in BGD 9971; this is considered an important trait in the alteration of chickpea plant type (Hegde, 2011; Ambika et al., 2021). In the F2 line of an inter-specific cross, ICC 5783 (*C. arietinum*) × ICCW 9 (*C. reticulatum*), 3 to 9 flowers per flowering node were observed; this is an important trait for improving chickpea plant type and yield (Gaur and Gour, 2002). Finally, to reduce harvesting and threshing time and cost, chickpea genotypes with the erect plant type were identified and are being used to develop chickpea cultivars suitable for mechanical harvesting (Vishnu et al., 2020). GBM 2 and NBeG 47 (Dheera) are the first two such chickpea varieties that have been released (http://dpd.gov.in/Varieties/Chickpea%20varieties.pdf) (Figure 3).

3.1.2 Biotic stress

In chickpea, the major diseases are fusarium wilt (*Fusarium oxysporum* f. sp. *ciceris*), ascochyta blight (*Ascochyta rabiei* (Pass.) Lab.), collar rot (*Sclerotium rolfsii* Sacc.), dry root rot (*Rhizoctonia bataticola* (Taub.) Butler), and botrytis gray mold (*Botrytis cinerea* Pers. Ex. Fr.), and the major pest is pod borer (*Helicoverpa armigera* Hubner). Chickpea germplasm screening programs have identified rather plentiful instances of donor germplasm resistant to fusarium wilt. However, robust resistant donor sources for dry root rot, botrytis gray mold, collar rot, and pod borer are lacking; thus, germplasm use could result in the identification of moderately resistant donors for these diseases (Pandey et al., 2004; Sharma M et al., 2015; Reddy et al., 2016). The ICAR–National Bureau of Plant Genetic Resources (ICAR-NBPGR) has evaluated over 2,500 accessions for resistance to botrytis gray mold, collar rot, and dry root rot under artificial inoculation and field conditions, but only a few moderately resistant accessions have been identified, such as IC244185, IC251727, ICC6881, and IC350842 for BGM; IC270930, IC95064, IC350829, IC95100, IC209375, IC83805, IC487359, and IC83991 for collar rot; and IC413984, IC397375, IC487359, IC506915, and ICC4295 for dry root rot (unpublished data). This is a sign of the narrow genetic base of the cultivated germplasm. Similarly, no resistance sources have yet been identified for pod borer. Gayacharan U. et al. (2020) have

TABLE 4 List of important resistance sources identified for various important biotic stresses in grain legume crops.

Crop	Trait	Screened germplasm	Screening method	Sources identified	References
Chickpea	*Fusarium* wilt resistance	13,500	Field and pot conditions	160 accessions	Haware et al. (1992)
		414 germplasm/varieties	Field conditions in sick plot	35 accessions	Chaudhry et al. (2007)
		1,915 accessions of Kabuli type	Field sick plot and laboratory conditions	110 accessions	Halila and Strange, (1997)
		5,174	Screened at ICARDA	110 accessions	Singh, (1997)
	Ascochyta blight resistance	1,970 diverse germplasm	Field conditions in sick plot, multiple seasons	IC275447, IC117744, EC267301, IC248147, and EC220109	Gayacharan et al., 2020c
		19,375 germplasm	Screened at ICARDA	32 accessions	Singh, (1997)
	Collar rot resistance	98	Greenhouse conditions	FLIP 97–132C, FLIP 97-85C, FLIP 98-53C, ILC -5263, and NCS 9905	Akram et al. (2008)
Green gram	MYMV resistance	100 germplasm lines	Field conditions	014043, 014133, 014249, 014250	Iqbal et al. (2011)
		81 germplasm lines	Field conditions	IC76361, IC119020-1, PLM490, IC75200, IC119020-2, CO7, CO8	Nainu and Murugan, (2020)
		120 germplasm lines	Field conditions	EC 398897, TM-11-07, TM-11-34, PDM-139, and 6 others	Mohan et al. (2014)
	Bruchid beetle tolerance	335 germplasm lines	'Free choice' and 'no choice' test method	LM 131, V 1123, LM 371, and STY 2633	Duraimurugan et al. (2014)
	Spotted pod borer (*Maruca vitrata*) tolerance	110 germplasm lines	Field conditions	KM-9-128, KM-9-136, RMG-492, LGG-527, and LGG-538	Sandhya et al. (2014)
	Bean fly (*Ophiomyia phaseoli*) tolerance	3,713 germplasm lines	Field conditions	28 accessions	Chiang and Talekar, (1980)
Black gram	MYMV resistance	344 germplasm lines	Field conditions and artificial agro-inoculation	IC144901 and IC001572	Bag et al. (2014)
		128 germplasm lines	Field conditions	KU 96-3, NDU 12-1, NIRB 002, NIRB 003, and NIRB 004	Kumari et al. (2020)
	ULCV resistance	87 germplasm lines	Field conditions	2cm-703, 90cm-015, 93cm-006, 94cm-019, 99cm-001, IAM 382-1, IAM382-9, IAM382-15, and IAM133	Ashfaq et al. (2007)
	Bruchid beetle tolerance	140 germplasm lines	'Free choice' and 'no choice' test method	UH 82-5, IC 8219, and SPS 143	Duraimurugan et al. (2014)
Moth bean	MYMV resistance	180 germplasm lines	Field conditions, multiple seasons	IC36522 and IC36217	Singh et al. (2020)
		204 diverse germplasm lines	Field conditions	PLMO 12, IC 36096, IC 415152, IC 129177, IC 129177, and 9 others	Meghwal et al. (2015)
	Leaf crinkle virus	180 germplasm lines	Field conditions, multiple seasons	IC39786 and IC39822	Singh et al. (2020)
		44 germplasm lines	Field conditions, multiple seasons	IC39786	Vir and Singh, (2015)
	Cercospora leaf spot resistance	180 germplasm lines	Field conditions, multiple seasons	IC16218	Singh et al. (2020)
Cowpea	Aphid (*Aphis craccivora*) resistance	105 cultivated and 92 wild germplasm	Greenhouse conditions	TVNu 1158	Souleymane et al. (2013)
	Bacterial blight (*Xanthomonas axonopodis* pv. *vignicola*) resistance	50 improved cultivars	Artificial inoculation	DANILA, IT00K-1263, IT03K-324-9, and 11 others	Boukar et al. (2019)
	Cowpea Mosaic Virus (CMV) resistance	225 germplasm lines	Field conditions, multiple seasons	IC202786, IC202809, and Bellary local	Deshpand et al. (2010)
	Cercospora leaf spot resistance	225 germplasm lines			

(Continued on following page)

TABLE 4 (*Continued*) List of important resistance sources identified for various important biotic stresses in grain legume crops.

Crop	Trait	Screened germplasm	Screening method	Sources identified	References
			Field conditions, multiple seasons	IC257420, IC27502, IC91556, IC198330, IC202797, IC219574, and IC202791	Deshpand et al. (2010)
	Cowpea rust (*Uromyces vignae*) resistance	225 germplasm lines	Field conditions, multiple seasons	IC206240, IC214834, IC214835, IC219871, Guntur local, and Bellary local	Deshpand et al. (2010)
	Bruchid (*Callosobruchus maculatus*) resistance	103 germplasm lines	No-choice test method	EC528425 and EC528387	Tripathi et al. (2020)
Lentil	Wilt (*Fusarium oxysporum* f.sp. *lentis*) resistance	196 landraces	Controlled and field conditions	BGE016363, BGE019696, BGE019698, BGE019708, and 8 others	Pouralibaba et al. (2015)
		93 diverse germplasm lines	Greenhouse and sick plot conditions	IG 69549 and IG 70238	Meena et al. (2017)
	Rust (*Uromyces fabae* (Pers.) de Bary) resistance	321 germplasm lines	Glasshouse and field conditions, multiple locations	Precoz, L 1534, L 2991, L 178, L 2297, L 24123, and HPLC 8868	Kumar et al. (1997)
		286 germplasm lines	Growth chamber conditions	RR-107, ILL7207, ILL7716, and ILL7618	Rubiales et al. (2013)
	Blight (*Stemphylium botryosum* Wallr.) resistance	70 germplasm lines including wild	Growth chamber, greenhouse, and field conditions	Various promising accessions identified	Podder et al. (2013)
	Seed weevil (*Bruchus spp.*) resistance	571 germplasm lines including wild	Field conditions with artificial release of insects	32 accessions	Laserna-Ruiz et al. (2012)
	Root-knot nematode (*Meloidogyne incognita*) resistance	300 germplasm lines	Pot conditions, artificial inoculation	EC223269, EC076551-C, EC267577-D, EC267555, EC255504, and 4 others	Khan et al. (2017)
Pigeon pea	*Fusarium* wilt and sterility mosaic disease resistance	146 germplasm accessions of a mini-core collection	Artificial field epiphytotic conditions, multiple seasons	ICP 6739, ICP 8860, ICP 11015, ICP 13304, and ICP 14819 have combined resistance	Sharma et al. (2012)
	Fusarium wilt (*Fusarium udum*) resistance	104 germplasm lines	Greenhouse and field conditions	VBG 42, VBG 52, and VBG 57	Okiror, (1999)
	Sterility mosaic disease resistance	976 accessions	Artificial epiphytotic conditions, multiple seasons	ICPLs 20094, 20106, 20098, and 20115	Sharma et al. (2015)
		88 germplasm lines	Field conditions at 10 locations	ICP 7867, ICP 10976, and ICP 10977	Nene et al. (1989)
		60 accessions of *C. carabaeoides*	Leaf-stapling followed by petiole grafting	ICP15684, ICP15688, ICP15692, ICP15695 and others	Kulkarni et al. (2003)
	Spotted pod borer (*Maruca vitrata*) tolerance	271 germplasm lines	Open field screening nursery	Promising accessions from four determinate and 12 nondeterminate types	Saxena et al. (2002)
Field pea	Pea weevil (*Bruchus pisorum* L.) tolerance	602 germplasm lines	Field conditions at three locations	Ethiopian gene bank acc. 32454, 235002, 226037, and 32410	Teshome et al. (2015)
	Powdery mildew (*Erysiphe pisi*) resistance	701 germplasm lines	Natural epiphytotic conditions	EC598655, EC598878, EC598704, IC278261, and IC218988	Rana et al. (2013)
	Rust (*Uromyces viciae-fabae*) resistance	250 lines consisting of released varieties, germplasm accessions, and advance breeding lines	Multilocation, field conditions, and further validation of 23 selected lines	IPF-2014-16, KPMR-936, and IPF-2014-13	Das et al. (2019)
Common bean	Angular leaf spot (*Phaeoisariopsis griseola*) resistance	1944 diverse germplasm lines including wild	Field and screenhouse conditions	Various resistance sources identified	Mahuku et al. (2003)
		300 germplasm lines	Field conditions	14 resistant accessions	Rezene and Mekonin, (2019)
	Damping-off (*Rhizoctonia solani*) resistance	274 germplasm lines	Artificial inoculation, pot conditions	PI 310668 and PI 533249	Peña et al. (2013)
	Fusarium root rot (*Fusarium cuneirostrum*) resistance	248 wild germplasm	Greenhouse, small pots, and artificial inoculation	PI417775 highly resistant; 21 others resistant	Haus et al. (2021)

(Continued on following page)

TABLE 4 (*Continued*) List of important resistance sources identified for various important biotic stresses in grain legume crops.

Crop	Trait	Screened germplasm	Screening method	Sources identified	References
	Fusarium wilt (*Fusarium oxysporum*) resistance	248 wild germplasm	Greenhouse, small pots, and artificial inoculation	PI661845 and PI535441 highly resistant; 16 others resistant	Haus et al. (2021)
	Bacterial wilt (*Curtobacterium flaccumfaciens* pv. *flaccumfaciens*) resistance	467 diverse germplasm	Pot conditions, artificial inoculation	PI 325691	Urrea & Harveson, (2014)

identified several robust resistance sources (viz., IC275447, IC117744, EC267301, IC248147, and EC220109) for ascochyta blight disease using the sick plot method following artificial inoculation in multiple environments and seasons; these are now being utilized in national chickpea breeding programs. Pande et al. (2006) have also identified several other promising chickpea accessions (*viz.*, ICC 17211, IG 69986, IG 70030, IG 70037, and IG 70038), which have shown combined tolerance against ascochyta blight and botrytis gray mold diseases. Finally, Singh (1997) has listed several of the important sources identified at the International Center for Agricultural Research in the Dry Areas (ICARDA) and at the ICRISAT. Various important sources of biotic stress resistance are listed by Singh et al. (2022) and are also presented in Table 4 and Figure 4.

3.1.3 Abiotic stress

In chickpea, the major abiotic stresses are terminal drought, terminal heat stress, and low temperatures during the late vegetative stage. The Northern Plains of India, which was once the most favorable zone for chickpea production, has faced a drastic decline in production of this crop due to a sharp rise in the minimum night temperature (Basu et al., 2009). Terminal heat stress is also a major challenge in the expansion of chickpea cultivation to rice-fallow land, of which there are around 11.7 million ha in the country (Singh N et al., 201a). Therefore, in order to tackle this problem, new sources of tolerance are being sought to enable the development of short-duration and heat-tolerant varieties. Basu and coworkers (2009) screened chickpea germplasm and identified several highly heat-tolerant chickpea lines, *viz.*, ICCV 92944 (JG14), ICCV 37, ICC67, JKG 1, GCP 101, and PG 12. A reference collection for heat stress tolerance has also been developed and screened at the reproductive stage (Krishnamurthy et al., 2011). The researchers observed broad genetic variation in heat-responsive traits, and later identified 10 heat stress tolerance lines under field conditions (Devasirvatham et al., 2015). A mini-core collection has also been screened for resistance to drought stress; five highly tolerant accessions (namely, ICC867, ICC 1923, ICC9586, ICC12947, and ICC14778) were identified (Krishnamurthy, et al., 2010). A germplasm line (ICC4958) developed by Saxena et al. (1993) has been extensively used for breeding drought-tolerant varieties. In-depth molecular analysis of the ICC4958 line has identified a QTL-hotspot region that harbors several traits related to drought tolerance (Bharadwaj et al., 2021). Certain other drought stress tolerant lines have been identified under field conditions (Ganjeali et al., 2011). A mini-core set was also screened for salinity stress resistance under pot conditions and artificial application of saline water (100 mM), which led to the identification of 10 highly tolerant accessions (Serraj et al., 2004). Genotypes ICCV 00104, ICCV 06101, CSG8962, and JG62 have also been identified as promising in terms of salinity tolerance (Kumar et al., 2016). Additionally, a total of 3,276 germplasm lines of chickpea were evaluated against cold stress at the ICARDA, Tel Hadya, Syria, between 1981 and 1987; 21 lines were found to be tolerant of cold stress (Singh et al., 1989). Choudhary and coworkers (2018) list several popular donors that represent the major sources for improvement of chickpea tolerance to abiotic stress. An extensive list of chickpea germplasm lines that have been identified as promising in relation to various abiotic stresses is also given in Table 5.

3.1.4 Nutritional Quality

The chickpea is well known for its nutritionally rich grains, which are widely used as an alternative source of supplementary nutrients. Chickpea grains contain 63% total carbohydrate, 21% protein, and 2.70%–6.48% total fat (Wang et al., 2021). The prominent minerals are K (1.2 g/100 g in desi type, 1.1 g/100 g in Kabuli type), P (0.38 g/100 g in desi, 0.5 g/100 g in Kabuli), Mg (169 mg/100 g in desi, 178 mg/100 g in Kabuli), and Ca (162 mg/100 g in desi, 107 mg/100 g in Kabuli; Wang et al., 2021). Chickpea grains are also a good source of vitamins C, B2, B3, B5, γ–tocopherol, E (α–tocopherol), and folic acid.

Large-scale nutritional profiling has not yet been carried out for legumes, primarily due to a lack of high-throughput nutritional profiling platforms. However, nutrient-specific donors with high mineral content have been identified, such as for Zn (MG–13, MG–17), Ca (PI518255, PI358934), and P (PI339154), and these can be used for biofortification of modern chickpea cultivars (Constantini et al., 2021). In an analysis of 79 accessions, one (LEGCA728) was identified as having high lutein content (28.32 μg g^{-1}), and distinct morphotypes were identified as superior in terms of high concentration of specific nutrients (Serrano et al., 2017). In this study, it was observed that nutritional variation is associated with seed morphology. Black and brown seeded varieties were found to have higher dietary fiber content, ranging from 18.0 to 22.1 g 100 g^{-1}, and higher polyunsaturated fatty acid (PUFA) content (67.0 g 100 g^{-1} of total fatty acids; Summo et al., 2019). Accessions with brown coloring also have high water absorption capacity (1.9 g water g^{-1} of flour), which makes these varieties suitable for mixing with cereal flours to produce nutritionally rich cereal-based food products. The vitamins, minerals, and fibers present in chickpea grains promote their utilization for many health benefits. Finally, carotenoid concentration (with the exception of lycopene) has been found to be higher in wild germplasm as compared to cultivated types (Jukanti et al., 2012).

3.2 Lentil

3.2.1 Agronomic traits

Lentil is one of the eight founder crops of agriculture (Ambika et al., 2022) and the most nutritious cool season legume cultivated in many farming systems worldwide. Lentil is divided into two categories

TABLE 5 List of important resistance sources identified for various important abiotic stresses in grain legume crops.

Crop	Trait	Germplasm screened	Screening method	Sources identified	References
Chickpea	Salinity tolerance	600 selected based on various strategies	Pot conditions	33 lines	Maliro et al. (2008)
		211 acc. of chickpea mini-core collection and 41 popular varieties	Pot conditions with 100 mM NaCl solution to field capacity of the soil	ICC 10755, ICC 13124, ICC 13357, ICC 15406, ICC 15697, and 5 others	Serraj et al. (2004)
		180 germplasm lines	Paper cup, greenhouse conditions, 8 ds/m electrical conductivity	ICCV 00104, ICCV 06101, CSG8962, and JG62	Kumar et al. (2016)
	Heat stress tolerance	References set of 200 accessions having very long duration	Late-sown, field conditions	18 lines	Krishnamurthy et al. (2011)
		167 accessions	Late- vs. timely-sown, field conditions	ICCV 95311, ICCV 98902, ICCV 07109, ICCV 92944, ICC 6969, and 5 others	Devasirvatham et al. (2015)
		35 early-maturing lines	Late- vs. timely-sown, field conditions	ICC 14346	Upadhyaya et al. (2011a)
	Drought tolerance	211 accessions of mini-core collection	Field conditions	ICC 867, ICC 1923, ICC 9586, ICC 12947, and ICC 14778	Krishnamurthy et al. (2010)
		1,500 diverse germplasm	Field conditions	ICC4958	Saxena et al. (1993)
		150 Kabuli type germplasm	Field conditions	MCC544, MCC696, and MCC693	Ganjeali et al. (2011)
	Cold tolerance	14 accessions	Field and controlled environments	ICCV 88502 and ICCV 88503	Srinivasan et al. (1998)
		3,276 germplasm and breeding lines	Field conditions	21 lines	Singh et al. (1989)
Green gram	Heat stress tolerance	41 elite lines	Late- vs. timely-sown, field conditions	EC693357, EC693358, EC693369, Harsha, and ML1299	Sharma et al. (2016)
	Drought tolerance	100 diverse germplasm	Hydroponics in controlled conditions	IC333090 and IC507340	Meena et al. (2021)
Black gram	Salt tolerance	48 genotypes	Various salinity levels at seedling stage	VNBG 017, AUB 3, and AUB 20	Priyadharshini et al. (2019)
	Waterlogging tolerance	290 germplasm lines	Pot conditions, 10 days of flooding 30 days after sowing	IC530491 and IC519330	Bansal et al. (2019)
Moth bean	Drought tolerance	32 diverse germplasm	Withdrawal of irrigation, field conditions	IC129177, IC103016, IC415139, IC 415155, IC36157, Maru moth, and Jadia	Malambane and Bhatt, (2014)
		15 diverse germplasm	Withdrawal of irrigation, field conditions	IC103016, IC36011, and IC36157	Sachdeva et al. (2016)
Cowpea	Drought tolerance	1,288 randomly selected lines	Withdrawal of irrigation, field conditions	TVu1436, TVu9693, TVu12115, TVu14632, and TVu15055	Fatokun et al. (2012)
	Salt tolerance	151 germplasm lines	Artificial conditions with 150 mM NaCl application at germination stage	PI582422, 09–529, PI293584, and PI582570	Ravelombola et al. (2017)
		155 germplasm lines	Artificial conditions with 200 mM NaCl application at seedling stage	PI354686, PI353270, PI354666, and PI354842	Dong et al. (2019)
		116 acc. at germination stage and 155 acc. at seedling stage	Artificial conditions with 150 and 200 mM NaCl application at germination and seedling stage screening, respectively	Trait-specific promising genotypes	Ravelombola et al. (2018)
	Heat tolerance	130 germplasm lines	Field conditions, multiple seasons	EC472250, EC472267, EC$&2285, EC472286, EC472289, and Pusa Komal	Mishra et al. (2005)
Lentil	Combined terminal heat and drought stress tolerance	166 selected through FIGS[#]	Field conditions at two contrasting locations	ILL 7835, ILL 6075, ILL 6362, ILL 7814, ILL 7835, and ILL 7804	Rajendran et al. (2020)

(Continued on following page)

TABLE 5 (*Continued*) List of important resistance sources identified for various important abiotic stresses in grain legume crops.

Crop	Trait	Germplasm screened	Screening method	Sources identified	References
	Boron tolerance	310 germplasm lines	Field conditions at seedling stage	ILL213A and ILL2024	Hobson et al. (2006)
	Salt tolerance	133 germplasm lines	Germination and seedling stage, NaCl application	ILL 5845, ILL 6451, ILL 6788, ILL 6793, and ILL 6796	Ashraf and Waheed, (1990)
Pigeon pea	Waterlogging tolerance	272 diverse accessions	*In vitro* laboratory conditions and natural field conditions	ICPH 2431, ICPH 2740, ICPH 2671, and 9 others	Sultana et al. (2013)
		146 accessions of mini-core collection	Pots placed in water tanks, multiple durations and seasons	24 accessions	Krishnamurthy et al. (2012)
Adzuki bean	Drought tolerance	80 germplasm lines	Mannitol-induced drought stress	-	Zhu et al. (2019)
Field pea	Cold tolerance	3,672 germplasm lines	Field conditions	214 accessions	Zhang et al. (2016)
	Frost tolerance	83 accessions collected from 34 countries	Controlled environmental chamber	ATC 104, ATC 377, ATC 968, ATC 3992, and ATC 4204	Shafiq et al. (2012)
	Salinity tolerance	780 globally distributed germplasm	Artificial conditions, using NaCl	ATC1836	Leonforte et al. (2013)
	High temperature tolerance	150 genotypes	Field conditions; timely, moderately late, and very late sowing	IPFD 11-5, Pant P-72, P-1544-1, and HUDP 11	Lamichaney et al. (2021)

#FIGS: focused identification of ermplasm strategy.

