frontiers | Research Topics

Biotechnological and genomic approaches for enhancing agronomic performance of crops

Edited by
Sreepriya Pramod, Ramsey Steven Lewis, Reyazul Rouf Mir and Himabindu Kudapa

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 978-2-83251-840-3
DOI 10.3389/978-2-83251-840-3

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

Biotechnological and genomic approaches for enhancing agronomic performance of crops

Topic editors

Sreepriya Pramod — Altria, United States

Ramsey Steven Lewis — North Carolina State University, United States

Reyazul Rouf Mir — Sher-e-Kashmir University of Agricultural Sciences and Technology, India

Himabindu Kudapa — International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Citation

Pramod, S., Lewis, R. S., Mir, R. R., Kudapa, H., eds. (2023). *Biotechnological and genomic approaches for enhancing agronomic performance of crops*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-840-3

Topic Editor Sreepriya Pramod is employed by the company Altria and holds patents related to the Research Topic subject. Topic Editor Ramsey Lewis has received funding from private companies and holds patents related to the Research Topic subject. All other Topic Editors declare no competing interests

Table of contents

05 Editorial: Biotechnological and genomic approaches for enhancing agronomic performance of crops
Reyazul R. Mir, Himabindu Kudapa, Sreepriya Pramod and Ramsey S. Lewis

13 Discovery of miRNAs and Development of Heat-Responsive miRNA-SSR Markers for Characterization of Wheat Germplasm for Terminal Heat Tolerance Breeding
Pooja Sihag, Vijeta Sagwal, Anuj Kumar, Priyanka Balyan, Reyazul Rouf Mir, Om Parkash Dhankher and Upendra Kumar

25 Precision Mapping of a Maize MAGIC Population Identified a Candidate Gene for the Senescence-Associated Physiological Traits
Marlon Caicedo, Eduardo D. Munaiz, Rosa A. Malvar, José C. Jiménez and Bernardo Ordas

38 Omics Technologies to Enhance Plant Based Functional Foods: An Overview
Spurthi N. Nayak, B. Aravind, Sachin S. Malavalli, B. S. Sukanth, R. Poornima, Pushpa Bharati, Kathleen Hefferon, Chittaranjan Kole and Naveen Puppala

56 Utilizing CRISPR-Cas in Tropical Crop Improvement: A Decision Process for Fitting Genome Engineering to Your Species
Kathleen A. Joo, Michael G. Muszynski, Michael B. Kantar, Ming-Li Wang, Xiaoling He and Angel R. Del Valle Echevarria

64 Comprehensive *In Silico* Characterization and Expression Profiling of *TCP* Gene Family in Rapeseed
Yunfei Wen, Ali Raza, Wen Chu, Xiling Zou, Hongtao Cheng, Qiong Hu, Jia Liu and Wenliang Wei

82 Genome-Wide Identification and Expression Analysis of the MADS-Box Gene Family in Sweet Potato [*Ipomoea batatas* (L.) Lam]
Zhengwei Shao, Minhong He, Zhipeng Zeng, Yanzhu Chen, Amoanimaa-Dede Hanna and Hongbo Zhu

96 Genome-Wide Analysis of the Fatty Acid Desaturase Gene Family Reveals the Key Role of *PfFAD3* in α-Linolenic Acid Biosynthesis in Perilla Seeds
Wu Duan, Yang Shi-Mei, Shang Zhi-Wei, Xu Jing, Zhao De-Gang, Wang Hong-Bin and Shen Qi

109 Genome-Wide Identification and Co-Expression Analysis of *ARF* and *IAA* Family Genes in *Euscaphis konishii*: Potential Regulators of Triterpenoids and Anthocyanin Biosynthesis
Bobin Liu, Juanli Zhu, Lina Lin, Qixin Yang, Bangping Hu, Qingying Wang, Xiao-Xing Zou and Shuang-Quan Zou

125 Genome-Wide Diversity of MADS-Box Genes in Bread Wheat is Associated with its Rapid Global Adaptability
Qasim Raza, Awais Riaz, Rana Muhammad Atif, Babar Hussain, Iqrar Ahmad Rana, Zulfiqar Ali, Hikmet Budak and Ibrahim A. Alaraidh

140 Comparative Transcriptome Analysis in *Taraxacum koksaghyz* to Identify Genes that Determine Root Volume and Root Length
Annika Wieghaus, Kai-Uwe Roelfs, Richard M. Twyman, Dirk Prüfer and Christian Schulze Gronover

154 Genome-Wide Evolution and Comparative Analysis of Superoxide Dismutase Gene Family in Cucurbitaceae and Expression Analysis of *Lagenaria siceraria* Under Multiple Abiotic Stresses
Shamsur Rehman, Arif Rashid, Muhammad Aamir Manzoor, Lingling Li, Weibo Sun, Muhammad Waheed Riaz, Dawei Li and Qiang Zhuge

174 CerealESTDb: A Comprehensive Resource for Abiotic Stress-Responsive Annotated ESTs With Predicted Genes, Gene Ontology, and Metabolic Pathways in Major Cereal Crops
Sanjeev Kumar, Jyotika Bhati, Arijit Saha, Shashi Bhushan Lal, Pankaj Kumar Pandey, Dwijesh Chandra Mishra, Mohammad Samir Farooqi, Anuj Kumar, Krishna Kumar Chaturvedi and Anil Rai

184 Global Survey, Expressions and Association Analysis of *CBLL* Genes in Peanut
Weifang Ren, Zhaocong Zeng, Sijian Wang, Juncheng Zhang, Jiahai Fang and Liyun Wan

197 Genome-Wide Identification and Expression Analysis of the Thioredoxin (*Trx*) Gene Family Reveals Its Role in Leaf Rust Resistance in Wheat (*Triticum aestivum* L.)
Ramesh Bhurta, Deepak T. Hurali, Sandhya Tyagi, Lekshmy Sathee, Sandeep Adavi B, Dalveer Singh, Niharika Mallick, Viswanathan Chinnusamy, Vinod and Shailendra K. Jha

216 An Update on Resistance Genes and Their Use in the Development of Leaf Rust Resistant Cultivars in Wheat
Kuldeep Kumar, Irfat Jan, Gautam Saripalli, P. K. Sharma, Reyazul Rouf Mir, H. S. Balyan and P. K. Gupta

227 Genotyping-by-Sequencing Based Molecular Genetic Diversity of Pakistani Bread Wheat (*Triticum aestivum* L.) Accessions
Shabbir Hussain, Madiha Habib, Zaheer Ahmed, Bushra Sadia, Amy Bernardo, Paul St. Amand, Guihua Bai, Nida Ghori, Azeem I. Khan, Faisal S. Awan and Rizwana Maqbool

241 Novel GBS-Based SNP Markers for Finger Millet and Their Use in Genetic Diversity Analyses
Haftom Brhane, Teklehaimanot Haileselassie, Kassahun Tesfaye, Rodomiro Ortiz, Cecilia Hammenhag, Kibrom B. Abreha and Mulatu Geleta

258 Heterologous Expression of Arabidopsis *AtARA6* in Soybean Enhances Salt Tolerance
Zhipeng Hong, Yang Li, Yang Zhao, Mingyu Yang, Xiaoming Zhang, Yuhan Teng, Linjie Jing, Danxun Kong, Tongxin Liu, Shuanglin Li, Fanli Meng, Qi Wang and Ling Zhang

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 October 2022
DOI 10.3389/fgene.2022.991630

Check for updates

OPEN ACCESS

EDITED BY
Zefeng Yang,
Yangzhou University, China

REVIEWED BY
Junliang Zhao,
Guangdong Academy of Agricultural Sciences (GDAAS), China

*CORRESPONDENCE
Reyazul R. Mir,
imrouf2006@gmail.com

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

RECEIVED 11 July 2022
ACCEPTED 15 September 2022
PUBLISHED 04 October 2022

CITATION
Mir RR, Kudapa H, Pramod S and Lewis RS (2022), Editorial: Biotechnological and genomic approaches for enhancing agronomic performance of crops.
Front. Genet. 13:991630.
doi: 10.3389/fgene.2022.991630

COPYRIGHT
© 2022 Mir, Kudapa, Pramod and Lewis. 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: Biotechnological and genomic approaches for enhancing agronomic performance of crops

Reyazul R. Mir[1]*, Himabindu Kudapa[2], Sreepriya Pramod[3] and Ramsey S. Lewis[4]

[1]Division of Genetics and Plant Breeding, Faculty of Agriculture, Sher-e-Kashmir University of Agricultural Sciences and Technology Srinagar, Srinagar, India, [2]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India, [3]Altria Client Services LLC, Center for Research and Technology, Richmond, VA, United States, [4]Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC, United States

KEYWORDS

biotechnology and bioinformatics, genomics (genetic tests and databases), crop improvement, gene, markers

Editorial on the Research Topic
Biotechnological and Genomic Approaches for Enhancing Agronomic Performance of Crops

Why do we need to care and what is the solution?

The knowledge of plant genetics and plant genomics is very important in advancing our understanding of the fundamental processes in the science of plant biology. Different biotic and abiotic stresses coupled with changing climatic conditions cause economic losses in most major crops worldwide. Pioneering strategies for improving plant health and defense mechanisms can inform vital priorities for maintaining agricultural production and supply a rapidly expanding global population. Despite the advances made in plant breeding, there is a need to support our cultivar development programs to understand the limitations in productivity and to direct efforts towards gaining precise understanding of the target genes responsible for controlling the traits of interest in any given crop. The advent of advanced genomic tools and technologies can now enable developing high-yield, stress-tolerant crop varieties (*see* Raza et al., 2021; Pazhamala et al., 2021).

The science of plant breeding is most important and helps in delivering new varieties with novel trait combinations and improved agronomic performance. The science of plant breeding is continuously evolving and one of the reasons for this continual progress is the integration of biotechnological/genomics tools and techniques with plant breeding. Great

success has been achieved over the years in trait phenotyping, genotyping approaches/methods, and biotechnological approaches and their use in crop improvement programs (Mir et al., 2015, 2019). The progress and new advancements made in these research areas continue to reduce the time and cost from discovery of genes/quantitative trait loci (QTLs) for traits of interest followed by their deployment in existing improved cultivars through marker-assisted selection (MAS). The broader research community has provided a wealth of genomics/transcriptomics resources, as evident by the fact that an increasing number of high-quality reference or draft genomes, pan-genomes, transcriptomes, and molecular markers are becoming available for almost all crop plants.

A variety of different kinds of molecular markers have been developed in almost all important crop plants in the world (Gupta et al., 2008, 2013a; Mir et al., 2013a,b; Mir and Varshney 2013; Tyagi et al., 2019, 2021; Kumar et al., 2021; Sagwal et al., 2022). More new types of molecular markers have been discovered and used for crop improvement programs. For instance, molecular markers have been used extensively in different activities including the study of genetic diversity of different crop germplasm resources (Banerjee et al., 2012; Choudhary et al., 2021; Shafi et al., 2021; Tahir et al., 2021), study of population structure, development of high-density genetic/linkage maps, and mapping genes through traditional quantitative trait loci (QTL) mapping/family mapping and somewhat modern genome-wide association studies (GWAS)/population mapping (Gupta et al., 2013b). In addition, these markers have been also used extensively in the development of improved crop varieties through modern breeding approaches like marker-assisted selection (MAS)/marker-assisted gene pyramiding, marker-assisted recurrent selection (MARS), and genomic selection (GS)/genome-wide selection (GWS) (Mir et al., 2012). For example, in rice through marker-assisted gene pyramiding two major bacterial blight (*Xa21* and *xa13*) and two major blast resistance genes (*Pi54* and *Pi1*) were introgressed into a mega variety Tellahamsa (Jamaloddin et al., 2020).

Newer emerging molecular markers, also known as next-generation molecular markers are expected to accelerate crop improvement programs (Mir et al., 2022). For instance, new markers including structural variations (SVs) and k-mers are believed to be extremely useful in GWAS and in other gene discovery programs (Mir et al., 2022). New genomics tools are continuously evolving. For instance, whole-genome re-sequencing (WGRS), QTL sequencing (QTL-Seq) and restriction site-associated DNA (RAD) genotyping have recently emerged and have revolutionized the gene discovery programs and therefore crop improvement programs (Mir et al., 2013a; Mir and Varshney 2013; Mir et al., 2022). The first set of tools for fine mapping of genes/QTLs influencing economically significant quantitative traits, such as biotic/abiotic stressors and agricultural yields, are reference genomes, transcriptomes, and molecular markers (Barmukh et al., 2022).

The study of genetics and development of improved varieties is quite easy for traits controlled by one of few genes (Qualitative traits). However, for quantitative traits, which are controlled by large number of minor genes, the improvement is not straightforward. For these quantitative traits, some sophisticated approaches have been developed for mapping genes/QTLs responsible for these traits followed by their deployment into cultivars. Several genome mapping approaches have been used in crop plants for decades (Mir et al., 2012; Gupta et al., 2013b). Among these approaches, QTL mapping is one of the most important traditional mapping approaches for mapping genes affecting quantitative traits. The QTL mapping method involves the use of bi-parental mapping populations developed using contrasting parental genotypes. Although used in many crop plants for genetic dissection, the QTL mapping method has several limitations and some of them are as follows: 1) involves the development of bi-parental mapping populations that can be time-intensive; 2) populations are derived using only two parental genotypes resulting in low mapping resolution and little diversity; 3) low rates of marker polymorphism. On the other hand, alternative approaches now widely used in genome mapping projects are genome-wide association studies (GWAS) or association mapping analyses. These methods have several advantages over traditional QTL mapping approaches. Some of these advantages are enumerated below: 1) diverse unrelated genotypes are used; 2) high mapping resolution due to a several cycles of historical recombination events; (iii) no marker is mono-morphic; and 4) there is no requirement to develop mapping populations. Additionally, several new advances such as gene-based trait mapping have aided gene discovery in crop plants. The increased speed of rapid sequencing has resulted in a decline in sequencing costs. These advancements facilitated the generation of draft genomes of >800 plant species, and the number continues to increase (Kulwal et al., 2022). The sequencing of a large number of plant genomes have not only facilitated the discovery of genes, alleles, markers, *etc.*, but also introduced the new concept of the plant pangenome, consisting of the "core genome" and the "dispensable genome" (*see* Kahn et al., 2020). Once identified, desirable allelic variability can be utilized in breeding programs through marker-assisted selection (MAS) including marker-assisted gene pyramiding. These breeding approaches rely on the use of markers associated with the QTL/gene of interest for foreground selection in segregating populations over the generations. Furthermore, whole genome screening combined with the use of novel breeding technologies, such as gene editing, enables more precise trait modification. The successful transfer of cultivars with improved agronomic performance from research settings to commercial growing situations continues to be the ultimate goal.

This Research Topic on '*Biotechnological and genomic approaches for enhancing agronomic performance of crops*' contributes to different aspects of plants genetics, genomics and biology. The Research Topic includes a total of 18 articles (fifteen research articles and three review articles) that cover recent advances in plant/crop research in the context of genomic resource development in different crops, functional genomics, gene discovery through different approaches, and understanding mechanisms of tolerance to different biotic/abiotic stresses. Several omics tools and technologies have been used to mine/characterize genes and gene families to facilitate their use in crop improvement. Initial characterization of genes/gene families has been limited to model plants - *Arabidopsis thaliana*, *Brachypodium distachyon*, *Medicago truncatula*, due to their typical features such as small genome size, ease of cultivation, dwarf plant phenotype and responsiveness to genetic transformation.

Next generation sequencing (NGS) has enabled the generation of comprehensive genomic and transcriptomic resources in many crops including model cereal (rice [*Oryza sativa*], wheat [*Triticum aestivum*], maize [*Zea mays*].) and legume (common bean [*Phaseolus vulgaris*], pea [*Pisum sativum*], soybean [*Glycine max*], chickpea [*Cicer arietinum*]) crops. In addition, the availability of these resources has facilitated their use in non-model crops leading to analysis of gene families, gene discovery, differential gene expression of various genes/gene families specific to stress responses, plant growth and development. To provide a comprehensive overview, the manuscripts in this Research Topic are categorized and summarized below under various sub-headings.

Marker development, germplasm characterization, and genome mapping/gene discovery/gene analysis

Discovery of molecular markers followed by their use in crop improvement programs is one of the most important activities in crop genomics/molecular breeding. A new class of molecular markers termed "micro-RNA derived SSR markers" has been developed and used by Sihag et al. in characterizing a set of wheat genotypes for important trait "heat tolerance". The study led to the identification of 104 heat-stress-responsive miRNAs already reported in different crop plants and validated a set of 70 miRNA-SSRs on 20 contrasting genotypes for heat tolerance. However, among these miRNA-derived SSRs markers, a set of only 19 were found polymorphic. These polymorphic markers were further tried for the study of genetic diversity and population structure of the selected wheat genotypes for heat tolerance. The analysis of genotypic data revealed that a smaller number of alleles (2.9 alleles per locus) were detected. Further analysis identified two important markers "miR159c and miR165b" that could discriminate between contrasting genotypes declared as promising diagnostic markers. We believe the results of the study are very important and may help in characterization of wheat germplasm for heat tolerance at the earliest stage and thus can save lot of time, efforts and money. The markers could be also used in wheat molecular breeding programs through MAS.

In addition to marker discovery, availability of all the marker resources developed for important traits in different crops in a single database would provide a useful catalog for marker assisted breeding activities. In this direction, Kumar S. et al. integrated, analyzed, and characterized the ESTs responsible for important abiotic stresses in some important cereals in the database "CerealESTdb". The data base will prove useful for scientists working in the area. The data base can prove useful to fetch the information about ESTs from important crops like rice, wheat, sorghum, and maize. The data base possess a set of 55,826 assembled EST sequences. In addition, a set of 51,791 predicted genes models and 254,609 gene ontology terms including extensive information on 1,746 associated metabolic pathways can be accessed in the data base. More efforts need to be made to help decipher complex biological phenomena under abiotic stresses. The "CerealESTdb" database can be publicly accessed at "URL http://cabgrid.res.in/CerealESTDb".

In one of the most important studies genetics-based senescence in maize, Caidedo et al. carried out GWAS to discover causative genetic variants influencing the major physiological measures of senescence. Plants use senescence as a defense mechanism against abiotic and biotic stresses. In this study, four physiological and two agronomic traits influencing senescence were recorded. The authors evaluated a set of 672 recombinant inbred lines (RILs) for two consecutive years. A set of 36 candidate genes were identified, including 11 genes supported by additional evidence for their involvement in senescence-related processes. Further, involvement of a key determinant of senescence "Zm00001d043586" in chlorophyll rate degradation was identified during late plant development stages. This study enhances our understanding of the genetic relationship between senescence and physiology related parameters in maize and provides new putative molecular markers for use in marker assisted selection.

The characterization of crop germplasm is one of the crucial steps in crop improvement programs. In this direction, an article by Hussain et al. presented characterization of Pakistani bread wheat germplasm Research Topic using genotyping by sequencing (GBS) approach. The study used a diverse germplasm of 184 Pakistani wheat genotypes. The genotypes were subjected to high-throughput genotyping (GBS genotyping) with a set of 123,596 high-quality single nucleotide polymorphism (SNP) markers. The study presented a detailed distribution of SNPs on different bread wheat genomes and

several genetic diversity parameters along with estimates of genome-wide linkage disequilibrium. The results of population structure, principal coordinate analysis, phylogenetic tree, and kinship analysis will prove useful in future Pakistani wheat breeding programs and in the study of marker-trait associations for important traits. Similarly, Brhane et al. used GBS based SNP markers for the study of genetic diversity in finger millet**.** The GBS method was used to simultaneously identify SNP markers as well as to genotype 288 finger millet genotypes collected from Ethiopia (228 landraces and 14 cultivars) and Zimbabwe (46 landraces). The 288 genotypes were further structured into seven sub-populations. Further analysis through AMOVA revealed genetic differentiation among sub-populations classified on the basis of geographical origin and trait data. The resultant genetic diversity from this study has the potential to improve future breeding programs for finger millet.

Leaf rust is one of the most devastating fungal diseases affecting wheat globally. More than 80 genes/QTLs have been identified for this disease on different chromosomes. In addition, some novel genes have also been identified for this disease. Kumar K. et al. presented a very comprehensive and timely review on leaf rust resistant genes and the development of resistant varieties. The review provided particularly valuable information on all of the leaf rust resistance genes including novel genes identified from different studies. Leaf rust resistance genes identified through QTL interval mapping and GWAS have been summarized in the review. The leaf rust resistance genes introgression/pyramiding into commercial/prominent cultivars from 18 different countries including India have been also tabulated. A comprehensive presentation of the challenges and future perspectives for breeding of leaf rust resistance will prove useful for global programs in wheat genetics and breeding. Another article in this Research Topic by Bhurta et al. demonstrated the importance of genome-wide identification and expression analysis of the Thioredoxin (*Trx*) gene family in wheat and examined the implication of *Trx* genes in leaf rust resistance. Thioredoxin (Trx) is a low molecular weight, thermostable, redox regulator. Redox regulators can control cellularhomeostasis viareactive oxygen species (ROS) and reactive nitrogen species (RNS) in response to pathogenic infections. The study identified a total of forty-two (42) wheat *Trx* genes (*TaTrx*) across the wheat genomes, with 12 on A, 16 on B, and 14 on D genomes, respectively. Further, eight of the 15 *TaTrx* genes selected based on *in silico* expression analysis, were further confirmed to be differentially targeted by microRNAs (miRNA). Gene expression analysis usingquantitative real-time PCR (qRT-PCR) revealed that *TaTrx* genes, *TaTrx11-5A*, *TaTrx13-5B*, *TaTrx14-5D*, and *TaTrx15-3B* are significantly induced in response to leaf rust infection in wheat (Bhurta et al.).

Functional genomics/transcriptomics

In plants, several studies reported an important role of transcription factor (TF) gene families in diverse processes such as developmental control and elicitation of defense/stress responses. For example, in rice, genome wide analysis of dehydrin (DHNs) gene family revealed that *DHNs* play an important role in combating dehydration stress (Verma et al., 2017). Similarly, the identification of AP2/ERF TF superfamily genes in five legumes *Medicago*, lotus, common bean, chickpea and pigeon-pea along with expression profiling in chickpea and pigeon-pea uncovered implications for *AP2/ERF* and *HSP90* genes in biotic and abiotic stresses (Agarwal et al., 2016). In this Research Topic, several articles are devoted to the significant role of functional genomics in the identification of gene families that control plant growth, development, biotic/abiotic stresses. For instance, the MADS-box gene family members serve a variety of roles in crop plant growth and development, and they hold a lot of promise for increasing grain yields under changing global conditions. An article in this issue by Shao et al. focused on the identification and expression analysis of MADS-Box gene family in sweet potato [*Ipomoea batatas* (L.) Lam]. MADS-box is one of the largest TF families in plants, with important roles in plant growth and development. This study provided a comprehensive analysis and identification of 95 MADS-box genes in sweet potato. These genes were analyzed and further categorized based on phylogenetic analysis with *Arabidopsis* MADS-box proteins. The MADS domain and core motifs of the sweet potato MADS-box genes were identified, with 19 genes showing collinear relationships and duplication events. Furthermore, RNA-Seq data suggested tissue-specific expression patterns for MADS box genes, including 34 being highly expressed in flowers and fruits, and 19 being highly expressed in the roots. Validation of these expressed genes though qRT-PCR revealed higher expression of five genes, *viz. IbMADS1*, *IbMADS18*, *IbMADS19*, *IbMADS79*, and *IbMADS90* in the tuberous root or fibrous root, and three genes *IbMADS18*, *IbMADS31*, and *IbMADS83* in the fruit. This study investigated the effects of MADS-box genes on the growth and development of sweet potato which provided molecular clues for use in sweet potato breeding.

In another article, the limited knowledge on MADS-box genes in the wheat genome is being rectified by a genome-wide investigation of the publicly accessible wheat reference genome undertaken by Raza et al. and has led to the identification of 300 MADS-box genes with high confidence. The identified genes were classified into 16 subfamilies after phylogenetic analysis with *Arabidopsis* and rice MADS-box genes. The *in silico* expression data suggested that MADS-box genes function as active plant protectors against pathogen insurgency and harsh climatic circumstances. The authors further presented a comprehensive set of MADS-box genes in the wheat genome, which could help speed up functional

genomics studies and bridge gaps between genotype-to-phenotype correlations by fine-tuning agronomically relevant features.

High-throughput sequencing technologies have been widely used to generate large amounts of genome-wide gene expression data (Kudapa et al., 2018). Gene expression patterns aid in better understanding of the functional elements of the genome, and also enable insights into development and disease resistance in plants. In addition to transcriptome analysis in a single genotype or tissue, comparative transcriptome analysis from multiple lines/growth stages/tissues can demonstrate numerous essential biochemical pathways and common genes in response to one or more stresses. Several prior studies have validated the effectiveness of comparative transcriptomics. For example, in wheat, comparative transcriptomics analysis was performed in susceptible and resistant wheat lines to identify genes and pathways that are targeted by obligate biotrophic fungal pathogens (Poretti et al., 2021). Similarly, in chickpea, comparative transcriptome analysis of control and inoculated roots of three resistant/susceptible genotypes revealed key genes associated with root-lesion nematode resistance (Channale et al., 2021).

In the current issue, comparative transcriptomic analysis was conducted in the Russian dandelion (*Taraxacum koksaghyz*) between plant populations differening in their root morphology (sizes and shapes), to identify key genetic determinants of root morphology (Wieghaus et al.). The genetics underlying the size and shape of roots of the Russian dandelion are key tragets for improvement as it is the primary storage location for natural rubber and inulin. Further, the Russian dandelion is considered an alternative source of natural rubber to the natural rubber tree (*Hevea brasiliensis*). The study identified differentially expressed genes correlating with different root morphotypes, resulting in the identification of multiple candidate genes, including a *glucan endo-1,3-β-d-glucosidase*, an *allene oxide synthase 3*, a TIFY10A/JAZ1 homolog. The results were further validated using qRT-PCR for two genes in Russian dandelion plants and its correlation with different root morphologies. These results confirm the effectiveness of comparative transcriptomics and its deployment in marker-assisted breeding of the Russian dandelion (Wieghaus et al.).

Our special topic also featured the elucidation of gene family research in crops, *Perilla frutescens* and *Euscaphis konishii* known for their medicinal values in Asia. Both studies were facilitated by the availability of publicly available genome and transcriptomic datasets of these species. Duan et al. conducted genome-wide analysis of the fatty acid desaturase (*FAD*) gene family in *Perilla* (*P. frutescens*) seeds known for their high α-Linolenic Acid (ALA) content and characterized 42 members of the *FAD* gene family which are known to be important players in the ALA biosynthesis pathway. This study also performed bioinformatic analyses of conserved domains, cis-regulatory elements, phylogenetic analyses, and characterized gene expression profiles in various tissue types. Further characterization of a *Perilla FAD* gene using overexpression in *Arabidopsis thaliana* demonstrated anincrease in ALA content in seeds.

The second study conducted by Liu et al. focused on characterization of 27 Auxin Response Factors (ARF) and 34 Auxin/Indole-3-acetic acid (Aux/IAA) protein families thus providing better understanding of their role in triterpenoid and anthocyanin biosynthesis in *Euscaphis konishii* fruits. This study also provides bioinformatic analyses of conserved domains, cis-regulatory elements, phylogenetic analyses and gene expression profiles characterized in various tissue types. The analysis of results suggested that *EkIAA* and *EkARF* genes remained conserved. Further, downstream analyses using qRT-PCR, yeast one-hybrid, and transient expression provide insight into the interaction between *EkARF5.1* and auxin response elements. This study provides gene targets including *EkARF5.1*that could help with the objective of crop improvement.

Under adverse conditions, superoxide dismutase (SOD) enzymes acts as the first line of defense in the plant antioxidant system, removing reactive oxygen species (ROS). SOD proteins are found throughout the plant kingdom and serve an important role in plant growth and development. Cucurbitaceae, the gourd family consisting of about 975 species is a significant economic and nutritional crop in the world. However, this crop has not undergone a full investigation of the *SOD* gene family. Rehman et al. conducted a genome-wide comparative analysis of the *SOD* gene family in Cucurbitaceae as well as gene expression analysis of *Lagenaria siceraria* under different abiotic stresses. The study identified 43 *SOD* genes from Cucurbitaceae species and analysis of these genes revealed that *SOD* genes could be divided into eight subfamilies. Further, the structural analysis of *SOD* gene family shows that the exon/intron structure and motifs are well preserved. Further, phylogenetic and population structure analysis revealed functional divergence of the Cucurbitaceae *SOD* gene family. Also, qRT-PCR of eight *Lsi SOD* genes shows differential response to various abiotic stresses, including droughtand cold stress. This study laid the groundwork for further detailed analysis of the role of the *SOD* gene family in Cucurbitaceae for improvement of this important crop.

Similarly, Wen et al. performed genome-wide analysis of *TCP* genes in the rapeseed genome, which included*in silico* characterization and gene expression analysis. The study led to the identification of 80 *BnTCP* genes and these genes could be grouped into two main classes (Class I - *PCF* and Class II - *CYC/TB1*) based on phylogenetic analysis. Further gene structural analysis shows diverse intron exon patterns as well as number of motifs, and phylogentic analysis shows a pattern of positive selection and duplications. The study also identified four hormone- and four stress-related responsive *cis*-elements in the promoter regions. Gene ontology enrichment analysis, shows enrichment within RNA/DNA

binding, metabolic processes, and transcriptional regulatory activities. Transcriptome widetissue-specific expression analyses indicates that only a few genes (*BnTCP9*, *BnTCP22, BnTCP25, BnTCP48, BnTCP52, BnTCP60, BnTCP66* and *BnTCP74*) are upregulated in root, stem, leaf, flower, seeds, and silique. Similarly, qRT-PCR-based expression analysis indicates that *BnTCP36, BnTCP39, BnTCP53, BnTCP59* and *BnTCP60* are upregulated at certain timepoints uponhormone treatment and abiotic stresses. These genes did not respond to drought and methyl jasmonate (MeJA). These insights pave the way for future research into importance of *BnTCP* gene family in rapeseed, particularly for tolerance to abiotic stresses.

Ren et al. studied the expression and association analysis of *CBLL* genes in peanut. In the study, *CBL*-like genes were classified as those with the Cys Met Meta PP domain. A set of 29 *CBLL* genes including 12 from cultivated peanuts and 17 from wild species were identified. The genes were found scattered at the ends of different chromosomes. These identified *CBLL* proteins were found to have five separate evolutionary branches based on evolution, gene structure, and motif analysis. The chromosome distribution patterns and synteny analyses identified whole-genome duplication (allo-polyploidization) leading to the expansion of *CBLL* genes. Further comparative genomics investigation showed that there were three shared collinear *CBLL* gene pairs between peanut, *Arabidopsis*, grape, and soybean, but none between peanut and rice. The authors also identified one polymorphic site in *AiCBLL7* through association analysis. This site is believed to be closely associated with pod length (PL), pod width (PW), hundred pod weight (HPW) and hundred seed weight (HSW).

Hong et al. performed a comparative transcriptomics study under different abiotic stresses, to better understand the role of *A. thaliana* gene *AtARA6* and its contribution to abiotic stress tolerance in soybean. The gene is known to play an important role in salt tolerance in *A. thaliana*. Transcriptome sequencing and subsequent differential expression analysis of transgenic soybean overexpressing *AtARA6* and wild-type, "Shen Nong 9" (SN9) under salt treatment and water treatment indicate several key categories of genes, including transcription factors such as WRKY86 that are significantly impacted after salt treatment. The results suggest the role of *AtARA6* gene in salt tolerance in transgenic soybean *via* regulation of soybean SNARE complexes. Thus, these investigations provide the gene expression changes mediated by abiotic stresses in transgenic soybean overexpressing *AtARA6* gene, and some insights into the mechanism mediating salt tolerance in soybean.

Genome editing and omics approaches for functional foods

Functional foods are plant-based foods that provide health advantages in addition to basic nourishment, such as fruits, vegetables, spices, cereals, millets, pulses, and oilseeds. Few crops have been studied to uncover their use as functional foods. Nayak et al. compiled a very important review article on recent advances made in different omics tools and technologies for enhancing plant based functional foods. A variety of omics approaches/integrated omics including genomics, transcriptomics, proteomics, metabolomics coupled with artificial intelligence and machine learning approaches have been proposed to improve crops plant yields and/or functional profiles. The authors have been made to compile literature and provide comprehensive insights into omics studies that have been carried out to uncover active components and crop improvement by enhancing the functional compounds in different crop plants. Further efforts are required to characterize functional foods used in traditional medicines, as well as application of this knowledge to improve biofortification of staple food crops in order to tackle malnutrition and hunger in the world.

Among the several recently emerging genomics tools and techniques, CRISPR-Cas is one of the most promising genome editing tools being used in crop improvement programs. Therefore, a comprehensive review was presented by Joo et al. on "Utilization of CRISPR-Cas in Tropical Crop Improvement". The authors provided a clear workflow with distinctprocesses with specific steps as well as decision points that should make practical application of the CRISPR-Cas gene editing platform more structured. The authors outlined different steps involved in the process with the aim of democratizing the technology and helping any lab planning to implement CRISPR-cas in their workflow.

Perspective

The contributions in this Research Topic provide a wealth of information on different aspects of genetics, genomics, crop biotechnology, and molecular plant biology. Genomic technologies, procedures, and statistical methodologies must continue to evolve in order to advance the discipline and breed genotypes that will fulfill future food needs as well as adaptability of crops to climate change. The development as well as use of new marker types derived from different sources, such as miRNA-derived SSRs, are essential for examining genetic linkages, population structure analyses, and marker-trait interactions with different important traits. Furthermore, developments in NGS technology have made it easier to construct and use pangenomes which will expand the number of DNA markers available for GWAS. These marker resources will aid in utilizing hitherto untapped genetic resources to accurately breed for the desired characteristic in a short period of time, which would have been otherwise difficult to achieve.

The utilization of high-throughput genotyping systems for the study of genetic diversity, population structure, functional genomics, transcriptomics, whole-genome scanning and

mapping genes/QTLs for critical features like biotic/abiotic stress tolerance, and climate resilience, were underscored in several publications in the Research Topic. Emphasis was also laid on genome-wide analysis of different gene families in important crop plants and in non-model crop plants. These gene families are responsible for governing important traits. Bi-parental mapping populations for QTL/gene discovery and varied germplasm resources for genome-wide association studies (GWAS) were utilized. However, study and analyses of multi-parental mapping populations can improve precision gene/QTL identification. Furthermore, new advanced gene discovery methods/multi-locus approaches are needed to uncover genes/QTLs, which can lead to the discovery of candidate genes for functional characterization by gene deletion, mutagenesis, and genome editing. The Research Topic aims to demonstrate how to use multi-parental populations and QTL identification to breed genotypes for different traits and encourages readers to apply this knowledge to breeding next-generation crop varieties.

Author contributions

RRM and HK prepared the first draft of the manuscript. SP and RL reviewed and corrected the draft. All authors read, reviewed, and approved the final draft of the editorial.

Funding

This work was supported by grants from the Department of Biotechnology (DBT), Government of India to RRM. HK acknowledges the Science and Engineering Research Board, Department of Science and Technology (CRG/2019/005966) for the support.

Acknowledgments

The editors would like to thank Andrew Paterson (Chief Editor Frontiers in Genetics section Plant Genomics) for approving this Research Topic.

Conflict of interest

SP is an employee of Altria Client Services LLC. RSL, RRM and HK declare no competing interests.

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

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

Banerjee, S., Das, S., Mir, R. R., Kundu, A., Topdar, N., Sarkar, D., et al. (2012). Assessment of genetic diversity and population structure in a selected germplasm collection of 292 jute genotypes by microsatellite (SSR) markers. *Mol. Plant Breed.* 3, 11–25. doi:10.5376/mpb.2012.03.0002

Barmukh, R., Roorkiwal, M., Garg, V., Khan, A. W., German, L., Jaganathan, D., et al. (2022). Genetic variation in CaTIFY4b contributes to drought adaptation in chickpea. *Plant Biotechnol. J.* 20, 1701–1715. doi:10.1111/pbi.13840

Channale, S., Kalavikatte, D., Thompson, J. P., Kudapa, H., Bajaj, P., Varshney, R. K., et al. (2021). Transcriptome analysis reveals key genes associated with root-lesion nematode *Pratylenchus thornei* resistance in chickpea. *Sci. Rep.* 11, 17491. doi:10.1038/s41598-021-96906-3

Choudhary, N., Anjali, G. M., Shafi, S., Jan, S., Mir, A., Singh, B., et al. (2021). Molecular diversity and nutriment studies of common bean (Phaseolus vulgaris L.) from the two hot-spots of western himalayas of Jammu and Kashmir. *Crop Pasture Sci.* 73, 249–262. doi:10.1071/CP21347

Gupta, P. K., Kulwal, P. L., and Mir, R. R. (2013b). "QTL mapping: Methodology and applications in cereal breeding," in *Cereal genomics II*. Editors P. K Gupta and R. K Varshney (Netherlands: Springer). doi:10.1007/978-94-007-6401-9_11

Gupta, P. K., Rustgi, S., and Mir, R. R. (2013a). "Array-based high-throughput dna markers and genotyping platforms for cereal genetics and genomics," in *Cereal genomics II*. Editors P. K. Gupta and R. K. Varshney (Netherlands: Springer). doi:10.1007/978-94-007-6401-9_2

Gupta, P. K., Rustgi, S., and Mir, R. R. (2008). Array-based high-throughput DNA markers for crop improvement. *Heredity* 101, 5–18. doi:10.1038/hdy.2008.35

Jamaloddin, M., Durga Rani, C. V., Swathi, G., Anuradha, C., Vanisri, S., Rajan, C. P., et al. (2020). Marker Assisted Gene Pyramiding (MAGP) for bacterial blight and blast resistance into mega rice variety "Tellahamsa". *PloS one* 15 (6), e0234088. doi:10.1371/journal.pone.0234088

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., and Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 25 (2), 148–158. doi:10.1016/j.tplants.2019.10.012

Kudapa, H., Garg, V., Chitikineni, A., and Varshney, R. K. (2018). The RNA-Seq-based high resolution gene expression atlas of chickpea (*Cicer arietinum* L.) reveals dynamic spatio-temporal changes associated with growth and development. *Plant Cell Environ.* 41 (9), 2209–2225. doi:10.1111/pce.13210

Kulwal, P. L., Mir, R. R., and Varshney, R. K. (2022). "Efficient breeding of crop plants," in *Fundamentals of field crop breeding*. Editor D. K. Yadav (Singapore: Springer). doi:10.1007/978-981-16-9257-4_14

Kumar, S., Kumar, M., Mir, R. R., Kumar, R., and Kumar, S. (2021). "Advances in molecular markers and their use in genetic improvement of wheat," in *Physiological, molecular, and genetic perspectives of wheat improvement*. Editors S. H. Wani, A. Mohan, and G. P. Singh (Springer), 139–174.

Mir, R. R., Choudhary, N., Singh, B., Khandy, I., Bawa, V., Sofi, P., et al. (2015). "Harnessing genomics through phenomics," in *Phenomics in crop plants: Trends, options and limitations*. Editor J. Kumar (India: Springer).

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi:10.1016/j.plantsci.2019.01.007

Mir, R. R., Rustgi, S., Zhang, Y-M., and Xu, C. (2022). Multi-faceted approaches for breeding nutrient-dense, disease resistant, and climate resilient crop varieties for food and nutritional security. *Heredity* 128, 387–390. doi:10.1038/s41437-022-00542-0

Mir, R. R., Saxena, R. K., Saxena, K. B., Upadhyaya, H. D., Kilian, A., Cook, D. R., et al. (2013b). Whole-genome scanning for mapping determinacy in pigeonpea (*Cajanus* spp.). *Plant Breed.* 132, 472–478. doi:10.1111/j.1439-0523.2012.02009.x

Mir, R. R., Hiremath, P. J., Riera-Lizarazu, O., and Varshney, R. K. (2013a). "Evolving molecular marker technologies in plants: From RFLPs to GBS," in *Diagnostics in plant breeding*. Editors T. Lübberstedt and R. K. Varshney (New York: Springer Science+Business), 229–247.

Mir, R. R., and Varshney, R. K. (2013). "Future prospects of molecular markers in plants," in *Molecular markers in plants*. Editor R. J. Henry (Oxford, UK: Blackwell Publishing Ltd), 169–190.

Mir, R. R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R., and &Varshney, R. K. (2012). Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theor. Appl. Genet.* 125, 625–645. doi:10.1007/s00122-012-1904-9

Pazhamala, L. T., Kudapa, H., Weckwerth, W., Millar, A. H., and Varshney, R. K. (2021). Systems biology for crop improvement. *Plant Genome* 14, e20098. doi:10.1002/tpg2.20098

Poretti, M., Sotiropoulos, A. G., Graf, J., Jung, E., Bourras, S., Krattinger, S. G., et al. (2021). Comparative transcriptome analysis of wheat lines in the field reveals multiple essential biochemical pathways suppressed by obligate pathogens. *Front. Plant Sci.* 12, 720462. doi:10.3389/fpls.2021.720462

Raza, A., Tabassum, J., Kudapa, H., Varshney, R. K., Sagwal, V., Sihag, P., et al. (2021). Can omics deliver temperature resilient ready-to-grow crops?Development and characterization of nitrogen and phosphorus use efficiency responsive genic and miRNA derived SSR markers in wheat. *HeredityHeredity* 41 (8), 391–401. doi:10.1038/s41437-022-00506-4

Sagwal, V., Sihag, P., Singh, Y., Mehla, S., Kapoor, P., Balyan, P., et al. (2022). Development and characterization of nitrogen and phosphorus use efficiency responsive genic and miRNA derived SSR markers in wheat. *Hered. Nat.* 128, 391–401. doi:10.1038/s41437-022-00506-4

Shafi, S., Tahir, M., Khan, M. A., Bhat, M. A., Kumar, U., Kumar, S., et al. (2021). Trait phenotyping and SSR markers characterization of wheat (*Triticum aestivum* L.) germplasm for breeding early maturing wheat's for Western-Himalayas. *Genet. Resour. Crop Evol.* 69, 755–770. doi:10.1007/s10722-021-01261-x

Tahir, M., Shafi, S., Khan, M. A., Sheikh, F. A., Bhat, M. A., Sofi, P. A., et al. (2021). Grain micronutrient and molecular characterization of wheat (*Triticum aestivum* L.) germplasm using genic and random SSR Markers. *Crop Pasture Sci.* 73, 93–103. doi:10.1071/CP21116

Tyagi, S., Kumar, A., Gautam, T., Pandey, R., Rustgi, S., and Mir, R. R. (2021). Development and use of miRNA-derived SSR markers for the study of genetic diversity, population structure, and characterization of genotypes for breeding heat tolerant wheat varieties. *PLoS One* 16 (2), e0231063. doi:10.1371/journal.pone.0231063

Tyagi, S., Sharma, S., Ganie, S. A., Tahir, M., Mir, R. R., and Pandey, R. (2019). PlantmicroRNAs: Biogenesis, gene silencing, web-based analysis tools and their use as molecular markers. *3 Biotech.* 9 (11), 413. doi:10.1007/s13205-019-1942-y

Verma, G., Dhar, Y. V., Srivastava, D., Kidwai, M., Chauhan, P. S., Bag, S. K., et al. (2017). Genome-wide analysis of rice dehydrin gene family: Its evolutionary conservedness and expression pattern in response to PEG induced dehydration stress. *PLoS One* 12 (5), 0176399. doi:10.1371/journal.pone.0176399

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 July 2021
doi: 10.3389/fgene.2021.699420

Discovery of miRNAs and Development of Heat-Responsive miRNA-SSR Markers for Characterization of Wheat Germplasm for Terminal Heat Tolerance Breeding

*Pooja Sihag[1], Vijeta Sagwal[1], Anuj Kumar[2], Priyanka Balyan[3], Reyazul Rouf Mir[4], Om Parkash Dhankher[5] and Upendra Kumar[1]**

[1] Department of Molecular Biology, Biotechnology and Bioinformatics, College of Basic Sciences and Humanities, CCS Haryana Agricultural University, Hisar, India, [2] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India, [3] Department of Botany, Deva Nagri College, Meerut, India, [4] Division of Genetics and Plant Breeding, Sher-e-Kashmir University of Agricultural Sciences and Technology, Srinagar, India, [5] Stockbridge School of Agriculture, University of Massachusetts, Amherst, MA, United States

OPEN ACCESS

Edited by:
Nidhi Rawat,
University of Maryland, College Park, United States

Reviewed by:
Jianfeng Weng,
Chinese Academy of Agricultural Sciences (CAAS), China
Charu Lata,
National Institute of Science Communication and Information Resources (CSIR), India

***Correspondence:**
Upendra Kumar
baliyan.upendra@gmail.com

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

Received: 23 April 2021
Accepted: 30 June 2021
Published: 28 July 2021

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

A large proportion of the Asian population fulfills their energy requirements from wheat (*Triticum aestivum* L.). Wheat quality and yield are critically affected by the terminal heat stress across the globe. It affects approximately 40% of the wheat-cultivating regions of the world. Therefore, there is a critical need to develop improved terminal heat-tolerant wheat varieties. Marker-assisted breeding with genic simple sequence repeats (SSR) markers have been used for developing terminal heat-tolerant wheat varieties; however, only few studies involved the use of microRNA (miRNA)-based SSR markers (miRNA-SSRs) in wheat, which were found as key players in various abiotic stresses. In the present study, we identified 104 heat-stress-responsive miRNAs reported in various crops. Out of these, 70 miRNA-SSR markers have been validated on a set of 20 terminal heat-tolerant and heat-susceptible wheat genotypes. Among these, only 19 miRNA-SSR markers were found to be polymorphic, which were further used to study the genetic diversity and population structure. The polymorphic miRNA-SSRs amplified 61 SSR loci with an average of 2.9 alleles per locus. The polymorphic information content (PIC) value of polymorphic miRNA-SSRs ranged from 0.10 to 0.87 with a mean value of 0.48. The dendrogram constructed using unweighted neighbor-joining method and population structure analysis clustered these 20 wheat genotypes into 3 clusters. The target genes of these miRNAs are involved either directly or indirectly in providing tolerance to heat stress. Furthermore, two polymorphic markers miR159c and miR165b were declared as very promising diagnostic markers, since these markers showed specific alleles and discriminated terminal heat-tolerant genotypes from the

susceptible genotypes. Thus, these identified miRNA-SSR markers will prove useful in the characterization of wheat germplasm through the study of genetic diversity and population structural analysis and in wheat molecular breeding programs aimed at terminal heat tolerance of wheat varieties.

Keywords: *Triticum aestivum* L., genetic diversity, simple sequence repeats, population structure, marker assisted breeding, heat responsive genes

INTRODUCTION

Triticum aestivum L. is the most commonly used hexaploid bread wheat (AABBDD; 6X = 2*n* = 42) with genome size of approximately 17 GB. It is derived via crossing among tetraploid *Triticum turgidum* (AABB) and diploid *Aegilops tauschii* (DD), whereas *T. turgidum* (AABB) was derived via crossing between *Triticum urartu* (AA) and *Aegilops speltoides* (BB) (Petersen et al., 2006). Wheat is the second principal staple food cereal of the world fulfilling the maximal carbohydrates dietary requirement (Kumar et al., 2013). To meet the food demand of the rapidly growing population, wheat productivity should be enhanced by 40–50% by 2050 (Sharma et al., 2021). Terminal heat stress is the serious issue in wheat-cultivating regions due to the rise in temperature every year because of the global warming effect (Mitra and Bhatia, 2008). Heat stress at the time of anthesis and grain filling adversely affects wheat quality and yield in terms of grain number and grain weight (Ferris et al., 1998; Kumar et al., 2013). A reduction of 4–6% in wheat grain yield was observed for every 1°C increase in temperature (Zhao et al., 2017). High temperature at the time of flowering reduces its fertility and disrupts photosynthesis mainly through leaf senescence, leading to heavy yield losses (Fábián et al., 2019). Substantial variation for heat tolerance is available among wheat cultivars and breeding lines but remains relatively unexploited. In comparison to the heat-susceptible lines, the genotypes maintaining yields under higher temperatures must contain genes that prevent degradation of chlorophyll molecules, maintain low canopy temperature, and encode enzymes that can maintain activity under elevated temperatures.

The heat tolerance capability of a genotype depends on different physiological, biochemical, and genetic parameters, which are directly or indirectly involved in providing defense against heat stress (Kumar R. R. et al., 2017). The microRNAs, transcription factors (TFs), signaling molecules, and heat-shock proteins (HSPs) play key roles in providing thermotolerance (Kumar et al., 2012). The microRNAs are small (19–24 bp), non-coding RNA molecules synthesized endogenously from *microRNA* (*miRNA*) genes and present in both prokaryotes and eukaryotes (Rhoades et al., 2002; Tyagi et al., 2019). It is well known that the miRNAs control the gene expression either by direct cleavage of the target mRNA or its suppression at translational level (Bej and Basak, 2014). These are vital regulatory elements intricate in several biotic and abiotic stress responses (Singh and Shah, 2014; Jatan and Lata, 2019). Numerous abiotic-stress-associated miRNAs have been recognized and characterized in plants, whereas only few reports are available in wheat (Ohama et al., 2017). In the case of abiotic stress, a strong interaction between miRNAs and TFs has been noticed. The miRNAs and TFs interaction regulate the expression of a particular gene either by its upregulation or downregulation in a significant manner (Chow et al., 2016). Earlier studies have identified miRNAs in response to different heat stress treatment at the grain filling stage of several heat-affected and heat-tolerant wheat varieties (Qin et al., 2008; Xin et al., 2010; Kumar et al., 2015; Ragupathy et al., 2016).

Previous studies have also shown the association of simple sequence repeats (SSRs) with heat-stress-responsive traits (Hassan, 2016; Sharma et al., 2018; Al-Ashkar et al., 2020; Saha et al., 2020). SSRs, also called as microsatellites are one- to six-nucleotide long tandem repeats distributed randomly throughout the genome (Gupta et al., 1996; Chen et al., 2009). SSRs have been identified in the non-coding, coding, and untranslated regions (UTRs), but their abundance was reported highest in non-coding regions of the genomes (Madsen et al., 2008). SSRs are widely used as molecular markers for genetic diversity analysis, marker-assisted selection, and linkage mapping among populations (Mir et al., 2012a,b, 2013; Gupta et al., 2013; Mir and Varshney, 2013; Kumar et al., 2021). The following properties of SSRs make them the best marker of interest in biological researches: (1) codominant nature (able to distinguish between homozygous and heterozygous); (2) divergence in number of short tandem repeats; (3) presence of multiple alleles; (4) polymorphic nature; and (5) high reproducibility (Brandström et al., 2008; Heesacker et al., 2008; Mir and Varshney, 2013; Mir et al., 2013). Genic SSR markers designed from heat-stress-associated genes were screened over heat-tolerant and heat-susceptible wheat genotypes to study the genetic diversity and association between SSR markers and traits of heat stress tolerance (Sharma et al., 2020; Manjunatha et al., 2021). Previous reports confirmed the presence of SSRs in pre-miRNA sequences of many plant species including *Arabidopsis*, rice, wheat, and *Brassica* in response to different abiotic stresses (Ganie and Mondal, 2015; Kumar A. et al., 2017; Singh et al., 2017; Tabkhkar et al., 2020). There are very few reports for genetic diversity studies in heat-tolerant and heat-responsive wheat genotypes that used miRNA-SSR as molecular markers. Tyagi et al. (2021) mined 96 heat-responsive miRNAs that led to the development of 13 miRNA-SSR markers. These markers were used to study the genetic diversity, population structure, and characterization of 37 wheat genotypes for heat tolerance breeding. Sharma et al. (2021) also designed 177 SSR markers from heat-responsive genes and pre-miRNAs (11) of wheat genome and screened over 36 wheat genotypes to study genetic diversity for heat tolerance (Sharma et al., 2021). The results of the previous

studies suggested the potential role of miRNA-derived SSR markers in marker-assisted breeding (MAB), which aimed to improve heat tolerance and adaptability developmental traits in crop plants including wheat (Sharma et al., 2021; Tyagi et al., 2021).

Keeping this in view, the present study was planned to develop more miRNA-derived SSR markers for the identification of additional promising markers for wheat molecular breeding programs. We mined heat-responsive miRNA from different plant species, including *Arabidopsis*, rice, and wheat, and developed miRNA-SSRs markers from the flanking region of pre-miRNA of wheat genome. The genetic diversity of 20 genotypes of wheat that differs in temperature response was evaluated using 70 miRNA-SSR markers. We have also studied the target genes of heat-responsive miRNAs and their role in heat stress. Our study will help in the categorization of the terminal heat-tolerant and heat-susceptible genotypes using miRNA-SSRs markers in the early stages of wheat development.

MATERIALS AND METHODS

Plant Material and DNA Extraction

Twenty wheat genotypes comprising 13 terminal heat-tolerant and 7 heat-susceptible genotypes were used in this study (**Table 1**). One hundred milligrams of fresh leaves was used for DNA isolation using cetyl trimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). The samples were immersed in liquid nitrogen and ground to fine powder using a Tissue Lyser (Retsch Mixer Mill MM400) machine. CTAB buffer (containing 2% CTAB, 1.4 M NaCl; pH 8, 20 mM EDTA, and 100 mM Tris Cl) was added to the ground powder and incubated at 65°C for 1 h. Chloroform/isoamyl alcohol (24:1) was mixed properly with lysed sample and centrifuged at 7,800 rpm for 10 min. The supernatant was collected in a fresh tube and added prechilled iso-propanol to precipitate the genomic DNA. After centrifugation, pellet was washed two times with 70% alcohol. The DNA pellet was dried and dissolved in 100 μl Tris–EDTA (TE) buffer. DNA quality and quantity were checked using NanoDrop spectrophotometer (Thermo Scientific, United States) along with 0.8% agarose gel separation. To perform polymerase chain reaction, DNA was diluted with nuclease-free water (NFW) to 100 ng/μl concentration.

PCR Amplification

To study the polymorphism, DNA of all 20 wheat genotypes was amplified using an Eppendorf thermal cycler. A reaction mixture of 20 μl containing 10 μl 2× Promega green master mix, 2 μl template DNA (200 ng), 0.4 μl of each forward and reverse primer (0.2 μM), and 7.2 μl NFW was prepared. PCR conditions were as follows: initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 50 s, annealing at 50–56°C (differ with primer pair) for 30 s, and extension at 72°C for 40 s, followed by the final extension at 72°C for 7 min. Amplified PCR product was resolved on 10% denaturing urea polyacrylamide gel electrophoresis unit (urea PAGE) at 120 V for 4 h with 1× Tris–acetate–EDTA (TAE) buffer. After staining with ethidium bromide (EtBr), the gel was visualized in Vilber Fusion Solo S gel documentation system.

Screening of Heat-Responsive miRNAs and Designing of miRNAs-SSRs Markers

Heat-responsive miRNAs reported in different plants including *Arabidopsis*, wheat, rice, maize, and *Brassica* were collected based on extensive literature survey. Mature and stem loop (pre-miRNA) sequences of these heat-responsive miRNAs were obtained from miRBase v22 database[1] in FASTA format (Kozomara et al., 2019). In Ensembl Plants database,[2] BLASTn search was carried out against reference genome data of wheat using pre-miRNA sequences to get the 500-bp downstream and 500-bp upstream sequence flanking pre-miRNA (Bolser et al., 2015). BLASTn search resulted into multiple hits; among them, the best sequence that aligned perfectly (100%) to the wheat genome and has an *e*-value of 10^{-10} was selected for designing of SSR markers. The microRNA-SSR markers were designed from the pre-miRNA flanking sequence using BatchPrimer3 v1.0.[3] For primer synthesis di-, tri-, tetra-, penta-, and hexa-SSRs of minimum 12 bp length were chosen. The following parameters were set for designing of SSR flanking primers: primer length of 18–25 bp, melting temperature (T_m) in 50–65°C range, 40–60% GC content, and product size of 100–300 bp length (You et al., 2008). A summary of the methods used for miRNA-SSR development and validation in wheat is represented in **Figure 1**.

Allele Scoring and Genetic Diversity Analysis

A total of 70 miRNA-SSR markers were screened on genomic DNA of selected 20 wheat genotypes to study the polymorphism pattern in terminal heat-tolerant and heat-susceptible wheat genotypes. The scoring of all the polymorphic primers was done based on the absence and presence of DNA band. The absence and presence of DNA band was indicated as 0 and 1, respectively. The size (bp) of each DNA band was determined using Evolution Edge software based on the size of Quick-Load® Purple 50 bp DNA Ladder bands. A binary (0/1) allele scoring matrix generated from polymorphic miRNA-SSR markers was used to compute the polymorphic information content (PIC) and genetic diversity among 20 wheat genotypes. PIC was calculated using the GeneCalc online bioinformatics tool[4] (Nagy et al., 2012) according to the equation: $PIC = 1-\sum pi^2$, where *pi* is the frequency of *i*th allele of a particular locus. The phylogenetic relations between terminal heat-tolerant and heat-susceptible genotypes were drawn using DARwin v6.0 software (Perrier et al., 2003). Dissimilarity matrix was computed and used for the grouping of 20 genotypes or for the construction of dendrogram using unweighted neighbor joining followed by bootstrapping with 1,000 replications (Gascuel, 1997). Principal coordinate analysis (PCA) was also computed to categorize these 20 genotypes into different groups using DARwin v6.0.

[1]http://www.mirbase.org/

[2]https://plants.ensembl.org/

[3]https://wheat.pw.usda.gov/demos/BatchPrimer3/0

[4]https://gene-calc.pl/

TABLE 1 | Details of 20 terminal heat-tolerant and heat-susceptible wheat genotypes used in this study.

S. no.	Genotypes	Tolerant/susceptible	Pedigree	Developing center
1	Cv. Raj3765	Tolerant	HD2402/VL639	RAU, Durgapur
2	WH730	Tolerant	CPAN2092/IMPROVED LOK1	CCSHAU, Hisar
3	WH1021	Tolerant	GW296/SONAK	CCSHAU, Hisar
4	WH1218	Tolerant	KA/NAC//TRCH/3/VORB	CCSHAU, Hisar
5	PBW533	Tolerant	PBW343/PBW138//PBW343	PAU, Ludhiana
6	WH789	Tolerant	–	CCSHAU, Hisar
7	WH1142	Tolerant	CHEN/Ae.Sq. (TAUS)//FCT/3/2*WEAVER	CCSHAU, Hisar
8	WH1124	Tolerant	MUNIA/CHTO/AMSEL	CCSHAU, Hisar
9	WH1133	Tolerant	BABAX/LR42//BABAX*2/3/VJV1TS1	CCSHAU, Hisar
10	WH787	Tolerant	–	CCSHAU, Hisar
11	WH147	Susceptible	E4870-C303/S339-PV18	CCSHAU, Hisar
12	WH711	Susceptible	ALD "S"'HUAC//HD2285/3/HFW-17	CCSHAU, Hisar
13	HD2967	Tolerant	ALD/COC//URES/HD2160M/HD2278	IARI, New Delhi
14	WH157	Susceptible	NP876/S308//CNO/8156	CCSHAU, Hisar
15	WH542	Susceptible	JUP/BJY "S"//URES	CCSHAU, Hisar
16	WH1140	Tolerant	WBLLI*2/VIVITSI	CCSHAU, Hisar
17	PBW550	Tolerant	WH594/RAJ3858/W485	PAU, Ludhiana
18	WH283	Susceptible	HD-1981/RAJ-821	CCSHAU, Hisar
19	HD2687	Susceptible	Gene introgressed from Sr31 Lr26 Yr9 Pm8	IARI, New Delhi
20	HD2329	Susceptible	HD1962/E4870/3/K65/5/HD1553/4/UP262	IARI, New Delhi

Population Structure Analysis

The total number of subpopulations present in a population of 20 wheat genotypes was identified using the Bayesian method with binary data matrix of 19 polymorphic miRNA-SSRs. This analysis was done using STRUCTURE 2.3.4 software (Pritchard et al., 2010). The following parameters were set: (1) burnin iteration length of 100,000, (2) Markov chain Monte Carlo (MCMC) value of 100,000, and (3) the value of subpopulations number (k) ranged from 1 to 10 with 3 repetitions (Chen et al., 2012). The best fit value of subpopulations was derived with the Evanno method (Evanno et al., 2005) using online STRUCTURE HARVESTER software[5] (Earl and Bridgett, 2012). Genotypes with more than 80% similarity were allocated in a distinct subpopulation, and those with <80% were designated as admixture.

miRNA Target Identification and Gene Ontology

The target genes of heat-responsive miRNAs were predicted using the psRNATarget server[6] (Dai et al., 2018). Mature miRNA sequences in fasta format extracted from miRBase v22 were used as a query to find the targets in the wheat cDNA library. The following default parameters were used in search of target genes: (1) expectation value was set 5; (2) range of translational inhibition as 10–11 nucleotides; (3) G:U pair penalty, 0.5; (4) seed region at 2–13 nucleotides and maximum two mismatches were allowed in the seed region; (5) complementarity scoring length (HSP size), 19; and (6) energy to unpair (UPE) the target sequence, 25. Identified targets of heat-responsive miRNAs were used for Gene Ontology (GO) studies with Web Gene Ontology Annotation Plot (WEGO 2.0[7]) database (Ye et al., 2018). The Gene Ontology ID of the target genes for biological process, molecular function, and cellular component were retrieved from the Ensemble plant database and used for GO analysis in WEGO 2.0 using default parameters. A summary of the methods used for miRNA target identification is represented in **Figure 1**.

Trait Data Recording and Analysis

The set of 20 genotypes including 13 heat-tolerant and 7 heat-susceptible genotypes were also phenotyped for 5 most important physiological/biochemical traits related to heat tolerance. The data on these traits were recorded under normal and late sown conditions at 10 and 20 days after anthesis. The canopy temperature depression (CTD) and chlorophyll content was recorded using an infrared thermometer and SPAD-502, respectively, whereas relative water content (RWC), electrolyte leakage to test cell membrane stability, and proline content were measured using established methods (Barr and Weatherley, 1962; Bates et al., 1973; Dionisio-Sese and Tobita, 1998).

RESULTS

Development of Heat-Responsive miRNA-SSRs

From the literature survey, 80 heat-responsive miRNA families consisting of 104 members were recognized in *Arabidopsis*, wheat, rice, maize, and sorghum crops (**Supplementary Table 1**). Due to the lack of research on wheat miRNAs, very few wheat miRNA sequences were deposited in the miRBase v22 database.

[5]http://taylor0.biology.ucla.edu/structureHarvester/

[6]http://plantgrn.noble.org/psRNATarget/

[7]https://biodb.swu.edu.cn/cgi-bin/wego/index.pl

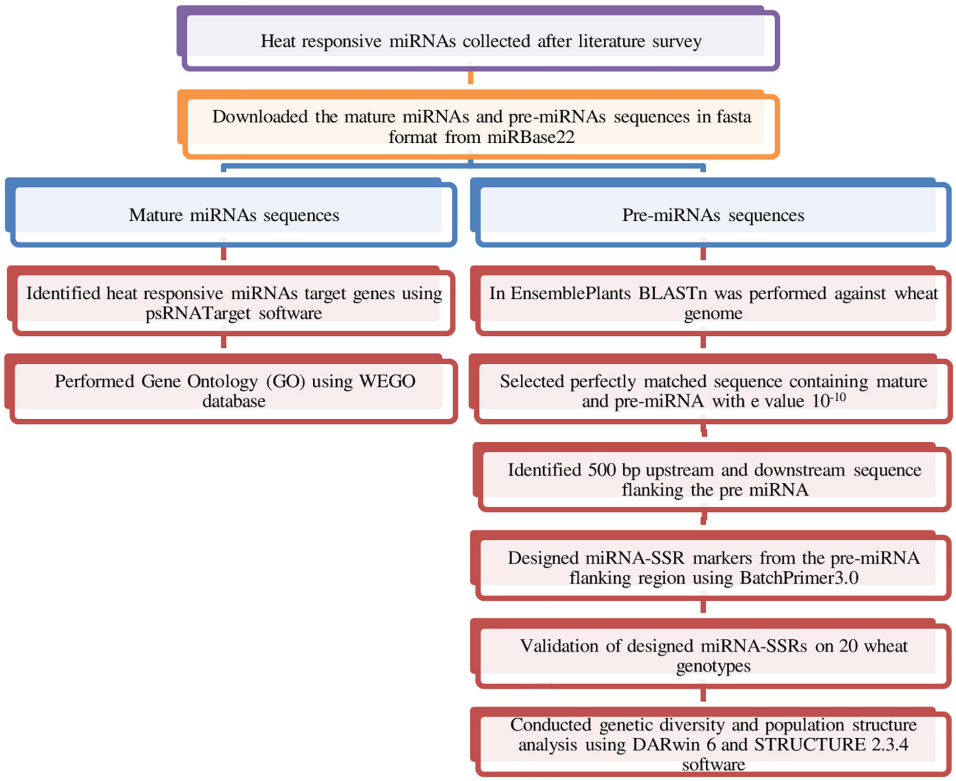


FIGURE 1 | Pipeline describing the process of identification and validation of miRNA-SSRs in wheat.

Therefore, along with *T. aestivum*, we also preferred the mature miRNAs and pre-miRNAs sequences deposited for *Oryza sativa (osa)*, *A. tauschii (ata)*, *Arabidopsis thaliana (ath)*, *Sorghum bicolor (sbi)*, and *Hordeum vulgare (hvu)* with frequency shown in **Figure 2A**. Mature sequences of five miRNAs (miR6941, miR3182, miR2012, miR2020, and miR2006) were not found in the miRBase v22 database. Primer designing using BatchPrimer3 v1.0 resulted in 70 significant miRNA-SSR primers pair, whereas no SSR flanking primers were obtained for two miRNAs (miR5071 and miR1122), and the absence of SSRs were noticed in 27 pre-miRNA sequences. The number of SSRs varies from dinucleotides to hexanucleotides with maximum dinucleotides frequency, i.e., up to 27 times $(NN)_{27}$ and tetranucleotides repeated up to 10 times $(NNNN)_{10}$. The maximum number of *miRNA* genes was found to possess tetranucleotides (47%) followed by trinucleotides (23%), pentanucleotides (14%), dinucleotides (9%), and hexanucleotides (7%) SSRs (**Figure 2B**). The highest number of miRNA-SSR markers was found to be located on 6B wheat chromosome and lowest on chromosome numbers 1D and 4D as shown in **Figure 2C**.

Validation of miRNA-SSR Markers and Genetic Diversity Analysis

Out of 70 screened miRNA-SSR markers, 64 markers resulted into clear, bright, and reproducible bands in all the 20 wheat genotypes (**Supplementary Table 2**). Among these, 19 polymorphic miRNA-SSR primers were chosen and used for genetic diversity studies of 20 terminal heat-tolerant and heat-susceptible genotypes (**Table 2**). Two polymorphic markers, miR159c and miR165b, were found to be able to distinguish terminal heat-tolerant and heat-susceptible wheat genotypes by the presence of different alleles in them (**Figure 3**). A total of 61 SSR loci were identified using 19 polymorphic primers pair with the range of 2–9 alleles, averaged 2.9 alleles per locus, and the amplicon size of polymorphic alleles ranging from 130 to 470 bp. The PIC value ranged from 0.10 to 0.87, with miR830 having the highest PIC (0.87) and miR159b and miR824 and miR2122 possessing the lowest PIC (0.10). The binary scoring matrix data obtained from 19 polymorphic miRNA-SSRs was used for phylogenetic studies of 20 terminal heat-tolerant and heat-susceptible wheat genotypes using DARwin6 software. The results clustered these 20 genotypes into 3 major clusters represented as clusters I, II, and III. Clusters I and II were found to consist of six terminal heat-susceptible and heat-tolerant wheat genotypes, respectively, whereas cluster III comprises of seven terminal heat-tolerant and one terminal heat-susceptible wheat genotype (**Figure 4A**). The principal coordinate analysis conducted using DARwin6 also distinctly separated the 20 wheat genotypes from each other (**Figure 4B**).

Population Structure Analysis

The population structure of 20 genotypes was studied using "STRUCTURE" software Bayesian clustering approach. The

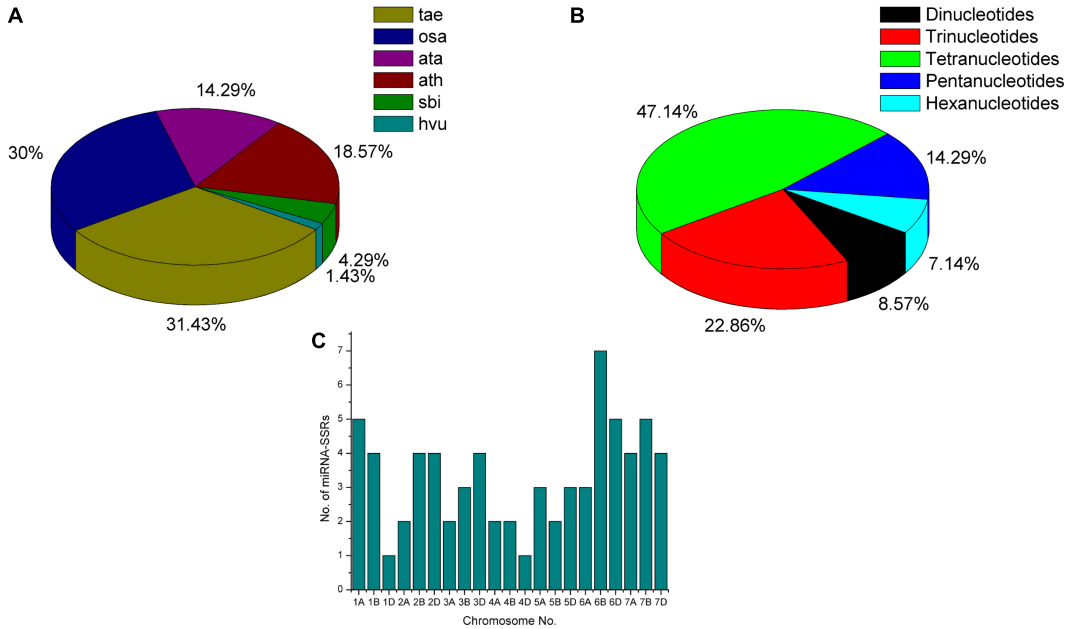


FIGURE 2 | Frequency and distribution of heat-responsive miRNA genes with SSR motifs. **(A)** Frequency of heat-responsive miRNAs submitted in miRBase reported in various crops. **(B)** Frequency of wheat miRNA-SSR repeats present in miRNA genes. **(C)** Distribution of heat-responsive miRNA-SSRs on wheat.

binary scoring matrix data of 19 polymorphic miRNA-SSRs screened on 20 genotypes was analyzed using this software, and the best "$K = 3$" value was selected based on ΔK value obtained from "Structure Harvester" (**Figure 5A**). This analysis distributed the population of 20 wheat genotypes into 3 subpopulations (P1, P2, and P3). Among these three subpopulations, each of P1 and P2 consists of six genotypes, and P3 contains four genotypes, whereas four genotypes were found to be admixed (**Figure 5B**). These results were obtained in accordance to the cluster analysis and PCA results.

Identification of miRNA Target Genes

The target genes of heat-responsive miRNAs were identified using psRNATarget server (**Supplementary Table 3**). The 80 target genes identified using this server will play a crucial role in providing tolerance against heat stress at various stages of wheat growth, whereas most of the targets were related to grain filling stages. These miRNAs regulate the target genes either by its cleavage or by suppressing the translational mechanism. The four miRNAs, namely, miR1130a, miR528, miR1137a, and miR159a, targeted the heat shock protein 17 (hsp17), heat shock protein 90 (hsp90), heat stress associated 32-kD protein (hsa32), and DnaJ heat shock family protein, respectively. Many heat-responsive genes were found to have the common target of more than two miRNAs such as peroxidase, glycosyltransferase, phytochrome, phenylalanine, flavin containing monooxygenase, and sucrose synthase. The miRNAs were also found to regulate the signaling molecules including the mitogen-activated protein kinase (MAPK), zinc finger protein (ZFP), cyclin-dependent kinase inhibitor (CDKI), serine/threonine-protein kinase, CASP-like protein, and TFs such as Auxin response factor (ARF), WRKY, NAC 6A, ethylene responsive factor 5a (ERF 5A), and R1R2R3-MYB. Heat stress induces oxidative damage, and plants adapt to this by synthesizing antioxidants including peroxidase (POX), superoxide dismutase (SOD), glutathione *S*-transferase (GST), and glutathione synthetase (GS), which were also targeted by the heat-responsive miRNAs.

Gene Ontology Analysis

Gene Ontology was conducted using WEGO to find the putative functions of miRNA-targeted genes expressed due to heat stress. These targeted genes were grouped into three classes: biological process (30 GO terms), molecular function (27 GO terms), and cellular component (28 GO terms) (**Figure 6**). GO analysis showed the role of these genes mainly in biosynthetic, metabolic, reproductive, and stress-responsive biological processes. Most of the genes were found to be associated with molecular functions such as carbohydrate binding, kinase regulator, oxidoreductase, peroxidase, DNA binding TF, signal transducer, SOD, thioredoxin disulfide reductase, transferase, and were found to be located mainly in the intracellular organelle part, apoplast, and in the membrane of the cell organelles (**Figure 6**). Therefore, the Gene Ontology analysis explained the role of miRNA-targeted genes in providing tolerance against terminal heat stress. The calculated *p*-values for GO terms are given in **Supplementary Table 4**.

Correlation of Trait and Genotypic Data

In order to understand the correlation between genotypic and trait data of five heat-tolerance-related traits, efforts have been made to analyze genotypic data of two most promising diagnostic markers (miR159c and miR165b) showing association

TABLE 2 | Details of 19 polymorphic miRNA-SSR markers used in the study of 20 terminal heat tolerant/susceptible wheat genotypes.

miRNA-SSR primer	SSR motif repeat	No. of alleles	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	T_m (°C)	Amplicon size (bp)	PIC value
miR159	(TTCT)4	4	CTCACCCCTCTATAAAACGAC	CTACATCTATGGGGCTAGGAG	50	140–260	0.62
miR159b	(CTC)4	2	ATTTTCCTTTCAATGACACCT	AAGAGATGGAACGGAAACTAC	51	165–169	0.10
miR159c	(CT)27	3	CTTTCCCTCGTGCTTGGAT	GCATAGTGATTTGATTTTCTTGTTAGC	52	180–250	0.61
miR159f	(CT)27	3	ACCTGTATAGGTTTTGCATGA	TTAGGTGCAGACTGAAAACAT	51.5	130–162	0.62
miR164a	(GCCC)3	7	ACTGCACTGCACGTGTTCTT	TTGAAGACGCATACCTCGTG	59.5	135–315	0.81
miR165b	(ATAC)10	6	CAACGGTGTGATTGTAAAAA	CGAAGTTTAATTTGGTTATGC	50	146–280	0.72
mir171b	(ATTG)3	3	CTGAACGCTACTGAGCCACT	CTACCAACACGGCAGCACTA	55	178–300	0.65
mir172c	(TTC)5	3	CCTCTCTTTGTCTTCATCCA	AAGAACCGACTGTGATCTGA	51.5	130–180	0.25
mir393a	(GAGG)4	2	CCTATATAAGGACCTCACTGGA	AGGCATTGTTGCTCTCTCT	52	157–175	0.49
mir396d	(GTGC)3	6	AAGTTATATCGGACCGTGTG	AGGAAGGGGTCGTATAAATAG	52	150–470	0.79
mir404	(TGCCGC)3	2	CTAAACCGGATAAAGGGTAGA	CAGAGGAACGCACGTAGT	52	140–162	0.49
mir830	(AGGGA)3	9	AGTACGCCTTGATCTCCTCT	CTACGTTACCTTCCTCTCTCC	53.5	130–320	0.87
mir824	(TTAA)3	2	GGCCTTTGACTGAATTAGTGT	CCAGAAAGGAATTATTTTGGA	50	134–140	0.10
miR857	(TTTA)3	2	TTCTTGCCTACTTGTTTCTG	GTCGCCGTCTTTGAATTT	50	267–272	0.10
mir1130a	(TGCA)3	2	AGTTGCACTGCTACAAGCTAC	ACTTTCGGATGTACTGTTCTG	52	146–165	0.48
mir2102	(GCC)4	2	ACCGCTGCTGTTGTATTG	TCAAGTCCTCTGCAAACAC	52	167–175	0.48
mir2122	(TTCTT)3	2	GGGTGGACAGTAAAATCAGA	CCATTTTTCAGGATCATTCTT	50	157–162	0.10
mir5384	(CCG)4	3	AGGGGATCCTCCTCAGAT	AGGTGGTTGGTGGCCAAG	52	280–350	0.66
mir9662b	(CGC)5	2	CCTTCACCAAACCCTCTT	GAGATCCAGCAGAAGGAGAT	53	227–235	0.18

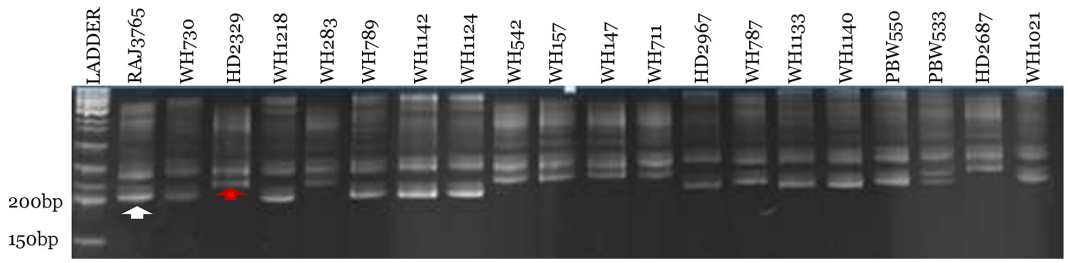


FIGURE 3 | Denaturing 10% urea-PAGE profile represents polymorphic pattern of miR159c miRNA-SSR marker over 20 terminal heat-tolerant/heat-susceptible wheat genotypes. First lane represents 50 bp DNA ladder, and the white and red arrow denote terminal heat-tolerant and heat-susceptible wheat genotypes, respectively.

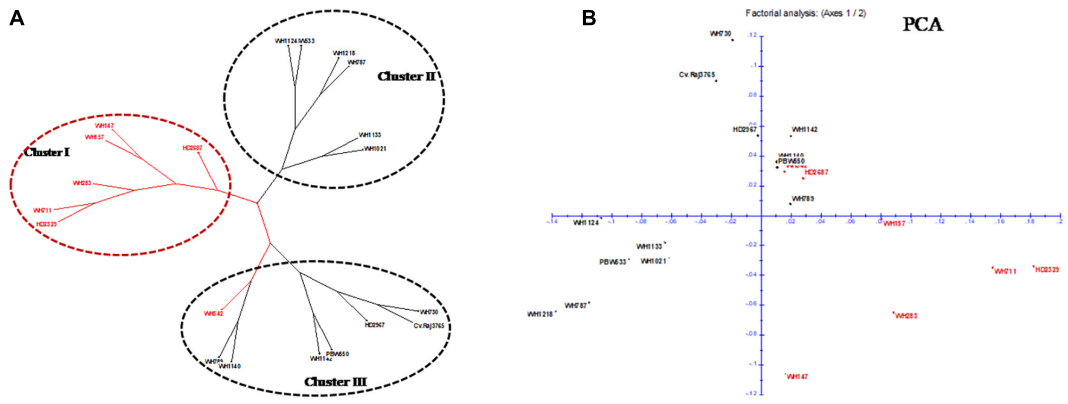


FIGURE 4 | Phylogenetic studies of wheat genotypes with heat-responsive miRNA-SSRs. **(A)** Dendrogram showing clustering of 20 wheat genotypes using polymorphic miRNA-SSR markers. Red color indicates heat-susceptible wheat genotypes, and black color indicates heat-tolerant wheat genotypes. **(B)** 2-D plot of principal coordinate analysis (PCA) for all 20 wheat genotypes. Red and black colors were used for heat-susceptible and heat-tolerant genotypes, respectively.

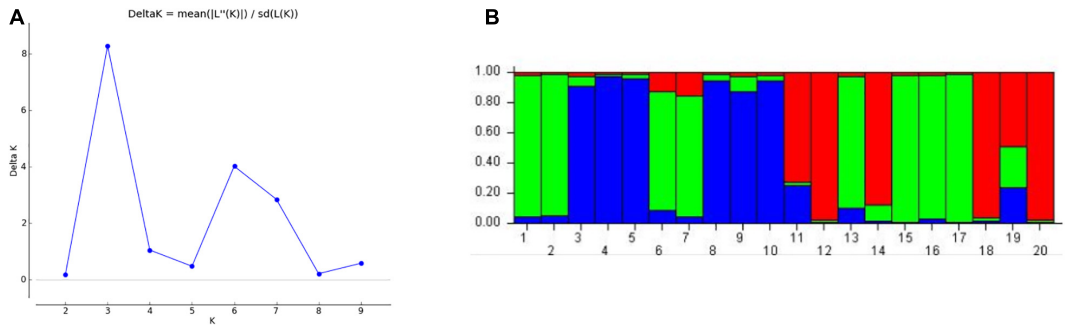


FIGURE 5 | Evaluation of number of subgroups based on "STRUCTURE" output. **(A)** Delta *K*, best value of $K = 3$. **(B)** Bar plot showing distribution of 20 wheat genotypes into 3 subgroups as shown in green, blue, and red colors.

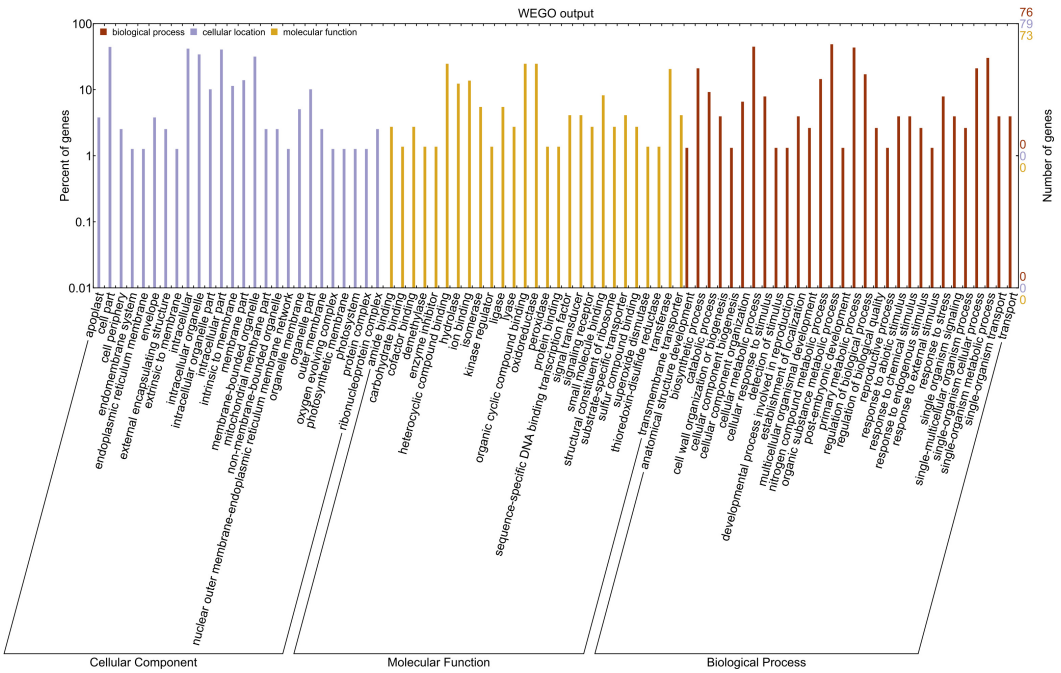


FIGURE 6 | The gene ontology analysis of the heat-responsive miRNAs-targeted genes in wheat. The red, yellow, and blue bars symbolize biological process, molecular functions, and cellular components, respectively. The *x*-axis represents three GO categories; the right side of *y*-axis represents the number of miRNA target genes, while the left side of *y*-axis represents the percentage of the target genes.

with heat tolerance. It was noticed that marker "miR159c" showed the presence of allele in all the heat-tolerant genotypes, while it is absent in heat-susceptible genotypes. Similarly, marker "miR165b" showed the absence of alleles in heat-tolerant genotypes, while it showed the presence of allele in heat-susceptible genotypes. The 100% specificity of alleles in these markers showing 100% association with heat tolerance is considered important, and this association was further confirmed by arranging trait data of heat-tolerant and heat-susceptible genotypes and testing their mean difference. The analysis of trait data of all the five traits (CTD, RWC, electrolyte leakage to test cell membrane stability, SPAD chlorophyll content, and proline content) recorded showed clear-cut mean difference between heat-tolerant vs. heat-susceptible genotypes. The heat-tolerant genotypes possessing specific alleles possess superior trait performance than heat-susceptible genotypes (**Supplementary Table 5**).

DISCUSSION

The environmental temperature increasing year by year due to the global warming emerged as a significant threat of terminal heat stress in wheat. Heat stress effect is more prominent when wheat was sown late in December instead of November due to delay in harvesting of wheat-preceding crop such as cotton and rice. Late-sown wheat experiences heat stress at the reproductive stages, which ultimately decreases the grain quality

and productivity. In order to reduce the yield loss occurring due to excessive heat effect, there is a strong need to develop enhanced terminal heat-tolerant wheat varieties. Wheat plants also adapt to terminal heat stress by shortening the grain filling period, which causes a decrease in grain number, weight, and quality. Global wheat production is estimated to have decreased by 6% with every 1°C arise in temperature (Asseng et al., 2015).

Among the molecular markers, SSR markers were preferred for genetic diversity analysis, markers-assisted breeding, quantitative trait locus (QTL) mapping, etc., because of their distribution throughout the genome (Mir and Varshney, 2013; Mir et al., 2013; Kumar et al., 2021). Previous reports have revealed the function of miRNAs in controlling the heat stress, and a distinctive difference in miRNA families and their expression level was found in control and heat-stressed wheat genotypes (Kumar et al., 2015; Sailaja et al., 2017; Ravichandran et al., 2019; Rangan et al., 2020). A vast number of SSRs were reported within the protein-coding regions and their UTRs. However, the data on SSRs from non-coding regions of genes including miRNA genes in wheat genome are very limited. In addition, some miRNAs families are extremely conserved from millions of years. Therefore, SSRs derived from these miRNAs (miRNA-SSR) can better serve as a functional marker to differentiate closely related wheat genotypes as compared to previously known markers including RAPD, RFLP, and SSR (Ganie and Mondal, 2015; Tyagi et al., 2019, 2021). The use of miRNA-SSR markers for genetic diversity studies have become the hot topic during the last decade. Many studies have been conducted to differentiate terminal heat-tolerant wheat genotypes using SSR markers (Kumar et al., 2016; Hassan, 2016; Al-Ashkar et al., 2020). Yet to our knowledge, only two studies screened the miRNA-SSR for genetic diversity analysis and for characterization of a terminal heat-tolerant and heat-susceptible wheat genotypes (Sharma et al., 2021; Tyagi et al., 2021).

Polymorphism or PIC values were found to be important parameters to study the genetic diversity among genotypes. If the PIC value is >0.5, the marker is highly informative, and if the PIC values between 0.25 and 0.5, the marker is assumed moderately informative; therefore, markers with PIC value > 0.7 are considered appropriate for diversity studies and genetic mapping (Botstein et al., 1980; Bandelj et al., 2004). In this study, the 19 polymorphic miRNA-SSR markers showed a high polymorphism with averaged 2.9 alleles per locus and the averaged PIC value of 0.48, which was found to be higher than that reported by Sharma et al. (2021) (averaged 2.58 alleles/locus and PIC value of 0.35/locus), who screened 37 markers on 36 wheat genotypes, and lower than that reported by Tyagi et al. (2021) (averaged 3.4 alleles/locus, however, lower PIC value ranged from 0.16 to 0.38), who conducted genetic diversity studies using 13 miRNA-SSRs on 37 wheat genotypes in response to heat stress.

The dendrogram and structure analysis performed with 19 heat-responsive polymorphic miRNA-SSRs divided the population of 20 wheat genotypes into 3 clusters based on their genetic makeup. Sharma et al. (2021) clustered 36 wheat genotypes into 4 clusters using 37 cgSSR and miRNA-SSR markers, whereas Tyagi et al. (2021) divided 37 wheat genotypes into 4 clusters with 7 polymorphic miRNA-SSR markers. Both dendrogram and population structure results showed the presence of both heat-tolerant and heat-susceptible wheat genotypes into one cluster, which can be either due to their common ancestry or that an inadequate number of miRNA-SSRs was used for genetic diversity analysis. Similar results were obtained by Tyagi et al. (2021). However, the report of Sharma et al. (2021) distinctly categorized the heat stress-tolerant and heat stress-susceptible wheat genotypes. Moreover, the presence of heat-susceptible wheat genotypes into cluster I and heat-tolerant genotypes in cluster II highlighted the usefulness of these polymorphic markers in screening of wheat genotypes for their heat tolerance level.

The expression of heat-stress-responsive genes was found to be regulated by various miRNAs in response to heat stress. For example, an overexpression of miR156 was reported in *Arabidopsis*, as it generates memory in response to heat stress (Stief et al., 2014), whereas a downregulation of miR159 was noticed under heat stress, as it regulates the MYB TFs. In this study, the miRNAs (miR159c and 165b) were found to target the peroxidase, phenylalanine ammonia-lyase *(PAL)*, RING-type E3 ubiquitin transferase, Xyloglucan endotransglucosylase/hydrolase (XTH), etc., which provides heat tolerance to plants by the process antioxidant metabolism and phenolics accumulations and increased the expression of heat shock proteins by degrading the suppressor proteins of heat-responsive genes, respectively (Moura et al., 2017; Peng et al., 2019). Transgenic rice plants overexpressing miR159 were more susceptible to heat stress as compared to control plants (Wang et al., 2012). Yan et al. (2016) showed the downregulation of miR165/miR166 in *Arabidopsis*; these miRNAs target the ABA-responsive genes, which are key players in providing tolerance to abiotic stress. In addition, Ravichandran et al. (2019) revealed the role of heat-responsive miR528 and miR9662 miRNAs in wheat in regulating the antioxidants and mitochondrial proteins, respectively. Our study also identified many miRNAs-targeted genes codes for heat-responsive proteins, antioxidants, and TFs including ARF, WRKY, hsp70, hsp17, POX, SOD, GST, and GS. These results were also supported by earlier studies (Goswami et al., 2014; Ravichandran et al., 2019; Su et al., 2019; Rangan et al., 2020). Therefore, the identified miRNA-SSR markers (miR159c and miR165b) that are able to differentiate heat-tolerant and heat-susceptible wheat genotypes will play a significant role in breeding programs. Additional support was provided by trait data analysis of five most important traits (CTD, RWC, electrolyte leakage to test cell membrane stability, SPAD chlorophyll content, and proline content) related to heat tolerance. The analysis clearly indicated that the marker loci miR159c and miR165b showed linkage with heat tolerance and related five traits. The results indicated that these two markers could be used in wheat molecular breeding programs aimed at enhancing heat tolerance of wheat varieties for the development of next-generation heat-tolerant wheat varieties.

In conclusion, our study identified 19 polymorphic miRNA-SSRs markers, but among these, only two miRNA-SSR markers (miR159c and miR165b) were able to differentiate terminal heat-tolerant genotypes from the susceptible one. Therefore, the identified miRNA-SSR markers enable the studying of genetic

diversity and MAB. As the physiological and phenotypic data are not sufficient to distinguish the tolerant and susceptible cultivars, the polymorphic miRNA-SSR markers will help the breeders to select terminal heat-tolerant wheat genotypes or marker-assisted selection at initial growth stages. This will further strengthen the release of new terminal heat-tolerant wheat 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/s.

AUTHOR CONTRIBUTIONS

PS and VS conducted the experiments. PS and AK performed bioinformatics data analysis. PS drafted the manuscript. UK served as principal investigator, obtained funding for this research, and also revised the manuscript. OD, PB, and RM reviewed the manuscript. All authors agreed to the final manuscript.

FUNDING

This research was partially supported by the Council of Scientific and Industrial Research (CSIR), New Delhi, India as JRF to PS [Grant No. 09/303 (0294)/2018 EMR-1].

ACKNOWLEDGMENTS

The corresponding author is highly grateful to the Head, Department of Molecular Biology, Biotechnology and Bioinformatics, CCS Haryana Agricultural University for providing all necessary facilities (Laboratory and Field) during the course of this work. PS thanks the Council of Scientific and Industrial Research (CSIR), New Delhi for Senior Research Fellow (SRF) award.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.699420/full#supplementary-material

REFERENCES

Al-Ashkar, I., Alotaibi, M., Refay, Y., Ghazy, A., Zakri, A., and Al-Doss, A. (2020). Selection criteria for high-yielding and early-flowering bread wheat hybrids under heat stress. *PloS One* 15:e0236351. doi: 10.1371/journal.pone.0236351

Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., et al. (2015). Rising temperatures reduce global wheat production. *Nat. Clim. Chang.* 5, 143–147. doi: 10.1038/nclimate2470

Bandelj, D., Jakše, J., and Javornik, B. (2004). Assessment of genetic variability of olive varieties by microsatellite and AFLP markers. *Euphytica* 136, 93–102.

Barr, H. D., and Weatherley, P. E. (1962). A re-examination of the relative turgidity technique for estimating water deficit in leaves. *Aust. J. Biol. Sci.* 15, 413–428.

Bates, L. S., Waldren, R. P., and Teare, I. D. (1973). Rapid determination of free proline for water-stress studies. *Plant Soil* 39, 205–207.

Bej, S., and Basak, J. (2014). MicroRNAs: the potential biomarkers in plant stress response. *Am. J. Plant Sci.* 5, 748–759.

Bolser, D. M., Kerhornou, A., Walts, B., and Kersey, P. (2015). Triticeae resources in Ensembl plants. *Plant Cell Physiol.* 56:e3. doi: 10.1093/pcp/pcu183

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314.

Brandström, M., Bagshaw, A. T., Gemmell, N. J., and Ellegren, H. (2008). The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Mol. Biol. Evol.* 25, 2579–2587. doi: 10.1093/molbev/msn201

Chen, M., Tan, Z., Jiang, J., Li, M., Chen, H., Shen, G., et al. (2009). Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. *FEBS Lett.* 583, 2959–2963. doi: 10.1016/j.febslet.2009.08.004

Chen, X., Min, D., Yasir, T. A., and Hu, Y. G. (2012). Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PloS One* 7:e44510. doi: 10.1371/journal.pone.0044510

Chow, C. N., Zheng, H. Q., Wu, N. Y., Chien, C. H., Huang, H. D., Lee, T. Y., et al. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* 44, D1154–D1160. doi: 10.1093/nar/gkv1035

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server (2017). *Nucleic Acids Res.* 46, W49–W54. doi: 10.1093/nar/gky316

Dionisio-Sese, M. L., and Tobita, S. (1998). Antioxidant responses of rice seedlings to salinity stress. *Plant Sci.* 135, 1–9.

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Earl, D. A., and Bridgett, 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.

Fábián, A., Sáfrán, E., Szabó-Eitel, G., Barnabás, B., and Jäger, K. (2019). Stigma functionality and fertility are reduced by heat and drought co-stress in wheat. *Front. Plant Sci.* 10:244. doi: 10.3389/fpls.2019.00244

Ferris, R., Ellis, R. H., Wheeler, T. R., and Hadley, P. (1998). Effect of high temperature stress at anthesis on grain yield and biomass of field-grown crops of wheat. *Ann. Bot.* 82, 631–639. doi: 10.1006/anbo.1998.0740

Ganie, S. A., and Mondal, T. K. (2015). Genome-wide development of novel miRNA-based microsatellite markers of rice (*Oryza sativa*) for genotyping applications. *Mol. Breed.* 35, 1–12. doi: 10.1007/s11032-015-0207-7

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14, 685–695.

Goswami, S., Kumar, R. R., and Rai, R. D. (2014). Heat-responsive microRNAs regulate the transcription factors and heat shock proteins in modulating thermo-stability of starch biosynthesis enzymes in wheat ('*Triticum aestivum*'L.) under the heat stress. *Aust. J. Crop Sci.* 8, 697–705.

Gupta, P. K., Balyan, H. S., Sharma, P. C., and Ramesh, B. (1996). Microsatellites in plants: a new class of molecular markers. *Curr. Sci.* 70, 45–54.

Gupta, P. K., Rustgi, S., and Mir, R. R. (2013). "Array-based high-throughput DNA markers and genotyping platforms for cereal genetics and genomics," in *Cereal Genomics II*, eds P. K. Gupta and R. K. Varshney (Dordrecht: Springer), 11–55. doi: 10.1007/978-94-007-6401-92

Hassan, M. I. (2016). Assessment of genetic diversity in bread wheat genotypes based on heat tolerance and SSR markers. *Assiut. J. Agric. Sci.* 47, 37–55.

Heesacker, A., Kishore, V. K., Gao, W., Tang, S., Kolkman, J. M., Gingle, A., et al. (2008). SSRs and INDELs mined from the sunflower EST database: abundance,

polymorphisms, and cross-taxa utility. *Theor. Appl. Genet.* 117, 1021–1029. doi: 10.1007/s00122-008-0841-0

Jatan, R., and Lata, C. (2019). Role of microRNAs in abiotic and biotic stress resistance in plants. *Proc. Indian Natl. Sci. Acad*. 85, 553–567. doi: 10.16943/PTINSA/2019/49586

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. doi: 10.1093/nar/gky1141

Kumar, A., Chauhan, A., Sharma, M., Kompelli, S. K., Gahlaut, V., Ijaq, J., et al. (2017). Genome-wide mining, characterization and development of miRNA-SSRs in *Arabidopsis thaliana*. *bioRxiv* [Preprint] doi: 10.1016/j.leaqua.2012.05.006

Kumar, R. R., Goswami, S., Shamim, M., Dubey, K., Singh, K., Singh, S., et al. (2017). Exploring the heat-responsive chaperones and microsatellite markers associated with terminal heat stress tolerance in developing wheat. *Funct. Integr. Genomics* 17, 621–640. doi: 10.1007/s10142-017-0560-1

Kumar, R. R., Goswami, S., Sharma, S. K., Singh, K., Gadpayle, K. A., Kumar, N., et al. (2012). Protection against heat stress in wheat involves change in cell membrane stability, antioxidant enzymes, osmolyte, H_2O_2 and transcript of heat shock protein. *Int. J. Plant Physiol. Biochem.* 4, 83–91. doi: 10.5897/IJPPB12.008

Kumar, R. R., Goswami, S., Sharma, S. K., Singh, K., Gadpayle, K. A., Singh, S. D., et al. (2013). Differential expression of heat shock protein and alteration in osmolyte accumulation under heat stress in wheat. *J. Plant Biochem. Biotechnol.* 22, 16–26. doi: 10.1007/s13562-012-0106-5

Kumar, R. R., Pathak, H., Sharma, S. K., Kala, Y. K., Nirjal, M. K., Singh, G. P., et al. (2015). Novel and conserved heat-responsive microRNAs in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* 15, 323–348. doi: 10.1007/s10142-014-0421-0

Kumar, S., Deepali, M. D., Bansal, R., Kumari, J., and Kumar, S. (2016). Molecular characterization of Indian wheat germplasm line for stay green and other heat tolerance genes using linked SSR markers. *Int. J. Biol. Sci. Biotech Today* 6, 90–94. doi: 10.5958/2322-0996.2016.00014.4

Kumar, S., Kumar, M., Mir, R. R., Kumar, R., and Kumar, S. (2021). "Advances in molecular markers and their use in genetic improvement of wheat," in *Physiological, Molecular, and Genetic Perspectives of Wheat Improvement*, eds S. H. Wani, A. Mohan, and G. Pratap Singh (Cham: Springer), 139–174.

Madsen, B. E., Villesen, P., and Wiuf, C. (2008). Short tandem repeats in human exons: a target for disease mutations. *BMC Genomics* 9:410. doi: 10.1186/1471-2164-9-410

Manjunatha, P. B., Sinha, N., Krishna, H., Chauhan, D., Kumar, P., Kumar, R. R., et al. (2021). Exploration of heat stress-responsive markers in understanding trait associations in wheat. *J. Plant Biol.* 64, 167–179. doi: 10.1007/s12374-020-09289-9

Mir, R. R., Hiremath, P. J., Riera-Lizarazu, O., and Varshney, R. K. (2013). "Evolving molecular marker technologies in plants: from RFLPs to GBS," in *Diagnostics in Plant Breeding*, eds R. K. Varshney and T. Lubberstedt (Dordrecht: Springer), 229–247.

Mir, R. R., Kumar, J., Balyan, H. S., and Gupta, P. K. (2012a). A study of genetic diversity among Indian bread wheat (*Triticum aestivum* L.) cultivars released during last 100 years. *Genet. Resour. Crop Evol.* 59, 717–726.

Mir, R. R., Kumar, N., Jaiswal, V., Girdharwal, N., Prasad, M., Balyan, H. S., et al. (2012b). Genetic dissection of grain weight (GW) in bread wheat through QTL interval and association mapping. *Mol. Breed.* 29, 963–972.

Mir, R. R., and Varshney, R. K. (2013). "Future prospects of molecular markers in plants," in *Molecular Markers in Plants*, ed. R. J. Henry (New York, NY: Wiley), 169–190.

Mitra, R., and Bhatia, C. R. (2008). Bioenergetic cost of heat tolerance in wheat crop. *Curr. Sci.* 94, 1049–1053.

Moura, D. S., Brito, G. G., Campos, ÂD., Moraes, ÍL., Fagundes, P. R., and Deuner, S. (2017). Phenylalanine ammonia-lyase and source-flow-sink related attributes in rice genotypes subjected to high night temperatures. *J. Agric. Sci.* 9:12. doi: 10.5539/jas.v9n12p268

Nagy, S., Poczai, P., Cernak, I., Gorgi, A., Hegedus, G., and Taller, J. (2012). PICcalc: an online programme to calculate polymorphic information content for molecular genetic studies. *Biochem. Genet.* 50, 670–672. doi: 10.1007/s10528-012-9509-1

Ohama, N., Sato, H., Shinozaki, K., and Yamaguchi-shinozaki, K. (2017). Transcriptional regulatory network of plant heat stress response. *Trends Plant Sci.* 22, 53–65. doi: 10.1016/j.tplants.2016.08.015

Peng, L., Wan, X., Huang, K., Pei, L., Xiong, J., Li, X., et al. (2019). AtPUB48 E3 ligase plays a crucial role in the thermotolerance of *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 509, 281–286. doi: 10.1016/j.bbrc.2018.12.123

Perrier, X., Flori, A., and Bonnot, F. (2003). "Data analysis methods," in *Genetic Diversity of Cultivated Tropical Plants*, eds P. Hamon, M. Seguin, X. Perrier, and J. C. Glaszmann (Enfield, NH: Science), 43–76.

Petersen, G., Seberg, O., Yde, M., and Berthelsen, K. (2006). Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol. Phylogenet. Evol.* 39, 70–82. doi: 10.1016/j.ympev.2006.01.023

Pritchard, J. K., Wen, W., and Falush, D. (2010). *Documentation for STRUCTURE Software: Version 2*. Chicago, IL: University of Chicago.

Qin, D., Wu, H., Peng, H., Yao, Y., Ni, Z., Li, Z., et al. (2008). Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using Wheat Genome Array. *BMC Genomics* 9:432. doi: 10.1186/1471-2164-9-432

Ragupathy, R., Ravichandran, S., Mahdi, M. S. R., Huang, D., Reimer, E., Domaratzki, M., et al. (2016). Deep sequencing of wheat sRNA transcriptome reveals distinct temporal expression pattern of miRNAs in response to heat, light and UV. *Sci. Rep.* 6, 1–15.

Rangan, P., Furtado, A., and Henry, R. (2020). Transcriptome profiling of wheat genotypes under heat stress during. *J. Cereal Sci.* 91:102895. doi: 10.1016/j.jcs.2019.102895

Ravichandran, S., Ragupathy, R., Edwards, T., Domaratzki, M., and Cloutier, S. (2019). MicroRNA-guided regulation of heat stress response in wheat. *BMC Genomics* 20:488. doi: 10.1186/s12864-019-5799-6

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell* 110, 513–520. doi: 10.1016/S0092-8674(02)00863-2

Saha, N. R., Islam, M. T., Islam, M. M., and Haque, M. S. (2020). Morpho-molecular screening of wheat genotypes for heat tolerance. *Afr. J. Biotechnol.* 19, 71–83. doi: 10.5897/AJB2019.17017

Sailaja, B., Mangrauthia, S. K., Voleti, S. R., Subrahmanyam, D., and Ravindra Babu, V. (2017). Expression analysis of novel microRNAs in rice during high temperature stress. *Bull. Environ. Pharmacol. Life Sci.* 6, 225–229.

Sharma, D., Jaiswal, J. P., Gahtyari, N. C., Chauhan, A., Chhabra, R., Saripalli, G., et al. (2020). Population structure, marker-trait association and identification of candidate genes for terminal heat stress relevant traits in bread wheat (*Triticum aestivum* L. em Thell). *Plant Genet. Resour.* 18, 168–178. doi: 10.1017/S1479262120000131

Sharma, D., Jaiswal, J. P., Singh, N. K., and Chauhan, A. (2018). Developing a selection criterion for terminal heat tolerance in bread wheat based on various mopho-physiological traits. *Int. J. Curr. Microbiol. App. Sci.* 7, 2716–2726.

Sharma, P., Mehta, G., Muthusamy, S. K., Singh, S. K., and Singh, G. P. (2021). Development and validation of heat–responsive candidate gene and miRNA gene based SSR markers to analysis genetic diversity in wheat for heat tolerance breeding. *Mol. Biol. Rep.* 48, 381–393. doi: 10.1007/s11033-020-06059-1

Singh, I., and Shah, K. (2014). Phytochemistry exogenous application of methyl jasmonate lowers the effect of cadmium-induced oxidative injury in rice seedlings. *Phytochemistry* 108, 57–66. doi: 10.1016/j.phytochem.2014.09.007

Singh, I., Smita, S., Mishra, D. C., Kumar, S., Singh, B. K., and Rai, A. (2017). Abiotic stress responsive miRNA-target network and related markers (SNP, SSR) in *Brassica juncea*. *Front. Plant Sci.* 8:1943. doi: 10.3389/fpls.2017.01943

Stief, A., Altmann, S., Hoffmann, K., Pant, B. D., Scheible, W. R., and Bäurle, I. (2014). *Arabidopsis* miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. *Plant Cell* 26, 1792–1807.

Su, P., Jiang, C., Qin, H., Hu, R., Feng, J., Chang, J., et al. (2019). Identification of potential genes responsible for thermotolerance in wheat under high temperature stress. *Genes* 10:174. doi: 10.3390/genes10020174

Tabkhkar, N., Rabiei, B., Lahiji, H. S., and Chaleshtori, M. H. (2020). Identification of a new set of drought-related miRNA-SSR markers and association analysis under drought stress in rice (*Oryza sativa* L.). *Plant Gene* 21:100220.

Tyagi, S., Kumar, A., Gautam, T., Pandey, R., Rustgi, S., and Mir, R. R. (2021). Development and use of miRNA-derived SSR markers for the study of genetic diversity, population structure, and characterization of genotypes for breeding heat tolerant wheat varieties. *PloS One* 16:e0231063. doi: 10.1371/journal.pone.0231063

Tyagi, S., Sharma, S., Ganie, S. A., Tahir, M., Mir, R. R., and Pandey, R. (2019). Plant microRNAs: biogenesis, gene silencing, web-based analysis tools and their use as molecular markers. *3 Biotech* 9:413.

Wang, Y., Sun, F., Cao, H., Peng, H., Ni, Z., Sun, Q., et al. (2012). TamiR159 directed wheat TaGAMYB cleavage and its involvement in anther development and heat response. *PloS One* 7:e48445. doi: 10.1371/journal.pone.0048445

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z., et al. (2010). Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 10:123. doi: 10.1186/1471-2229-10-123

Yan, J., Zhao, C., Zhou, J., Yang, Y., Wang, P., Zhu, X., et al. (2016). The miR165/166 mediated regulatory module plays critical roles in ABA homeostasis and response in *Arabidopsis thaliana*. *PLoS Genet.* 12:e1006416. doi: 10.1371/journal.pgen.1006416

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., et al. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* 46, W71–W75. doi: 10.1093/nar/gky400

You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D., et al. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*. 9:253. doi: 10.1186/1471-2105-9-253

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., et al. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. U.S.A.* 114, 9326–9331. doi: 10.1073/pnas.1701762114

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

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fgene.2021.716821

Precision Mapping of a Maize MAGIC Population Identified a Candidate Gene for the Senescence-Associated Physiological Traits

Marlon Caicedo[1], Eduardo D. Munaiz[2], Rosa A. Malvar[2], José C. Jiménez[3] and Bernardo Ordas[2]**

[1] Instituto Nacional de Investigaciones Agropecuarias (INIAP), Quito, Ecuador, [2] National Research Council of Spain (CSIC) Misión Biológica de Galicia, Pontevedra, Spain, [3] National Institute of Forestry, Agriculture and Livestock Research (INIFAP), Cuauhtémoc, Mexico

OPEN ACCESS

Edited by:
Ramsey Steven Lewis,
North Carolina State University,
United States

Reviewed by:
Bode A. Olukolu,
The University of Tennessee,
Knoxville, United States
Ryan Andres,
North Carolina State University,
United States
Marcio Resende,
University of Florida, United States

***Correspondence:**
Bernardo Ordas
bordas@mbg.csic.es
Eduardo D. Munaiz
d.munaiz.eduardo@gmail.com

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

Received: 29 May 2021
Accepted: 16 August 2021
Published: 04 October 2021

Citation:
Caicedo M, Munaiz ED, Malvar RA, Jiménez JC and Ordas B (2021) Precision Mapping of a Maize MAGIC Population Identified a Candidate Gene for the Senescence-Associated Physiological Traits. Front. Genet. 12:716821. doi: 10.3389/fgene.2021.716821

Senescence is an important trait in maize (*Zea mais* L.), a key crop that provides nutrition values and a renewable source of bioenergy worldwide. Genome-wide association studies (GWAS) can be used to identify causative genetic variants that influence the major physiological measures of senescence, which is used by plants as a defense mechanism against abiotic and biotic stresses affecting its performance. We measured four physiological and two agronomic traits that affect senescence. Six hundred seventy-two recombinant inbred lines (RILs) were evaluated in two consecutive years. Thirty-six candidate genes were identified by genome-wide association study (GWAS), and 11 of them were supported by additional evidence for involvement in senescence-related processes including proteolysis, sugar transport, and sink activity. We identified a candidate gene, Zm00001d043586, significantly associated with chlorophyll, and independently studied its transcription expression in an independent panel. Our results showed that Zm00001d043586 affects chlorophyl rate degradation, a key determinant of senescence, at late plant development stages. These results contribute to better understand the genetic relationship of the important trait senescence with physiology related parameters in maize and provide new putative molecular markers that can be used in marker assisted selection for line development.

Keywords: maize, senescence, chlorophyll index, physiological, phenotyping, MAGIC, GWAS

INTRODUCTION

Genetic studies of delay of leaf senescence indicate that it is a multigenic trait (Guiboileau et al., 2010; Kant et al., 2015), as a result of the interacting of several metabolic paths that involve nutrient absorption and remobilization (Hörtensteiner and Feller, 2002; Lim and Nam, 2005), and the disassembly of the photosynthesis apparatus (Balazadeh, 2014). These biochemical processes of chlorophyll breakdown during leaf senescence lead to physiological changes in the plant phenotype (Balazadeh, 2014). Phenotypic changes due to delay senescence has been described in several species, e.g., *Zea mais L.* (Gentinetta et al., 1986; Zhang et al., 2019), *Oryza sativa* (Ramkumar et al., 2019), or *Glycine max* (Guiamét et al., 1990). Delay of senescence has a favorable influence on

crop production: positively associated with plant lodging resistance (Duvick and Cassman, 1999; Munaiz et al., 2021), increasing biomass yield (Borrell et al., 2000), favoring grain yield in plants grown under drought conditions (Borrell et al., 2014), and heat stress tolerance (Pinto et al., 2016). One important aspect is that late senescence allows for a longer photosynthesis period during the life cycle of the plant and larger dry matter accumulation during grain filling, leading to crops with higher yield performance (Valentinuz and Tollenaar, 2004).

Quantitative trait loci (QTL) have been identified using traditional linkage mapping, multiparent advanced generation intercrosses (MAGIC) populations and association mapping in diverse panels. MAGIC's advantage compared to biparental populations is the possibility of simultaneous studies of more than two alleles, usually eight alleles (Valdar et al., 2006). Traditionally, QTL mapping has used biparental populations, e.g., Glowinski and Flint-Garcia (2018); however, multiple inbred founder methods, which directionally intermate several times to obtain a unique combination of the genetic material from all founders in a single inbred line with a unique mosaic of small allelic haplotypes, have been used recently. Thus, providing sufficient genetic base, the higher number of parents and recombination events of the MAGIC population are advantageous in comparison with biparental population, while keeping pedigrees and genetic structure known in both cases. The first one allows for higher allelic diversity, whereas the later can benefit from having smaller haplotypes regions. The main limitation of biparental QTL methods is the reduce resolution, primarily because of the smaller number of crossing-over events accumulated over a few generations. On the other hand, diverse association panels made up of nonrelated genotypes that have accumulated a higher number of recombination events since the last shared progenitor could increase this resolution and, thus, display a more restricted linkage disequilibrium between pairs of neighboring molecular markers (Hyten et al., 2007) benefiting of a larger number of marker density and coverage of the entire genome.

MAGIC populations have been used in QTL identification analysis in a wide range of crops including rice (Bandillo et al., 2013), wheat (Huang et al., 2012), fava bean (Sallam and Martsch, 2015), maize (Dell'Acqua et al., 2015), barley (Sannemann et al., 2018), and sorghum (Ongom and Ejeta, 2018). In addition, multiparental populations were proven to successfully identify QTL associated with physiologically related senescence genes (Camargo et al., 2016). One important aspect is that late senescence allows a longer time in the field and shows agronomic advantages, accumulates more dry matter during the grain filling, and leads to higher yields (Valentinuz and Tollenaar, 2004). In turn, the grain filling depends on two sources such as carbon and nitrogen: the first one is provided by the transfer of photo assimilates directly to the grain from the photosynthetically active leaves, and the second one is supplied by the redistribution of photo assimilates stored in the reserve tissues before or after flowering (Yang and Zhang, 2006). Therefore, high-throughput screening of physiological features serves as proxy for QTL identification of biochemical changes during leaf aging.

Leaf senescence is a highly complex process that requires the expression of specific genes (Penfold and Buchanan-Wollaston, 2014) during aging. Genes whose transcript expression are upregulated during senescence are referred as senescence-associated genes "SAGs" and have been extensively studied and identified (Weaver et al., 1998; Munaiz et al., 2020). Senescence-associated genes functional diversity studies suggests that leaf senescence is programmed involving various cellular events, including degradation processes (Quirino et al., 2000), chlorophyll degradation, coupled with chloroplast dismantle (Sakuraba et al., 2014; Jagadish et al., 2015). However, SAGs are not only expressed during aging natural senescence but also respond to other environment cues, such as stress response and hormones, which induce early senescence (Weaver et al., 1998). Senescence-associated gene genetic studies used dark incubation treatment to induce leaf senescence. Senescence-associated genes expressed under this treatment condition were reported as primarily associated to stress conditions (Kleber-Janke and Krupinska, 1997; Park et al., 1998) and were not as highly expressed during natural senescence (Becker and Apel, 1993; Weaver et al., 1998). On the other hand, other SAGs have been identified as primarily expressed during natural senescence and were not highly expressed due to other factors. One of these genes is SAG12 in *Arabidopsis*, a gene that encodes a cysteine protease influenced by auxin, cytokinin, and sugars. SAG12 homologous in barley (*Hordeum vulgare*), the HvS40 gene, has been used as a molecular marker of age-mediated leaf senescence because its expression inversely correlates with the photosynthetic efficiency decline in barley (Humbeck et al., 1996; Jehanzeb et al., 2017).

In this study, we analyzed phenotypic profiles derived from senescence physiological trait screening of a large, eight-founder, "MAGIC" maize population to evaluate the genetics underlying the temporal changes in key developmental traits such as chlorophyl, PII quenching, and fluorescence. Time differences at silking and 2 months later were measured and were mapped to the maize genome. Molecular markers were identified, and the transcript expression of candidate genes were individually assessed in a diverse panel of genotypes. We demonstrated that physiological traits helped to identify crucial putative markers that could not be captured otherwise. The candidate gene expression analysis approach provided further genetic evidence of the role of this gene during the plant crop development, and it may serve as an important gene for breeding senescence-related genotypes.

MATERIALS AND METHODS

Plant Material

The parents of the MAGIC population belong to the no stiff stalk heterotic group and most of them derived directly from different European landraces. This multiparent population was developed as described in Butron et al. (2019) and Jiménez-Galindo et al. (2019), and it is formed of 672 recombinant inbred lines (RILS) over six cycles of recombination in an eight-way cross synthetic. The diverse genetic base and the significant number of RILs

allows for a high resolution of putative QTL-associated regions as shown in other studies, for example, corn borer and fusarium ear rot resistance (Butron et al., 2019; Jiménez-Galindo et al., 2019), forage digestibility (Lopez-Malvar et al., 2021), maize stover yield and saccharification efficiency (López-Malvar et al., 2021), and early development in cold conditions (Yi et al., 2020).

Experimental Design and Trait Evaluation

In 2014 and 2015, a subset of the 672 RILs and the 8 founder inbred lines were evaluated at Pontevedra, Spain. The experiments were arranged in a modified augmented design with 16 blocks and 50 lines within each block, 42 RILs, and 8 testers. Each experimental plot was composed of 17 plants and a total of 2.45 m^2. Two seeds were sown per position at 0.18 m between plants and 0.80 m between rows. Plantlets were manually thinned to a single plant per position obtaining a final density at 70,000 plants ha^{-1}. Standard agronomical practices were carried out. Physiological traits were measured from the middle part of each ear leaf at two time points, silking stage (_1) and two months later (_2): chlorophyll index (Chlo) calculates the total chlorophyll content of a leaf, minimum chlorophyll fluorescence (F0) measures the light re-emitted by leaf chlorophyll molecules in the absence of photosynthetic light, and maximum quantum yield of photosystem II (Fv/Fm) that is a measure of the intrinsic (or maximum) efficiency of PSII. The trait measurements were taken from three plants of each plot using a chlorophyll content meter (CCM200, Opti-Sciences, Hudson, NH, United States) and chlorophyll fluorometer (OS-30p, Opti Sciences Inc., Hudson, NH, United States), respectively. The chlorophyll fluorescence was measured on leaves by the saturation pulse method, a method in which leaves are adapted in dark conditions for 20 min prior to the measurement. In addition, agronomic traits were evaluated: visual score (visual) was recorded according to a subjective visual scale of 1–5 (1 = dead leaves and 5 = completely green and healthy leaves), grain yield ($t{\cdot}ha^{-1}$), and flowering in days to silking.

Minimal level of fluorescence, F0, indicates the minimum energy required to excite chlorophyll and reactor centers to activate electron transport via PSII (Murchie and Lawson, 2013). The rise in F0 is usually associated with heat losses from the PSII and the inactivation of the PSII reaction centers leading to oxidative damage (reduction of Fv/Fm) (Long et al., 1994). Rate changes in Fv/Fm are due to a modification in the efficiency of non-photochemical quenching that can be seen as an indicator of plant photosynthetic performance. The plant will show photosynthesis inhibition under stress conditions when values are below 0.83, which is considered the optimum target value for most plant species (Björkman and Demmig, 1987; Johnson et al., 1993).

DNA Isolation and Genotyped by Sequencing

Genomic DNA was isolated from 672 RILs along with their 8 parental lines. Coleoptiles were collected and used for genomic DNA extraction using DNasy Plant Minikit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Quality and quantity of DNA was checked in a Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, United States) and on agarose gels. The complete set, progenitor, and the 672 RILs were genotyped by sequencing (GBS) (Elshire et al., 2011) at the Biotechnological Institute of the Cornell University with a total of 955,690 single nucleotide polymorphisms (SNPs). Chromosomes and positions were located according to the version 2 of the maize reference genome B73 (Sen et al., 2010). These SNPs were filtered using the Tassel program version 5.2.40 (Bradbury et al., 2007); subsequently, polymorphic SNPs that were sequenced in at least 50% of the RIL were selected. Minor allele frequency was set at 0.05 to eliminate monomorphic SNPs or with rare alleles. In addition, SNPs with more than two alleles and deletions and insertions were eliminated. Genotypes of both RIL or founders with more than 5% of heterozygous SNP were considered as lost data in the analysis. After filtering, a database was generated with 465 RIL common lines in both locations and the seven parents (except parent EP17 due to seed quality) with a total of 224,363 SNPs, which were used in the subsequent analyzes.

Statistical Analyses

Analysis of variance were performed in a two-way approach in SAS software 9.4 [version 9.4 (SAS Institute Inc., Cary, NC)], by environment and the two environments together to account for genotype by environment interaction. As detailed in the field experimental design section, RILs were grown in an augmented design. To calculate the variance attributable to each factor, we modeled each phenotype considering inbred lines as fixed effects, and environments and blocks as random effects. The model used was

$$y_{ijk} = \mu + g_i + \beta_k + t_j + gt_{ij} + \varepsilon_{ijk}$$

where μ is the overall mean, g_i is the fixed effect of the ith maize line, β_k is the random effect of kth block, t_j is the randomized effect of the jth environment, gt_{ij} is the effect of the ijth genotype by environment, and ε_{ijk} are residuals. Models were fitted in SAS 9.4 [version 9.4 (SAS Institute Inc., Cary, NC)] using mixed model procedure (PROC MIXED). Estimations of genotypic effects were recorded as the best linear unbiased estimator (BLUEs), and based on the combined analysis, heritabilities (h^2) were estimated for each trait according to Holland et al. (2003). In addition, pairwise genetic correlations (r_g) and phenotypic correlations (r_p) were calculated as described in Holland (2006).

Association Mapping

Genome-wide association analysis (GWAS) was performed in Tassel 5.2.40 (Bradbury et al., 2007) following the mixed linear model (MLM):

$$y = \mathbf{X}\beta + \mathbf{Z}\gamma + e$$

where y is the vector of the phenotypes (BLUEs of the RILs), β is a vector of fixed effects, including the SNP marker tested, γ is a vector of random additive effects (inbred lines), $\boldsymbol{X}$ and $\boldsymbol{Z}$ represent design matrices that relate and with β and γ, respectively, and e is a random residuals vector. The variances of the random lines were modeled as:

$$Var\ (\gamma) = \mathbf{K}\sigma_a^2$$

where **K** is the matrix n × n of kinship coefficients, and σ_a^2 is the estimated additive genetic variance (Yu et al., 2006). Estimates of maximum restricted likelihood of the variance components were obtained using the optimal compression level and P3D population parameters described in Zhang et al. (2010). Significantly associated SNPs were determined as the point where the observed and expected F test statistics deviated in the Q–Q plot, and the threshold was –logp > 4 (Gowda et al., 2015). To test for marker association, we used a threshold (–logp) > 4 and $p < 0.05$ as a cutoff value in the multiple QTL analysis.

The linkage disequilibrium (LD) measure (r^2) was determined in the regions containing each of the SNPs significantly associated with each phenotypic trait. For each significant SNP, linkage blocks were established using the Haploview software (Barrett, 2009). Within the LD blocks (≤1 Mbp), candidate genes were identified and characterized in the MaizeGBD genome browser (Harper et al., 2011). Additive effects of each haplotype were calculated by adding the phenotypic values of the haplotypes that showed higher values for the trait and subtracting the phenotypic values of the haplotypes that showed lower values for that trait.

Transcriptome Expression for QTL Validation

To validate the putative genes associated with the physiological traits, we used RNA-seq data described in Caicedo Villafuerte (2018). Briefly, RNA-sequencing library was prepared based on Illumina standard instruction according to TruSeq Stranded RNA LT (Illumina, San Diego, CA, United States). Quality control of the library DNA was evaluated, checking the concentration and the size distribution with the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, United States) to meet Illumina HiSeq 4000 PE100 platform system before the sequencing was performed. To assess read sequencing quality, we used FastQC (Andrews, 2010). Adapters contained in the reads were removed using the Cutadap program[1] (Martin, 2011). The complementary DNA (cDNA) sequence from B73 file was used as reference, downloaded from www.maizegdb.org, derived from Maize B73 genome assembly (Sen et al., 2010), using the Bowtie 2, version 2.3.0 (Langmead and Salzberg, 2012) and Tophat, version 2.1.1 (Trapnell et al., 2009) tools. Fold change differences and p-values of differential expressed genes were calculated in R environment (version 3.6.3[2]) using EdgeR package (version: 3.28.1) (Robinson et al., 2010). The gene transcript expression was evaluated in six inbred that differed by the senescence delay (**Supplementary Table 1**), inbred NC292 with early senescence, the inbred PHT10 with mid-early senescence and the inbred PHW52 with middle senescence phenotypes, the inbred PHHB9 with middle senescence phenotype, the inbred PA8637 with mid-late senescence phenotype, and the inbred PHW79 with late senescence phenotype. Time sequence samples from each genotype were taken at silking, 15 days after silking, 30 days after silking, 45 days after silking, 65 days after silking, and 90 days after silking depending on senescence earliness as described in **Supplementary Table 1**.

[1]http://cutadapt.readthedocs.io/en/stable/guide.html#illumina-truseq

[2]https://www.r-project.org/

RESULTS

Phenotypic Variation

Analysis of traits across times indicated, in general, that among the MAGIC lines and their parents, there was a substantial variation in senescence performance. Analysis of variance for all phenotypes evaluated showed significant differences at silking stage (_1) and 2 months (_2) after silking among RILs and founders ($p < 0.05$). All traits showed a decrease between silking stage (_1) and two months after silking (_2). For visual scores, senescence showed a 2.41-fold change decrease; for chlorophyll index (Chlo), a 5.25-fold decrease; for minimum fluorescence (F0), a 1.18-fold decrease; and for the PII quenching ratio, a decline of 2.15-fold (**Table 1**).

Founder parents of the MAGIC population also showed significant differences among them ($p < 0.05$) for all traits. For visual assessment, founder lines had a score above 3 units at silking stage, whereas the scoring 2 months later was significantly lower at ∼1 (**Table 1**). All other physiological values changed in ranking between the silking stage and the second measurement 2 months after silking, indicating the phenotypic variability of the parental lines of the MAGIC population. For example, Chlo ranged from 31.23 to 62.27 at silking, while 2 months later, this range was much lower, between 1.78 and 10.04. For the trait F_0, founders at silking stage had a minimum value at 57.86 and a maximum at 83.67, whereas 2 months after silking, F_0_2 ranged from 18.82 to 80.35. For PSII quenching, the values observed across all founder parents were at 0.740, except for EP43 that had the lowest rate at 0.679. However, 2 months after silking, F_v/F_m_2, the parent F473 had the highest value at 0.539.

Recombinant inbred lines (RILs) had a large phenotypic variability in both scoring times. At silking evaluation, for visual scoring, inbreds showed a phenotypic range between 2 and 5; Chlo range was between 6.8 and 74.6; F_0 ranged from 33.0 to 96.0; and F_v/F_m had a variation ranging from 0.446 to 0.811 units. At scoring time, 2 months after silking, visual phenotypes ranged from 1 to 4; Chlo phenotypic variation was between 0.7 and 56.9 units; F_0 ranged from 0.0 to 230.0; and for F_v/F_m, the phenotypic variation ranged from 0.006 to 0.787 units (**Table 1**).

Heritability and Correlation Analysis

Heritability was high for all the physiological traits. F_0_1 had the highest value at 0.81 followed by Chlo_1 at 0.75 and F_v/F_m _1 that had a 0.58 value. Heritability values were slightly lower for these traits when measured 2 months after silking at 0.63, 0.49, and 0.42, for Chlo_2, F_v/F_m _2, and F_0_2, respectively. However, visual aspect evaluated at silking point had the lowest heritability at 0.27, and increased substantially after 2 months with a 0.69 value for visual_2 (**Table 2**, above diagonal). We calculated pairwise genetic correlations among the four traits evaluated at two different time points obtaining 11 out of 28 correlations among pair of traits significantly correlated with $r > |0.70|$. Senescence visual assessment at silking stage (visual_1) showed positive relationships $r > 0.62$ within five out of eight pairwise comparisons. In addition, minimum chlorophyll fluorescence (F_0_2) evaluated 2 months after silking

TABLE 1 | Averages values of 465 RIL and means of founder parents for three physiological traits and senescence visual scores evaluated at two phenological stages in the MAGIC population across two environments in Pontevedra, Spain.

Genetic source	Physiological traits[y]							
	Silking evaluation[z]				Two months after silking evaluation[z]			
	Chlo_1	F_0_1	F_v/F_m _1	Visual_1	Chlo_2	F_0_2	F_v/F_m _2	Visual_2
RILs								
Means	43.54 ± 0.41	63.74 ± 0.20	0.740 ± 0.001	3.74 ± 0.01	8.18 ± 0.35	54.28 ± 1.29	0.344 ± 0.010	1.55 ± 0.03
Range	6.8–74.6	33.0–96.0	0.446–0.811	2–5	0.7–56.9	0.0–230.30	0.006–0.787	1–4
Parents								
A509	38.23	60.61	0.744	3.51	1.78	18.82	0.102	1.08
EP125	31.28	62.15	0.743	3.56	2.92	40.18	0.268	1.19
EP43	41.33	83.67	0.679	3.79	5.88	44.11	0.160	1.13
EP53	62.27	64.75	0.746	3.70	10.04	57.73	0.366	1.52
EP86	38.07	63.78	0.738	3.97	7.21	48.34	0.375	1.69
F473	53.45	67.01	0.729	3.72	4.64	80.35	0.418	1.37
PB130	44.19	57.86	0.755	3.60	8.85	61.32	0.539	1.50
EP17	–	–	–	–	–	–	–	–
LSD ($p > 0.05$)	8.30	1.74	0.02	0.54	8.17	36.16	0.27	0.39

[y] *Physiological traits where Chlo is chlorophyll index, F_0 is minimum fluorescence, F_v/F_m is maximum quantum yield of PSII, and Visual is visual aspect with scale 1 (senescence) to 5 (green).*
[z] *Subindex followed by _1 indicates that the trait was evaluated at silking point, and that followed by _2 means that it was evaluated 2 months after silking.*

was significantly correlated with six out of eight of the other traits compared, with correlation coefficients at $r > 0.23$ for the trait F_0_1 to a correlation r at 1.00 for visual_1 (**Table 2**). As for the phenotypic correlations, only two pairwise comparison had $r > 0.60$, visual_2 with Chlo_2 and Fv/Fm_2, while the remaining correlation coefficients were relatively low varying between 0.08 and 0.55 (**Table 2**, below diagonal). Among all the phenotypic correlations Fv/Fm_2–Visual_2 was the highest at 0.7.

Pearson's correlations of flowering time with the physiological traits at silking time point varied from –0.31 to 0.12; visual_1 had the lowest correlation at –0.31, and the trait F0_1 had the highest value at 0.12. Furthermore, correlations evaluated 2 months later ranged from 0.15 to 0.40, being Chlo_2 the lowest and visual_2 the highest values, respectively. For yield performance, Pearson's correlations with the physiological traits at silking time point ranged between –0.16 and 0.48, corresponding to the traits F0_1 and visual_1, respectively. Yield evaluated 2 months after silking varied from 0.13 to 0.24, corresponding to F0_2 and Chlo_2, respectively.

Quantitative Trait Loci

We developed a genome-wide association studies (GWAS) using the RILs population previously genotyped utilizing a GBS technology and assessed individual phenotypes at both time point, silking stage (_1) and 2 months after silking (_2) stage, to identify loci associated with all the traits. Each candidate SNP associated to the trait was then studied in Haploview to account for LD and to identify the haplotype interval linked to the trait of interest.

For visual senescence, GWAS analysis detected eight significant SNPs associated at silking stage on chromosome (chr) 2, chr3, chr4, and chr8 (**Supplementary Figure 1** and **Supplementary Table 2**). The percentage of phenotypic variance explained for each putative marker ranged from 5.1 to 7.7%. Visual senescence studied 2 months after silking resulted in 27 SNPs localized on chr3, chr6, chr7, chr9, and chr10. For these putative QTLs, the proportion of phenotypic variance explained ranged between 4.8 and 8.8% (**Supplementary Figure 2** and **Supplementary Table 2**). We observed that peak QTLs became highly significant haplotypes at time point, 2 months after silking (**Table 3**), detecting four QTL interval haplotypes. These intervals corresponded to the markers on chr3 with an interval region of 87 kbp between 88 and 89 Mb (90% confidence region), another marker on chr3 with an interval region of 7 kbp between 90.3 and 90.4 Mb, the marker on chr6 with the interval region of 471 kbp at 56–57 Mb, and the QTL at marker on chr10, S10_2034863 (**Table 3**).