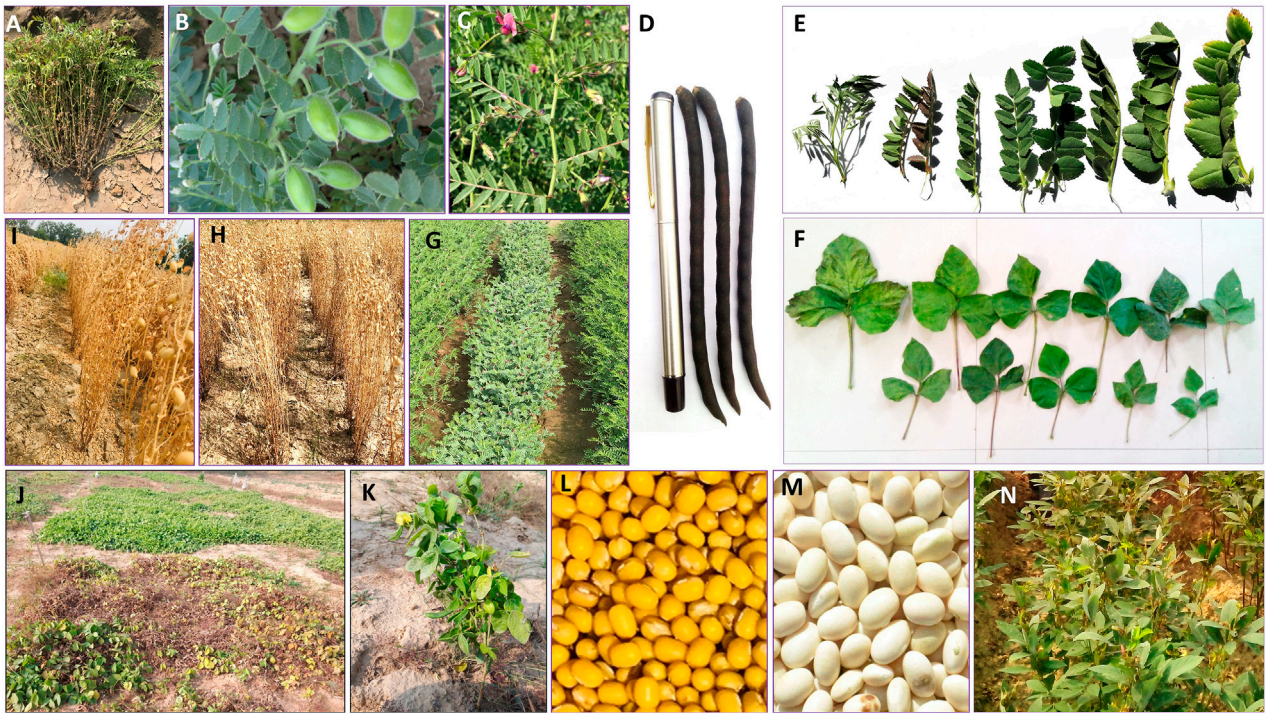


FIGURE 3
Highlights of various important agro-morphological variations. Genotype (ICC16358) with a large number of branches per plant **(A)**; genotype (IC486088) having upright peduncle and pods **(B)**; genotype (ICC15559) with two to three flowers/peduncle **(C)**; genotype (EC398937) with greater pod length (>15 cm) and a higher number of seeds/pod **(D)**; leaflet size variation **(E)**; leaf size variation **(F)**; genotype with short internode length and compact phenotype **(G)**; genotype (IC24417) with erect and tall growth habit **(H)**; an erect genotype (NBeG 47) in chickpea **(I)**; early-maturing (IC347181) **(J)** and erect type (VLG 39) **(K)** genotypes of horse gram; sona mung with bright yellow seeds having superior visual appeal **(L)** in green gram; a common bean germplasm (IC341862) having pea-shaped, bright white-colored seed with superior visual appeal **(M)**; and a pigeon pea genotype with determinate growth habit **(N)**.

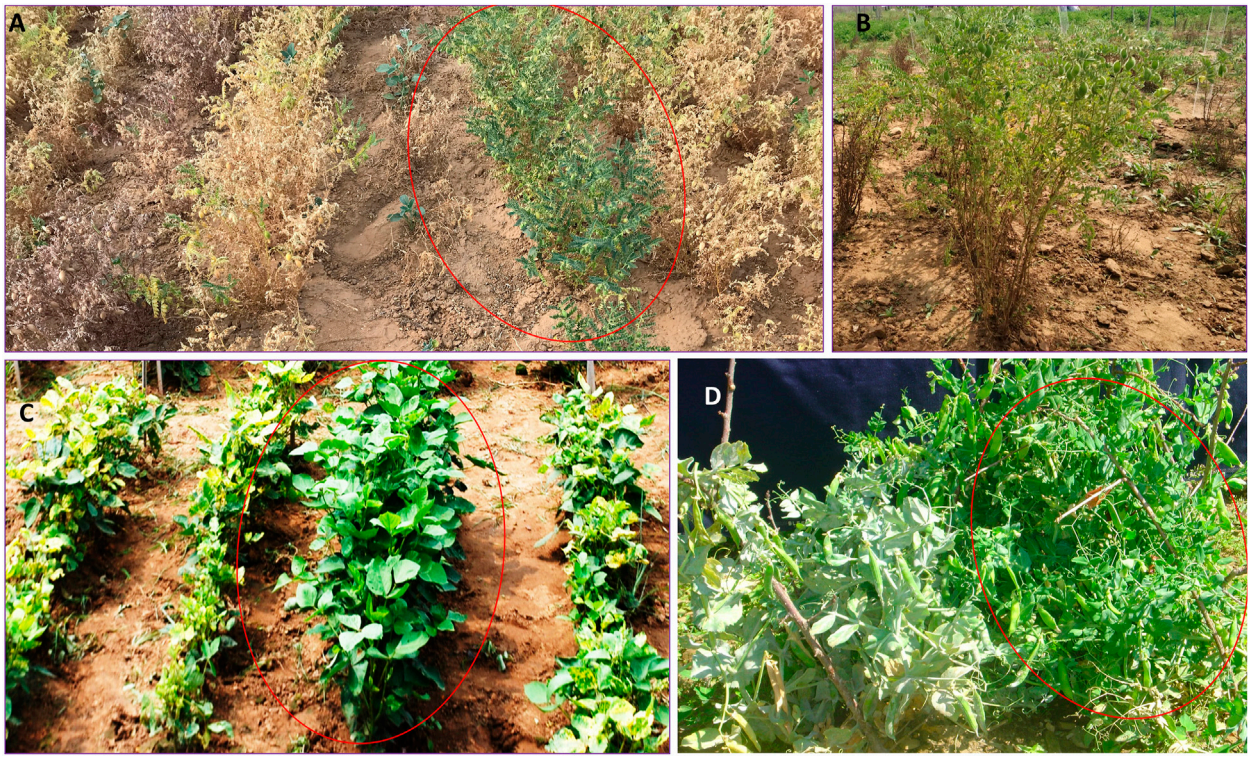


FIGURE 4
Several promising newly identified resistance donors: IC486215 **(A)** for resistance against dry root rot in chickpea; IC275447 **(B)** for resistance against ascochyta blight in chickpea; IC118998 **(C)** for yellow mosaic disease (YMD) resistance in green gram; and IC278261 **(D)** for powdery mildew resistance in field pea.

based on seed size, i.e., microsperma (seed diameter 2–6 mm) and macrosperma (seed diameter 6–9 mm), with 100 seed weight ranging from 1.5 to 8.0 g. In order to identify new germplasm for various agro-morphological traits, extensive germplasm exploration, characterization, and evaluation programs have been undertaken globally. As a result, several trait-specific donors have been identified and used to develop improved varieties. High biomass, good plant standing, higher seed weight, and number of pods/peduncle are considered to be crucial traits for yield enhancement in lentil. With the availability of such genotypes having tall (>30 cm), erect growth habits and good standing ability with good ground clearance (>15 cm), high-yielding varieties that are suitable for mechanical harvesting have been developed, such as ILL590, ILL1005, ILL6037, ILL6212, ILL6994, ILL7155, and ILL7947 (Sarker and Erskine, 2006; Kumar et al., 2013). Germplasm lines for early flowering (IC560333, IC559639, IC560111, and IC560148), high biomass (IC559744, IC559608, IC559767, and IC560040), a large number of primary branches (IC559870, IC318881, IC398688, and IC560182), and high yield (IC398094, IC560212, IC560332, and IC560206) have also been identified through gene bank germplasm characterization (Gautam et al., 2013). Mishra et al. (2022b) identified PMF-1, PMF-2, PMF-3, and PMF-4 as producing multiple flowers per peduncle, which is an important trait in lentil breeding. The GRIS portal indicates the registration of accessions for a range of important unique traits, such as extended funiculus, which helps with rapid water uptake (IC317520); multiple flowers and pods per peduncle (IC241473); early flowering and maturity (IC241532); and extra bold seeds (EC499760). A core set of 287 accessions was developed for lentil using diversity documentation on 3,068 accessions conserved at the WRPIS, USDA, by Simon and Hannan (1995). Promising germplasm lines have been identified for various agronomic traits, such as seedling vigor, earliness, number of pods/peduncle, number of pods/plant, and seed weight (Singh., 1995). In another large-scale characterization conducted at the ICAR–NBPGR, accessions were characterized on 26 agro-morphological traits, and a core set of 170 accessions was developed (Tripathi et al., 2021a). Kumar et al. (2013) have also highlighted important lines for traits such as winter hardiness, short duration type, mechanical harvesting, and higher seed weight.

3.2.2 Biotic stress

The major diseases in the lentil crop are fusarium wilt (*Fusarium oxysporum* f. sp. *lentis* (Fol)), root rot complex, rust (*Uromyces viciae-fabae*), stemphylium blight (*Stemphylium botryosum*), powdery mildew (*Erysiphe* spp.), and ascochyta blight (*Ascochyta lentis*); the major pests are pod borer (*Etiella zinkenella*), aphids (*Aphis craccivora*), and seed weevil (*Bruchus spp.*). Several studies have been conducted to identify resistant donors, some of which are listed in Table 4. For example, 12 accessions were identified as resistant to fusarium wilt out of 196 landraces screened under both field and controlled conditions (Pouralibaba et a., 2015). In another study, 93 accessions were screened under three different screening conditions (specifically, a hotspot location, field sick plot, and artificial greenhouse

conditions) for resistance to wilt, and two highly resistant germplasm lines (viz., IG 69549 and IG 70238) were identified (Meena et al., 2017). Sources for fusarium disease resistance, such as ILL5883, ILL5588, ILL4400, and ILL590, and for resistance to other important diseases, such as rust (ILL358, ILL4605, ILL5604, ILL6002, and ILL6209), ascochyta blight (Indianhead, ILL358, ILL857, ILL5562, ILL5588, ILL5684, ILL5883, and ILL6024), and Stemphylium blight (ILL 4605), have been highlighted by Kumar et al. (2013). Additionally, 4 lines (RR–107, ILL7207, ILL7716, and ILL7618) have been identified as resistant to rust (*Uromyces fabae* (Pers.) de Bary) out of 286 accessions screened under controlled conditions (Rubiales et al., 2013). Blight (*Stemphylium botryosum* Wallr.) resistance has also been identified in wild lentil germplasm (Podder et al., 2013). Seed weevil (*Bruchus spp*.) is another major threat to lentil grains; therefore, 571 accessions of lentil originating from 27 different countries were evaluated under natural field conditions in central Spain, with wide variation (0%–70%) being observed in infestation rate in the lentil germplasm (Laserna–Ruiz et al., 2012). In this study, a total of 32 accessions with lower levels of infestation were identified. In a separate screening of 300 lentil accessions against root-knot nematode (*Meloidogyne incognita*), 9 accessions were identified as tolerant (Khan et al., 2017; Table 4). Furthermore, these have been registered with the GRIS portal to enhance the utilization of such important sources in lentil breeding programs. Examples of such accessions include IC296883 for multiple resistance against *Meloidoyne incognita*, *M. javanica*, Botrytis gray mold, and pod borer; IC567650 for rust resistance; and IC559673 and IC559890 for nematode resistance.

3.2.3 Abiotic stress

Drought, heat, cold, frost, salinity, and waterlogging are the major abiotic stresses affecting lentil cultivation around the world. Several studies have been conducted to identify germplasm tolerant to these stresses (Table 5). In one such study, 166 lentil accessions were screened under field conditions, and six lines (ILL 7835, ILL 6075, ILL 6362, ILL 7814, ILL 7835, and ILL 7804) were identified for combined heat and drought stress tolerance (Rajendran et al., 2020). The Focused Identification of Germplasm Strategy (FIGS) was used to select 162 accessions for screening against heat and combined heat–drought stresses under field conditions at two locations (El Haddad et al., 2020); one germplasm line (IC621470) has been registered for drought tolerance on the GRIS portal. Based on a salt stress tolerance index, several promising accessions (ILL 5845, ILL 6451, ILL6788, ILL 6793, and ILL 6796) were identified in a screening of 133 accessions under artificial conditions (Ashraf and Waheed, 1990). Accessions ILL52, ILL465, ILL 1878, ILL 1918, ILL7115, ILL7155, ILL468, ILL590, ILL662, ILL669, ILL780, ILL857, ILL975, WA8649041, and WA8649090 have been selected for winter hardiness (Erskine et al., 1981; Hamdi et al., 1996). Additionally, Stoddard et al. (2006) have identified ILL5865 and ILL1878 as lines with good levels of tolerance to freezing. In Australia, in order to expand lentil cultivation, 310 accessions were screened in soils with a high boron concentration; accessions originating from Afghanistan and Ethiopia were found to perform comparatively well under these conditions. Boron-tolerant accessions ILL213A and ILL2024 were also recorded as having higher biomass than boron-intolerant accessions (Hobson et al., 2006).

3.2.4 Nutritional quality

Lentil has a high protein content (20%–27%; Zaccardelli et al., 2012) and contains 2%–3% fat, 50%–65% starch, and 8%–9% soluble sugars (Jood et al., 1998). Lentil protein is considered to be among the most beneficial, as it has good Leu/Ile and Leu/Lys ratios (1.24–1.98 and 1.08–2.03, respectively), high digestibility (~83%), and strong potential for use in food products (Jarpa-Parra, 2018). Among pulses, lentil is also one of the richest sources of Zn and Fe. A screening of over 2000 cultivated and wild germplasm has revealed a wide range of variation in Fe (42–168 ppm) and Zn (22–101 ppm; Mehra et al., 2018), with accessions EC78933 and EC 78414 found to have particularly high Fe and Zn content, respectively. In one of the experiments conducted by Kumar et al. (2014), 41 genotypes were examined for stability of Zn and Fe content over three locations; L 4704 (136.91 mg/kg grain) and VL 141 (81.542 mg/kg grain) were found to be promising in relation to Fe and Zn, respectively. A germplasm line (IC317520) with an extended funicle has also been identified; this is expected to be associated with shorter cooking time (Tripathi et al., 2021b). Several genotypes have been registered with GRIS: IC208329 and IC208326 for high protein content (27.4%–28.5%), and IC0616579 for high iron 136.91 (mg/kg grain) and zinc (71.69 mg/kg grain) content.

3.3 Common bean

3.3.1 Agronomic traits

Common bean is an economically important legume and is cultivated worldwide. In order to assess phenotypic variability in the *ex-situ* collections of the Indian National Gene Bank, 4,274 accessions were characterized on 22 traits, and a good range of variation was observed in leaf length, leaf width, pod length, number of pods per plant, number of seeds per pod, and seed weight (Rana et al., 2015). Promising accessions were identified for early flowering (IC370764), pod length (IC328871, EC271552), pods/plant (EC500299), early maturity (EC0944456), a large number of seeds/pod (IC383008), *etc.* In another study, 203 accessions of a core collection were examined for seed quality traits and to identify promising germplasm lines (Saba et al., 2016). The Andean Diversity Panel (ADP), a regional core collection comprising 396 accessions, with the majority originating from the Andean region, was established in order to enhance germplasm utilization in the region's common bean improvement program (Cichy et al., 2015a). The ADP consists primarily of popular cultivars, breeding lines, and landraces. The CIAT gene bank conserves over 40,000 common bean accessions, making it the largest collection in the world. In an evaluation of a core set of 1,414 accessions, Amirul et al. (2006) observed wide variability in their morphological, biochemical, and nutritional traits. Through 12 multi-environment trials, a recent study has also identified four specific germplasm from 481 breeding lines with notable agronomic traits; the authors also developed a model to predict genotypic performance under different environmental conditions (Keller et al., 2020).

3.3.2 Biotic stress

The common bean is affected by many bacterial, fungal, and viral diseases, as well as insect pests. Several studies have been conducted to identify resistant sources in common bean germplasm (Table 4). A recent study has identified 14 accessions resistant to angular leaf spot

(*Phaeoisariopsis griseola*) under field conditions (Rezene and Mekonin, 2019). Peña et al. (2013) screened 274 germplasm lines under artificial and pot conditions and identified two lines (PI 310668 and PI 533249) showing resistance against damping-off disease (*Rhizoctonia solani*). A set of 248 accessions of wild bean (*Phaseolus* spp.) were screened under greenhouse, pot, and artificial conditions against fusarium root rot (*Fusarium cuneirostrum*) and fusarium wilt (*Fusarium oxysporum*), resulting in the identification of 21 and 16 lines resistant to fusarium root rot and fusarium wilt, respectively (Haus et al., 2021). Urrea and Harveson (2014) carried out screening of 467 germplasm lines against bacterial wilt (*Curtobacterium flaccumfaciens pv. flaccumfaciens*) under pot and artificial conditions, and identified PI 325691 as a resistant line to the disease. The GRIS portal also indicates that several accessions have been registered as resistant to important diseases, such as anthracnose (IC0341862, IC635031, and IC635032), white mold disease (EC271515 and IC278744), and bean common mosaic virus (IC340947 and IC0360831).

3.3.3 Abiotic stress

The common bean is severely affected by abiotic stresses, such as cold, drought, heat, and salinity, and not a great deal of work has been carried out to identify trait-specific donors, with a few exceptions. Urrea and Porch (2009) screened 277 accessions of *P. vulgaris* and *P. acutifolius* under terminal drought stress conditions at Mitchell. The G35346 line has been identified for aluminum (Al) tolerance and used to transfer Al tolerance to common bean varieties (Butare et al., 2012). Tepary bean (*P. acutifolius* Gray), a relative of common bean, is known to have comparatively better tolerance for drought and sub-zero temperatures; on this basis and through preliminary screening of tepary bean accessions, one accession (W6 15578) has been identified as a potential donor for tolerance of both these stresses (Souter et al., 2017). Additionally, Dasgan and Koc (2009) screened 64 lines at 125 mM NaCl to identify salt-tolerant donor lines; a good level of variation was observed, and five highly tolerant genotypes were identified: Yalova 5, TR68587, Kibris Amerikan, Magnum, and Yerhammadisi.

3.3.4 Nutritional quality

Common bean is an excellent source of protein, dietary fiber, vitamins, and minerals. Its grains are a rich source of water-soluble vitamins, particularly thiamin, riboflavin, niacin, and folic acid. Analysis of a Chilean bean core collection of 246 accessions revealed protein content ranging from 183.5 to 259.7 g kg^{-1}, Fe content from 68.9 to 152.4 mg kg^{-1}, and Zn content from 27.9 to 40.7 mg kg^{-1} (Paredes et al., 2009). Kaur et al. (2009) also studied the physicochemical, hydration, textural, and cooking properties of common bean, observing a wide range of variation in terms of seed density (0.51–2.15 g/ml), hydration capacity (0.03–0.62 g/seed), hydration index (0.16–0.97), swelling capacity (1.24–1.93 ml/seed), cooking time (50–120 min), and amylose content (0.09%–5.02%). Another study revealed the ranges of variation in common bean for antioxidant activity (5.5%–44.9%), starch content (17.4%–40.7%), size of starch granules (1.64–176 μm), rapidly digestible starch (11.1%–19.5%), slowly digestible starch (8.5%–17.3%), and resistant starch (63.9%–76.1%; Sharma et al., 2015). Common bean is well known for rich diversity in seed coat color, and this color plays a major role in the selection, taste, and palatability of particular genotypes. Therefore, to investigate the relationship between color and protein and mineral content, a study was conducted in 100 genotypes having carioca, black, and other grain color patterns (Silva et al., 2012). The results indicated that black-colored beans are richer in protein, iron, and zinc; carioca grains are richer in manganese and magnesium; and grains of other colors are rich in calcium. Ciat-A-257, Bolinha, Iapar 81, Linea 29, and Roxo PV were found to be rich in protein (28.95%–30.40%). Additionally, 206 accessions from the Andean Diversity Panel were evaluated on cooking time, and five accessions (ADP0367, ADP0521, ADP0469, ADP0518, and ADP0452) were identified as promising in terms of shorter cooking time (Cichy et al., 2015b). Germplasm was also compared on nutritional composition and cooking characteristics with its closely related cultivated species, the tepary bean (*Phaseolus acutifolius*), in order to identify superior donors, as the latter species is highly tolerant to abiotic stresses (Porch et al., 2017). The results of this study indicated that there were no species-level differences on most nutritional parameters, with the exception of shorter cooking times for tepary bean accessions (Porch et al., 2017).

3.4 Pigeon pea

3.4.1 Agronomic traits

Pigeon pea is a legume of Indian origin (Ambika et al., 2022), and India remains its largest producer and consumer (Bohra et al., 2012). Pigeon pea is a resource-rich crop in terms of genetic and genomic resources, whole genome sequencing information, availability of trait-specific germplasm, genetic stocks, *etc.* The largest collection of pigeon pea germplasm is currently conserved at the ICRISAT gene bank (13,632 acc.), followed by the Indian National Gene Bank, ICAR–NBPGR (11,210 acc.); these collections are the major resources for trait identification and crop improvement. To enhance germplasm utilization, a set of 1,290 pigeon pea accessions (Reddy et al., 2005) has been developed, followed by a mini-core set of 146 accessions (Upadhyaya et al., 2006) and a composite core set of 1,000 accessions, plus a reference set of the most diverse 300 accessions (Upadhyaya et al., 2011a). An exceptionally good level of phenotypic variation has been observed for traits such as pods/plant, number of racemes, plant height, seed yield/plant, and days to maturity (Reddy et al., 2005). Promising accessions included in the pigeon pea composite core set have been listed for important economic traits, such as early flowering, a large number of pods/plant, seed weight, and yield/plant (Upadhyaya et al., 2011b). A vast amount of variability in flowering period has been observed, and a number of genotypes have been reported to show exceptionally short and long flowering durations. ICPL 90011 is reported to be an extra-short duration genotype with the lowest photoperiod sensitivity (Silim et al., 2007). Diverse trait-specific germplasms have been identified for use as potential sources for improvement programs (Upadhyaya et al., 2007c; Mir et al., 2014; Yohane et al., 2020). As of December 2022, 55 pigeon pea germplasms have been registered with the GRIS portal for a range of unique traits, including genetic male sterility (IC296750), cytoplasmic genetic male sterility (IC471860, IC471861, IC296590, IC296592, IC555904, *etc.*), cytoplasmic male sterility (IC296625, IC296623, *etc.*), fertility restoration (IC296805, IC296806, IC296807, *etc.*), early maturity (IC0587711, IC0587712), open flower (IC0573418, IC0573419, IC0573420), determinate growth habit (IC296589), and several other important traits.

3.4.2 Biotic stress

Pigeon pea production is adversely affected by many insects and diseases, such as wilt (*Fusarium udum* Butler), sterility mosaic virus (PPSMV) disease, phytophthora blight (*Phytophthora drechsleri* f. sp. *cajani*), Gram pod borer (*Helicoverpa armigera*), pod fly [*Melanagromyza obtusa* (Malloch)], and spotted pod borer [*Maruca vitrata* (Geyer)]. In one experiment, Saxena et al. (2002) evaluated 271 accessions under natural field conditions, and found that disease severity scores in pigeon pea germplasm ranged from 3 to 9. Screening against fusarium wilt and sterility mosaic disease (SMD) was carried out under artificial conditions for multiple seasons, resulting in the identification of several resistant accessions, viz., ICP 6739, ICP 8860, ICP 11015, ICP 13304, and ICP 14819 (Sharma et al., 2012). Accession ICPW 94 of the wild species *C. scarabaeoides* has been identified as resistant to all isolates of SMD, and is used in crossing programs (Hema et al., 2014). Earlier similar studies were also conducted using petiole grafting and artificial conditions in search of donors for SMD resistance in pigeon pea (Kulkarni et al., 2003; Sharma et al., 2015). Several sources of resistance for various biotic stresses in pigeon pea are listed in a review by Sultana et al. (2021) and in Table 4.

3.4.3 Abiotic stress

Pigeon pea is considered to be a drought-tolerant crop due to its deep root system and wide range in maturity period, which allows it to fit into a wide range of environments and cropping systems (Choudhary et al., 2011). Major abiotic stresses limiting pigeon pea productivity are waterlogging, drought, low temperatures (<10°C), and photoperiod sensitivity. Through several germplasm evaluation programs in pigeon pea, a number of popular donors have been identified; these are major sources for abiotic stress tolerance (Choudhary et al., 2018). In one study, 96 pigeon pea accessions were identified for early flowering; these are considered potential sources for the breeding of early-maturing pigeon pea varieties in order to avoid terminal drought and heat stress (Upadhyaya et al., 2011b). Sultana et al. (2013) screened 272 pigeon pea lines for waterlogging stress tolerance under laboratory and field conditions, and identified 12 lines tolerant to waterlogging. Similarly, in another study conducted under pot conditions for multiple seasons, 24 pigeon pea accessions were identified as waterlogging-tolerant (Krishnamurthy et al., 2012). Various other sources for resistance to important abiotic stresses are also listed in a review by Sultana et al. (2021) and in Table 5.

3.4.4 Nutritional quality

Pigeon pea contains approximately 86.6%–88.0% dry matter, 19.0%–21.7% crude protein, 1.2%–1.3% crude fat, 9.8%–13.0% crude fiber, and 3.9%–4.3% ash content (Amarteifio et al., 2002). Pigeon pea mineral content (mg/100 g dry matter) ranges are as follows: 1845–1941 K, 163–293 P, 120–167 Ca, 113–127 Mg, 11.3–12.0 Na, 7.2–8.2 Zn, 2.5–4.7 Fe, and 1.6–1.8 Cu. However, these values vary with genotype and across different studies (Talari and Shakappa, 2018). Biochemical evaluation of a total of 55 genotypes comprising advanced lines, improved cultivars, and landraces resulted in the identification of variation in four parameters: crude protein content (16.7%–28.4%), total phenol (21.9–84.4 mg/100 g), total flavonoid (16.4–33.4 mg/100 g), and total antioxidant activity (19.2–82.5 mg/100 g) (Cheboi et al., 2019).

3.5 Field pea

3.5.1 Agronomic traits

Field pea is cultivated in over 100 countries for fresh and dry grains and for fodder. Over 31,000 germplasm accessions of *Pisum* are conserved *ex situ* in various gene banks, including the Australian Grains Genebank, Australia; the Western Regional Plant Introduction Station, USDA, United States of America; the Leibniz Institute of Plant Genetics and Crop Plant Research, Germany; and the ICAR–NBPGR, New Delhi, India. Although a limited number of large-scale studies have been conducted on the agro-morphological characterization of field pea and for trait identification, several studies nevertheless indicate a substantial amount of phenotypic variability on qualitative as well as quantitative traits, such as days to 50% flowering, seed weight, plant height, and number of pods/plant (Azmat et al., 2011; Bhuvaneswari et al., 2017). The accessions IPF–5–19, EC 8495, HUDP–15, and DDR–30 have been found to show promise in terms of seed yield (Bhuvaneswari et al., 2017). Singh et al. (2010) evaluated 71 accessions on agronomic performance and seed and flower characteristics, identifying promising accessions in terms of early flowering (IC279013), early maturity (IC394017), a large number of pods/cluster (IC279195), longer pods (IC279013), pods/plant (IC219027), seed yield (IC279082), and seeds/pod (IC394028). Genotypes with five flowers per peduncle (VRPM–901–5) and three flowers per peduncle at multiple flowering nodes have been reported in garden pea, which could be highly useful in field pea improvement (Devi et al., 2018). Several unique trait-specific pea accessions have been registered in the GRIS portal, such as IC296677 (leafletless, dual purpose, and high-yielding), IC296678 (dwarf, leafletless), IC296737 (male sterile line governed by a single gene), IC279125 (bold seed with 50.14 g 100 seed weight), IC0610501 and IC630592 (≥ three pods/peduncle), IC636671 and IC640781 (extra-early flowering), and EC414478 (extended funicle).

3.5.2 Biotic stress

The major biotic stresses affecting field pea are powdery mildew (*Erysiphe pisi*), rust (*Uromyces viciae-fabae*), ascochyta blight (complex of *Ascochyta* spp.), white rot (*Sclerotinia sclerotiorum* (Lib) de Bary), wilt (*Fusarium oxysporum* f. sp. *pisi*), root rot (many pathogenic fungi), and collar rot (*Sclerotium rolfsii*). Screening against pea weevil (*Bruchus pisorum* L.) in 602 field pea lines, primarily from the Ethiopian Institute of Biodiversity (EIB), Addis Ababa, Ethiopia, resulted in the identification of four resistant lines: 32454, 235002, 226037, and 32410 (Teshome et al., 2015). Large-scale germplasm screening against powdery mildew disease under natural epiphytotic conditions has also been carried out, with the germplasm lines EC598655, EC598878, EC598704, IC278261, and IC218988 being identified as promising (Rana et al., 2013). Nisar et al. (2006) also reported three germplasm lines (Fallon, PS99102238, and PS0010128) to be highly resistant against powdery mildew.

3.5.3 Abiotic stress

Cold, frost, salinity, and heat stresses are the major sources of abiotic stress in field pea crop production. Many studies have taken up the aim of developing lines tolerant to abiotic stresses. In one such study, five field pea germplasm (ATC 104, ATC 377, ATC 968, ATC 3992, and ATC 4204) were identified as frost-tolerant at the reproductive stage through screening of 84 accessions under controlled environmental conditions (Shafiq et al., 2012). Screening of 3,672 pea germplasm lines under field conditions led to the

identification of 214 cold-tolerant lines (Zhang et al., 2016). Additionally, 780 accessions were screened for salinity stress tolerance under artificial conditions (Leonforte et al., 2013). Finally, in a recent study, IPFD 11–5, Pant P–72, P–1544–1, and HUDP 11 were identified as heat-tolerant lines based on evaluation under timely- and late-sown field conditions (Lamichaney et al., 2021).

3.5.4 Nutritional quality

Field peas in general have lower protein content (~25%), very low fat content (~0.1%), and very high carbohydrate content (~70%). Major yield-attributing traits in field pea are pods/plant, number of grains/pod, and seed weight. In one study, 94 pea genotypes were examined for pea carotenoid content; higher carotenoid content (10–27 μg/g) was observed in accessions with green cotyledons, and comparatively low carotenoid content (5–17 μg/g) in accessions with yellow cotyledons (Ashokkumar et al., 2015). Pea grains have comparatively higher antioxidant activity than chickpeas. Promising field pea accessions have also been identified in terms of mineral content, such as Zn (IG52442, IG134828), Cu (IG116297, IG52442) and Ca (IG51520, IG52442), by Costantini et al. (2021). Additionally; Singh et al. (2010) have identified lines with shorter cooking time (IC260344) and observed that the genotypes that absorb more water and swell more during soaking require less cooking time. The authors have also identified IC320964 as superior in terms of ash content (3.73%), and several other accessions as promising in terms of their physicochemical properties. In a nutritional analysis of 96 accessions from diverse collections at the USDA National Germplasm Center, Pullman, WA, a wide range of variation was observed in mineral micronutrient content (Hacisalihoglu et al., 2021). An atypical morphotype having extended funicle (EC0414478) was identified in pea germplasm, and this accession was found to be associated with faster water uptake in comparison to the checks included (Tripathi et al., 2021b); this is likely to help with the development of pea cultivars with shorter cooking times.

3.6 Cowpea

3.6.1 Agronomic traits

Cowpea is a multi-purpose grain legume (yielding grains, green pods, and leaves) and is widely cultivated in Asia, Africa, and America. It is considered to be one of the best suited crops for hotter, semi-arid agro-climatic conditions, as it requires less water and also grows well in sandy soils. The germplasm conserved in various gene banks has exhibited a good amount of genetic variability, which enables it to grow in various agro-climatic regions and in various soil types. To enhance the utilization of cowpea germplasm, over 12,000 accessions of cowpea were characterized on 28 agro–botanical descriptors at the International Institute of Tropical Agriculture (IITA), Ibadan, and a core set of 2,062 accessions was developed (Mahalakshmi et al., 2007). In another study, 4,000 accessions were characterized in multi-location trials by the ICAR-National Bureau of Plant Genetic Resources (unpublished records). A great deal of variability was observed in plant and seed morphology. Gerrano et al. (2015) identified germplasm lines having desirable grain yield characteristics, such as Fahari, IT93K129-4, Glenda, and vegetable cowpea dakama cream; Nkhoma et al. (2020) identified lines Bubebe, CP411, CP421, CP645, Chimponogo, and MS1–8–1-4 as high-yielding and genetically divergent among 90 genotypes studied, making them ideal parental lines. Cowpea genotypes IT96D-604, 93K-619-1, IT97K-569-9, and IT99K-1060 have also been identified as high-yielding (Goa et al., 2022).

3.6.2 Biotic stress

The major diseases affecting cowpea are cowpea mosaic virus (CpMV) disease, *Cercospora* leaf spot (CLS), brown blotch (*Colletotrichum capsici*), and bacterial blight (*Xanthomonas axonopodis* pv. *vignicola*), while the major pests are pod borer, aphids, thrips and bruchids. The severity of these biotic factors varies with agro-climatic zone and growing conditions. Although cowpea is one of the more prominent legume crops and the largest of the Vigna group, it has not received commensurate research attention. As a result, cowpea improvement has suffered from a lack of reliable donors for resistance to many of these biotic factors. Nonetheless, efforts have recently been undertaken in this direction, and several important and promising donors for resistance to a small number of these biotic stresses have been identified; these are listed in Table 4. In a study that aimed to identify resistant donors for aphid (*Aphis craccivora*), cultivated germplasm (105 accessions) and wild germplasm (92 accessions) were screened under greenhouse conditions; only a single accession (TVNu 1,158) was identified as a resistant line (Souleymane et al., 2013). The findings of this study also indicated that both the cultivated and the wild relatives of this crop have poor genetic bases. Boukar et al. (2019) identified 14 lines having resistance to bacterial blight (*Xanthomonas axonopolis* pv. *vignicola*) under artificial inoculation. In another study, 225 germplasm lines were screened against CpMV, CLS, and cowpea rust (*Uromyces vignae*), resulting in the identification of promising accessions for resistance to these pathogens (Deshpand et al., 2010; Table 4). Finally, Tripathi et al. (2020) identified EC528425 and EC528387 as tolerant to bruchid (*Callosobruchus maculatus*) through the screening of 103 cowpea lines using a 'no-choice' test method.

3.6.3 Abiotic stress

The major abiotic stresses are drought, heat stresses, and poor soil fertility, especially in sub-Saharan Africa (SSA), where cowpea is grown as a major crop, as well as soil salinity in almost all irrigated areas worldwide (Horn and Shimelis, 2020). Several studies have been conducted to identify resistant donors (see Table 5). Five lines with superior drought stress tolerance (viz., TVu1436, TVu9693, TVu12115, TVu14632, and TVu15055) have been identified using the water withdrawal method under field conditions (Fatokun et al., 2012), while Dagupan Pangasinan, UCR 369, and Negro have been identified as tolerant to waterlogging at the seedling stage (Olorunwa et al., 2022). Accessions EC472250, EC472267, EC472285, EC472286, EC472289, and Pusa Komal have been identified as tolerant to heat stress through screening in multiple seasons under field conditions (Mishra et al., 2005). Accessions PI582422, 09–529, PI293584, and PI582570 have been identified as tolerant to salinity stress under artificial screening conditions through imposition of salinity stress (150 mM NaCl) at the seed germination stage (Ravelombola et al., 2017). Other similar studies have also identified lines tolerant to salt stress using different NaCl concentrations (150 mM and 120 mM) at the germination and seedling stages (Ravelombola et al., 2018; Dong et al., 2019). Based on screening of 155 cowpea lines in 200 mM NaCl, several promising lines were identified as salt tolerant, i.e., PI354686, PI353270,

PI354666, PI354842 PI548785, PI582466, PI339599, and 09-697 (Dong et al., 2019).

3.6.4 Nutritional quality

Cowpea is a major source of nutrition in sub-Saharan Africa, Asia, and Latin America. Based on nutritional profiling of 100 breeding lines on a dry weight basis, a significant range of variation has been observed in terms of protein content (22.9%–32.5%), ash content (2.9%–3.9%), fat content (1.4%–2.7%), and carbohydrate content (59.7%–71.6%; Nielsen et al., 1993). Genotypes also vary in 50% cooking time, which ranges from 21.1 to 61.9 min, and promising donors have been identified, such as IT83S–872 for protein content, IT84S–2085 and IT86D–466 for ash content, and IT85F–2805 for shortest cooking time (Nielsen et al., 1993). In a study aiming to investigate nutritional variability in immature cowpea pods, 22 genotypes were analyzed on various nutritional composition parameters; genotypes such as ITOOK-1060, TVU-14196, and 98K–5301 were found to be superior on such parameters as Mg, Na, Mn, Boron, Al, Zn, Cu, K, P, and protein (Gerrano et al., 2017). The fresh young leaves of cowpea are also consumed in several countries; therefore, analyses have been conducted of the nutritional composition of 15 varieties and the sensory attributes of 10 varieties (Ahenkora et al., 1998). In this study, nutrient concentration in cowpea leaves on a dry weight basis was found to range from 303.8 to 468.9 mg/100 g for phosphorus, from 33.5 to 148.0 mg/100 g for ascorbic acid, and from 27.1% to 34.7% for protein.

3.7 Black gram

3.7.1 Agronomic traits

Black gram is a grain legume of Indian origin, primarily cultivated in South Asian regions. Although black gram is an important legume, its productivity level is very poor compared to that of other legumes, which can mainly be attributed to a lack of good plant ideotypes and resistance sources for major diseases in its cultivated gene pool (Kumar and Singh, 2014; Shanthi et al., 2019; Subramaniyan et al., 2022). Therefore, to identify donors for desired agro-morphological traits, 484 accessions have been characterized on qualitative and quantitative traits; a good deal of variation was observed in flowering and maturity period and in yield (Ghafoor et al., 2001). Recently, 840 accessions of black gram were also characterized, resulting in the identification of promising germplasm lines in terms of early flowering (IC343936, IC436615), synchronous flowering (IC73523, IC396032, IC485444), pod length (IC438379), number of seeds/pod (IC472051_2, IC565238), and seed weight (IC485605_2, IC485588) (Gayacharan et al., 2022). For novel trait generation in black gram, gamma-irradiated mutants were generated using black gram cultivars ADT 3, Co 6, and TU 17-9, which have exhibited high plant yield (Dhasarathan et al., 2021). Additionally, RBU1012 and Pant U-19 have been found to be the most stable genotypes in terms of yield when evaluated under field conditions (Singh N. P et al., 2016). The GRIS portal indicates that unique germplasms of black gram such as IC296878 (dwarf with ground pod bearing habit), IC553269 (brown pods with yellow seeds), IC594172 (male sterile flowers with protruded stigma and crumpled petals), IC594173 (sympodial pod-bearing habit), IC426765 (photosensitive), and IC636672 (extra-early maturing) have been registered for important traits.

3.7.2 Biotic stress

Urdbean leaf crinkle disease (ULCD) and mungbean yellow mosaic disease (MYMD) are the two major diseases of the black gram crop. Yield losses may reach or exceed 60%, depending on the susceptibility of the host plant, if the crop is affected in its early vegetative stage. Nevertheless, unlike green gram, black gram has a high level of resistance against MYMD in its cultivated gene pool, as revealed in a large-scale evaluation under field and artificial conditions conducted during 2019 and 2020 (unpublished records). Several black gram sources of MYMD resistance, identified on the basis of field screening, are highlighted in a review published by Mishra et al. (2020a). Urdbean leaf crinkle virus (ULCV) disease has spread across all the cropping systems in India, and yield losses can reach 100% if the disease outbreak occurs at the early growth stage under favorable weather and host genotype conditions (Biswas et al., 2009). Resistance sources for ULCV have been reported by several researchers (Ashfaq et al., 2007; Gautam et al., 2016); several such sources for this and other diseases are listed in Table 4. In the GRIS database, accessions IC0570267, IC0570268, IC0570269, IC11613, IC636672, IC0144901, and IC485638 are registered as MYMD resistant, and IC0585931 as bruchid resistant. Powdery mildew (*Erysiphe polygoni*) and *Cercospora* leaf spot (*Cercospora canescens*) are the other major diseases of black gram. The major pests affecting this crop include spotted pod borer (*Maruca testulalis* r), whitefly (*Bemicia tabaci*), bruchids (*Callosobruchus chinensis*. and *C. maculatus*), and nematodes (*Meloodogyne incognita*, *M. javanica*, and *Heterodera cajani*), for which reliable sources of resistance are lacking. Bruchids begin infesting the crop during the pod maturity stage, and they are the cause of up to 90% of produce losses (Soundararajan et al., 2012).

3.7.3 Abiotic stress

The crop is grown in a rainfed environment under tropical and sub-tropical climatic conditions. Therefore, terminal drought and heat, as well as waterlogging, are the major constraints on black gram production. Salinity is another problem, particularly in arid and semi-arid regions. Only a small number of studies have examined the potential for improvement of black gram in terms of resistance to abiotic stresses. Saline-tolerant lines, such as BARI Mash-1 (Hasan et al., 2017), VNBG 017, AUB 3, and AUB 20 (Priyadharshini et al., 2019), have been identified as promising under artificial screening conditions. Under natural waterlogging conditions during a germplasm characterization program, a small number of germplasm lines have been identified as tolerant; these were further evaluated under artificial waterlogging conditions, and accessions IC530491 and IC519330 were found to be tolerant to waterlogging (Bansal et al., 2019). In another study, 26 genotypes were analyzed under waterlogging stress; a large amount of variation was observed in various quantitative traits, and BU Acc 25, BU Acc 17, and BU Acc 24 were identified as the strongest performers in terms of yield (Rana et al., 2019). In terms of drought stress tolerance, cultivars VBN4 and K1 have been identified as promising based on protein and biochemical analyses (Sai and Chidambaranathan, 2019).

3.7.4 Nutritional quality

Black gram grains are a rich source of protein (22%–26%) and moderately high in calories (ca. 350 cal/100 g), carbohydrates (ca. 56.6%), and fat (1.1%–1.2%) (Panhwar, 2005; Suneja et al., 2011). They also contain vitamins, viz., Vit. B1 (0.42 mg/100 g), Vit. B2 (0.37 mg/100 g), Niacin (2 mg/100 g), and minerals, viz., Ca (185 mg/100 g), Fe

(8.7 mg/100 g), and P (345 mg/100 g) (Panhwar, 2005). However, a limited amount of germplasm has been nutritionally profiled for the identification of nutrient-rich lines. Black gram is reported to exhibit a substantial amount of variation in nutrient content between the whole grain and its milled fraction (Girish et al., 2012). A small number of genotypes among 26 investigated, such as Shekhar 2, have been found to have high Fe and Zn content, and genotypes Yakubpur, PU 31, IPU 99–200, and PDU 1 have been found to have high polyphenol content (Singh J et al., 2017). There is a need for large-scale nutritional profiling to develop an understanding of nutritional variability in the germplasm and to identify superior genotypes with minimal levels of anti-nutritional factors to enhance the palatability of the crop.