For chlorophyll, we detected 32 SNPs at silking stage, distributed across all chromosomes, except chr2, chr6, and chr9 (**Supplementary Figure 1** and **Supplementary Table 3**). The percentage of phenotypic variance explained by a single SNP ranged from 4.0 to 8.3%, on average. The total number of unfavorable alleles detected across founder lines was twice as much of the favorable alleles: for instance, EP125, the founder with the lowest chlorophyll index at silking stage, had the highest number of unfavorable alleles (22) across all the parental lines. At time point 2 months after silking, 19 SNPs were detected, and they were located on chr1, chr3, chr4, chr5, and chr6 (**Supplementary Figure 2** and **Supplementary Table 3**). The percentage of phenotypic variance explained by a single SNP was, on average, slightly higher than at silking stage ranging between 4.7 and 8.7%. As expected, the favorable alleles positively controlling chlorophyl decreased as plant senescence developed resulting in ∼1/4 of the total alleles (21/85). After studying the linkage blocks for each of these putative significant SNPs, we identified three QTL interval haplotypes on chr1 at silking

TABLE 2 | Genotypic and phenotypic correlation coefficient matrix of physiological traits and senescence visual appearance evaluated in the MAGIC population EPS21 across two environments in Pontevedra, Spain.

Phenotype[x,w]	Chlo_1[z]	F_0_1	F_v/F_m_1	Visual_1	Chlo_2[z]	F_0_2	F_v/F_m_2	Visual_2
Chlo_1		–0.01	0.36*	0.70*	0.43*	0.32*	0,30*	0,17*
F_0_1	–0.08		–0.61	–0.03	0.05	0.23*	–0.03	–0.03
F_v/F_m_1	0.26*	–0.47		0.62*	0.26*	0.17	0.27*	0.29*
Visual_1	0.20*	0.01	0.13*		0.93	1.00*	0.96*	0.76*
Chlo_2	0.26*	–0.01	0.13*	0.20		0.75*	0.87*	0.83
F_0_2	0.14*	0.12*	0.08*	0.24*	0.47*		0.85*	0.89*
F_v/F_m_2	0.10*	0.02	0.10*	0.19*	0.55*	0.83*		1.00*
Visual_2	0.09*	–0.02	0.14*	0.24*	0.64	0.51*	0.60*	
h^2	0.75	0.81	0.58	0.27	0.63	0.42	0.49	0.69
Flow[y]	–0.15	0.12	–0.01	–0.31	0.15	0.36	0.31	0.40
Yield	0.18	–0.16	0.20	0.48	0.24	0.13	0.15	0.23

Below the line, heritabilities (h^2) and Pearson correlations with flowering time (flow) and yield.

[x] *Physiological traits where Chlo is chlorophyll index, F_0 is minimum fluorescence, F_v/F_m is maximum quantum yield of PSII. and Visual is visual aspect with scale 1 (senescence) to 5 (green).*

[y] *Flow is flowering time at silking measured in days after sowing, and yield was measured in tons per hectare (t/ha).*

[w] *Genetic correlations above the diagonal (r_g) and phenotypic correlations below the diagonal (r_p).*

[z] *Subindex followed by 1 and 2 indicates that the trait was evaluated at silking point or evaluated 2 months after silking, respectively.*

**Correlation coefficient exceeded two times their standard error.*

stage, and three QTL haplotypes when evaluation took place 2 months after silking, namely, two QTLs on chr1 and one QTL on chr3, altogether resulting in seven of the significant putative markers (**Table 3**). The QTL region on chr1 was significant at both evaluation times. However, chr3 was detected only at 2 months after silking evaluation time. On chr3, the significant SNP (p = 3.86E–6) was located at 201,538,092 bp explaining 5.3% of the phenotypic variation (**Supplementary Figure 2B**). These findings suggest that chr3 region may have more importance at late stages of plant development as the plant matures.

For PSII quenching, 19 SNPs were detected to be significantly associated at silking stage evaluation. There were six putative markers on chr1, in addition to one marker on each chr4, chr6, and chr10, respectively (**Supplementary Figure 1** and **Supplementary Table 4**). On average, the percentage of the phenotypic variance explained by a single SNP ranged from 4.2 to 8.3%, and the additive allelic substitution effect fluctuated between 4.97 and 10.16 units. The Manhattan plot depicted a high peak on chr1 (**Supplementary Figure 1C**) with the most significant SNP at position 47,775,156 bp (p = 9.35E–6) explaining 5.4% of the phenotypic variance. Quantitative trait loci analysis of PSII quenching evaluated 2 months after silking detected two SNPs significantly associated, one SNP located on chr3 and another one on chr6 (**Supplementary Figure 2** and **Supplementary Table 4**). The phenotypic variance explained by a single SNP was, on average, lower than the earlier stage, ranging between 5.4 and 5.5%. However, additive effects at loci were much higher than at silking point with 79.98 and 103.79 units (**Supplementary Table 4**). For the 2 months after silking measurement, we identified a QTL associated to a haplotype on the chromosomal region of chr3. This allelic region corresponding to marker S3_207022252 had an additive effect at 85.5, and it was located within an interval region of 0.04 kbp flanking markers at 20.7 Mb position (**Table 3**). This QTL was not significant at silking stage but was highly significant at later maturity of the plant development.

For fluorescence (F0), QTL analysis detected 43 SNPs significantly associated to chlorophyl fluorescence at silking stage evaluation. Among these putative QTL markers, 35 SNPs were found on chr1, 2 SNPs were detected on chr2, 3 SNPs on chr3, and 1 putative marker on chr5, ch6, and chr7, respectively (**Supplementary Figure 1** and **Supplementary Table 5**). On average, the percentage of phenotypic variation explained by a single SNP ranged between 0.043 and 0.108. The Manhattan plot depicted a high peak on chr1 (**Supplementary Figure 1D**) with the most significant SNP at position 26,037,654 bp (p = 8.44E–07) explaining 6.8% of the phenotypic variance. Fluorescence measured at 2 months after silking, we reported 20 putative markers significantly associated with this trait: 18 SNPs located on chr3, 1 marker on chr 5, and another 1 on chr8 (**Supplementary Figure 2**). On average, the percentage of the phenotypic variance explained by a single SNP ranged from 4.6 to 8.45%, and the additive effect ranged between 6.82 and 14.61, on average (**Supplementary Table 5**). The QTL peak had a significant SNP on chr3 located at 201,538,222 bp (p = 2.02E–6) explaining 6.3% of the phenotypic variation (**Supplementary Figure 2D**).

Candidate Genes and Expression Analysis

We used the putative QTLs significantly associated with all the traits to search for candidate genes annotated developing a BLAST search on the maize genome browser. These searches resulted in candidate genes for all the phenotypic traits studied. For visual senescence evaluated at matured stage (Visual_2), there were seven candidates on chr3. For chlorophyl, there were three candidate genes at silking stage (Chlo_1) located on chr1

TABLE 3 | QTLs mapped with significant SNPs associated across traits and evaluation time point.

Trait	Time	SNP	Flanking markers	A509	EP125	EP17	EP43	EP53	EP86	F473	PB130	TQTL	ΣEAT	ΣEAR	EAP
Visual	2MAS	S3_8974495	S3_8888862–S3_8976187	–0.8	–0.1, –0.2	0.4, –0.2	–0.1, –0.7	–0.1, –0.6	0.5, 0.2	0.4, –0.7	0.2, –0.7	87	1.4	0.8	0.1
Visual	2MAS	S3_9038739	S3_9035419–S3_9042531	0.04, 0.02	–0.1	0.2, –0.1	0.1, –0.2	–0.1, –0.2	0.6, 0.1	0.3, –0.1	0.4, –0.1	7	0.9	0.7	0.3
Visual	2MAS	S6_56810498	S6_56809291–S6_57280494	–0.4	–0.3, –0.4	–0.3, –0.4	–0.4	0.3, 0.2	–0.4	0.3, 0	–0.1, –0.3	471	0.7	0.2	0.0
Visual	2MAS	S10_2034863	S10_2034863–S10_2034863	0.0	0.0	0.2, 0	0.2	0.2, 0	0.0	0.0	0.0	0	0.2	0.2	0.2
Chlorophyl	Silking	S1_278163516	S1_278163476–S1_278352686	–10.7, –19.4	–16.5, –19.4	–8.2, –11	–3.6, –8	6.5, 2.8	3.1, –5.1	6.5	6.5	189	32.4	11.6	5.3
Chlorophyl	Silking	S1_278356935	S1_278353505–S1_278357906	3.5, –5.3	-2.8	7.9, 4.6	3.1, –3	7.9, 2.3	7.9, –3.9	7.2	7.2	4	13.3	9.5	0.3
Chlorophyl	Silking	S1_278992401	S1_278909244–S1_279068377	–1.6, –3.8	–4.5	–2.2, –4.5	–3.8	–2.4	–3.8, –4.5	–0.2	1.4	159	5.9	1.9	1.9
Chlorophyl	2MAS	S1_279901234	S1_279901234–S1_279901243	4.5	0.0	0.0	4.5	0.0	0.0	4.5	4.5	0.009	4.5	2.3	2.3
Chlorophyl	2MAS	S1_279901243	S1_279901234–S1_279901243	4.5	0.0	0.0	4.5	0.0	0.0	4.5	4.5	0.009	4.5	2.3	2.3
Chlorophyl	2MAS	S3_201194856	S3_201189615–S3_201228263	–2.6, –7	–5.9, –11.1	12.4, –7.5	1, –0.1	–2.1, –6.6	11.1, 5	–6.2, –11.1	–6.3	39	24.6	3.3	4.0
Chlorophyl	2MAS	S3_201538092	S3_201538092–S3_201545231	–19.1	–19.1	1.9, –19.1	1.9, –1.6	–19.1	–1.6, –3.5	–19.1	–19.1	7	10.5	1.5	1.5
Chlorophyl	2MAS	S3_201538103	S3_201538092–S3_201545231	–19.1	–19.1	1.9, –19.2	1.9, –1.7	–18.1	–1.6, –3.6	–19.1	–19.1	7	10.5	1.5	1.5
Chlorophyl	2MAS	S3_201538104	S3_201538092–S3_201545231	–19.1	–19.1	1.9, –19.3	1.9, –1.8	–17.1	–1.6, –3.7	–19.1	–19.1	7	10.5	1.5	1.5
Chlorophyl	2MAS	S3_201538113	S3_201538092–S3_201545231	–19.1	–19.1	1.9, –19.4	1.9, –1.9	–16.1	–1.6, –3.8	–19.1	–19.1	7	10.5	1.5	1.5
PSII Quenching	2MAS	S3_207022252	S3_207022212–S3_207022252	171.1, 11.1	11.1	171.1, 0	0.0	171.1, 11.1	171.1, 0	11.1	11.1	0.04	85.5	49.5	49.5

Linkage blocks with flanking markers, QTL length, haplotype that increases or decreases the effect, sum of the real and theoretical additive effect of the favorable haplotype, and additive effect of the parents for each SNP or SNPs group significantly associated with Chlo_1, Chlo_2, and Fv/Fm_2, evaluated in the EPS21 MAGIC population, in Pontevedra, Spain.

TQTL, QTL size in kbp; HF, haplotype that increases the additive effect; HD, haplotype that reduces the additive effect; ΣEAT, sum of the theoretical additive effect, resulting from the sum of the additive effects of all significant SNPs in the linkage block; ΣEAR, sum of the real additive effect, resulting from the sum of the additive effects of all significant SNPs in the linkage block and subtracted from the total average of the entire population; EAP, additive effect based on the genotype of the parental.

and five candidate genes on chr3 that were associated with the trait measured 2 months after silking (Chlo_2). Finally, for PSII quenching, we observed two candidate genes associated with this trait at late plant maturity (F_v/F_m_2) (**Table 4**).

The comparative genomics search resulted in a candidate gene associated with Chlo_2 (Zm00001d043586) that was annotated as an *Arabidopsis* homologous; this gene encodes for a protein involved in senescence regulation of the S40 family. To evaluate how plant aging affects the gene expression of Zm00001d043586, located on chromosome 3 at position 201,538,092 bp, we compared the transcript expression obtained from RNA-seq at different time points after silking on seven different genotypes. Time sequence samples from each genotype were measured at silking (d0), 15 days after silking (d15), 30 days after silking (d30), 45 days after silking (d45), 65 days after silking (d65), and 90 days after silking (d90) depending on senescence earliness as described in **Supplementary Table 1**. First, we compared gene expression between each time point relative to d0 (**Table 5**). Early senescence line NC292 showed highly significant differences at time point d15 with a log-fold change (FC) at 5.08 (p = 1.91E–14). Mid-early phenotypes such as PHT10 started to be significantly different at d30 with a logFC at 4.45 (2.22E–03). Mid-late phenotype, PA8637, showed gene expression difference to silking time at d45 with a logFC 4.76 (p = 3.25E–02). Finally, late senescence phenotype, PHW79, had a significant gene expression at 60 days after silking (d60) with a logFC 6.45 (p = 1.33E–04). Second, we studied the transcription differences between pairwise time sequence for each genotype (**Figure 1**). Remarkably, gene expression in NC292 was significantly different between d0 and d15, and remained not significant between d15 and d30, and d30 and d45. Mid-early and middle senescence phenotypes, such as PHT10, PHW79, and PHHB9, increased steadily the gene expression. However, genotypes PHW52 and PA8637 increased their expression later at d30 and declined after the d65 stage (**Figure 1**).

TABLE 4 | Candidate genes and physical positions for each significant SNP associated with Chlo_1, Chlo_2, and F_v/F_m_2, visual senescence, and their respective descriptions.

Chr[a]	Trait	SNP physical position (bp)	Gene ID and position (B73 RefGen_v2)	TAIR database description
1	Chlo_1	278,163,516	Zm00001d034071 (S1_278113793–S1_278126352)	Nuclear-encoded chloroplast stromal cyclophilin CYP20-3 (also known as ROC4). Protein is tyrosine phosphorylated, and its phosphorylation state is modulated in response to ABA in *Arabidopsis thaliana* seeds.
1	Chlo_1	278,356,935	Zm00001d034076 (S1_278418921–S1_278425448)	Cyclic nucleotide-regulated ion channel family. "defense, no death" gene (DND1) encodes a mutated cyclic nucleotide-gated cation channel; Same as CNGC2 (article ID 229): Cyclic nucleotide gated channel, activated by cAMP, conducts K^+ and other monovalent cations but excludes Na^+. Conducts Ca^{2+} into cells, which is linked to the generation of NO and the NO signaling pathway involved in the innate immune response to pathogens. CNGC2 could be the key step mediating bulk Ca^{2+} influx into leaf cells after unloading from the vascular and have no direct roles in the leaf development and HR.
1	Chlo_1	278,992,401	Zm00001d034099 (S1_278996225–S1_278999894)	Encodes a leaf-type ferredoxin: NADP(H) oxidoreductase.
1	Chlo_2	279,901,234	Zm00001d034137 (S1_279742698–S1_279744047)	Encodes a chloroplast localized subunit of casien kinase4, and the casein kinase II. CK2-mediated phosphorylation enhances the light-induced degradation of PIF1 to promote photomorphogenesis.
1	Chlo_2	279,901,243	GRMZM2G003421 (S1_279824465–S1_279826154)	Encodes a member of F-box proteins that includes two other proteins in Arabidopsis (ZTL and FKF1). Overexpression results in arrhythmic phenotypes for a number of circadian clock outputs in both constant light and constant darkness, long hypocotyls under multiple fluences of both red and blue light, and a loss of photoperiodic control of flowering time.
3	Chlo_2	201,194,856	Zm00001d043574 (S3_201165018–S3_201167249)	Chloroplastic NifS-like protein that can catalyze the conversion of cysteine into alanine and elemental sulfur [S (0)] and of selenocysteine into alanine and elemental Se [Se (0)].
3	Chlo_2	201,538,092	Zm00001d043586 (S3_201322694–S3_201323761)	Senescence regulator (protein of unknown function, DUF584)
3	Chlo_2	201,538,103	Zm00001d043589 (S3_201371185–S3_201375806)	MADS box gene negatively regulated by APETALA1
3	Chlo_2	201,538,104	Zm00001d043607 (S3_201861549–S3_201871591)	Encodes a hexokinase (HXK1) in the plant glucose-signaling network. Functions as a glucose sensor to interrelate nutrient, light, and hormone signaling networks for controlling growth and development in response to the changing environment
3	Chlo_2	201,538,113	Zm00001d043609 (S3_201966295–S3_201973589)	MAP kinase kinase 6. Encodes a member of the MAP Kinase Kinase family of proteins. It can phosphorylate MPK12 *in vitro*, and it can be dephosphorylated by MKP2 *in vitro* (ANQ1, ATMKK6, MKK6).
3	F_v/F_m_2	207,022,252	Zm00001d043800 (S3_207029164–S3_207037370)	Serine/threonine kinase family catalytic domain protein
3	F_v/F_m_2	207,022,252	Zm00001d043801 (S3_207037486–S3_207041394)	Alpha/beta-hydrolases superfamily protein
3	Visual_2	8,974,495	Zm00001d039566 (S3_8887555–S3_8888747)	(HSP17.6II) 17.6 kDa class II heat shock protein
3	Visual_2	8,974,495	Zm00001d039575 (S3_8980485–S3_8982638)	A single copy gene that encodes a protein with sequence similarity to tomato E8 (ACC oxidase, the last step in ethylene biosynthesis) involved in ethylene synthesis and fruit ripening in tomato. This gene is not induced by ethylene in siliques. The transcript is found in siliques, etiolated seedlings, leaves, stems, and flowers.
3	Visual_2	9,038,739	Zm00001d039578 (S3_9037860–S3_9042655)	Histidyl-tRNA synthetase, key player in translation and act early in protein synthesis by mediating the attachment of amino acids to their cognate tRNA molecules
3	Visual_2	9038,739	Zm00001d039579 (S3_9045175–S3_9047443)	phd34—PHD-transcription factor 34, RING/FYVE/PHD zinc finger superfamily protein
3	Visual_2	9,038,739	Zm00001d039581 (S3_9072834–S3_9073771)	ca5p15—CCAAT-HAP5-transcription factor 515, nuclear factor Y, subunit C13
6	Visual_2	56,810,498	GRMZM2G306677 (S6_56815895–S6_56820763)	A member of ARF GAP domain (AGD), *A. thaliana* has 15 members, grouped into four classes.
10	Visual_2	2,034,863	Zm00001d023261 (S10_2008922–S10_2010657)	Nucleotide-sugar transporter family protein
10	Visual_2	2,034,863	Zm00001d023262 (S10_2011814–S10_2017183)	brk3—brick3, (GRL, NAP1, NAPP) transcription activators
10	Visual_2	2,034,863	Zm00001d023263 (S10_2039154–S10_2041723)	Glutamine-fructose-6-phosphate transaminase that likely plays a role in UDP-N-acetylglucosamine biosynthesis.

[a]*Chr, chromosome; Chlo_1 and Chlo_2 = chlorophyll index evaluated at silking and two months after silking; F_v/F_m_2 = maximum quantum yield of photosystem II evaluated 2 months after silking; and Visual_2 = visual scoring senescence evaluated 2 months after silking.*

TABLE 5 | Expression profiles fold change and p-value significant differences of six divergent stay-green genotypes studied for the GRMZM2G311347 gene at seven different time points in days (d) after the compared flowering time (d0) as references, d15, d30, d45, d60, d75, and d90 evaluated in Galicia, Spain.

Genotype	SG[x]	Tagwise	LogFC [z]					
	Phenotype	Dispersion	d15 [y]	d30	d45	d60	d75	d90
NC292	Early	0.14	5.08	4.77	5.26	NA	NA	NA
			1.91E-14	1.95E-12	1.37E-14	NA	NA	NA
PHT10	Mid Early	0.17	2.20	4.45	6.05	11.07	NA	NA
			1.78E-01	2.22E-03	1.01E-06	1.48E-32	NA	NA
PHHB9	Medium	0.21	0.00	3.95	7.12	7.44	9.14	NA
			9.99E-01	3.56E-02	9.08E-07	1.89E-07	8.43E-12	NA
PHW52	Medium	0.24	0.00	3.28	5.86	9.88	8.39	NA
			9.97E-01	8.63E-02	1.42E-04	2.11E-16	1.24E-09	NA
PA8637	Mid late	0.81	–0.01	–0.01	4.76	8.37	7.11	NA
			9.78E-01	9.89E-01	3.25E-02	8.59E-05	9.02E-04	NA
PHW79	Late	0.48	0.00	0.00	0.00	6.45	7.63	7.63
			9.99E-01	9.96E-01	9.99E-01	1.33E-04	1.81E-06	1.81E-06

Shaded values represent significant fold change expression compare to d0, flowering time.
[x] Visual phenotypes representing early, mid early, medium mid late and late senescence.
[y] Denoted as d0 = flowering time, d15 = 15 days after flowering, d30 = 30 days after flowering, d45 = 45 days after flowering, d60 = 60 days after flowering, d75 = 75 days after flowering, and d90 = 90 days after flowering.
[z] LogFC denotes log-fold change expression with p-values underneath each value having references d0.

DISCUSSION

This research aimed to understand the genetic regions influencing senescence traits and the temporal changes between two key phenotypic stages silking (_1) and 2 months after silking (_2). We studied senescence in key physiological traits associated with plant aging: Chlo, calculates the total chlorophyll content of a leaf, F_0 measures the minimum chlorophyll fluorescence, and Fv/Fm that represents the maximum quantum yield of photosystem II. This strategy resulted in the detection of QTLs linked with novel markers, suggesting that physiological screenings should be further investigated in the future.

Visual phenotypic correlations were low, indicating that selections based on visual scores may not be effective. Pairwise genetic correlation between all traits identified high positive (≥0.60) correlations between Chlo_1, Fv/Fm_1, and visual_1, and chlo_2, F_0_2, Fv/Fm_2, and visual_2. For example, in maize, Yang et al. (2017) reported phenotypic correlations at 0.77 for Fv/Fm and visual scores. Other studies in sorghum obtained lower correlations at 0.15 (Sukumaran et al., 2016), perhaps influenced by the difference in plant morphology of these two species or the conditions of drought stress of the sorghum experiments. However, Sukumaran et al. obtained similar visual scoring range at flowering times between 1.3 and 4.6. In our study, there was also a negative correlation between F_0_1 and Fv/Fm_1, indicating that the light exposure at early stage and its efficiency may not be sufficient at early stages of flowering time. However, that was not the case at later development, 2 months after flowering, with a high positive correlation at 0.85, suggesting that the source-sink apparatus may be more efficient toward the grain filling.

Heritability results showed that F_0_1 had the highest value at 0.81, followed by Chlo_1 and Chlo_2, at 0.75 and 0.62, respectively. This results are in agreement with those obtained by Yang et al. (2017) at 0.66 and are higher than the ones observed in Cai et al. (2012), between 0.36 and 0.57. These differences may be due to the phenotypic diversity of the founder parents used in this work whose chlorophyll content did not differ more than 57.3 points. However, Yang et al. (2017) included founder female parent "Zheng58" with functional stay green phenotypes and still observed the similar range of values as in this research. This high F0 heritability would allow for selections with reduce minimum energy to excite chlorophyll and reactor centers with sufficient intensity to induce electron transport through PSII, a desirable trait in maize germplasm. Genetic correlations were high at early stage between Fv/Fm_1 with F0_1 (0.61) and moderate with Chlo_1 (0.36). Similarly, genetic correlations were high between the physiological 2 months after silking, with values at 0.73 and 0.87 for combinations F0_2 with Chlo_2 and Chlo_2 with Fv/Fm, respectively. Because of the high heritabilities and the genetic correlations, this would suggest that Chlo, F0, and Fv/Fm would be a reliable phenotypic technique to assess plant senescence. F0_1 would be of particular importance due to its negative Pearson correlation with yield (–0.15) and the possibility of evaluation at early stage of the plant cycle. In the association analysis, we identified a total of 227 individual associated SNPs for all the evaluated traits (data not shown); however, among all these detected SNPs, based on the number of homozygous lines for a given variant, the selection of QTLs should be carefully made, providing that these observations are below 58 (465 RIL/8 parent lines = 58 lines) and that could generate false positive results due to the unequal composition of the data and the reduce accuracy, suggesting that natural selection has probably existed. Therefore, the total QTL identified for each trait was smaller because two or more of these significant SNPs can form a single linkage block but only one QTL. For instance,

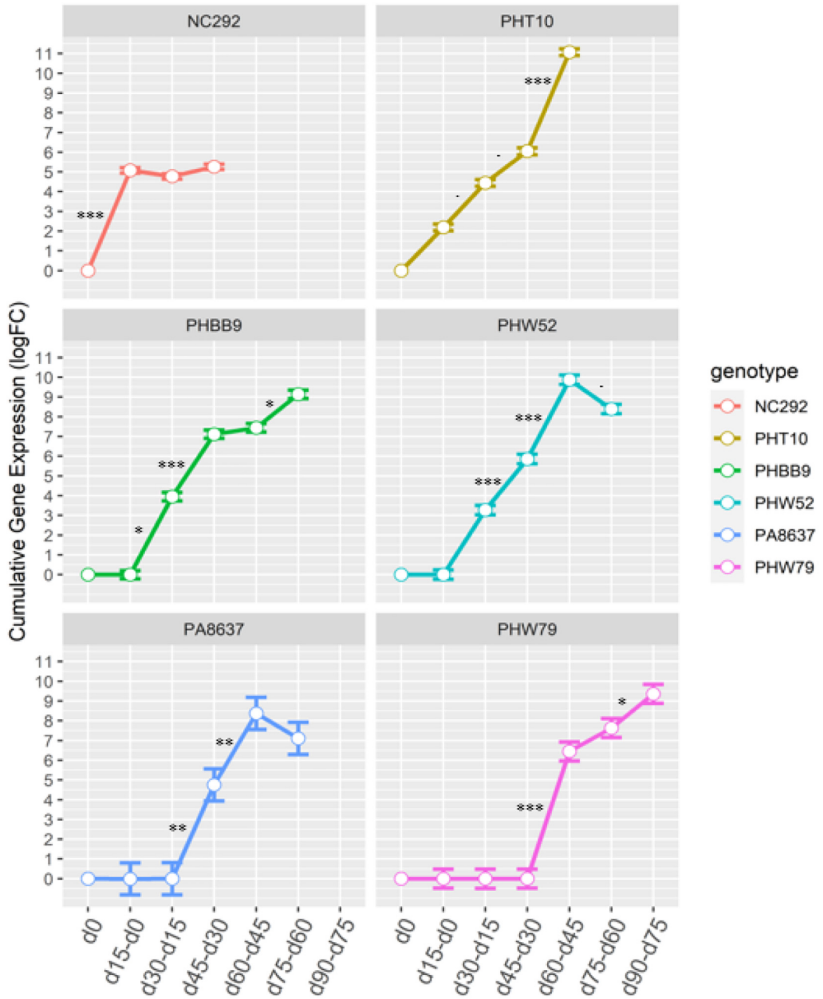


FIGURE 1 | Cumulative fold-change transcript expression of SNP "Zm00001d043586" at seven time points measured in six genotypes NC292, PHT10, PHHB9, PHW52, PA8637, and PHW79 evaluated in Pontevedra, Spain. Error bars represent the tagwise dispersion of the SNP for each genotype. [z]Denoted as d0 = flowering time; d15–d0 = expression between d15 and d0; d30–d15 = expression between d30 and d15; d45–d30 = expression between d45 and d30; d60–d45 = expression between d60 and d45; d75–d60 = expression between d75 and d60; d90–d75 = expression days after flowering..*, **, *** denotes significance at $p = 0.10$, $p = 0.05$, $p = 0.01$, $p = 0.001$, respectively.

there were two QTLs on chr1 in the traits PSII_1 and F0_1, at silking stage, but the candidate SNPs are located at 47,775,156 bp and 26,037,654 bp, respectively, which suggests that they are different haplotypes (**Supplementary Figure 1**). There was no common significant QTL (SNPs) for traits evaluated at both time points, silking and 2 months later, suggesting that each QTL was specific to the plant development time point at silking and senescence stages. However, the same QTL peak was detected for traits F0_2 and Chlo_2 on the chromosome 3 genomic region at approximately 201,538,092 bp, the same region as the Zm00001d043586 candidate gene, once more suggesting the importance of this region at late developmental stages.

Our research provided novel QTLs that have not been previously reported. These molecular markers were primarily found with a flint European origin, likely because it was the first time that senescence physiological traits were studied using this genetic background. We found QTLs associated with Chlo located on chromosomes 1, 3, 4, 5, 6, 7, 8, and 10, both at silking and 2 months after silking. Similarly, Yang et al. (2017) reported six QTL on chromosomes 1, 4, 6, 8, and 9 associated with this same trait. For the Fv/Fm trait, we detected associated QTLs positioned on chromosomes 1, 3, 4, 6, and 10, whereas other researchers found nine QTLs on chromosomes 1, 2, 3, 4, 5, 6, and 8 (Yang et al., 2017).

We found two genes associated with chlorophyll index (Chlo), represented by the SNPs on chromosome 1 significantly associated with Chlo_1 (S1_278163516, S1_278356935, and S1_278992401) located within the candidate gene Zm00001d034073 in the version 2 of the B73 genome and corresponding to the gene Zm00001eb057820 in the version 5. This gene is homologous to the WRKY DNA-BINDING

PROTEIN 57 gene (AT1G69310) in *Arabidopsis*, playing a role as a convergence point between jasmonic acid and auxin-mediated signaling during jasmonic acid-induced foliar senescence (Jiang et al., 2014). Although the SNP associated with this gene was only significantly associated with chlorophyll content at silking stage, and not at senescence stage, it was upregulated at later development based on the gene expression profiles studied in the seven maize inbred lines, confirming that Zm00001d034073 plays a role during naturally occurring leaf senescence.

On the other hand, SNPs significantly associated with Chlo_2 on chromosome 3 (S3_201538092, S3_201538103, S3_201538104, and S3_201538113) were located within the candidate gene Zm00001d043586. The homologous gene in *Arabidopsis* AT1G29640.1, also known as AtS40-1 (Fischer-Kilbienski et al., 2010), has been described as a senescence regulator. Proteins of the S40 family are induced during natural senescence and may also be regulated in response to hormone regulation (Krupinska et al., 2002; Zheng et al., 2019) because some proteins may only be expressed during dark-induced senescence (Kleber-Janke and Krupinska, 1997). For example, the SAG12 encoding protein identified in barley (*Hordeum vulgare*) produced by the nucleus gene HvS40 was associated with the degeneration of chloroplasts occurring during naturally occurred senescence (Krupinska et al., 2002). In our research, based on the gene Zm00001d043586 expression profiles studied in seven maize inbred lines, we detected that transcript expression was upregulated during senescence in six out of the seven lines studied when taking flowering time expression as reference. This suggests that the overall importance of this gene was due to natural induced senescence (**Table 5** and **Figure 1**). However, we observed that the pace of the senescence response regarding to the chlorophyll degradation showed clear differences based on the RNA-seq analysis. Early senescence lines showed high levels of expression from silking stage up to d45 (**Table 5**); however, for the middle senescence lines, the upregulation was initiated at d45 until d75, and for the late-senescence line, PHW79, transcription started at d60 to the end of the cycle, d90 (**Figure 1**). Thus, our results suggest that the gene Zm00001d043586 is associated with the speed of chlorophyll dismantle at late plant cycle stages. This may be due to a difference in the chloroplast degradation (Krupinska et al., 2002). Our results also suggest that this chloroplast degradation is germplasm dependent, depending upon the type of the diversity of the genotypes, because these changes happened at later stages of plant development within the different material. The general model of senescence proposed that the initial decline rates of photosynthesis may be the senescence-initiating factor (Hensel et al., 1993), and this is not the case of the SAG12 gene because it does not rapidly respond to senescence-promoting stresses, and it is only expressed in the leaves as they age (Weaver et al., 1998). Thus, we observed a similarly trend in maize represented in **Figure 1**, as reported by Noh and Amasino (1999) in *Arabidopsis* and by Zheng et al. (2019) in rice.

This study provides a systematic phenological characterization of a maize MAGIC population on two temporal time points conditions. We were able to identify not previously detected QTL loci for all the traits studied related to the onset of senescence at flowering time and also at late developmental stage, 2 months later. By using a multifounder maize population, we were able to capture much of the genetic variation present in elite cultivars and in particular the flint European genetic background. Thus, fine-scaling genomic regions of the developmental and physiological patterns can be exploited for senescence line development. We showed that the Zm00001d043586 gene was significantly associated with the chlorophyll rate, and it was of particular importance at later plant development due to its rate of gene expression response. This research supports the use of MAGIC population for QTL mapping that can assist in pyramidal selection using complementary alleles for crop enhancement.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: National Center for Biotechnology Information (NCBI) BioProject, https://www.ncbi.nlm.nih.gov/bioproject/, PRJNA746402.

AUTHOR CONTRIBUTIONS

BO: conceptualization. MC and BO: data collection. EM, JJ, RM, MC, and BO: data analysis. EM and MC: writing original draft. EM, RM, BO, JJ, and MC: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

We gratefully acknowledge the support of the Ministry of Innovation and Universities (MCIU), the Estate Agent of Innovation (AEI) of Spain, and the European Fund for Regional Development (FEDER) with the project PID2019-108127RB-I00, and the Ramon y Cajal foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.716821/full#supplementary-material

REFERENCES

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Bioinformatics, Babraham Institute.

Balazadeh, S. (2014). Stay-green not always stays green. *Mol. Plant* 7, 1264–1266. doi: 10.1093/mp/ssu076

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross

(MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6, 1–15.

Barrett, J. C. (2009). Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb. Protocols* 2009:db.i71.

Becker, W., and Apel, K. (1993). Differences in gene expression between natural and artificially induced leaf senescence. *Planta* 189, 74–79.

Björkman, O., and Demmig, B. (1987). Photon yield of O 2 evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. *Planta* 170, 489–504. doi: 10.1007/bf00402983

Borrell, A. K., Hammer, G. L., and Henzell, R. G. (2000). Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. *Crop Sci.* 40, 1037–1048. doi: 10.2135/cropsci2000.4041037x

Borrell, A. K., Mullet, J. E., George-Jaeggli, B., van Oosterom, E. J., Hammer, G. L., Klein, P. E., et al. (2014). Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *J. Exp. Bot.* 65, 6251–6263. doi: 10.1093/jxb/eru232

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

Butron, A., Santiago, R., Cao, A., Samayoa, L., and Malvar, R. (2019). QTLs for resistance to fusarium ear rot in a multiparent advanced generation intercross (MAGIC) maize population. *Plant Dis.* 103, 897–904. doi: 10.1094/pdis-09-18-1669-re

Cai, H., Chu, Q., Yuan, L., Liu, J., Chen, X., Chen, F., et al. (2012). Identification of quantitative trait loci for leaf area and chlorophyll content in maize (Zea mays) under low nitrogen and low phosphorus supply. *Mol. Breeding* 30, 251–266. doi: 10.1007/s11032-011-9615-5

Caicedo Villafuerte, M. (2018). *Mejora Genética de Maíz Para Senescencia Retrasada "Stay Green"*. Spain: PhD Universidad de Santiago de Compostela, Santiago de Compostela.

Camargo, A. V., Mott, R., Gardner, K. A., Mackay, I. J., Corke, F., Doonan, J. H., et al. (2016). Determining phenological patterns associated with the onset of senescence in a wheat MAGIC mapping population. *Front. Plant Sci.* 7:1540. doi: 10.3389/fpls.2016.01540

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.

Duvick, D. N., and Cassman, K. G. (1999). Post–green revolution trends in yield potential of temperate maize in the North-Central United States. *Crop Sci.* 39, 1622–1630. doi: 10.2135/cropsci1999.3961622x

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.

Fischer-Kilbienski, I., Miao, Y., Roitsch, T., Zschiesche, W., Humbeck, K., and Krupinska, K. (2010). Nuclear targeted AtS40 modulates senescence associated gene expression in Arabidopsis thaliana during natural development and in darkness. *Plant Mol. Biol.* 73, 379–390. doi: 10.1007/s11103-010-9618-3

Gentinetta, E., Ceppl, D., Lepori, C., Perico, G., Motto, M., and Salamini, F. (1986). A major gene for delayed senescence in maize. Pattern of photosynthates accumulation and inheritance. *Plant Breeding* 97, 193–203.

Glowinski, A., and Flint-Garcia, S. (2018). "Germplasm resources for mapping quantitative traits in maize," in *The Maize Genome*, eds J. Bennetzen, S. Flint-Garcia, C. Hirsch and R. Tuberosa (Berlin: Springer), 143–159. doi: 10.1007/978-3-319-97427-9_10

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., et al. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theoretical Appl. Genet.* 128, 1957–1968. doi: 10.1007/s00122-015-2559-0

Guiamét, J. J., Teeri, J. A., and Noodén, L. D. (1990). Effects of nuclear and cytoplasmic genes altering chlorophyll loss on gas exchange during monocarpic senescence in soybean. *Plant Cell Physiol.* 31, 1123–1130.

Guiboileau, A., Sormani, R., Meyer, C., and Masclaux-Daubresse, C. (2010). Senescence and death of plant organs: nutrient recycling and developmental regulation. *Comptes Rendus Biol.* 333, 382–391. doi: 10.1016/j.crvi.2010.01.016

Harper, L. C., Schaeffer, M. L., Thistle, J., Gardiner, J. M., Andorf, C. M., Campbell, D. A., et al. (2011). The MaizeGDB Genome Browser tutorial: one example of database outreach to biologists via video. *Database* 2011:bar016. doi: 10.1093/database/bar016

Hensel, L. L., Grbiæ, V., Baumgarten, D. A., and Bleecker, A. B. (1993). Developmental and age-related processes that influence the longevity and senescence of photosynthetic tissues in Arabidopsis. *Plant Cell* 5, 553–564. doi: 10.2307/3869710

Holland, J. B. (2006). Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. *Crop Sci.* 46, 642–654. doi: 10.2135/cropsci2005.0191

Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breed. Rev.* 22, 9–112.

Hörtensteiner, S., and Feller, U. (2002). Nitrogen metabolism and remobilization during senescence. *J. Exp. Bot.* 53, 927–937. doi: 10.1093/jexbot/53.370.927

Huang, B. E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., Morell, M. K., et al. (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol. J.* 10, 826–839. doi: 10.1111/j.1467-7652.2012.00702.x

Humbeck, K., Quast, S., and Krupinska, K. (1996). Functional and molecular changes in the photosynthetic apparatus during senescence of flag leaves from field-grown barley plants. *Plant, Cell Environ.* 19, 337–344. doi: 10.1111/j.1365-3040.1996.tb00256.x

Hyten, D. L., Choi, I.-Y., Song, Q., Shoemaker, R. C., Nelson, R. L., Costa, J. M., et al. (2007). Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175, 1937–1944. doi: 10.1534/genetics.106.069740

Jagadish, K. S., Kishor, K., Polavarapu, B., Bahuguna, R. N., von Wirén, N., and Sreenivasulu, N. (2015). Staying alive or going to die during terminal senescence-an enigma surrounding yield stability. *Front. Plant Sci.* 6:1070. doi: 10.3389/fpls.2015.01070

Jehanzeb, M., Zheng, X., and Miao, Y. (2017). The role of the S40 gene family in leaf senescence. *Int. J. Mol. Sci.* 18:2152. doi: 10.3390/ijms18102152

Jiang, Y., Liang, G., Yang, S., and Yu, D. (2014). Arabidopsis WRKY57 functions as a node of convergence for jasmonic acid–and auxin-mediated signaling in jasmonic acid–induced leaf senescence. *Plant Cell* 26, 230–245. doi: 10.1105/tpc.113.117838

Jiménez-Galindo, J. C., Malvar, R. A., Butrón, A., Santiago, R., Samayoa, L. F., Caicedo, M., et al. (2019). Mapping of resistance to corn borers in a MAGIC population of maize. *BMC Plant Biol.* 19:431. doi: 10.1186/s12870-019-2052-z

Johnson, G., Young, A., Scholes, J., and Horton, P. (1993). The dissipation of excess excitation energy in British plant species. *Plant Cell Environ.* 16, 673–679. doi: 10.1111/j.1365-3040.1993.tb00485.x

Kant, S., Burch, D., Badenhorst, P., Palanisamy, R., Mason, J., and Spangenberg, G. (2015). Regulated expression of a cytokinin biosynthesis gene IPT delays leaf senescence and improves yield under rainfed and irrigated conditions in canola (Brassica napus L.). *PLoS One* 10:e0116349. doi: 10.1371/journal.pone.0116349

Kleber-Janke, T., and Krupinska, K. (1997). Isolation of cDNA clones for genes showing enhanced expression in barley leaves during dark-induced senescence as well as during senescence under field conditions. *Planta* 203, 332–340. doi: 10.1007/s004250050199

Krupinska, K., Haussühl, K., Schäfer, A., van der Kooij, T. A., Leckband, G., Lörz, H., et al. (2002). A novel nucleus-targeted protein is expressed in barley leaves during senescence and pathogen infection. *Plant Physiol.* 130, 1172–1180. doi: 10.1104/pp.008565

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

Lim, P. O., and Nam, H. G. (2005). The molecular and genetic control of leaf senescence and longevity in arabidopsis. *Curr. Topics Dev. Biol.* 67, 50–85.

Long, S., Humphries, S., and Falkowski, P. G. (1994). Photoinhibition of photosynthesis in nature. *Annu. Rev. Plant Biol.* 45, 633–662.

López-Malvar, A., Butron, A., Malvar, R., McQueen-Mason, S., Faas, L., Gómez, L., et al. (2021). Association mapping for maize stover yield and saccharification efficiency using a multiparent advanced generation intercross (MAGIC) population. *Sci. Rep.* 11:3425.

Lopez-Malvar, A., Malvar, R. A., Butron, A., Revilla, P., Pereira-Crespo, S., and Santiago, R. (2021). Genetic dissection for maize forage digestibility traits

in a Multi-Parent Advanced Generation Intercross (MAGIC) population. *Agronomy* 11:104. doi: 10.3390/agronomy11010104

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Munaiz, E. D., Albrecht, K. A., and Ordas, B. (2021). Genetic diversity for dual use maize: grain and second-generation biofuel. *Agronomy* 11:230. doi: 10.3390/agronomy11020230

Munaiz, E. D., Martínez, S., Kumar, A., Caicedo, M., and Ordás, B. (2020). The Senescence (Stay-Green)—An important trait to exploit crop residuals for bioenergy. *Energies* 13:790. doi: 10.3390/en13040790

Murchie, E. H., and Lawson, T. (2013). Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64, 3983–3998. doi: 10.1093/jxb/ert208

Noh, Y.-S., and Amasino, R. M. (1999). Identification of a promoter region responsible for the senescence-specific expression of SAG12. *Plant Mol. Biol.* 41, 181–194.

Ongom, P. O., and Ejeta, G. (2018). Mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (Sorghum bicolor (L.) Moench). *G3: Genes Genomes Genet.* 8, 331–341. doi: 10.1534/g3.117.300248

Park, J.-H., Oh, S. A., Kim, Y. H., Woo, H. R., and Nam, H. G. (1998). Differential expression of senescence-associated mRNAs during leaf senescence induced by different senescence-inducing factors in Arabidopsis. *Plant Mol. Biol.* 37, 445–454.

Penfold, C. A., and Buchanan-Wollaston, V. (2014). Modelling transcriptional networks in leaf senescence. *J. Exp. Bot.* 65, 3859–3873. doi: 10.1093/jxb/eru054

Pinto, R. S., Lopes, M. S., Collins, N. C., and Reynolds, M. P. (2016). Modelling and genetic dissection of staygreen under heat stress. *Theoretical Appl. Genet.* 129, 2055–2074. doi: 10.1007/s00122-016-2757-4

Quirino, B. F., Noh, Y.-S., Himelblau, E., and Amasino, R. M. (2000). Molecular aspects of leaf senescence. *Trends Plant Sci.* 5, 278–282.

Ramkumar, M., Senthil Kumar, S., Gaikwad, K., Pandey, R., Chinnusamy, V., Singh, N. K., et al. (2019). A novel stay-green mutant of rice with delayed leaf senescence and better harvest index confers drought tolerance. *Plants* 8:375. doi: 10.3390/plants8100375

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

Sakuraba, Y., Lee, S.-H., Kim, Y.-S., Park, O. K., Hörtensteiner, S., and Paek, N.-C. (2014). Delayed degradation of chlorophylls and photosynthetic proteins in Arabidopsis autophagy mutants during stress-induced leaf yellowing. *J. Exp. Bot.* 65, 3915–3925. doi: 10.1093/jxb/eru008

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, 501–514. doi: 10.1007/s10709-015-9848-z

Sannemann, W., Lisker, A., Maurer, A., Léon, J., Kazman, E., Cöster, H., et al. (2018). Adaptive selection of founder segments and epistatic control of plant height in the MAGIC winter wheat population WM-800. *BMC Genomics* 19:559. doi: 10.1186/s12864-018-4915-3

Sen, T. Z., Harper, L. C., Schaeffer, M. L., Andorf, C. M., Seigfried, T. E., Campbell, D. A., et al. (2010). Choosing a genome browser for a Model Organism Database: surveying the maize community. *Database* 2010:baq007.

Sukumaran, S., Li, X., Li, X., Zhu, C., Bai, G., Perumal, R., et al. (2016). QTL mapping for grain yield, flowering time, and stay-green traits in sorghum with genotyping-by-sequencing markers. *Crop Sci.* 56, 1429–1442. doi: 10.2135/cropsci2015.02.0097

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120

Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172, 1783–1797. doi: 10.1534/genetics.104.039313

Valentinuz, O. R., and Tollenaar, M. (2004). Vertical profile of leaf senescence during the grain-filling period in older and newer maize hybrids. *Crop Sci.* 44, 827–834. doi: 10.2135/cropsci2004.0827

Weaver, L. M., Gan, S., Quirino, B., and Amasino, R. M. (1998). A comparison of the expression patterns of several senescence-associated genes in response to stress and hormone treatment. *Plant Mol. Biol.* 37, 455–469.

Yang, J., and Zhang, J. (2006). Grain filling of cereals under soil drying. *N. Phytol.* 169, 223–236. doi: 10.1111/j.1469-8137.2005.01597.x

Yang, Z., Li, X., Zhang, N., Wang, X., Zhang, Y., Ding, Y., et al. (2017). Mapping and validation of the quantitative trait loci for leaf stay-green-associated parameters in maize. *Plant Breed.* 136, 188–196. doi: 10.1111/pbr.12451

Yi, Q., Malvar, R., Álvarez-Iglesias, L., Ordás, B., and Revilla, P. (2020). Dissecting the genetics of cold tolerance in a multiparental maize population. *Theoretical Appl. Genet.* 133, 503–516. doi: 10.1007/s00122-019-03482-2

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Zhang, J., Fengler, K. A., Van Hemert, J. L., Gupta, R., Mongar, N., Sun, J., et al. (2019). Identification and characterization of a novel stay-green QTL that increases yield in maize. *Plant Biotechnol. J.* 17, 2272–2285. doi: 10.1111/pbi.13139

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

Zheng, X., Jehanzeb, M., Zhang, Y., Li, L., and Miao, Y. (2019). Characterization of S40-like proteins and their roles in response to environmental cues and leaf senescence in rice. *BMC Plant Biol.* 19:174. doi: 10.1186/s12870-019-1767-1

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

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

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

frontiers in Genetics

REVIEW
published: 08 November 2021
doi: 10.3389/fgene.2021.742095

Omics Technologies to Enhance Plant Based Functional Foods: An Overview

Spurthi N. Nayak[1], B. Aravind[1], Sachin S. Malavalli[1], B. S. Sukanth[1], R. Poornima[1], Pushpa Bharati[2], Kathleen Hefferon[3], Chittaranjan Kole[4] and Naveen Puppala[5]**

[1]Department of Biotechnology, University of Agricultural Sciences, Dharwad, India, [2]Department of Food Science and Nutrition, University of Agricultural Sciences, Dharwad, India, [3]Department of Microbiology, Cornell University, Ithaca, NY, United States, [4]President, International Phytomedomics and Nutriomics Consortium (ipnc.info), Daejeon, South Korea, [5]New Mexico State University-Agricultural Science Center at Clovis, New Mexico, NM, United States

Functional foods are natural products of plants that have health benefits beyond necessary nutrition. Functional foods are abundant in fruits, vegetables, spices, beverages and some are found in cereals, millets, pulses and oilseeds. Efforts to identify functional foods in our diet and their beneficial aspects are limited to few crops. Advances in sequencing and availability of different omics technologies have given opportunity to utilize these tools to enhance the functional components of the foods, thus ensuring the nutritional security. Integrated omics approaches including genomics, transcriptomics, proteomics, metabolomics coupled with artificial intelligence and machine learning approaches can be used to improve the crops. This review provides insights into omics studies that are carried out to find the active components and crop improvement by enhancing the functional compounds in different plants including cereals, millets, pulses, oilseeds, fruits, vegetables, spices, beverages and medicinal plants. There is a need to characterize functional foods that are being used in traditional medicines, as well as utilization of this knowledge to improve the staple foods in order to tackle malnutrition and hunger more effectively.

Keywords: nutrition, genomics, transgene, functional foods, nutraceuticals

OPEN ACCESS

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

Reviewed by:
Mehanathan Muthamilarasan, University of Hyderabad, India
Seetha Anitha, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India
Salej Sood, Indian Council of Agricultural Research (ICAR), India

***Correspondence:**
Naveen Puppala
npuppala@nmsu.edu
Chittaranjan Kole
ckoleorg@gmail.com

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

Received: *15 July 2021*
Accepted: *13 October 2021*
Published: *08 November 2021*

Citation:
Nayak SN, Aravind B, Malavalli SS, Sukanth BS, Poornima R, Bharati P, Hefferon K, Kole C and Puppala N (2021) Omics Technologies to Enhance Plant Based Functional Foods: An Overview. Front. Genet. 12:742095. doi: 10.3389/fgene.2021.742095

INTRODUCTION

To address global food and nutritional security, there is a need to increase the agricultural production and nutritive value of food. Assured access to nutritionally adequate and safe food is essential for attaining the nutritional security. With urbanization and changing food habits, "smart foods with higher nutrition per bite" is the need. The awareness of utilization of these foods for prevention and treatment of certain diseases prompted the researchers to discover active compounds that render health benefits. Foods that have an additional physiological benefits besides providing basic nutritional needs were first referred to as "functional foods" in Japan in the mid-1980s. Broadly, functional foods can be categorized according to the active components that have health benefits. Based on their origin, they can be classified as naturally derived products (plant or animal sources) or synthetic products (synbiotics, nutraceuticals) (Mohanty and Singhal, 2018). The functional products from plant origin include phytochemicals such as polyphenolic compounds, alkaloids, flavonoids, carotenoids, saponins, allyl sulfides, catechins, nutraceuticals, etc. (**Table 1**; **Figure 1**). There are clear evidences from epidemiological studies and clinical trials that a plant-based diet can reduce the risk of chronic diseases and disorders such as cancer (Velmurugan et al., 2005; Aghajanpour et al., 2017; Sayeed et al., 2017), diabetes (Hannan et al., 2007; Ballali and Lanciai,

TABLE 1 | Functional compounds and their health benefits.

Sl. No	Compound	Health benefits	References
1	Tocopherols, β-carotene	Antioxidants, reduce the risk of heart diseases and few types of cancers and protect from age-related muscular degeneration	Sies and Stahl, 1995; Gul et al., 2015, Jacobo-Valenzuela et al., 2011
2	α-linolenic acid	Cardioprotective in nature, modulation of an inflammatory response, and improves central nervous system functions	Stark et al. (2008)
3	Astaxanthin	Antioxidant and anti-inflammatory improves blood circulation and brain functions, promote an integrated immune response	Kidd, (2011)
4	Anthocyanins	Acts as dietary antioxidants helps to prevent neural diseases, cardiovascular problems, diabetes, inflammation and many other diseases	Yousuf et al. (2016)
5	Tannins	Antioxidant, anti-inflammatory, anticancerous, antiallergic, antihelminthic and antimicrobial activities	Sharma et al. (2019)
6	β-glucan	Beneficial role in insulin resistance, dyslipidemia, hypertension, and obesity	El Khoury et al. (2012)
7	Lycopene	Antioxidant, anticancer, protect against cardiovascular diseases, modulation of inflammatory responses, cholesterol reduction	Thies et al. (2017)
8	Flavonoids	Antioxidant, prevention of coronary heart diseases, hepatoprotective and anti-cancer activity	Yao et al., 2004; Yadav et al., 2020, Basu et al., 2018
9	Vitamin C	Prevent scurvy, coronary heart diseases stroke and cancer	Granger and Eck (2018)
10	Alkaloids	Analgesic, antipyretic, antioxidants, anti-inflammatory, improves brain functioning, antidiabetic and helps to treat gastroenteritis and chronic diseases	Derosa et al., 2016; Street et al., 2017, Adams et al., 2014
11	Saponins	Lowers blood lipids, lower blood glucose response and cholesterol levels, reduce cancer risks	Shi et al. (2004)
12	Eugenol	Antioxidants, anti-inflammation, helps to control hyperglycemia, elevated cholesterol levels, neural disorders and cancer. Also, possess antimicrobial agent	Khalil et al. (2017)
13	Polyphenols	Neuroprotectve, anti-aging, Antioxidant, anti-inflammatory	Lau et al. (2005)
14	Isothiocyanates	Lowers the risk of liver, breast, lung cancers	Aghajanpour et al., 2017; Mohanty and Singhal, 2018; Kartikey et al., 2019
15	Phytosterols	Anticancer, antibacterial, antiviral, and cholesterol-lowering activity	Chongtham et al. (2011)

2012; Alkhatib et al., 2017), obesity (Hill and Peters, 2002; Riccardi et al., 2005; Baboota et al., 2013), cardiovascular ailments (Alissa and Ferns, 2012; Hamid and Abd Hamid, 2019) and other effects on human health (Lobo et al., 2010). Most of the functional foods with scientific supporting evidence are the native/familiar foods that were used in traditional medicine for generations (Hasler, 1998; Fokunang et al., 2011; Abbott, 2014).

The scientific advances and next generation sequencing technologies available in recent years have impacted significantly on crop breeding and food science (Varshney et al., 2009; Kato et al., 2011). There is a lot of scope to utilize these technologies to understand the functional compounds that have health benefits and to improve the crops with respect to its nutritional status along with productivity related traits. Utilization of different omics technologies in research related to food and nutrition with the objective of improving the human health and well-being is referred as foodomics (Capozzi and Bordoni, 2013). An integrated use of omics technologies approaches to increase the nutrient potential of any crop, further applications in food processing and formulations can influence the nutritional security to greater extent (Bagchi et al., 2015; Tian et al., 2016).

The omics discipline comprises of four major broad areas like genomics, transcriptomics, proteomics and metabolomics. Integrated use of the omics technologies provides a holistic approach to study the systems biology (Pazhamala et al., 2021). Genomics includes the sequencing of whole genomes, assembly and annotation of the sequences, study of the genes, identification and development of molecular markers and quantitative trait loci (QTLs) for target traits, genomics assisted breeding, genomic selection, etc. (Varshney et al., 2005; Kole et al., 2015; McGuire et al., 2020). Transcriptomics deals with the dynamic expression of gene products in specific tissue at particular stage. The study of differential expression is quantified by using different molecular biology tools such as RNA sequencing, microarrays, Serial analysis of Gene Expression (SAGE), qRT-PCR, etc. While microarray, SAGE and qRT-PCR technologies determine the abundance of defined transcripts, the RNA-sequencing utilizes the advantage of high-throughput sequencing to identify the novel transcripts (Lowe et al., 2017). Proteomics can be effectively used to study protein structure, function, and interaction with other proteins or ligands such as bioactive compounds. Advanced techniques like Matrix-assisted laser desorption/ionization Time of flight (MALDI-TOF) and Liquid chromatography coupled to mass spectrometry (LC-MS) are able to detect expression of specific proteins. Metabolomics identifies and quantifies specific metabolites present in a sample. Metabolomics can be beneficial for quantification of biologically active compounds, food fingerprinting, and food profiling. Techniques like Gas chromatography coupled to mass spectrometry (GC-MS), Liquid chromatography coupled to mass spectrometry (LC-MS), Inductive couple plasma (ICP), nuclear magnetic resonance (NMR), Near infrared spectrometry (NIR) have been used for characterization of metabolites (Prakash et al., 2018; Kumar et al., 2019). Besides these omics approaches, genome editing tools like RNAi, CRISPR/Cas9, TALENs, ZFNs



FIGURE 1 | An infographics showing the functional components of the food and their potential health benefits to human beings.

can be utilized to improve the crop plants. Use of computational and bioinformatics tools is indispensable while using all the above mentioned technologies. Advances in data science with applications of artificial intelligence and machine learning has enabled deep learning of the data for better understanding of the biological processes and crop prediction modelling in genomic selections (**Figure 2**). In this review, we discuss about the utilization of omics technologies in determining and enhancing the active food compounds in major crop plants including cereals, millets, pulses, oil seeds, fruits, vegetables, spices and medicinal plants.

Cereals

Cereals are the major part of our daily diet and source of carbohydrates but lack an adequate amount of nutrition in terms of vitamins, and essential amino acids (Munck, 1972). Hence there is a need to improve the quality and nutritional parameters of cereals. Recent advances in genomics and genetic engineering are useful in targeted improvements especially by improving the quality and nutritional value in crop plants (Sedeek et al., 2019). Several omics technologies have been used to improve rice, wheat, barley especially for disease resistance and improving the yield of the crops (Zenda et al., 2021). However, there are only a few reports related to deciphering the functional compounds in cereal crops using modern biotechnological tools (**Table 2**). For instance, rice is improved with higher carotenoid content leading to increased Vitamin A (Dubock, 2019) and biofortified with micronutrients like Fe and Zn (Welch and Graham, 2004; Trijatmiko et al., 2016). A genetic engineering approach was successfully used to develop "Golden Rice" with significant levels of β-carotene that will help to combat vitamin A deficiency. Ye et al. (2000) and Shao et al.

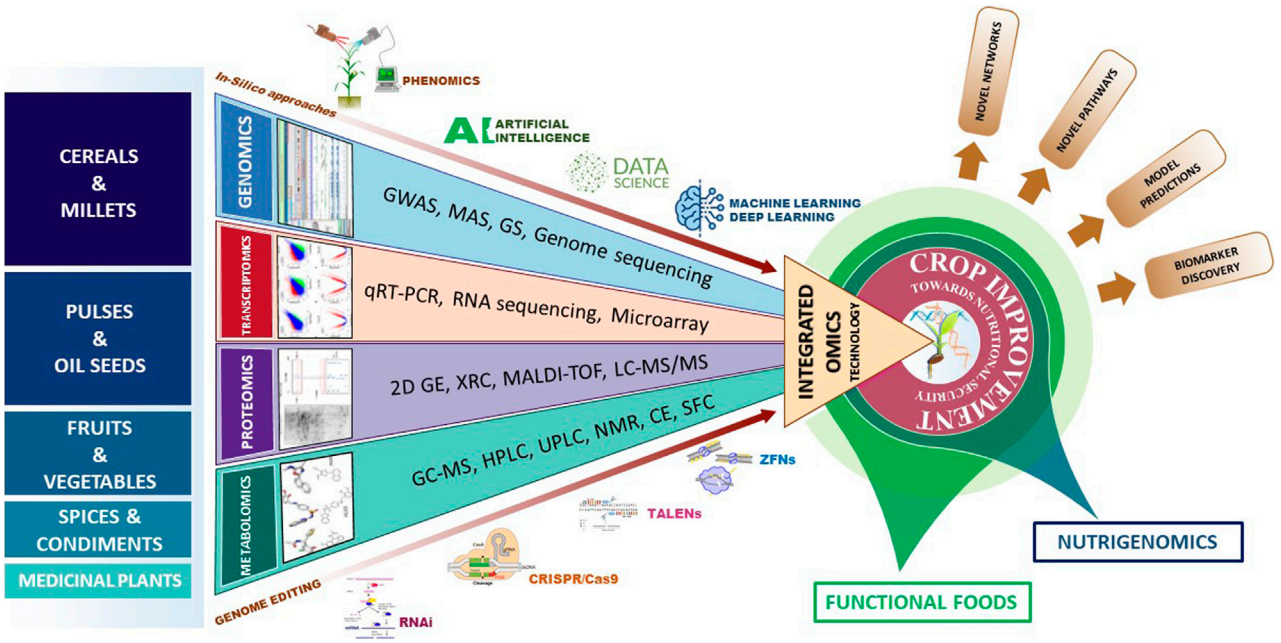


FIGURE 2 | Integrated omics approaches for enhancing functional foods. In this figure, the crop categories specified are orderly arranged in descending manner with respect to the utilization of omics technologies to improve functional foods. Abbreviations: GWAS, Genome-Wide Association Studies; MAS, Molecular Assisted Selection; GS, Genomic Selection; qRT-PCR, Quantitative Real Time polymerase Chain Reaction; 2D GE, 2-Dimensional Gel electrophoresis; XRC, X-Ray Crystallography; MALDI-TOF, Matrix-Assisted Laser Desorption/Ionization-Time Of Flight; LC-MS/MS, Liquid Chromatography—Mass Spectrophotometry; GC-MS, Gas Chromatography—Mass Spectrophotometry, HPLC: High Performance Liquid Chromatography; UPLC, Ultra Performance Liquid Chromatography; NMR, Nuclear Magnetic Resonance; CE, Capillary Electrophoresis and SFC, Supercritical Fluid Chromatography; RNAi, RNA interference; TALENs, Transcription Activator-Like Effector Nucleases; ZFNs, Zinc Finger Nucleases.

(2011) reported the marker loci/QTLs underlying the naturally occurring variations of grain color and nutritional quality traits in 416 rice germplasm accessions, including 361 white rice, 50 red rice, and six black rice across 41 marker loci. These markers could be further used for marker-assisted breeding to improve rice for nutritional qualities. The efforts were also made to dissect the nutrient traits especially Fe, Zn and anthocyanin content using genome-wide association studies on diversity panel consisting of 156 accessions of colored rice (Descalsota-Empleo et al., 2019). QTLs for functional components like phenolic content, flavonoid content and antioxidant capacity were identified using 127 double haploid lines developed through anther culture (Jin et al., 2009). Genome editing tools like CRISPR-Cas9 have also been utilized to enhance the amylose content (Sun Y. et al., 2017b). There are also efforts to develop fragrant rice by knocking out of betaine aldehyde dehydrogenase (*BADH2*) gene using TALEN technology (Shan et al., 2015).

Similar to rice, there are many reports to utilize the omics approaches in wheat to improve the nutrition and functional components. To list a few, an enriched wheat with high vitamin A content was developed by transforming the two bacterial carotenoid biosynthetic genes *CrtB* and *CrtI* into wheat cultivar Bobwhite (Wang et al., 2014). In addition to this, candidate genes involved in carotenoid biosynthesis and catabolism have been elucidated using GWAS studies in wheat (Colasuonno et al., 2017). Genomic regions for the color, carotenoids, and polyphenol oxidase activity of flour in wheat have been studied using linkage-based QTL analysis (Zhao et al., 2013). Hussain et al. (2017) reported QTLs for several nutrients, including Zn, Fe, Mn, Cu, Ca, Mg, etc. under saline conditions. In wheat, candidate genes for enhancing the grain Zn content have been identified by GWAS using high-density genotyping arrays on 369 wheat genotypes (Alomari et al., 2018). Genetic improvement in grain quality and micronutrients has been instrumental in quality breeding for wheat (Distelfeld et al., 2006; Balyan et al., 2013; Pu et al., 2014). Apart from this, there is also a need to utilize genomics approaches to decrease heavy metal (for example, Cadmium) uptake (Knox et al., 2009) and improve digestibility with reduced flatulence (Sharma et al., 2002). Genetically modified maize and wheat have showed increased accumulation of folate (Vitamin B9) levels (Liang et al., 2019).

Many economically backward countries rely on crops such as sorghum and maize as their staple food. In maize, biofortification with micronutrients like Zn and Fe (Zhao, 2007), enhanced β-carotene (Muthusamy et al., 2014), and amino acids like Lysine (Mertz et al., 1964; Shetti et al., 2020) have been carried out to ensure the nutritional security. To dissect the genomic regions for various metabolites in maize (*Zea mays*), a metabolome-based GWAS was carried out (Zhou et al., 2019). An integrated omics-based mapping to unravel flavonoid biosynthesis was also attempted in maize (Jin et al., 2017).

TABLE 2 | Study of functional foods in cereals and millets using biotechnological approaches.