3.8 Green gram

3.8.1 Agronomic traits

Green gram is a highly nutritious and palatable grain legume cultivated in Asia, primarily for its grains. Green gram cultivation faces constraints such as a narrow genetic base in the cultivated gene pool, a lack of ideal plant type, and many biotic and abiotic stresses. In order to identify new donors, 1,532 *ex situ* collections of green gram conserved in the Indian National Gene Bank were characterized, potential donors for certain agro-morphological traits were identified, and a core set of 152 accessions was also developed (Bisht et al., 1998). A good level of variation was observed in branch length, nodulation, number of pods bearing a peduncle, number of pods per plant, and yield per plant. The World Vegetable Center, Taiwan, holds over 6,700 accessions of green gram, which have been utilized for the development of a core set of 1,481 accessions based on geographic stratification and clustering of genotypes on eight phenotypic traits (Schafleitner et al., 2015). This core set was genotyped using 20 microsatellite markers, and a mini-core set of 289 accessions was developed; this is now extensively utilized for trait identification. In another large-scale characterization and preliminary evaluation of green gram germplasm, promising germplasm lines were identified in terms of early flowering (EC398944, EC398883), synchronous flowering (EC396115, IC76414, and IC488968), greater pod length (EC398937), seed weight (EC398903, EC398884, and EC396413), *etc.* (Gayacharan K. et al., 2020). Recently, the entire green gram *ex situ* collection of the Indian National Gene Bank has been characterized in multi-location trials, and a diverse core set of 400 accessions has been developed (unpublished records). Photoperiod-insensitive genotypes (EC 318985–319057) have also been identified in green gram (Pratap et al., 2014; Pratap et al., 2019). The GRIS portal lists a number of accessions with unique traits, such as a photosensitive nature (IC546478), high seed weight (IC418452), early maturity (IC0589309, IC589310, IC39289, and IC639796), and penta-foliate leaves (IC296679).

3.8.2 Biotic stress

Green gram production is affected by biotic stresses such as yellow mosaic disease (YMD), pod borers, and storage pests. YMD is a comparatively new disease in green gram and is spreading rapidly into new areas, which is a cause for concern. In YMD-susceptible genotypes of green gram, yield losses up to 85% are reported (Karthikeyan et al., 2014), but it has been observed that losses may reach 100% if the crop is infected at seedling stage. Resistance sources are lacking in the entire cultivated gene pool of the crop, as revealed in a field screening of 4,100 accessions at New Delhi (a YMD hotspot location). However, variability in the severity of the disease is observed according to multiple factors, such as genotypic constitution, vector population load, weather conditions, presence of multiple virus strains, *etc.* (unpublished record). Similar reports have also made by other researchers based on germplasm screening (Shad et al., 2006). Several resistant sources for YMD are listed in a review (Mishra et al., 2020a). There are also several reports of YMD resistance in green gram under field conditions (Iqbal et al., 2011; Mohan et al., 2014; Nainu and Murugan, 2020). Aside from YMD, Duraimurugan et al. (2014) identified four lines (viz. LM 131, V 1123, LM 371, and STY 2633) as resistant against bruchid beetle based on a 'free choice' and 'no choice' test method. Spotted pod borer (*Maruca vitrata*) also causes severe damage to the crop, if not controlled at the appropriate stage of crop growth, and there are no resistant sources available for this pest. Sandhya et al. (2014) have reported that KM–9–128, KM–9–136, RMG–492, LGG–5, and LGG–538 are tolerant to *Maruca vitrata* following field screening of 110 genotypes.

3.8.3 Abiotic stress

Green gram is primarily grown under rainfed conditions; thus, abiotic stresses such as drought, waterlogging, heat, and salinity affect crop production (Singh and Singh, 2011). In general, reliable tolerant donors for these abiotic stresses are lacking in this crop. Forty-one elite lines were screened for heat stress tolerance under late-sown conditions; of these, five lines (viz., EC693357, EC693358, EC693369, Harsha, and ML1299) showed heat stress tolerance (Sharma et al., 2016). Additionally, IC333090 and IC507340 were found to be drought tolerant, out of 100 lines screened under hydroponics conditions (Meena et al., 2021). Mung bean lines OBGG-2013-9 and OBGG-2013-14 have also been reported to exhibit cold tolerance (Kabi et al., 2017).

3.8.4 Nutritional quality

Green gram is nutrient-rich and possesses various health benefits, such as antioxidant, anti-cancer, anti-inflammatory and hypolipidemic activity (Sudhakaran and Bukkan, 2021). Because of its high nutritional value, green gram is regarded as "green pearl" (Nair et al., 2013). It contains approximately 19.7%–29.1% protein (Bartwal et al., 2022), 61%–63% carbohydrates, 1.1%–2.3% fat, 3.2%–4.2% ash, 0.03–0.06 g Fe kg^{-1}, and 0.02–0.04 g Zn kg^{-1} (Nair et al., 2013; Sudhakaran and Bukkan, 2021). The nutritional composition of green gram and black gram is very similar, but green gram is reported to have higher moisture, fat, and protein content (Shaheen et al., 2012). A small number of accessions with particularly high nutritional value are listed in the GRIS portal; these could potentially function as donors for nutritional improvement of the crop. Specifically, accessions IC296771 (27.8%) and IC573456 (25.8%) are registered for high protein content, and IC573449, IC573450, IC573451, IC573453, and IC573454 for high Fe and Zn content.

3.9 Horse gram

3.9.1 Agronomic traits

Horse gram is one of the least utilized and least studied legumes. The crop is known for its nutritional and therapeutic value, and is

primarily cultivated in hill states and dry areas of southern India. This crop has failed to attract the attention of breeders and researchers due to a lack of ideal ideotypes and morphological variation. A small number of characterization and evaluation studies have been conducted, indicating comparatively wide variation in terms of plant height, pod length, seed test weight, and pods per plant (Singh et al., 2019). Additionally, genotypes CRHG-6 and CRHG-8 are of the non-shattering type, which has been developed through mutation breeding (Salini et al., 2014). In a characterization study examining seven qualitative and quantitative traits in 66 horse gram genotypes, a good amount of variability was observed for pod length and pods per plant (Gomashe et al., 2018). Priyanka et al. (2021) studied 12 quantitative traits across 252 horse gram genotypes, reporting that the highest yield was 65.61 g per plant. Another characterization and evaluation study of 51 accessions led to the identification of several promising accessions (viz., S44/L23, S56/L29, S8/L4, S96/L49, and S29/L14) in terms of early flowering and maturity (Kaundal and Kumar, 2021). In the GRIS portal, only one accession (IC587788) is registered for high fodder yield.

3.9.2 Biotic stress

Horse gram is susceptible to various biotic stresses and still lacks resistant donors for use in crop improvement programs. Only a small number of studies have been carried out to identify resistant donors for selected diseases and pests, such as YMD, wilt, anthracnose, and storage pests. Parimala et al. (2011) identified horse gram accessions AK-38, HG-14, HG-52, HG-59, HG-63, HG-75 as having resistance against horse gram YMD, and AK-38 and HG-46 as resistant to powdery mildew disease. In another study, accessions HG 63, HG 58, HG 50, and Palem 2 were identified as resistant to wilt disease (Durg, 2012). Accession IC470275 has also been identified as resistant to anthracnose disease (Colletotrichum dematium) (Sankar et al., 2015). Finally, horse gram lines Palem-1, Palem-2, AK-21, and NSB-27 have been identified as resistant against Callosobruchus chinensis, a storage pest (Divya et al., 2012; Divya et al., 2013). Accession IC587786 is registered on the GRIS portal as resistant to anthracnose disease.

3.9.3 Abiotic stress

Horse gram germplasm have been screened against abiotic stresses, such as drought, salinity, moisture, and heavy metal stress. Several germplasm lines, such as M-249 and HPK-4, have shown resistance against drought stress (Bhardwaj and Yadav, 2021). Yasin et al. (2014) identified a line tolerant to moisture stress, namely D13. The variety Paiyur-2 was found to have high proline and glycine betaine content and lower lipid peroxidation under salinity stress (Kanagaraj and Sathish, 2017). This genotype was further tested for antioxidant activity, and was found to exhibit enhanced antioxidant activity under salinity stress (Desingh and Kanagaraj, 2019). Separately, the same Paiyur-2 variety was found to be promising for salinity tolerance, and heavy metal tolerance was observed in Madhu (for chromium) and in HGR-4 (for nickel; Dhali et al., 2021; Edulamudi et al., 2021). Based on a screening of 88 germplasm lines for biochemical parameters, accessions TCR491, IC110286, IC56145, and IC53641 were identified as suitable for environments imposing drought stress (Sharma and Chahota, 2022).

3.9.4 Nutritional quality

Horse gram is used as a food and fodder crop and is known for its medicinal and therapeutic uses. It provides protein (17.9%–25.3%), carbohydrates (51.9%–60.9%), lipids (0.58%–2.06%), and vitamins, such as riboflavin, niacin, and vitamin C (Jha et al., 2022). The protein content of horse gram is relatively high compared to that of green gram, black gram, dry peas, the kidney bean, chickpea, or pigeon pea (Longvah et al., 2017). As consumers have become more health conscious, consumption of sprouted horse gram seeds has increased. In the seeds, albumin–globulin is a major contributor (~75.27–78.76%) to the total protein content. The seeds are low in fatty acid content but rich in dietary fiber, required for proper functioning of the lower intestine (Kawale et al., 2005). Horse gram seeds also exhibit antioxidant activity and radical scavenging activity. In a recent study, metabolic profiling was conducted for 96 accessions of horse gram, which were derived from 700 accessions spread across the entirety of India (Gautam and Chahota, 2022). Tremendous variability in protein content was observed, with the lowest protein content (13%) being found in IC120837 and TCR-1439, while a related wild species (*Macrotyloma sar-gharwalensis*) had the highest protein content (40%). Accessions IC280031 and IC139356 were found to be most nutritive, as the largest number of metabolites (44) was observed for these among the 96 lines selected in an analysis using ^{1}H NMR spectroscopy (Gautam and Chahota, 2022). An earlier study had also identified the species *Macrotyloma sar-gharwalensis* (IC212722) as containing 34.88% protein (Yadav et al., 2004).

3.10 Moth bean

3.10.1 Agronomic traits

Moth bean is considered to be a hardy crop suitable for hot, arid regions. In addition to this, it also helps to reduce soil erosion, particularly in sandy deserts, due to its extensive root system and profuse foliage cover. Most local cultivars continue to have wild traits, such as pod-shattering, a trailing and spreading growth habit, asynchronous maturity, and a photo-sensitive nature. Few studies have been conducted to explore the existing variability in the gene pool. However, a good amount of phenotypic variability in moth bean was reported by Chaudhary et al. (2021) in an evaluation of 40 genotypes on 10 morphological traits. In another study, accessions IC 36607, IC 39675, IC 251908, IC 36563, and IC 36245 exhibited higher seed yield as compared to checks (Vir and Singh, 2015). Similarly, 50 genotypes of moth bean were phenotyped on 12 quantitative traits and exhibited high levels of variability (Sahoo et al., 2019). Additionally, Singh S et al. (2017) have developed a variety (RMO 257) with superior agronomic traits, such as plant height, dry matter accumulation, and seed yield. A small number of moth bean genotypes are registered in the GRIS portal for early maturity (IC432859 and IC120963) and for single stem formation (IC432859).

3.10.2 Biotic stress

The major biotic stresses are YMD, leaf crinkle disease (LCD), bacterial leaf spot (*Xanthomonas phaseoli*), Cercospora leaf spot (*Cercospora dolichi*), charcoal rot (*Macrophomina phaseolina* Tassi), pod borer, and bruchids. Accessions IC36522 and IC36217 have been identified as YMD resistant as a result of screening under field conditions in multiple seasons (Singh et al., 2020). Meghwal et al. (2015) screened 204 germplasm lines and identified 14 accessions resistant to YMD. Resistance to leaf crinkle virus and *Cercospora* leaf spot has also been reported in the crop

(Singh et al., 2020). Vir and Singh (2015) identified LCD resistance in moth bean under field conditions in multiple seasons.

3.10.3 Abiotic stress

Moth bean is one of the best-suited crops for arid and semi-arid environments, and is highly tolerant to drought and heat stress. Although only a small number of moth bean accessions have been investigated, a good amount of variation in the germplasm has been revealed in terms of resistance to drought and heat stress; this could be exploited for crop improvement in order to sustain its productivity amid climate change (Vir and Singh, 2015; Pal et al., 2020). Accessions such as IC129177, IC103016, IC415139, IC 415155, IC36157, Maru moth, and Jadia, which have been identified as tolerant to drought stress, can serve as donors for crop improvement (Malambane and Bhatt, 2014). Additionally, in a separate study, lines IC103016, IC36011, and IC36157 have been identified as promising for drought tolerance (Sachdeva et al., 2016).

4 Role of genomics in enhancing grain legume germplasm utilization and attaining higher genetic gains

The first step toward enhancing the utilization of grain legume crop germplasm accessions for trait discovery and subsequent genetic improvement requires thorough and extensive genotypic and phenotypic characterization of such accessions using large-scale data (Rasheed et al., 2017). The numerous germplasm resources (including landraces, wild accessions, cultivated varieties, and breeding lines) available for diverse grain legume crop species, representing diverse agro-climatic regions of the world, have been stored efficiently at various national and international gene banks and repository centers. Considering the difficulties involved in genotypic and phenotypic characterization of this vast set of available germplasm resources, efforts are currently being made to develop core and mini-core collections in the case of several legumes by identifying the greatest amount of genetic diversity that can be represented with a minimal number of accessions (Table 3). This is where the crucial role of genomics comes into play, especially as a means of producing realistic estimates of the level of molecular diversity existing among germplasm accessions, which enables efficient screening of unduplicated authentic accessions in the process of constructing core collections of grain legumes.

Tremendous technological advances made over the last decade in sequencing and other high-throughput sequence- and array-based genotyping assays have supplied much-needed momentum to germplasm characterization. Draft and reference whole genomes, resequencing information, and global transcriptome information have now been decoded for many important grain legume crop plants using first-generation Sanger sequencing and next-generation sequencing (NGS)-based second-generation short read and third-generation long read sequencing assays; the results of these are now publicly accessible (Table 6; Michael and Jackson, 2013; http://www.embl.de; http://www.ebi.ac.uk; http://www.ddbj.nig.ac.jp; http://www.phytozome.org). Grain legumes were once considered to be resource-poor crops, but recent efforts by national and international organizations has altered this trend. As a result, a vast amount of genome sequence information, including whole genome sequences, is available in public databases (Table 6). This sequence information has since been used to understand the genomic features and evolutionary characteristics of the crops in question, and also to develop a vast range of genomic resources, including molecular markers (Garg et al., 2022). Among several sequence-based molecular markers that have been made available, simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs) have occupied a central position, finding extensive use in allelic diversity screening and other genomics-assisted crop improvement programs due to their genome-wide distribution, multi/bi-allelic nature, and amenability to high-throughput detection and genotyping assays (Lateef, 2015). The availability of several high-throughput genotyping platforms and the rapid evolution of the chemical techniques that they employ has enhanced the precision and pace of large-scale mining and genotyping of SSR and SNP markers (Kujur et al., 2015). High-throughput genotyping of SSR and SNP markers in a larger set of germplasm accessions and core or mini-core collections of grain legumes has been expedited *via* the use of various array-based and NGS assays, especially automated fragment analyzer (ABI3730xl automated DNA sequencer), Illumina GoldenGate, and Infinium assays; the Fluidigm dynamic array; KASP (KBioScience Allele-Specific Polymorphism) profiling; MALDI–TOF; the Affymetrix GeneTitan SNP Chip array; and Genotyping-By-Sequencing (GBS) assay (Varshney et al., 2013a; Jain et al., 2013). Among these, the MALDI–TOF, Illumina GoldenGate, and Infinium assays, SNP Chip Array, and KASP profiling have come to be considered highly advantageous and are utilized widely for high-throughput genotyping of previously mined SNP markers in many crop plants (Gaur et al., 2012; Hiremath et al., 2012). In particular, GBS assay has been extensively utilized for simultaneous genome-wide discovery and genotyping of SNPs in diverse plant species (Sonah et al., 2013; Spindel et al., 2013). Its development has thus expedited the mining of novel functional allelic variants and their large-scale validation and genotyping at the whole genome level for efficient trait association mapping of diverse small- and large-genome grain legume crop plants.

Using the aforementioned high-throughput marker-based genotyping strategies, along with large-scale multi-environment phenotyping information, sets of 211, 146, 184, and 289 germplasm accessions belonging to core or mini-core collections of chickpea, pigeon pea, groundnut, and green gram, respectively, have been developed. These have been collated based on the 16,991, 13,632, 15,490, and 6700 accessions available for these respective crop species as a result of the efforts of international institutes such as the IRRI, ICRISAT, USDA, and the World Vegetable Center (Upadhyaya et al., 2001; 2002; Zhang et al., 2011; Schafleitner et al., 2015). These core or mini-core germplasm resources, readily available for many grain legume crop plants, are the primary sources of trait discovery once these collections have been thoroughly characterized genotypically and phenotypically for diverse traits of agronomic importance, including yield, (a)biotic stress tolerance, and nutritional quality traits. Under this perspective, candidate gene-based association mapping and genome-wide association studies (GWAS) relying on the large-scale genotyping of informative SNP markers and robust field phenotyping information on these core or mini-core germplasm lines (i.e., association panels) are now considered to be an effective approach for the identification of major and minor genes/QTLs and alleles regulating both simple qualitative and complex quantitative traits in grain legume crop plants (Varshney et al., 2011). Candidate gene-based association mapping, which is carried out using genotyping information from SNPs in various coding and regulatory sequence components of

TABLE 6 Whole genome sequence information available for grain legume crops.

Crop	Project ID	Cultivar	Assembly level	Assembly size (Mb)	Scaffold N50 (Mb)	Sequencing chemistry	Genome coverage %	No. of predicted protein-coding genes	References
Chickpea (*Cicer arietinum*)	ASM33114v1	CDC Frontier	Chromosome	532.29	39.99	Illumina Hiseq 2000	73.8	28,269	Varshney et al. (2013b)
Chickpea (*Cicer arietinum*)	ASM34727v4	ICC4958	Chromosome	511.02	39.90	454; Illumina GAIIx	94	30,257	Parween et al. (2015)
Chickpea (*Cicer reticulatum*)	ASM368901v2	PI489777	Chromosome	416.9	39.84	Illumina HiSeq	78	25,680	Gupta et al. (2017)
Cowpea (*Vigna unguiculata*)	ASM411807v1	IT97K-499-35	Chromosome	519.43	1.64	PacBio; Bionano	91	29,773	Lonardi et al. (2019)
Asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*)	ASM395868v2	Xiabao II	Chromosome	632.8	2.7	Illumina HiSeq 4000	342	42,609	Xia et al. (2019)
Green gram (*Vigna radiata*)	Vradiata_ver6	VC 1973A	Chromosome	431	1.52	Illumina HiSeq2000	80	22,427	Kang et al. (2014)
Green gram (*Vigna radiata*)	SRRS9994113	VC 1973A	Chromosome	476	5.2	PacBio RS II	87.1	30,958	Ha et al. (2021)
Black gram (*Vigna mungo*)	ASM1909614v1	Pant U-31	Scaffold	474	1.42	Illumina HiSeq; Oxford Nanopore GridION	82	42,115	Jegadeesan et al. (2021)
Black gram (*Vigna mungo*)	ASM1342719v1	Chai Nat 80	Chromosome	499	43.17	Illumina HiSeq	92	29,411	Pootakham et al. (2021)
Rice bean (*Vigna umbellata*)	ASM1883591v1	Himshakti	Scaffold	414	0.08	Illumina; PacBio	-	31,276	Kaul et al. (2019)
Pigeon pea (*Cajanus cajan*)	PRJNA72815	ICPL 87119 (Asha)	Chromosome	605.78	0.52	Illumina Hiseq 2000; Sanger sequencing	72.7	46,750	Varshney et al. (2011)
Pigeon pea (*Cajanus cajan*)	PRJNA68667	ICPL 87119 (Asha)	Contig	510.8	-	FLX 454; Illumina HiSeq	75.6	47,004	Singh et al. (2012)
Adzuki bean (*Vigna angularis*)	PRJNA261643	Jingnong 6	Chromosome	450	1.29	Illumina HiSeq 2000	83	34,183	Yang et al. (2015)
Adzuki bean (*Vigna angularis*)	PRJDB3778	Shumari	Chromosome	462	3.0	PacBio RSII; Illumina HiSeq2500	85.6	30,507	Sakai et al. (2015)
Adzuki bean (*Vigna angularis*)	PRJNA253346	Kyungwonpat	Chromosome	443	0.7	Illumina Roche	75	26,857	Kang et al. (2015)
Pea (*Pisum sativum*)	PRJEB31320	Caméor	Chromosome	3920	0.41	Illumina	88	44,756	Kreplak et al. (2019)
Common bean (*Phaseolus vulgaris*)	PRJNA41439	G19833	Chromosome	472.5	50.3	ABI 3730; 454 FLX; Illumina GAII	98	27,197	Schmutz et al. (2014)
Common bean (*Phaseolus vulgaris*)	PRJNA221782	BAT93	Chromosome	458.2	0.43	454; SOLiD; Sanger	81	30,491	Vlasova et al., 2016
Lima bean (*Phaseolus lunatus*)	PRJNA596114	G27455	Chromosome	541.5	47.8	PacBio Sequel; Illumina HiSeq		28,326	Garcia et al. (2021)
Lima bean (*Phaseolus lunatus*)	PRJNA647124	Bridgeton-DES4	Scaffold	597.4	2.9	Illumina HiSeq	91	64,541	

(Continued on following page)

10.3389/fgene.2023.996828

TABLE 6 (*Continued*) Whole genome sequence information available for grain legume crops.

Crop	Project ID	Cultivar	Assembly level	Assembly size (Mb)	Scaffold N50 (Mb)	Sequencing chemistry	Genome coverage %	No. of predicted protein-coding genes	References
Grass pea (*Lathyrus sativus*)	PRJEB33571	LS007	Scaffold	6200.8	0.06	Illumina; Oxford Nanopore	89.8	33,819	Wisser et al. (2021)
Horse gram (*Macrotyloma uniflorum*)	PRJDB5374	HPK4	Scaffold	259.2	2.8	Illumina HiSeq 2000; Illumina MiSeq	89	36,105	Emmrich et al. (2020)
Horse gram (*Macrotyloma uniflorum*)	PRJNA400556	PHG-9	Scaffold	279.12	1.1	Illumina HiSeq; PacBio	83.53	24,521	Shirasawa et al. (2021)

Source: NCBI database (https://www.ncbi.nlm.nih.gov/).

genes in a trait-specific association panel, plays a significant role in the identification of genes/QTLs controlling yield, nutritional quality, and stress tolerance traits in grain legume crops. With the availability of high-throughput genome-wide SSR and SNP marker-based genotyping information on germplasm lines belonging to an association panel, the GWAS has now become a routine approach for high-resolution scanning of the whole genome to identify target genomic regions, including genes/QTLs (both major and minor) associated with traits of agricultural importance in many grain legume crops (Varshney et al., 2017; Varshney et al., 2021). The trait-influencing molecular signatures once identified using trait association mapping are significant for their potential utilization for genomics- (marker-)assisted crop improvement programs.

The delineated molecular signatures regulating traits of agronomic importance have been utilized for introgression, combining, and pyramiding into selected grain legume crop genotypes of interest through traditional and advanced genomics-assisted breeding approaches in order to develop superior crop varieties in terms of yield and stress tolerance (Varshney, 2016). Recently, a chickpea cultivar 'Pusa JG16' has been released in India as a drought-hardy cultivar; this was developed through genomics-assisted breeding utilizing a QTL-hotspot region from ICC4958 (Bhardwaj et al., 2021). Introgressions of functional natural genetic variations and of favorable genes, QTLs, alleles, and chromosomal segments identified from a larger set of grain legume germplasm accessions (including landraces and wild species), particularly for yield and stress component traits, have been transferred into the cultivated genetic background for improvement of the relevant crop through the use of such approaches as introgression lines (ILs), advanced-backcross QTL (AB–QTL) analysis, association genetics, and multi-parent advanced generation intercross (MAGIC) populations (Roorkiwal et al., 2020; Bohra et al., 2021). For example, the 'Geletu' chickpea variety was developed through marker-assisted back-crossing (MABC) and released in Ethiopia; it provided a yield advantage of 15% over the check variety 'Teketay' and 78% over a local check (https://www.icrisat.org/). The MABC technique has been used in the development of introgression of QTLs into elite cultivars in order to develop introgressed lines (Bharadwaj and Yadav, 2012; Varshney et al., 2013b; Barmukh et al., 2022). The 'Pusa Chickpea 20211' variety is another example in which resistance genes for multiple races (*foc* 1,2,3,4, and 5) of fusarium wilt have been stacked through MABC in the mega desi chickpea variety 'Pusa 391' (Bharadwaj et al., 2022). Molecular tags associated with major effects on qualitative and quantitative trait regulation have now been transferred into diverse grain legume crop genotypes for their genetic enhancement through marker-assisted selection (MAS), including MABC and marker-assisted foreground and background selection. The identification of a QTL-hotspot region in linkage group 4 (CaLG04) in chickpea that harbors major QTLs for multiple drought adaptive traits, followed by its introgression into elite chickpea cultivars, is an excellent example of genomics-assisted breeding (Barmukh et al., 2022). This region accounts for 58.2% of explained phenotypic variation and a 16% yield enhancement under drought conditions in introgressed lines, which is primarily attributed to improvements in root traits, such root length, density, surface area, and volume (Varshney et al., 2013b; Roorkiwal et al., 2020; Bharadwaj et al., 2021).

Complications in the domain of genetic background effects/epistasis and linkage drag of QTLs, as well as minor effects of both minor and major QTLs/genes on complex trait regulation, have

impeded the use of the traditional MAS (QTL–MAS) approach in the genetic enhancement of grain legume crops on complex quantitative traits. To overcome these complexities, many novel advanced genomics-assisted breeding approaches are currently emerging, such as marker-assisted recurrent selection (MARS), MAGIC, and genomic/genome-wide (haplotype) selection. These involve the transferal and pyramiding of the favorable alleles of minor-effect genes/QTLs controlling complex quantitative traits for the genetic enhancement of grain legume crop plants in terms of yield, nutritional quality, and stress tolerance. Varshney et al. (2021) have identified superior haplotypes through whole genome sequencing (WGS) of 3,336 accessions, both cultivated (3,171) and wild (195), for important traits relating to yield enhancement; these can then be introgressed into elite chickpea cultivars. The study also identified target genomic regions for the purging of deleterious alleles, which can be achieved through genomics-assisted breeding and/or gene editing (Varshney et al., 2021). Similarly, superior haplotypes have been identified in pigeon pea based on the 292 pigeon pea genotypes of a reference set that included breeding lines, landraces, and wild species (Sinha et al., 2020). In this study, haplotype–phenotype association analysis for drought-responsive traits resulted in the identification of promising haplotypes (*C. cajan_23080*-H2, *C. cajan_30211*-H6, *C. cajan_26230*-H11, and *C. cajan_26230*-H5) for three genes regulating five drought component traits (Sinha et al., 2020). Genomic selection (GS) and integrated genomic–enviromic prediction (iGEP) are other promising strategies that can be used to improve genetic gain in legume crops (Heffner et al., 2009; Xu et al., 2022). GS uses a smaller training population that is well genotyped and phenotyped, while iGEP uses additional data on genotype–environment interactions to build a prediction model. These models are then used to predict the true breeding values of selecting particular candidates based on multi-omics data, big data technology, and artificial intelligence (Xu et al., 2022). Thus, genomics plays an integral role in improving genetic gain in modern agricultural practices and has immense potential to expedite future grain legume crop improvement programs. The traditional and novel genomics-assisted breeding approaches that are now available provide clues to the quantitative dissection of complex trait regulation, and thus have potential to expedite studies of the genetic enhancement of complex traits in diverse grain legume crops.

5 Future outlook

Grain legumes are a major source of food and nutrition globally. However, it has not been possible to enhance gain legume production to the required level, primarily due to the narrow genetic base of most of the legume crops, coupled with changing climatic conditions. Most grain legumes are lacking in desired plant ideotypes and resistance sources for various biotic and abiotic stresses. The genetic diversity conserved in gene banks globally is a major resource for crop breeding programs, but it is utilized only marginally. Therefore, in order to broaden the genetic base of grain legume crops and enhance the genetic gains made in improvement programs, conventional approaches and modern scientific tools should be integrated in a phased and carefully judged manner. The first phase should focus on the search for desired traits and the infusion of diversity into the cultivated gene pool through use of landraces and CWRs. The second phase should focus on the utilization of advanced selection tools, such as genomics, high-throughput precision phenotyping, and artificial intelligence, to exploit the hidden potential of the available genetic diversity; and the third phase should involve technologies such as mutational breeding, genome editing, and transgenic technologies to improve, modify, or introgress any novel or alien traits that are not available in the entire crop gene pool. We presume that enrichment of the genetic diversity of cultivated grain legume gene pools, along with simultaneous improvements in their yield and plant type with the aid of advanced scientific tools, will enhance grain legume crop yields to the required level.

Author contributions

JR conceptualized the topic addressed in the manuscript. Gayacharan, SP, HV, and RY collated the literature and wrote the sections on their respective areas of expertise. NM extended her assistance in grammatical correction, reference editing, and the creation of figures. Gayacharan compiled the contributions received from all authors, and finally the manuscript was edited by JR and Gayacharan. All authors reviewed the final draft of the manuscript and gave their approval.

Funding

The authors thank the GEF project “Mainstreaming agricultural biodiversity conservation and utilization in agricultural sector to ensure ecosystem services and reduce vulnerability,” implemented in India by the UN Environment (Project Code: A 1265), for providing funding support.

Acknowledgments

The authors duly acknowledge the ICAR-National Bureau of Plant Genetic Resources, New Delhi; the DBT–National Institute of Plant Genome Research, New Delhi; Shivaji College, University of Delhi; and the Alliance of Biodiversity International and CIAT, India Office, NASC Complex, New Delhi for providing the necessary support.

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.

References

Ahenkora, K., AduDapaah, H. K., and Agyemang, A. (1998). Selected nutritional components and sensory attributes of cowpea (*Vigna unguiculata* [L.] Walp) leaves. *Plant Foods Hum. Nutr.* 52, 221–229. doi:10.1023/a:1008019113245

Akram, A., Iqbal, S. M., Rauf, C. A., and Aleem, R. (2008). Detection of resistant sources for collar rot disease in chickpea germplasm. *Pak. J. Bot.* 40, 2211–2215.

Amarteifio, J., Munthali, D., Karikari, S., and Morake, T. K. (2002). The composition of pigeon peas (*Cajanus cajan* (L. Millsp.) grown in Botswana. *Plant Foods Hum. Nutr.* 57, 173–177. doi:10.1023/A:1015248326920

AmbikaGayacharanHamwieh, A., Talukdar, A., Kumar, S., Sharma, B. B., et al. (2022). Unraveling origin, history, genetics, and strategies for accelerated domestication and diversification of food legumes. *Front. Genet.* 13, 932430. doi:10.3389/fgene.2022.932430

AmbikaHegde, V., Nimmy, M. S., Bharadwaj, C., Tripathi, S., Singh, R. K., et al. (2021). Unraveling genetics of semi-determinacy and identification of markers for indeterminate stem growth habit in chickpea (*Cicer arietinum* L.). *Sci. Rep.* 11 (1), 21837–21838. doi:10.1038/s41598-021-01464-3

Amirul, I. F. M., Basford, K. E., Redden, R. J., and Beebe, S. (2006). Preliminary evaluation of the common bean core collection at CIAT. *Plant Genet. Resour. newslett. (IPGRI/FAO)* 145, 29.

Archak, S., Tyagi, R. K., Harer, P. N., Mahase, L. B., Singh, N., Dahiya, O. P., et al. (2016). Characterization of chickpea germplasm conserved in the Indian National Genebank and development of a core set using qualitative and quantitative trait data. *Crop J* 4, 417–424. doi:10.1016/j.cj.2016.06.013

Ashfaq, M., Khan, M. A., Mughal, S. M., Javed, N., Mukhtar, T., and Bashir, M. (2007). Evaluation of urdbean germplasm for resistance against urdbean leaf crinkle virus. *Pak. J. Bot.* 39, 2103.

Ashokkumar, K., Diapari, M., Jha, A. B., Tar'an, B., Arganosa, G., and Warkentin, T. D. (2015). Genetic diversity of nutritionally important carotenoids in 94 pea and 121 chickpea accessions. *J. Food Compos. Anal.* 43, 49–60. doi:10.1016/j.jfca.2015.04.014

Ashraf, M., and Waheed, A. (1990). Screening of local/exotic accessions of lentil (Lens culinaris Medic.) for salt tolerance at two growth stages. *Plant Soil* 128, 167–176. doi:10.1007/bf00011106

Azmat, M. A., Nawab, N. N., Khan, A. A., Ashraf, M., Niaz, S., and Mahmood, K. (2011). Characterization of pea germplasm. *Int. J. Veg. Sci.* 17 (3), 246–258. doi:10.1080/19315260.2010.544380

Bag, M. K., Gautam, N. K., Prasad, T. V., Pandey, S., Dutta, M., and Roy, A. (2014). Evaluation of an Indian collection of blackgram germplasm and identification of resistance sources to Mungbean yellow mosaic virus. *Crop Prot.* 61, 92–101. doi:10.1016/j.cropro.2014.03.021

Bansal, R., Sharma, S., Tripathi, K., Gayacharanand Kumar, A. (2019). Waterlogging tolerance in blackgram [*Vigna mungo* (L.) Hepper] is associated with chlorophyll content and membrane integrity. *Indian J. biochem. Biophys.* 5, 81.

Barmukh, R., Roorkiwal, M., Dixit, G. P., Bajaj, P., Kholova, J., Smith, M. R., et al. (2022). Characterization of 'QTL-hotspot' introgression lines reveals physiological mechanisms and candidate genes associated with drought adaptation in chickpea. *J. Exp. Bot.* 73 (22), 7255–7272. doi:10.1093/jxb/erac348

Bartwal, A., John, R., Padhi, S. R., Suneja, P., Bhardwaj, B., Gayacharan, et al. (2022). NIR spectra processing for developing efficient protein prediction model in mungbean. *J. Food Compos. Anal.* 105087, 105087. doi:10.1016/j.jfca.2022.105087

Basu, P. S., Ali, M., and Chaturvedi, S. K. (2009). "Terminal heat stress adversely affects chickpea productivity in Northern India–strategies to improve thermotolerance in the crop under climate change," in *W3 workshop proceedings: Impact of climate change on agriculture*, 189–193.

Bharadwaj, C., Jorben, J., Rao, A., Roorkiwal, M., Patil, B. S., Ahammed, S. K., et al. (2022). Development of high yielding Fusarium wilt resistant cultivar by pyramiding of "genes" through marker-assisted backcrossing in chickpea (*Cicer arietinum* L.). *Front. Genet.* 1747, 924287. doi:10.3389/fgene.2022.924287

Bharadwaj, C., Tripathi, S., Soren, K. R., Thudi, M., Singh, R. K., Sheoran, S., et al. (2021). Introgression of "QTL-hotspot" region enhances drought tolerance and grain yield in three elite chickpea cultivars. *Plant Genome* 14 (1), e20076. doi:10.1002/tpg2.20076

Bhardwaj, J., and Yadav, S. K. (2021). Comparative study on biochemical parameters and antioxidant enzymes in a drought tolerant and a sensitive variety of Horsegram (*Macrotylomauniflorum*) under drought stress. *Am. J. Plant Physiol.* 7 (1), 17–29. doi:10.3923/ajpp.2012.17.29

Bhuvaneswari, S., Sharma, S. K., Punitha, P., Shashidhar, K. S., Naveenkumar, K. L., and Prakash, N. (2017). Evaluation of morphological diversity of field pea [*Pisum sativum* subsp. arvense (L.)] germplasm under sub–tropical climate of Manipur. *Legume Res.* 40 (2), 215. doi:10.18805/lr.v0iof.10756

Bisht, I. S., Mahajan, R. K., and Patel, D. P. (1998). The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. *Genet. Resour. Crop Evol.* 45, 127–133. doi:10.1023/a:1008670332570

Biswas, K. K., Tarafdar, A., Kumar, A., Dikshit, H. K., and Malathi, V. G. (2009). Multiple infection in urdbean (*Vigna mungo*) in natural condition by begomovirus, tospovirus and urdbean leaf crinkle virus complex. *Indian Phytopathol.* 62 (1), 75–82.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S. R., et al. (2021). Reap the crop wild relatives for breeding future crops. *Trends Biotechnol.* 40, 412–431. doi:10.1016/j.tibtech.2021.08.009

Bohra, A., Saxena, R. K., Gnanesh, B. N., Saxena, K., Byregowda, M., Rathore, A., et al. (2012). An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor. Appl. Genet.* 125 (6), 1325–1338. doi:10.1007/s00122-012-1916-5

Boukar, O., Belko, N., Chamarthi, S., Togola, A., Batieno, J., Owusu, E., et al. (2019). Cowpea (*Vigna unguiculata*): Genetics, genomics and breeding. *Plant Breed.* 138, 415–424. doi:10.1111/pbr.12589

Burdon, J. J., Barrett, L. G., Rebetzke, G., and Thrall, P. H. (2014). Guiding deployment of resistance in cereals using evolutionary principles. *Evol. Appl.* 7 (6), 609–624. doi:10.1111/eva.12175

Butare, L., Rao, I., Lepoivre, P., Cajiao, C., Polania, J., Cuasquer, J., et al. (2012). Phenotypic evaluation of interspecific recombinant inbred lines (RILs) of Phaseolus species for aluminium resistance and shoot and root growth response to aluminium–toxic acid soil. *Euphytica* 186 (3), 715–730. doi:10.1007/s10681-011-0564-1

Celmeli, T., Sari, H., Canci, H., Sari, D., Adak, A., Eker, T., et al. (2018). The nutritional content of common bean (*Phaseolus vulgaris* L.) landraces in comparison to modern varieties. *Agronomy* 8, 166. doi:10.3390/agronomy8090166

Chaudhary, N. B., Patel, A. M., Chaudhari, R. H., Viradiyam, Y. A., and Patel, H. S. (2021). Genetic analysis for yield and quality related traits in moth bean [*Vigna aconitifolia* (Jacq.) Marechal]. *Int. J. Chem. Stud.* 9 (1), 2658–2661. doi:10.22271/chemi.2021.v9.i1ak.11629

Chaudhry, M. A., Muhammad, F., and Afzal, M. (2007). Screening of chickpea germplasm against Fusarium wilt. *J. Agric. Res.* 44 (4), 307.

Chauhan, J. S., Singh, B. B., and Gupta, S. (2016). Enhancing pulses production in India through improving seed and variety replacement rates. *Indian J. Genet. Plant Breed.* 76 (4), 410–10. doi:10.5958/0975-6906.2016.00060.2

Cheboi, J. J., Kinyua, M. G., Kimurto, P. K., Kiplagat, O. K., Nganga, F., and Ghimire, S. R. (2019). Biochemical composition of pigeonpea genotypes in Kenya. *Aust. J. Crop Sci.* 13 (11), 1848–1855. doi:10.21475/ajcs.19.13.11.p1886

Chiang, H. S., and Talekar, N. S. (1980). Identification of sources of resistance to the beanfly and two other agromyzid flies in soybean and mungbean. *J. Econ. Entomol.* 73 (2), 197–199. doi:10.1093/jee/73.2.197

Choudhary, A. K., Sultana, R., Pratap, A., Nadarajan, N., and Jha, U. C. (2011). Breeding for abiotic stresses in pigeonpea. *J. Food Legum.* 24 (3), 165.

Choudhary, A. K., Sultana, R., Vales, M. I., Saxena, K. B., Kumar, R. R., and Ratnakumar, P. (2018). Integrated physiological and molecular approaches to improvement of abiotic stress tolerance in two pulse crops of the semi–arid tropics. *Crop J* 6 (2), 99–114. doi:10.1016/j.cj.2017.11.002

Cichy, K. A., Porch, T. G., Beaver, J. S., Cregan, P., Fourie, D., Glahn, R. P., et al. (2015a). A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Sci.* 55 (5), 2149–2160. doi:10.2135/cropsci2014.09.0653

Cichy, K. A., Wiesinger, J. A., and Mendoza, F. A. (2015b). Genetic diversity and genome–wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 128 (8), 1555–1567. doi:10.1007/s00122-015-2531-z

Costantini, M., Summo, C., Centrone, M., Rybicka, I., D'Agostino, M., Annicchiarico, P., et al. (2021). Macro– and micro–nutrient composition and antioxidant activity of chickpea and pea accessions. *Pol. J. Food Nutr. Sci.* 71 (2), 177–185. doi:10.31883/pjfns/135813

Das, A., Parihar, A. K., Saxena, D., Singh, D., Singha, K. D., Kushwaha, K. P. S., et al. (2019). Deciphering genotype-by-environment interaction for targeting test environments and rust resistant genotypes in field pea (Pisum sativum L.). *Front. Plant Sci.* 10, 825. doi:10.3389/fpls.2019.00825

Dasgan, H. Y., and Koc, S. (2009). Evaluation of salt tolerance in common bean genotypes by ion regulation and searching for screening parameters. *J. Food Agric. Environ.* 7 (2), 363.

Davis, D. R., Epp, M. D., and Riordan, H. D. (2004). Changes in USDA food composition data for 43 garden crops, 1950 to 1999. *J. Am. Coll. Nutr.Vigna Unguiculata.* 231 (64), 669. doi:10.1080/07315724.2004.10719409

Deshpand, K. S., Patil, B. R., Salimath, P. M., Nidagundi, J. M., and Karthigeyan, S. (2010). Evaluation of native and collected germplasm for earliness seed traits and resistance to rust, CMV and leaf spot in cowpea [*Vigna unguiculata* (L.) Walp]. *Electron. J. Plant Breed.* 1, 384

Desingh, R., and Kanagaraj, G. (2019). Differential responses to salt stress on antioxidant enzymatic activity of two horse gram [*Macrotyloma uniflorum* (Lam.) Verdc] varieties. *Int. J. Res. Anal. Rev.* 6, 425.

Devasirvatham, V., Gaur, P. M., Raju, T. N., Trethowan, R. M., and Tan, D. K. Y. (2015). Field response of chickpea (*Cicer arietinum* L.) to high temperature. *Field Crops Res.* 172, 59–71. doi:10.1016/j.fcr.2014.11.017

Devi, J., Mishra, G. P., Sanwal, S. K., Dubey, R. K., Singh, P. M., and Singh, B. (2018). Development and characterization of penta-flowering and triple-flowering genotypes in garden pea (Pisum sativum L. var. hortense). *PloS one* 13 (7), e0201235. doi:10.1371/journal.pone.0201235

Dhali, S., Pradhan, M., Sahoo, R. K., Mohanty, S., and Pradhan, C. (2021). Alleviating Cr(VI) stress in horse gram (*Macrotyloma uniflorum* Var. Madhu) by native Cr-tolerant nodule endophytes isolated from contaminated site of Sukinda. *Environ. Sci. Pollut. Res.* 28, 31717–31730. doi:10.1007/s11356-021-13009-2

Dhasarathan, M., Geetha, S., Karthikeyan, A., Sassikumar, D., and Meenakshiganesan, N. (2021). Development of novel blackgram (*Vigna mungo* (L.) Hepper) mutants and deciphering Genotype × Environment interaction for yield–related traits of mutants. *Agronomy* 11, 1287. doi:10.3390/agronomy11071287

Divya, P., Durga, K. K., and Udayababu, P. (2013). Relative susceptibility of horse gram (*Macrotyloma uniflorum*) accessions to the attack of *Callosobruchus chinensis* (Fab.) (Coleoptera: Buchidae). *Madras Agric. J.* 100, 597.