S. No	Crop	Functional food	Gene(s)/QTL(s)	Methodology	References
1	Rice	β-carotene	Daffodil &*crtI* gene G*t*HMG1, G*Zm*Psy1 and G*Pa*CrtI1 genes	Transgenic and expression studies	Beyer et al. (2002) Tian et al. (2019)
		Fe and Zn	Ferritin	Transgenic and interval mapping	Lucca et al., 2002; Vasconcelos et al., 2003; Zhang et al., 2014
			genes (*Os*MTP6, *Os*NAS3, *Os*MT2D, *Os*VIT1, and *Os*NRAMP7) and 7 QTLs for each Fe and Zn	GWAS	Descalsota et al. (2018)
			48 MQTLs and 8 genes related to grain Fe and Zn concentration	MQTL analysis	Raza et al. (2019)
		α-linolenic acid rich	chimeric gene consisting of a maize *Ubi1-P-int* and a soybean *GmFAD3* cDNA	Transgenics	Anai et al. (2003)
		Astaxanthin	*s*ZmPSY1, *s*PaCrtI, *s*CrBKT, and *s*HpBHY genes	Transgenics	Zhu et al. (2018)
		α-tocopherol	*Os*GGR2 gene	RNA interference	Kimura et al. (2018)
		Phytic acid	OsITP5/6K-1 gene	RNA interference	Karmakar et al. (2020)
		Resistant starch	*sbe3-rs* gene	MAS	Yang et al. (2020)
2	Wheat	Micronutrients and Vitamins	*Gpc-B1* gene and DArT markers	MAS	Distelfield et al., 2006; Uauy et al., 2006
		Zn, Fe, Cu, Mn, Se rich	QTLs for Zn, Fe, Cu, Mn, Se	Interval mapping	Pu et al. (2014)
		Anthocyanins	*Ba* gene *Pp3 and Pp-D1* genes	MAS	Gordeeva et al., 2019; Gordeeva et al., 2020
3	Wheat and Barley	PUFAs	Artificial D6-desaturase gene	Transgenics using the biolistic method	Čertík et al. (2013)
4	Sorghum	Lysine	*BHL-9*	Transgenics	Zhao et al. (2003)
		Protein	*hl* gene and P721 opaque gene	Mutation breeding and MAS	Axtell et al., 1979; Welch and Graham, 2004
		Vitamin A	Prolamin and lysine alpha-ketoglutarate reductase genes z	Transgenics	Lipkie et al. (2013)
		Fe and Zn	QTLs and candidate genes like *CYP71B34, ZFP 8*	QTL mapping	Kotla et al. (2019)
		starch and amylose	Grain quality/starch pathway genes *Sh2, Bt2, Sssl, Ae1, and Wx*	GWAS	De Alencar Figueiredo et al. (2010)
		Tannin antioxidant	*tan-1, tan-1a* and *tan-1b*	GWAS and expression analysis	Wu et al. (2012)
		β-carotene andZeaxanthin	3 QTLs for β carotene and 4 QTLs forZeaxanthin	GWAS	Cruet-Burgos et al. (2020)
5	Maize	Vitamin C and E rich	DHAR cDNA	Transgenics	Chen et al. (2003)
		pro-vitamin A	*crtB* and *crtI*	Transgenics	Aluru et al. (2008)
			lcyE	MAS	Yang et al. (2018)
			*lcy*E, crtRB1, and o2	MABB	Sagare et al. (2019)
		Fe and Zn	SNPs associated with kernel Fe and Zn content	GWAS and QTL mapping	Hindu et al. (2018)
6	Barley	Hordothionin rich	Hordothionin	Mutation breeding and MAS	Rao et al. (1994)
		β-glucan	*Csl* genes and QTLs	GWAS and paired-end-RNA sequencing-based transcriptome	Cai et al., 2013; Chen et al., 2014; Shu and Rasmussen 2014
		Malting protein	13–30 candidate genes like *metallothionein, α-amylase, α-glucosidase, limit dextrinase, and β-ketoacyl synthase*	cDNA array-based gene expression analysis and SAGE	White et al., 2006; Lapitan et al., 2009
		Palatable and easily digestible	starch branching enzymes SBEIIa and SBEIIb	RNAi technology	Regina et al. (2010)
		—	20 QTLs associated with TPC, FLC and AOA	GWAS	Han et al. (2018)
7	Oats	β-glucan	QTL's for β-glucan	GWAS, MAS, QTL mapping	Gazal et al. (2014)
8	Pearl millet	Fe and Zn	QTLs (11 for Fe and 8 QTLs for Zn)	QTL mapping	Kumar et al. (2018)
9	Finger millet	Calcium	Calmodulin and *Cax1* transporter genes	Differential expression/ accumulation	Kumar et al. (2014)
10	Foxtail millet (*Setaria italica*)	storage associated genes	storage associated genes and noncoding RNAs	Transcriptome analysis	Qi et al. (2013)

There have been extensive efforts to breed for quality protein maize (QPM) with nearly as twice lysine and tryptophan content than the usual (Gibbon and Larkins, 2005). In addition to this, a transgenic approach has been used to increase protein by reducing zein content (Huang et al., 2006), and increasing Provitamin A content (Aluru et al., 2008). CRISPR-Cas9 and TALEN approaches have been used in maize to reduce phytic acid content, a food inhibitor that chelates micronutrients and prevents their bioavailability for mono gastric animals, including humans (Liang et al., 2014).

The reports related to the use of omics approaches to enhance functional compounds in barley (*Hordeum vulgare*) and oats

TABLE 3 | Study of functional foods in pulses and oilseeds using biotechnological approaches.

S. No	Crop	Functional food	Gene(s)/QTL(s)	Methodology	References
1	Soybean	Vitamin E	21 QTLs	QTL mapping	Li et al. (2010)
		α-tocopherol	6 QTLs associated with α-tocopherol content	QTL mapping	Park et al. (2019)
		—	19 QTLs were identified	GWAS	Sui et al. (2020)
		Tocopherol and tocotrienol	At-VTE3 co-expressed with At-VTE4	Transgenics	Van-Eenennaam et al. (2003)
2	Chickpea	β-carotene, leutin rich	1-4QTLs	QTL mapping	Abbo et al. (2005)
3	Groundnut	β-carotene, lutein and cryptoxanthin	Phytoene synthase 1 (*psy1*)	Transgenics	Bhatnagar et al. (2010)
		Anthocyanidin	a putative candidate gene and linked marker InDel02	eQTLmapping	Huang et al. (2006)
		Oleic acid	*ahFAD2* gene	MAS	Bera et al. (2018)
		Resveratrol	9 QTLs identified	ddRAD sequencing and High-Density genetic map	Luo et al. (2021)
4	Safflower (*Carthamus tinctorious*)	Gamma linolenic acid (GLA)	*Delta-6- desaturase* gene	Transgenics	Devi et al. (2008)
5	Mustard (*Brassica* spp.)	δ-tocopherol	*gamma-TMT* gene	Transgenics	Yusuf and Sarin, (2007)
		Carotenoid	*crtB*	Transgenics	Shewmaker et al. (1999)
		β-carotene, zeaxanthin, violaxanthin and lutein	Epsilon cyclase gene	RNAi technology	Yu et al. (2008)
6	Sunflower (*Helianthus annuus*)	Oleic acid	*FAD2*	Transgenics	Smith et al. (2007)
7	Canola	Protein-rich	*ACC7* gene	Transgenics	Roesler et al. (1997)

(*Avena sativa*) are limited as compared to major cereals. In barley, the β-glucan content greatly improves the malting properties and its presence has been found to increase palatability (Chen et al., 2014). A gene expression study using SAGE analysis identified six proteins associated with the malting property (White et al., 2006). Other functional compounds like total polyphenols, flavonoids, and antioxidant properties were studied in 67 cultivated and 156 Tibetan wild barley accessions using GWAS (Han et al., 2018). In oats, a GWAS study was conducted in a global germplasm collection to identify molecular markers associated with β-glucan content (Newell et al., 2012; Gazal et al., 2014).

Sorghum and Millets

Sorghum and millets are small-grained cereals and seed grasses that are traditional staple foods in African and Asian countries. In recent times, sorghum and millets are being utilized as an alternative to major cereals because of their higher nutritional, mineral, dietary fiber content along with climate-resilient nature. Besides, they are gluten-free and play a pivotal role in preventing and curing several lifestyle health issues like diabetes (Anitha et al., 2021). The genetic and genomic resources have been developed in some of the small millets (Vetriventhan et al., 2020) and efforts to utilize the genomic tools to improve the nutrient components are underway (**Table 2**). The nutraceutical property is mainly based on the kernel color in these crops. In this regard, QTL analysis of endosperm color and carotenoid (provitamin A) content in sorghum grains utilized in breeding high provitamin sorghum crop (Fernandez et al., 2008). Another effort using GWAS analysis with 404,628 SNP markers identified novel marker-trait association for polyphenols in a global diversity panel of 381 sorghum accessions (Rhodes et al., 2014).

There are limited efforts to utilize genomics tools in all other millets for improving the nutritional properties. The biofortification of millets seemed to be a good option for improving the nutritionally rich millets (Vinoth and Ravindhran, 2017). The QTLs controlling the content of micronutrients like Zn and Fe were identified in pearl millet (Kumar et al., 2018; Govindraj et al., 2019). Finger millet (*Eleusine coracana*) has been studied at various stages of growth and development using transcriptomics and was found to have high absorption and accumulation of calcium during grain development (Mirza et al., 2014). Glucosinolates in millets were found to reduce carcinogen-DNA interaction resulting in detoxification. Similarly, the isoflavones (phytoestrogens), genistein, and daidzein were found to reduce the incidence of many cancers, coronary heart diseases, and osteoporosis (Bandyopadhyay et al., 2017). The bioavailability of the micronutrients present in millets needs to be elucidated and utilized in crop improvement.

Pulses and Oilseeds

Pulses are a rich source of protein, with low fat, high fiber content and low glycemic index. Soluble fiber helps to decrease blood cholesterol levels and control blood sugar levels, and insoluble fiber helps with digestion. The biotechnological application for nutritional improvement mainly concentrates on enriching micronutrients and vitamins in pulses (**Table 3**). Pulses are known for the functional component saponins and several health benefits associated with them (Singh et al., 2017). Although pulses have been studied for several biotic and abiotic stresses at the molecular level, there are very few reports related to the genetic dissection of antioxidant activity

TABLE 4 | Study of functional foods in fruits and vegetables using biotechnological approaches.

S. No	Crop	Functional food	Gene	Methodology	References
1	Orange	Lycopene	Carotenoid and MEP pathway genes	Mutation breeding	Alquezar et al. (2008)
2	Pummelo	Naringin	Naringin	*In vitro* and *In vivo* studies followed by molecular docking	Cheng et al. (2020)
3	Apple	Astaxanthin	*bkt* and *crt*R-B genes	Transgenics	Jia et al. (2019)
		Flavanols, anthocyanins and hydroxycinnammic acid	79 QTLs identified for 17 polyphenolic content	QTL mapping and candidate gene mapping	Chagné et al., 2012
4	Grapes	Flavonols, anthocyanin and tannins	*VviGST1, VviGST3, and VviGST4*	Transgenic	Pérez-Díaz et al. (2016)
5	Watermelon	Lycopene	2 candidate genes Cla005011 and Cla005012	MAS	Wang et al. (2019a)
6	Walnut	Walnut Protein Hydrolysate (WPH)	Walnut Protein Hydrolysate (WPH)	*Invitro* and *In vivo* studies	Wang et al. (2019b)
7	Strawberry	Total flavonoids	7 QTLs and 2 candidate genes (*Fa*MYB1 and *Fa*F3′H) controlling flavonoid content identified	QTL analysis and Expression studies	*Karmakar et al.* (2020)
8	Tomato	Anthocyanin	Anthocyanin 1 (*ANT1*)	TALENs and CRISPR/Cas9 achieved gene	Čermák et al. (2015)
		Anthocyanin	Phytoene desaturase (*SlPDS*),	CRISPR/Cas9	Pan et al. (2016)
		Lycopene	*Lycopene β/ε -Cyclase*	RNAi technology and Agrobacterium-mediated gene transformation	Ma et al. (2011)
		Carotenoid	*Brassicajuncea*3-Hydroxy-3-methylglutaryl-coenzyme asynthase (*Bj*HMGS)	Mutation breeding	Galpaz et al., 2008, Liao et al. (2018)
		Carotenoids, Vit-C, Vit-E and Phenolic acids	7 QTLs for carotenoids, 6 for Vit-C, 5 for Vit-E, 3 for Glutathione, and a total of 43 QTLs for phenolic acids were identified	QTL mapping	Colak et al. (2020)
9	Cabbage	Anthocyanin	Purple (*Pr*) gene (flavonoid 3′-hydroxylase, dihydroflavonol 4-reductase, and leucoanthocyanidindioxygenase)	Transgenics	Chiu et al. (2010)
10	Carrot	Carotenoid	*DCAR 032551* gene	Genome assembly and Transcriptomics	Iorizzo et al. (2016)
11	Bell pepper	Alkaloid compound-Capsaicinoids	Deaminase (TD) and prephenate aminotransferase enzyme identified	*de novo* transcriptome assembly	Liu et al., 2013, Bennett and Kirby, 1968
12	Potato	Essential amino acid-rich protein and rich in methionine	*AmA1*	Transgenics	Chakraborty et al. (2010)
		Inulin producing	Constitutive expression of the *1-SST* and *1-FFT* (genes of globe artichoke)	Transgenics	Hellwege et al. (2000)
		β-carotene and lutein rich	*crtB* gene	Transgenics	Ducreux et al. (2005)
13	Chilli	β-carotene	lycopene beta-cyclase (β-Lcy) gene	Transgenics	El Nagar, (2018)
14	Brinjal	β-carotene	*crt*B gene	Transgenics	Mishiba et al. (2020)
16	Sweet potato	α-tocopherol	tocopherol cyclase (IbTC)	Transgenics	Kim et al. (2019)
17	Broccoli	Sulforaphane	MAM1, myrosinase and FMO_{GS-OX2} genes	Transgenics	Cao et al. (2021)

and nutrition-related traits. Biofortification of pulses with Fe and Zn in lentils, chickpeas and field pea (*Pisum sativum*) has been carried out to address global malnutrition and micronutrient deficiencies (Thavarajah and Gupta, 2014). In chickpea, the GWAS study conducted in 94 diverse chickpea genotypes showed eight SNPs associated with Fe and Zn content in the seeds (Diapari et al., 2014). Similarly, a GWAS study in lentils identified two tightly linked SNP markers for Fe and Zn content (Khazaei et al., 2017).

Omega-3 fatty acids are considered to be essential for brain development, which is mainly available through oil seed crops in human diet. The areas of crop improvement in terms of nutrition in oilseeds rely on improving oil quality, resveratrol content and improved shelf life (Pandey et al., 2014; Qi et al., 2014; Shasidhar et al., 2017; Luo et al., 2021). Profiling of nutraceutical properties of 60 groundnut cultivars differentiating in kernel colors has been carried out and marker-trait association studies have been carried out (Nayak et al., 2020). The expression of phytoene synthase showed 50-fold increased levels of carotenoids in rapeseed using genetic engineering (Shewmaker et al., 1999). An increased expression of zeaxanthin, violaxanthin and lutein by targeting the downregulation of the epsilon cyclase gene using RNAi technology) has been reported in mustard (Yu et al., 2008). The molecular mapping and QTL analysis of flavonoid genes was also elucidated in rapeseed (Qu et al., 2016), soybean (Li et al., 2016), and groundnut (Mondal et al., 2015). Efforts are being carried out to use advanced biotechnological applications to improve oilseeds nutritionally for further crop improvement (**Table 3**).

Fruits

Fruits are promoted as functional foods as they are a rich source of several antioxidants, polyphenols, minerals, soluble fibers, vitamins especially C, A and E. They primarily consist of

flavonoids including flavonols, flavones, isoflavones, flavanones and anthocyanins, and non-flavonoid polyphenolics including phenolic acids, lignans and stilbenes (Joy et al., 2018). Though fruits are the major source of functional foods, systematic experimental reports on the utilization of omics technologies to improve functional components are limited to a few fruit crops (**Table 4**).

In citrus, GWAS studies were conducted on 787 different citrus fruits using 1,841 SNP markers, and marker-trait associations were studied on fruit quality traits, including acid %, taste, and aroma (Minamikawa et al., 2017). Specific locus amplified fragment (SLAF) sequencing was performed over *C. reticulata* × *P. trifoliata* F_1 pseudo testcross population and have constructed a high density integrated genetic map with 3,817 markers. This study has identified 17 significant QTLs of which three colocalized genomic regions were observed for multiple carotenoid constituents (Zheng et al., 2018). In another study, a navel orange (*Citrus sinensis* L. Osbeck) mutant ("Cara Cara") was developed with bright red pulp with presence of lycopene (Alquezar et al., 2008). The expression analysis of genes involved in the carotenoid pathway using HPLC, northern hybridization, and RT-PCR indicated the increased accumulation of lycopene content in the mutant compared to navel orange. To elucidate the basis of lycopene accumulation in Cara Cara, the carotenoid profile and expression of three isoprenoids and nine carotenoid genes in flavedo and pulp of Cara Cara and Navel fruits throughout development and maturation were studied. The results indicated the accumulation of lycopene along with phytoene and phytofluene from early developmental stages in pulp as well as peel (Alquezar et al., 2008). Lemons are known for several functional components, including phenolics, vitamins, minerals, dietary fiber, essential oils and carotenoids (González-Molina et al., 2010). In the case of Sicilian blood oranges, retrotransposons were shown to induce seed-specific accumulation of anthocyanins during cold stress (Butelli et al., 2012). Edmunds et al. (2012) reported that the anti-inflammatory property of kiwifruit extract is due to the changes in the expression level of genes involved in the immune signaling pathway and metabolic processes using microarray technique.

The king of fruits "mango" (*Mangifera indica*) is a rich source of various polyphenolic compounds and is found in all the parts of the plant including pulp, peel, seed, bark, leaf, and flower. Mango polyphenols, especially mangiferin, acts as an antioxidant and has several health benefits (Masibo and He, 2008). The transcriptomics and proteomics studies in mango have predicted the involvement of genes involved in the anthocyanin biosynthesis pathway during the fruit development stage of mango (Wu et al., 2014). There is little effort towards the use of biotechnological approaches to improve the functional components of mango.

Red grapes are significant sources of anthocyanins, the main compounds responsible for the color of red grapes and wine (Mazza and Francis, 1995). Metabolite profiling of bioactive components of grapes especially flavonols, anthocyanins, and tannins indicated the presence of several bioactive compounds. The quercetin and kaempferol content was found to be greater in white grapes than red ones, but the red grapes were reservoirs of other bioactive components such as myricetin, laricitrin, syringetin and isorhamnetin (Mattivi et al., 2006). Resveratrol, an antioxidant that is known to lower blood pressure, and act as a chemopreventive with antiaging benefits are present in grapes. These flavonoids not only provide health benefits to humans but also help plants to fight against several biotic and abiotic stresses. For instance, transformation of bHLH transcription factor gene, *VvbHLH1* from grapes into *Arabidopsis,* resulted in an increased accumulation of flavonoids and enhanced salt and drought tolerance (Wang et al., 2016).

In Japanese plum (*Prunus salicina*), the molecular marker associated with transcription factors found in the flavonoid pathway was used to study population diversity (González et al., 2016). Date palm (*Phoenix dactylifera*) fruits are composed of minerals (Se, Cu, K, and Mg), vitamins (C, A, B6, B9, B2, B3) besides being a good source of total phenolics and natural antioxidants (such as anthocyanins, ferulic acid). Phenolic compounds and selenium present in date fruit impart antioxidant activity (Guizani, 2013). Similarly, transcriptome sequencing in Indian gooseberry (*Phyllanthus emblica*) revealed the genes involved in flavonoid and vitamin C biosynthesis (Kumar et al., 2016). In many fruits, biotechnological approaches, including "omics" studies and use of molecular markers for trait mapping to improve bioactive components are very limited.

Vegetables

Among vegetables, most of the genomics studies have been carried out in tomatoes as this crop is considered to be one of the model plants in genetic transformation and other genomics studies. The most critical functional component present in the tomato is carotenoids, especially lycopene and anthocyanins. To obtain lycopene-rich tomatoes, the genes encoding lycopene β/ε-cyclase, responsible for the conversion of lycopene to carotenoid, were silenced using RNAi technology. Significant increases in lycopene content were observed in transgenic plants (Ma et al., 2011). A mutation breeding approach was also used to increase the carotenoid content of tomatoes by 30%. Abscisic acid-deficient mutants in tomatoes have been shown to increase the lycopene content (Galpaz et al., 2008). Further, vegetables rich in anthocyanins were developed by overexpression of specific genes of the carotenoid biosynthesis pathway that induced a purple color, especially in tomato and cauliflower (*Brassica oleracea* var. botrytis) (Gonzali et al., 2009; Chiu et al., 2010).

Genome-editing technologies, especially CRISPR-Cas9, has potential use in horticultural crops (Karkute et al., 2017). Recently, this technique was used to edit five genes that are involved in the carotenoid pathway to increase lycopene content by inhibiting the conversion from lycopene to β- and α-carotene in tomatoes that increased lycopene content by five-folds (Li et al., 2018). In another study, intense purple-colored tomato plants were obtained by overexpressing an Anthocyanin mutant 1 (*ANT1*) gene that encodes for Myb transcription factors using TALENs and CRISPR/Cas9 approaches (Čermák et al., 2015). Furthermore, phytoene desaturase (*S1PDS*), an essential

TABLE 5 | Study of functional foods in beverages, spices and condiments using biotechnological approaches.

S. No	Crop	Functional food	Gene	Methodology	References
1	Coffee	Caffeine	N-methyltransferase genes, CaMXMT1	RNA interference method, Transgenics	Ashihara et al. (2008)
			65 caffeine associated SNPs identified	Genome sequencing and KEGG pathway-based analysis	Tran et al. (2018)
2	Tea	epigallocatechingallate, epigallocatechin, epicatechingallate	*CsANR1* and *CsANR2*	Expression in *E. coli*	Pang et al. (2013) Kim et al. (2014)
		Catechins and polyphenols	Demethylase gene	Transgenics followed by metabolic engineering	Yadav et al. (2020)
		Caffeine	27 QTLS were mapped to 8 linkage groups	2b-RAD Sequencing and High-Density genetic mapping	Xu et al. (2018)
3	Cocoa	Catechins and proanthocyanidins	Glycerol-3-phosphate acyltransferase (*GPAT*) genes, and lysophospholipid acyltransferase (*LPAT*) genes	Expression studies in yeast	Wei et al. (2018)
4	Cardamom	d-limonene	d-limonene	RNA sequencing Transcriptomics	Nadiya et al. (2017)
5	Clove	Eugenol and eugenyl acetate	Metabolites extracts	Gas chromatography/mass spectrometry	Lee and Shibamoto et al. (2001)
6	Black Pepper	Piperine	Piperine	Transcriptomics	Hu et al. (2015)
7	Garlic	Organic sulfur compounds	Acetolactate synthase (*ALS*) gene	Transgenics using the biolistic method	Park et al. (2002) Santhosha et al. (2013) Al- Safadi et al. (2000)
8	Fenugreek	Saponins	*diosgenin*	Gene expression studies	Ciura et al. (2017)
9	Saffron	Crocin made up of Apo carotenoids	Carotenoids	Induced mutation (gamma rays and chemical mutation)	Khan et al., 2011; Kyriakoudi et al., 2015

enzyme in carotenoid biosynthesis, and phytochrome interaction factor PIF 4 (*S1PIF4*) were targeted using gRNAs with the stable transformed CRISPR/Cas9 system (Pan et al., 2016). Most of the flavonoids in tomatoes are present in the peel of the fruit. Hence, a holistic approach of pathway engineering to increase the content of novel flavonoids especially stilbenes in the flesh of the tomato fruit was reported (Schijlen et al., 2006).

In carrot (*Daucus carota* subsp. Carota), a candidate gene, *DCAR_032551* that is responsible for carotenoid accumulation in carrot taproot and is co-expressed with several isoprenoid biosynthetic genes was identified from genome assembly and transcriptomic studies (Iorizzo et al., 2016). A candidate gene-based association study was carried out in carrots using 109 SNPs in 17 candidates/carotenoid biosynthesis genes over 380 diverse carrot cultivars, indicated the association of carotenoid content with the root color (Jourdan et al., 2015). A terpene synthase gene family of carrot was studied using QTL analysis and candidate gene-based association on a panel of carrot diversity set of 85 cultivars. GBS approach was used to genotype the panel with >168,000 SNPs (Keilwagen et al., 2017). Similarly, in bell pepper, several putative candidate genes are involved in the biosynthesis of capsaicinoids, such as Dihydroxyacid dehydratase (*DHAD*), Thr deaminase (*TD*) and Prephenate aminotransferase (*PAT*) were predicted from *de novo* transcriptome assembly (Liu et al., 2013).

Besides, several transcriptomics studies related to functional foods are available in crops such as lettuce (*Lactuca sativa*) (Zhang et al., 2017). In general, there is much scope to use genomics approaches to understand the molecular mechanisms and to increase the functional components in fruits and vegetables as evident by reports (**Table 4**).

Spices and Condiments

In spices and condiments, several studies have been carried out to profile metabolites, especially flavonoids, tannins, and alkaloids (Lee and Shibamoto, 2001; Shahidi and Ambigaipalan, 2015). In cinnamon (*Cinnamomum verum*), DART-QToF-MS method was utilized to discriminate true cinnamon from other species (Avula et al., 2015). There are limited reports on trait mapping in the case of spices (**Table 5**).

Most of the research in spices is related to the discovery of functional components. For instance, garlic has organic sulfur compounds as primary functional foods that have medicinal properties to reduce common cold, blood pressure and harmful cholesterol levels (Martin-Lagos et al., 1995). The functional food in turmeric is referred as curcumin, which acts as a acid neutralizer, blood purifier, tonic and antiseptic. The biological properties of curcumin were explored by using protein expression studies (Fang et al., 2011). The functional component of cardamom (*Elettaria cardamomum*) is d-limonene with antibacterial, anti-inflammatory, analgesic, and antispasmodic activities (Nadiya et al., 2017). Similarly, eugenol and eugenyl acetate, the functional components of clove (*Syzygium aromaticum*) are natural oxidants (Lee and Shibamoto, 2001). Coumarin, a functional component of cinnamon at lower doses has blood-thinning, anti-fungicidal and anti-tumor activities (Kawatra and Rajagopalan, 2015). Piperine from black pepper (*Piper nigrum*) has antioxidant, anti-inflammatory, and anticancer properties (Gorgani et al., 2017). Cumin (*Cuminum*

cyminum) has cuminaldehyde that enhances appetite, taste perception, digestion, vision, strength, and lactation. It is also used to treat diseases such as fever, loss of appetite, diarrhea, vomiting, abdominal distension, edema and puerperal disorders (Sowbhagya, 2013). Ginger (*Zingiber officinale*) has gingerols, shagols, and paradols with antioxidant, antimicrobial, and anti-inflammatory potential (Butt and Sultan, 2011). Nutmeg (*Myristica fragrans*) has tannin, flavonoid, and terpenoid which are natural antioxidants (Assa et al., 2014). Coriander (*Coriandrum sativum*) has carotenoids, polyphenols and essential oils, which provides vitamin A and vitamin C (Laribi et al., 2015). Fenugreek (*Trigonella foenum-graecum*) has quercetin, kaempferol and vitexin derivatives which are anti-diabetic and anti-nociceptive properties (Ghosh et al., 2015). Saffron (*Crocus sativus*) has crocins, picrocrocin, and safranal, which is antispasmodic, eupeptic, gingival sedative, carminative, diaphoretic activities (Melnyk et al., 2010). Using mutation breeding in saffron has increased yields (Khan et al., 2011) that in turn increases the overall bioactive components per plant.

Transgenic research is still an emerging area in spices and condiments. In garlic and turmeric (*Curcuma longa*), genetic engineering approaches were utilized for developing herbicide tolerant plants (Park et al., 2002; Shirgurkar et al., 2006). Although there are some reports on the transcriptome of black pepper fruits (Hu et al., 2015), ginseng (Panax ginseng) (Rai et al., 2016), and cardamom (Nadiya et al., 2017) to study global transcriptome, there are no reports related to functional components in most of the spices. There is tremendous potential to use genomics approaches including trait mapping, transcriptomics, whole-genome studies and allele mining in case of spices to demonstrate and increase the functional components.

Beverages

Beverage crops produce potable beverages other than water. Major beverage crops include Coffee (*Coffea* spp.), Tea (*Camelia sinensis*), Cocoa (*Theobroma cacao*), and Lemongrass (*Cymbopogon citratus*). Coffee has caffeine as the primary phenolic compound and is known to reduce the risk of stroke and cancer. Caffeine in higher doses is harmful as it may lead to insomnia, nervousness, restlessness, irritability, an upset stomach, a fast heartbeat, and even muscle tremors. As a result, there are efforts to improve decaffeinated coffee plants using RNAi technology (Ashihara et al., 2008). Tea has catechins and epicatechin as primary functional foods, and they are known to possess chemopreventive activities against prostate and ovarian cancers, anti-obesity and anti-diabetic effects. Efforts are underway to elucidate the proanthocyanidin pathway, also to reduce caffeine content (Pang et al., 2013). Lemongrass has citral as its primary functional food which has antimicrobial and medicinal properties. Little research has been performed on this crop.

QTLs for flavonoid-related traits in a tea were identified using a high-density genetic map (Xu et al., 2018). Several transcriptomics studies have been carried out in tea to elucidate genes involved in polyphenol synthesis, Catechin biosynthesis and other regulatory networks (Mamati et al., 2006; Wu et al., 2016; Sun P. et al., 2017). To knock down the expression of the genes involved in caffeine biosynthesis, RNAi was used to repress the expression of the gene encoding theobromine synthase (*CaMXMT1*) that reduced the caffeine content in the transgenic coffee plants up to 70% (Ogita et al., 2003). In the case of tea, the functional characterization of the proanthocyanidin pathway and potential applications in metabolic engineering was elucidated (Pang et al., 2013). Cocoa rich in catechins and proanthocyanidins has a promising effect on lowering blood pressure, boosting moods, and sharpening memory. Metabolic engineering of yeast for cocoa butter production was attempted by cloning the genes involved in triglycerol synthesis viz., glycerol-3-phosphate acyltransferase (GPAT), lysophospholipid acyltransferase (LPAT) from cocoa into yeast (Wei et al., 2018). Efforts are being made to develop lemongrass varieties such as Jor Lab L-8 with higher amounts of essential oil and herbage production (Mohan et al., 2016). The biotechnological applications have not been effectively utilized to increase the functional components in beverages and there are few reports related to this (**Table 5**).

Medicinal Plants

Medicinal plants are called so because of their antibiotic, antidiabetic, antihyperglycemic, and antihyperlipidemic properties. Most medicinal plants are not consumed as staple foods, but as preventive medicines for several diseases ranging from the common cold to complex diseases like cancer. Herbal genomics has high potential to explore, though there are few efforts related to molecular breeding and genetic engineering in the medicinal crops (Chakraborty, 2018). However, metabolite profiling of some medicinal plants has been studied. In the case of a famous Ayurvedic crop Haritaki (*Terminalia chebula*), a component of Triphala (an ayurvedic composition), the metabolite profiling of polyphenols and evaluation of the decoction as a chemopreventive agent was studied (Pellati et al., 2013). Similarly, metabolite profiling was examined in a highly traded South African medicinal plant commonly known as pain brush lily (*Scadoxus puniceus*) and the bioactive compounds were isolated (Naidoo et al., 2018). Efforts for profiling polyphenols, alkaloids and other bioactive compounds are being carried out in other Asian medicinal plants (Gibon et al., 2006; Vega-Gálvez et al., 2011; Gantait et al., 2014; Hao and Xiao, 2015; Saito, 2018). For instance, in the case of Candyleaf (*Stevia rebaudiana*), the water extracts from leaf and calli were shown to have antioxidant activity and contain bioactive compounds including folic acid, vitamin C, catechin, quercetin and pyrogallol. Higher reactive oxygen species (ROS) scavenging activities were found in leaf extracts (Kim et al., 2011). Transcriptomics studies have also been carried out in some of the important medicinal plants including Ashwagandha (*Withania somnifera*) to understand the secondary metabolites which have therapeutic utilization (Tripathi et al., 2020).

Recent advances in metabolite and pathway engineering and their utilization in medicinal plant research have positively contributed to herbal genomics research. Most of the molecular studies in medicinal plants involved either discovery of the genes/enzymes/pathways related to secondary metabolites or increasing the production of the secondary metabolites using

elicitors, hairy root cultures or metabolite engineering approaches.

FUTURE PROSPECTS

Current approaches in crop sciences using integrated omics platform aims at providing a nutritionally rich, diverse balanced diet to the society. Several leading edge technologies in understanding and manipulating different segments of scientific research areas *viz.* genomics, proteomics, metabolomics *etc.* has enabled the researchers to enhance contents of key nutrients in crop plants. Not just nutrition, but reducing the unflavorful compounds (phytic acid, acrylamide-forming amino acids, *etc.*) in food crops has allowed people to consume a wide range of food crops. Bio fortification has potential to solve nutrition deficiencies and in this view several food crops *viz.* rice, maize, wheat, *etc.* have been biofortified to have enhanced amounts of Fe, Zn, *etc.* (Ye et al., 2000; Gil-Humanes et al., 2014; Mugode et al., 2014; Trijatmiko et al., 2016). Crop improvement with new advancements in field phenomics, employing applications of machine learning (Niazian and Niedbala., 2020), nanotechnology and artificial intelligence (Ben Ayed and Hanana, 2021; Zhang et al., 2021), biosensors like lidar (Jin et al., 2021) followed by statistical analysis using data science (Tong and Nikoloski, 2021) approaches will enable researchers to precisely assess traits for plant breeding and development (Deery and Jones, 2021).

In order to ensure the nutrional security, along with enhancing the nutritional value, we need to work on reduced food-wastes that has a significant economic, environmental and social impact (FAO, 2019). Several initiatives in estimating food waste and prevention has been proposed (Moraes et al., 2021), however, devising methodologies in estimating and reducing food wastage is still a paradox (Richards et al., 2021). This can be featured as an opportunity to overcome malnutrition in addition to food waste reduction and stabilize bio-economy with sustainable processing of food waste into bio-based products (Sharma et al., 2021). The innovative technologies for extraction and microencapsulation of bioactives using novel technologies in metabolomics can be utilized in enhancing plant based functional foods (Pattnaik et al., 2021).

The research in nutrition and omics technologies in food science with epidemiological techniques should be classically established (Palou et al., 2004). In the future, the advances in foodomics and nutrigenomics can enable to achieve nutritional security in most of the crops. Utilization of omics technologies to identify the functional components in less explored crops like fruits, vegetables, spices and medicinal plants is essential to improve the functional components. There is a need to integrate multi-omics technologies in functional food research to elucidate and enhance the nutrition components in plants. Nutrigenomics can provide insights into the interaction of functional foods in human health and would provide allusion towards scientifically personalized diet.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SNN along with BA, SSM, and RP wrote the first draft. BSS, PB, KH, CK, and NP revised the manuscript. All the authors have read and approved the manuscript.

FUNDING

USDA-NIFA-Hatch funds provided to New Mexico Agricultural Experiment Station, New Mexico State University. SNN, BA and BSS acknowledge DBT-GoI for funds and fellowship.

REFERENCES

Abbo, S., Molina, C., Jungmann, R., Grusak, M. A., Berkovitch, Z., Reifen, R., et al. (2005). Quantitative Trait Loci Governing Carotenoid Concentration and Weight in Seeds of Chickpea (*Cicer Arietinum* L.). *Theor. Appl. Genet.* 111 (2), 185–195. doi:10.1007/s00122-005-1930-y

Abbott, R. (2014). *Documenting Traditional Medical Knowledge.* Ryan Abbott, Documenting Traditional Medical Knowledge. Geneva, Switzerland: World Intellectual Property Organization.

Adams, G. G., Imran, S., Wang, S., Mohammad, A., Kok, M. S., Gray, D. A., et al. (2014). The Hypoglycemic Effect of Pumpkin Seeds, Trigonelline (TRG), Nicotinic Acid (NA), and D-Chiro-Inositol (DCI) in Controlling Glycemic Levels in Diabetes Mellitus. *Crit. Rev. Food Sci. Nutr.* 54 (10), 1322–1329. doi:10.1080/10408398.2011.635816

Aghajanpour, M., Nazer, M. R., Obeidavi, Z., Akbari, M., Ezati, P., and Kor, N. M. (2017). Functional Foods and Their Role in Cancer Prevention and Health Promotion: a Comprehensive Review. *Am. J. Cancer Res.* 7 (4), 740–769.

Al-Safadi, B., Mir, A., and Arabi, M. (2000). Improvement of Garlic (*Allium Sativum* L.) Resistance to white Rot and Storability Using Gamma Irradiation Induced Mutations. *J. Genet. Breed.* 54 (3), 175–182.

Alissa, E. M., and Ferns, G. A. (2012). Functional Foods and Nutraceuticals in the Primary Prevention of Cardiovascular Diseases. *J. Nutr. Metab.* doi:10.1155/2012/569486

Alkhatib, A., Tsang, C., Tiss, A., Bahorun, T., Arefanian, H., Barake, R., et al. (2017). Functional Foods and Lifestyle Approaches for Diabetes Prevention and Management. *Nutrients* 9 (12), 1310. doi:10.3390/nu9121310

Alomari, D. Z., Eggert, K., von Wirén, N., Alqudah, A. M., Polley, A., Plieske, J., et al. (2018). Identifying Candidate Genes for Enhancing Grain Zn Concentration in Wheat. *Front. Plant Sci.* 9, 1313. doi:10.3389/fpls.2018.01313

Alquezar, B., Rodrigo, M. J., and Zacarías, L. (2008). Regulation of Carotenoid Biosynthesis during Fruit Maturation in the Red-Fleshed orange Mutant Cara Cara. *Phytochemistry* 69 (10), 1997–2007. doi:10.1016/j.phytochem.2008.04.020

Aluru, M., Xu, Y., Guo, R., Wang, Z., Li, S., White, W., et al. (2008). Generation of Transgenic maize with Enhanced Provitamin A Content. *J. Exp. Bot.* 59 (13), 3551–3562. doi:10.1093/jxb/ern212

Anai, T., Koga, M., Tanaka, H., Kinoshita, T., Rahman, S. M., and Takagi, Y. (2003). Improvement of rice (*Oryza Sativa* L.) Seed Oil Quality through Introduction of a Soybean Microsomal omega-3 Fatty Acid Desaturase Gene. *Plant Cel Rep.* 21 (10), 988–992. doi:10.1007/s00299-003-0609-6

Anitha, S., Kane-Potaka, J., Tsusaka, T. W., Botha, R., Rajendran, A., Givens, D. I., et al. (2021). A Systematic Review and Meta-Analysis of the Potential of Millets and Sorghum for Managing and Preventing Diabetes Mellitus. *Front. Nutr.* 386.

Ashihara, H., Sano, H., and Crozier, A. (2008). Caffeine and Related Purine Alkaloids: Biosynthesis, Catabolism, Function and Genetic Engineering. *Phytochemistry* 69 (4), 841–856. doi:10.1016/j.phytochem.2007.10.029

Assa, J. R., Widjanarko, S. B., Kusnadi, J., and Berhimpon, S. (2014). Antioxidant Potential of Flesh, Seed and Mace of Nutmeg (*Myristica Fragrans* Houtt). *Int. J.ChemTech Res.* 6 (4), 2460–2468.

Avula, B., Smillie, T. J., Wang, Y.-H., Zweigenbaum, J., and Khan, I. A. (2015). Authentication of True Cinnamon (*Cinnamon Verum*) Utilising Direct Analysis in Real Time (DART)-QToF-MS. *Food Additives & Contaminants: A* 32 (1), 1–8. doi:10.1080/19440049.2014.981763

Axtell, J., Jd, A., Van Scoyoc, S., and Pj, C. (1979). "Current Status of Protein Quality Improvement in Grain Sorghum," in International Symposium on Seed Protein Improvement in Cereals and Grain Legumes, Neuherberg (Germany, FR), 4 Sep 1978 (Vienna, Austria: International Atomic Energy Agency).

Baboota, R. K., Bishnoi, M., Ambalam, P., Kondepudi, K. K., Sarma, S. M., Boparai, R. K., et al. (2013). Functional Food Ingredients for the Management of Obesity and Associated Co-morbidities - A Review. *J. Funct. Foods* 5 (3), 997–1012. doi:10.1016/j.jff.2013.04.014

Bagchi, D., Swaroop, A., and Bagchi, M. (2015). *Genomics, Proteomics and Metabolomics in Nutraceuticals and Functional Foods*. John Wiley & Sons.

Ballali, S., and Lanciai, F. (2012). Functional Food and Diabetes: a Natural Way in Diabetes Prevention? *Int. J. Food Sci. Nutr.* 63 (Suppl. 1), 51–61. doi:10.3109/09637486.2011.637487

Balyan, H. S., Gupta, P. K., Kumar, S., Dhariwal, R., Jaiswal, V., Tyagi, S., et al. (2013). Genetic Improvement of Grain Protein Content and Other Health-related Constituents of Wheat Grain. *Plant Breed* 132 (5), 446–457.

Bandyopadhyay, T., Jaiswal, V., and Prasad, M. (2017). "Nutrition Potential of Foxtail Millet in Comparison to Other Millets and Major Cereals," in *The Foxtail Millet Genome* (Springer), 123–135. doi:10.1007/978-3-319-65617-5_10

Basu, A., Schell, J., and Scofield, R. H. (2018). Dietary Fruits and Arthritis. *Food Funct.* 9 (1), 70–77. doi:10.1039/c7fo01435j

Ben Ayed, D. J., and Hanana, M. (2021). Artificial Intelligence to Improve the Food and Agriculture Sector. *J. Food Qual.* 2021.

Bennett, D. J., and Kirby, G. W. (1968). Constitution and Biosynthesis of Capsaicin. *J. Chem. Soc. C*, 442–446. doi:10.1039/j39680000442

Bera, S. K., Kamdar, J. H., Kasundra, S. V., Dash, P., Maurya, A. K., Jasani, M. D., et al. (2018). Improving Oil Quality by Altering Levels of Fatty Acids through Marker-Assisted Selection of Ahfad2 Alleles in Peanut (*Arachis hypogaea* L.). *Euphytica* 214, 1–15. doi:10.1007/s10681-018-2241-0

Beyer, P., Al-Babili, S., Ye, X., Lucca, P., Schaub, P., Welsch, R., et al. (2002). Golden Rice: Introducing the β-Carotene Biosynthesis Pathway into Rice Endosperm by Genetic Engineering to Defeat Vitamin A Deficiency. *J. Nutr.* 132 (3), 506S–510S. doi:10.1093/jn/132.3.506s

Bhatnagar, M., Prasad, K., Bhatnagar-Mathur, P., Lakshmi Narasu, M., Waliyar, F., and Sharma, K. K. (2010). An Efficient Method for the Production of Marker-free Transgenic Plants of Peanut (*Arachis hypogaea* L.). *Plant Cel Rep* 29 (5), 495–502. doi:10.1007/s00299-010-0838-4

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., et al. (2012). Retrotransposons Control Fruit-specific, Cold-dependent Accumulation of Anthocyanins in Blood Oranges. *Plant Cell* 24 (3), 1242–1255. doi:10.1105/tpc.111.095232

Butt, M. S., and Sultan, M. T. (2011). Ginger and its Health Claims: Molecular Aspects. *Crit. Rev. Food Sci. Nutr.* 51 (5), 383–393. doi:10.1080/10408391003624848

Cai, S., Yu, G., Chen, X., Huang, Y., Jiang, X., Zhang, G., et al. (2013). Grain Protein Content Variation and its Association Analysis in Barley. *BMC Plant Biol.* 13 (1), 35–11. doi:10.1186/1471-2229-13-35

Cao, H., Liu, R., Zhang, J., Liu, Z., Fan, S., Yang, G., et al. (2021). Improving Sulforaphane Content in Transgenic Broccoli Plants by Overexpressing MAM1, FMO GS–OX2, and Myrosinase. *PCTOC* 146, 1–11. doi:10.1007/s11240-021-02079-2

Capozzi, F., and Bordoni, A. (2013). Foodomics: a New Comprehensive Approach to Food and Nutrition. *Genes Nutr.* 8 (1), 1–4. doi:10.1007/s12263-012-0310-x

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, Precise Modification of the Tomato Genome. *Genome Biol.* 16 (1), 1–15.

Čertík, M., Klempová, T., Guothová, L., Mihálik, D., and Kraic, J. (2013). Biotechnology for the Functional Improvement of Cereal-based Materials Enriched with PUFA and Pigments. *Eur. J. Lip. Sci. Tech.* 115 (11), 1247–1256.

Chakraborty, P. (2018). Herbal Genomics as Tools for Dissecting New Metabolic Pathways of Unexplored Medicinal Plants and Drug Discovery. *Biochimie Open* 6, 9–16. doi:10.1016/j.biopen.2017.12.003

Chakraborty, S., Chakraborty, N., Agrawal, L., Ghosh, S., Narula, K., Shekhar, S., et al. (2010). Next-generation Protein-Rich Potato Expressing the Seed Protein geneAmA1is a Result of Proteome Rebalancing in Transgenic Tuber. *Proc. Natl. Acad. Sci. USA* 107 (41), 17533–17538. doi:10.1073/pnas.1006265107

Chen, X., Long, H., Gao, P., Deng, G., Pan, Z., Liang, J., et al. (2014). Transcriptome Assembly and Analysis of Tibetan Hulless Barley (*Hordeum Vulgare* L. Var. Nudum) Developing Grains, with Emphasis on Quality Properties. *PLoS ONE* 9 (5), e98144. doi:10.1371/journal.pone.0098144

Chen, Z., Young, T. E., Ling, J., Chang, S.-C., and Gallie, D. R. (2003). Increasing Vitamin C Content of Plants through Enhanced Ascorbate Recycling. *Proc. Natl. Acad. Sci.* 100 (6), 3525–3530. doi:10.1073/pnas.0635176100

Cheng, L., Zheng, W., Li, M., Huang, J., Bao, S., Xu, Q., et al. (2020). Citrus Fruits Are Rich in Flavonoids for Immunoregulation and Potential Targeting ACE2. *Preprints*, 2020020313.

Chiu, L.-W., Zhou, X., Burke, S., Wu, X., Prior, R. L., and Li, L. (2010). The Purple Cauliflower Arises from Activation of a MYB Transcription Factor. *Plant Physiol.* 154 (3), 1470–1480. doi:10.1104/pp.110.164160

Chongtham, N., Bisht, M. S., and Haorongbam, S. (2011). Nutritional Properties of Bamboo Shoots: Potential and Prospects for Utilization as a Health Food. *Compr. Rev. Food Sci. Food Saf.* 10 (3), 153–168. doi:10.1111/j.1541-4337.2011.00147.x

Ciura, J., Szeliga, M., Grzesik, M., and Tyrka, M. (2017). Next-generation Sequencing of Representational Difference Analysis Products for Identification of Genes Involved in Diosgenin Biosynthesis in Fenugreek (*Trigonella Foenum-graecum*). *Planta* 245 (5), 977–991. doi:10.1007/s00425-017-2657-0

Colasuonno, P., Lozito, M. L., Marcotuli, I., Nigro, D., Giancaspro, A., Mangini, G., et al. (2017). The Carotenoid Biosynthetic and Catabolic Genes in Wheat and Their Association with Yellow Pigments. *BMC Genomics* 18 (1), 122–218. doi:10.1186/s12864-016-3395-6

Cruet-Burgos, C., Cox, S., Ioerger, B. P., Perumal, R., Hu, Z., Herald, T. J., et al. (2020). Advancing Provitamin A Biofortification in Sorghum: Genome-wide Association Studies of Grain Carotenoids in Global Germplasm. *Plant Genome* 13, e20013. doi:10.1002/tpg2.20013

de Alencar Figueiredo, L. F., Sine, B., Chantereau, J., Mestres, C., Fliedel, G., Rami, J.-F., et al. (2010). Variability of Grain Quality in Sorghum: Association with Polymorphism in Sh2, Bt2, SssI, Ae1, Wx and O2. *Theor. Appl. Genet.* 121 (6), 1171–1185. doi:10.1007/s00122-010-1380-z

Deery, D. M., and Jones, H. G. (2021). Field Phenomics: Will it Enable Crop Improvement? *Plant Phenomics* 2021, 1–16.

Derosa, G., Maffioli, P., and Sahebkar, A. (2016). Piperine and its Role in Chronic Diseases. *Adv. Exp. Med. Biol.* 928, 173–184. doi:10.1007/978-3-319-41334-1_8

Descalsota, G. I. L., Swamy, B. P. M., Zaw, H., Inabangan-Asilo, M. A., Amparado, A., Mauleon, R., et al. (2018). Genome-wide Association Mapping in a rice MAGIC Plus Population Detects QTLs and Genes Useful for Biofortification. *Front. Plant Sci.* 9, 1347. doi:10.3389/fpls.2018.01347

Descalsota-Empleo, G. I., Noraziyah, A. A. S., Navea, I. P., Chung, C., Dwiyanti, M. S., Labios, R. J. D., et al. (2019). Genetic Dissection of Grain Nutritional Traits and Leaf Blight Resistance in rice. *Genes* 10 (1), 30. doi:10.3390/genes10010030

Devi, I., Ansari, N., and Kumar, V. (2008). "Biofortification of Safflower Oil with Gamma Linolenic Acid through Transgenic Approach Using delta-6-desaturase Gene from Borago Officinalis," in Safflower: unexploited potential and world adaptability. 7th International Safflower Conference, Wagga Wagga, New South Wales, Australia, 3-6 November, 2008 (Bendigo, VIC: Agri-MC Marketing and Communication), 1–5.

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

Distelfeld, A., Uauy, C., Fahima, T., and Dubcovsky, J. (2006). Physical Map of the Wheat High-grain Protein Content geneGpc-B1and Development of a High-throughput Molecular Marker. *New Phytol.* 169 (4), 753–763. doi:10.1111/j.1469-8137.2005.01627.x

Dubock, A. (2019). Golden Rice: To Combat Vitamin A Deficiency for Public Health. *Vitamin A* 1–21. doi:10.5772/intechopen.84445

Ducreux, L. J., Morris, W. L., Hedley, P. E., Shepherd, T., Davies, H. V., Millam, S., et al. (2005). Metabolic Engineering of High Carotenoid Potato Tubers Containing Enhanced Levels of Beta-Carotene and Lutein. *J. Exp. Bot.* 56 (409), 81–89. doi:10.1093/jxb/eri016

Edmunds, S. J., Roy, N. C., Davy, M., Cooney, J. M., Barnett, M. P., Zhu, S., et al. (2012). Effects of Kiwifruit Extracts on Colonic Gene and Protein Expression Levels in IL-10 Gene-Deficient Mice. *Br. J. Nutr.* 108 (1), 113–129. doi:10.1017/S0007114511005241

El Khoury, D., Cuda, C., Luhovyy, B., and Anderson, G. (2012). Beta Glucan: Health Benefits in Obesity and Metabolic Syndrome. *J. Nutr. Metab.* doi:10.1155/2012/851362

El Nagar, M. (2018). Genetic Engineering to Improve β-carotene Content in Pepper. *Ann. Agric. Sci. Moshtohor* 56, 119–126. doi:10.21608/assjm.2018.65121

Fang, H. Y., Chen, S. B., Guo, D. J., Pan, S. Y., and Yu, Z. L. (2011). Proteomic Identification of Differentially Expressed Proteins in Curcumin-Treated MCF-7 Cells. *Phytomedicine* 18 (8-9), 697–703. doi:10.1016/j.phymed.2010.11.012

FAO (2019). The State of Food and Agriculture. Availableat: http://www.fao.org/state-of-food-agriculture/2019/en/(Accessed September, 2020).

Fernandez, M. G. S., Hamblin, M. T., Li, L., Rooney, W. L., Tuinstra, M. R., and Kresovich, S. (2008). Quantitative Trait Loci Analysis of Endosperm Color and Carotenoid Content in Sorghum Grain. *Crop Sci.* 48 (5), 1732–1743. doi:10.2135/cropsci2007.12.0684

Fokunang, C. N., Ndikum, V., Tabi, O. Y., Jiofack, R. B., Ngameni, B., Guedje, N. M., et al. (2011). Traditional Medicine: Past, Present and Future Research and Development Prospects and Integration in the National Health System of Cameroon. *Afr. J. Tradit Complement. Altern. Med.* 8 (3), 284–295. doi:10.4314/ajtcam.v8i3.65276

Galpaz, N., Wang, Q., Menda, N., Zamir, D., and Hirschberg, J. (2008). Abscisic Acid Deficiency in the Tomato Mutant High-Pigment 3 Leading to Increased Plastid Number and Higher Fruit Lycopene Content. *Plant J.* 53 (5), 717–730. doi:10.1111/j.1365-313x.2007.03362.x

Gantait, S., Debnath, S., and Nasim Ali, M. (2014). Genomic Profile of the Plants with Pharmaceutical Value. *3 Biotech.* 4 (6), 563–578. doi:10.1007/s13205-014-0218-9

Gazal, A., Dar, Z. A., Zaffar, G., Lone, A. A., Abidi, I., Shabir, A., et al. (2014). Trends in Breeding Oat for Nutritional Grain Quality - an Overview. *Jans* 6 (2), 904–912. doi:10.31018/jans.v6i2.552

Ghosh, B., Chandra, I., and Chatterjee, S. (2015). Fenugreek (*Trigonella Foenum-graecum* L.) and its Necessity. *Fire J. Eng. Technol.* 1 (1), 66–67.

Gibbon, B. C., and Larkins, B. A. (2005). Molecular Genetic Approaches to Developing Quality Protein maize. *Trends Genet.* 21 (4), 227–233. doi:10.1016/j.tig.2005.02.009

Gibon, Y., Usadel, B., Blaesing, O. E., Kamlage, B., Hoehne, M., Trethewey, R., et al. (2006). Integration of Metabolite with Transcript and Enzyme Activity Profiling during Diurnal Cycles in Arabidopsis Rosettes. *Genome Biol.* 7 (8), R76–R23. doi:10.1186/gb-2006-7-8-R76

Gil-Humanes, J., Pistón, F., Altamirano-Fortoul, R., Real, A., Comino, I., Sousa, C., et al. (2014). Reduced-gliadin Wheat Bread: an Alternative to the Gluten-free Diet for Consumers Suffering Gluten-Related Pathologies. *PLoS ONE* 9 (3), e90898. doi:10.1371/journal.pone.0090898

González, M., Salazar, E., Castillo, J., Morales, P., Mura-Jornet, I., Maldonado, J., et al. (2016). Genetic Structure Based on EST-SSR: a Putative Tool for Fruit Color Selection in Japanese Plum (Prunus Salicina L.) Breeding Programs. *Mol. Breed.* 36 (6), 68. doi:10.1007/s11032-016-0491-x

González-Molina, E., Domínguez-Perles, R., Moreno, D. A., and García-Viguera, C. (2010). Natural Bioactive Compounds of Citrus limon for Food and Health. *J. Pharm. Biomed. Anal.* 51 (2), 327–345. doi:10.1016/j.jpba.2009.07.027

Gonzali, S., Mazzucato, A., and Perata, P. (2009). Purple as a Tomato: towards High Anthocyanin Tomatoes. *Trends Plant Sci.* 14 (5), 237–241. doi:10.1016/j.tplants.2009.02.001

Gordeeva, E., Badaeva, E., Yudina, R., Shchukina, L., Shoeva, O., and Khlestkina, E. (2019). Marker-Assisted Development of a Blue-Grained Substitution Line Carrying the Thinopyrum Ponticum Chromosome 4Th(4D) in the Spring Bread Wheat Saratovskaya 29 Background. *Agronomy* 9, 723. doi:10.3390/agronomy9110723

Gordeeva, E., Shamanin, V., Shoeva, O., Kukoeva, T., Morgounov, A., and Khlestkina, E. (2020). The Strategy for Marker-Assisted Breeding of Anthocyanin-Rich Spring Bread Wheat (*Triticum aestivum* L.) Cultivars in Western Siberia. *Agronomy* 10, 1603. doi:10.3390/agronomy10101603

Gorgani, L., Mohammadi, M., Najafpour, G. D., and Nikzad, M. (2017). Piperine-The Bioactive Compound of Black Pepper: From Isolation to Medicinal Formulations. *Compr. Rev. Food Sci. Food Saf.* 16 (1), 124–140. doi:10.1111/1541-4337.12246

Govindaraj, M., Rai, K. N., Cherian, B., Pfeiffer, W. H., Kanatti, A., and Shivade, H. (2019). Breeding Biofortified Pearl Millet Varieties and Hybrids to Enhance Millet Markets for Human Nutrition. *Agriculture* 9 (5), 106.

Granger, M., and Eck, P. (2018). Dietary Vitamin C in Human Health. *Adv. Food Nutr. Res.* 83281–310.

Guizani, N. (2013). Date palm Fruits as Functional Foods. *Int. J. Nutr. Pharmacol. Neurol. Dis.* 3 (3), 161. doi:10.4103/2231-0738.114830

Gul, K., Tak, A., Singh, A. K., Singh, P., Yousuf, B., and Wani, A. A. (2015). Chemistry, Encapsulation, and Health Benefits of β-carotene - A Review. *Cogent Food Agric.* 1, 1018696. doi:10.1080/23311932.2015.1018696

Gürbüz Çolak, N., Eken, N. T., Ülger, M., Frary, A., and Doğanlar, S. (2020). Mapping of Quantitative Trait Loci for Antioxidant Molecules in Tomato Fruit: Carotenoids, Vitamins C and E, Glutathione and Phenolic Acids. *Plant Sci.* 292, 110393. doi:10.1016/j.plantsci.2019.110393

Hamid, S. B., and Hamid, A. F. A. (2019). Roles of Nutraceuticals and Functional Food in Prevention of Cardiovascular Disease. *Nutraceutical and Functional Foods in Disease Prevention*. Hershey, PA: IGI Global, 126–165. doi:10.4018/978-1-5225-3267-5.ch005

Han, Z., Zhang, J., Cai, S., Chen, X., Quan, X., and Zhang, G. (2018). Association Mapping for Total Polyphenol Content, Total Flavonoid Content and Antioxidant Activity in Barley. *BMC Genomics* 19 (1), 81–10. doi:10.1186/s12864-018-4483-6

Hannan, J. M. A., Ali, L., Rokeya, B., Khaleque, J., Akhter, M., Flatt, P. R., et al. (2007). Soluble Dietary Fibre Fraction ofTrigonella Foenum-Graecum(fenugreek) Seed Improves Glucose Homeostasis in Animal Models of Type 1 and Type 2 Diabetes by Delaying Carbohydrate Digestion and Absorption, and Enhancing Insulin Action. *Br. J. Nutr.* 97 (3), 514–521. doi:10.1017/s0007114507657869

Hao, D. C., and Xiao, P. G. (2015). Genomics and Evolution in Traditional Medicinal Plants: Road to a Healthier Life. *Evol. Bioinform Online* 11, 197–212. doi:10.4137/EBO.S31326

Hasler, C. M. (1998). Functional Foods: Their Role in Disease Prevention and Health Promotion. *Food Technology-Champaign then Chicago* 52, 63–147.

Hellwege, E. M., Czapla, S., Jahnke, A., Willmitzer, L., and Heyer, A. G. (2000). Transgenic Potato (*Solanum tuberosum*) Tubers Synthesize the Full Spectrum of Inulin Molecules Naturally Occurring in globe Artichoke (*Cynara Scolymus*) Roots. *Proc. Natl. Acad. Sci.* 97 (15), 8699–8704. doi:10.1073/pnas.150043797

Hill, J. O., and Peters, J. C. (2002). Biomarkers and Functional Foods for Obesity and Diabetes. *Br. J. Nutr.* 88 (S2), S213–S218. doi:10.1079/bjn2002685

Hindu, V., Palacios-Rojas, N., Babu, R., Suwarno, W. B., Rashid, Z., Usha, R., et al. (2018). Identification and Validation of Genomic Regions Influencing Kernel Zinc and Iron in maize. *Theor. Appl. Genet.* 131, 1443–1457. doi:10.1007/s00122-018-3089-3

Hu, L., Hao, C., Fan, R., Wu, B., Tan, L., and Wu, H. (2015). De Novo assembly and Characterization of Fruit Transcriptome in Black Pepper (*Piper Nigrum*). *PLoS ONE* 10 (6), e0129822. doi:10.1371/journal.pone.0129822

Huang, S., Frizzi, A., Florida, C. A., Kruger, D. E., and Luethy, M. H. (2006). High Lysine and High Tryptophan Transgenic Maize Resulting from the Reduction of Both 19- and 22-kD α-zeins. *Plant Mol. Biol.* 61 (3), 525–535. doi:10.1007/s11103-006-0027-6

Hussain, B., Lucas, S. J., Ozturk, L., and Budak, H. (2017). Mapping QTLs Conferring Salt Tolerance and Micronutrient Concentrations at Seedling Stagein Wheat. *Sci. Rep.* 7 (1), 15662–15714. doi:10.1038/s41598-017-15726-6

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., et al. (2016). A High-Quality Carrot Genome Assembly Provides New Insights into

Carotenoid Accumulation and Asterid Genome Evolution. *Nat. Genet.* 48 (6), 657–666. doi:10.1038/ng.3565

Jacobo-Valenzuela, N., Maróstica-Junior, M. R., Zazueta-Morales, J. d. J., and Gallegos-Infante, J. A. (2011). Physicochemical, Technological Properties, and Health-Benefits of Cucurbita Moschata Duchense vs. Cehualca. *Food Res. Int.* 44 (9), 2587–2593. doi:10.1016/j.foodres.2011.04.039

Jia, D., Fan, L., Shen, J., Qin, S., Li, F., and Yuan, Y. (2019). Genetic Transformation of the Astaxanthin Biosynthetic Genes *Bkt* and *crtR-B* into Apple Tree to Increase Photooxidation Resistance. *Scientia Horticulturae* 243, 428–433. doi:10.1016/j.scienta.2018.08.058

Jin, L., Xiao, P., Lu, Y., Shao, Y., Shen, Y., and Bao, J. (2009). Quantitative Trait Loci for Brown rice Color, Phenolics, Flavonoid Contents, and Antioxidant Capacity in rice Grain. *Cereal Chem. J.* 86 (6), 609–615. doi:10.1094/cchem-86-6-0609

Jin, M., Zhang, X., Zhao, M., Deng, M., Du, Y., Zhou, Y., et al. (2017). Integrated Genomics-Based Mapping Reveals the Genetics Underlying maize Flavonoid Biosynthesis. *BMC Plant Biol.* 17 (1), 17. doi:10.1186/s12870-017-0972-z

Jin, S., Sun, X., Wu, F., Su, Y., Li, Y., Song, S., et al. (2021). Lidar Sheds New Light on Plant Phenomics for Plant Breeding and Management: Recent Advances and Future Prospects. *ISPRS J. Photogrammetry Remote Sensing* 171, 202–223. doi:10.1016/j.isprsjprs.2020.11.006

Jourdan, M., Gagné, S., Dubois-Laurent, C., Maghraoui, M., Huet, S., Suel, A., et al. (2015). Carotenoid Content and Root Color of Cultivated Carrot: a Candidate-Gene Association Study Using an Original Broad Unstructured Population. *PLoS One* 10 (1), e0116674. doi:10.1371/journal.pone.0116674

Joy, P. P., Anjana, R., Rashida-Rajuva, A., and Ratheesh, A. (2018). "Fruits as a Functional Food," in *State-of-the-art Technologies in Food Science: Human Health, Emerging Issues and Specialty Topics* (Boca Raton, FL: CRC Press), 111–168.

Karkute, S. G., Singh, A. K., Gupta, O. P., Singh, P. M., and Singh, B. (2017). CRISPR/Cas9 Mediated Genome Engineering for Improvement of Horticultural Crops. *Front. Plant Sci.* 8, 1635. doi:10.3389/fpls.2017.01635

Karmakar, A., Bhattacharya, S., Sengupta, S., Ali, N., Sarkar, S. N., Datta, K., et al. (2020). RNAi-Mediated Silencing of ITPK Gene Reduces Phytic Acid Content, Alters Transcripts of Phytic Acid Biosynthetic Genes, and Modulates mineral Distribution in rice Seeds. *Rice Sci.* 27, 315–328. doi:10.1016/j.rsci.2020.05.007

Kartikey, K., Singh, G., Sah, D., Singh, R. B., Singh, A. K., Takahashi, T., et al. (2019). "Functional Food Security for Osteoporosis, Carcinogenesis, Atherosclerosis and Brain Degeneration," in *The Role of Functional Food Security in Global Health* (Elsevier), 639–651. doi:10.1016/b978-0-12-813148-0.00037-2

Kato, H., Takahashi, S., and Saito, K. (2011). Omics and Integrated Omics for the Promotion of Food and Nutrition Science. *J. Traditional Complement. Med.* 1 (1), 25–30. doi:10.1016/s2225-4110(16)30053-0

Kawatra, P., and Rajagopalan, R. (2015). Cinnamon: Mystic powers of a Minute Ingredient. *Pharmacognosy Res.* 7 (Suppl. 1), S1–S6. doi:10.4103/0974-8490.157990

Keilwagen, J., Lehnert, H., Berner, T., Budahn, H., Nothnagel, T., Ulrich, D., et al. (2017). The Terpene Synthase Gene Family of Carrot (*Daucus Carota* L.): Identification of QTLs and Candidate Genes Associated with Terpenoid Volatile Compounds. *Front. Plant Sci.* 8, 1930. doi:10.3389/fpls.2017.01930

Khalil, A. A., Rahman, U. u., Khan, M. R., Sahar, A., Mehmood, T., and Khan, M. (2017). Essential Oil Eugenol: Sources, Extraction Techniques and Nutraceutical Perspectives. *RSC Adv.* 7, 32669–32681. doi:10.1039/c7ra04803c

Khan, M. A., Nagoo, S., Naseer, S., Nehvi, F., and Zargar, S. M. (2011). Induced Mutation as a Tool for Improving Corm Multiplication in Saffron (*Crocus Sativus* L.). *J. Phytol.* 3 (7).

Khazaei, H., Podder, R., Caron, C. T., Kundu, S. S., Diapari, M., Vandenberg, A., et al. (2017). Marker-Trait Association Analysis of Iron and Zinc Concentration in Lentil (Lens Culinaris Medik.) Seeds. *Plant Genome* 10 (2). doi:10.3835/plantgenome2017.02.0007

Kidd, P. (2011). Astaxanthin, Cell Membrane Nutrient with Diverse Clinical Benefits and Anti-aging Potential. *Altern. Med. Rev.* 16, 355–364.

Kim, I.-S., Yang, M., Lee, O.-H., and Kang, S.-N. (2011). The Antioxidant Activity and the Bioactive Compound Content of *Stevia rebaudiana* Water Extracts. *LWT - Food Sci. Technology* 44 (5), 1328–1332. doi:10.1016/j.lwt.2010.12.003

Kim, S.-E., Lee, C.-J., Ji, C. Y., Kim, H. S., Park, S.-U., Lim, Y.-H., et al. (2019). Transgenic Sweetpotato Plants Overexpressing Tocopherol Cyclase Display Enhanced α-tocopherol Content and Abiotic Stress Tolerance. *Plant Physiol. Biochem.* 144, 436–444. doi:10.1016/j.plaphy.2019.09.046

Kim, S. J., Li, M., Jeong, C. W., Bae, H. B., Kwak, S. H., Lee, S. H., et al. (2014). Epigallocatechin-3-gallate, a green tea Catechin, Protects the Heart against Regional Ischemia-Reperfusion Injuries through Activation of RISK Survival Pathways in Rats. *Arch. Pharm. Res.* 37 (8), 1079–1085. doi:10.1007/s12272-013-0309-x

Kimura, E., Abe, T., Murata, K., Kimura, T., Otoki, Y., Yoshida, T., et al. (2018). Identification of OsGGR2, a Second Geranylgeranyl Reductase Involved in α-tocopherol Synthesis in rice. *Sci. Rep.* 8, 1870–1878. doi:10.1038/s41598-018-19527-3

Knox, R. E., Pozniak, C. J., Clarke, F. R., Clarke, J. M., Houshmand, S., and Singh, A. K. (2009). Chromosomal Location of the Cadmium Uptake Gene (Cdu1) in Durum Wheat. *Genome* 52 (9), 741–747. doi:10.1139/g09-042

Kole, C., Muthamilarasan, M., Henry, R., Edwards, D., Sharma, R., Abberton, M., et al. (2015). Application of Genomics-Assisted Breeding for Generation of Climate Resilient Crops: Progress and Prospects. *Front. Plant Sci.* 6, 563. doi:10.3389/fpls.2015.00563

Kotla, A., Phuke, R., Hariprasanna, K., Mehtre, S. P., Rathore, A., Gorthy, S., et al. (2019). Identification of QTLs and Candidate Genes for High Grain Fe and Zn Concentration in Sorghum [Sorghum Bicolor (L.)Moench]. *J. Cereal Sci.* 90, 102850. doi:10.1016/j.jcs.2019.102850

Kumar, A., Kujur, A., Singh, P. P., and Prakash, B. (2019). Nanoencapsulated Plant-Based Bioactive Formulation against Food-Borne Molds and Aflatoxin B1 Contamination: Preparation, Characterization and Stability Evaluation in the Food System. *Food Chem.* 287, 139–150. doi:10.1016/j.foodchem.2019.02.045

Kumar, A., Kumar, S., Bains, S., Vaidya, V., Singh, B., Kaur, R., et al. (2016). De Novo transcriptome Analysis Revealed Genes Involved in Flavonoid and Vitamin C Biosynthesis in *Phyllanthus Emblica* (L.). *Front. Plant Sci.* 7, 1610. doi:10.3389/fpls.2016.01610

Kumar, A., Mirza, N., Charan, T., Sharma, N., and Gaur, V. S. (2014). Isolation, Characterization and Immunolocalization of a Seed Dominant CaM from finger Millet (*Eleusine Coracana* L. Gartn.) for Studying its Functional Role in Differential Accumulation of Calcium in Developing Grains. *Appl. Biochem. Biotechnol.* 172 (6), 2955–2973. doi:10.1007/s12010-013-0714-0

Kumar, S., Hash, C., Nepolean, T., Mahendrakar, M., Satyavathi, C., Singh, G., et al. (2018). Mapping Grain Iron and Zinc Content Quantitative Trait Loci in an Iniadi-Derived Immortal Population of Pearl Millet. *Genes* 9 (5), 248. doi:10.3390/genes9050248

Kyriakoudi, A., Ordoudi, S., Roldán-Medina, M., and Tsimidou, M. (2015). Saffron, a Functional Spice. *Austin. J. Nutri. Food Sci.* 3 (1), 1059.