Divya, P., Durga, K. K., Sunil, N., Rajasri, M., and Udayababu, P. (2012). Screening of horse gram accessions against pulse beetle, *Callosobruchus chinensis*. *Indian J. Plant Prot.* 40, 280.

Dong, L., Ravelombola, W., Weng, Y., Qin, J., Bhattarai, G., Zia, B., et al. (2019). Seedling salt tolerance for above ground–related traits in cowpea (*Vigna unguiculata* (L.) Walp). *Euphytica* 215 (3), 53–22. doi:10.1007/s10681-019-2379-4

Duraimurugan, P., Pratap, A., Singh, S. K., and Gupta, S. (2014). Evaluation of screening methods for bruchid beetle (*Callosobruchus chinensis*) resistance in greengram (*Vigna radiata*) and blackgram (*Vigna mungo*) genotypes and influence of seed physical characteristics on its infestation. *Vegetos* 27 (1), 60–67. doi:10.5958/j.2229-4473.27.1.011

Durga, K. K. (2012). Variability and divergence in horse gram (*Dolichos uniflorus*). *J. Arid. Land* 4 (1), 71–76. doi:10.3724/sp.j.1227.2012.00071

Ebert, A. W., Chang, C. H., Yan, M. R., and Yang, R. Y. (2017). Nutritional composition of mungbean and soybean sprouts compared to their adult growth stage. *Food Chem.* 237, 15–22. doi:10.1016/j.foodchem.2017.05.073

Edulamudi, P., Masilamani, A. J. A., Vanga, U. R., Divi, V. R. S. G., and Konada, V. M. (2021). Nickel tolerance and biosorption potential of rhizobia associated with horse gram [*Macrotyloma uniflorum* (Lam.) Verdc.]. *Int. J. Phytoremediation* 23 (11), 1184–1190. doi:10.1080/15226514.2021.1884182

El–Haddad, N., Rajendran, K., Smouni, A., Es–Safi, N. E., Benbrahim, N., Mentag, R., et al. (2020). Screening the FIGS set of lentil (*Lens culinaris* Medikus) germplasm for tolerance to terminal heat and combined drought–heat stress. *Agronomy* 10 (7), 1036. doi:10.3390/agronomy10071036

Emmrich, P. M., Sarkar, A., Njaci, I., Kaithakottil, G. G., Ellis, N., Moore, C., et al. (2020). *A draft genome of grass pea (*Lathyrus sativus*), a resilient diploid legume*. Biorxiv [Preprint]. Available at:Accessed May 15, 2022) https://www.biorxiv.org/content/10.1101/2020.04.24.058164v1.

Erskine, W., Meyveci, K., and Izgin, N. (1981). Screening a world lentil collection for cold tolerance. *LENS Newsl.* 8, 5.

Fan, M. S., Zhao, F. J., Fairweather–Tait, S. J., Poulton, P. R., Dunham, S. J., and McGrath, S. P. (2008). Evidence of decreasing mineral density in wheat grain over the last 160 years. *J. Trace Elem. Med. Biol.* 22 (4), 315–324. doi:10.1016/j.jtemb.2008.07.002

Fatokun, C. A., Boukar, O., and Muranaka, S. (2012). Evaluation of cowpea (*Vigna unguiculata* (L.) Walp.) germplasm lines for tolerance to drought. *Plant Genet. Resour.* 10 (3), 171–176. doi:10.1017/s1479262112000214

Frankel, O. H. (1984). "Genetic perspective of germplasm conservation," in *Genetic manipulations: Impact on man and society*. Editor W. Arber (England: Cambridge Univ. Press), 161–170.

Ganjeali, A., Porsa, H., and Bagheri, A. (2011). Assessment of Iranian chickpea (*Cicer arietinum* L.) germplasms for drought tolerance. *Agric. Water Manag.* 98 (9), 1477–1484. doi:10.1016/j.agwat.2011.04.017

Garcia, T., Duitama, J., Zullo, S. S., Gil, J., Ariani, A., Dohle, S., et al. (2021). Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nat. Commun.* 12 (1), 702–717. doi:10.1038/s41467-021-20921-1

Garg, V., Dudchenko, O., Wang, J., Khan, A. W., Gupta, S., Kaur, P., et al. (2022). Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement. *J. Adv. Res.* 42, 315–329. doi:10.1016/j.jare.2021.10.009

Gaur, P. M., and Gour, V. K. (2002). A gene producing one to nine flowers per flowering node in chickpea. *Euphytica* 128 (2), 231–235. doi:10.1023/a:1020845815319

Gaur, R., Azam, S., Jeena, G., Khan, A. W., Choudhary, S., Jain, M., et al. (2012). High–throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). *DNA Res.* 19, 357–373. doi:10.1093/dnares/dss018

Gautam, M., and Chahota, R. K. (2022). Metabolite profiling and protein quantification to a large library of 96 horsegram (*Macrotyloma uniflorum*) germplasm. *Sci. Rep.* 12 (1), 7865–7869. doi:10.1038/s41598-022-11962-7

Gautam, N. K., Kumar, K., and Prasad, M. (2016). Leaf crinkle disease in urdbean (*Vigna mungo* L. Hepper): An overview on causal agent, vector and host. *Protoplasma* 253 (3), 729–746. doi:10.1007/s00709-015-0933-z

Gautam, N. K., Singh, M., Khan, Z., Roy, A., Akhtar, J., and Ram, B. (2013). Assessment of lentil germplasm with respect to agromonic performance and major biotic stress. *Legume Res.* 36 (3), 214–219.

GayacharanGupta, K., Gupta, V., Tyagi, V., and Singh, K. (2020a). "Mungbean genetic resources and utilization," in *The mungbean genome, compendium of plant genomes*. Editors R. Nair, R. Schafleitner, and S. H. Lee (Cham: Springer), 9–25. doi:10.1007/978–3–030–20008–4_2

GayacharanRani, U., Singh, S., Basandrai, A. K., Rathee, V. K., Tripathi, K., et al. (2020b). Identification of novel resistant sources for ascochyta blight (*Ascochyta rabiei*) in chickpea. *Plos One* 15 (10), e0240589. doi:10.1371/journal.pone.0240589

GayacharanTripathi, K., Aski, M. S., Singh, N., Kumar, A., and Lal, H. (2022). Understanding genetic diversity in blackgram [*Vigna mungo* (L.) Hepper] collections of Indian National Genebank. *Genet. Resour. Crop Evol.* 69 (3), 1229–1245. doi:10.1007/s10722-021-01301-6

Gerrano, A. S., Adebola, P. O., Rensburg, W. S. J., and Laurie, S. M. (2015). Genetic variability in cowpea (*Vigna unguiculata* (L.) Walp.) genotypes. *S. Afr. J. Plant Soil* 32 (3), 165–174. doi:10.1080/02571862.2015.1014435

Gerrano, A. S., Rensburg, W. S. J., Adebola, P. O., et al. (2017). Nutritional composition of immature pods in selected cowpea (*Vigna unguiculata* (L.) Walp.) genotypes in South Africa. *Aust. J. Crop Sci.* 11 (2), 134–141. doi:10.21475/ajcs.17.11.02.p72

Ghafoor, A., Sharif, A., Ahmad, Z., Zahid, M. A., and Rabbani, M. A. (2001). Genetic diversity in blackgram (*Vigna mungo* L. Hepper). *Field Crops Res.* 69 (2), 183–190. doi:10.1016/s0378-4290(00)00141-6

Girish, T. K., Pratape, V. M., and Rao, U. P. (2012). Nutrient distribution, phenolic acid composition, antioxidant and alpha–glucosidase inhibitory potentials of blackgram (*Vigna mungo* L.) and its milled by–products. *Food Res. Int.* 46 (1), 370–377. doi:10.1016/j.foodres.2011.12.026

Goa, Y., Mohammed, H., Worku, W., and Urage, E. (2022). Genotype by environment interaction and yield stability of cowpea (*Vigna unguiculata* (L.) Walp.) genotypes in moisture limited areas of Southern Ethiopia. *Heliyon* 8 (3), e09013. doi:10.1016/j.heliyon.2022.e09013

Gomashe, S. S., Dikshit, N., Chand, D., and Shingane, S. N. (2018). Assessment of genetic diversity using morpho-agronomical traits in horse gram. *Int. J. Curr. Microbiol. Appl. Sci.* 7 (5), 2095–2103. doi:10.20546/ijcmas.2018.705.244

Guo, Y. (2022). Disentangling the breakdown of resistance. *Nat. Food* 3 (8), 566. doi:10.1038/s43016-022-00582-z

Gupta, S., Nawaz, K., Parween, S., Roy, R., Sahu, K., Kumar, A., et al.(2017). Draft genome sequence of *Cicer reticulatum* L., the wild progenitor of chickpea provides a resource for agronomic trait improvement. *DNA Res.* 24 (1), 1–10. doi:10.1093/dnares/dsw042

Ha, J., Satyawan, D., Jeong, H., Lee, E., Cho, K. H., Kim, M. Y., et al. (2021). A near-complete genome sequence of mungbean (*Vigna radiata* L.) provides key insights into the modern breeding program. *Plant Genome* 14 (3), e20121. doi:10.1002/tpg2.20121

Hacisalihoglu, G., Beisel, N. S., and Settles, A. M. (2021). Characterization of pea seed nutritional value within a diverse population of *Pisum sativum*. *PloS One* 16 (11), e0259565. doi:10.1371/journal.pone.0259565

Halila, M. H., and Strange, R. N. (1997). Screening of Kabuli chickpea germplasm for resistance to Fusarium wilt. *Euphytica* 96 (2), 273–279. doi:10.1023/a:1003051100683

Hamdi, A., Kusmenoglu, I., and Erskine, W. (1996). Sources of winter-hardiness in wild lentil. *Genet. Resour. Crop Evol.* 43, 63–67. doi:10.1007/bf00126942

Hannan, R. M., Kaiser, W. J., and Muehlbauer, F. J. (1994). *Development and utilization of the USDA chickpea germplasm core collection*. Madison, WI: Agronomy AbstractsASA, 217.

Hasan, M. K., Sabagh, A., Sikdar, M. S. I., Alam, M. J., Ratnasekera, D., Barutcular, C., et al. (2017). Comparative adaptable agronomic traits of blackgram and mungbean for saline lands. *Plant Arch.* 17 (1), 589–593.

Haus, M. J., Pierz, L. D., Jacobs, J. L., Wiersma, A. T., Awale, H. E., Chilvers, M. I., et al. (2021). Preliminary evaluation of wild bean (*Phaseolus* spp.) germplasm for resistance to *Fusarium cuneirostrum* and *Fusarium oxysporum*. *Crop Sci.* 61 (5), 3264–3274. doi:10.1002/csc2.20495

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 crop. Res.* 30, 147–154. doi:10.1016/0378-4290(92)90063-f

Hawkes, J. G. (1983). *The diversity of crop plants*. Cambridge, Massachusetts: Harvard University Press.

Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49 (1), 1–12. doi:10.2135/cropsci2008.08.0512

Hegde, V. S. (2011). Morphology and genetics of a new found determinate genotype in chickpea. *Euphytica* 182 (1), 35–42. doi:10.1007/s10681-011-0447-5

Hema, M., Sreenivasulu, P., Patil, B. L., Kumar, P. L., and Reddy, D. V. (2014)., 90. Academic Press, 431–505.Tropical food legumes: Virus diseases of economic importance and their control*Adv. virus Res.*

Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10, 716–732. doi:10.1111/j.1467-7652.2012.00710.x

Hobson, K., Armstrong, R., Nicolas, M., Connor, D., and Materne, M. (2006). Response of lentil (*Lens culinaris*) germplasm to high concentrations of soil boron. *Euphytica* 151 (3), 371–382. doi:10.1007/s10681-006-9159-7

Horn, Lydia N., and Shimelis, H. (2020). Production constraints and breeding approaches for cowpea improvement for drought prone agro-ecologies in Sub-Saharan Africa. *Ann. Agric. Sci.* 651, 83–91. doi:10.1016/j.aoas.2020.03.002

Hu, Z. J., Huang, Y. Y., Lin, X. Y., Feng, H., Zhou, S. X., Xie, Y., et al. (2022). Loss and natural variations of blast fungal avirulence genes breakdown rice resistance genes in the sichuan basin of China. *Front. Plant Sci.* 13, 788876. doi:10.3389/fpls.2022.788876

Iqbal, U., Iqbal, S. M., Afzal, R., Jamal, A., Farooq, M. A., and Zahid, A. (2011). Screening of mungbean germplasm against mungbean yellow mosaic virus (MYMV) under field conditions. *Pak. J. Phytopathol.* 23 (1), 48–51.

Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant. J.* 74, 715–729. doi:10.1111/tpj.12173

Jarpa-Parra, M. (2018). Lentil protein: A review of functional properties and food application. An overview of lentil protein functionality. *Int. J. Food Sci. Technol.* 53 (4), 892–903. doi:10.1111/ijfs.13685

Jegadeesan, S., Raizada, A., Dhanasekar, P., and Suprasanna, P. (2021). Draft genome sequence of the pulse crop blackgram [*Vigna mungo* (L.) Hepper] reveals potential R-genes. *Sci. Rep.* 11, 11247. doi:10.1038/s41598-021-90683-9

Jha, U. C., Nayyar, H., Parida, S. K., and Siddique, K. H. (2022). *Horse gram, an underutilized climate–resilient legume: Breeding and genomic approach for improving future genetic gain" in developing climate resilient grain and forage legumes*. Singapore: Springer, 167.

Jood, S., Bishnoi, S., and Sharma, A. (1998). Chemical analysis and physico-chemical properties of chickpea and lentil cultivars. *Food/Nahrung* 42 (02), 71–74. doi:10.1002/(sici)1521-3803(199804)42:02<71::aid-food71>3.3.co;2-u

Jukanti, A. K., Gaur, P. M., Gowda, C. L. L., and Chibbar, R. N. (2012). Nutritional quality and health benefits of chickpea (*Cicer arietinusm* L.): A review. *Br. J. Nutr.* 108 (S1), S11–S26. doi:10.1017/S0007114512000797

Kabi, M., Das, T. R., and &Baisakh, B. (2017). Screening of superior genotypes for cold tolerance and MYMV resistance in greengram (Vigna radiata). *Int. J. Currmicrobiol. Sci.* 12, 2270. doi:10.20546/ijcmas

Kanagaraj, G., and Sathish, C. (2017). Salt Stress induced changes in growth, pigments and protein contents in two horse gram varieties. *J. Sci. Agric.* 08.

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5 (1), 5443–5449. doi:10.1038/ncomms6443

Kang, Y. J., Satyawan, D., Shim, S., Lee, T., Lee, J., Hwang, W. J., et al. (2015). Draft genome sequence of adzuki bean, *Vigna angularis*. *Sci. Rep.* 5, 8069. doi:10.1038/srep08069

Karthikeyan, A., Shobhana, V. G., Sudha, M., Raveendran, M., Senthil, N., Pandiyan, M., et al. (2014). Mungbean yellow mosaic virus (MYMV): A threat to greengram (*Vigna radiata*) production in Asia. *Int. J. Pest Manag.* 60 (4), 314–324. doi:10.1080/09670874.2014.982230

Kaul, T., Eswaran, M., Thangaraj, A., Meyyazhagan, A., Nehra, M., Raman, N. M., et al. (2019). Rice bean (*Vigna umbellata*) draft genome sequence: Unravelling the late flowering and unpalatability related genomic resources for efficient domestication of this underutilized crop. *BioRxiv* (Preprint)Available at: https://www.biorxiv.org/content/10.1101/816595v1.abstract (Accessed May 15, 2022).

Kaundal, S. P., and Kumar, R. (2021). Preliminary evaluation for most promising accessions from procured germplasm of forgotten medicinal legume Horsegram {*Macrotyloma uniflorum* (Lam.) Verdc} in Jalandhar Region of Punjab. *J. Res. Agric. Anim. Sci.* 8 (12), 09.

Kaur, S., Singh, N., Sodhi, N. S., Rana, J. C., Kawale, S. B., Kadam, S. S., et al. (2009). Diversity in properties of seed and flour of kidney bean germplasmEffect of processing on insoluble dietary fiber and resistant starch in kidney bean and horse gram. *Food Chem. Food Sci. Technol.* 11742 (2), 282361.

Keller, B., Ariza–Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla–Benavides, A. E., Buendia, H. F., et al. (2020). Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. *Front. Plant Sci.* 11, 1001. doi:10.3389/fpls.2020.01001

Khan, Z., Gautam, N. K., Gawade, B. H., and Dubey, S. C. (2017). Screening of lentil (*Lens culinaris*Medik.) germplasm for resistance to root–knot nematode, *Meloidogyneincognita*. *Indian J. Genet.* 77 (3), 408–413. doi:10.5958/0975-6906.2017.00055.4

Khine, N. A., Kundu, K. K., Malik, D. P., and Devi, M. (2021). Production and trade performance of blackgram (*Vigna mungo*) and greengram (*Vigna radiata*) in India and Myanmar. *Asian J. Agric. Ext. Econ. Sociol.* 39 (10), 231–243. doi:10.9734/ajaees/2021/v39i1030687

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 (9), 1411–1422. doi:10.1038/s41588-019-0480-1

Krishnamurthy, L., Gaur, P. M., Basu, P. S., Chaturvedi, S. K., Tripathi, S., Vadez, V., et al. (2011). Large genetic variation for heat tolerance in the reference collection of chickpea (*Cicer arietinum* L.) germplasm. *Plant Genet. Resour.* 9 (1), 59–69. doi:10.1017/s1479262110000407

Krishnamurthy, L., Kashiwagi, J., Gaur, P. M., Upadhyaya, H. D., and Vadez, V. (2010). Sources of tolerance to terminal drought in the chickpea (*Cicer arietinum* L.) minicore germplasm. *Field crop. Res.* 119, 322–330. doi:10.1016/j.fcr.2010.08.002

Krishnamurthy, L., Upadhyaya, H. D., Saxena, K. B., and Vadez, V. (2012). Variation for temporary waterlogging response within the mini core pigeonpea germplasm. *J. Agric. Sci.* 150 (3), 357–364. doi:10.1017/s0021859611000682

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Front. Plant Sci.* 6, 162. doi:10.3389/fpls.2015.00162

Kulkarni, N. K., Reddy, A. S., Kumar, P. L., Vijaynarasimha, J., Rangaswamy, K. T., Muniyappa, V., et al. (2003). Broad–based resistance to pigeonpea sterility mosaic disease in accessions of Cajanus scarabaeoides (L.) Benth. *Indian J. Plant Prot.* 31 (1), 6–11.

Kumar, H., Dikshit, H. K., Singh, A., Jain, N., Kumari, J., Singh, A. M., and Prabhu, K. V. (2014). Characterization of grain iron and zinc in lentil ('Lens culinaris' Medikus' culinaris') and analysis of their genetic diversity using SSR markers. *Aust. J. Crop Sci.* 8 (7), 1005.

Kumar, M., and Singh, P. S. (2014). Screening of blackgram genotypes against major insect pests. *Indian J. Entomology* 76 (1), 84.

Kumar, N., Bharadwaj, C., Pal, M., Satyavathi, C. T., Kumar, T., Singhal, T., et al. (2016). Screening of chickpea (*cicer arietinum* L.) genotypes for seedling salinity tolerance. *J. Trop. Agric.* 34 (3), 667.

Kumar, S., Barpete, S., Kumar, J., Gupta, P., and Sarker, A. (2013). Global lentil production: Constraints and strategies. *SATSA Mukhapatra–Annual Tech.* 17, 1.

Kumar, S., Gupta, S., Chandra, S., and Singh, B. B. (2004). "How wide is the genetic base of pulse crops?" in *Pulses in new perspective*. Editors M. Ali, B. B. Singh, S. Kumar, and V. Dhar (Kanpur, India: Indian Institute of Pulse Research), 211.

Kumar, V., Singh, B. M., Singh, S., and Sugha, S. K. (1997). Evaluation of lentil germplasm for resistance to rust. *Lentil Exp. News Serv.* 24 (1–2), 21–22.

Kumari, S., Sahni, S., and Prasad, B. D. (2020). Screening of urdbean germplasm for high seed yield coupled with resistance to Mungbean yellow mosaic virus (MYMV). *Curr. Appl. Sci. Technol.* 39 (1), 100–106. doi:10.9734/cjast/2020/v39i130485

Lamichaney, A., Parihar, A. K., Hazra, K. K., Dixit, G. P., Katiyar, P. K., Singh, D., et al. (2021). Untangling the influence of heat stress on crop phenology, seed set, seed weight, and germination in field pea (*Pisum sativum* L.). *Front. Plant Sci.* 12, 635868. doi:10.3389/fpls.2021.635868

Laserna–Ruiz, I., De–Los, M. P., Santana–Méridas, O., Sánchez–Vioque, R., and Rodríguez–Conde, M. F. (2012). Screening and selection of lentil (Lens Miller) germplasm resistant to seed bruchids (*Bruchus* spp.). *Euphytica* 188 (2), 153–162. doi:10.1007/s10681-012-0752-7

Lateef, D. D. (2015). DNA marker technologies in plants and applications for crop improvements. *J. Biosci. Med.* 3 (05), 7–18. doi:10.4236/jbm.2015.35002

Leonforte, A., Forster, J. W., Redden, R. J., Nicolas, M. E., and Salisbury, P. A. (2013). Sources of high tolerance to salinity in pea (*Pisum sativum* L.). *Euphytica* 189 (2), 203–216. doi:10.1007/s10681-012-0771-4

Logozzo, G., Donnoli, R., Macaluso, L., Papa, R., Knuppfer, H., and Zeuli, P. S. (2007). Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genet. Resour. Crop Evol.* 54 (8), 1763–1779. doi:10.1007/s10722-006-9185-2

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S., Lo, S., et al. (2019). The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J* 98 (5), 767–782. doi:10.1111/tpj.14349

Longvah, T., Ananthan, R., Bhaskarachary, K., and Venkaiah, K. (2017). *Indian food composition tables*. Hyderabad: National Institute of Nutrition Indian Council of Medical Research Department of Health Research Ministry of Health and Family Welfare, Government of India, 501.