Lapitan, N. L. V., Hess, A., Cooper, B., Botha, A.-M., Badillo, D., Iyer, H., et al. (2009). Differentially Expressed Genes during Malting and Correlation with Malting Quality Phenotypes in Barley (Hordeum Vulgare L.). *Theor. Appl. Genet.* 118 (5), 937–952. doi:10.1007/s00122-008-0951-8

Laribi, B., Kouki, K., M'Hamdi, M., and Bettaieb, T. (2015). Coriander (*Coriandrum Sativum* L.) and its Bioactive Constituents. *Fitoterapia* 103, 9–26. doi:10.1016/j.fitote.2015.03.012

Lau, F. C., Shukitt-Hale, B., and Joseph, J. A. (2005). The Beneficial Effects of Fruit Polyphenols on Brain Aging. *Neurobiol. Aging* 26 (1), 128–132. doi:10.1016/j.neurobiolaging.2005.08.007

Lee, K.-G., and Shibamoto, T. (2001). Antioxidant property of aroma extract isolated from clove buds [Syzygium aromaticum (L.) Merr. et Perry]. *Food Chem.* 74 (4), 443–448. doi:10.1016/s0308-8146(01)00161-3

Li, H., Liu, H., Han, Y., Wu, X., Teng, W., Liu, G., et al. (2010). Identification of QTL Underlying Vitamin E Contents in Soybean Seed Among Multiple Environments. *Theor. Appl. Genet.* 120 (7), 1405–1413. doi:10.1007/s00122-010-1264-2

Li, M.-W., Muñoz, N. B., Wong, C.-F., Wong, F.-L., Wong, K.-S., Wong, J. W.-H., et al. (2016). QTLs Regulating the Contents of Antioxidants, Phenolics, and Flavonoids in Soybean Seeds Share a Common Genomic Region. *Front. Plant Sci.* 7, 854. doi:10.3389/fpls.2016.00854

Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., et al. (2018). Lycopene Is Enriched in Tomato Fruit by CRISPR/Cas9-mediated Multiplex Genome Editing. *Front. Plant Sci.* 9, 559. doi:10.3389/fpls.2018.00559

Liang, Q., Wang, K., Liu, X., Riaz, B., Jiang, L., Wan, X., et al. (2019). Improved Folate Accumulation in Genetically Modified maize and Wheat. *J. Exp. Bot.* 70 (5), 1539–1551. doi:10.1093/jxb/ery453

Liang, Z., Zhang, K., Chen, K., and Gao, C. (2014). Targeted Mutagenesis in *Zea mays* Using TALENs and the CRISPR/Cas System. *J. Genet. Genomics* 41 (2), 63–68. doi:10.1016/j.jgg.2013.12.001

Liao, P., Chen, X., Wang, M., Bach, T. J., and Chye, M.-L. (2018). Improved Fruit α-tocopherol, Carotenoid, Squalene and Phytosterol Contents through Manipulation of Brassica Juncea 3-HYDROXY-3-METHYLGLUTARYL-COA SYNTHASE1 in Transgenic Tomato. *Plant Biotechnol. J.* 16, 784–796. doi:10.1111/pbi.12828

Lipkie, T. E., De Moura, F. F., Zhao, Z.-Y., Albertsen, M. C., Che, P., Glassman, K., et al. (2013). Bioaccessibility of Carotenoids from Transgenic Provitamin A Biofortified Sorghum. *J. Agric. Food Chem.* 61 (24), 5764–5771. doi:10.1021/jf305361s

Liu, S., Li, W., Wu, Y., Chen, C., and Lei, J. (2013). De Novo transcriptome Assembly in Chili Pepper (*Capsicum Frutescens*) to Identify Genes Involved in the Biosynthesis of Capsaicinoids. *PLoS ONE* 8 (1), e48156. doi:10.1371/journal.pone.0048156

Lobo, V., Patil, A., Phatak, A., and Chandra, N. (2010). Free Radicals, Antioxidants and Functional Foods: Impact on Human Health. *Phcog Rev.* 4 (8), 118. doi:10.4103/0973-7847.70902

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics Technologies. *Plos Comput. Biol.* 13 (5), e1005457. doi:10.1371/journal.pcbi.1005457

Lucca, P., Hurrell, R., and Potrykus, I. (2002). Fighting Iron Deficiency Anemia with Iron-Rich rice. *J. Am. Coll. Nutr.* 21 (Suppl. 3), 184S–190S. doi:10.1080/07315724.2002.10719264

Luo, H., Guo, J., Yu, B., Chen, W., Zhang, H., Zhou, X., et al. (2021). Construction of ddRADseq-Based High-Density Genetic Map and Identification of Quantitative Trait Loci for Trans-resveratrol Content in Peanut Seeds. *Front. Plant Sci.* 12, 438. doi:10.3389/fpls.2021.644402

Ma, C., Ma, B., He, J., Hao, Q., Lu, X., and Wang, L. (2011). Regulation of Carotenoid Content in Tomato by Silencing of Lycopene β/ε-Cyclase Genes. *Plant Mol. Biol. Rep.* 29 (1), 117–124. doi:10.1007/s11105-010-0211-3

Mamati, G. E., Liang, Y., and Lu, J. (2006). Expression of Basic Genes Involved in tea Polyphenol Synthesis in Relation to Accumulation of Catechins and Total tea Polyphenols. *J. Sci. Food Agric.* 86 (3), 459–464. doi:10.1002/jsfa.2368

Martín-Lagos, R. A., Serrano, M. F. O., and Lopez, M. D. R. (1995). Determination of Organic sulphur Compounds in Garlic Extracts by Gas Chromatography and Mass Spectrometry. *Food Chem.* 53 (1), 91–93. doi:10.1016/0308-8146(95)95792-5

Masibo, M., and He, Q. (2008). Major Mango Polyphenols and Their Potential Significance to Human Health. *Compr. Rev. Food Sci. Food Saf.* 7 (4), 309–319. doi:10.1111/j.1541-4337.2008.00047.x

Mattivi, F., Guzzon, R., Vrhovsek, U., Stefanini, M., and Velasco, R. (2006). Metabolite Profiling of Grape: Flavonols and Anthocyanins. *J. Agric. Food Chem.* 54 (20), 7692–7702. doi:10.1021/jf061538c

Mazza, G., and Francis, F. J. (1995). Anthocyanins in Grapes and Grape Products. *Crit. Rev. Food Sci. Nutr.* 35 (4), 341–371. doi:10.1080/10408399509527704

McGuire, A. L., Gabriel, S., Tishkoff, S. A., Wonkam, A., Chakravarti, A., Furlong, E. E. M., et al. (2020). The Road Ahead in Genetics and Genomics. *Nat. Rev. Genet.* 21 (10), 581–596. doi:10.1038/s41576-020-0272-6

Melnyk, J. P., Wang, S., and Marcone, M. F. (2010). Chemical and Biological Properties of the World's Most Expensive Spice: Saffron. *Int. Food Res.* 43 (8), 1981–1989.

Mertz, E. T., Bates, L. S., and Nelson, O. E. (1964). Mutant Gene that Changes Protein Composition and Increases Lysine Content of maize Endosperm. *Science* 145 (3629), 279–280. doi:10.1126/science.145.3629.279

Minamikawa, M. F., Nonaka, K., Kaminuma, E., Kajiya-Kanegae, H., Onogi, A., Goto, S., et al. (2017). Genome-wide Association Study and Genomic Prediction in Citrus: Potential of Genomics-Assisted Breeding for Fruit Quality Traits. *Sci. Rep.* 7 (1), 4721–4813. doi:10.1038/s41598-017-05100-x

Mirza, N., Taj, G., Arora, S., and Kumar, A. (2014). Transcriptional Expression Analysis of Genes Involved in Regulation of Calcium Translocation and Storage in finger Millet (*Eleusine Coracana* L. Gartn.). *Gene* 550 (2), 171–179. doi:10.1016/j.gene.2014.08.005

Mishiba, K.-I., Nishida, K., Inoue, N., Fujiwara, T., Teranishi, S., Iwata, Y., et al. (2020). Genetic Engineering of Eggplant Accumulating β-carotene in Fruit. *Plant Cell Rep.* 39, 1029–1039. doi:10.1007/s00299-020-02546-8

Mohan, L., Sukriti, D., and Bhattacharyya, P. (2016). Development of a High Yielding Variety, Jor Lab L-8 of Lemongrass (*Cymbopogon Flexuosus* L.). *Ann. Agri Bio Res* 21 (1), 22–23.

Mohanty, S., and Singhal, K. (2018). "Functional Foods as Personalised Nutrition: Definitions and Genomic Insights," in *Functional Food and Human Health* (Springer), 513–535. doi:10.1007/978-981-13-1123-9_22

Mondal, S., Phadke, R. R., and Badigannavar, A. M. (2015). Genetic Variability for Total Phenolics, Flavonoids and Antioxidant Activity of Testaless Seeds of a Peanut Recombinant Inbred Line Population and Identification of Their Controlling QTLs. *Euphytica* 204 (2), 311–321. doi:10.1007/s10681-014-1324-9

Moraes, N. V., Lermen, F. H., and Echeveste, M. E. S. (2021). A Systematic Literature Review on Food Waste/loss Prevention and Minimization Methods. *J. Environ. Manage.* 286, 112268. doi:10.1016/j.jenvman.2021.112268

Mugode, L., Ha, B., Kaunda, A., Sikombe, T., Phiri, S., Mutale, R., et al. (2014). Carotenoid Retention of Biofortified Provitamin A maize (*Zea mays* L.) after Zambian Traditional Methods of Milling, Cooking and Storage. *J. Agric. Food Chem.* 62 (27), 6317–6325. doi:10.1021/jf501233f

Munck, L. (1972). Improvement of Nutritional Value in Cereals. *Hereditas* 72, 1–128.

Muthusamy, V., Hossain, F., Thirunavukkarasu, N., Choudhary, M., Saha, S., Bhat, J. S., et al. (2014). Development of β-Carotene Rich Maize Hybrids through Marker-Assisted Introgression of β-carotene Hydroxylase Allele. *PLoS One* 9 (12), e113583. doi:10.1371/journal.pone.0113583

Nadiya, F., Anjali, N., Thomas, J., Gangaprasad, A., and Sabu, K. K. (2017). Transcriptome Profiling of *Elettaria Cardamomum* (L.) Maton (Small Cardamom). *Genomics Data* 11, 102–103. doi:10.1016/j.gdata.2016.12.013

Naidoo, D., Slavětínská, L. P., Aremu, A. O., Gruz, J., Biba, O., Doležal, K., et al. (2018). Metabolite Profiling and Isolation of Biologically Active Compounds fromScadoxus Puniceus, a Highly Traded South African Medicinal Plant. *Phytotherapy Res.* 32 (4), 625–630. doi:10.1002/ptr.6000

Nayak, S. N., Hebbal, V., Bharati, P., Nadaf, H. L., Naidu, G. K., and Bhat, R. S. (2020). Profiling of Nutraceuticals and Proximates in Peanut Genotypes Differing for Seed Coat Color and Seed Size. *Front. Nutr.* 7, 45. doi:10.3389/fnut.2020.00045

Newell, M. A., Asoro, F. G., Scott, M. P., White, P. J., Beavis, W. D., and Jannink, J.-L. (2012). Genome-wide Association Study for Oat (*Avena Sativa* L.) Beta-Glucan Concentration Using Germplasm of Worldwide Origin. *Theor. Appl. Genet.* 125 (8), 1687–1696. doi:10.1007/s00122-012-1945-0

Niazian, M., and Niedbała, G. (2020). Machine Learning for Plant Breeding and Biotechnology. *Agriculture* 10 (10), 436. doi:10.3390/agriculture10100436

Ogita, S., Uefuji, H., Yamaguchi, Y., Koizumi, N., and Sano, H. (2003). Producing Decaffeinated Coffee Plants. *Nature* 423 (6942), 823. doi:10.1038/423823a

Palou, A., Picó, C., and Bonet, M. L. (2004). Food Safety and Functional Foods in the European Union: Obesity as a Paradigmatic Example for Novel Food Development. *Nutr. Rev.* 62 (Suppl. l_2), S169–S181. doi:10.1111/j.1753-4887.2004.tb00089.x

Pan, C., Ye, L., Qin, L., Liu, X., He, Y., Wang, J., et al. (2016). CRISPR/Cas9-mediated Efficient and Heritable Targeted Mutagenesis in Tomato Plants in the First and Later Generations. *Sci. Rep.* 6 (1), 24765–24769. doi:10.1038/srep24765

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., et al. (2014). Identification of QTLs Associated with Oil Content and Mapping FAD2 Genes and Their Relative Contribution to Oil Quality in Peanut (*Arachis hypogaea* L.). *BMC Genet.* 15 (1), 133–214. doi:10.1186/s12863-014-0133-4

Pang, Y., Abeysinghe, I. S. B., He, J., He, X., Huhman, D., Mewan, K. M., et al. (2013). Functional Characterization of Proanthocyanidin Pathway Enzymes from Tea and Their Application for Metabolic Engineering. *Plant Physiol.* 161 (3), 1103–1116. doi:10.1104/pp.112.212050

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, 510–515. doi:10.1186/s12870-019-2117-z

Park, M. Y., Yi, N. R., Lee, H. Y., Kim, S. T., Kim, M., Park, J.-H., et al. (2002). Generation of Chlorsulfuron-Resistant Transgenic Garlic Plants (*Allium*

Sativum L.) by Particle Bombardment. *Mol. Breed.* 9 (3), 171–181. doi:10.1023/a:1019702705996

Pattnaik, M., Pandey, P., Martin, G. J. O., Mishra, H. N., and Ashokkumar, M. (2021). Innovative Technologies for Extraction and Microencapsulation of Bioactives from Plant-Based Food Waste and Their Applications in Functional Food Development. *Foods* 10 (2), 279. doi:10.3390/foods10020279

Pazhamala, L. T., Kudapa, H., Weckwerth, W., Millar, A. H., and Varshney, R. K. (2021). Systems Biology for Crop Improvement. *Plant Genome (TSI)*, 1–23. doi:10.1002/tpg2.20098

Pellati, F., Bruni, R., Righi, D., Grandini, A., Tognolini, M., Pio Prencipe, F., et al. (2013). Metabolite Profiling of Polyphenols in a Terminalia Chebula Retzius Ayurvedic Decoction and Evaluation of its Chemopreventive Activity. *J. Ethnopharmacology* 147 (2), 277–285. doi:10.1016/j.jep.2013.02.025

Pérez-Díaz, R., Madrid-Espinoza, J., Salinas-Cornejo, J., González-Villanueva, E., and Ruiz-Lara, S. (2016). Differential Roles for VviGST1, VviGST3, and VviGST4 in Proanthocyanidin and Anthocyanin Transport in Vitis Vinífera. *Front. Plant Sci.* 7, 1166. doi:10.3389/fpls.2016.01166

Prakash, B., Kujur, A., Yadav, A., Kumar, A., Singh, P. P., and Dubey, N. K. (2018). Nanoencapsulation: An Efficient Technology to Boost the Antimicrobial Potential of Plant Essential Oils in Food System. *Food Control* 89, 1–11. doi:10.1016/j.foodcont.2018.01.018

Pu, Z.-e., Yu, M., He, Q.-y., Chen, G.-y., Wang, J.-r., Liu, Y.-x., et al. (2014). Quantitative Trait Loci Associated with Micronutrient Concentrations in Two Recombinant Inbred Wheat Lines. *J. Integr. Agric.* 13 (11), 2322–2329. doi:10.1016/s2095-3119(13)60640-1

Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J. (2013). Genome-wide Annotation of Genes and Noncoding RNAs of Foxtail Millet in Response to Simulated Drought Stress by Deep Sequencing. *Plant Mol. Biol.* 83 (4-5), 459–473. doi:10.1007/s11103-013-0104-6

Qi, Z., Han, X., Hou, M., Xin, D., Wang, Z., Zhu, R., et al. (2014). QTL Analysis of Soybean Oil Content under 17 Environments. *Can. J. Plant Sci.* 94 (2), 245–261. doi:10.4141/cjps2013-219

Qu, C., Zhao, H., Fu, F., Zhang, K., Yuan, J., Liu, L., et al. (2016). Molecular Mapping and QTL for Expression Profiles of Flavonoid Genes in *Brassica Napus*. *Front. Plant Sci.* 7, 1691. doi:10.3389/fpls.2016.01691

Rai, A., Yamazaki, M., Takahashi, H., Nakamura, M., Kojoma, M., Suzuki, H., et al. (2016). RNA-seq Transcriptome Analysis of Panax Japonicus, and its Comparison with Other Panax Species to Identify Potential Genes Involved in the Saponins Biosynthesis. *Front. Plant Sci.* 7, 481. doi:10.3389/fpls.2016.00481

Rao, A. G., Hassan, M., and Hemple, C. (1994). Structure-function Validation of High Lysine Analogs of α-hordothionin Designed by Protein Modeling. *Protein Eng. Des. Sel.* 7 (12), 1485–1493. doi:10.1093/protein/7.12.1485

Raza, Q., Riaz, A., Sabar, M., Atif, R. M., and Bashir, K. (2019). Meta-analysis of Grain Iron and Zinc Associated QTLs Identified Hotspot Chromosomal Regions and Positional Candidate Genes for Breeding Biofortified rice. *Plant Sci.* 288, 110214. doi:10.1016/j.plantsci.2019.110214

Regina, A., Kosar-Hashemi, B., Ling, S., Li, Z., Rahman, S., and Morell, M. (2010). Control of Starch Branching in Barley Defined through Differential RNAi Suppression of Starch Branching Enzyme IIa and IIb. *J. Exp. Bot.* 61 (5), 1469–1482. doi:10.1093/jxb/erq011

Rhodes, D. H., Hoffmann, L., Jr, Rooney, W. L., Ramu, P., Morris, G. P., and Kresovich, S. (2014). Genome-wide Association Study of Grain Polyphenol Concentrations in Global Sorghum [*Sorghum Bicolor* (L.) Moench] Germplasm. *J. Agric. Food Chem.* 62 (45), 10916–10927. doi:10.1021/jf503651t

Riccardi, G., Capaldo, B., and Vaccaro, O. (2005). Functional Foods in the Management of Obesity and Type 2 Diabetes. *Curr. Opin. Clin. Nutr. Metabol. Care* 8 (6), 630–635. doi:10.1097/01.mco.0000171126.98783.0c

Richards, C., Hurst, B., Messner, R., and O'Connor, G. (2021). The Paradoxes of Food Waste Reduction in the Horticultural Supply Chain. *Ind. Mark. Manag.* 93, 82–491. doi:10.1016/j.indmarman.2020.12.002

Roesler, K., Shintani, D., Savage, L., Boddupalli, S., and Ohlrogge, J. (1997). Targeting of the Arabidopsis Homomeric Acetyl-Coenzyme A Carboxylase to Plastids of Rapeseeds. *Plant Physiol.* 113 (1), 75–81. doi:10.1104/pp.113.1.75

Sagare, D. B., Shetti, P., Surender, M., and Reddy, S. (2019). Marker-assisted Backcross Breeding for Enhancing β-carotene of QPM Inbreds. *Mol. Breed.* 39, 1–12. doi:10.1007/s11032-019-0939-x

Saito, K. (2018). Development of Plant Metabolomics and Medicinal Plant Genomics. *Yakugaku Zasshi* 138 (1), 1–18. doi:10.1248/yakushi.17-00193

Santhosha, S. G., Jamuna, P., and Prabhavathi, S. N. (2013). Bioactive Components of Garlic and Their Physiological Role in Health Maintenance: A Review. *Food Biosci.* 3, 59–74. doi:10.1016/j.fbio.2013.07.001

Sayeed, M. A., Bracci, M., Lazzarini, R., Tomasetti, M., Amati, M., Lucarini, G., et al. (2017). Use of Potential Dietary Phytochemicals to Target miRNA: Promising Option for Breast Cancer Prevention and Treatment? *J. Funct. Foods* 28, 177–193. doi:10.1016/j.jff.2016.11.008

Schijlen, E., Ric de Vos, C. H., Jonker, H., Van Den Broeck, H., Molthoff, J., Van Tunen, A., et al. (2006). Pathway Engineering for Healthy Phytochemicals Leading to the Production of Novel Flavonoids in Tomato Fruit. *Plant Biotechnol. J* 4 (4), 433–444. doi:10.1111/j.1467-7652.2006.00192.x

Sedeek, K. E. M., Mahas, A., and Mahfouz, M. (2019). Plant Genome Engineering for Targeted Improvement of Crop Traits. *Front. Plant Sci.* 10, 114. doi:10.3389/fpls.2019.00114

Shahidi, F., and Ambigaipalan, P. (2015). Phenolics and Polyphenolics in Foods, Beverages and Spices: Antioxidant Activity and Health Effects - A Review. *J. Funct. Foods* 18, 820–897. doi:10.1016/j.jff.2015.06.018

Shan, Q., Zhang, Y., Chen, K., Zhang, K., and Gao, C. (2015). Creation of Fragrant rice by Targeted Knockout of theOsBADH2gene Using TALEN Technology. *Plant Biotechnol. J.* 13 (6), 791–800. doi:10.1111/pbi.12312

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 (5), 1005–1016. doi:10.1007/s00122-010-1505-4

Sharma, H. C., Crouch, J. H., Sharma, K. K., Seetharama, N., and Hash, C. T. (2002). Applications of Biotechnology for Crop Improvement: Prospects and Constraints. *Plant Sci.* 163 (3), 381–395. doi:10.1016/s0168-9452(02)00133-4

Sharma, K., Kumar, V., Kaur, J., Tanwar, B., Goyal, A., Sharma, R., et al. (2019). Health Effects, Sources, Utilization and Safety of Tannins: A Critical Review. *Toxin Rev.*, 1–13. doi:10.1080/15569543.2019.1662813

Sharma, P., Gaur, V. K., Sirohi, R., Varjani, S., Hyoun Kim, S., and Wong, J. W. C. (2021). Sustainable Processing of Food Waste for Production of Bio-Based Products for Circular Bioeconomy. *Bioresour. Technology* 325, 124684. doi:10.1016/j.biortech.2021.124684

Shasidhar, Y., Vishwakarma, M. K., Pandey, M. K., Janila, P., Variath, M. T., Manohar, S. S., et al. (2017). Molecular Mapping of Oil Content and Fatty Acids Using Dense Genetic Maps in Groundnut (*Arachis hypogaea* L.). *Front. Plant Sci.* 8, 794. doi:10.3389/fpls.2017.00794

Shetti, P., Sagare, D. B., Surender, M., and Reddy, S. S. (2020). Development of Lysine and Tryptophan Rich maize (Zea mays) Inbreds Employing Marker Assisted Backcross Breeding. *Plant Gene* 23, 100236. doi:10.1016/j.plgene.2020.100236

Shewmaker, C. K., Sheehy, J. A., Daley, M., Colburn, S., and Ke, D. Y. (1999). Seed-specific Overexpression of Phytoene Synthase: Increase in Carotenoids and Other Metabolic Effects. *Plant J.* 20 (4), 401–412. doi:10.1046/j.1365-313x.1999.00611.x

Shi, J., Arunasalam, K., Yeung, D., Kakuda, Y., Mittal, G., and Jiang, Y. (2004). Saponins from Edible Legumes: Chemistry, Processing, and Health Benefits. *J. Med. Food* 7, 67–78. doi:10.1089/109662004322984734

Shirgurkar, M. V., Naik, V. B., von Arnold, S., Nadgauda, R. S., and Clapham, D. (2006). An Efficient Protocol for Genetic Transformation and Shoot Regeneration of Turmeric (*Curcuma Longa* L.) via Particle Bombardment. *Plant Cel Rep* 25 (2), 112–116. doi:10.1007/s00299-005-0033-1

Shu, X., and Rasmussen, S. r. K. (2014). Quantification of Amylose, Amylopectin, and Î²-Glucan in Search for Genes Controlling the Three Major Quality Traits in Barley by Genome-wide Association Studies. *Front. Plant Sci.* 5, 197. doi:10.3389/fpls.2014.00197

Sies, H., and Stahl, W. (1995). Vitamins E and C, Beta-Carotene, and Other Carotenoids as Antioxidants. *AM. J. Clin. Nutr.* 62, 1315S–1321S. doi:10.1093/ajcn/62.6.1315s

Singh, B., Singh, J. P., Singh, N., and Kaur, A. (2017). Saponins in Pulses and Their Health Promoting Activities: A Review. *Food Chem.* 233, 540–549. doi:10.1016/j.foodchem.2017.04.161

Smith, S. A., King, R. E., and Min, D. B. (2007). Oxidative and thermal Stabilities of Genetically Modified High Oleic sunflower Oil. *Food Chem.* 102 (4), 1208–1213. doi:10.1016/j.foodchem.2006.06.058

Sowbhagya, H. B. (2013). Chemistry, Technology, and Nutraceutical Functions of Cumin (*Cuminum Cyminum* L): an Overview. *Crit. Rev. Food Sci. Nutr.* 53 (1), 1–10. doi:10.1080/10408398.2010.500223

Stark, A. H., Crawford, M. A., and Reifen, R. (2008). Update on Alpha-Linolenic Acid. *Nutr. Rev.* 66, 326–332. doi:10.1111/j.1753-4887.2008.00040.x

Street, R. A., Prinsloo, G., and Mcgaw, L. J. (2017). "Alkaloids Potential Health Benefits and Toxicity," in *Utilisation of Bioactive Compounds from Agricultural and Food Production Waste* (Boca Raton, FL: CRC Press), 60–85.

Sui, M., Jing, Y., Li, H., Zhan, Y., Luo, J., Teng, W., et al. (2020). Identification of Loci and Candidate Genes Analyses for Tocopherol Concentration of Soybean Seed. *Front. Plant Sci.* 11, 539460. doi:10.3389/fpls.2020.539460

Sun, P., Cheng, C., Lin, Y., Zhu, Q., Lin, J., and Lai, Z. (2017a). Combined Small RNA and Degradome Sequencing Reveals Complex microRNA Regulation of Catechin Biosynthesis in tea (*Camellia Sinensis*). *PLoS One* 12 (2), e0171173. doi:10.1371/journal.pone.0171173

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017b). Generation of High-Amylose rice through CRISPR/Cas9-mediated Targeted Mutagenesis of Starch Branching Enzymes. *Front. Plant Sci.* 8, 298. doi:10.3389/fpls.2017.00298

Thavarajah, D., Thavarajah, P., and Gupta, D. S. (2014). Pulses Biofortification in Genomic Era: Multidisciplinary Opportunities and Challenges. *Legumes in the Omic Era*, 207–220. doi:10.1007/978-1-4614-8370-0_10

Thies, F., Mills, L. M., Moir, S., and Masson, L. F. (2017). Cardiovascular Benefits of Lycopene: Fantasy or Reality? *Proc. Nutr. Soc.* 76, 122–129. doi:10.1017/s0029665116000744

Tian, J., Bryksa, B. C., and Yada, R. Y. (2016). Feeding the World into the Future - Food and Nutrition Security: the Role of Food Science and Technology. *Front. Life Sci.* 9 (3), 155–166. doi:10.1080/21553769.2016.1174958

Tian, Y. S., Wang, B., Peng, R. H., Xu, J., Li, T., Fu, X. Y., et al. (2019). Enhancing Carotenoid Biosynthesis in rice Endosperm by Metabolic Engineering. *Plant Biotechnol. J.* 17, 849–851. doi:10.1111/pbi.13059

Tong, H., and Nikoloski, Z. (2021). Machine Learning Approaches for Crop Improvement: Leveraging Phenotypic and Genotypic Big Data. *J. Plant Physiol.* 257, 153354. doi:10.1016/j.jplph.2020.153354

Tran, H. T. M., Ramaraj, T., Furtado, A., Lee, L. S., and Henry, R. J. (2018). Use of a Draft Genome of Coffee (C Offea Arabica) to Identify SNP S Associated with Caffeine Content. *Plant Biotechnol. J.* 16, 1756–1766. doi:10.1111/pbi.12912

Trijatmiko, K. R., Dueñas, C., Tsakirpaloglou, N., Torrizo, L., Arines, F. M., Adeva, C., et al. (2016). Biofortified Indica rice Attains Iron and Zinc Nutrition Dietary Targets in the Field. *Sci. Rep.* 6 (1), 19792–19813. doi:10.1038/srep19792

Tripathi, S., Sangwan, R. S., Mishra, B., Jadaun, J. S., and Sangwan, N. S. (2020). Berry Transcriptome: Insights into a Novel Resource to Understand Development Dependent Secondary Metabolism in Withania Somnifera (Ashwagandha). *Physiol. Plantarum* 168 (1), 148–173. doi:10.1111/ppl.12943

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., and Dubcovsky, J. (2006). A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. *Science* 314 (5803), 1298–1301. doi:10.1126/science.1133649

Van Eenennaam, A. L., Lincoln, K., Durrett, T. P., Valentin, H. E., Shewmaker, C. K., Thorne, G. M., et al. (2003). Engineering Vitamin E Content: from Arabidopsis Mutant to Soy Oil. *Plant Cell* 15 (12), 3007–3019. doi:10.1105/tpc.015875

Varshney, R., Graner, A., and Sorrells, M. (2005). Genomics-assisted Breeding for Crop Improvement. *Trends Plant Sci.* 10 (12), 621–630. doi:10.1016/j.tplants.2005.10.004

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 (9), 522–530. doi:10.1016/j.tibtech.2009.05.006

Vasconcelos, M., Datta, K., Oliva, N., Khalekuzzaman, M., Torrizo, L., Krishnan, S., et al. (2003). Enhanced Iron and Zinc Accumulation in Transgenic rice with the Ferritin Gene. *Plant Sci.* 164 (3), 371–378. doi:10.1016/s0168-9452(02)00421-1

Vega-Gálvez, A., Miranda, M., Aranda, M., Henriquez, K., Vergara, J., Tabilo-Munizaga, G., et al. (2011). Effect of High Hydrostatic Pressure on Functional Properties and Quality Characteristics of Aloe Vera Gel (*Aloe Barbadensis* Miller). *Food Chem.* 129 (3), 1060–1065. doi:10.1016/j.foodchem.2011.05.074

Velmurugan, B., Mani, A., and Nagini, S. (2005). Combination of S-Allylcysteine and Lycopene Induces Apoptosis by Modulating Bcl-2, Bax, Bim and Caspases during Experimental Gastric Carcinogenesis. *Eur. J. Cancer Prev.* 14 (4), 387–393. doi:10.1097/00008469-200508000-00012

Vetriventhan, M., Azevedo, V. C. R., Upadhyaya, H. D., Nirmalakumari, A., Kane-Potaka, J., Anitha, S., et al. (2020). Genetic and Genomic Resources, and Breeding for Accelerating Improvement of Small Millets: Current Status and Future Interventions. *Nucleus* 63, 217–239. doi:10.1007/s13237-020-00322-3

Vinoth, A., and Ravindhran, R. (2017). Biofortification in Millets: a Sustainable Approach for Nutritional Security. *Front. Plant Sci.* 8, 29. doi:10.3389/fpls.2017.00029

Wang, C., Qiao, A., Fang, X., Sun, L., Gao, P., Davis, A. R., et al. (2019a). Fine Mapping of Lycopene Content and Flesh Color Related Gene and Development of Molecular Marker-Assisted Selection for Flesh Color in Watermelon (*Citrullus lanatus*). *Front. Plant Sci.* 10, 1240. doi:10.3389/fpls.2019.01240

Wang, C., Zeng, J., Li, Y., Hu, W., Chen, L., Miao, Y., et al. (2014). Enrichment of Provitamin A Content in Wheat (*Triticum aestivum* L.) by Introduction of the Bacterial Carotenoid Biosynthetic Genes *CrtB* and *CrtI*. *J. Exp. Bot.* 65 (9), 2545–2556. doi:10.1093/jxb/eru138

Wang, M., Amakye, W. K., Guo, L., Gong, C., Zhao, Y., Yao, M., et al. (2019b). Walnut-Derived Peptide PW5 Ameliorates Cognitive Impairments and Alters Gut Microbiota in APP/PS1 Transgenic Mice. *Mol. Nutr. Food Res.* 63, 1900326. doi:10.1002/mnfr.201900326

Wang, Q., Reddy, V. A., Panicker, D., Mao, H.-Z., Kumar, N., Rajan, C., et al. (2016). Metabolic Engineering of Terpene Biosynthesis in Plants Using a Trichome-specific Transcription factorMsYABBY5from Spearmint (Mentha Spicata). *Plant Biotechnol. J.* 14 (7), 1619–1632. doi:10.1111/pbi.12525

Wei, Y., Bergenholm, D., Gossing, M., Siewers, V., and Nielsen, J. (2018). Expression of cocoa Genes in *Saccharomyces cerevisiae* Improves cocoa Butter Production. *Microb. Cel Fact* 17 (1), 11. doi:10.1186/s12934-018-0866-2

Welch, R. M., and Graham, R. D. (2004). Breeding for Micronutrients in Staple Food Crops from a Human Nutrition Perspective. *J. Exp. Bot.* 55 (396), 353–364. doi:10.1093/jxb/erh064

White, J., Pacey-Miller, T., Crawford, A., Cordeiro, G., Barbary, D., Bundock, P., et al. (2006). Abundant Transcripts of Malting Barley Identified by Serial Analysis of Gene Expression (SAGE). *Plant Biotechnol. J.* 4 (3), 289–301. doi:10.1111/j.1467-7652.2006.00181.x

Wu, H.-x., Jia, H.-m., Ma, X.-w., Wang, S.-b., Yao, Q.-s., Xu, W.-t., et al. (2014). Transcriptome and Proteomic Analysis of Mango (*Mangifera Indica* Linn) Fruits. *J. Proteomics* 105, 19–30. doi:10.1016/j.jprot.2014.03.030

Wu, Q., Chen, Z., Sun, W., Deng, T., and Chen, M. (2016). De Novo Sequencing of the Leaf Transcriptome Reveals Complex Light-Responsive Regulatory Networks in Camellia Sinensis Cv. Baijiguan. *Front. Plant Sci.* 7, 332. doi:10.3389/fpls.2016.00332

Wu, Y., Li, X., Xiang, W., Zhu, C., Lin, Z., Wu, Y., et al. (2012). Presence of Tannins in Sorghum Grains Is Conditioned by Different Natural Alleles of Tannin1. *Proc. Natl. Acad. Sci.* 109 (26), 10281–10286. doi:10.1073/pnas.1201700109

Xu, L.-Y., Wang, L.-Y., Wei, K., Tan, L.-Q., Su, J.-J., and Cheng, H. (2018). High-density SNP Linkage Map Construction and QTL Mapping for Flavonoid-Related Traits in a tea Plant (Camellia Sinensis) Using 2b-RAD Sequencing. *BMC Genomics* 19, 955. doi:10.1186/s12864-018-5291-8

Yadav, M., Sehrawat, N., Singh, M., Upadhyay, S. K., Aggarwal, D., and Sharma, A. K. (2020). Cardioprotective and Hepatoprotective Potential of Citrus Flavonoid Naringin: Current Status and Future Perspectives for Health Benefits. *Ajbls* 9, 1–5. doi:10.5530/ajbls.2020.9.1

Yang, R., Piao, Z., Wan, C., Lee, G., Ruan, X., and Bai, J. (2020). Breeding for Three-Line Japonica Hybrid rice Combinations with High Resistant Starch Content Using Molecular Marker-Assisted Selection. *Breed. Sci.* 70, 409–414. doi:10.1270/jsbbs.20005

Yang, R., Yan, Z., Wang, Q., Li, X., and Feng, F. (2018). Marker-assisted Backcrossing of *lcyE* for Enhancement of proA in Sweet Corn. *Euphytica* 214, 1–12. doi:10.1007/s10681-018-2212-5

Yao, L. H., Jiang, Y. M., Shi, J., Datta, N., Singanusong, R., Chen, S. S., et al. (2004). Flavonoids in Food and Their Health Benefits. *Plant Foods Hum. Nutr.* 59, 113–122. doi:10.1007/s11130-004-0049-7

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., et al. (2000). Engineering the Provitamin A (-Carotene) Biosynthetic Pathway into (Carotenoid-free) Rice Endosperm. *Science* 287 (5451), 303–305. doi:10.1126/science.287.5451.303

Yousuf, B., Gul, K., Wani, A. A., and Singh, P. (2016). Health Benefits of Anthocyanins and Their Encapsulation for Potential Use in Food Systems: a

Review. *Crit. Rev. Food Sci. Nutr.* 56, 2223–2230. doi:10.1080/10408398.2013.805316

Yu, B., Lydiate, D. J., Young, L. W., Schäfer, U. A., and Hannoufa, A. (2008). Enhancing the Carotenoid Content of *Brassica Napus* Seeds by Downregulating Lycopene Epsilon Cyclase. *Transgenic Res.* 17 (4), 573–585. doi:10.1007/s11248-007-9131-x

Yusuf, M. A., and Sarin, N. B. (2007). Antioxidant Value Addition in Human Diets: Genetic Transformation of Brassica Juncea with γ-TMT Gene for Increased α-tocopherol Content. *Transgenic Res.* 16 (1), 109–113. doi:10.1007/s11248-006-9028-0

Zenda, T., Liu, S., Dong, A., and Duan, H. (2021). Advances in Cereal Crop Genomics for Resilience under Climate Change. *Life* 11 (6), 502. doi:10.3390/life11060502

Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., et al. (2017). RNA Sequencing Provides Insights into the Evolution of Lettuce and the Regulation of Flavonoid Biosynthesis. *Nat. Commun.* 8 (1), 2264–2312. doi:10.1038/s41467-017-02445-9

Zhang, M., Pinson, S. R. M., Tarpley, L., Huang, X.-Y., Lahner, B., Yakubova, E., et al. (2014). Mapping and Validation of Quantitative Trait Loci Associated with Concentrations of 16 Elements in Unmilled rice Grain. *Theor. Appl. Genet.* 127 (1), 137–165. doi:10.1007/s00122-013-2207-5

Zhang, P., Guo, Z., Ullah, S., Melagraki, G., Afantitis, A., and Lynch, I. (2021). Nanotechnology and Artificial Intelligence to Enable Sustainable and Precision Agriculture. *Nat. Plants* 7 (7), 864–876. doi:10.1038/s41477-021-00946-6

Zhao, Y., Sun, H. Y., Wang, Y. Y., Pu, Y. Y., Kong, F. M., and Li, S. S. (2013). QTL Mapping for the Color, Carotenoids and Polyphenol Oxidase Activity of Flour in Recombinant Inbred Lines of Wheat. *Austr. J. Crop Sci.* 7 (3), 328.

Zhao, Z.-y., Glassman, K., Sewalt, V., Wang, N., Miller, M., Chang, S., et al. (2003). "Nutritionally Improved Transgenic Sorghum," in *Plant Biotechnology 2002 and beyond* (Springer), 413–416. doi:10.1007/978-94-017-2679-5_85

Zhao, Z.-y. (2007). "The Africa Biofortified Sorghum Project- Applying Biotechnology to Develop Nutritionally Improved Sorghum for Africa," in *Biotechnology and Sustainable Agriculture 2006 and beyond* (Springer), 273–277. doi:10.1007/978-1-4020-6635-1_41

Zheng, X., Tang, Y., Ye, J., Pan, Z., Tan, M., Xie, Z., et al. (2018). SLAF-based Construction of a High-Density Genetic Map and its Application in QTL Mapping of Carotenoids Content in Citrus Fruit. *J. Agric. Food Chem.* 67 (3), 994–1002. doi:10.1021/acs.jafc.8b05176

Zhou, S., Kremling, K. A., Bandillo, N., Richter, A., Zhang, Y. K., Ahern, K. R., et al. (2019). Metabolome-scale Genome-wide Association Studies Reveal Chemical Diversity and Genetic Control of maize Specialized Metabolites. *Plant Cell* 31 (5), 937–955. doi:10.1105/tpc.18.00772

Zhu, Q., Zeng, D., Yu, S., Cui, C., Li, J., Li, H., et al. (2018). From Golden Rice to aSTARice: Bioengineering Astaxanthin Biosynthesis in rice Endosperm. *Mol. Plant* 11, 1440–1448. doi:10.1016/j.molp.2018.09.007

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

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

Copyright © 2021 Nayak, Aravind, Malavalli, Sukanth, Poornima, Bharati, Hefferon, Kole and Puppala. 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

PERSPECTIVE
published: 12 November 2021
doi: 10.3389/fgene.2021.786140

Utilizing CRISPR-Cas in Tropical Crop Improvement: A Decision Process for Fitting Genome Engineering to Your Species

*Kathleen A. Joo[1], Michael G. Muszynski[1], Michael B. Kantar[1], Ming-Li Wang[2], Xiaoling He[2] and Angel R. Del Valle Echevarria[1,2]**

[1]Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, Honolulu, HI, United States, [2]Hawaii Agriculture Research Center, Waipahu, HI, United States

Adopting modern gene-editing technologies for trait improvement in agriculture requires important workflow developments, yet these developments are not often discussed. Using tropical crop systems as a case study, we describe a workflow broken down into discrete processes with specific steps and decision points that allow for the practical application of the CRISPR-Cas gene editing platform in a crop of interest. While we present the steps of developing genome-edited plants as sequential, in practice parts can be done in parallel, which are discussed in this perspective. The main processes include 1) understanding the genetic basis of the trait along with having the crop's genome sequence, 2) testing and optimization of the editing reagents, development of efficient 3) tissue culture and 4) transformation methods, and 5) screening methods to identify edited events with commercial potential. Our goal in this perspective is to help any lab that wishes to implement this powerful, easy-to-use tool in their pipeline, thus aiming to democratize the technology.

Keywords: non-commodity, science democratization, orphan crop, genome engineering, CRISPR-cas, tropical crop

OPEN ACCESS

Edited by:
Ramsey Steven Lewis, North Carolina State University, United States

Reviewed by:
Jaindra Nath Tripathi, International Institute of Tropical Agriculture (IITA), Kenya
Kedong George Da, North Carolina State University, United States

****Correspondence:***
Angel R. Del Valle Echevarria
aechevarria@harc-hspa.com

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

Received: *29 September 2021*
Accepted: *29 October 2021*
Published: *12 November 2021*

Citation:
Joo KA, Muszynski MG, Kantar MB, Wang M-L He X and Del Valle Echevarria AR (2021) Utilizing CRISPR-Cas in Tropical Crop Improvement: A Decision Process for Fitting Genome Engineering to Your Species. Front. Genet. 12:786140. doi: 10.3389/fgene.2021.786140

INTRODUCTION

Since its proposal as a eukaryotic gene-editing tool (Jinek et al., 2012), the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-Associated protein (Cas) technology has been widely applied in microorganisms, animals, and plants to study gene function (Haque et al., 2018) due to its simplicity in design and straightforward execution. Numerous CRISPR-Cas system reviews explain its discovery and aspects to consider when using this powerful molecular tool, so we suggest reviewing Anzalone et al., 2020 for details on specifics, such as the diverse engineered Cas nucleases. For this perspective, the reader must know that the CRISPR-Cas tool is composed of a small guide RNA (sgRNA) complementary to a DNA target sequence and a Cas endonuclease (i.e., Cas9 and Cas12a). The Cas endonuclease recognizes a protospacer adjacent motif (PAM) sequence that is upstream [5′-TTTV-(22 nt of target sequence)-3′] for Cas12a or downstream [5′-(20 nt of target sequence)-NGG-3′] for Cas9 of the target sequence (Jinek et al., 2012; Zetsche et al., 2015). The Cas nuclease associates with the sgRNA to form a ribonucleoprotein (RNP) complex, which scans the genome for the PAM sequence and, by complementation, the DNA target sequence (Jinek et al., 2012; Zetsche et al., 2015). The RNP complex catalyzes a double-strand break (DSB) in the target DNA, triggering

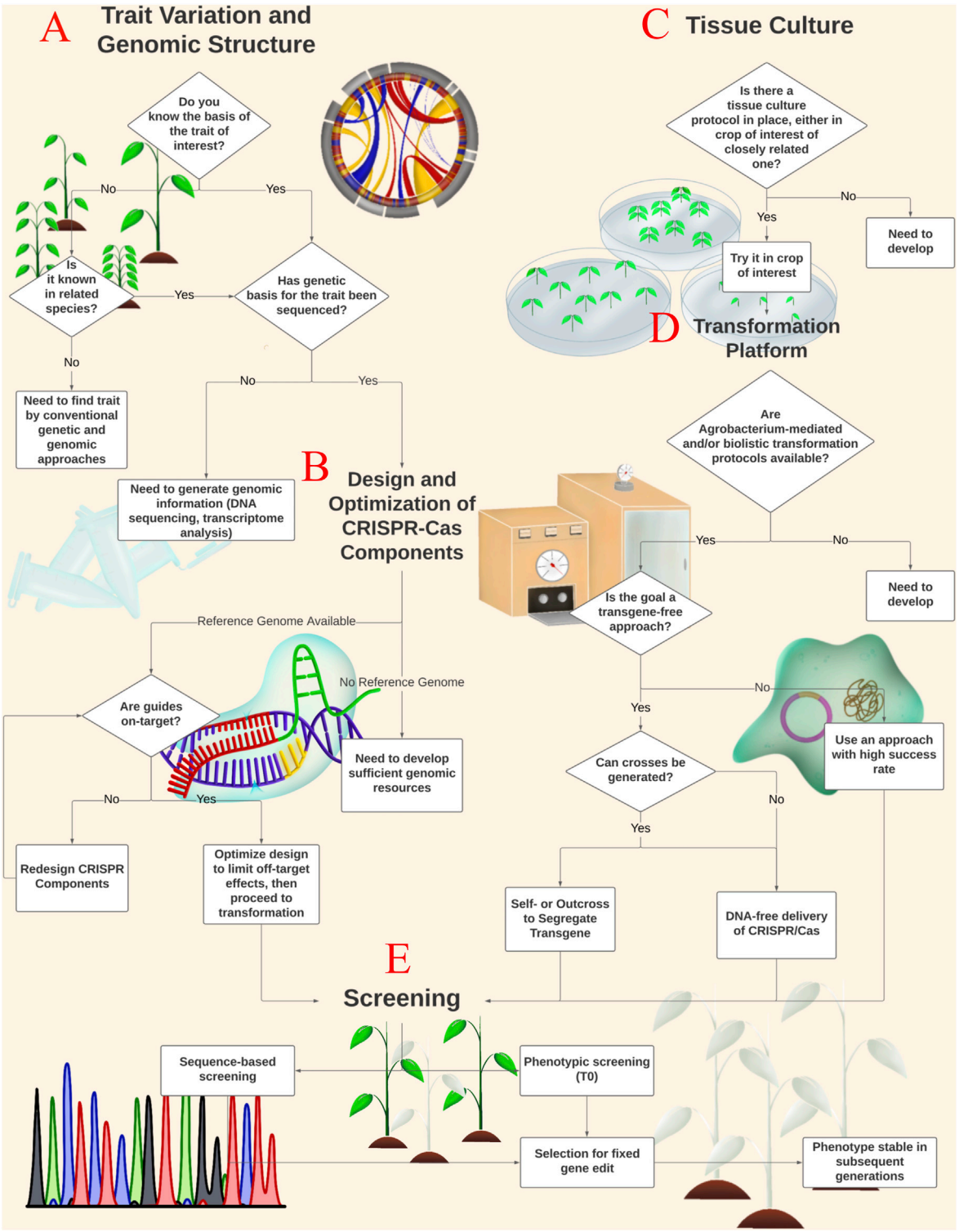


FIGURE 1 | CRISPR-Cas Gene Editing Workflow. Here we present a schematic for implementing CRISPR-Cas gene editing technology in a crop of interest. Diamonds represent decision points, while boxes represent important processes. In order to develop a gene editing protocol using CRISPR-Cas technology, the genetic basis for the trait of interest must be identified and well described **(A)**. If genomic information is non-existent or limited for the varietal of interest, it must be extrapolated from existing data on similar genotypes or from related species. Adequate genomic information is essential for the design and optimization **of** CRISPR-Cas sgRNA, nuclease selection, and off-target analysis **(B)**. Identification of the genetic basis for a trait **(A)** and design of CRISPR-Cas components **(B)** constitute the genomic aspect of the workflow, and consideration for both can take place independently of tissue culture **(C)** and transformation **(D)**. We suggest testing various tissue culture protocols

(*Continued*)

FIGURE 1 | when they do not exist for the varietal of interest but do exist for a different genotype or related species **(C).** CRISPR-Cas reagents may be introduced to cultured cells in several ways including *Agrobacterium*-mediated transformation, transient expression from transfected plasmids, and biolistic transformation with RNP or RNA complexes. Selection of the transformation method will depend on the goals of the research. Following CRISPR-Cas design **(B)** and transformation **(D)**, explants should be screened, field tested, and propagated **(E)** to generate a phenotypically stable, CRISPR-Cas gene edited population. Screening for edited events **(E)** will involve sequencing for allelic differences and phenotypic selection. Depending on the goals of the research, marker-guided selection and resistance genes may also be used to screen for edited events.

the cell's error-prone DNA repair mechanism, which results in the creation of a mutation that may generate a desirable change in a trait of interest (Woo et al., 2015).

In agriculture, CRISPR-Cas technology has been used to introduce added-value traits (Tian et al., 2018; Kaur et al., 2020; Yoon et al., 2020) and to recapitulate domestication processes (Soyk et al., 2017, 2019; Lemmon et al., 2018). Based on these examples, it is clear that the application of the CRISPR-Cas tool, coupled with traditional breeding practices, has tremendous potential in alleviating threats to food security, production, and sustainability (Jung and Till, 2021). Numerous published studies state that the CRISPR-Cas tool can be used in any crop due to its simplicity, but there are key aspects to consider when applying this technology to specific types of crops.

Tropical crops present an appropriate case study where implementation of the CRISPR-Cas technology would significantly contribute to its improvement. However, a lack of genomic resources, complex tissue culture procedures, and reproductive compatibility constraints in many tropical crops make utilizing this robust tool challenging (Atkins and Voytas, 2020; Maher et al., 2020). Although promising advances in tropical crop improvement using CRISPR-Cas have recently been achieved, most focus on proof-of-concept studies (Odipio et al., 2017; Naim et al., 2018; Fan et al., 2020; Ntui et al., 2020; Eid et al., 2021; Syombua et al., 2021; Zhao et al., 2021) and few on value-added traits (Gomez et al., 2017; Bull et al., 2018; Fister et al., 2018; Mehta et al., 2019; Oz et al., 2021). In this perspective we outline critical points to consider when executing a gene-editing project in a tropical crop species (**Figure 1**). This perspective will outline key considerations at the main decision points throughout the gene editing workflow.

Trait Variation and Genomic Structure

The foundation of any gene-editing project relies on understanding the genetic mechanism(s) underlying trait variation within the species of interest and identifying where to source specific varietal material (i.e., genebanks, elite lines). There are numerous international and regional institutions dedicated to conserving tropical collections (FAO, 2010). However, inherent properties of common tropical crops, such as recalcitrant seeds (Bourdeix et al., 2020), low seed number or viability (Batte et al., 2019; Kallow et al., 2020; Mertens et al., 2021), and laborious maintenance of vegetatively propagated crops (Balogun, 2009; Gaba and Singer, 2009; Panis et al., 2020), tend to limit collection size.

Plant breeding efforts in tropical crops often focus on domestication traits first, followed by value-added commercialization traits (Ramstein et al., 2019; Bernardo, 2020). Several challenges for efficient breeding in tropical crops include polyploidy, clonal propagation, and obligate outcrossing biology (Santantonio et al., 2020). Furthermore, lengthy growth cycles (Batugal and Bourdeix, 2005; Wickramasuriya and Dunwell, 2018) and unpredictable or asynchronous flowering (Amadi et al., 2012; Darkwa et al., 2020) in some tropical crops make breeding intractable. Tropical crops that have not become global commodities are often clonal, and therefore have traits associated with domestication syndrome that are not "fixed" (Denham et al., 2020). The lack of research on many of these crops also impacts the identification of genotypes with more desirable characteristics (Varshney et al., 2012). Outbreeding tropical crops tend to be highly heterozygous (Ceballos et al., 2004; Wickramasuriya and Dunwell, 2018; Darkwa et al., 2020), making inbred and double-haploid lines difficult to generate (Ceballos et al., 2004). The inability to create inbred lines, which are required for straightforward prediction of genetic gain and consistent improvement of breeding material (Bernardo, 2014; Cobb et al., 2019), limits the efficiency of conventional breeding in many tropical crops and often makes knowing the exact genetic basis of a trait difficult. However, genome engineering becomes a very enticing technology for crop improvement when the genetic basis has been identified, often in a tractable model species.

Critical to the development of a gene-editing workflow is understanding the genetic variation that underlies trait expression (**Figure 1A**). Genetic mapping is typically used to identify the relative location of genes associated with specific traits on a chromosome, though the process becomes more complex in polyploid organisms. To date, success in gene editing has focused on traits following qualitative, Mendelian inheritance [i.e., few, large effect gene(s)] since many domestication traits behave in this manner. For example, plant architecture (Li et al., 2016; Lemmon et al., 2018; Zhang et al., 2020), flowering time (Soyk et al., 2017), and some disease resistance traits (Macovei et al., 2018) have been studied in detail in model organisms which identified their genetic bases, making them attractive targets for similar modification in tropical crops. Therefore, the identity of the underlying genetic cause of the trait is key to being able to "introduce" such traits into the crop of interest.

Another important workflow component is an annotated reference genome, which allows for quick identification of target sequences and homologous sites for genome editing (Bortesi and Fischer, 2015). In cases where little to no genomic information is available, information should be extrapolated from existing genomic resources of closely related species, an approach Lemmon et al. (2018) used to target domestication traits in the orphan crop "groundcherry." Though genome size and ploidy level often cause challenges in genome assembly, reference genomes do exist for several tropical crops (Ming et al., 2008; Argout et al., 2011; Prochnik et al., 2012; Xiao et al., 2017; Lantican et al., 2019; Sahu et al., 2020; Wang et al., 2020; Bally et al., 2021; Strijk et al., 2021). Pan-genomes,

if available, offer a more robust view of a cultivar's genetic variation since they consist of a core genome shared by all sequenced individuals, and reveal genetic variations that are present or absent in re-sequenced genomes (Kyriakidou et al., 2018; Marschall et al., 2018).

Design and Optimization of CRISPR-Cas Components

Sequence information about the gene or region in the chromosome associated with the trait of interest is critical in designing the appropriate gRNA respective to the Cas endonuclease (**Figure 1B**). Having the genomic sequence allows for identification of the PAM sequence relevant to the Cas enzyme to be used (i.e., 5′-NGG-3′ for Cas9 (Jinek et al., 2012) or 5′-TTTV for Cas12a (Zetsche et al., 2015), where "N" is any nucleotide and "V" is A, G, or C) and subsequent gRNA design. Several software programs are available that screen for PAM sequences based on the Cas enzyme of interest and suggest guide designs that target the region with a quality score. If a reference genome is available, the software can also identify potential off-target sites (Brazelton et al., 2015; Doench et al., 2016). The guides can be synthesized in-house via *in vitro* transcription (Liang et al., 2017; 2018b) or ordered commercially (Banakar et al., 2019) to test the reliability of their design.

Once obtained, the guides should be assessed for editing efficiency either *in vitro* or *in vivo* prior to tissue culture and transformation. In both instances, the gRNA is incubated with their compatible Cas nuclease to form the RNP complex before testing (Woo et al., 2015; Li et al., 2016; Liang et al., 2017, 2018b; Banakar et al., 2019). The *in vitro* assay is an enzymatic reaction similar to a restriction enzyme digestion, wherein the target sequence has been amplified or cloned and subsequently incubated with the RNP complex to determine target-cleaving efficiency, while the *in vivo* assay typically relies on transfection of protoplasts with the RNP complex. *In vitro* gRNA design assays are a fast, cost-effective, and reliable method for gRNA optimization. Post-transfection, the target region of the genome is amplified and assessed enzymatically (*in vitro* assay) or sequenced [i.e., Sanger, next-generation sequencing (NGS)]], depending on resource availability and short-term goals (Woo et al., 2015). If the *in vitro* assay is performed post-transfection, the absence of cleavage is indicative of high efficiency gRNA design. Sanger sequencing addresses whether an editing event occurred, as well as the types and proportion of alleles produced. The Sanger sequence data can be analyzed by different software such as TIDE (Brinkman et al., 2018). When performed post-transfection, NGS can quantify the different edited alleles produced, and, if a reference genome is available, the relative number of off-target events.

Off-target events are unintended genetic modifications that can arise during gene-editing and are usually due to target sequence similarity (few or no mismatches) (Brazelton et al., 2015; Doench et al., 2016). The risk of off-target editing can be addressed at the gRNA design stage using various genomic strategies that typically depend on the availability of a reference genome (Manghwar et al., 2020). Without a reference genome, off-target editing sites can be unpredictable and limit the widespread application of gene-editing technology for commercial purposes. CIRCLE-seq is an *in vitro* screen for genome-wide off-target cleavage sites which has previously demonstrated potential as a genome-independent method of off-target analysis, though the technology is limited to the CRISPR-Cas9 system (Tsai et al., 2017; Lee et al., 2019). Notably, off-target effects have been observed more frequently in edited plants produced by T-DNA transformation compared to "DNA-free" RNA or RNP methods, wherein the likelihood of undesirable edits dramatically decreases as more mismatches are present between the target sequence and off-target regions (Modrzejewski et al., 2020). Therefore, introducing the CRISPR-Cas system as RNP or RNA may reduce the probability of off-target events.

Tissue Culture

Gene editing reagents need to be introduced into plant cells using tissue culture (**Figure 1C**) and transformation methods (Sandhya et al., 2020). A major challenge for CRISPR-Cas genome editing in tropical crops is often the lack of efficient tissue culture and transformation protocols due to their lengthy generation times (Haque et al., 2018; van Eck, 2018). Development of efficient tissue culture and regeneration protocols will depend on the research group's resources and the crop of interest. Thus far, there are two ways of performing tissue culture and subsequent regeneration post-transformation/transfection: chemical-based and molecular genetic-based.

The most common approach is the chemical-based strategy and is based on testing different ratios of auxin and cytokinin, two important hormones in plant development (Gaba, 2004; Chin and Tan, 2018). In this approach, plant regeneration is achieved through direct or indirect routes. The direct route involves the induction of shoots or roots directly from differentiated explant tissue, resulting in genetically stable clonal plants at a low rate of efficiency. On the other hand, the indirect route involves dedifferentiation of somatic tissue into a callus phase and subsequent production of somatic embryos occurring at a higher efficiency (Chin and Tan, 2018). Genotype-dependency, rate of somaclonal variation, and low rate of plantlet regeneration present challenges for efficient tissue culture and regeneration using this approach (Chin and Tan, 2018).

Molecular genetic-based tissue culture and transformation methods fall broadly into three categories: developmental regulators (DR), morphogenic factors, and Growth-Regulatory plus GRF-Interacting Factors (GRF-GIF). The DR-based strategy involves the expression of meristem-organizing genes, either *in vitro* or ectopically, to generate new shoots that give rise to fertile plantlets from somatic tissue (Maher et al., 2020). This method, however, is not generalizable to all crops or regulatory systems (Nasti and Voytas, 2021). On the other hand, the application of morphogenic factors is genotype-independent and generates a larger number of edited plants in less time (Lowe et al., 2016). Briefly, the overexpression of morphogenic genes, such as *Baby boom* (Boutilier et al., 2002) and *Wuschel* (Arroyo-Herrera et al., 2008), can be used to efficiently induce somatic embryogenesis directly from explant tissue without the need for a callus phase. The GRF-GIF approach uses the overexpression of specific chimeric GRF and GIF fusion proteins (Kong et al., 2020) to produce stable transformants with increased regeneration efficiency. This approach was instrumental in producing edited genotypes that were recalcitrant to previous transformation methods.

Transformation Platform

There are two general approaches to transformation using CRISPR-Cas: transgenesis and transfection (**Figure 1D**). Transgenic techniques introduce exogenous DNA fragments, or transgenes, to target tissue genomes via bombardment or co-cultivation with disarmed *Agrobacterium tumefaciens* strains that insert T-DNA from binary vector systems (Bower and Birch, 1992; Fitch et al., 1993; May et al., 1995; Nasti and Voytas, 2021). The advantage of the transgene approach is that it uses selection to identify events that carry the transgene versus those that do not, facilitating the screening of CRISPR-Cas positive events. However, the integration of transgenes is non-specific and sometimes unstable (Jaganathan et al., 2018). Transfection techniques are possible because the gene editing reagents can be assembled as RNP complexes or RNA molecules *in vitro* (Zhang Y. et al., 2021) and subsequently delivered to embryogenic calli via particle bombardment or introduced in isolated protoplasts via transfection (Liang et al., 2018b). The advantage of using the transfection approach is that the edited events are transgene-free, which may or may not influence regulatory processes required for commercial release of gene edited crops. However, the disadvantage is that hundreds of events must be screened due to the absence of a selectable marker (Fister et al., 2018).

Screening

Once gene-editing reagents have been inserted into plant cells and cellular repair mechanisms have created edits, cells carrying these genetic changes need to be identified, so that regenerated plantlets that are chimeric or homozygous for the desired edit can be recovered (**Figure 1E**). Edited events may be phenotypically screened and subsequently verified through sequencing, or vice versa depending on the trait of interest (Kaur et al., 2018). Many methods are available (Slatko et al., 2018), but large-scale editing projects typically require a high-throughput approach. Recent advancements such as PacBio technology and Nanopore sequencing can generate long DNA (and RNA) reads at unprecedented volume, in contrast to first and second-generation sequencing which mostly produce short-read sequences (Slatko et al., 2018). The drawback of high-throughput approaches is their relatively high error rate, though the large computational capacity circumvents this problem. In the transgenic approach, selectable markers, such as antibiotic (Hardegger and Sturm, 1998) or herbicide resistance (Zhao et al., 2000), have been essential in identifying and propagating edited plants. However, concern over the presence of transgenes in the final product and the prospects of successive rounds of transformation and gene silencing have made marker elimination a more attractive approach (Veluthambi et al., 2003). Elimination of transgenes, T-DNA, and selectable markers depends on the efficiency of progeny regeneration and the ability to segregate the T-DNA from the edited allele through crosses (Russell et al., 1992; Hohn et al., 2001; Veillet et al., 2019). Furthermore, edited lines should be assessed in various field environments to ensure that the desired phenotype is heritable and stable for commercial application. Breeders must consider gene-environment interactions (GEI) and trait stability to recommend ideal varieties to growers and maximize success (Huang et al., 2016).

In the case of vegetatively propagated crops, or where traditional breeding is cumbersome or not possible, screening must be approached differently. Transgene-free methods remove the need to segregate transgenic material out of desired lines and still retain the edited event but rely on genomic analyses for selection. Briefly, polymerase chain reaction (PCR)-based and DNA sequencing-based methods are highly sensitive and specific, provided an adequate reference is available (Grohmann et al., 2019). Although limited in their application, alternative approaches include DNA hybridization assays, protein- and metabolite-based methods (Grohmann et al., 2019), and restriction enzyme assays (Kim et al., 2014; Shan et al., 2014; Vouillot et al., 2015) combined with bioinformatic analysis tools (Liang et al., 2018a). Probe-based quantitative PCR (qPCR) is a simple, robust, and rapid approach that can be applied to a broad range of genotypes (Falabella et al., 2017; Peng et al., 2018) and the use of a locked nucleic acid (LNA) probe may offer increased specificity by reducing off-target amplification (Zhang H. et al., 2021). We suggest reviewing Grohmann et al., 2019 for an overview of screening and selection methods to aid in protocol development.

CONCLUSION

The promise in genome engineering technology is clear, especially in under-resourced regional tropical crops. Here, we have outlined a clear workflow to operationalize CRISPR-Cas technology in any species of interest, though it is important to understand the different cultural relevance of underutilized crops to appropriately develop resources (Gordon et al., 2021). The processes outlined here can be developed independently at different times, but all need to be in place in order to establish a CRISPR-Cas gene editing platform for improvement of your target crop.

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

Conceptualization: KJ and AD. Drafting: KJ and AD. Figure Creation: KJ. Supervision: MM and AD. Writing and Reviewing: KJ, MM, MK, M-LW, XH, and AD. Funding acquisition: M-LW, XH, and AD.

FUNDING

This work was supported by a United States Department of Agriculture (USDA) National Institutes of Food and Agriculture (NIFA) Agriculture through Gene Editing Grant 2021-67013-34545 to AD. The funder had no role in the development of this work.

REFERENCES

Amadi, C. O., Mbanaso, E. N. A., and Chukwu, G. O. (2012). A Review Cocoyam Breeding in Nigeria: Achievements, Challenges and Prospects. Nigeria Agric. J. 43. , 2012 Available at: https://www.ajol.info/index.php/naj/article/view/110109 (Accessed September 15, 2021)

Anzalone, A. v., Koblan, L. W., and Liu, D. R. (2020). Genome Editing with CRISPR-Cas Nucleases, Base Editors, Transposases and Prime Editors. *Nat. Biotechnol.* 38, 824–844. doi:10.1038/s41587-020-0561-9

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M. J., Droc, G., Gouzy, J., et al. (2011). The Genome of Theobroma Cacao. *Nat. Genet.* 43, 101–108. doi:10.1038/ng.736

Arroyo-Herrera, A., Ku Gonzalez, A., Canche Moo, R., Quiroz-Figueroa, F. R., Loyola-Vargas, V. M., Rodriguez-Zapata, L. C., et al. (2008). Expression of WUSCHEL in Coffea Canephora Causes Ectopic Morphogenesis and Increases Somatic Embryogenesis. *Plant Cel Tiss Organ. Cult* 94, 171–180. doi:10.1007/s11240-008-9401-1

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

Bally, I. S. E., Bombarely, A., Bombarely, A., Chambers, A. H., Cohen, Y., Dillon, N. L., et al. (2021). The 'Tommy Atkins' Mango Genome Reveals Candidate Genes for Fruit Quality. *BMC Plant Biol.* 21. doi:10.1186/s12870-021-02858-1

Balogun, M. O. (2009). Microtubers in Yam Germplasm Conservation and Propagation : The Status , the Prospects and the Constraints. *Biotechnol. Mol. Biol. Rev.* 4

Banakar, R., Eggenberger, A. L., Lee, K., Wright, D. A., Murugan, K., Zarecor, S., et al. (2019). High-frequency Random DNA Insertions upon Co-delivery of CRISPR-Cas9 Ribonucleoprotein and Selectable Marker Plasmid in rice. *Sci. Rep.* 9. doi:10.1038/s41598-019-55681-y

Batte, M., Swennen, R., Uwimana, B., Akech, V., Brown, A., Tumuhimbise, R., et al. (2019). Crossbreeding East African highland Bananas: Lessons Learnt Relevant to the Botany of the Crop after 21 Years of Genetic Enhancement. *Front. Plant Sci.* 10. doi:10.3389/fpls.2019.00081

Batugal, P., and Bourdeix, R. (2005). Conventional Coconut Breeding. *Coconut Genet. Resour.*, 327–375.

Bernardo, R. (2014). Genomewide Selection when Major Genes Are Known. *Crop Sci.* 54, 68–75. doi:10.2135/cropsci2013.05.0315

Bernardo, R. (2020). Reinventing Quantitative Genetics for Plant Breeding: Something Old, Something New, Something Borrowed, Something BLUE. *Heredity* 125, 375–385. doi:10.1038/s41437-020-0312-1

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 System for Plant Genome Editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi:10.1016/j.biotechadv.2014.12.006

Bourdeix, R., Adkins, S., Johnson, V., Perera, L., and Sisunandar (2020). "*In Situ* and *Ex Situ* Conservation of Coconut Genetic Resources," In *Coconut Biotechnology: Towards the Sustainability Of the "Tree Of Life"*. Springer, 51–75. doi:10.1007/978-3-030-44988-9_4

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., et al. (2002). Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth. *Plant Cell* 14, 1737–1749. doi:10.1105/tpc.001941

Bower, R., and Birch, R. G. (1992). Transgenic Sugarcane Plants via Microprojectile Bombardment. *Plant J.* 2, 409–416. doi:10.1111/j.1365-313X.1992.00409.x

Brazelton, V. A., Zarecor, S., Wright, D. A., Wang, Y., Liu, J., Chen, K., et al. (2015). A Quick Guide to CRISPR sgRNA Design Tools. *GM Crops & Food* 6, 266–276. doi:10.1080/21645698.2015.1137690

Brinkman, E. K., Kousholt, A. N., Harmsen, T., Leemans, C., Chen, T., Jonkers, J., et al. (2018). Easy Quantification of Template-Directed CRISPR/Cas9 Editing. *Nucleic Acids Res.* 46, e58. doi:10.1093/nar/gky164

Bull, S. E., Seung, D., Chanez, C., Mehta, D., Kuon, J.-E., Truernit, E., et al. (2018). Accelerated *Ex Situ* Breeding of GBSS - and PTST1 -edited Cassava for Modified Starch. *Sci. Adv.* 4. doi:10.1126/sciadv.aat6086

Ceballos, H., Iglesias, C. A., Pérez, J. C., and Dixon, A. G. O. (2004). Cassava Breeding: Opportunities and Challenges. *Plant Mol. Biol.* 56, 503–516. doi:10.1007/s11103-004-5010-5

Chin, C., and Tan, H. (2018). The Use of Proteomic Tools to Address Challenges Faced in Clonal Propagation of Tropical Crops through Somatic Embryogenesis. *Proteomes* 6, 21. doi:10.3390/PROTEOMES6020021

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., et al. (2019). Enhancing the Rate of Genetic Gain in Public-Sector Plant Breeding Programs: Lessons from the Breeder's Equation. *Theor. Appl. Genet.* 132, 627–645. doi:10.1007/s00122-019-03317-0

Darkwa, K., Olasanmi, B., Asiedu, R., and Asfaw, A. (2020). Review of Empirical and Emerging Breeding Methods and Tools for Yam (Dioscorea spp.) Improvement: Status and Prospects. *Plant Breed* 139, 474–497. doi:10.1111/pbr.12783

Denham, T., Barton, H., Castillo, C., Crowther, A., Dotte-Sarout, E., Florin, S. A., et al. (2020). The Domestication Syndrome in Vegetatively Propagated Field Crops. *Ann. Bot.* 125, 581–597. doi:10.1093/aob/mcz212

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi:10.1038/nbt.3437

Eid, A., Mohan, C., Sanchez, S., Wang, D., and Altpeter, F. (2021). Multiallelic, Targeted Mutagenesis of Magnesium Chelatase with CRISPR/Cas9 Provides a Rapidly Scorable Phenotype in Highly Polyploid Sugarcane. *Front. Genome Ed.* 3. doi:10.3389/fgeed.2021.654996

Falabella, M., Sun, L., Barr, J., Pena, A. Z., Kershaw, E. E., Gingras, S., et al. (2017). Single-step qPCR and dPCR Detection of Diverse CRISPR-Cas9 Gene Editing Events *In Vivo*. *G3: Genes, Genomes, Genet.* 7, 3533–3542. doi:10.1534/g3.117.300123

Fan, Y., Xin, S., Dai, X., Yang, X., Huang, H., and Hua, Y. (2020). Efficient Genome Editing of Rubber Tree (Hevea Brasiliensis) Protoplasts Using CRISPR/Cas9 Ribonucleoproteins. *Ind. Crops Prod.* 146, 112146. doi:10.1016/j.indcrop.2020.112146

FAO (2010). *The Contribution of Plant Genetic Resources for Food and Agriculture to Food Security and Sustainable Agricultural developmentThe Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome, Italy: FAO, 399

Fister, A. S., Landherr, L., Maximova, S. N., and Guiltinan, M. J. (2018). Transient Expression of CRISPR/Cas9 Machinery Targeting TcNPR3 Enhances Defense Response in Theobroma Cacao. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00268

Fitch, M. M., Manshardt, R., Gonsalves, D., and Slightom, J. (1993). Transgenic Papaya Plants from Agrobacterium-Mediated Transformation of Somatic Embryos. *Plant Cel Rep.* 12. doi:10.1007/BF00237128

Gaba, V. (2004). Plant Growth Regulators in Plant Tissue Culture and Development. *Plant Development Biotechnol.*, 87–99. doi:10.1201/9780203506561.ch8

Gaba, V., and Singer, S. (2009). Propagation of Sweetpotatoes, *In Situ* Germplasm Conservation and Conservation by Tissue Culture. *The Sweetpotato*, 65–80. doi:10.1007/978-1-4020-9475-0_6

Gomez, M. A., Lin, Z. D., Moll, T., Chauhan, R. D., Renninger, K., Beyene, G., et al. (2017). Simultaneous CRISPR/Cas9-Mediated Editing of Cassava eIF4E Isoforms nCBP-1 and nCBP-2 Reduces Cassava Brown Streak Disease Symptom Severity and Incidence, *Plant Biotechnol. J.*, 17(2):421–434. doi:10.1101/209874

Gordon, D. R., Jaffe, G., Doane, M., Glaser, A., Gremillion, T. M., and Ho, M. D. (2021). Responsible Governance of Gene Editing in Agriculture and the Environment. *Nat. Biotechnol.* 39, 1055–1057. doi:10.1038/s41587-021-01023-1

Grohmann, L., Keilwagen, J., Duensing, N., Dagand, E., Hartung, F., Wilhelm, R., et al. (2019). Detection and Identification of Genome Editing in Plants: Challenges and Opportunities. *Front. Plant Sci.* 10. doi:10.3389/fpls.2019.00236

Haque, E., Taniguchi, H., Hassan, M. M., Bhowmik, P., Karim, M. R., Śmiech, M., et al. (2018). Application of CRISPR/Cas9 Genome Editing Technology for the Improvement of Crops Cultivated in Tropical Climates: Recent Progress, Prospects, and Challenges. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00617

Hardegger, M., and Sturm, A. (1998). Transformation and Regeneration of Carrot (Daucus Carota L.). *Mol. Breed.* 4, 119–127. doi:10.1023/A:1009681725540

Hohn, B., Levy, A. A., and Puchta, H. (2001). Elimination of Selection Markers from Transgenic Plants. *Curr. Opin. Biotechnol.* 12, 139–143. doi:10.1016/S0958-1669(00)00188-9

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., van Sanford, D., Costa, J., et al. (2016). Genomic Selection for Wheat Traits and Trait Stability. *Theor. Appl. Genet.* 129, 1697–1710. doi:10.1007/s00122-016-2733-z

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for Crop Improvement: An Update Review. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.00985

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821. doi:10.1126/science.1225829

Jung, C., and Till, B. (2021). Mutagenesis and Genome Editing in Crop Improvement: Perspectives for the Global Regulatory Landscape. *Trends Plant Sci.*. doi:10.1016/j.tplants.2021.08.002

Kallow, S., Longin, K., Sleziak, N. F., Janssens, S. B., Vandelook, F., Dickie, J., et al. (2020). Challenges for *Ex Situ* Conservation of Wild Bananas: Seeds Collected in papua new guinea Have Variable Levels of Desiccation Tolerance. *Plants* 9, 1243. doi:10.3390/plants9091243

Kaur, N., Alok, A., ShivaniKaur, N., Kaur, N., Pandey, P., Awasthi, P., et al. (2018). CRISPR/Cas9-mediated Efficient Editing in Phytoene Desaturase (PDS) Demonstrates Precise Manipulation in Banana Cv. Rasthali Genome. *Funct. Integr. Genomics* 18, 89–99. doi:10.1007/s10142-017-0577-5

Kaur, N., Alok, A., ShivaniKumar, P., Kumar, P., Kaur, N., Awasthi, P., et al. (2020). CRISPR/Cas9 Directed Editing of Lycopene Epsilon-Cyclase Modulates Metabolic Flux for β-carotene Biosynthesis in Banana Fruit. *Metab. Eng.* 59, 76–86. doi:10.1016/j.ymben.2020.01.008

Kim, J. M., Kim, D., Kim, S., and Kim, J.-S. (2014). Genotyping with CRISPR-Cas-Derived RNA-Guided Endonucleases. *Nat. Commun.* 5. doi:10.1038/ncomms4157

Kong, J., Martin-Ortigosa, S., Finer, J., Orchard, N., Gunadi, A., Batts, L. A., et al. (2020). Overexpression of the Transcription Factor GROWTH-REGULATING FACTOR5 Improves Transformation of Dicot and Monocot Species. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.572319

Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Strömvik, M. v. (2018). Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01660

Lantican, D. v., Strickler, S. R., Canama, A. O., Gardoce, R. R., Mueller, L. A., and Galvez, H. F. (2019). De Novo Genome Sequence Assembly of Dwarf Coconut (Cocos Nucifera L. 'Catigan Green Dwarf') Provides Insights into Genomic Variation between Coconut Types and Related Palm Species. *G3: Genes, Genomes, Genet.* 9, 2377–2393. doi:10.1534/g3.119.400215

Lee, K., Zhang, Y., Kleinstiver, B. P., Guo, J. A., Aryee, M. J., Miller, J., et al. (2019). Activities and Specificities of CRISPR/Cas9 and Cas12a Nucleases for Targeted Mutagenesis in maize. *Plant Biotechnol. J.* 17, 362–372. doi:10.1111/pbi.12982

Lemmon, Z. H., Reem, N. T., Dalrymple, J., Soyk, S., Swartwood, K. E., Rodriguez-Leal, D., et al. (2018). Rapid Improvement of Domestication Traits in an Orphan Crop by Genome Editing. *Nat. Plants* 4, 766–770. doi:10.1038/s41477-018-0259-x

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., et al. (2016). Reassessment of the Four Yield-Related Genes Gn1a, DEP1, GS3, and IPA1 in rice Using a CRISPR/Cas9 System. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.00377

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., et al. (2017). Efficient DNA-free Genome Editing of Bread Wheat Using CRISPR/Cas9 Ribonucleoprotein Complexes. *Nat. Commun.* 8. doi:10.1038/ncomms14261

Liang, Z., Chen, K., Yan, Y., Zhang, Y., and Gao, C. (2018a). Genotyping Genome-Edited Mutations in Plants Using CRISPR Ribonucleoprotein Complexes. *Plant Biotechnol. J.* 16, 2053–2062. doi:10.1111/pbi.12938

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J.-L., et al. (2018b). Genome Editing of Bread Wheat Using Biolistic Delivery of CRISPR/Cas9 *In Vitro* Transcripts or Ribonucleoproteins. *Nat. Protoc.* 13, 413–430. doi:10.1038/nprot.2017.145

Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M.-J., et al. (2016). Morphogenic Regulators Baby Boom and Wuschel Improve Monocot Transformation. *Plant Cell* 28, 1998–2015. doi:10.1105/tpc.16.00124

Macovei, A., Sevilla, N. R., Cantos, C., Jonson, G. B., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel Alleles of rice eIF4G Generated by CRISPR/Cas9-targeted Mutagenesis Confer Resistance to Rice Tungro Spherical Virus. *Plant Biotechnol. J.* 16, 1918–1927. doi:10.1111/pbi.12927

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

Manghwar, H., Li, B., Ding, X., Hussain, A., Lindsey, K., Zhang, X., et al. (2020). CRISPR/Cas Systems in Genome Editing: Methodologies and Tools for sgRNA Design, Off-Target Evaluation, and Strategies to Mitigate Off-Target Effects. *Adv. Sci.* 7, 1902312. doi:10.1002/advs.201902312

Marschall, T., Marz, M., Abeel, T., Dijkstra, L., Dutilh, B. E., Ghaffaari, A., et al. (2018). Computational Pan-Genomics: Status, Promises and Challenges. *Brief Bioinform* 19, bbw089. doi:10.1093/bib/bbw089

May, G. D., Afza, R., Mason, H. S., Wiecko, A., Novak, F. J., and Arntzen, C. J. (1995). Generation of Transgenic Banana (Musa Acuminata) Plants via Agrobacterium-Mediated Transformation. *Nat. Biotechnol.* 13, 486–492. doi:10.1038/nbt0595-486

Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S.-e. -A., Hirsch-Hoffmann, M., Gruissem, W., et al. (2019). Linking CRISPR-Cas9 Interference in Cassava to the Evolution of Editing-Resistant Geminiviruses. *Genome Biol.* 20. doi:10.1186/s13059-019-1678-3

Mertens, A., Swennen, R., Rønsted, N., Vandelook, F., Panis, B., Sachter-Smith, G., et al. (2021). Conservation Status Assessment of Banana Crop Wild Relatives Using Species Distribution Modelling. *Divers. Distrib* 27, 729–746. doi:10.1111/ddi.13233

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., et al. (2008). The Draft Genome of the Transgenic Tropical Fruit Tree Papaya (Carica Papaya Linnaeus). *Nature* 452, 991–996. doi:10.1038/nature06856

Modrzejewski, D., Hartung, F., Lehnert, H., Sprink, T., Kohl, C., Keilwagen, J., et al. (2020). Which Factors Affect the Occurrence of Off-Target Effects Caused by the Use of CRISPR/Cas: A Systematic Review in Plants. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.574959

Naim, F., Dugdale, B., Kleidon, J., Brinin, A., Shand, K., Waterhouse, P., et al. (2018). Gene Editing the Phytoene Desaturase Alleles of Cavendish Banana Using CRISPR/Cas9. *Transgenic Res.* 27, 451–460. doi:10.1007/s11248-018-0083-0

Nasti, R. A., and Voytas, D. F. (2021). Attaining the Promise of Plant Gene Editing at Scale. *Proc. Natl. Acad. Sci. USA* 118, e2004846117. doi:10.1073/pnas.2004846117

Ntui, V. O., Tripathi, J. N., and Tripathi, L. (2020). Robust CRISPR/Cas9 Mediated Genome Editing Tool for Banana and Plantain (Musa spp.). *Curr. Plant Biol.* 21, 100128. doi:10.1016/j.cpb.2019.100128

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D. A., Bart, R., and Taylor, N. J. (2017). Efficient CRISPR/cas9 Genome Editing of Phytoene Desaturase in Cassava. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.01780

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

Panis, B., Nagel, M., Van den houwe, I., and van, I. (2020). Challenges and Prospects for the Conservation of Crop Genetic Resources in Field Genebanks, in *In Vitro* Collections And/or in Liquid Nitrogen. *Plants* 9, 1634. doi:10.3390/plants9121634

Peng, C., Wang, H., Xu, X., Wang, X., Chen, X., Wei, W., et al. (2018). High-throughput Detection and Screening of Plants Modified by Gene Editing Using Quantitative Real-Time Polymerase Chain Reaction. *Plant J.* 95, 557–567. doi:10.1111/tpj.13961

Prochnik, S., Marri, P. R., Desany, B., Rabinowicz, P. D., Kodira, C., Mohiuddin, M., et al. (2012). The Cassava Genome: Current Progress, Future Directions. *Trop. Plant Biol.* 5, 88–94. doi:10.1007/s12042-011-9088-z

Ramstein, G. P., Jensen, S. E., and Buckler, E. S. (2019). Breaking the Curse of Dimensionality to Identify Causal Variants in Breeding 4. *Theor. Appl. Genet.* 132, 559–567. doi:10.1007/s00122-018-3267-3

Russell, S. H., Hoopes, J. L., and Odell, J. T. (1992). Directed Excision of a Transgene from the Plant Genome. *Mol. Gen. Genet.* 234, 49–59. doi:10.1007/BF00272344

Sahu, S., Liu, M., Yssel, A., Kariba, R., Muthemba, S., Jiang, S., et al. (2020). Draft Genomes of Two Artocarpus Plants, Jackfruit (A. Heterophyllus) and Breadfruit (A. Altilis). *Genes* 11, 27. doi:10.3390/genes11010027

Sandhya, D., Jogam, P., Allini, V. R., Abbagani, S., and Alok, A. (2020). The Present and Potential Future Methods for Delivering CRISPR/Cas9 Components in Plants. *J. Genet. Eng. Biotechnol.* 18. doi:10.1186/s43141-020-00036-8

Santantonio, N., Atanda, S. A., Beyene, Y., Varshney, R. K., Olsen, M., Jones, E., et al. (2020). Strategies for Effective Use of Genomic Information in Crop Breeding Programs Serving Africa and South Asia. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.00353

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome Editing in rice and Wheat Using the CRISPR/Cas System. *Nat. Protoc.* 9, 2395–2410. doi:10.1038/nprot.2014.157

Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* 122. doi:10.1002/cpmb.59

Soyk, S., Lemmon, Z. H., Sedlazeck, F. J., Jiménez-Gómez, J. M., Alonge, M., Hutton, S. F., et al. (2019). Duplication of a Domestication Locus Neutralized a Cryptic Variant that Caused a Breeding Barrier in Tomato. *Nat. Plants* 5, 471–479. doi:10.1038/s41477-019-0422-z

Soyk, S., Müller, N. A., Park, S. J., Schmalenbach, I., Jiang, K., Hayama, R., et al. (2017). Variation in the Flowering Gene SELF PRUNING 5G Promotes Day-Neutrality and Early Yield in Tomato. *Nat. Genet.* 49, 162–168. doi:10.1038/ng.3733

Strijk, J. S., Hinsinger, D. D., Roeder, M. M., Chatrou, L. W., Couvreur, T. L. P., Erkens, R. H. J., et al. (2021). Chromosome-level Reference Genome of the Soursop (Annona Muricata): A New Resource for Magnoliid Research and Tropical Pomology. *Mol. Ecol. Resour.* 21, 1608–1619. doi:10.1111/1755-0998.13353

Syombua, E. D., Zhang, Z., Tripathi, J. N., Ntui, V. O., Kang, M., George, O. O., et al. (2021). A CRISPR/Cas9-based Genome-editing System for Yam (Dioscorea spp.). *Plant Biotechnol. J.* 19, 645–647. doi:10.1111/pbi.13515

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., et al. (2018). Engineering Herbicide-Resistant Watermelon Variety through CRISPR/Cas9-mediated Base-Editing. *Plant Cel Rep* 37, 1353–1356. doi:10.1007/s00299-018-2299-0

Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. v., Aryee, M. J., and Joung, J. K. (2017). CIRCLE-seq: A Highly Sensitive *In Vitro* Screen for Genome-wide CRISPR-Cas9 Nuclease Off-Targets. *Nat. Methods* 14, 607–614. doi:10.1038/nmeth.4278

van Eck, J. (2018). Genome Editing and Plant Transformation of Solanaceous Food Crops. *Curr. Opin. Biotechnol.* 49, 35–41. doi:10.1016/j.copbio.2017.07.012

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, 1172–1176. doi:10.1038/nbt.2440

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.-P., Guyon-Debast, A., Chauvin, J.-E., et al. (2019). Transgene-free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. *Ijms* 20, 402. doi:10.3390/ijms20020402

Veluthambi, K., Gupta, A. K., and Sharma, A. (2003). The Current Status of Plant Transformation Technologies. *Curr. Sci.* 84

Vouillot, L., Thélie, A., and Pollet, N. (2015). Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3: Genes, Genomes, Genet.* 5, 407–415. doi:10.1534/g3.114.015834

Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., et al. (2020). The Genome Evolution and Domestication of Tropical Fruit Mango. *Genome Biol.* 21. doi:10.1186/s13059-020-01959-8

Wickramasuriya, A. M., and Dunwell, J. M. (2018). Cacao Biotechnology: Current Status and Future Prospects. *Plant Biotechnol. J.* 16, 4–17. doi:10.1111/pbi.12848

Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., et al. (2015). DNA-free Genome Editing in Plants with Preassembled CRISPR-Cas9 Ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi:10.1038/nbt.3389

Xiao, Y., Xu, P., Fan, H., Baudouin, L., Xia, W., Bocs, S., et al. (2017). The Genome Draft of Coconut (Cocos Nucifera). *GigaScience* 6. doi:10.1093/gigascience/gix095

Yoon, Y.-J., Venkatesh, J., Lee, J.-H., Kim, J., Lee, H.-E., Kim, D.-S., et al. (2020). Genome Editing of eIF4E1 in Tomato Confers Resistance to Pepper Mottle Virus. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.01098

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, 759–771. doi:10.1016/j.cell.2015.09.038

Zhang, H., Li, J., Zhao, S., Yan, X., Si, N., Gao, H., et al. (2021a). An Editing-site-specific Pcr Method for Detection and Quantification of Cao1-Edited rice. *Foods* 10, 1209. doi:10.3390/foods10061209

Zhang, J., Zhang, X., Chen, R., Yang, L., Fan, K., Liu, Y., et al. (2020). Generation of Transgene-free Semidwarf Maize Plants by Gene Editing of Gibberellin-Oxidase20-3 Using CRISPR/Cas9. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.01048

Zhang, Y., Iaffaldano, B., and Qi, Y. (2021b). CRISPR Ribonucleoprotein-Mediated Genetic Engineering in Plants. *Plant Commun.* 2, 100168. doi:10.1016/j.xplc.2021.100168

Zhao, Y., Karan, R., and Altpeter, F. (2021). Error-free Recombination in Sugarcane Mediated by Only 30 Nucleotides of Homology and CRISPR/Cas9 Induced DNA Breaks or Cre-recombinase. *Biotechnol. J.* 16, 2000650. doi:10.1002/biot.202000650

Zhao, Z.-y., Cai, T., Tagliani, L., Miller, M., Wang, N., Pang, H., et al. (2000). Agrobacterium-mediated Sorghum Transformation. *Plant Mol. Biol.* 44, 789–798. doi:10.1023/A:1026507517182

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

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 November 2021
doi: 10.3389/fgene.2021.794297

Comprehensive *In Silico* Characterization and Expression Profiling of *TCP* Gene Family in Rapeseed

Yunfei Wen[1,2†], *Ali Raza*[2,3†], *Wen Chu*[2†], *Xiling Zou*[2], *Hongtao Cheng*[2], *Qiong Hu*[2], *Jia Liu*[2*] *and Wenliang Wei*[1*]

[1]College of Agriculture, Yangtze University, Jingzhou, China, [2]Key Laboratory for Biological Sciences and Genetic Improvement of Oil Crops, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Ministry of Agriculture and Rural Affairs, Wuhan, China, [3]Fujian Provincial Key Laboratory of Crop Molecular and Cell Biology, Center of Legume Crop Genetics and Systems Biology/College of Agriculture, Oil Crops Research Institute, Fujian Agriculture and Forestry University (FAFU), Fuzhou, China

TCP proteins are plant-specific transcription factors that have multipurpose roles in plant developmental procedures and stress responses. Therefore, a genome-wide analysis was performed to categorize the *TCP* genes in the rapeseed genome. In this study, a total of 80 *BnTCP* genes were identified in the rapeseed genome and grouped into two main classes (PCF and CYC/TB1) according to phylogenetic analysis. The universal evolutionary analysis uncovered that *BnTCP* genes had experienced segmental duplications and positive selection pressure. Gene structure and conserved motif examination presented that Class I and Class II have diverse intron-exon patterns and motifs numbers. Overall, nine conserved motifs were identified and varied from 2 to 7 in all *TCP* genes; and some of them were gene-specific. Mainly, Class II (PCF and CYC/TB1) possessed diverse structures compared to Class I. We identified four hormone- and four stress-related responsive *cis*-elements in the promoter regions. Moreover, 32 bna-miRNAs from 14 families were found to be targeting 21 *BnTCPs* genes. Gene ontology enrichment analysis presented that the *BnTCP* genes were primarily related to RNA/DNA binding, metabolic processes, transcriptional regulatory activities, etc. Transcriptome-based tissue-specific expression analysis showed that only a few genes (mainly *BnTCP9, BnTCP22, BnTCP25, BnTCP48, BnTCP52, BnTCP60, BnTCP66,* and *BnTCP74*) presented higher expression in root, stem, leaf, flower, seeds, and silique among all tested tissues. Likewise, qRT-PCR-based expression analysis exhibited that *BnTCP36, BnTCP39, BnTCP53, BnTCP59,* and *BnTCP60* showed higher expression at certain time points under various hormones and abiotic stress conditions but not by drought and MeJA. Our results opened the new groundwork for future understanding of the intricate mechanisms of *BnTCP* in various developmental processes and abiotic stress signaling pathways in rapeseed.

Keywords: abiotic stress, *cis*-elements, evolution, genomics, miRNA, phytohomones, segmental duplication

OPEN ACCESS

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

Reviewed by:
Muhammad Azhar Nadeem, Sivas University of Science and Technology, Turkey
Manu Kumar, Dongguk University Seoul, South Korea
Muhammad Waseem, South China Agricultural University, China
Muhammad Azhar Hussain, Hainan University, China

***Correspondence:**
Jia Liu
liujia02@caas.cn
Wenliang Wei
whwenliang@163.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: *13 October 2021*
Accepted: *01 November 2021*
Published: *17 November 2021*

Citation:
Wen Y, Raza A, Chu W, Zou X, Cheng H, Hu Q, Liu J and Wei W (2021) Comprehensive In Silico Characterization and Expression Profiling of TCP Gene Family in Rapeseed. Front. Genet. 12:794297. doi: 10.3389/fgene.2021.794297

INTRODUCTION

Transcription factors (TFs) contribute to plant growth and development, regulate gene expression, and play a significant role in several cellular/biological processes (Hoang et al., 2017; Karaaslan et al., 2020). The Teosinte Branched 1, Cycloidea, and Proliferating Cell Factors (*TCP*) gene family is a group of plant-specific TFs limited to higher plants (Cubas et al., 1999). These TFs have numerous roles in controlling plant growth and developmental procedures by directing cell proliferation (Cubas et al., 1999; Martín-Trillo and Cubas, 2010). Notably, these TFs are categorized by an extremely conserved 59-amino-acid basic helix–loop–helix (bHLH) motif at the N-terminus nominated as the TCP domain (Aggarwal et al., 2010). This domain is accountable for nuclear targeting, DNA binding and also contributes to protein-protein interactions (Kosugi and Ohashi, 2002). Due to the dissimilarity in the TCP domain, *TCP* gene family members are grouped into two main classes, i.e., Class I (PCF) and Class II (CYC/TB1 and CIN sub-classes) (Navaud et al., 2007; Viola et al., 2012). Along with the TCP domain, several members of Class II retain 18–20-residue arginine-rich motifs (Viola et al., 2012; Sarvepalli and Nath, 2018). This supposed R domain was anticipated to develop a hydrophilic-helix or α-coiled-coil structure that facilitates the protein-protein interactions in plants (Cubas et al., 1999).

Over the past few years, *TCP* gene families have been reported in several plant species. For example, 16 *TCP* members in Moso bamboo (*Phyllostachys edulis*) (Liu et al., 2018); 18 *TCP* members in sweet potato (*Ipomoea batatas* L.) (Ren et al., 2021); 23 *TCP* members in orchids (*Phalaenopsis equestris*) (Lin et al., 2016); 24 *TCP* members in *Arabidopsis thaliana* (Yao et al., 2007); 27 *TCP* members in cucumber (*Cucumis sativus* L.) (Wen et al., 2020); 28 *TCP* members in rice (*Oryza sativa* L.) (Yao et al., 2007); 38 *TCP* members in cotton (*Gossypium raimondii* L.) (Ma et al., 2014); 39 *TCP* members in turnips (*Brassica rapa* ssp. rapa) (Du et al., 2017); 63 *TCP* members in *Brassica juncea* var. tumida (He et al., 2020), 73 *TCP* members in allotetraploid cotton (*Gossypium hirsutum* L.) (Yin et al., 2018), etc. Several studies have shown that *TCP* genes play significant roles in plant growth and developmental processes, including trichome formation (Wang et al., 2013), seed germination (Zhang et al., 2019), petal development (van Es et al., 2018), floral development (Li et al., 2019), shoot branching (Braun et al., 2012), and leaf anatomical morphology (Ren et al., 2021). These findings indicate the diverse role of *TCP* genes in regulating plant developmental processes.

Likewise, several *TCP* genes expression is also regulated by various hormones and abiotic stress treatments. For instance, in cotton, the *GhTCP14* gene is considered a vital switch in an auxin-facilitated extension of fiber cells (Wang et al., 2013). In *A. thaliana*, the expression level of *AtTCP15* was induced by gibberellic acid (Gastaldi et al., 2020), and *AtTCP1* was positively regulated by brassinosteroids treatment (Gao et al., 2015). Furthermore, the different *TCP* genes also participate in ethylene, strigolactone, and cytokinins signaling pathways (Braun et al., 2012; van Es et al., 2018). Under abiotic stress conditions, the expression level of *AtTCP15* was induced by heat stress (Wu et al., 2016), and *TCP20/22* was positively regulated by the circadian clock in *A. thaliana* (Wu et al., 2016). In cotton, several *GhTCP* genes were significantly up-regulated under heat, salt, and drought stresses (Yin et al., 2018). In cucumber, many *CsTCP* genes significantly responded to temperature and photoperiod and were also induced by gibberellic acid and ethylene treatments (Wen et al., 2020). *TCP* genes are also expressed in different tissues, such as many *TCP* genes mainly expressed in flower, leaf, and stem, flower, bud differentiation, and gametophyte development in *Prunus mume* (Zhou et al., 2016).

Rapeseed (*Brassica napus* L.) is the second most crucial oilseed crop and possesses a complex genome. However, its growth and productivity are significantly affected at physiological, biochemical, and molecular levels in response to various plant hormones and abiotic stress conditions (Raza et al., 2021a; He et al., 2021; Raza, 2021; Su et al., 2021). To date, the *TCP* gene family and key *TCP* genes responding to abiotic stress have not been fully characterized in rapeseed. Consequently, this is the first genome-wide study to identify the *TCP* genes in the rapeseed genome. To boost our understanding into *TCP* genes evolution in rapeseed, their phylogenetic relationships, synteny study, gene structures, conserved motifs, *cis*-elements, miRNA predictions, and functional annotation were evaluated. Additionally, their expression profile in numerous tissues/organs and under various phytohormones and abiotic stress conditions were largely appraised, which deeply improved our understanding of the *TCP* genes in rapeseed.

MATERIAL AND METHODS

Identification and Characterization of *TCP* Genes in Rapeseed

As described in our recent studies (Raza et al., 2021a; Su et al., 2021), two methods were used to recognize *TCP* genes in the *B. napus* genome, i.e., BLASTP and the Hidden Markov Model (HMM). The genome of *B. napus* (ZhongShuang 11; ZS11 variety) was downloaded from the BnPIR database (http://cbi.hzau.edu.cn/bnapus/index.php). For BLASTP, the amino acid sequences of 24 *AtTCPs* were used as a query with an e-value set to $1e^{-5}$. The amino acid sequences of 24 *AtTCPs* were attained from the TAIR *Arabidopsis* genome database (http://www.arabidopsis.org/). Secondly, the local software HMMER 3.1 (http://www.hmmer.org/) was used to search the *TCP* genes with default constraints. Then, the HMM file of the TCP domain (PF03634) was downloaded from the Pfam protein domain database (http://pfam.xfam.org/). Finally, 80 *BnTCP* genes were found by merging the two processes in the rapeseed genome. Likewise, *TCP* genes were also identified in *Brassica rapa* and *Brassica oleracea* genomes using the same method. Their genome sequences were downloaded from JGI Phytozome 12.0 database (https://phytozome.jgi.doe.gov/pz/portal.html).

The physico-chemical properties of *BnTCPs* were evaluated with the help of ProtParam (http://web.expasy.org/protparam/). The subcellular localization of BnTCP proteins was prophesied

from the WoLF PSORT server (https://wolfpsort.hgc.jp/). Genes structures were created *via* TBtools software (V 1.068; https://github.com/CJ-Chen/TBtools). The conserved motifs of BnTCP protein sequences were recognized employing the MEME website (https://meme-suite.org/meme/db/motifs).

Phylogenetic Tree and Synteny Analysis of BnTCP Proteins

To explore the evolutionary relationship of the *BnTCP*s, a phylogenetic tree between *B. napus, B. oleracea, B. rapa,* and *A. thaliana* protein sequences were constructed. The sequence alignment was executed using MEGA 7 software (https://megasoftware.net/home). The neighbor-joining process was accomplished to create a phylogenetic tree with 1,000 bootstrap replicates using the Evolview v3 website (https://www.evolgenius.info/evolview) to display the phylogenetic tree. Synteny relationships of *TCP* genes between *B. napus, B. oleracea, B. rapa,* and *A. thaliana* were performed using the python-package, JCVI (https://github.com/tanghaibao/jcvi) and Multiple Collinearity Scan Toolkit (MCScanX; https://github.com/wyp1125/MCScanX). The Ka/Ks ratios of all *TCPs* were calculated using KaKs_Calculator 2.0 (https://sourceforge.net/projects/kakscalculator2/).

Analysis of *Cis*-Regulatory Elements in the *BnTCPs* Promoters

To investigate the putative *cis*-elements in the *BnTCPs* promoters, we extracted the 2Kb sequence upstream of start codons in the *B. napus* genome. Moreover, the promoter sequence of each gene was examined applying the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), and the figure was drawn with the help of TBtools software.

Prediction of Putative miRNA Targeting *BnTCP* Genes and GO Annotation Analysis

The CDS of BnTCPs was employed to distinguish possible target miRNAs in the psRNATarget database (https://www.zhaolab.org/psRNATarget/home) with default parameters. The interaction network figure between the miRNAs and BnTCP genes was constructed with Cytoscape software (V3.8.2; https://cytoscape.org/download.html). Gene ontology (GO) annotation analysis was accomplished by submitting all BnTCPs protein sequences to the eggNOG website (http://eggnog-mapper.embl.de/). At the same time, GO enrichment analysis was performed with TBtools software.

Expression Profiling of *BnTCP* Genes in Diverse Tissues

For tissue-specific expression profiling, the rapeseed RNA-seq data was downloaded from the National Genomics Data Center (https://ngdc.cncb.ac.cn/gsa/; BioProject ID: PRJCA001495). The data were analyzed as described in a recent study (Raza et al., 2021a). In short, Cuffquant and Cuffnorm were used to produce normalized counts in transcripts per million (TPM) values. Due to the large difference in the expression profiles, we used the log10 method to calculate the expression results to better visualize differently expressed genes in diverse tissues. Based on log10 values, the expression heat map was created using GraphPad Prism 9.0.0 software (https://www.graphpad.com/).

Plant Material and Stress Conditions

A widely used variety, ZhongShuang 11 (ZS11), was used for stress treatments. Seeds of ZS11 were provided by the OCRI-CAAS, Wuhan, China. The stress treatments were carried out as described in our recent work (Raza et al., 2021a). All the treatments were executed with three biological replicates. All the samples were immediately frozen in liquid nitrogen and were stored at −80°C for further experimentation.

RNA Extraction and qRT-PCR Analysis

Total RNA was extracted using TransZol Up Plus RNA Kit (TransGen Biotech, Beijing, China), and cDNA was synthesized using cDNA SuperMix kit (TransGen Biotech, Beijing, China) according to developer instructions. The complete data of qRT-PCR reaction has been explained in our recent work (Raza et al., 2021a). The expression data were investigated using the $2^{-\Delta\Delta CT}$ method as described by Wang et al. (2014). All the primers used for qRT-PCR are listed in **Supplementary Table S1**. The graphs were developed using GraphPad Prism 9.0.0 software.

RESULTS

Identification and Characterization of *TCP* Genes in Rapeseed

To identify *TCP* family genes in rapeseed, 24 AtTCP protein sequences were used as queries for a BlastP search against the rapeseed genome. As a result, a sum of 80 *BnTCP* genes was identified with the TCP domain (**Table 1**). Hereafter, these genes are named as "*BnTCP1–BnTCP80.*" Among which 38 genes were positioned in the A subgenome, and 42 genes were positioned in the C subgenome (**Table 1**). Comprehensive data of the 80 *BnTCP* genes are shown in **Table 1**. Briefly, gene length varied from 531 bp (*BnTCP10*) to 9,673 bp (*BnTCP40*), the CDS length varied from 531 bp (*BnTCP10*) to 5,193 bp (*BnTCP40*), and the amino acid length varied from 176 (*BnTCP10*) to 1,730 (*BnTCP40*) amino acids. The number of exons varied from one (*BnTCP3–BnTCP6, BnTCP8–BnTCP10, BnTCP12–BnTCP15*) to 11 (*BnTCP40*) (**Table 1**). Notably, only two genes (*BnTCP40* and *BnTCP66*) had the highest number of introns (11 and 8, respectively), and several genes lack introns, such as *BnTCP3-BnTCP6, BnTCP8–BnTCP10, BnTCP12–BnTCP15* (**Table 1**). The forecasted molecular weights of the 80 BnTCP proteins extended from 18.60 kDa (BnTCP10) to 194.78 kDa (BnTCP40), and the isoelectric points prolonged from 5.31 (BnTCP28/45) to 11.12 (BnTCP49). The differences in molecular weights and isoelectric points are mainly due to the unusual high content of basic amino acids and post-translational modifications. The

TABLE 1 | Information of the identified *TCP* genes in rapeseed genome.

Gene ID	Gene name	Genomic position (bp)	Gene length (bp)	CDS length (bp)	Exon	Intron	Protein length (aa)	Molecular weight (kDa)	Theoretical isoelectric point (pi)	Subcellular localization
BnaA01T0093200ZS	*BnTCP1*	A01: 5350331–5352930: –	2,600	1,077	2	1	358	39.2194	8.32	Nucleus
BnaA01T0326500ZS	*BnTCP2*	A01: 30234583–30236602: –	2,020	1,284	4	3	427	48.7818	7.19	Nucleus
BnaA01T0351100ZS	*BnTCP3*	A01: 31838550–31839623: –	1,074	1,074	1	0	357	38.9368	7.45	Nucleus
BnaA02T0029000ZS	*BnTCP4*	A02: 1746997–1747686: +	690	690	1	0	229	23.9396	10.2	Nucleus
BnaA02T0147800ZS	*BnTCP5*	A02: 8342653–8343498: +	846	846	1	0	281	30.2123	5.42	Nucleus
BnaA02T0165200ZS	*BnTCP6*	A02: 9824501–9825541: –	1,041	1,041	1	0	346	39.2804	5.48	Nucleus
BnaA02T0176600ZS	*BnTCP7*	A02: 10751996–10753110: –	1,115	1,011	2	1	336	38.2344	9.04	Nucleus
BnaA02T0183300ZS	*BnTCP8*	A02: 11226562–11227527: +	966	966	1	0	321	33.8707	7.73	Nucleus
BnaA02T0196800ZS	*BnTCP9*	A02: 12284709–12285833: +	1,125	1,125	1	0	374	39.0581	8.99	Nucleus
BnaA02T0389800ZS	*BnTCP10*	A02: 33966392–33966922: –	531	531	1	0	176	18.6008	8.29	Mitochondria
BnaA02T0391800ZS	*BnTCP11*	A02: 34141978–34143043: –	1,066	906	2	1	301	33.862	7.76	Nucleus
BnaA03T0030500ZS	*BnTCP12*	A03: 1438953–1439693: +	741	741	1	0	246	27.3701	7.02	Nucleus
BnaA03T0032200ZS	*BnTCP13*	A03: 1501650–1502354: +	705	705	1	0	234	24.2451	10.87	Nucleus
BnaA03T0221500ZS	*BnTCP14*	A03: 11512211–11513155: –	945	945	1	0	314	33.4158	10.47	Nucleus
BnaA03T0286400ZS	*BnTCP15*	A03: 15203307–15204224: +	918	918	1	0	305	34.0894	7.89	Nucleus
BnaA03T0339700ZS	*BnTCP16*	A03: 18046255–18048379: +	2,125	1,218	3	2	405	44.1964	8.02	Nucleus
BnaA03T0354000ZS	*BnTCP17*	A03: 18816141–18817922: +	1,782	1,278	4	3	425	48.4559	8.5	Nucleus
BnaA03T0405400ZS	*BnTCP18*	A03: 21941344–21942444: –	1,101	1,101	1	0	366	40.739	6.68	Nucleus
BnaA03T0447400ZS	*BnTCP19*	A03: 24280135–24281286: –	1,152	1,152	1	0	383	42.3337	7.69	Nucleus
BnaA04T0122600ZS	*BnTCP20*	A04: 13913091–13913765: +	675	675	1	0	224	24.7676	7.82	Mitochondria
BnaA05T0054400ZS	*BnTCP21*	A05: 2969609–2971094: –	1,486	975	1	0	324	34.5098	10.18	Nucleus
BnaA05T0129300ZS	*BnTCP22*	A05: 7820651–7821986: +	1,336	1,047	2	1	348	38.6262	6.85	Peroxisome
BnaA05T0402900ZS	*BnTCP23*	A05: 39177839–39180026: –	2,188	1,227	2	1	408	44.3164	7.55	Nucleus
BnaA06T0174300ZS	*BnTCP24*	A06: 11013702–11015537: –	1,836	1,437	1	0	478	51.5508	7.21	Nucleus
BnaA06T0320300ZS	*BnTCP25*	A06: 40732993–40733748: +	756	756	1	0	251	27.1051	10.33	Nucleus
BnaA06T0386200ZS	*BnTCP26*	A06: 44783008–44783931: +	924	924	1	0	307	32.3907	7.96	Nucleus
BnaA07T0270200ZS	*BnTCP27*	A07: 25405486–25406442: –	957	957	1	0	318	33.7336	8.08	Mitochondria
BnaA07T0280800ZS	*BnTCP28*	A07: 26138757–26139815: +	1,059	1,059	1	0	352	39.9883	5.31	Nucleus
BnaA07T0291800ZS	*BnTCP29*	A07: 26854045–26855100: –	1,056	1,056	1	0	351	39.9273	7.2	Nucleus
BnaA07T0312100ZS	*BnTCP30*	A07: 27924720–27925685: +	966	966	1	0	321	34.0659	7.47	Nucleus
BnaA07T0330100ZS	*BnTCP31*	A07: 29148894–29149970: +	1,077	1,077	1	0	358	37.3864	7.47	Nucleus
BnaA08T0015500ZS	*BnTCP32*	A08: 1202449–1203435: +	987	987	1	0	328	35.6779	7.24	Nucleus
BnaA08T0205600ZS	*BnTCP33*	A08: 22895827–22896753: +	927	834	2	1	277	30.8253	6.95	Nucleus
BnaA09T0037700ZS	*BnTCP34*	A09: 2358891–2359727: –	837	837	1	0	278	29.6245	6.63	Nucleus
BnaA09T0073000ZS	*BnTCP35*	A09: 4327191–4327838: –	648	648	1	0	215	23.0186	9.87	Chloroplast
BnaA09T0175800ZS	*BnTCP36*	A09: 11575172–11576356: +	1,185	1,185	1	0	394	41.4311	6.46	Nucleus

(Continued on following page)

TABLE 1 | (*Continued*) Information of the identified *TCP* genes in rapeseed genome.

Gene ID	Gene name	Genomic position (bp)	Gene length (bp)	CDS length (bp)	Exon	Intron	Protein length (aa)	Molecular weight (kDa)	Theoretical isoelectric point (pi)	Subcellular localization
BnaA09T0412200ZS	*BnTCP37*	A09: 47278416–47282262: +	3,847	957	3	2	318	35.2881	8.13	Nucleus
BnaA10T0253900ZS	*BnTCP38*	A10: 24336051–24336761: –	711	711	1	0	236	24.7343	8.55	Nucleus
BnaC01T0114300ZS	*BnTCP39*	C01: 7792345–7794974: –	2,630	1,080	2	1	359	39.2785	8.32	Nucleus
BnaC01T0126300ZS	*BnTCP40*	C01: 8666079–8675751: +	9,673	5,193	11	10	1730	194.7861	8.16	Nucleus
BnaC01T0404000ZS	*BnTCP41*	C01: 46796515–46798601: –	2,087	1,293	4	3	430	49.2565	7.4	Nucleus
BnaC01T0435000ZS	*BnTCP42*	C01: 50033768–50035625: –	1,858	1,089	2	1	362	39.202	7.25	Nucleus
BnaC02T0032000ZS	*BnTCP43*	C02: 2159588–2160265: +	678	678	1	0	225	23.5673	9.85	Nucleus
BnaC02T0189600ZS	*BnTCP44*	C02: 15598917–15599783: +	867	867	1	0	288	30.9993	6.15	Nucleus
BnaC02T0211700ZS	*BnTCP45*	C02: 18337482–18338522: –	1,041	1,041	1	0	346	39.4164	5.31	Nucleus
BnaC02T0229100ZS	*BnTCP46*	C02: 20787472–20790829: –	3,358	1,005	4	3	334	38.0556	7.95	Nucleus
BnaC02T0241600ZS	*BnTCP47*	C02: 22073752–22074708: +	957	957	1	0	318	33.7857	7.71	Nucleus
BnaC02T0261700ZS	*BnTCP48*	C02: 24613854–24614969: +	1,116	1,116	1	0	371	38.7778	9.38	Nucleus
BnaC02T0521300ZS	*BnTCP49*	C02: 62454902–62455594: –	693	693	1	0	230	24.9241	11.12	Nucleus
BnaC02T0524000ZS	*BnTCP50*	C02: 62771638–62772713: –	1,076	906	2	1	301	33.7318	6.72	Nucleus
BnaC03T0037100ZS	*BnTCP51*	C03: 1909961–1910695: +	735	735	1	0	244	27.341	6.91	Nucleus
BnaC03T0039100ZS	*BnTCP52*	C03: 2018328–2019035: +	708	708	1	0	235	24.258	10.87	Nucleus
BnaC03T0260500ZS	*BnTCP53*	C03: 16121718–16122599: –	882	882	1	0	293	30.929	10.32	Nucleus
BnaC03T0344300ZS	*BnTCP54*	C03: 23421941–23422903: +	963	963	1	0	320	35.7522	7.08	Nucleus
BnaC03T0408900ZS	*BnTCP55*	C03: 28003605–28005738: +	2,134	1,215	3	2	404	44.0973	8.04	Nucleus
BnaC03T0429200ZS	*BnTCP56*	C03: 29713443–29715268: +	1,826	1,278	4	3	425	48.6133	8.66	Nucleus
BnaC03T0620400ZS	*BnTCP57*	C03: 59527128–59528955: +	1,828	1,422	1	0	473	51.1025	7.44	Nucleus
BnaC03T0663600ZS	*BnTCP58*	C03: 64144410–64145333: –	924	924	1	0	307	34.4323	8.09	Nucleus
BnaC04T0060500ZS	*BnTCP59*	C04: 5408796–5409773: –	978	978	1	0	325	34.6679	10.03	Chloroplast
BnaC04T0168800ZS	*BnTCP60*	C04: 15546097–15547417: +	1,321	1,047	2	1	348	38.6283	7.04	Peroxisome
BnaC04T0410600ZS	*BnTCP61*	C04: 53526218–53526877: +	660	660	1	0	219	23.7855	6.89	Mitochondria
BnaC05T0259700ZS	*BnTCP62*	C05: 21147596–21148549: –	954	954	1	0	317	35.0458	8.13	Nucleus
BnaC05T0413400ZS	*BnTCP63*	C05: 46522363–46523833: –	1,471	1,248	3	2	415	47.8023	8.78	Nucleus
BnaC05T0452400ZS	*BnTCP64*	C05: 50135134–50137315: –	2,182	1,215	2	1	404	44.2776	8.38	Nucleus
BnaC05T0566500ZS	*BnTCP65*	C05: 57728974–57732707: +	3,734	1,047	3	2	348	38.7422	7.42	Nucleus
BnaC05T0569300ZS	*BnTCP66*	C05: 57878862–57886495: +	7,634	1,944	9	8	647	72.5415	8.6	Chloroplast
BnaC06T0305200ZS	*BnTCP67*	C06: 40942818–40943765: –	948	948	1	0	315	33.6966	8.08	Mitochondria
BnaC06T0321700ZS	*BnTCP68*	C06: 42719251–42720315: +	1,065	1,065	1	0	354	40.2017	5.86	Nucleus
BnaC06T0335100ZS	*BnTCP69*	C06: 44122410–44123480: –	1,071	1,071	1	0	356	40.574	5.56	Nucleus

(Continued on following page)

TABLE 1 | (*Continued*) Information of the identified *TCP* genes in rapeseed genome.

Gene ID	Gene name	Genomic position (bp)	Gene length (bp)	CDS length (bp)	Exon	Intron	Protein length (aa)	Molecular weight (kDa)	Theoretical isoelectric point (pi)	Subcellular localization
BnaC06T0353200ZS	*BnTCP70*	C06: 45538660–45539631: –	972	972	1	0	323	36.4448	8.94	Nucleus
BnaC06T0362700ZS	*BnTCP71*	C06: 46363762–46364718: +	957	957	1	0	318	34.0369	7.65	Nucleus
BnaC06T0386600ZS	*BnTCP72*	C06: 48098081–48099548: +	1,468	1,068	1	0	355	37.2392	7.17	Nucleus
BnaC07T0302100ZS	*BnTCP73*	C07: 44542787–44543746: –	960	960	1	0	319	33.6842	8	Nucleus
BnaC07T0373600ZS	*BnTCP74*	C07: 49902745–49903473: –	729	729	1	0	242	26.0729	10.11	Nucleus
BnaC07T0378100ZS	*BnTCP75*	C07: 50224221–50225315: –	1,095	1,095	1	0	364	40.5408	6.68	Nucleus
BnaC07T0422500ZS	*BnTCP76*	C07: 53049841–53052375: –	2,535	1,125	2	1	374	41.2696	7.62	Nucleus
BnaC09T0023100ZS	*BnTCP77*	C09: 1586085–1587093: –	1,009	831	2	1	276	29.3745	6.75	Mitochondria
BnaC09T0063000ZS	*BnTCP78*	C09: 4101151–4101798: –	648	648	1	0	215	22.9824	9.87	Chloroplast
BnaC09T0197900ZS	*BnTCP79*	C09: 17390038–17391216: +	1,179	1,179	1	0	392	41.1277	6.51	Mitochondria
BnaC09T0567700ZS	*BnTCP80*	C09: 64952104–64952814: –	711	711	1	0	236	24.7502	7.81	Nucleus

In the genomic position, the positive (+) and negative (–) sign shows the presence of a gene on the positive and negative strand of that specific marker correspondingly.

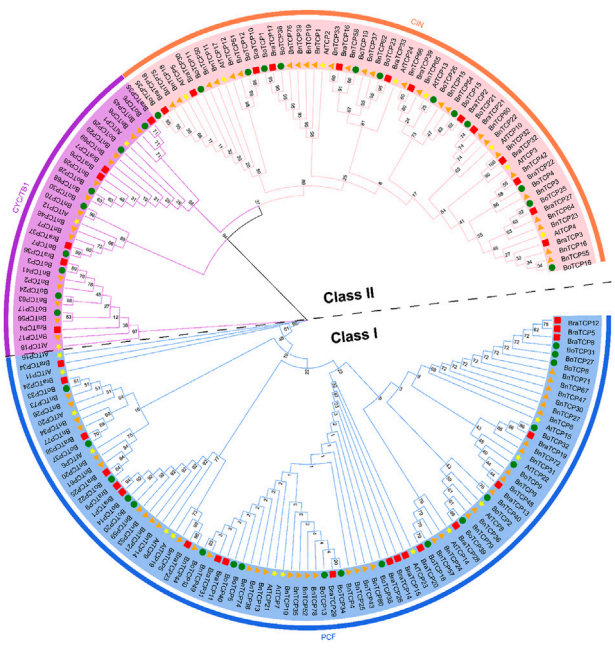


FIGURE 1 | A neighbor-joining phylogenetic tree assessment of *TCP* genes from *B. napus, B. oleracea, B. rapa*, and *A. thaliana*. Overall, 80 *BnTCPs* from *B. napus* (orange triangle), 40 *BoTCPs* from *B. oleracea* (green circles), 39 *BraTCPs* from *B. rapa* (red boxes), and 24 *AtTCPs* from *A. thaliana* (yellow stars) were clustered into two major classes, denoted by exclusive colors.

results of *in silico* subcellular localization prophesied that two BnTCP proteins were located on the peroxisome, four BnTCP proteins on the chloroplast, seven BnTCP proteins on the mitochondria, and remaining 67 BnTCP proteins were located on the nucleus (**Table 1**). Meanwhile, 24 genes (*AtTCP1–AtTCP24*) from *A. thaliana*, 39 genes (*BraTCP1–BraTCP39*) from *B. rapa*, and 40 genes (*BoTCP1–BoTCP41*) from *B. oleracea* genomes were also identified (**Supplementary Table S2**).

Phylogenetic Relationship of *TCP* Genes

To unreveal the evolutionary relationship between the *BnTCPs, BoTCPs, BraTCPs,* and *AtTCPs* genes, an unrooted phylogenetic tree was built by a multiple sequence alignment of the prophesied TCP protein sequences from *B. napus, B. oleracea, B. rapa,* and *A. thaliana* (**Figure 1**). Based on the standard classification of *Arabidopsis TCP* genes (Yao et al., 2007), the 184 *TCPs* genes from four plant species were distributed into two main classes, i.e., Class I (PCF) and Class II. Further, Class II was further divided into two sub-classes (CYC/TB1 and CIN) (**Figure 1**). The *TCPs* from the four plant species were scattered in nearly all clades, suggesting that the *TCP* family genes differentiated before the separation of these ancestral plant species. The findings discovered that Class I (PCF) was encompassed 91 *TCP* members (39 *BnTCPs*, 19 *BoTCPs*, 20 *BraTCPs,* and 13 *AtTCPs*). Class II contained a sum of 92 *TCP* members. Particularly, sub-class CYC/TB1 was comprised of 31 *TCP* members (14 *BnTCPs*, eight *BoTCPs*, six *BraTCPs,* and three *AtTCPs*), and CIN was comprised of 61 *TCP* members (27 *BnTCPs*, 13 *BoTCPs*, 13 *BraTCPs,* and eight *AtTCPs*) (**Figure 1**; **Supplementary Table S3**). In a nutshell, *TCPs* grouping into the same sub-class may retain corresponding functions. It is worth highlighting that *BnTCP* genes were almost equally scattered in both classes with homologs from other plant species (**Figure 1**; **Supplementary Table S3**). Mainly,

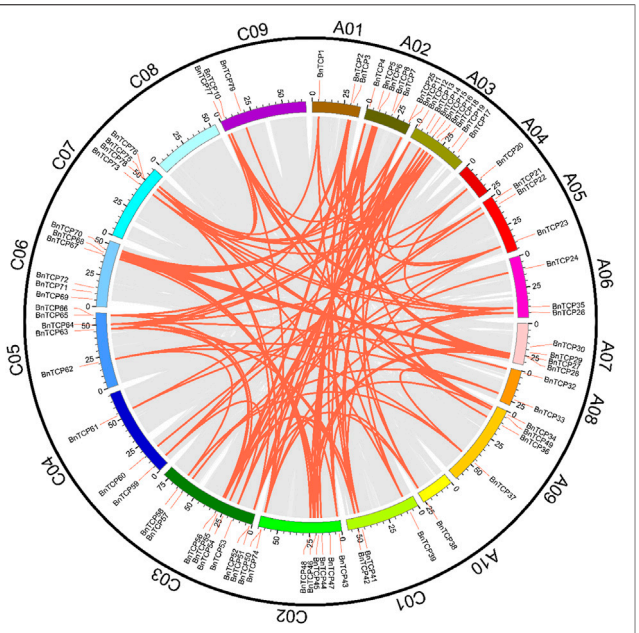


FIGURE 2 | Chromosomal locations and inter-chromosomal relations of *BnTCP* genes. Grey lines in the background present all syntenic blocks in the *B. napus* genome, and the red lines present syntenic *BnTCPs* gene pairs.

Class I (PCF) was observed to have a greater number of *BnTCPs* (39), followed by CIN (27) and CYC/TB1 (14). Moreover, it was observed that the *BnTCPs* have a more closely phylogenetic relationship with the *BoTCPs* and *BraTCPs* in each class.

Chromosomal Locations and Synteny Evaluation of *TCP* Genes

The chromosol location of 38 *BnTCPs* gene pairs was observed (**Figure 2**; **Supplementary Table S4**), and 18 out of 19 chromosomes held *BnTCPs* genes (**Figure 2**). Mainly, chromosomes A04 and A10 possessed only one gene, chromosome A08 possessed two genes, chromosomes A01, A05, A06, C01, C04, and C09 possessed three genes, chromosomes A07, A09, and C07 possessed four genes, chromosomes A02 and C05 possessed five genes, chromosome C06 possessed six genes, chromosomes C02 and C03 possessed eight genes. Chromosome A03 had the highest number of genes (10). Astonishingly, the C08 chromosome did not contain any *BnTCP* genes (**Figure 2**). The dataset displays that segmental duplications have contributed to the development of *BnTCP* genes (**Supplementary Table S4**). Notably, one putative paralog pair, i.e., *BnTCP16/BnTCP55*, was produced by tandem duplication (**Supplementary Table S4**). These results showed the large-scale segmental duplication events involved in the evolution of the *BnTCP* gene family in rapeseed.

Collinearity analysis was used to appraise the evolutionary relationship of the *TCP* genes between *B. napus* and the other heritable plant species (**Figure 3**; **Supplementary Table S4**). In general, in the A subgenome, several *B. napus* genes showed syntenic networks with various *BraTCPs,* and *AtTCP*. Notably, one *BnTCP* gene at chromosome A10 also displayed a syntenic relationship with a *BoTCP* gene at chromosome C09. Likewise, in the C subgenome, many *B. napus* genes revealed syntenic connections with different *BraTCPs*, and *AtTCPs* (**Figure 3**). Mainly, three *BnTCP* genes at chromosome C03 and C05 also showed a syntenic relationship with three *BrTCP* genes at chromosome A03, A05 and A08 (**Figure 3**). On the other hand, several homologues of *A. thaliana, B. rapa,* and *B. oleracea* showed a syntenic alliance with *BnTCPs* genes in the rapeseed genome, suggesting that whole-genome or segmental duplication events represented a crucial role in the evolution of *BnTCPs* gene family in the rapeseed genome (**Figure 3**; **Supplementary Table S4**).

The Ka/Ks ratio is regarded as a considerable indicator in assessing the sequence evolution in terms of selection pressures and duplication types (Hurst, 2002). Hence, to comprehend the evolutionary history of the *BnTCPs*, the Ka, Ks, and Ka/Ks ratio was determined for *B. napus* and the other three plant species (**Supplementary Table S4**). The calculations disclosed that all of the duplicated *BnTCP* gene pairs had a Ka/Ks ratio of <1 (**Supplementary Table S4**), indicating that the *BnTCP* genes may have undergone intense purifying selective pressure during the evolution process (**Supplementary Table S4**). Similar outcomes were also examined in *B. rapa, B. oleracea,* and *A. thaliana* (**Supplementary Table S4**).

Assessment of Gene Structures and Conserved Motifs of *BnTCP* Genes

The exon-intron configurations of the *BnTCP* genes were examined to boost our knowledge of the evolution of the rapeseed *TCP* family genes. The results showed that the number of introns and exons ranged from 0 to 10 and 1–11, respectively (**Figure 4B**; **Table 1**). Overall, 55 genes have one exon and zero intron; 13 genes have two exons and one intron; five genes have three exons and two introns; five genes have four exons and three introns; one gene has nine exons and eight introns; and one gene has 11 exons and 10 introns (**Figure 4**; **Table 1**). Notably, Class I (PCF) and Class II (CIN and CYC/TB1) possessed almost similar structures except for a few genes (**Figure 4B**). Losses or gains of exons were discovered throughout the evolution of the *TCP* family genes. Our outcomes advised that *TCP* genes retained a comparatively constant exon-intron arrangement during the evolution of the rapeseed genome. Moreover, *BnTCP* gene members inside a Class had extremely matching gene structures, constant with their phylogenetic groups.

Furthermore, the full-length protein sequences were examined to identify their conserved motifs (**Figure 4C**). The conserved motif of the *TCP* genes varied from 2 to 7. Overall, nine conserved motifs were identified, and the complete data (such as name, sequence, width, and E-value) of discovered motifs are presented in **Supplementary Table S5**. The motif dispersals were also comparable within the Classes (**Figure 4C**). For example, motifs 1, 2, 6, and 7 were specific to Class I; however, motifs 6 were also present in some genes belonging to Class II (mainly CYC/TB1). In Class I, only one gene (*BnTCP40*) was found to have a greater number of motifs (seven). In Class II (mainly CYC/

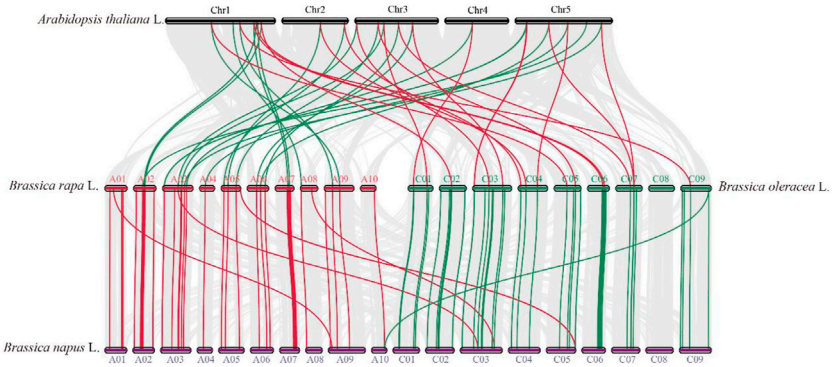


FIGURE 3 | Collinearity analysis of *TCP* genes in *B. napus, B. rapa, B. oleracea,* and *A. thaliana* chromosomes. Grey lines in the background display the collinear blocks within *B. napus* and other plant genomes. However, the red and green lines signify the syntenic *TCP* gene pairs. Genes located on the *B. napus* A subgenome are syntenic with *B. rapa, B. oleracea,* and *A. thaliana*, while genes located on *B. napus* C subgenome are mainly syntenic with *B. rapa, B. oleracea,* and *A. thaliana.*

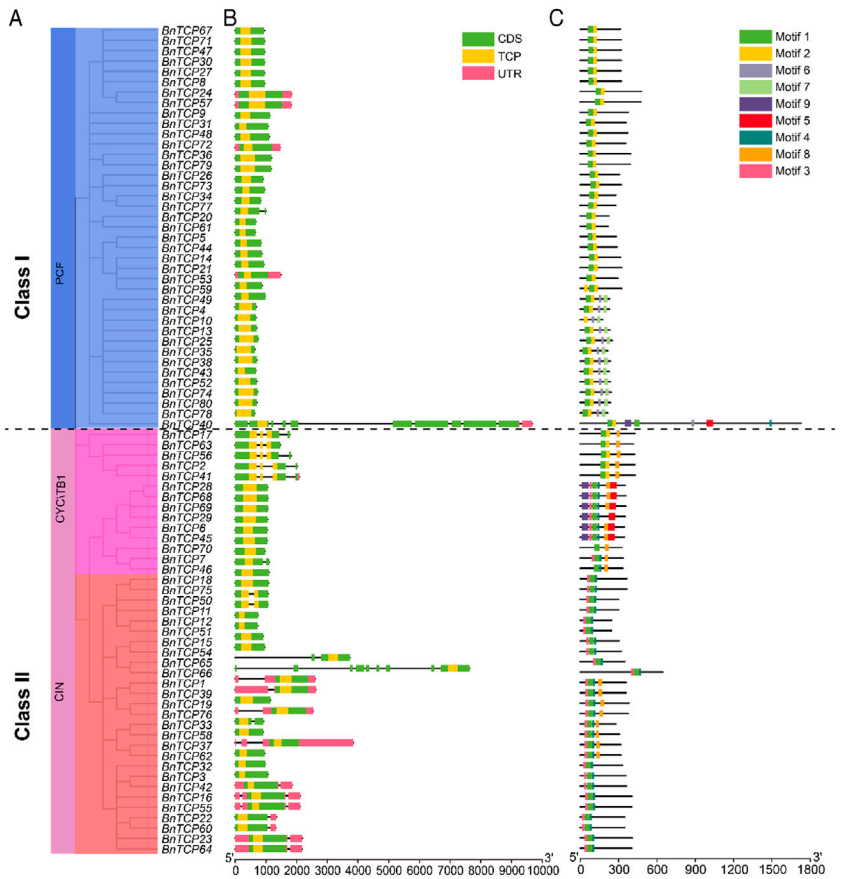


FIGURE 4 | The gene structure and motif analysis of *BnTCPs* genes. **(A)** Based on phylogenetic relationships, the *BnTCPs* were clustered into two major classes. **(B)** Gene structure of *BnTCPs* genes. The pink color signifies UTR regions, the light green color denotes exon or CDS, the yellow color represents TCP domain, and the black horizontal line symbolizes introns. **(C)** Conserved motif compositions detected in *BnTCPs*. Different color boxes represent different motifs.