Mahalakshmi, V., Ng, Q., Lawson, M., and Ortiz, R. (2007). Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical descriptors. *Plant Genet. Resour.* 5 (3), 113–119. doi:10.1017/s1479262107837166

Mahesh, H. B., Prasannakumar, M. K., Manasa, K. G., Perumal, S., Khedikar, Y., Kagale, S., et al. (2021). Genome, transcriptome, and germplasm sequencing uncovers functional

variation in the warm–season grain legume Horsegram *Macrotyloma uniflorum* (Lam.) Verdc. *Front. Plant Sci.* 12, 758119. doi:10.3389/fpls.2021.758119

Mahuku, G. S., Jara, C., Cajiao, C., and Beebe, S. (2003). Sources of resistance to angular leaf spot (*Phaeoisariopsisgriseola*) in common bean core collection, wild *Phaseolus vulgaris* and secondary gene pool. *Euphytica* 130 (3), 303–313. doi:10.1023/a:1023095531683

Malambane, G., and Bhatt, K. V. (2014). Agro–morphological performance of moth bean (*Vigna aconitifolia*) lines in severe drought stressed and rain–fed condition. *Direct Res. J. Agric. Food Sci.* 2, 107.

Maliro, M. F., McNeil, D., Redden, B., Kollmorgen, J. F., and Pittock, C. (2008). Sampling strategies and screening of chickpea (*Cicer arietinum* L.) germplasm for salt tolerance. *Genet. Resour. Crop Evol.* 55 (1), 53–63. doi:10.1007/s10722-007-9214-9

Mbinda, W., and Masaki, H. (2021). Breeding strategies and challenges in the improvement of blast disease resistance in finger millet. A current review. *Front. Plant Sci.* 11, 602882. doi:10.3389/fpls.2020.602882

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 (4), 499–515. doi:10.1007/s10722-011-9699-0

Meena, J. K., Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., Srinivasa, N., et al. (2017). Screening of lentil (*Lens culinaris*Medikus sub sp. *culinaris*) germplasm against Fusarium wilt (*Fusarium oxysporum* f. sp. lentis). *Int. J. Curr. Microbiol. App. Sci.* 6 (11), 2533–2541. doi:10.20546/ijcmas.2017.611.298

Meena, S. K., Pandey, R., Sharma, S., Kumar, T., Singh, M. P., Dikshit, H. K., et al. (2021). Physiological basis of combined stress tolerance to low phosphorus and drought in a diverse set of mungbean germplasm. *Agronomy* 11 (1), 99. doi:10.3390/agronomy11010099

Meghwal, R., Joshi, U., Kumar, S., and Sharma, R. (2015). Screening of moth bean (*Vigna aconitifolia*) core collection against yellow mosaic virus. *Indian J. agri. Sci.* 85, 571–575.

Mehra, R., Sarker, A., Dixit, H. K., Aski, M., Khandia, R., and Munjal, A. (2018). Genetic diversity iron and zinc content in lentil (Lens culinaris Medikus subsp. culinaris) as assessed by SSR marker. *Life Sci. Inf. Publ.* 4, 440.

Michael, T. P., and Jackson, S. (2013). The first 50 plant genomes. *Plant Genome* 6. doi:10.3835/plantgenome2013.03.0001in

Mir, R. R., Kudapa, H., Srikanth, S., Saxena, R. K., Sharma, A., Azam, S., et al. (2014). Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.). *Theor. Appl. Genet.* 127 (12), 2663–2678. doi:10.1007/s00122-014-2406-8

Mishra, G. P., Dikshit, H. K., Kumari, J., Tripathi, K., Devi, J., Aski, M., et al. (2020a). Identification and characterization of novel penta-podded genotypes in the cultivated lentil. *Crop Sci.* 60 (4), 1974–1985. doi:10.1002/csc2.20156

Mishra, G. P., Dikshit, H. K., Sv, R., Tripathi, K., Kumar, R. R., Aski, M., et al. (2020b). Yellow mosaic disease (YMD) of mungbean (*Vigna radiata* (L.) Wilczek): Current status and management opportunities. *Front. Plant Sci.* 11, 1064. doi:10.3389/fpls.2020.01064

Mishra, S. K., Singh, B. B., Daya, C., and Meena, K. N. (2005). Sources for high temperature tolerant germplasm in cowpea (*Vigna unguiculata* L. Walp.). *J. Arid. Legum.* 2 (1), 78.

Mohan, S., Sheeba, A., Murugan, E., and Ibrahim, S. M. (2014). Screening of mungbean germplasm for resistance to mungbean yellow mosaic virus under natural condition. *Indian. J. Sci. Technol.* 7 (7), 891–896. doi:10.17485/ijst/2014/v7i7.1

Nainu, A., and Murugan, S. (2020). Screening of mungbean [*Vigna radiata* (L.) Wilczek] genotypes under field condition for resistance against mungbean yellow mosaic virus (MYMV). *Plant Arch.* 20 (2), 6867.

Nair, R. M., Yang, R. Y., Easdown, W. J., Thavarajah, D., Thavarajah, P., Hughes, J. D. A., et al. (2013). Biofortification of mungbean (*Vigna radiata*) as a whole food to enhance human health. *J. Sci. Food Agric.* 93 (8), 1805–1813. doi:10.1002/jsfa.6110

Nene, Y. L., Reddy, M. V., Beniwal, S. P. S., Mahmood, M., Zote, K. K., Singh, R. N., et al. (1989). Multilocational testing of pigeonpea for broad–based resistance to sterility mosaic in India. *Indian Phytopathol.* 42 (3), 444.

Nielsen, S. S., Brandt, W. E., and Singh, B. B. (1993). Genetic variability for nutritional composition and cooking time of improved cowpea lines. *Crop Sci.* 33 (3), 469–472. doi:10.2135/cropsci1993.0011183x003300030010x

Nisar, M., Ghafoor, A., Khan, M. R., and Qureshi, A. S. (2006). Screening of *Pisum sativum* L. Germplasm against *Erysiphe pisi* syd. *Acta Biol. Crac. Ser. Bot.* 48 (2), 33.

Nkhoma, N., Shimelis, H., Laing, M. D., Shayanowako, A., and Mathew, I. (2020). Assessing the genetic diversity of cowpea [Vigna unguiculata (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genet.* 21 (1), 110–116. doi:10.1186/s12863-020-00914-7

Okiror, M. A. (1999). Evaluation of pigeonpea (*Cajanus cajan*) germplasm for resistance to Fusarium wilt. *Indian J. Agric. Sci.* 69 (8), 600.

Olorunwa, O. J., Adhikari, B., Shi, A., and &Barickman, T. C. (2022). Screening of cowpea (Vigna unguiculata (L.) Walp.) genotypes for waterlogging tolerance using morpho-physiological traits at early growth stage. *Plant Sci.* 315, 111136. doi:10.1016/j.plantsci.2021.111136

Pal, S., Sharma, N. K., Sharma, A. K., Kumar, A., and Meena, S. (2020). Study of genetic variability parameters for seed yield and component traits in moth bean [*Vigna aconitifolia* (Jacq) Marechal]. *Int. J. Chem. Stud.* 8 (4), 2638–2641. doi:10.22271/chemi.2020.v8.i4ae.10038

Pande, S., Kishore, G. K., and Rao, J. N. (2004). Evaluation of chickpea lines for resistance to dry root rot caused by *Rhizoctonia*. *Int. Chickpea PigeonpeaNewslett.* 11 (2), 37.

Pande, S., Ramgopal, D., Kishore, G. K., Mallikarjuna, N., Sharma, M., Pathak, M., et al. (2006). Evaluation of wild *Cicer* species for resistance to Ascochyta blight and *Botrytis graymold* in controlled environment at ICRISAT, Patancheru, India. *J. Sat. Agric. Res.* 2 (1), 1.

Panhwar, F. (2005). *Chemistry and biochemistry of legume, with cultural activities in Sindh"*. Editor L. Chem (Germany: Digitalverlag GmbH).

Paredes, M., Becerra, V., and Tay, J. (2009). Inorganic nutritional composition of common bean (*Phaseolus vulgaris* L.) genotypes race Chile. *Chil. J. Agric. Res.* 69 (4), 486. doi:10.4067/s0718-58392009000400002

Parimala, K., Meenakumari, K. V. S., Sudhakar, R., and Durga, K. K. (2011). Screening of horsegram genotypes against yellow mosaic virus and powdery mildew diseases. *Indian J. Plant. Prot.* 39, 160.

Parween, S., Nawaz, K., Roy, R., Pole, A. K., Venkata, S., B., Misra, G., et al. (2015). An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5 (1), 12806–12814. doi:10.1038/srep12806

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

Pengelly, B. C., and Maass, B. L. (2001). *Lablab purpureus* (L.) Sweet–diversity, potential use and determination of a core collection of this multi–purpose tropical legume. *Genet. Resour. Crop Evol.* 48 (3), 261. doi:10.1023/A:1011286111384

Podder, R., Banniza, S., and Vandenberg, A. (2013). Screening of wild and cultivated lentil germplasm for resistance to stemphylium blight. *Plant Genet. Resour.* 11 (1), 26–35. doi:10.1017/s1479262112000329

Pootakham, W., Nawae, W., Naktang, C., Sonthirod, C., Yoocha, T., Kongkachana, W., et al. (2021). A chromosome-scale assembly of the blackgram (*Vigna mungo*) genome. *Mol. Ecol. Resour.* 21 (1), 238–250. doi:10.1111/1755-0998.13243

Porch, T. G., Cichy, K., Wang, W., Brick, M., Beaver, J. S., Santana–Morant, D., et al. (2017). Nutritional composition and cooking characteristics of tepary bean (*Phaseolus acutifolius* Gray) in comparison with common bean (*Phaseolus vulgaris* L.). *Genet. Resour. Crop Evol.* 64 (5), 935–953. doi:10.1007/s10722-016-0413-0

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 (2), 399–405. doi:10.1007/s10658-015-0692-x

Pratap, A., Basu, P. S., Gupta, S., Malviya, N., Rajan, N., Tomar, R., et al. (2014). Identification and characterization of sources for photo-and thermo-insensitivity in *Vigna* species. *Plant Breed.* 133 (6), 756–764. doi:10.1111/pbr.12215

Pratap, A., Gupta, S., Basu, P. S., Tomar, R., Dubey, S., Rathore, M., et al. (2019). "Towards development of climate smart mungbean: Challenges and opportunities," in *Genomic designing of climate–smart pulse crops*. Editor C. Kole (Cham): Springer), 235–264.

Pratap, A., Kumar, S., Polowick, P. L., Blair, M. W., and Baum, M. (2022). Editorial: Accelerating genetic gains in pulses. *Front. Plant Sci.* 13, 879377. doi:10.3389/fpls.2022.879377

Priyadharshini, B., Vignesh, M., Prakash, M., and Anandan, R. (2019). Evaluation of blackgram genotypes for saline tolerance at seedling stage. *Indian J. Agric. Res.* 53 (1), 83. doi:10.18805/ijare.a-5118

Priyanka, S., Sudhagar, R., Vanniarajan, C., and Ganesamurthy, K. (2021). Assessment of genetic divergence in horse gram (*Macrotylum uniflorum* (Lam) Verdc.) using quantitative traits. *Legume Res.* 44, 261–267. doi:10.18805/LR–4103

Rajendran, K., Smouni, A., Es–Safi, N. E., Benbrahim, N., Mentag, R., Nayyar, H., et al. (2020). Screening the FIGS set of lentil (*Lens culinaris* Medikus) germplasm for tolerance to terminal heat and combined drought–heat stress. *Agronomy* 10 (7), 1036. doi:10.3390/agronomy10071036

Rana, J. C., Banyal, D. K., Sharma, K. D., Sharma, M. K., Gupta, S. K., and Yadav, S. K. (2013). Screening of pea germplasm for resistance to powdery mildew. *Euphytica* 189 (2), 271–282. doi:10.1007/s10681-012-0798-6

Rana, J. C., Gautam, N. K., GayacharanSingh, M., Yadav, R., Tripathi, K., Yadav, S. K., et al. (2016). Genetic resources of pulse crops in India: An overview. *Indian J. Genet. Pl. Breed.* 76 (4), 420–436. doi:10.5958/0975-6906.2016.00061.4

Rana, J. C., Sharma, T. R., Tyagi, R. K., Chahota, R. K., Gautam, N. K., Singh, M., et al. (2015). Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. *Euphytica* 205 (2), 441–457. doi:10.1007/s10681-015-1406-3

Rana, M., Hossain, M., Urmi, T., Ahmed, S., Haque, M., and Islam, M. (2019). Evaluation of Blackgram (*Vigna mungo* L.) genotypes for their tolerance to flooding. *Agric.* 17 (1–2), 89–101. doi:10.3329/agric.v17i1–2.44699

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., et al. (2017). Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Mol. plant* 10 (8), 1047–1064. doi:10.1016/j.molp.2017.06.008

Ravelombola, W., Shi, A., Weng, Y., Mou, B., Motes, D., Clark, J., et al. (2018). Association analysis of salt tolerance in cowpea (*Vigna unguiculata* (L.) Walp) at

germination and seedling stages. *Theor. Appl. Genet.* 131 (1), 79–91. doi:10.1007/s00122-017-2987-0

Ravelombola, W. S., Shi, A., Weng, Y., Clark, J., Motes, D., Chen, P., et al. (2017). Evaluation of salt tolerance at germination stage in cowpea [*Vigna unguiculata* (L.) Walp]. *HortSci* 52 (9), 1168–1176. doi:10.21273/hortsci12195-17

Reddy, A. T., Gowda, J. E., Ravikumar, R., Rao, A. M., Ramesh, S., and Saifulla, M. (2016). Resistance source identification for dry root rot disease in chickpea. *Adv. Life Sci.* 5 (21), 9767.

Reddy, L. J., Upadhyaya, H. D., Gowda, C. L. L., and Singh, S. (2005). Development of core collection in pigeonpea [*Cajanus cajan* (L.) Millspaugh] using geographic and qualitative morphological descriptors. *Genet. Resour.CropEvol.* 52 (8), 1049–1056. doi:10.1007/s10722-004-6152-7

Rex Consortium (2016). Combining selective pressures to enhance the durability of disease resistance genes. *Front. Plant Sci.* 7, 1916. doi:10.3389/fpls.2016.01916

Rezene, Y., and Mekonin, S. (2019). Screening common bean (*Phaseolus vulgaris* L.) germplasm for resistance against angular leaf spot (*Pseudocercosporagriseola*) disease under field condition. *J. Plant Stud.* 8 (1), 30. doi:10.5539/jps.v8n1p30

Rodiño, A. P., Santalla, M., De Ron, A. M., and Singh, S. P. (2003). A core collection of common bean from the Iberian peninsula. *Euphytica* 131 (2), 165–175.

Roorkiwal, M., Bharadwaj, C., Barmukh, R., Dixit, G. P., Thudi, M., Gaur, P. M., et al. (2020). Integrating genomics for chickpea improvement: Achievements and opportunities. *Theor. Appl. Genet.* 133 (5), 1703–1720. doi:10.1007/s00122-020-03584-2

Roy, D., Joshi, P. K., and Chandra, R. (2017). Pulses for nutrition in India: Changing patterns from farm to fork. *Intl. Food Policy Res. Inst.*

Rubiales, D., Rojas-Molina, M. M., and Sillero, J. C. (2013). Identification of pre-and posthaustorial resistance to rust (*Uromycesviciae fabae*) in lentil (*Lens culinaris*) germplasm. *Plant Breed.* 132 (6), 676–680. doi:10.1111/pbr.12096

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 (3), 359.

Sachdeva, S., Sharma, V., Bhat, K. V., and Sen, S. (2016). Morphological and physiological response studies of moth bean (*Vigna Aconitifolia* L.) genotypes under drought stress. *Int. J. Appl. Sci. Eng.* 5 (7).

Sahoo, S., Sanadya, S. K., Kumari, N., and Baranda, B. (2019). Estimation of the various genetic variability parameters for seed yield and its component traits in Mothbean germplasm [Vigna aconitifolia (Jacq.) Marechal]. *J. Pharmcog. Phytochem* SP3, 49.

Sai, C. B., and Chidambaranathan, P. (2019). Reproductive stage drought tolerance in blackgram is associated with role of antioxidants on membrane stability. *Plant Physiol. Rep.* 24 (3), 399–409. doi:10.1007/s40502-019-00471-x

Sakai, H., Naito, K., Ogiso–Tanaka, E., Takahashi, Y., Iseki, K., Muto, C., et al. (2015). The power of single molecule real–time sequencing technology in the *de novo* assembly of a eukaryotic genome. *Sci. Rep.* 5 (1), 16780–16813. doi:10.1038/srep16780

Salini, K., Maruthi, V., Maheswari, M., and Sarkar, B. (2014). "Genetic improvement of horse gram through mutation breeding," in *3rd international conference on agriculture and horticulture* (India: Hyderabad International Convention Centre). October 27–29.

Sandhya, R. C., Rao, G. R., Chalam, M. S. V., Kumar, P. A., and Rao, V. S. (2014). Field screening of greengram genotypes against *Marucavitrata* in Summer. *J. Agric. Sci.* 1, 53.

Sankar, A. U., Anitha, K., Sivaraj, N., Kumari, K. V. S., Sunil, N., and Chakrabarty, S. K. (2015). Screening of horsegram germplasm collected from Andhra Pradesh against anthracnose. *Legume Res.–An Int. J.* 38 (6), 753–757. doi:10.18805/lr.v38i6.6719

Sarker, A., and Erskine, W. (2006). Recent progress in the ancient lentil. *J. Agric. Sci.* 144 (1), 19–29. doi:10.1017/s0021859605005800

Saxena, K. B., Chandrasena, G. D. S. N., Hettiarachchi, K., Iqbal, Y. B., Fonseka, H. H. D., and Jayasekera, S. J. B. A. (2002). Evaluation of pigeonpea accessions and selected lines for reaction to Maruca. *Crop Sci.* 42 (2), 615–618. doi:10.2135/cropsci2002.6150

Saxena, N. P., Krishnamurthy, L., and Johansen, C. (1993). Registration of a drought–resistant chickpea germplasm. *Crop Sci.* 33 (6), 1424. doi:10.2135/cropsci1993.0011183x003300060088x

Schafleitner, R., Nair, R. M., Rathore, A., Wang, Y. W., Lin, C. Y., Chu, S. H., et al. (2015). The AVRDC–The World Vegetable Center mungbean (*Vigna radiata*) core and mini core collections. *BMC Genomics* 16, 344. doi:10.1186/s12864-015-1556-7

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 (7), 707–713. doi:10.1038/ng.3008

Schreinemachers, P., Sequeros, T., Rani, S., Rashid, M. A., Gowdru, N. V., Rahman, M. S., et al. (2019). Counting the beans: Quantifying the adoption of improved mungbean varieties in South Asia and Myanmar. *Food secur.* 11 (3), 623–634. doi:10.1007/s12571-019-00926-x

Serraj, R., Krishnamurthy, L., and Upadhyaya, H. D. (2004). Screening chickpea mini–core germplasm for tolerence to soil salinity. *Int. Chickpea PigeonpeaNewslett.* 11, 29–32.

Serrano, C., Carbas, B., Castanho, A., Soares, A., Patto, M. C. V., and Brites, C. (2017). Characterisation of nutritional quality traits of a chickpea (*Cicer arietinum*) germplasm collection exploited in chickpea breeding in Europe. *Crop Pasture Sci.* 68 (11), 1031. doi:10.1071/CP17129

Shad, N., Mughal, S. M., Farooq, K., and Bashir, M. (2006). Evaluation of mungbean germplasm for resistance against mungbean yellow mosaic begomovirus. *Pak. J. Bot.* 38 (2), 449.

Shafiq, S., Mather, D. E., Ahmad, M., and Paull, J. G. (2012). Variation in tolerance to radiant frost at reproductive stages in field pea germplasm. *Euphytica* 186 (3), 831–845. doi:10.1007/s10681-012-0625-0

Shaheen, S., Harun, N., Khan, F., Hussain, R. A., Ramzan, S., Rani, S., et al. (2012). Comparative nutritional analysis between *Vigna radiata* and *Vigna mungo* of Pakistan. *Afr. J. Biotechnol.* 11 (25), 6694. doi:10.5897/AJB11.3496

Shanthi, P., Ganesan, K. N., Manivannan, N., and Natarajan, C. (2019). Correlation and path analysis in Blackgram (Vigna mungo L.). *Electron. J. Plant Breed.* 10 (3), 1218–1222. doi:10.5958/0975-928x.2019.00154.6

Sharma, A., and Chahota, R. K. (2022). Comparative analysis of biochemical parameters in response to stress in horsegram (*Macrotyloma uniflorum* L.) germplasm. *Himachal J. Agric. Res.* 47 (2), 143–148.

Sharma, H. C., Mukuru, S. Z., Manyasa, E., and Were, J. W. (1999). Breakdown of resistance to sorghum midge, Stenodiplosissorghicola. *Euphytica* 109 (2), 131–140. doi:10.1023/a:1003724217514

Sharma, L., Priya, M., Bindumadhava, H., Nair, R. M., and Nayyar, H. (2016). Influence of high temperature stress on growth, phenology and yield performance of mungbean [*Vigna radiata* (L.) Wilczek] under managed growth conditions. *Sci. Hortic.* 213, 379–391. doi:10.1016/j.scienta.2016.10.033

Sharma, M., Ghosh, R., and Pande, S. (2015). Dry root rot (*Rhizoctonia bataticola* (taub.) butler): An emerging disease of chickpea–where do we stand? *Arch. Phytopathol. Plant Prot.* 48, 797–812. doi:10.1080/03235408.2016.1140564

Sharma, M., Rathore, A., Mangala, U. N., Ghosh, R., Sharma, S., Upadhyay, H. D., et al. (2012). New sources of resistance to Fusarium wilt and sterility mosaic disease in a mini–core collection of pigeonpea germplasm. *Eur. J. Plant Pathol.* 133 (3), 707–714. doi:10.1007/s10658-012-9949-9

Sharma, M, M., Telangre, R., Ghosh, R., and Pande, S. (2015). Multi–environment field testing to identify broad, stable resistance to sterility mosaic disease of pigeonpea. *J. Gen. Plant Pathol.* 81 (3), 249–259. doi:10.1007/s10327-015-0585-z

Sharma, S, S., Singh, N., Virdi, A. S., and Rana, J. C. (2015). Himalayan kidney bean germplasm: Grain-flour characteristics, structural-functional properties and *in-vitro* digestibility of starches. *Food Res. Int.* 77, 498–505. doi:10.1016/j.foodres.2015.08.030

Shirasawa, K., Chahota, R., Hirakawa, H., Nagano, S., Nagasaki, H., Sharma, T., et al. (2021). A Chromosome–scale draft genome sequence of horsegram (*Macrotyloma uniflorum*). *Gigabyte* 2021, 1–23. doi:10.46471/gigabyte.30

Silim, S. N., Gwataa, E. T., Coeb, R., and Omanga, P. A. (2007). Response of pigeonpea genotypes of differrent maturity duration to temperature and photoperiod in Kenya. *Afr. Crop Sci. J.* 15 (2). doi:10.4314/acsj.v15i2.54420

Silva, C. A., Abreu, Â. D. F. B., Ramalho, M. A. P., and Maia, L. G. S. (2012). Chemical composition as related to seed color of common bean. *Crop Breed. Appl. Biotechnol.* 12, 132–137. doi:10.1590/s1984-70332012000200006

Simon, C. J., and Hannan, R. M. (1995). Development and use of core subsets of cool–season food legume germplasm collections. *Hort. Sci.* 30 (4), 907C–907C. doi:10.21273/hortsci.30.4.907c

Singh, D. P. (1995). Lentil germplasm: Evaluation and utilization, a review. *Indian J. Genet. Plant Breed.* 55 (02), 173.