TB1), motifs 5 and 9 were specific and absent in other genes. Likewise, motifs 4, 8, and 3 were specific to Class II (CIN and CYC/TB1) (**Figure 4C**). In short, the class organization's consistency was convincingly sustained by evaluating the conserved motifs composition, gene structures, and phylogenetic interactions, indicating that the TCP proteins

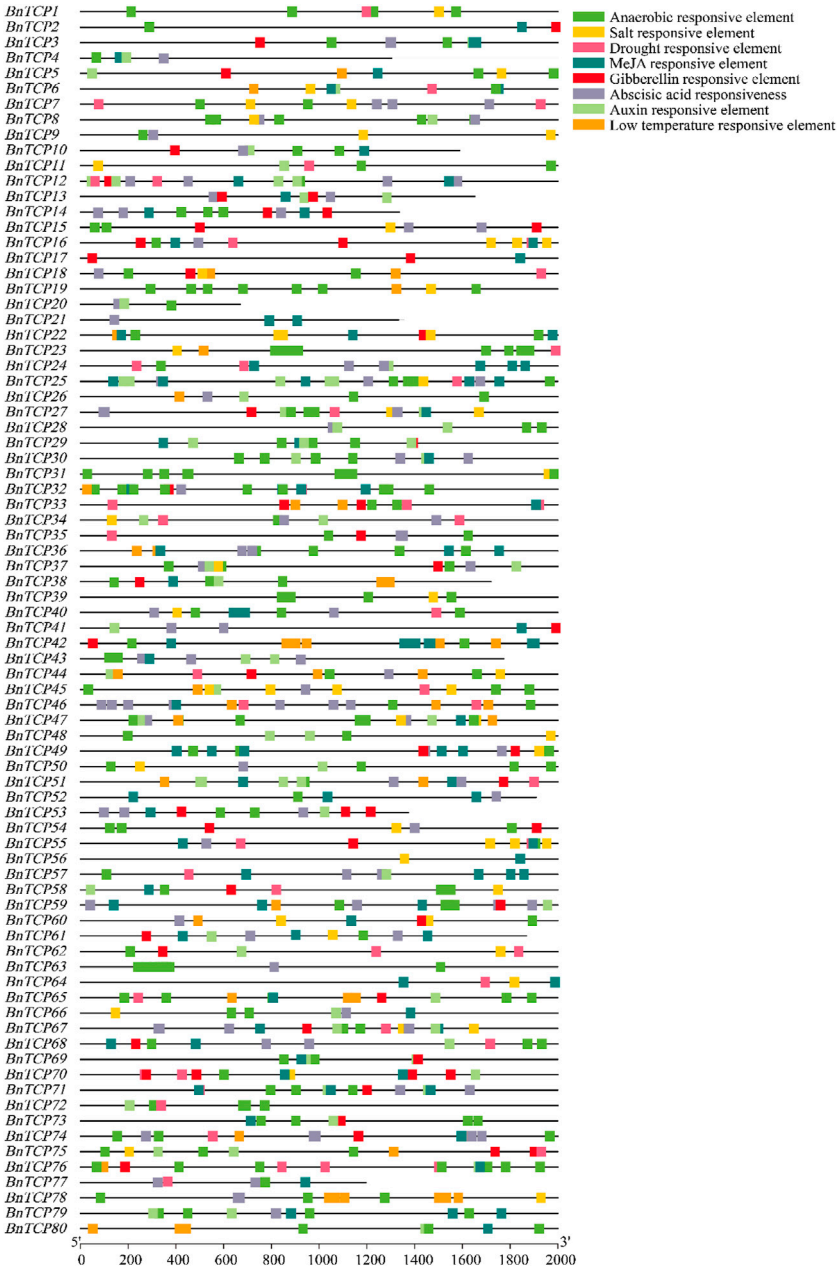


FIGURE 5 | *Cis*-elements in the promoter regions of the *BnTCP* genes are linked with different hormone- and stress-responsive elements. Different color boxes show different identified elements. See **Supplementary Table S6** for more information.

have tremendously well-maintained amino acid remains and members within a class may possess parallel roles.

Examination of *Cis*-Elements in Promoters of *BnTCP* Genes

The *cis*-acting elements in the gene promoter generally regulate gene expression levels and functions. Therefore, *cis*-acting elements in *BnTCPs* promoter localities were analyzed to characterize the gene functions and regulatory roles. The comprehensive data of *cis*-elements are shown in **Supplementary Table S6**. Overall, we focused on and identified eight types of elements. Particularly, four phytohormone-correlated [(abscisic acid (ABA), auxin, methyl jasmonate (MeJA), and gibberellin (GA)] responsive elements consist of ABRE, TGA-element, TGACG-motif, CGTCA-motif, P-box, GARE-motif, TATC-box, were distinguished (**Figure 5**; **Supplementary Table S6**). Specifically, several phytohormone-correlated elements were anticipated to be limited to some genes and extensively

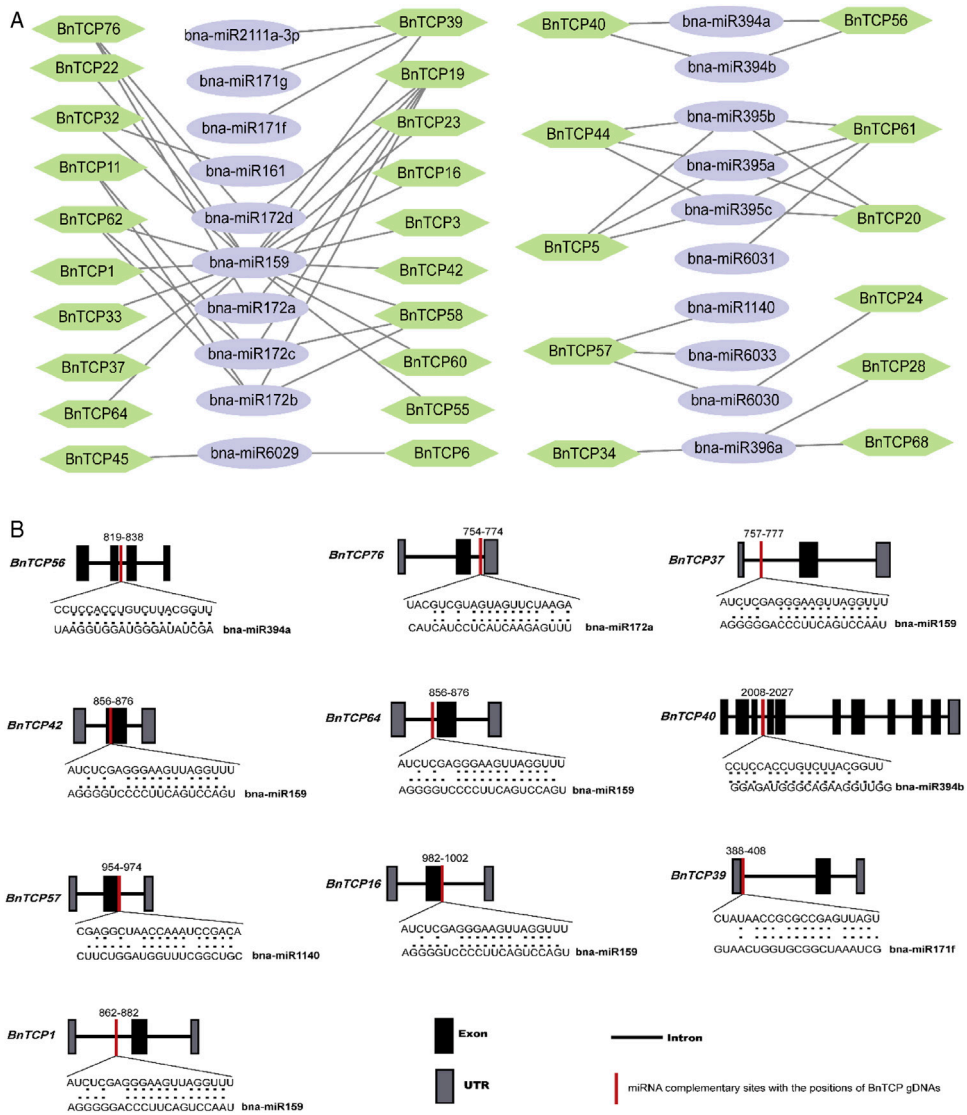


FIGURE 6 | miRNA targeting *BnTCP* genes. **(A)** Network figure of anticipated miRNA targeting *BnTCPs* genes. Green hexagon colors correspond to *BnTCPs* genes, and bluish ellipse shapes represent miRNAs. **(B)** The schematic illustration signifies the *BnTCPs* targeted by miRNAs. The RNA sequence of each complementary site from 5′ to 3′ and the predicted miRNA sequence from 3′ to 5′ are exposed in the long-drawn-out areas. See **Supplementary Table S7** for the detailed data of all predicted miRNAs.

dispersed (**Figure 5**; **Supplementary Table S6**), indicating the essential role of these genes in phytohormone-arbitration.

Additionally, we identified four stress-associated (salt, drought, low-temperature, and anaerobic) responsive elements, including TCA-element, MBS, LTR, ARE, suggesting their contribution to stress stimulation (**Figure 5**; **Supplementary Table S6**). TCA-element associated with salt stress responsiveness was mainly present in 38 *BnTCP* genes and widely distributed in four genes (*BnTCP16, BnTCP22, BnTCP45,* and *BnTCP55*) (**Figure 5**; **Supplementary Table S6**). Drought stress responsive element (MBS) was mainly present in 36 genes and widely circulated in 12 genes (**Figure 5**; **Supplementary Table S6**). Low-temperature responsive element (LTR) was particularly present in 26 genes and widely scattered in six genes (*BnTCP42, BnTCP44, BnTCP46, BnTCP65, BnTCP78,* and *BnTCP80*) (**Figure 5**; **Supplementary Table S6**). ARE element associated with anaerobic responsiveness was mainly present in 74 genes and largely dispersed in 35 genes (**Figure 5**; **Supplementary Table S6**). Generally, these findings recommended that *BnTCPs* expression profiles may differ under phytohormone and abiotic stress conditions.

Genome-Wide Prediction of miRNA Targeting *BnTCP* Genes

During the past few years, various investigations have uncovered that miRNA-induced regulation participates in the stress responses in plants. Consequently, for a profound knowledge of miRNA-mediated post-

transcriptional adjustment of *BnTCP genes*, we forecasted 32 miRNAs targeting 21 *BnTCP* genes (**Figure 6A**; **Supplementary Table S7**). Some of the miRNA-targeted sites are shown in **Figure 6B**, while the comprehensive data of all miRNAs targeted genes/sites is given in **Supplementary Table S7**. Overall, the results demonstrated that one member of the bna-miR1140 family targeted one gene (*BnTCP57*); one member of the bna-miR161 family targeted one gene (*BnTCP32*); one member of the bna-miR162 family targeted one gene (*BnTCP77*); one member of the bna-miR2111a-3p targeted one gene (*BnTCP39*); one member of the bna-miR6031 targeted one gene (*BnTCP61*); one member of the bna-miR6033 targeted one gene (*BnTCP57*); one member of bna-miR396 family targeted three genes (*BnTCP28, BnTCP34,* and *BnTCP68*); and one member of bna-miR159 family targeted 17 genes. Two members of the bna-miR171 family targeted one gene (*BnTCP39*); two members of the bna-miR6029 family targeted two genes (*BnTCP6* and *BnTCP45*); two members of the bna-miR6030 family targeted two genes (*BnTCP24* and *BnTCP57*); and two members of the bna-miR394 family targeted two genes (*BnTCP40* and *BnTCP56*). Three members of the bna-miR395 family targeted four genes (*BnTCP5, BnTCP61, BnTCP20*, and *BnTCP44*). Four members of the bna-miR172 family targeted five genes (*BnTCP11, BnTCP19, BnTCP58, BnTCP62*, and *BnTCP76*) (**Figure 6A**; **Supplementary Table S7**). In general, *BnTCP5, BnTCP19, BnTCP20, BnTCP39, BnTCP44, BnTCP57, BnTCP58, BnTCP61,* and *BnTCP76* were predicted to be targeted by a greater number (three and four) of miRNAs (**Figure 6A**; **Supplementary Table S7**). In further examination, the expression levels of these miRNAs and their targeted genes require validation to preside their biological functions in the rapeseed genome.

Functional Annotation Study of *BnTCP* Genes

To further recognize the contribution of *BnTCP* genes, we performed GO annotation and enrichment analysis based on biological process (BP), molecular function (MF), and cellular component (CC) classes. These terms can boost our understanding of the diverse function of genes. The results of BP, MF, and CC exhibited several significantly enriched terms (**Supplementary Table S8**). For instance, in GO-BP class, 49 highly enriched terms were identified, including cellular process (GO: 0009987), RNA biosynthetic process (GO: 0032774), regulation of nucleic acid-templated transcription (GO: 1903506), heterocycle metabolic process (GO: 0046483), organic substance biosynthetic process (GO: 1901576), regulation of metabolic process (GO:0019222), nitrogen compound metabolic process (GO: 0006807) (**Supplementary Table S8**). The GO-CC enrichment outcomes distinguished ten highly enriched terms, including intracellular part (GO: 0044424), cell part (GO: 0044464), cellular_component (GO: 0005575), organelle (GO: 0043226), membrane-bounded organelle (GO: 0043227), intracellular (GO: 0005622), nucleus (GO: 0005634) (**Supplementary Table S8**). Similarly, the GO-MF results projected eight significantly enriched terms such as transcription regulator activity (GO: 0140110), molecular_function (GO: 0003674), DNA binding transcription factor activity (GO: 0003700), organic cyclic compound binding (GO: 0097159), nucleic acid binding (GO: 0003676) (**Supplementary Table S8**). Overall, GO enrichment analysis validates the functional role of *BnTCP* genes in several cellulars, molecular, and biological processes related to RNA/DNA binding, metabolic processes, transcriptional regulatory activities associated with rapeseed growth and developmental processes.

Expression Profiles of *BnTCP* Genes Under Various Tissues

Tissue-specific expression profiles of *BnTCPs* genes were examined in six diverse tissues and organs, i.e., roots, stems, leaves, flowers, seeds, and silique using RNA-seq data from *B. napus* (ZS11 variety) (BioProject ID: PRJCA001495). Overall, the results showed that only a few genes showed higher expression in some specific tissues (**Figure 7**). For instance, some genes showed higher expression levels in the root, including *BnTCP9, BnTCP13, BnTCP25, BnTCP31, BnTCP35, BnTCP36, BnTCP48, BnTCP52, BnTCP72,* and *BnTCP74*). In stem, a few genes showed higher expression levels, including *BnTCP9, BnTCP26, BnTCP31, BnTCP36, BnTCP48, BnTCP52, BnTCP66, BnTCP67, BnTCP72, BnTCP73,* and *BnTCP74*). In leaf, *BnTCP9, BnTCP13, BnTCP16, BnTCP19, BnTCP22, BnTCP23, BnTCP35, BnTCP39, BnTCP52, BnTCP55, BnTCP60, BnTCP64, BnTCP65, BnTCP66, BnTCP67, BnTCP74,* and *BnTCP76* genes display higher expression patterns. In flower, *BnTCP16, BnTCP19, BnTCP23, BnTCP35, BnTCP42, BnTCP55, BnTCP60, BnTCP64, BnTCP52, BnTCP55, BnTCP60, BnTCP64, BnTCP66, BnTCP74, BnTCP76,* and *BnTCP79* genes exhibited higher expression patterns. In seeds, several genes showed higher expression in seeds at different time points, such as *BnTCP3, BnTCP9, BnTCP22, BnTCP25, BnTCP35, BnTCP36 BnTCP60, BnTCP74, BnTCP78,* and *BnTCP79*). Likewise, in silique, many genes including *BnTCP9, BnTCP22, BnTCP24, BnTCP25, BnTCP27, BnTCP30, BnTCP31, BnTCP35, BnTCP39, BnTCP48, BnTCP52, BnTCP60, BnTCP64, BnTCP66, BnTCP67, BnTCP71, BnTCP74, BnTCP78,* and *BnTCP80,* showed higher expression patterns at different time points (**Figure 7**). Notably, some genes also showed moderate expression levels in various tissues. Overall, expression data show that only a few genes may significantly contribute to rapeseed growth and development. Therefore, the precise function of these genes could be discovered in future investigations.

Expression Profiles of *BnTCP* Genes Under Phytohormones and Abiotic Stress Conditions

In this study, qRT-PCR-based expression profiles of ten randomly selected *BnTCPs* genes were evaluated in response to abiotic (cold, waterlogging, salinity, and drought), and phytohormones (ABA, GA, IAA, and MeJA) stresses at different time points (**Figures 8**, **9**). Under abiotic stress conditions, most of the genes exhibited comparatively low expression levels, apart from some genes. For example, in response to cold stress, *BnTCP53* showed significantly higher



FIGURE 7 | Expression profiling of *BnTCP* genes in various developmental organs/tissues. The 7, 10, 14, and 45 d tags specified the time-points when the samples were harvested. The red, black, and green colors display high to low expression levels. The expression heat map was created by taking log10 of transcripts per million (TPM) values.

expression at 4 h (1.521 fold), 6 h (2.081 fold), and 8 h (1.616 fold); *BnTCP59* showed higher expression at 2 h (2.312 fold), 4 h (2.409 fold), 6 h (2.873 fold), and 8 h (2.187 fold); and *BnTCP60* showed higher expression mainly at 6 h (1.254 fold), and 8 h (1.598 fold), compared to CK. Under waterlogging stress, *BnTCP39* showed higher expression at 2 h (11.684 fold), 8 h (13.369 fold); and *BnTCP53* showed higher expression at 2 h (10.865 fold), compared to CK. Under salt stress, almost all genes showed relatively higher expression at CK; however, *BnTCP53* display substantially higher expression at 2 h (1.402 fold), 4 h (1.395 fold), 6 h (1.109 fold), and 8 h (1.197 fold), compared to CK and other genes. In response to drought stress, only *BnTCP7* was up-regulated at 4 h (29.946 fold), 6 h (31.192 fold), and 8 h (28.428 fold) (**Figure 8**).

Under phytohormones treatment, almost all genes showed significantly higher expression except MeJA. Particularly, *BnTCP39, BnTCP64, BnTCP75, BnTCP25, BnTCP36, BnTCP53*, and *BnTCP7* display higher expression levels in response to ABA and GA treatments. These genes also possess ABA and GA-

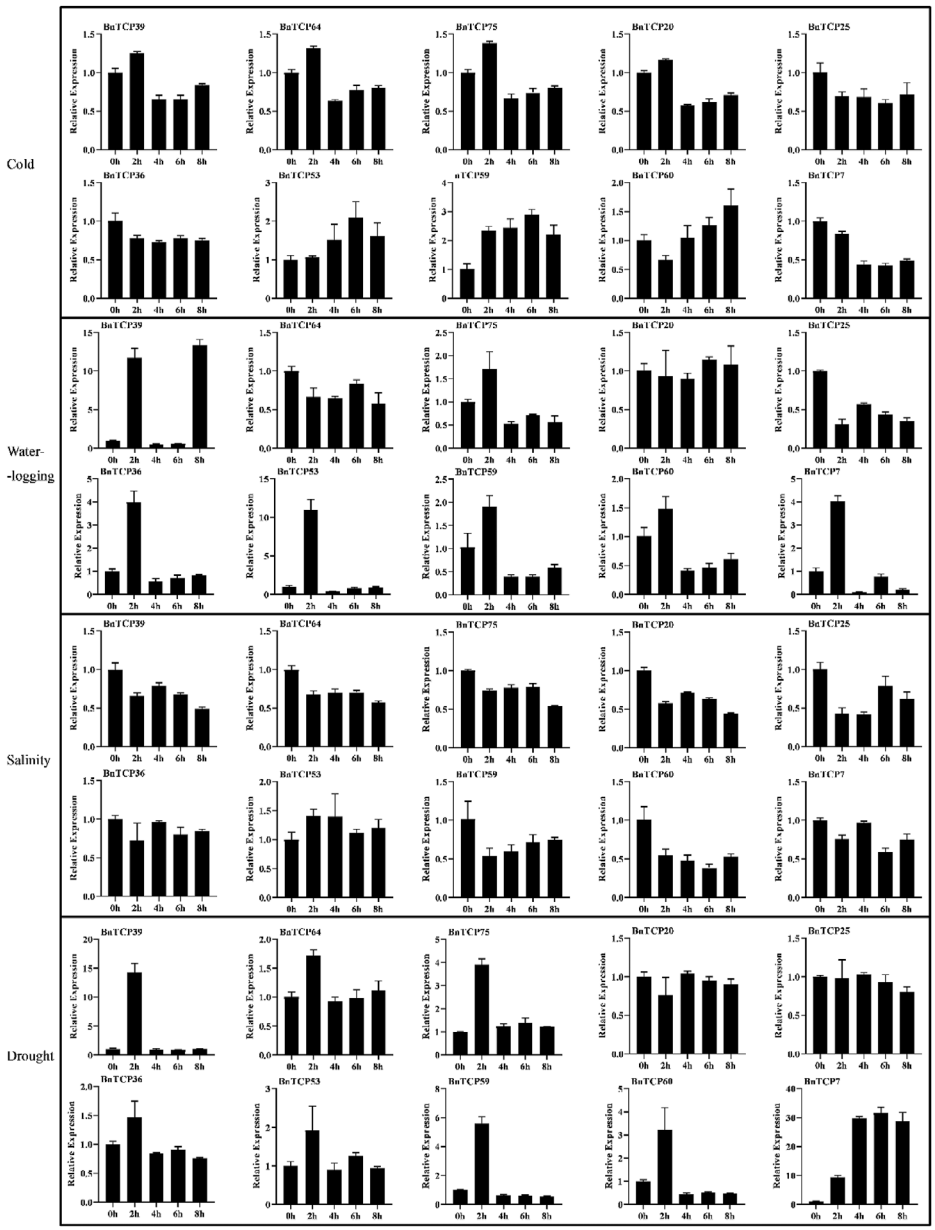


FIGURE 8 | Expression patterns of the *BnTCP* genes under different abiotic stress conditions. The 0 h (CK), 2, 4, 6, and 8 h labels presented the time points (hours) when the samples were harvested for expression study after the stress treatment. Error bars represent the SD (n = 3).

associated *cis*-elements in their promoter regions (**Supplementary Table S6**). Furthermore, *BnTCP39* under IAA treatment showed higher expression at 2 h (1.568 fold); *BnTCP64* significantly showed higher expression at 2 h (2.209 fold), 6 h (1.384 fold), and 8 h (1.590 fold) compared to CK. Notably, *BnTCP53* displays higher expression at 6 h (4.487 fold) in response to MeJA treatment than CK (**Figure 9**). However, *BnTCP60* did not reveal any significant expression to any stress conditions except cold stress, and *BnTCP53* was found to be more active in response to both hormones and abiotic conditions. These results are also consistent with the identified cis-elements in the promoter regions of genes.

DISCUSSION

Characterization and Evolution of *TCP* Gene Family in Rapeseed

Rapeseed is an allotetraploid crop that practices broad genome replication and combination activities (Song et al., 2020). Nevertheless, rapeseed yield is pretentious by numerous environmental factors such as cold, heat, drought, salinity, and heavy metals toxicity (Raza et al., 2021a; Raza et al., 2021b; He et al., 2021; Raza, 2021; Su et al., 2021). The TCP TFs are antique plant-specific proteins (Cubas et al., 1999). So far, these TFs are yet to

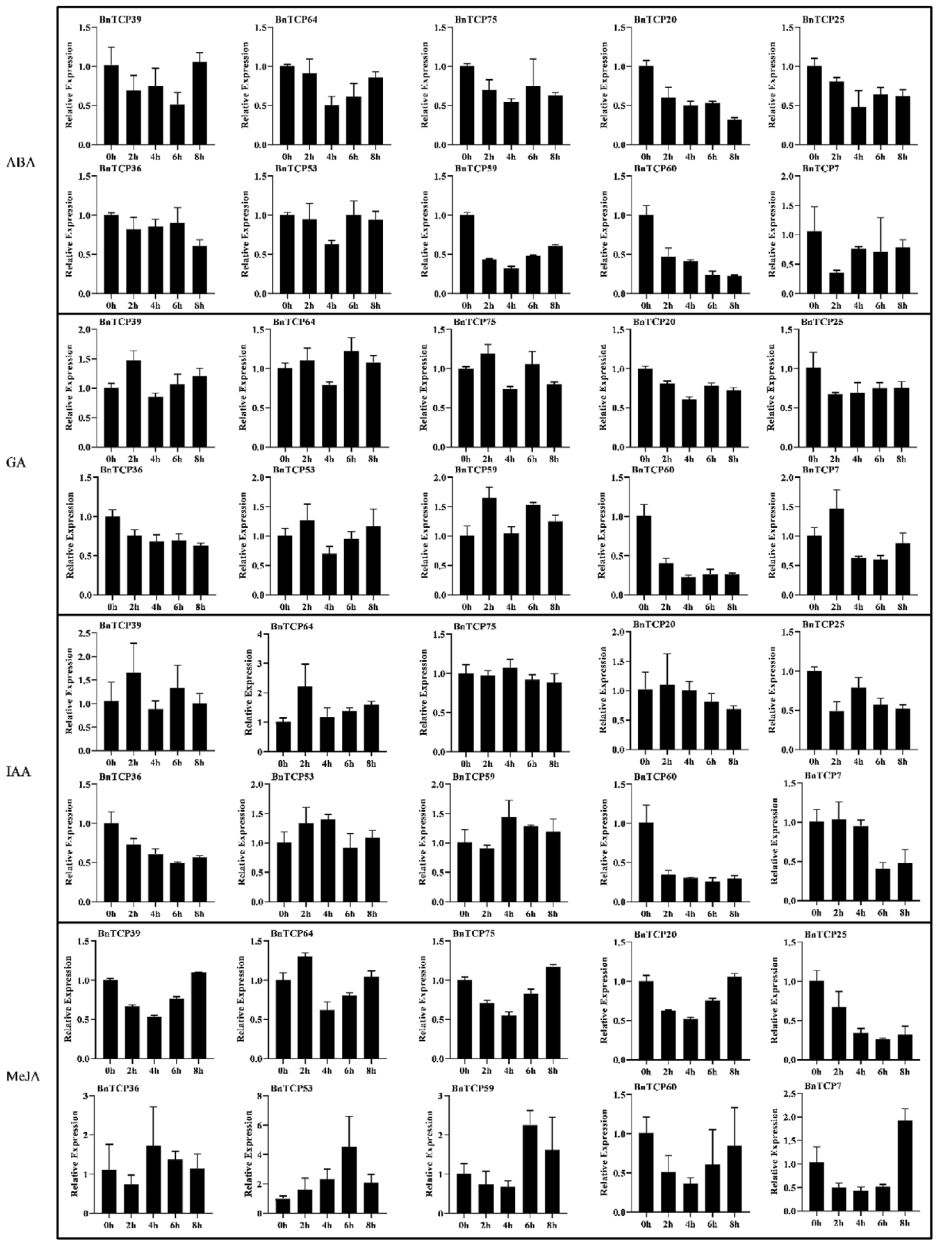


FIGURE 9 | Expression patterns of the *BnTCP* genes under different hormones treatments. The 0 h (CK), 2, 4, 6, and 8 h labels presented the time points (hours) when the samples were harvested for expression study after the stress treatment. Error bars represent the SD (n = 3).

be reported in unicellular algae; and moss, pluricellular green algae, lycophyte, and ferns possess five-six *TCP* genes in their genomes (Martín-Trillo and Cubas, 2010). Owing to the development, duplications, and divergence of *TCP* genes, the gymnosperms and angiosperms plants possess bigger *TCP* gene families containing more than ten *TCP* members (Martín-Trillo and Cubas, 2010). Over the past few years, *TCP* gene families have been reported in various crop plants such as *Arabidopsis thaliana* (Yao et al., 2007), Moso bamboo (Liu et al., 2018), sweet potato (Ren et al., 2021), cucumber (Wen et al., 2020), rice (Yao et al., 2007), *Brassica rapa* (Du et al., 2017); *Brassica juncea* (He et al., 2020); and cotton (Ma et al., 2014; Yin et al., 2018), etc. To our best knowledge, the *TCP* gene family has not been fully characterized in the rapeseed genome. The accessibility of the rapeseed genome sequences offers resources for genome-wide identification of rapeseed genes (Song et al., 2020). Here, we recognized 80 *BnTCP* genes in the rapeseed genome. Notably, this is the largest *TCP* gene family found in the rapeseed genome due to whole-genome duplication (WGD; polyploidy). Previously, the highest number of *TCP* genes (73 members) was reported in allotetraploid cotton (Yin et al., 2018).

In the rapeseed genome, the *BnTCP* gene family may be evolved by WGD or segmental events, together with numerous tandem

duplications. Notably, WGD differs from tandem duplication in that WGD enhances the quantity of all genes altogether and produces duplicate, theoretically redundant copies of all the genes within a genome. In contrast, tandem duplication is another vital way for gene expansion. Genes expanded by tandem duplication are always scattered all together as a cluster in chromosomes (Sémon and Wolfe, 2007; Fang et al., 2012). Gene duplication events of *TCP* genes were also spotted in several plant species (Ma et al., 2014; Lin et al., 2016; Yin et al., 2018; Wen et al., 2020). Apart from duplication events, exon-intron associations and numbers can also explain the evolutionary story of the gene family (Xu et al., 2012; Raza et al., 2021a; Su et al., 2021). The *BnTCPs* gene structure compared with the same group presented that *TCP* genes contribute to analogous exon/intron allocation in terms of exon and intron numbers; in the meantime, *BnTCPs* with the same class exhibited comparable motif patterns. These results imply that these *TCP* members may play identical roles in response to a variety of abiotic stresses. Similar observations have been reported in previous *TCP* gene families where both classes showed comparable diverse intron/exon distribution and motifs patterns (Ma et al., 2014; Lin et al., 2016; Yin et al., 2018; Wen et al., 2020).

The Contribution of *TCP* Genes to Stress Responses and Tolerance Mechanisms

To get further insights into the contribution of *BnTCP* genes against various environmental cues, *cis*-elements in the promoter regions were envisaged. The findings demonstrated that two types of *cis*-elements were documented, i.e., stress (salt, drought, low-temperature, and anaerobic) and hormones (auxin, ABA, MeJA, and GA) associated. Consistent with earlier studies, *cis*-elements participate in the plant stress responses (Yamaguchi-Shinozaki and Shinozaki, 1994; Liu et al., 2014).

For instance, in rice, TCA-element between −563 and −249 bp upstream of *OsPIANK1* was found to be associated with resistance to *Magnaporthe oryzae* infection and exogenous SA treatment (Mou et al., 2013). ABREs are a class of *cis*-elements capable of binding to sturdily conserved ABA-dependent transcription factors, which exist in the promoter region of numerous stress-resistant genes and normalize the expression level of associated genes under abiotic stress conditions (Choi et al., 2000; Barbosa et al., 2013; Fujita et al., 2013). In wheat, the expression pattern of *TaGAPC1* was regulated under osmotic and salinity stress conditions. The promoter region of *TaGAPC1* contains various stress-related *cis*-regulatory elements (including MBS, DRE, GT1, and LTR), and methylation changes were observed during the stress treatments, leading to stress tolerance and regulation in gene expression (Fei et al., 2017). In the current study, these results were further authenticated by the GO annotation evaluation. For instance, GO enrichment analysis validates the functional role of *BnTCP* genes in several cellulars, molecular, and biological processes associated with RNA/DNA binding, metabolic processes, transcriptional regulatory activities related to rapeseed growth and developmental processes under stress conditions. Likewise, numerous scholars stated comparable outcomes in diverse crop plant species where *TCP* genes were found to play a substantial part in plant growth and developmental processes and responses to different abiotic stress episodes (Yao et al., 2007; Lin et al., 2016; Zhou et al., 2016; Wen et al., 2020). These results can enhance our knowledge of *BnTCP* genes under different stress conditions and rapeseed developmental processes at a molecular level.

Moreover, 10 randomly selected *BnTCP* genes' expression profiling was appraised against various hormones and abiotic stress treatments. Particularly, *BnTCP7, BnTCP53,* and *BnTCP59* exhibited higher expression in response to cold, waterlogging, drought, and salinity stresses. Under hormones treatment, many genes showed higher expression, mainly by ABA, IAA, and GA treatments. These results are in agreement with recent reports where several *TCP* genes displayed higher expression levels under stress conditions. For example, in cotton, 41 *GhTCP* genes responded significantly to heat, salinity, and drought stresses (Yin et al., 2018). In cucumber, several *TCP* genes were up-regulated in response to photoperiod and temperature stresses; likewise, some genes were also induced by phytohormones like ethylene and gibberellin (Wen et al., 2020). In Moso bamboo, several *TCP* genes were substantially up-regulated in response to various hormones such as ABA, MeJA, and SA (Liu et al., 2018). The overexpression of maize *ZmTCP42* gene in *A. thaliana* led to sensitivity to ABA in seed germination and improved drought stress tolerance in transgenic *A. thaliana* (Ding et al., 2019). In another study, the overexpression of the *MeTCP4* gene regulated the cold tolerance by mediating ROS production and scavenging in transgenic *A. thaliana*. It also enhances the expression levels of stress- and ROS-scavenging-associated genes in the cold stressed *A. thaliana* plants (Cheng et al., 2019). Overexpressed *PeTCP10* gene improved salt and ABA tolerance of transgenic *A. thaliana* at the vegetative growth phase (Xu et al., 2021). These results indicate that *TCP* genes pointedly contribute to hormone signaling pathways and abiotic stress tolerance in various plant species.

Expression Pattern of *TCP* Genes in Different Developmental Tissues

In the present study, tissue-specific expression profiles of 80 *BnTCPs* genes were assessed in six unique tissues using a publicly available RNA-seq dataset (https://ngdc.cncb.ac.cn/gsa/; BioProject ID: PRJCA001495). The aforementioned studies have illustrated that *TCP* genes may exhibit diverse expression profiles in various developmental tissues. For instance, the RNA-seq data was used to analyze the tissue-specific expression levels in cotton. The results showed that many *GhTCP* genes exhibited higher expression in various tissues, including mature leaves, stem, root, torus, petal, stamen, pistil, calycle (bracts), ovules, and fibers (Yin et al., 2018). In cucumber, qRT-PCR-based expression analysis showed that almost all 27 *CsTCP* genes exhibited higher expression in diverse tissues comprising root, stem, leaf,

cotyledon, tendril, and flower (Wen et al., 2020). In *B. rapa*, some *BrTCP* genes showed higher expression in roots, leaves, and flower buds (Du et al., 2017). These results agree with our results where some *TCP* genes showed higher expression in all studied tissues (roots, stems, leaves, flowers, seeds, and silique), demonstrating that these genes may participate in rapeseed growth and development.

miRNA: Key Performers in the Adaptation of Stress Responses

MicroRNAs (miRNAs) are a class of small-non-coding RNAs produced from single-strand hairpin RNA precursors. These miRNAs control gene expression through binding to complementary sequences within target mRNAs (Shen et al., 2015; Wani et al., 2020). Substantial development has been made in discovering the targets of rapeseed miRNAs that act to diverse stresses and also participate in developmental processes (Buhtz et al., 2008; Huang et al., 2010; Körbes et al., 2012; Zhou et al., 2012; Shen et al., 2015; Chen et al., 2018; Fu et al., 2019). In the present study, we anticipated 32 miRNAs targeting 21 *BnTCPs* genes. Recent reports support our results, e.g., five *TCP* genes, including *TCP2, TCP3, TCP4, TCP10,* and *TCP24,* were targeted by miR319 and have been involved in normalizing leaf morphogenesis in *A. thaliana* (Schommer et al., 2008). In cotton, three *TCP* (*GhTCP21, GhTCP31,* and *GhTCP54*) genes were targeted by miR319 (Wen et al., 2020). The overexpression of the miR319-associated *TCP* gene improves the growth, and it also increases the tolerance to salinity and drought stresses in transgenic bentgrass (*Agrostis stolonifera* L.) (Zhou et al., 2013).

The overexpression of soybean miR172c improves tolerance to drought and salinity stress, and it also enhances sensitivity to ABA in transgenic *A. thaliana* plants (Li et al., 2016). Recently, Cheng et al. (2021) found that miR172s (mainly miR172a and miR172b) act as a positive regulator of salt tolerance in rice and wheat. Further, it also maintains the ROS homeostasis during salt stress, generally by balancing the expression of several ROS-scavenging-related genes in both wheat and rice (Cheng et al., 2021). The transgenic *B. napus* plants overexpressing miR395 indicated a lower degree of cadmium-induced oxidative stress compared to wild-type plants (Zhang et al., 2013). Another miRNA (miR394) was reported to enhance the multiple abiotic stress tolerance, including salinity and drought (Song et al., 2013) and cold stress (Song et al., 2016), in *A. thaliana* plants by ABA-dependent manner. These discoveries stipulated that miRNA might play critical roles in plant growth, development, and abiotic stress tolerance by regulating the expression levels of their target and stress-responsive genes.

CONCLUSION

In summary, for the first time, we identified 80 putative *BnTCP* genes in the rapeseed genome, which are distributed on 18 chromosomes. Genome-wide *in silico* analysis such as characterization, evolution, gene structures, conserved motifs, *cis*-elements, miRNA prediction, and GO annotation was performed to gain insights into *TCP* genes in rapeseed. Furthermore, their expression profiling was also appraised in different tissues and against diverse hormones and abiotic stresses. In a nutshell, this study laid the basis for additional functional studies (such as overexpression, knockout *via* CRISPR/Cas system, etc.) of the *BnTCP* genes in rapeseed breeding programs under adverse environmental conditions.

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

YW, AR, and WC conceived the idea, analyzed the data, and wrote the manuscript. QH, WW, XZ, HC, and JL supervised the research, reviewed and improved the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by the Central Public-interest Scientific Institution Basal Research Fund (Y2020YJ03), the Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ZDRW202105), and China Agriculture Research System of MOF and MARA (CARS-12) supported this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.794297/full#supplementary-material

REFERENCES

Aggarwal, P., Das Gupta, M., Joseph, A. P., Chatterjee, N., Srinivasan, N., and Nath, U. (2010). Identification of Specific DNA Binding Residues in the TCP Family of Transcription Factors in Arabidopsis. *Plant Cell* 22 (4), 1174–1189. doi:10.1105/tpc.109.066647

Barbosa, E. G. G., Leite, J. P., Marin, S. R. R., Marinho, J. P., de Fátima Corrêa Carvalho, J., Fuganti-Pagliarini, R., et al. (2013). Overexpression of the ABA-Dependent AREB1 Transcription Factor from *Arabidopsis thaliana* Improves Soybean Tolerance to Water Deficit. *Plant Mol. Biol. Rep.* 31 (3), 719–730. doi:10.1007/s11105-012-0541-4

Braun, N., de Saint Germain, A., Pillot, J.-P., Boutet-Mercey, S., Dalmais, M., Antoniadi, I., et al. (2012). The Pea TCP Transcription Factor PsBRC1 Acts Downstream of Strigolactones to Control Shoot Branching. *Plant Physiol.* 158 (1), 225–238. doi:10.1104/pp.111.182725

Buhtz, A., Springer, F., Chappell, L., Baulcombe, D. C., and Kehr, J. (2008). Identification and Characterization of Small RNAs from the Phloem of *Brassica Napus*. *Plant J.* 53 (5), 739–749. doi:10.1111/j.1365-313x.2007.03368.x

Chen, L., Chen, L., Zhang, X., Liu, T., Niu, S., Wen, J., et al. (2018). Identification of miRNAs that Regulate Silique Development in *Brassica Napus*. *Plant Sci.* 269, 106–117. doi:10.1016/j.plantsci.2018.01.010

Cheng, Z., Lei, N., Li, S., Liao, W., Shen, J., and Peng, M. (2019). The Regulatory Effects of MeTCP4 on Cold Stress Tolerance in *Arabidopsis thaliana*: a Transcriptome Analysis. *Plant Physiol. Biochem.* 138, 9–16. doi:10.1016/j.plaphy.2019.02.015

Cheng, X., He, Q., Tang, S., Wang, H., Zhang, X., and Lv, M. (2021). The miR172/IDS1 Signaling Module Confers Salt Tolerance through Maintaining ROS Homeostasis in Cereal Crops. *New Phytol.* 230 (3), 1017–1033. doi:10.1111/nph.17211

Choi, H.-i., Hong, J.-h., Ha, J.-o., Kang, J.-y., and Kim, S. Y. (2000). ABFs, a Family of ABA-Responsive Element Binding Factors. *J. Biol. Chem.* 275 (3), 1723–1730. doi:10.1074/jbc.275.3.1723

Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The TCP Domain: a Motif Found in Proteins Regulating Plant Growth and Development. *Plant J.* 18 (2), 215–222. doi:10.1046/j.1365-313x.1999.00444.x

Ding, S., Cai, Z., Du, H., and Wang, H. (2019). Genome-wide Analysis of TCP Family Genes in Zea mays L. Identified a Role for ZmTCP42 in Drought Tolerance. *Ijms* 20 (11), 2762. doi:10.3390/ijms20112762

Du, J., Hu, S., Yu, Q., Wang, C., Yang, Y., Sun, H., et al. (2017). Genome-wide Identification and Characterization of BrrTCP Transcription Factors in Brassica Rapa Ssp. Rapa. *Front. Plant Sci.* 8, 1588. doi:10.3389/fpls.2017.01588

Fang, L., Cheng, F., Wu, J., and Wang, X. (2012). The Impact of Genome Triplication on Tandem Gene Evolution in Brassica Rapa. *Front. Plant Sci.* 3, 261. doi:10.3389/fpls.2012.00261

Fei, Y., Xue, Y., Du, P., Yang, S., and Deng, X. (2017). Expression Analysis and Promoter Methylation under Osmotic and Salinity Stress of TaGAPC1 in Wheat (*Triticum aestivum* L). *Protoplasma* 254 (2), 987–996. doi:10.1007/s00709-016-1008-5

Fu, Y., Mason, A. S., Zhang, Y., Lin, B., Xiao, M., Fu, D., et al. (2019). MicroRNA-mRNA Expression Profiles and Their Potential Role in Cadmium Stress Response in *Brassica Napus*. *BMC Plant Biol.* 19 (1), 570. doi:10.1186/s12870-019-2189-9

Fujita, Y., Yoshida, T., and Yamaguchi-Shinozaki, K. (2013). Pivotal Role of the AREB/ABF-SnRK2 Pathway in ABRE-Mediated Transcription in Response to Osmotic Stress in Plants. *Physiol. Plantarum* 147 (1), 15–27. doi:10.1111/j.1399-3054.2012.01635.x

Gao, Y., Zhang, D., and Li, J. (2015). TCP1 Modulates DWF4 Expression *via* Directly Interacting with the GGNCCC Motifs in the Promoter Region of DWF4 in *Arabidopsis thaliana*. *J. Genet. Genomics* 42 (7), 383–392. doi:10.1016/j.jgg.2015.04.009

Gastaldi, V., Lucero, L. E., Ferrero, L. V., Ariel, F. D., and Gonzalez, D. H. (2020). Class-I TCP Transcription Factors Activate the SAUR63 Gene Subfamily in Gibberellin-dependent Stamen Filament Elongation. *Plant Physiol.* 182 (4), 2096–2110. doi:10.1104/pp.19.01501

He, J., He, X., Chang, P., Jiang, H., Gong, D., and Sun, Q. (2020). Genome-wide Identification and Characterization of TCP Family Genes in Brassica Juncea Var. Tumida. *PeerJ* 8, e9130. doi:10.7717/peerj.9130

He, H., Lei, Y., Yi, Z., Raza, A., Zeng, L., Yan, L., et al. (2021). Study on the Mechanism of Exogenous Serotonin Improving Cold Tolerance of Rapeseed (Brassica Napus L.) Seedlings. *Plant Growth Regul.* 94, 161–170. doi:10.1007/s10725-021-00700-0

Hoang, X. L. T., Nhi, D. N. H., Thu, N. B. A., Thao, N. P., and Tran, L. P. (2017). Transcription Factors and Their Roles in Signal Transduction in Plants under Abiotic Stresses. *Curr. Genomics* 18 (6), 483–497. doi:10.2174/1389202918666170227150057

Huang, S. Q., Xiang, A. L., Che, L. L., Chen, S., Li, H., Song, J. B., et al. (2010). A Set of miRNAs from *Brassica Napus* in Response to Sulphate Deficiency and Cadmium Stress. *Plant Biotechnol. J.* 8 (8), 887–899. doi:10.1111/j.1467-7652.2010.00517.x

Hurst, L. D. (2002). The Ka/Ks Ratio: Diagnosing the Form of Sequence Evolution. *Trends Genet.* 18 (9), 486–487. doi:10.1016/s0168-9525(02)02722-1

Karaaslan, E. S., Wang, N., Faiß, N., Liang, Y., Montgomery, S. A., Laubinger, S., et al. (2020). Marchantia TCP Transcription Factor Activity Correlates with Three-Dimensional Chromatin Structure. *Nat. Plants* 6 (10), 1250–1261. doi:10.1038/s41477-020-00766-0

Körbes, A. P., Machado, R. D., Guzman, F., Almerão, M. P., de Oliveira, L. F. V., Loss-Morais, G., et al. (2012). Identifying Conserved and Novel microRNAs in Developing Seeds of *Brassica Napus* Using Deep Sequencing. *PLoS One* 7 (11), e50663. doi:10.1371/journal.pone.0050663

Kosugi, S., and Ohashi, Y. (2002). DNA Binding and Dimerization Specificity and Potential Targets for the TCP Protein Family. *Plant J.* 30 (3), 337–348. doi:10.1046/j.1365-313x.2002.01294.x

Li, W., Wang, T., Zhang, Y., and Li, Y. (2016). Overexpression of Soybean miR172c Confers Tolerance to Water Deficit and Salt Stress, but Increases ABA Sensitivity in Transgenic *Arabidopsis thaliana*. *Exbotj* 67 (1), 175–194. doi:10.1093/jxb/erv450

Li, D., Zhang, H., Mou, M., Chen, Y., Xiang, S., Chen, L., et al. (2019). Arabidopsis Class II TCP Transcription Factors Integrate with the FT-FD Module to Control Flowering. *Plant Physiol.* 181 (1), 97–111. doi:10.1104/pp.19.00252

Lin, Y.-F., Chen, Y.-Y., Hsiao, Y.-Y., Shen, C.-Y., Hsu, J.-L., Yeh, C.-M., et al. (2016). Genome-wide Identification and Characterization ofTCPgenes Involved in Ovule Development ofPhalaenopsis Equestris. *Exbotj* 67 (17), 5051–5066. doi:10.1093/jxb/erw273

Liu, J.-H., Peng, T., and Dai, W. (2014). Critical Cis-Acting Elements and Interacting Transcription Factors: Key Players Associated with Abiotic Stress Responses in Plants. *Plant Mol. Biol. Rep.* 32 (2), 303–317. doi:10.1007/s11105-013-0667-z

Liu, H.-L., Wu, M., Li, F., Gao, Y.-M., Chen, F., and Xiang, Y. (2018). TCP Transcription Factors in Moso Bamboo (Phyllostachys Edulis): Genome-wide Identification and Expression Analysis. *Front. Plant Sci.* 9, 1263. doi:10.3389/fpls.2018.01263

Ma, J., Wang, Q., Sun, R., Xie, F., Jones, D. C., and Zhang, B. (2014). Genome-wide Identification and Expression Analysis of TCP Transcription Factors in Gossypium Raimondii. *Sci. Rep.* 4 (1), 6645. doi:10.1038/srep06645

Martín-Trillo, M., and Cubas, P. (2010). TCP Genes: a Family Snapshot Ten Years Later. *Trends Plant Sci.* 15 (1), 31–39. doi:10.1016/j.tplants.2009.11.003

Mou, S., Liu, Z., Guan, D., Qiu, A., Lai, Y., and He, S. (2013). Functional Analysis and Expressional Characterization of rice Ankyrin Repeat-Containing Protein, OsPIANK1, in Basal Defense against Magnaporthe Oryzae Attack. *PLoS One* 8 (3), e59699. doi:10.1371/journal.pone.0059699

Navaud, O., Dabos, P., Carnus, E., Tremousaygue, D., and Hervé, C. (2007). TCP Transcription Factors Predate the Emergence of Land Plants. *J. Mol. Evol.* 65 (1), 23–33. doi:10.1007/s00239-006-0174-z

Raza, A., Su, W., Gao, A., Mehmood, S. S., Hussain, M. A., Nie, W., et al. (2021a). Catalase (CAT) Gene Family in Rapeseed (Brassica Napus L.): Genome-wide Analysis, Identification, and Expression Pattern in Response to Multiple Hormones and Abiotic Stress Conditions. *Ijms* 22 (8), 4281. doi:10.3390/ijms22084281

Raza, A., Su, W., Hussain, M. A., Mehmood, S. S., Zhang, X., Cheng, Y., et al. (2021b). Integrated Analysis of Metabolome and Transcriptome Reveals Insights for Cold Tolerance in Rapeseed (*Brassica Napus* L.). *Front. Plant Sci.* 12, 721681. doi:10.3389/fpls.2021.721681

Raza, A. (2021). Eco-physiological and Biochemical Responses of Rapeseed (*Brassica Napus* L.) to Abiotic Stresses: Consequences and Mitigation Strategies. *J. Plant Growth Regul.* 40, 1368–1388. doi:10.1007/s00344-020-10231-z

Ren, L., Wu, H., Zhang, T., Ge, X., Wang, T., Zhou, W., et al. (2021). Genome-Wide Identification of TCP Transcription Factors Family in Sweet Potato Reveals Significant Roles of miR319-Targeted TCPs in Leaf Anatomical Morphology. *Front. Plant Sci.* 12, 686698. doi:10.3389/fpls.2021.686698

Sarvepalli, K., and Nath, U. (2018). CIN-TCP Transcription Factors: Transiting Cell Proliferation in Plants. *Iubmb Life* 70 (8), 718–731. doi:10.1002/iub.1874

Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E. E., et al. (2008). Control of Jasmonate Biosynthesis and Senescence by miR319 Targets. *Plos Biol.* 6 (9), e230. doi:10.1371/journal.pbio.0060230

Sémon, M., and Wolfe, K. H. (2007). Consequences of Genome Duplication. *Curr. Opin. Genet. Dev.* 17 (6), 505–512. doi:10.1016/j.gde.2007.09.007

Shen, E., Zou, J., Hubertus Behrens, F., Chen, L., Ye, C., Dai, S., et al. (2015). Identification, Evolution, and Expression Partitioning of miRNAs in allopolyploidBrassica Napus. *Exbotj* 66 (22), 7241–7253. doi:10.1093/jxb/erv420

Song, J. B., Gao, S., Sun, D., Li, H., Shu, X. X., and Yang, Z. M. (2013). miR394 and LCR Are Involved in Arabidopsis Salt and Drought Stress Responses in an

Abscisic Acid-dependent Manner. *BMC Plant Biol.* 13 (1), 210–216. doi:10.1186/1471-2229-13-210

Song, J. B., Gao, S., Wang, Y., Li, B. W., Zhang, Y. L., and Yang, Z. M. (2016). miR394 and its Target Gene LCR Are Involved in Cold Stress Response in Arabidopsis. *Plant Gene* 5, 56–64. doi:10.1016/j.plgene.2015.12.001

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., et al. (2020). Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of *Brassica Napus*. *Nat. Plants* 6 (1), 34–45. doi:10.1038/s41477-019-0577-7

Su, W., Raza, A., Gao, A., Jia, Z., Zhang, Y., Hussain, M. A., et al. (2021). Genome-Wide Analysis and Expression Profile of Superoxide Dismutase (SOD) Gene Family in Rapeseed (*Brassica Napus* L.) under Different Hormones and Abiotic Stress Conditions. *Antioxidants* 10 (8), 1182. doi:10.3390/antiox10081182

van Es, S. W., Silveira, S. R., Rocha, D. I., Bimbo, A., Martinelli, A. P., Dornelas, M. C., et al. (2018). Novel Functions of the Arabidopsis Transcription factorTCP5in Petal Development and Ethylene Biosynthesis. *Plant J.* 94 (5), 867–879. doi:10.1111/tpj.13904

Viola, I. L., Reinheimer, R., Ripoll, R., Manassero, N. G. U., and Gonzalez, D. H. (2012). Determinants of the DNA Binding Specificity of Class I and Class II TCP Transcription Factors. *J. Biol. Chem.* 287 (1), 347–356. doi:10.1074/jbc.m111.256271

Wang, M.-Y., Zhao, P.-M., Cheng, H.-Q., Han, L.-B., Wu, X.-M., Gao, P., et al. (2013). The Cotton Transcription Factor TCP14 Functions in Auxin-Mediated Epidermal Cell Differentiation and Elongation. *Plant Physiol.* 162 (3), 1669–1680. doi:10.1104/pp.113.215673

Wang, Z., Chen, Y., Fang, H., Shi, H., Chen, K., Zhang, Z., et al. (2014). Selection of Reference Genes for Quantitative Reverse-Transcription Polymerase Chain Reaction Normalization in *Brassica Napus* under Various Stress Conditions. *Mol. Genet. Genomics* 289 (5), 1023–1035. doi:10.1007/s00438-014-0853-1

Wani, S. H., Kumar, V., Khare, T., Tripathi, P., Shah, T., Ramakrishna, C., et al. (2020). miRNA Applications for Engineering Abiotic Stress Tolerance in Plants. *Biologia* 75 (7), 1063–1081. doi:10.2478/s11756-019-00397-7

Wen, H., Chen, Y., Du, H., Zhang, L., Zhang, K., He, H., et al. (2020). Genome-Wide Identification and Characterization of the TCP Gene Family in Cucumber (Cucumis Sativus L.) and Their Transcriptional Responses to Different Treatments. *Genes* 11 (11), 1379. doi:10.3390/genes11111379

Wu, J. F., Tsai, H. L., Joanito, I., Wu, Y. C., Chang, C. W., Li, Y. H., et al. (2016). LWD-TCP Complex Activates the Morning Gene CCA1 in Arabidopsis. *Nat. Commun.* 7 (1), 13181. doi:10.1038/ncomms13181

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of Duplicate Genes in Exon-Intron Structure. *Proc. Natl. Acad. Sci.* 109 (4), 1187–1192. doi:10.1073/pnas.1109047109

Xu, Y., Liu, H., Gao, Y., Xiong, R., Wu, M., Zhang, K., et al. (2021). The TCP Transcription Factor PeTCP10 Modulates Salt Tolerance in Transgenic Arabidopsis. *Plant Cel Rep.* 40 (10), 1971–1987. doi:10.1007/s00299-021-02765-7

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994). A Novel Cis-Acting Element in an Arabidopsis Gene Is Involved in Responsiveness to Drought, Low-Temperature, or High-Salt Stress. *Plant Cell* 6 (2), 251–264. doi:10.1105/tpc.6.2.251

Yao, X., Ma, H., Wang, J., and Zhang, D. (2007). Genome-Wide Comparative Analysis and Expression Pattern of TCP Gene Families in *Arabidopsis thaliana* and Oryza Sativa. *J. Integr. Plant Biol.* 49 (6), 885–897. doi:10.1111/j.1744-7909.2007.00509.x

Yin, Z., Li, Y., Zhu, W., Fu, X., Han, X., Wang, J., et al. (2018). Identification, Characterization, and Expression Patterns of Tcp Genes and Microrna319 in Cotton. *Ijms* 19 (11), 3655. doi:10.3390/ijms19113655

Zhang, L. W., Song, J. B., Shu, X. X., Zhang, Y., and Yang, Z. M. (2013). miR395 Is Involved in Detoxification of Cadmium in Brassica Napus. *J. Hazard. Mater.* 250-251, 204–211. doi:10.1016/j.jhazmat.2013.01.053

Zhang, W., Cochet, F., Ponnaiah, M., Lebreton, S., Mathéron, L., Pionneau, C., et al. (2019). The MPK 8- TCP 14 Pathway Promotes Seed Germination in Arabidopsis. *Plant J.* 100 (4), 677–692. doi:10.1111/tpj.14461

Zhou, Z. S., Song, J. B., and Yang, Z. M. (2012). Genome-wide Identification of *Brassica Napus* microRNAs and Their Targets in Response to Cadmium. *J. Exp. Bot.* 63 (12), 4597–4613. doi:10.1093/jxb/ers136

Zhou, M., Li, D., Li, Z., Hu, Q., Yang, C., Zhu, L., et al. (2013). Constitutive Expression of a miR319 Gene Alters Plant Development and Enhances Salt and Drought Tolerance in Transgenic Creeping Bentgrass. *Plant Physiol.* 161 (3), 1375–1391. doi:10.1104/pp.112.208702

Zhou, Y., Xu, Z., Zhao, K., Yang, W., Cheng, T., Wang, J., et al. (2016). Genome-wide Identification, Characterization and Expression Analysis of the TCP Gene Family in Prunus Mume. *Front. Plant Sci.* 7, 1301. doi:10.3389/fpls.2016.01301

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

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 November 2021
doi: 10.3389/fgene.2021.750137

Genome-Wide Identification and Expression Analysis of the MADS-Box Gene Family in Sweet Potato [*Ipomoea batatas* (L.) Lam]

*Zhengwei Shao, Minhong He, Zhipeng Zeng, Yanzhu Chen, Amoanimaa-Dede Hanna and Hongbo Zhu**

College of Coastal Agricultural Sciences, Guangdong Ocean University, Zhanjiang, China

MADS-box gene, one of the largest transcription factor families in plants, is a class of transcription factors widely present in eukaryotes. It plays an important role in plant growth and development and participates in the growth and development of flowers and fruits. Sweet potato is the seventh most important food crop in the world. Its tuberous roots, stems, and leaves contain a large number of proteins, lipids, carotenoids, anthocyanins, conjugated phenolic acids, and minerals, which have high edible, forage, and medicinal value, and is also an important energy crop. At present, MADS-box genes in sweet potato have rarely been reported, and there has been no study on the genome-wide identification and classification of MADS-box genes in *Ipomoea batatas*. This study provided the first comprehensive analysis of sweet potato MADS-box genes. We identified 95 MADS-box genes, analyzed the structure and protein of sweet potato MADS-box genes, and categorized them based on phylogenetic analysis with *Arabidopsis* MADS-box proteins. Chromosomal localization indicated an unequal number of MADS-box genes in all 14 chromosomes except LG3, with more than 10 MADS-box genes located on chromosomes LG7, LG11, and LG15. The MADS domain and core motifs of the sweet potato MADS-box genes were identified by motif analysis. We identified 19 MADS-box genes with collinear relationships and analyzed duplication events. *Cis*-acting elements, such as light-responsive, auxin-responsive, drought-inducible, and MeJA-responsive elements, were found in the promoter region of the MADS-box genes in sweet potato, which further indicates the basis of MADS-box gene regulation in response to environmental changes and hormones. RNA-seq suggested that sweet potato MADS-box genes exhibit tissue-specific expression patterns, with 34 genes highly expressed in sweet potato flowers and fruits, and 19 genes highly expressed in the tuberous root, pencil root, or fibrous root. qRT-PCR again validated the expression levels of the 10 genes and found that *IbMADS1*, *IbMADS18*, *IbMADS19*, *IbMADS79*, and *IbMADS90* were highly expressed in the tuberous root or fibrous root, and *IbMADS18*, *IbMADS31*, and *IbMADS83* were highly expressed in the fruit. In this study, the molecular basis of MADS-box genes of sweet potato was analyzed from various angles. The effects of MADS-box genes on the growth and development of sweet potato were investigated, which may provide a certain theoretical basis for molecular breeding of sweet potato.

Keywords: sweet potato, MADS-box, MIKC type, bioinformatics analysis, growth development, expression analysis

OPEN ACCESS

Edited by:
Himabindu Kudapa, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Muhammad Ramzan Khan, Quaid-i-Azam University, Pakistan
Jin Xiao, Nanjing Agricultural University, China

****Correspondence:***
Hongbo Zhu
tdzhu@126.com

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

Received: *30 July 2021*
Accepted: *05 October 2021*
Published: *17 November 2021*

Citation:
Shao Z, He M, Zeng Z, Chen Y, Hanna A-D and Zhu H (2021) Genome-Wide Identification and Expression Analysis of the MADS-Box Gene Family in Sweet Potato [Ipomoea batatas (L.) Lam]. Front. Genet. 12:750137. doi: 10.3389/fgene.2021.750137

1 INTRODUCTION

MADS-box family gene is a kind of transcription factor (TF) widely existing in eukaryotes. MADS is the first letter abbreviation of Mini chromosome maintenance 1 (MCM1) gene in *Saccharomyces cerevisiae* (Passmore et al., 1988), AGAMOUS (AG) gene in *Arabidopsis thaliana* (Yanofsky et al., 1990), DEFICIENS (DEF) gene in *Antirrhinum majus* (Sommer et al., 1990), and serum response factor (SRF) gene in human serum (Norman et al., 1988). It is one of the largest TF families in plants and plays an important role in plant growth and development. Based on phylogenetic analysis, MADS-box genes in plants are categorized into type I and type II (Parenicova et al., 2003). Type I MADS-box gene, also known as M Type, usually has one to two exons, encoding proteins with highly conserved MADS domain. According to the differences in MADS domain, it is further divided into Mα, Mβ, and Mγ subfamilies (De Bodt et al., 2003). Type II gene, also known as MIKC type MADS-box gene, is a kind of plant-specific MADS-box gene with six to eight exons. The encoded protein contains four conserved domains: MADS-box (M domain), Intervening domain (I domain), Keratin-like domain (K domain), and C-terminal domain (Nam et al., 2005).

The completion of the whole genome sequences of many plant species has led to the identification and characterization of important gene families. MADS-box genes are integral to the ABCDE model of flowering regulation. Fan et al. identified three flower development class E genes, *AGL2*, *AGL4*, and *AGL9* (Fan et al., 1997). Class E genes cooperate with several other classes to control calyx, petal, stamen, carpel, and ovule development (Theissen, 2001). *AP3* homolog genes in *Chrysanthemum* play a role in the development of ray floret and disc floret in chrysanthemum (Won et al., 2021), and nitrate signaling-associated TFs such as *AGL8*, *AGL21*, and *LBD29* are involved in CmANR1-modulated control of root development. In addition, *CmANR1* also acts as a positive regulator to control shoot growth and development (Sun et al., 2021). Wang et al. found 15 tomato MADS-box genes involved in floral organ identification and five tomato MADS-box genes related to fruit development (Wang et al., 2019). *AGL17* towards the elongation zone is expressed in lateral root cap epidermal cells; *AGL12* and *AGL21* are also expressed in the central cylinder of differentiated roots. Both are expressed in developing embryos of *Arabidopsis* (Burgeff et al., 2002).

Sweet potato [*Ipomoea batatas* (L.) Lam.], 2n = 6*x* = 90, is an annual herb, and its underground parts are round, oval, or spindle-shaped tuberous roots. The shape, skin color, and flesh color of the tuberous root vary with varieties or soil (Wu et al., 1998). Sweet potato is also the seventh important food crop in the world, with high yield and wide use. The tuberous roots, stems, and leaves contain a large number of proteins, lipids, carotenoids, anthocyanin, conjugated phenolic acids, and minerals, which have high edible, feed, and medicinal values, and is also essential energy crop (Wang et al., 2016). In 2002, Kim et al. cloned two MADS-box genes named *IbMADS3* and *IbMADS4* from sweet potato for the first time and found that these two MADS-box genes were significantly expressed in pencil roots and tuberous roots of sweet potato (Kim et al., 2002). In 2005, Kim et al. reported three MADS-box genes of *IbMADS79*, *IbAGL17*, and *IbAGL20*; and they found that *IbMADS79* was only expressed in roots of sweet potato (Kim et al., 2005). Subsequently, Ku et al. (2008) cloned *IbMADS1*, indicating that *IbMADS1* was involved in the initiation of tuberous root differentiation of sweet potato. Dong et al. (2018) cloned nine MADS-box TFs in 2018 and found that *IbMBP2*, *IbMBP3*, *IbMBP4*, and *IbMBP9* showed high expression during storage roots development and were upregulated by storage roots development-related hormones. In 2020, Zhu et al. identified 37 MADS-box genes in *Ipomoea trifida*, the close wild ancestor of sweet potato. They found that MIKCC genes may be significant for regulating the floral organ development in *I. trifida* (Zhu et al., 2020). The MADS-box genes within the sweet potato genome have not been completely mined; the function and mechanism of action need to be verified; and the identification of more MADS-box genes is beneficial to elucidating the molecular mechanism of growth and development of sweet potato, providing a certain theoretical basis for molecular breeding.

2 MATERIALS AND METHODS

2.1 Identification and Gene Structure Analysis of MADS-Box Family Members in Sweet Potato

The sweet potato genomic information was derived from *I. batatas* "Taizhong 6" genomic data provided by the Ipomoea genome hub (https://ipomoea-genome.org/). The MADS-box family HMM model files SRF-TF (PF00319) and K-box (PF01486) were downloaded from the Pfam database (https://pfam.xfam.org/). The whole genome MADS-box family genes of sweet potato were identified by HMMER software (Krejci et al., 2016), and the MADS-box genes were identified by the SMART website (http://smart.embl-heidelberg.de/) again. The genes containing the MADS-box domain were named from *IbMADS1* to *IbMADS95*, and the three genes matching the published sequences of *IbMADS1* (Ku et al., 2008), *IbMADS4* (Kim et al., 2002), and *IbMADS79* (Kim et al., 2005) were still named as *IbMADS1*, *IbMADS4*, and *IbMADS79*, respectively. *IbMBP1*, *IbMBP4*, and *IbMBP8* (Dong et al., 2018) were named as *IbMADS47*, *IbMADS38*, and *IbMADS51*, respectively, in this study. According to the sweet potato genome gff3 annotation information, GSDS online website was used (http://gsds.gao-lab.org/). The MADS-box gene structure of sweet potato was analyzed visually. TBtools (Chen et al., 2020) was used to plot the innovative prediction results visually. TBtools (Chen et al., 2020) was used for structure visualization. According to the annotation information of the sweet potato genome, the MADS-box genes structure of sweet potato was visualized using GSDS online website tool (http://gsds.gao-lab.org/).

2.2 MADS-Box Phylogenetic Tree Construction

The *Arabidopsis* MADS-box protein sequence was obtained from the Plant TFDB website (http://planttfdb.gao-lab.org/index.php) (Jin et al., 2017). *I. trifida* MADS-box proteins were obtained according to the research of Zhu et al. (Zhu et al., 2020). The MADS-box genes of sweet potato were divided into two categories, type I and type II, based on the phylogenetic analysis of the MADS gene with *Arabidopsis*. ClustalW was used to compare the two types of MADS-box proteins; a neighbor-joining (NJ) tree was constructed using bootstrap analysis with 1,000 replicates, the amino acid p-distance model, missing data treatment option set at partial deletion, and a site coverage cutoff of 60%; and MEGA-X (Kumar et al., 2018) was used to construct the NJ phylogenetic tree (Saitou and Nei, 1987). The parameters for constructing the evolutionary tree were as follows: Poisson model, uniform rates, Sam (homogeneous), and pairwise deletion. According to the classification method of Sheng et al. (2019), the MADS-box gene subfamily of sweet potato was categorized based on the MADS-box gene subfamily of *Arabidopsis*, using Evolview online tools (https://www.evolgenius.info/evolview/) to beautify the tree.

2.3 Chromosomal Localization of MADS-Box Genes

According to the annotation information of the sweet potato genome downloaded from the Ipomoea genome hub, the positions of 95 sweet potato MADS-box genes on chromosomes were obtained, and the chromosome mapping was performed using Map Chart (Voorrips, 2002) software. Origin 8.5 (https://www.originlab.com) was used to draw the distribution cake map of sweet potato MADS-box genes.

2.4 Conserved Motif Analysis of MADS-Box Proteins

The sweet potato MADS-box protein sequence was submitted to MEME online tool (http://meme-suite.org/) (Bailey et al., 2009) for the motif prediction. The discovery number of Motif was set to be 8, the width of motif was 6–50, and other default settings were used. The MADS-box motif domain of sweet potato was visualized by TBtools.

2.5 *Cis*-Acting Elements Analysis of MADS-Box Genes

According to the annotation information of the sweet potato genome and sweet potato genome sequence, referring to Sumiya (2021) and Zhang et al. (2021), the 2-kb upstream sequence of sweet potato MADS-box genes was submitted to PlantCare website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002) for *cis*-acting element prediction. The PlantCare analysis results were simplified, and TBtools was used for visualization.

2.6 Collinearity Analysis of MADS-Box Genes

According to the annotation information and the whole genome protein sequence of sweet potato, MCScanX (Wang et al., 2012) software was used to analyze the MADS-box collinearity of sweet potato, and Circos (Krzywinski et al., 2009) software was used for visualization.

2.7 Expression Analysis of MADS-Box Genes

The tuberous root (greater than 10 mm in diameter), pencil root (2–5 mm in diameter), fibrous root (diameter of less than 1.5 mm), flower (an incompletely opened bud), fruit (the peel is green and the fruit form full), and stem tip (1–2 rachises of the rolled leaf at the stem apex, excluding the rachis where the expanded leaf is located) of "Jishu 26" (*I. batatas* cv. Ji 26) cultivated in an open field for 100 days were selected. After rapid freezing in liquid nitrogen, with dry ice sent to Biomarker Technologies Company (Beijing, China, https://www.biocloud.net) for total RNA extraction, library construction, and transcriptome sequencing, raw data were uploaded to National Center for Biotechnology Information (NCBI) project PRJNA744414; the expression of sweet potato MADS-box genes was analyzed based on genome information and local blast; the FPKM value (Florea et al., 2013) was calculated to reflect the gene expression; and the calculation formula is as follows: FPKM = {cDNA Fragments\over{Mapped Fragments(Millions) * TranscriptLength(kb)}}. Pheatmap package and ggplot2 package (Ito and Murphy, 2013) of R Statistics (https://www.r-project.org) were used for data processing and visualization.

2.8 Verification of Gene Expression by qRT-PCR

The tuberous root is the main harvest organ of sweet potato. According to the expression profile of the MADS-box gene in different parts, several genes highly expressed in tuberous root were selected for qRT-PCR expression verification. The double-stranded cDNA of fibrous root (F.R, sampling details were the same as 1.7), tuberous root (T.R, sampling details were the same at 1.7), leaf (L, expanded leaf 1–2 at the apex of stem), stem (S, 1–2 rachises of the expanded leaf at the stem apex, excluding the rachis where the rolled leaf is located), stem tip (S.T, sampling details were the same at 1.7), fruit (F.T, sampling details were the same at 1.7), and other tissues cultivated in open field for 100 days "Jishu 26" were used as templates; *β-actin* was used as a reference gene, using the CFX96™ quantitative PCR instrument to perform the fluorescent quantitative PCR. The reaction program was as follows: predenaturation at 95°C for 3 min, 95°C for 5 s, 60°C for 30 s, 40 cycles; and 95°C for 15 s, 60°C for 6 s, 95°C for 15 s. The reaction system was as follows: 2× universal SYBR green fast qPCR mix (ABclonal Technology, Wuhan, China) for 5 μl, ddH_2O for 3 μl, cDNA for 1 μl, forward primer for 0.5 μl, and reverse primer for 0.5 μl. Primers were designed using Primer premier 6 (Singh et al., 1998) based on the sequence of the amplified CDs of the

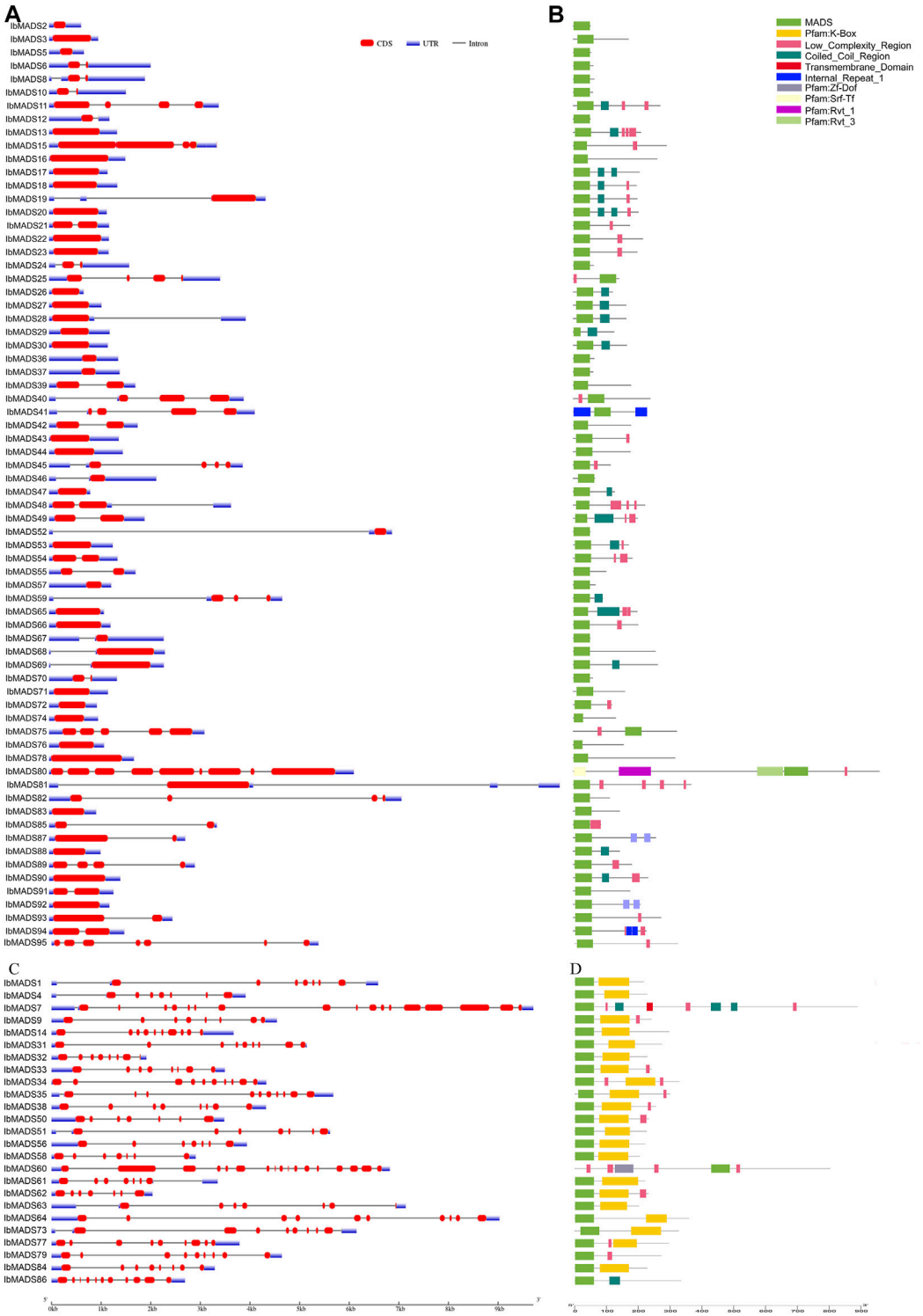


FIGURE 1 | SMART identification and structure of sweet potato MADS-box gene. **(A)** Type I gene structure. **(B)** Type I protein structure prediction by SMART. **(C)** Type II gene structure. **(D)** Type II protein structure prediction by SMART.

MADS-box genes, and the primer sequences are listed in **Supplementary Table S1**. Three biological replicates and three technical replicates were set up, and relative gene expression was calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmitten, 2001). Data processing and analysis were performed using SPSS software, and column graphs were drawn using origin 8.5.

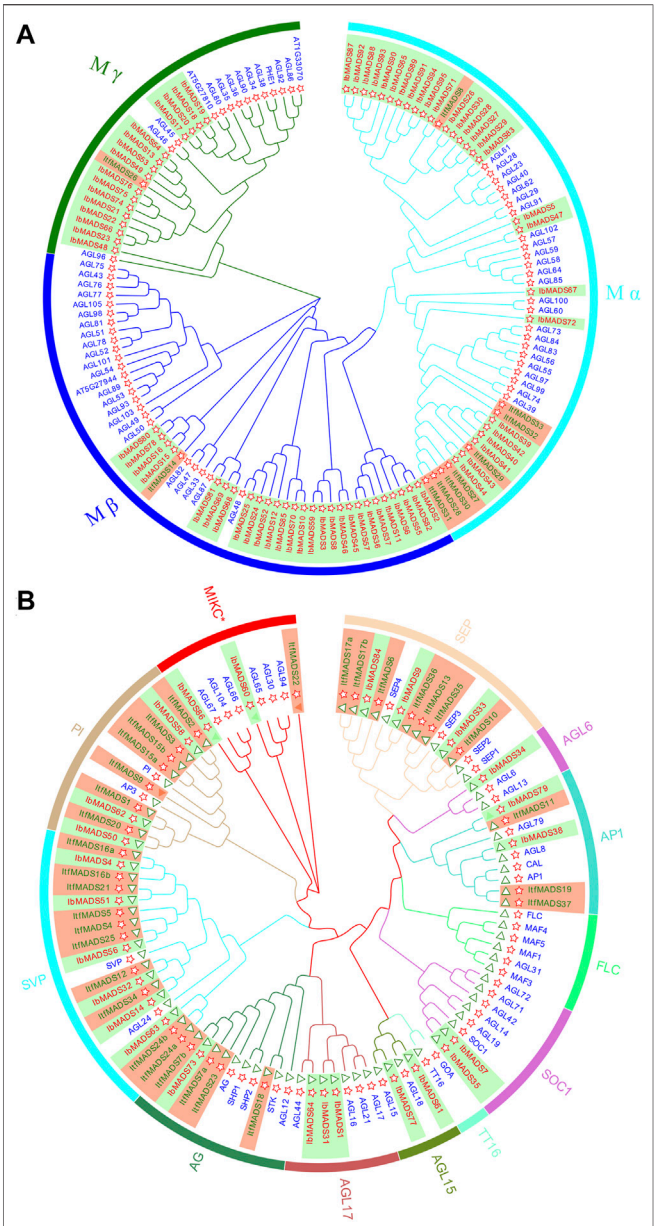


FIGURE 2 | Phylogenetic tree of type I **(A)** and type II **(B)** MADS-box proteins in *Ipomoea batatas*, *Ipomoea trifida*, and *Arabidopsis*. Red ID indicates *I. batatas* gene, green ID indicates *I. trifida* gene, and blue ID indicates *Arabidopsis* gene. The red star indicates the sequence containing the SRF-box domain, and the green triangle indicates the sequence containing K-Box domain.

3 RESULTS AND ANALYSIS

3.1 Identification and Structural Analysis of the MADS-Box Family of Sweet Potato

A total of 95 MADS-box genes were identified from the whole genome of sweet potato and named according to *IbMADS1–IbMADS95* (**Supplementary Table S2**). According to the gene structure, the sweet potato MADS-box genes were divided into two categories: type I and type II (**Figure 1**). The different types of MADS-box gene structures of sweet potato have apparent differentiation. Most type I genes are less than 2 kb in length and have one to two long fragment exons. On the other hand, the type II gene is usually over 3 kb and contains more than six exons, and most exonic fragments are short. Smart protein structure prediction found that the sweet potato MADS-box gene contains other domains in addition to the MADS domain. Most type I genes contained variable numbers of low-complexity region (LCR) structures and coiled-coil region (CCR) structures. The 21 type II genes contained a K-box domain; the K-box domain was commonly found to be associated with SRF-type TFs, also the signature domain of type II MADS-box genes. And the type II genes without a K-box domain usually belong to the MIKC* subfamily.

3.2 Phylogenetic Analysis of MADS Proteins

The NJ phylogenetic tree of MADS-box proteins from *I. batatas*, *I. trifida*, and *Arabidopsis* were constructed by MEGA-X. Based on the subfamily classification of *Arabidopsis* MADS proteins, sweet potato type I MADS proteins were divided into three subfamilies: Mα, Mβ, and Mγ (**Figure 2A**). In comparison, type II MADS proteins were divided into 12 subfamilies (**Figure 2B**). The type I MADS proteins of sweet potato and *Arabidopsis* were cross-distributed in each subfamily. Among them, both Mα subfamily and Mβ subfamily had 27 proteins each, and Mγ subfamily was less abundant with 16 proteins, which was similar to the distribution of the *Arabidopsis* protein subfamily. However, the type I proteins of *I. trifida* were mainly concentrated in Mα subfamily, with only one gene each in the Mβ and Mγ subfamilies.

The type II gene of *I. batatas* is phylogenetically closer to that of *I. trifida*, as indicated by gene number and subfamily distribution. The type II gene of *I. batatas* did not appear in the FLC subfamily or TT16 subfamily, and other subfamilies were distributed, with one to six proteins varying in number. The proteins for *I. trifida* are mainly distributed in subfamilies such as SEP, PI, and SVP and show deletions in several subfamilies. Twenty-two type II proteins contain K-box region, while three sweet potato type II proteins have no K-box region, one of them belongs to the AP1 subfamily, and the other two belong to the MIKC* subfamily.

3.3 Chromosomal Localization of MADS-Box Genes

Chromosome localization of 95 MADS-box genes in sweet potato was performed (**Figure 3**). It was found that except LG3 chromosome, the distribution of MADS-box genes in sweet potato varied in quantity in the other 14 chromosomes. More than 10 MADS-box genes were distributed on LG7, LG8, LG11, and LG15 chromosomes. LG1, LG5, and LG12 chromosomes contain fewer MADS-box genes, only one to three. Mα subfamily is mainly distributed on chromosomes LG7, LG8, and LG15, with more than five genes. In contrast, the Mβ subfamily (blue) is more fragmented, with one to four genes distributed in all chromosomes except LG16 and LG12. Seven MADS-box genes on LG6 were Mγ subfamily, and the rest were sporadically distributed on LG4, LG9, LG10, LG12, and LG14. The MIKC subfamily genes are widely distributed, except for LG1, LG3, LG6, LG9, and LG12, which are more numerous on chromosome LG11, with seven genes belonging to

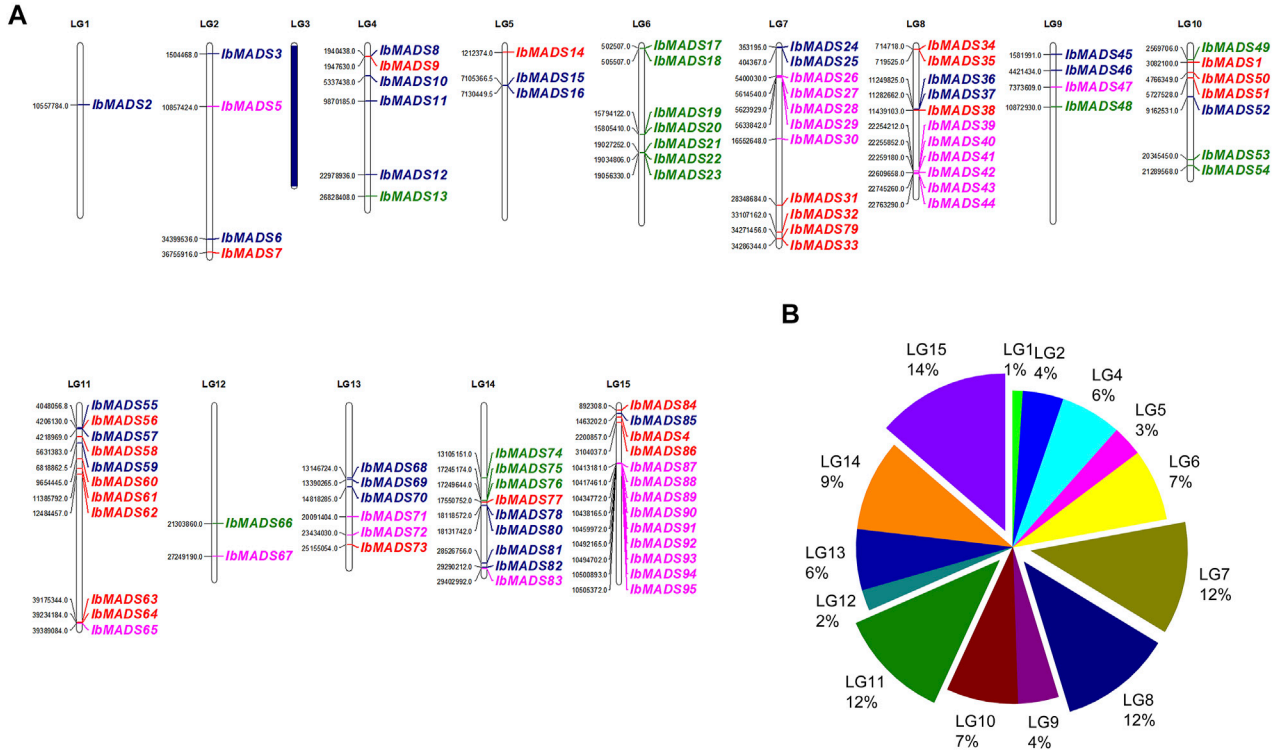


FIGURE 3 | Chromosome location **(A)** and distribution **(B)** of MADS-box gene in sweet potato. The purple name represents Mα subfamily, the blue name represents Mβ subfamily, the green name represents Mγ subfamily, and the red name represents type II MADS-box genes. LG3 has no MADS-box gene distribution.

the MIKC subfamily. Gene cluster exists for LG6, LG7, LG8, LG14, and LG15, especially in the interval 10413181 to 10505372 of LG15 containing nine MADS-box genes. Gene clusters were present on chromosomes LG6, LG7, LG8, LG14, and LG15. In particular, the interval 10413181 to 10505372 of LG15 contains nine MADS-box gene presence distributed centrally.

3.4 Analysis of MADS Proteins Conserve Motif

Conserve motif analysis of sweet potato MADS-protein sequence was performed by online MEME tool. There are some differences in the number of MADS-protein motif in sweet potato. Each protein contains two to six motifs (**Figure 4**). All genes have motif 1 domain encoding 28 amino acids, and more than 90 genes contain motif 2 and motif 5. The 51 amino acids encoded by these three motifs are the core motifs of MADS-box genes, which are related to the SRF-box region. In addition, the motif 3 domain also plays an important role, and more than 50 proteins contain highly conserved motif 3 domains encoding 29 amino acids, including 22 type II MADS proteins.

3.5 Analysis of *Cis*-Acting Elements of MADS-Box Gene Promoter Regions

The 2-kb upstream CDS sequences were extracted from the promoter regions of 95 MADS-box genes of sweet potato. The *cis*-acting elements in the promoter region of sweet potato MADS-box were predicted by PlantCARE online tool. All sweet potato MADS-box genes contain three to 12 light-responsive elements, including Box 4, G-box, GT1-motif, and TCT-motif. In addition, the promoter regions of sweet potato MADS-box genes are mainly more than 200 MeJA-responsive elements, including CGTCA-motif and TGACG-motif. And ABA-responsive elements (ABRE), auxin-responsive elements (TGA-element and AuxRR-core), and about 100 MYB response elements are involved in drought inducibility and light-responsiveness. In addition, there are a small number of regulatory elements such as circadian control elements, meristem expression elements, salicylic acid-responsive elements, low-temperature-responsive elements, gibberellin-responsive elements, defense, and stress-responsive elements. They indicated that MADS-box genes could respond to various environmental changes and coordinate the average growth and development of sweet potato (**Figure 5**).

3.6 Collinearity Analysis of MADS-Box Gene in Sweet Potato

Based on the collinearity analysis by MCScanX, a large number of genes were found to be collinear relational (light pink and light green) between and within chromosomes in sweet potato. There are 19 sweet potato MADS-box genes with collinearity, of which 15 genes belonged to inter-chromosome segmental duplications and four genes belonged to tandem duplications. Among them, *IbMADS4*, *IbMADS14*, *IbMADS32*, and *IbMADS56* genes share a

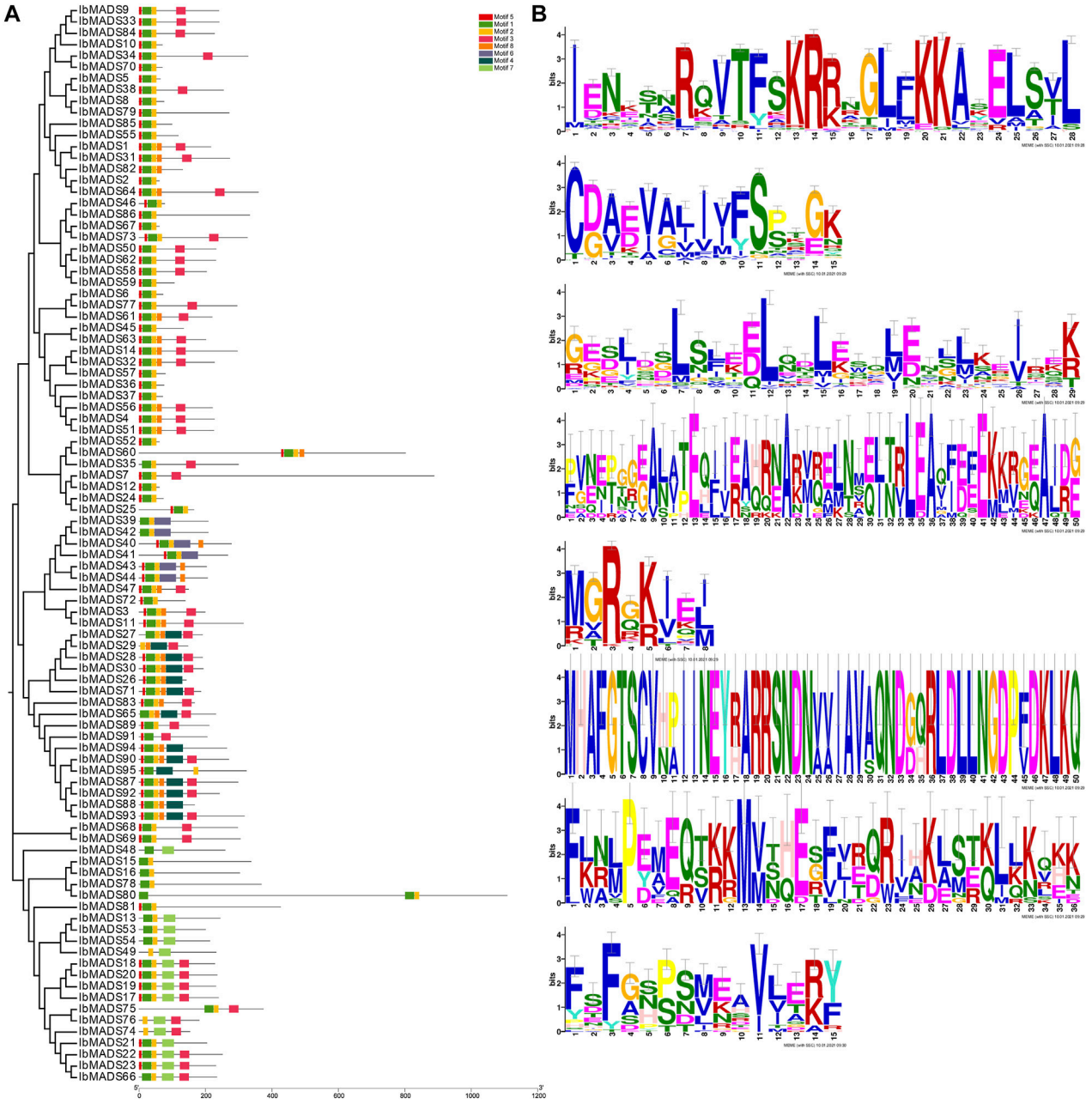


FIGURE 4 | Analysis of MADS-protein motif domain in sweet potato **(A)** and motifs **(B)**.

reciprocal segmental duplication relationship, with *IbMADS51* showing a segmental duplication relationship with *IbMADS56* and *IbMADS4* but not with *IbMADS14* or *IbMADS32*. There are tandem duplications between *IbMADS58* and *IbMADS59*, and between *IbMADS68* and *IbMADS69*. And there are segmental duplications in the other 10 intergenic pairs. The most MADS-box genes with collinear relationships were on chromosome LG11 with five, and the remaining chromosomes had fewer than two or no MADS-box genes with collinear relationships (**Figure 6**).

3.7 Expression Profiles of MADS-Box Genes in Sweet Potato

To investigate the expression patterns of MADS-box genes in sweet potato, the 75 MADS-box gene expression profiles were detected in six different tissues by transcriptome sequencing. The results showed that sweet potato MADS-box genes were expressed and differentiated in different tissues (**Figure 7**). Cluster analysis clustered the sweet potato MADS-box gene expression profiles into seven classes, with class I, II, and III genes mainly expressed in different types of roots. Among them,

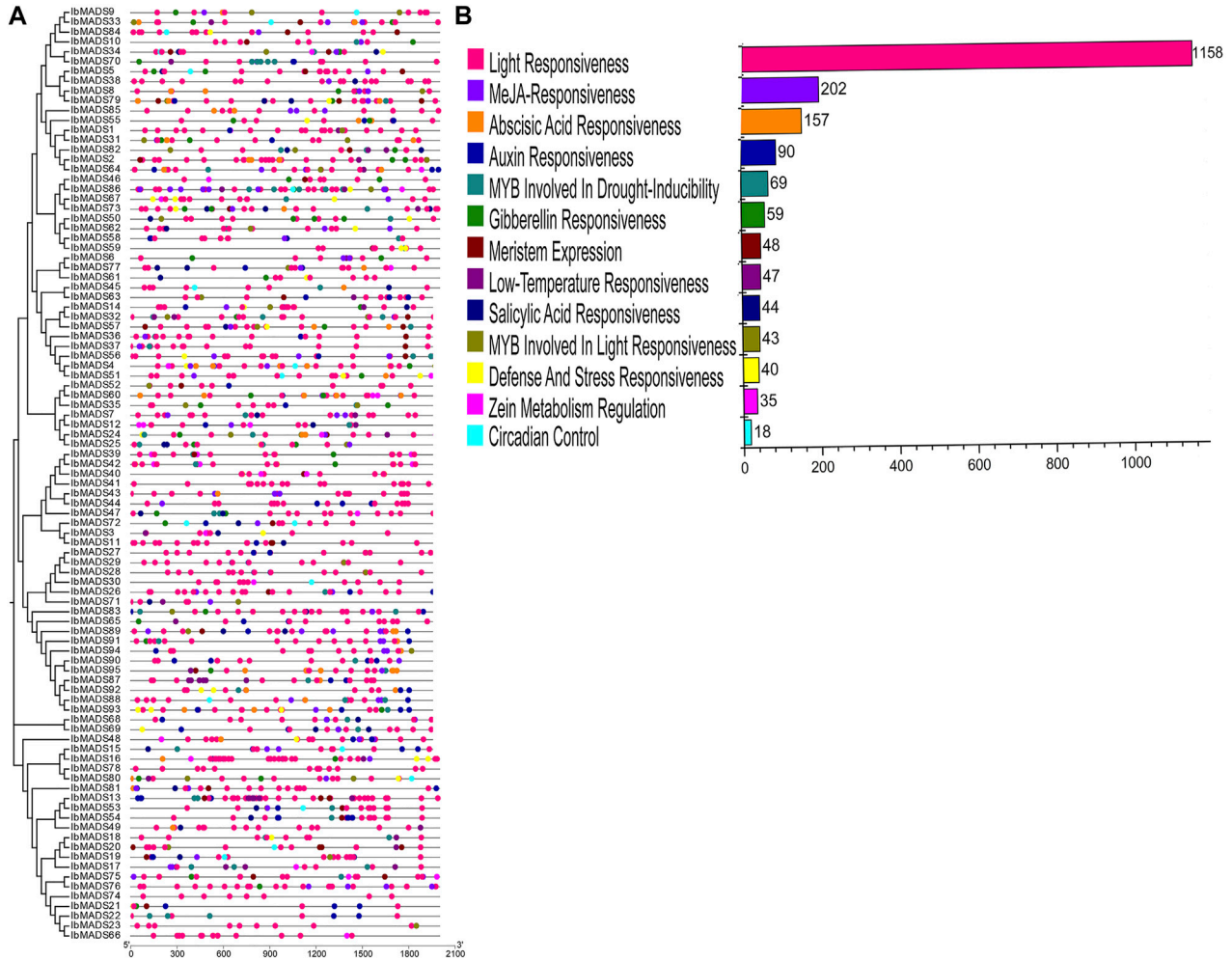


FIGURE 5 | *Cis*-acting elements analysis of sweet potato MADS-box. **(A)** *Cis*-acting element distribution. **(B)** *Cis*-acting element numbers.

seven genes in class II were highly and specifically expressed in the tuberous root. The class IV genes were expressed in the stem tip and fibrous root. The class V genes were highly and specifically expressed in the stem tip. The class VI genes were predominantly expressed in flower, and a small proportion of these genes were also somewhat expressed in fruits. The class VII genes are highly and specifically expressed in fruit. Overall, 22 MADS-box genes were highly expressed in flowers, 19 were highly expressed in stem tips, 12 were highly expressed in fruits, and the remaining genes were differentially expressed in different types of roots.

3.8 Expression Analysis of MADS-Box Genes by qRT-PCR

qRT-PCR was used to further validate the expression profiles of the 10 genes, which were more highly expressed in roots than other tissues, according to transcriptome sequencing of MADS-box genes in different tissues (**Figure 8**). Ten genes showed tissue-specific expression at different levels in all tissues; and *IbMADS1*, *IbMADS18*, *IbMADS19*, *IbMADS79*, and *IbMADS90* were highly expressed in the fibrous root or tuberous root. Seven genes were more highly expressed in the leaf, only *IbMADS15* was highly expressed in the stem, and the remaining genes were relatively poorly expressed in the stem. Ten genes were all expressed at lower levels in the stem tip. *IbMADS18*, *IbMADS31*, and *IbMADS83* were significantly expressed in the fruit. Some genes were highly expressed only in a single tissue, and others were not significantly expressed, such as *IbAMDS1*, which was significantly expressed only in the fibrous root; *IbMADS17* and *IbMADS20*, which were highly expressed only in the leaf; and *IbMADS31* and *IbMADS83*, which were significantly expressed only in the fruit. The five remaining genes were more highly expressed in two to three tissues.

4 DISCUSSION

MADS-box genes are widely present in eukaryotes. There are 108, 72, 168, and 160 MADS-box genes identified in *Arabidopsis* (Livak and Schmittgen, 2001), *Oryza sativa* (Ray et al., 2007),

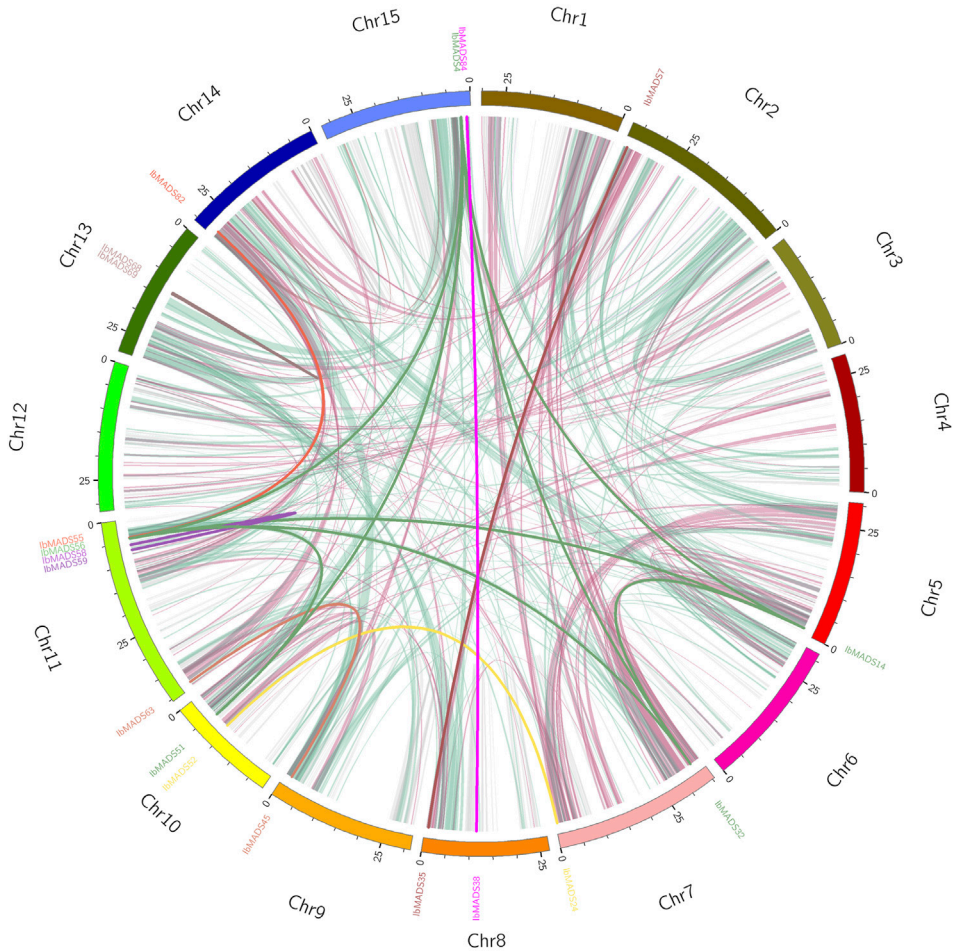


FIGURE 6 | Collinearity analysis of MADS-box gene in sweet potato. The same color name indicates that the gene has a tandem repeat.

Nicotiana tabacum (Bai et al., 2019), and *Brassica rapa* (Duan et al., 2015). Interestingly, *I. trifida*, a diploid plant belonging to the genus *Ipomoea*, has 37 MADS-box genes (Zhu et al., 2020). In this study, 95 sweet potato MADS-box genes were identified, the number of which was approximately three times that of *I. trifida*. Phylogenetic analysis found that the MADS-box genes of *I. batatas* are not simply tripled by *I. trifida*. We discovered more type I MADS genes in *I. batatas*, and the distribution of type II genes was also more abundant than that of *I. trifida*. Nevertheless, *I. batatas* are missing only in the FLC subfamily and the TT16 subfamily. It is suggested that FLC genes are repressors of flowering which affect multiple pathways for controlling the flowering time (Michaels and Amasino, 1999; Scott et al., 2003). Loss of FLC subfamily genes may indicate that sweet potato vernalization is not regulated by its repression. Consistently, a previous study also showed that the emergence of FLC subfamily members is not detected in *I. trifida* (Zhu et al., 2020). *TT16* encodes a TF involved in the transcriptional regulation of proanthocyanidin biosynthesis in the seed coat (Nesi et al., 2002). The absence of sweet potato TT16 subfamily genes suggests the potential involvement of other pathways for regulating proanthocyanidin biosynthesis in sweet potato.

Moreover, we again found differentiation in the MADS-box gene structure, which may be attributed to the presence of I-domain, K-domain, and C-domain. Type II genes typically exhibit a longer gene length and a greater number of exons, which is also one of the basis for classifying subtypes of MADS-box genes (Kaufmann et al., 2005). Consistent with the structure feature, the type I genes of sweet potato have one to two exons, and most of them range at 1 kb, while the type II genes are generally larger than 3 kb and have more than six exons. In addition, the MADS-box genes of sweet potato are distributed more dispersedly on the chromosome. Their numbers differ from those of chromosomes, which are similar to the chromosome distribution of *Arabidopsis* (Lamesch et al., 2012), *Jatropha curcas* (Tang et al., 2020), *Cardamine hirsuta* (Ghorbani et al., 2020), and *I. trifida*. The MADS-box proteins in sweet potato have a strong motif conservation among various subfamilies, with different numbers and motifs. The core motifs are motif 1, motif 2, and motif 5, which encode about 50 amino acids, and they are related to the SRF-box domain (Ng et al., 2001). Sweet

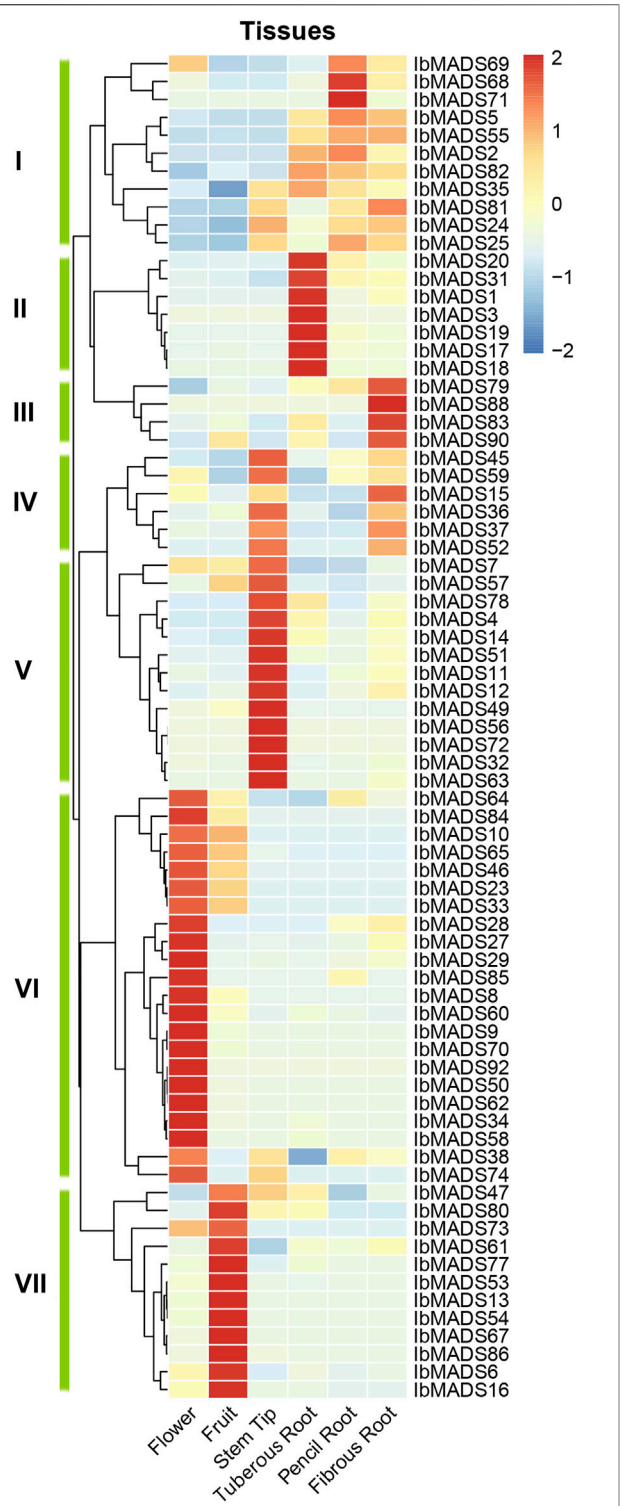


FIGURE 7 | Expression analysis of MADS-box gene in different organisms of sweet potato.

potato is a hexaploid plant with both homologous and heterologous chromosomes, which has a complex genetic background (Yang et al., 2017). MCScanX analysis showed that a large number of genes had collinear relationships within the sweet potato genome, and the MADS-box genes also had a certain homology of evolutionary relationship. Among them, 19 MADS-box-genes had collinear relationships. In particular, reciprocal segmental duplications existed in *IbMADS4*, *IbMADS14*, *IbMADS32*, and *IbMADS56*. *IbMAD51* shared a segmental duplication relationship with *IbMADS4* and *IbMADS56*, while it did not have a segmental duplication relationship with *IbMADS14* and *IbMADS32*. It is presumed that *IbMADS4*, *IbMADS14*, *IbMADS32*, and *IbMADS56* were derived from a single duplication event, while *IbMAD51* was derived from another separate duplication event, possibly a single segmental duplication event.

The MADS-box family is considered to be as a key family of genes responsible for regulating reproductive growth and vegetative organ development, which was initially characterized as homeotic genes in floral organs. Also, the importance of the MADS-box family in the morphogenesis and growth of other organs of plants, especially roots and fruits, has been highlighted (Riechmann et al., 1997; Zhao et al., 2019). Multiple types of *cis*-acting elements have been identified in the promoter region of the MADS-box gene from sweet potato, mainly including light-responsive elements, auxin-responsive elements, MeJA-responsive elements, abscisic acid-responsive elements, and drought-inducible elements. It is indicated that MADS-box genes in sweet potato respond to both hormones and environmental factors, and they are closely related to the normal growth and development of sweet potato. Expression levels of MADS-box genes can reveal the relevant physiological functions, Hasebe et al. (1998) found that most MADS-box genes in ferns are expressed in reproductive and vegetative organs, while most MADS-box genes in seed plants are only expressed in one organism, and only a few are expressed in reproductive and vegetative organisms. Transcriptome sequencing revealed that the majority of MADS-box genes in sweet potato exhibit the expression specialization. Thirty-four genes were specifically expressed in sweet potato flowers and fruits, while the remaining were mainly expressed in vegetative tissues like the root and stem tip. The expression profiles of 10 genes in different tissues of sweet potato were further verified by qRT-PCR. *IbMADS1* was highly expressed in the fibrous root; and *IbMADS79*, *IbMADS18*, *IbMADS19*, and *IbMADS90* were highly expressed in the tuberous root. It has been reported that *IbMADS1* and *IbMADS79* are associated with the root development of sweet potato. In particular, *IbMADS1* is an important integrator at the initiation of tuberization (Kim et al., 2005; Ku et al., 2008), which is consistent with our finding. *Cis*-acting elements like the light-responsive elements, Zein metabolism-regulation elements, and auxin-responsive elements were found in the promoter region of *IbMADS1*; and light-responsive elements, meristem-express elements, MeJA-responsive elements, and salicylic acid-responsive element were found in the promoter region of *IbMADS79*. *IbMADS1* and *IbMADS79* may play an important role in the root development process and may participate in the formation of the tuberous root. The homologous genes *IbMADS18* and *IbMADS19* of *FEM11* were highly expressed not only in the

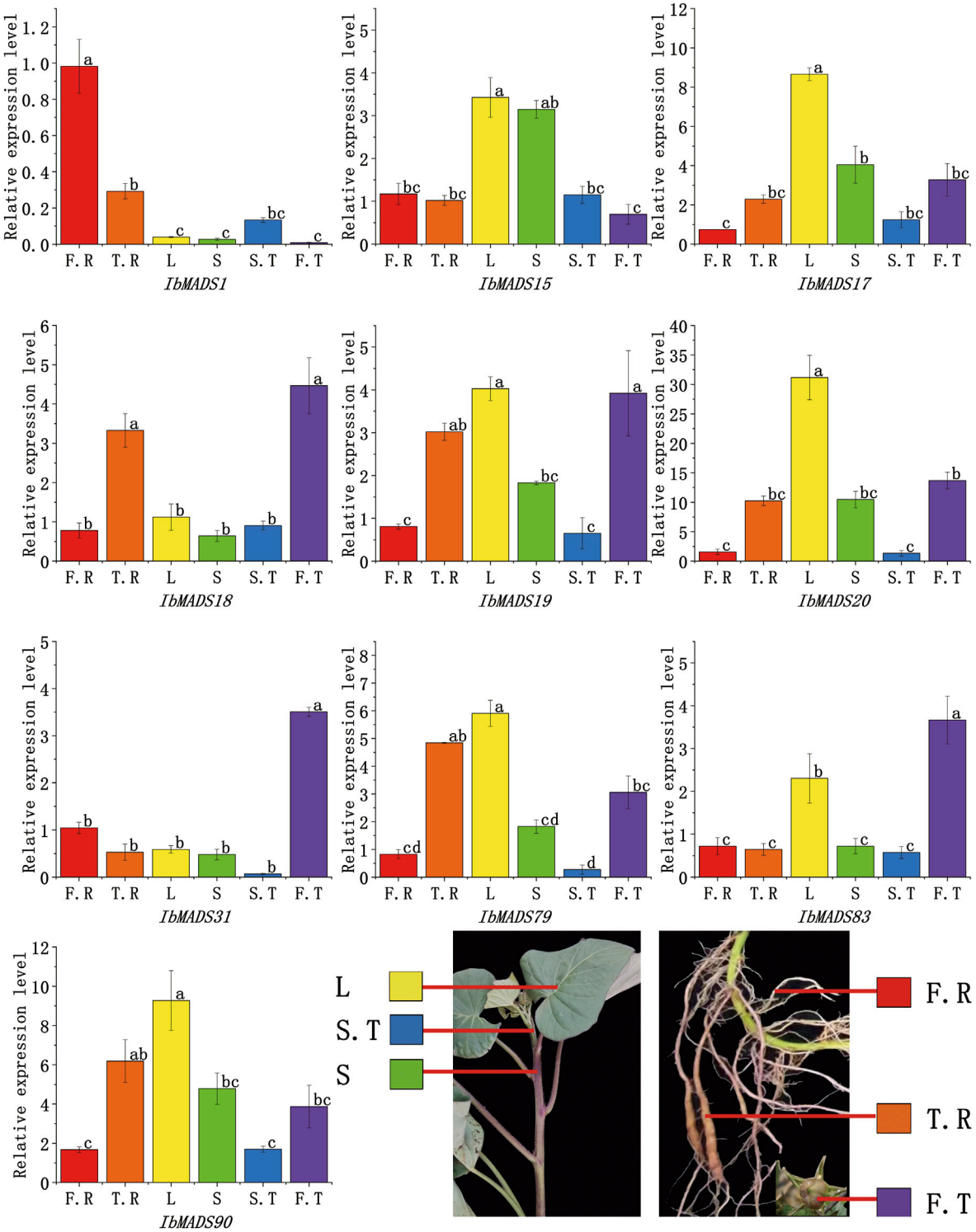


FIGURE 8 | The gene expression levels of some MADS-box genes in different tissues of sweet potato.

fruit but also in the tuberous root. And the homologous gene *IbMADS90* of *AGL62* was highly expressed in the tuberous root and leaf. *FEM111* and *AGL62* are associated with the flower and fruit development in *Arabidopsis* (Kang et al., 2008; Portereiko et al., 2006; Steffen et al., 2008). *IbMADS18*, *IbMADS19*, and *IbMADS90* are involved in the tuberous root development of sweet potato and the fruit development to some extent. In addition, *IbMADS31* and *IbMADS83* are also specifically

expressed in the fruit. *AGL61*, a homolog of *IbMADS83*, functions as a TF that controls the expression of downstream genes during central cell development and is involved in the seed endosperm development (Steffen et al., 2008). And *AGL17*, the homolog of *IbMADS31*, is a significant downstream target of CURLY LEAF in floral transition control (Shu et al., 2020). We found that *cis*-acting elements like the MeJA-responsive elements, auxin-responsive elements, abscisic acid-responsive elements, and gibberellin-responsive elements in the promoter region of *IbMADS31* and *IbMADS83* provided indirect evidences for the involvement of *IbMADS31* and *IbMADS83* in the flower and fruit development of sweet potato. *IbMADS15*, *IbMADS17*, *IbMADS19*, *IbMADS20*, *IbMADS79*, and *IbMADS90* were highly expressed in the leaf or stem, which were all identified as type I genes. In addition, a large number of type I genes were detected mainly expressed in the stem tip by RNA-seq. It is suggested that type I genes were involved in the nutritional growth process of the aerial parts of sweet potato, to some extent.

Overall, the sweet potato MADS-box genes are involved in the growth and development of reproductive organs like the flower and fruit of sweet potato. They were vital in the process of nutritional growth and development such as the root, stem, and leaf of sweet potato.

5 CONCLUSION

In this study, the 95 MADS-box family genes were identified from the sweet potato genome by Hmmer, SMART, and other tools, and they were divided into type I and type II subgroups according to the gene structure. The phylogenetic relationship of MADS-box genes between sweet potato and *Arabidopsis* was analyzed based on the NJ phylogenetic tree. The type I genes of sweet potato were further divided into three subfamilies, Mα, Mβ, and Mγ, whereas the type II genes were divided into 12 subfamilies.

The distribution of the sweet potato MADS-box genes on chromosomes was analyzed by chromosomal localization, and the analysis of the motif identified the MADS-box conserved domain motifs. Collinearity between and within chromosomes of sweet potato was analyzed using MCScanX to identify 19 MADS-box genes with segmental duplications or tandem duplications. Analysis of the MADS-box promoter region of sweet potato revealed that the MADS-box genes contain light-responsive elements, auxin-responsive elements, MeJA-responsive elements, abscisic acid-responsive elements, drought induction elements, and other *cis*-acting elements that respond to multiple environmental changes and coordinate growth and development. Furthermore, we analyzed the expression patterns of MADS-box genes in various tissues of sweet potato by transcriptome sequencing and qRT-PCR, and we found that the sweet potato MADS-box genes are involved in the growth and development of reproductive organs such as the flower and fruit, but also the growth and development of nutritional organs such as the root, stem, and leaf of sweet potato.

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

ZS conceived and designed the experiments. ZS and MH performed the experiments. ZZ and YC prepared the materials. ZS analyzed the data and wrote the paper. A-DH and HZ helped to revise the paper. All authors read and approved the final articles.

FUNDING

This research was funded by the NSFC-Guangdong Natural Science Foundation Joint Project (U1701234).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.750137/full#supplementary-material

REFERENCES

Bai, G., Yang, D. H., Cao, P., Yao, H., Zhang, Y., Chen, X., et al. (2019). Genome-Wide Identification, Gene Structure and Expression Analysis of the MADS-Box Gene Family Indicate Their Function in the Development of Tobacco (Nicotiana Tabacum l). *Int. J. Mol. Sci.* 20 (20), 5043. doi:10.3390/ijms20205043

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 (2), W202–W208. doi:10.1093/nar/gkp335

Burgeff, C., Liljegren, S. J., Tapia-López, R., Yanofsky, M. F., and Alvarez-Buylla, E. R. (2002). MADS-Box Gene Expression in Lateral Primordia, Meristems and Differentiated Tissues of *Arabidopsis thaliana* Roots. *Planta* 214 (3), 365–372. doi:10.1007/s004250100637

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 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouz, P., Theien, G., et al. (2003). Genomewide Structural Annotation and Evolutionary Analysis of the Type I MADS-Box Genes in Plants. *J. Mol. Evol.* 56 (5), 573–586. doi:10.1007/s00239-002-2426-x

Dong, T., Song, W., Tan, C., Zhou, Z., Yu, J., Han, R., et al. (2018). Molecular Characterization of Nine Sweet Potato (Ipomoea Batatas Lam.) MADS-Box Transcription Factors during Storage Root Development and Following Abiotic Stress. *Plant Breed* 137 (5), 790–804. doi:10.1111/pbr.12613

Duan, W., Song, X., Liu, T., Huang, Z., Ren, J., Hou, X., et al. (2015). Genome-wide Analysis of the MADS-Box Gene Family in Brassica Rapa (Chinese Cabbage). *Mol. Genet. Genomics.* 290 (1), 239–255. doi:10.1007/s00438-014-0912-7

Fan, H.-Y., Hu, Y., Tudor, M., and Ma, H. (1997). Specific Interactions between the K Domains of AG and AGLs, Members of the MADS Domain Family of DNA Binding Proteins. *Plant J.* 12 (5), 999–1010. doi:10.1046/j.1365-313X.1997.12050999.x

Florea, L., Song, L., and Salzberg, S. L. (2013). Thousands of Exon Skipping Events Differentiate Among Splicing Patterns in Sixteen Human Tissues. *F1000Res* 2, 188. doi:10.12688/f1000research.2-188.v2

Ghorbani Marghashi, M., Bagheri, H., and Gholami, M. (2020). Genome-Wide Study of Flowering-Related MADS-Box Genes Family in Cardamine Hirsuta. *3 Biotech.* 10 (12), 518. doi:10.1007/s13205-020-02521-w

Hasebe, M., Wen, C.-K., Kato, M., and Banks, J. A. (1998). Characterization of MADS Homeotic Genes in the Fern Ceratopteris Richardii. *Proc. Natl. Acad. Sci.* 95 (11), 6222–6227. doi:10.1073/pnas.95.11.6222

Ito, K., and Murphy, D. (2013). Application of Ggplot2 to Pharmacometric Graphics. *CPT: Pharmacometrics Syst. Pharmacol.* 2, 79. doi:10.1038/psp.2013.56

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: Toward a Central Hub for Transcription Factors and Regulatory Interactions in Plants. *Nucleic Acids Res.* 45 (D1), D1040–D1045. doi:10.1093/nar/gkw982

Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). The AGL62 MADS Domain Protein Regulates Cellularization during Endosperm Development inArabidopsis. *Plant Cell* 20 (3), 635–647. doi:10.1105/tpc.107.055137

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-Type MADS-Domain Proteins: Structural Modularity, Protein Interactions and Network Evolution in Land Plants. *Gene* 347 (2), 183–198. doi:10.1016/j.gene.2004.12.014

Kim, S.-H., Hamada, T., Otani, M., and Shimada, T. (2005). Isolation and Characterization of MADS Box Genes Possibly Related to Root Development in Sweetpotato(Ipomoea Batatas L. Lam). *J. Plant Biol.* 48 (4), 387–393. doi:10.1007/BF03030580

Kim, S.-H., Mizuno, K., and Fujimura, T. (2002). Isolation of MADS-Box Genes from Sweet Potato (Ipomoea Batatas (L.) Lam.) Expressed Specifically in Vegetative Tissues. *Plant Cel Physiol* 43 (3), 314–322. doi:10.1093/pcp/pcf043

Krejci, A., Hupp, T. R., Lexa, M., Vojtesek, B., and Muller, P. (2016). Hammock: A Hidden Markov Model-Based Peptide Clustering Algorithm to Identify Protein-Interaction Consensus Motifs in Large Datasets. *Bioinformatics* 32 (1), btv522–16. doi:10.1093/bioinformatics/btv522

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* 19 (9), 1639–1645. doi:10.1101/gr.092759.109

Ku, A. T., Huang, Y.-S., Wang, Y.-S., Ma, D., and Yeh, K.-W. (2008). IbMADS1 (Ipomoea Batatas MADS-Box 1 Gene) Is Involved in Tuberous Root Initiation in Sweet Potato (Ipomoea Batatas). *Ann. Bot.* 102 (1), 57–67. doi:10.1093/aob/mcn067

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 (6), 1547–1549. doi:10.1093/molbev/msy096

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): Improved Gene Annotation and New Tools. *Nucleic Acids Res.* 40 (Database issue), D1202–D1210. doi:10.1093/nar/gkr1090

Lescot, M., Dehais, 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 (1), 325–327. doi:10.1093/nar/30.1.325

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 (4), 402–408. doi:10.1006/meth.2001.1262

Michaels, S. D., and Amasino, R. M. (1999). FLOWERING LOCUS C Encodes a Novel MADS Domain Protein that Acts as a Repressor of Flowering. *Plant Cell* 11 (5), 949–956. doi:10.1105/tpc.11.5.949

Michaels, S. D., He, Y., Scortecci, K. C., and Amasino, R. M. (2003). Attenuation of FLOWERING LOCUS C Activity as a Mechanism for the Evolution of Summer-Annual Flowering Behavior in Arabidopsis. *Proc. Natl. Acad. Sci.* 100 (17), 10102–10107. doi:10.1073/pnas.1531467100

Nam, J., Kaufmann, K., Theissen, G., and Nei, M. (2005). A Simple Method for Predicting the Functional Differentiation of Duplicate Genes and its Application to MIKC-Type MADS-Box Genes. *Nucleic Acids Res.* 33 (2), e12. doi:10.1093/nar/gni003

Nesi, N., Debeaujon, I., Jond, C., Stewart, A. J., Jenkins, G. I., Caboche, M., et al. (2002). The Transparent Testa16 Locus Encodes the Arabidopsis Bsister Mads Domain Protein and Is Required for Proper Development and Pigmentation of the Seed Coat. *Plant Cell* 14 (10), 2463–2479. doi:10.1105/tpc.004127

Ng, M., and Yanofsky, M. F. (2001). Function and Evolution of the Plant MADS-Box Gene Family. *Nat. Rev. Genet.* 2 (3), 186–195. doi:10.1038/35056041

Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and Properties of cDNA Clones Encoding SRF, a Transcription Factor that Binds to the C-Fos Serum Response Element. *Cell* 55 (6), 989–1003. doi:10.1016/0092-8674(88)90244-9

Pařenicová´, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and Phylogenetic Analyses of the Complete MADS-Box Transcription Factor Family in Arabidopsis. *Plant Cell* 15 (7), 1538–1551. doi:10.1105/tpc.011544

Passmore, S., Maine, G. T., Elble, R., Christ, C., and Tye, B.-K. (1988). *Saccharomyces cerevisiae* Protein Involved in Plasmid Maintenance Is Necessary for Mating of MATα Cells. *J. Mol. Biol.* 204 (3), 593–606. doi:10.1016/0022-2836(88)90358-0

Portereiko, M. F., Lloyd, A., Steffen, J. G., Punwani, J. A., Otsuga, D., and Drews, G. N. (2006). AGL80Is Required for Central Cell and Endosperm Development inArabidopsis. *Plant Cell* 18 (8), 1862–1872. doi:10.1105/tpc.106.040824

Ray, S., Agarwal, P., Arora, R., Kapoor, S., and Tyagi, A. K. (2007). Expression Analysis of Calcium-Dependent Protein Kinase Gene Family during Reproductive Development and Abiotic Stress Conditions in rice (Oryza Sativa L. Ssp. Indica). *Mol. Genet. Genomics.* 278 (5), 493–505. doi:10.1007/s00438-007-0267-4

Riechmann, J. L., and Meyerowitz, E. M. (1997). MADS Domain Proteins in Plant Development. *Biol. Chem.* 378 (10), 1079–1101. doi:10.1515/bchm.1997.378.10.1079

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4 (4), 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Sheng, X.-G., Zhao, Z.-Q., Wang, J.-S., Yu, H.-F., Shen, Y.-S., Zeng, X.-Y., et al. (2019). Genome Wide Analysis of MADS-Box Gene Family in *Brassica oleracea* Reveals Conservation and Variation in Flower Development. *BMC Plant Biol.* 19 (1), 106. doi:10.1186/s12870-019-1717-y

Shu, J., Chen, C., Kohalmi, S. E., and Cui, Y. (2020). Evidence that AGL17 Is a Significant Downstream Target of CLF in floral Transition Control. *Plant Signaling Behav.* 15 (7), 1766851. doi:10.1080/15592324.2020.1766851

Singh, V. K., Mangalam, A. K., Dwivedi, S., and Naik, S. (1998). Primer Premier: Program for Design of Degenerate Primers from a Protein Sequence. *Biotechniques* 24 (2), 318–319. doi:10.2144/98242pf02

Sommer, H., Beltrán, J. P., Huijser, P., Pape, H., Lönnig, W. E., Saedler, H., et al. (1990). Deficiens, a Homeotic Gene Involved in the Control of Flower Morphogenesis in Antirrhinum Majus: The Protein Shows Homology to Transcription Factors. *EMBO J.* 9 (3), 605–613. doi:10.1002/j.1460-2075.1990.tb08152.x

Steffen, J. G., Kang, I.-H., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). AGL61 Interacts with AGL80 and Is Required for Central Cell Development in Arabidopsis. *Plant Physiol.* 148 (1), 259–268. doi:10.1104/pp.108.119404

Sumiya, N. (2021). Cis-Acting Elements Involved in the G2/M-Phase-Specific Transcription of the Cyclin B Gene in the Unicellular Alga Cyanidioschyzon Merolae. *J. Plant Res.* 134 (6), 1301–1310. doi:10.1007/s10265-021-01334-z

Sun, C.-H., Wang, J.-H., Gu, K.-D., Zhang, P., Zhang, X.-Y., Zheng, C.-S., et al. (2021). New Insights into the Role of MADS-Box Transcription Factor Gene CmANR1 on Root and Shoot Development in chrysanthemum (Chrysanthemum Morifolium). *BMC Plant Biol.* 21 (1), 79. doi:10.1186/s12870-021-02860-7

Tang, Y., Wang, J., Bao, X., Wu, Q., Yang, T., Li, H., et al. (2020). Genome-Wide Analysis of Jatropha Curcas MADS-Box Gene Family and Functional Characterization of the JcMADS40 Gene in Transgenic rice. *BMC Genomics* 21 (1), 325. doi:10.1186/s12864-020-6741-7

Theissen, G. (2001). Development of floral Organ Identity: Stories from the MADS House. *Curr. Opin. Plant Biol.* 4 (1), 75–85. doi:10.1016/s1369-5266(00)00139-4

Voorrips, R. E. (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

Wang, S., Nie, S., and Zhu, F. (2016). Chemical Constituents and Health Effects of Sweet Potato. *Food Res. Int.* 89 (Pt 1), 90–116. doi:10.1016/j.foodres.2016.08.032

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40 (7), e49. doi:10.1093/nar/gkr1293

Wang, Y., Zhang, J., Hu, Z., Guo, X., Tian, S., and Chen, G. (2019). Genome-Wide Analysis of the MADS-Box Transcription Factor Family in Solanum Lycopersicum. *Int. J. Mol. Sci.* 20 (12), 2961. doi:10.3390/ijms20122961

Won, S. Y., Jung, J.-A., and Kim, J. S. (2021). Genome-Wide Analysis of the MADS-Box Gene Family in Chrysanthemum. *Comput. Biol. Chem.* 90, 107424. doi:10.1016/j.compbiolchem.2020.107424

Wu, Z., Raven, P. H., and Hong, D. (1998). *Flora of china*. Beijing: Science Press.

Yang, J., Moeinzadeh, M.-H., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., et al. (2017). Haplotype-Resolved Sweet Potato Genome Traces Back its Hexaploidization History. *Nat. Plants* 3 (9), 696–703. doi:10.1038/s41477-017-0002-z

Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The Protein Encoded by the Arabidopsis Homeotic Gene Agamous Resembles Transcription Factors. *Nature* 346 (6279), 35–39. doi:10.1038/346035a0

Zhang, S., Guo, Y., Zhang, Y., Guo, J., Li, K., Fu, W., et al. (2021). Genome-wide Identification, Characterization and Expression Profiles of the CCD Gene Family in Gossypium Species. *3 Biotech.* 11 (5), 249. doi:10.1007/s13205-021-02805-9

Zhao, H.-B., Jia, H.-M., Wang, Y., Wang, G.-Y., Zhou, C.-C., Jia, H.-J., et al. (2019). Genome-Wide Identification and Analysis of the MADS-Box Gene Family and its Potential Role in Fruit Development and Ripening in Red Bayberry (Morella Rubra). *Gene* 717, 144045. doi:10.1016/j.gene.2019.144045

Zhu, P., Dong, T., Xu, T., and Kang, H. (2020). Identification, Characterisation and Expression Analysis of MADS-Box Genes in Sweetpotato Wild Relative Ipomoea Trifida. *Acta Physiol. Plant* 42 (11), 163. doi:10.1007/s11738-020-03153-6

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

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

Copyright © 2021 Shao, He, Zeng, Chen, Hanna and Zhu. 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: 24 November 2021
doi: 10.3389/fgene.2021.735862

Genome-Wide Analysis of the Fatty Acid Desaturase Gene Family Reveals the Key Role of *PfFAD3* in α-Linolenic Acid Biosynthesis in Perilla Seeds

Wu Duan[1,2], *Yang Shi-Mei*[1], *Shang Zhi-Wei*[3], *Xu Jing*[3], *Zhao De-Gang*[2,3], *Wang Hong-Bin*[1] *and Shen Qi*[1]*

[1]Institute of Medical Plant Physiology and Ecology, School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou, China, [2]Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region, College of Life Sciences, Guizhou University, Guiyang, China, [3]Guizhou Rapeseed Institute, Guizhou Academy of Agricultural Sciences, Guiyang, China

Perilla (*Perilla frutescens*), a traditional medicinal and oilseed crop in Asia, contains extremely high levels of polyunsaturated α-linolenic acid (ALA) (up to 60.9%) in its seeds. ALA biosynthesis is a multistep process catalyzed by fatty acid desaturases (FADs), but the *FAD* gene family in perilla has not been systematically characterized. Here, we identified 42 *PfFADs* in the perilla genome and classified them into five subfamilies. Subfamily members of *PfFADs* had similar exon/intron structures, conserved domain sequences, subcellular localizations, and cis-regulatory elements in their promoter regions. *PfFAD*s also possessed various expression patterns. *PfFAD3.1* was highly expressed in the middle stage of seed development, whereas *PfFAD7/8.3* and *PfFAD7/8.5* were highly expressed in leaf and later stages of seed development, respectively. Phylogenetic analysis revealed that the evolutionary features coincided with the functionalization of different subfamilies of PUFA desaturase. Heterologous overexpression of *PfFAD3.1* in *Arabidopsis thaliana* seeds increased