Singh, D. P., and Singh, B. B. (2011). Breeding for tolerance to abiotic stresses in mungbean. *J. Food Leg.* 24 (2), 83.

Singh, K. B. (1997). Chickpea (*Cicer arietinum* L.). *Field Crops Res.* 53, 161–170. doi:10.1016/s0378-4290(97)00029-4

Singh, K. B., Malhotra, R. S., and Saxena, M. C. (1989). Chickpea evaluation for cold tolerance under field conditions. *Crop Sci.* 29 (2), 282–285. doi:10.2135/cropsci1989.0011183x002900020009x

Singh, K., Shekhawat, N., Singh, O. V., Choudhary, M., and Ram, D. (2020). Identification for sources of resistance to biotic stresses in mothbean [*Vigna aconitifolia* (Jaeq.) Marcchal] Germplasm. *Int. J. Curr. Microbiol. Appl. Sci.* 9 (5), 1243–1249. doi:10.20546/ijcmas.2020.905.138

Singh, M., Dahiya, O. P., Singh, N., Nizar, A., Dutta, M., Tyagi, R. K., et al. (2013). Identification of an upright peduncle and podding genotype in chickpea germplasm conserved in the National Genebank. *Indian J. Plant Genet. Resour.* 26 (2), 166–168.

Singh, N., Kaur, N., Rana, J. C., and Sharma, S. K. (2010). Diversity in seed and flour properties in field pea (*Pisum sativum*) germplasm. *Food Chem.* 122 (3), 518–525. doi:10.1016/j.foodchem.2010.02.064

Singh, N. K., Gupta, D. K., Jayaswal, P. K., Mahato, A. K., Dutta, S., Singh, S., et al. (2012). The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol.* 21 (1), 98–112. doi:10.1007/s13562-011-0088-8

Singh, R. K., Singh, C., Chandana, B. S., Mahto, R. K., Patial, R., Gupta, A., et al. (2022). Exploring chickpea germplasm diversity for broadening the genetic base utilizing genomic resourses. *Front. Genet.* 13, 905771. doi:10.3389/fgene.2022.905771

Singh, R., Salam, J. L., Mandavi, N. C., Saxena, R. R., and Rao, A. (2019). Genetic diversity estimation in horsegram [*Macrotyloma uniflorum* (L) Verdcout] genotypes collected from Bastar plateau. *Int. J. Curr. Microbiol. App. Sci.* 8 (12), 613–620. doi:10.20546/ijcmas.2019.812.080

Singh, J, J., Kanaujia, R., Sivastava, A. K., Dixit, G. P., and Singh, N. P. (2017). Genetic variability for iron and zinc as well as antinutrients affecting bioavailability in blackgram (*Vigna mungo* (L.) Hepper). *J. Food Sci. Technol.* 54 (4), 1035–1042. doi:10.1007/s13197-017-2548-1

Singh, N, N., Nongthombam, R., SeyieKonjengbamKhoyumthem, K. N. P., and Naorem, B. (2016). Stability analysis of high yielding varieties of blackgram (*Vigna Mungo* L. Hepper). *Int. J. Agric. Environ. Biotechnol.* 9, 17–24. doi:10.5958/2230–732X.2016.00003.6

Singh, N. P, N. P., Praharaj, C. S., and Sandhu, J. S. (2016). Utilizing untapped potential of rice fallow of East and North–east India through pulse production. *Indian J. Genet. Plant Breed.* 76 (4), 388–398. doi:10.5958/0975-6906.2016.00058.4

Singh, S, S., Gupta, V., Singh, S. P., and Yadava, N. S. (2017). Growth and productivity of Moth bean [*Vigna aconitifolia* (Jacq.) Marechal] in response to different varieties and phosphorus levels. *J. Pharm. Phytochem.* 6 (3), 811. doi:10.1007/s13197-017-2548-1

Sinha, P., Singh, V. K., Saxena, R. K., Khan, A. W., Abbai, R., Chitikineni, A., et al. (2020). Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (Cajanus cajan L.). *Plant Biotechnol. J.* 18 (12), 2482–2490. doi:10.1111/pbi.13422

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Legare, 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

Souleymane, A., Aken'Ova, M. E., Fatokun, C. A., and Alabi, O. Y. (2013). Screening for resistance to cowpea aphid (*Aphis craccivora* Koch) in wild and cultivated cowpea (*Vigna unguiculata* L. Walp.) accessions. *Int. J. Sci. Environ. Technol.* 2, 611.

Soundararajan, R. P., Chitra, N., Geetha, S., and Poorani, J. (2012). Biological control of bruchid *Callosobruchus maculatus* (F.) in blackgram. *J. Biopestic.* 5, 192.

Souter, J. R., Gurusamy, V., Porch, T. G., and Bett, K. E. (2017). Successful introgression of abiotic stress tolerance from wild tepary bean to common bean. *Crop Sci.* 57 (3), 1160–1171. doi:10.2135/cropsci2016.10.0851

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., et al. (2013). Erratum to: Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* 126, 201–202. doi:10.1007/s00122-015-2618-6

Srinivasan, A., Johansen, C., and Saxena, N. P. (1998). Cold tolerance during early reproductive growth of chickpea (*Cicer arietinum* L.): Characterization of stress and genetic variation in pod set. *Field Crops Res.* 57 (2), 181–193. doi:10.1016/s0378-4290(97)00118-4

Stoddard, F. L., Balko, C., Erskine, W., Khan, H. R., Link, W., and Sarker, A. (2006). Screening techniques and sources of resistance to abiotic stresses in cool-season food legumes. *Euphytica* 147 (1), 167–186. doi:10.1007/s10681-006-4723-8

Subramaniyan, R., Narayana, M., Krishnamoorthy, I., Natarajan, G., and Gandhi, K. (2022). Novel and stable QTL regions conferring resistance to MYMV disease and its inheritance in blackgram (*Vigna mungo* (L.) Hepper). *J. Genet.* 101 (1), 18–27. doi:10.1007/s12041-022-01359-w

Sudhakaran, S. M. N., and Bukkan, D. S. (2021). A review on nutritional composition, antinutritional components and health benefits of greengram (*Vigna radiata* (L.) Wilczek). *J. Food Biochem.* 45 (6), e13743. doi:10.1111/jfbc.13743

Sultana, R., Saxena, K. B., Kumar, R. R., Kumar, D., and Kirti, M. (2021). *Pigeonpea" in* the Beans and the peas. Woodhead Publishing, 217–240.

Sultana, R., Vales, M. I., Saxena, K. B., Rathore, A., Rao, S., Rao, S. K., et al. (2013). Waterlogging tolerance in pigeonpea (*Cajanus cajan* (L.) millsp.): Genotypic variability and identification of tolerant genotypes. *J. Agric. Sci.* 151 (5), 659–671. doi:10.1017/s0021859612000755

Summo, C., De Angelis, D., Ricciardi, L., Caponio, F., Lotti, C., Pavan, S., et al. (2019). Nutritional, physico–chemical and functional characterization of a global chickpea collection. *J.Food Compos. Anal.* 84, 103306. doi:10.1016/j.jfca.2019.103306

Suneja, Y., Kaur, S., Gupta, A. K., and Kaur, N. (2011). Levels of nutritional constituents and antinutritional factors in blackgram (*Vigna mungo* L. Hepper). *Food Res. Int.* 44 (2), 621–628. doi:10.1016/j.foodres.2010.12.020

Talari, A., and Shakappa, D. (2018). Role of pigeon pea (Cajanus cajan L.) in human nutrition and health: A review. *Asian J. Dairy Food Res.* doi:10.18805/AJDFR.DR-1379

Teshome, A., Mendesil, E., Geleta, M., Andargie, D., Anderson, P., Rämert, B., et al. (2015). Screening the primary gene pool of field pea (*Pisum sativum* L. subsp. sativum) in Ethiopia for resistance against pea weevil (*Bruchuspisorum* L.). *Genet. Resour. Crop Evol.* 62 (4), 525–538. doi:10.1007/s10722-014-0178-2

Tripathi, K., Gore, P. G., Bansal, R., Gayacharan, C., Shubha, K., Kumar, V., et al. (2021b). Identification and revealing the potential traits of the unique germplasm with extended funiculus in pea (*Pisum sativum* L.). *Genet. Res. Crop Evol.* 68 (8), 3125–3132. doi:10.1007/s10722-021-01275-5

Tripathi, K., Kumari, J., Gore, P. G., Mishra, D. C., Singh, A. K., Mishra, G. P., et al. (2021a). Agro–morphological characterization of lentil germplasm of Indian National genebank and development of a core set for efficient utilization in lentil improvement programs. *Front. Plant Sci.* 12, 751429. doi:10.3389/fpls.2021.751429

Tripathi, K., Prasad, T. V., Bhardwaj, R., Jha, S. K., Semwal, D. P., Gore, P. G., et al. (2020). Evaluation of diverse germplasm of cowpea [*Vigna unguiculata* (L.) Walp.] against bruchid [*Callosobruchus maculatus* (Fab.)] and correlation with physical and biochemical parameters of seed. *Plant Genet. Resour.* 18 (3), 120–129. doi:10.1017/s1479262120000180

Upadhyaya, H. D., Bajaj, D., Srivastava, R., Daware, A., Basu, U., Tripathi, S., et al. (2017). Genetic dissection of plant growth habit in chickpea. *Funct. Integr. Genomics* 17 (6), 711–723. doi:10.1007/s10142-017-0566-8

Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* 41 (1), 206–210. doi:10.2135/cropsci2001.411206x

Upadhyaya, H. D., Dronavalli, N., Gowda, C. L. L., and Singh, S. (2011a). Identification and evaluation of chickpea germplasm for tolerance to heat stress. *Crop Sci.* 51 (5), 2079–2094. doi:10.2135/cropsci2011.01.0018

Upadhyaya, H. D., Dwivedi, S. L., Gowda, C. L. L., and Singh, S. (2007b). Identification of diverse germplasm lines for agronomic traits in a chickpea (*Cicer arietinum* L.) core collection for use in crop improvement. *Field Crops Res.* 100 (2–3), 320–326. doi:10.1016/j.fcr.2006.08.008

Upadhyaya, H. D., and Ortiz, R. (2001). A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102 (8), 1292–1298. doi:10.1007/s00122-001-0556-y

Upadhyaya, H. D., Ortiz, R., Bramel, P. J., Singh, S., et al. (2002). Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. *Euphytica* 123, 333–342. doi:10.1023/a:1015088417487

Upadhyaya, H. D., Reddy, K. N., Sastry, D. V. S. S. R., and Gowda, C. L. L. (2007c). Identification of photoperiod insensitive sources in the world collection of pigeonpea at ICRISAT. *J. Sat. Agric. Res.* 3 (1), 1.

Upadhyaya, H. D., Reddy, K. N., Sharma, S., Varshney, R. K., Bhattacharjee, R., Singh, S., et al. (2011b). Pigeonpea composite collection and identification of germplasm for use in crop improvement programmes. *Plant Genet. Resour.* 9 (1), 97–108. doi:10.1017/s1479262110000419

Upadhyaya, H. D., Reddy, L. J., Gowda, C. L. L., Reddy, K. N., and Singh, S. (2006). Development of a mini core subset for enhanced and diversified utilization of pigeonpea germplasm resources. *Crop Sci.* 46 (5), 2127–2132. doi:10.2135/cropsci2006.01.0032

Upadhyaya, H. D., Salimath, P. M., Gowda, C. L. L., and Singh, S. (2007a). New early–maturing germplasm lines for utilization in chickpea improvement. *Euphytica* 157 (1), 195–208. doi:10.1007/s10681-007-9411-9

Upadhyaya, H. D., Yadav, D., Dronavalli, N., Gowda, C. L. L., and Singh, S. (2010). Mini core germplasm collections for infusing genetic diversity in plant breeding programs. *Electron. J. Plant Breed.* 1 (4), 1294.

Urrea, C. A., and Harveson, R. M. (2014). Identification of sources of bacterial wilt resistance in common bean (*Phaseolus vulgaris*). *Plant Dis.* 98 (7), 973–976. doi:10.1094/PDIS-04-13-0391-RE

Urrea, C. A., and Porch, T. (2009). *Phenotypic evaluation of a subset of the* Phaseolusvulgaris *core collections and the* P. acutifolius *germplasm collection, and advanced common bean lines for drought tolerance in Nebraska. Cooperative* 104

Van de Wouw, A. P., Sheedy, E. M., Ware, A. H., Marcroft, S. J., and &Idnurm, A. (2022). Independent breakdown events of the Brassica napus Rlm7 resistance gene including via the off-target impact of a dual-specificity avirulence interaction. *Mol. Plant Pathol.* 23, 997–1010. doi:10.1111/mpp.13204

Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. K. (2021). Designing future crops: Genomics-assisted breeding comes of age. *Trends Plant Sci.* 26 (6), 621–630. doi:10.1016/j.tplants.2005.10.004

Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2011). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource–poor farmers. *Nat. Biotechnol.* 30 (1), 83–89. doi:10.1038/nbt.2022

Varshney, R. K. (2016). Exciting journey of 10 years from genomes to fields and markets: Some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* 242, 98–107. doi:10.1016/j.plantsci.2015.09.009

Varshney, R. K., Gaur, P. M., Chamarthi, S. K., Krishnamurthy, L., Tripathi, S., Kashiwagi, J., et al. (2013b). Fast-track introgression of "QTL-hotspot" for root traits and other drought tolerance traits in JG 11, an elite and leading variety of chickpea. *Plant Genome* 6 (3), plantgenome2013–07. doi:10.3835/plantgenome2013.07.0022

Varshney, R. K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., et al. (2021). A chickpea genetic variation map based on the sequencing of 3, 366 genomes. *Nature* 599 (7886), 622–627. doi:10.1038/s41586-021-04066-1

Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082–1088. doi:10.1038/ng.3872

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013a). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 1 (3), 240–246. doi:10.1038/nbt.2491

Vir, O., and Singh, A. K. (2015). Moth bean [*Vigna aconitifolia* (Jacq.) Marechal] germplasm: Evaluation for genetic variability and inter characters relationship in hot arid climate of Western Rajasthan, India. *Legume Res.* 38 (6), 748–752.

Vishnu, B., Jayalakshmi, V., Rani, M. S., Vlasova, A., Capella–Gutiérrez, S., Rendón–Anaya, M., et al. (2020). Genetic diversity studies among chickpea (Cicer arietinum L.) genotypes under rainfed and irrigated conditions for yield attributing and traits related to mechanical harvestingGenome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and

temporal specialization of genes. *Legume Research-An Int. JournalGenome Biol.* 4317 (21), 321–19418. doi:10.1186/s13059-016-0883-6

Wang, J., Li, Y., Li, A., Liu, R. H., Gao, X., Li, D., et al. (2021). Nutritional constituent and health benefits of chickpea (*Cicer arietinum* L.): A review. *Food Res. Int.* 150, 110790. doi:10.1016/j.foodres.2021.110790

White, P. J., Bradshaw, J. E., Finlay, M., Dale, B., Ramsay, G., Hammond, J. P., et al. (2009). Depolarization-activated calcium channels shape the calcium signatures induced by low-temperature stress. *HortSci* 44 (1), 6–8. doi:10.1111/j.1469-8137.2009.02857.x

WIEWS (2022). WIEWS - World Information and Early Warning System on Plant Genetic Resources for Food and Agriculture Available at: https://www.fao.org/wiews/data/ex–situ–sdg–251/search/en/?no_cache=1 (Accessed on February 2, 2022).

Wisser, R. J., Oppenheim, S. J., Ernest, E. G., Mhora, T. T., Dumas, M. D., Gregory, N. F., et al. (2021). Genome assembly of a mesoamerican derived variety of lima bean: A foundational cultivar in the mid–atlantic USA. *G3* 11 (11), jkab207. doi:10.1093/g3journal/jkab207

Xia, Q., Pan, L., Zhang, R., Ni, X., Wang, Y., Dong, X., et al. (2019). The genome assembly of asparagus bean, *Vigna unguiculata* ssp. *Sci. Data* 6 (1), 124. doi:10.1038/s41597-019-0130-6

Xu, H., Jing, T., Tomooka, N., Kaga, A., Isemura, T., and Vaughan, D. A. (2008). Genetic diversity of the azuki bean (*Vigna angularis* (Willd.) Ohwiand Ohashi) gene pool as assessed by SSR markers. *Genome* 51 (9), 728–738. doi:10.1139/G08-058

Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M. S., et al. (2022). Smart breeding driven by big data, artificial intelligence and integrated genomic-enviromic prediction. *Mol. Plant* 15, 1664–1695. doi:10.1016/j.molp.2022.09.001

Xu–Xiao, Z., Jian–Ping, G., Wang, S., Liu, Q., Redden, R. R., and Ford, R. (2008). Genetic diversity and core collection of alien *Pisum sativum* L. germplasm. *Acta Agron. Sin.* 34 (9), 1518–1528. doi:10.1016/s1875-2780(09)60003-1

Yadav, S., Negi, K. S., and Mandal, S. (2004). Protein and oil rich wild horsegram. *Genet. Resour. Crop Evol.* 51 (6), 629–633. doi:10.1023/b:gres.0000024650.44318.2d

Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., et al. (2015). Genome sequencing of adzuki bean (*Vigna angularis*) provides insight into high starch and lowfat accumulation and domestication. *Proc. Natl. Acad. Sci.* 112 (43), 13213–13218. doi:10.1073/pnas.1420949112

Yasin, J. K., Nizar, M. A., Rajkumar, S., Verma, M., Verma, N., Pandey, S., et al. (2014). Existence of alternate defense mechanisms for combating moisture stress in horse gram [*Macrotyloma uniflorum* (Lam.) Verdc.]. *Legume Res. Int. J.* 37 (2), 145–154. doi:10.5958/j.0976-0571.37.2.022

Yohane, E. N., Shimelis, H., Laing, M., Mathew, I., and Shayanowako, A. (2020). Phenotypic divergence analysis in pigeonpea [*Cajanus cajan* (L.) Millspaugh] germplasm accessions. *Agronomy* 10 (11), 1682. doi:10.3390/agronomy10111682

Zaccardelli, M., Lupo, F., Piergiovanni, A. R., Laghetti, G., Sonnante, G., Daminati, M. G., et al. (2012). Characterization of Italian lentil (*Lens culinaris*Medik.) germplasm by agronomic traits, biochemical and molecular markers. *Genet. Resour. Crop Evol.* 59 (5), 727–738. doi:10.1007/s10722-011-9714-5

Zeven, A. C., and Zhukovsky, P. M. (1975). *Dictionary of cultivated plants and their centres of diversity*. Wageningen: PUDOC.

Zhang, H., Zhang, D., Wang, M., Sun, J., Qi, Y., Li, J., et al. (2011). A core collection and mini core collection of Oryza sativa L. in China. *Theor. Appl. Genet.* 122, 49–61. doi:10.1007/s00122-010-1421-7

Zhang, X., Wan, S., Hao, J., Hu, J., Yang, T., and Zong, X. (2016). Large–scale evaluation of pea (*Pisum sativum* L.) germplasm for cold tolerance in the field during winter in Qingdao. *Crop J* 4 (5), 377–383. doi:10.1016/j.cj.2016.06.016

Zhu, Z., Chen, H., Liao, F., Li, L., Liu, C., Liu, L., et al. (2019). Evaluation and screening of adzuki bean germplasm resources for drought tolerance during germination stage. *South. J. Agric. Econ.* 50 (6), 1183–1190.

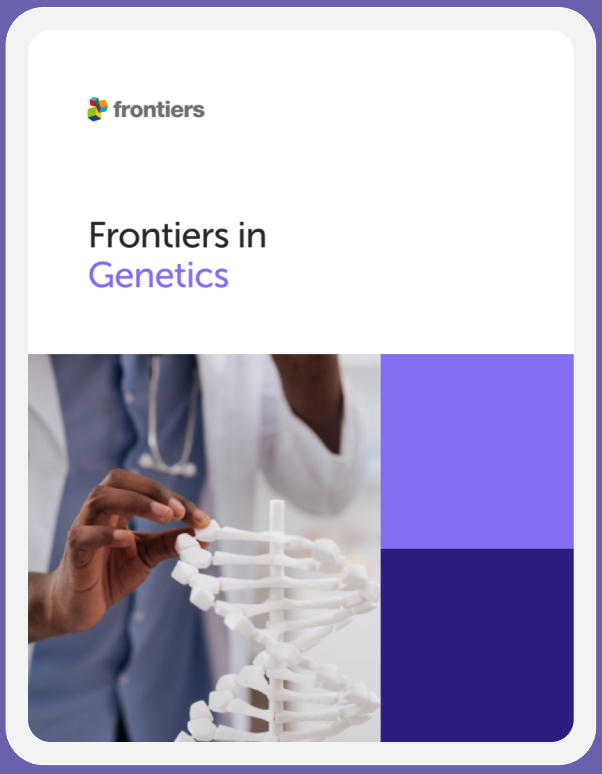
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact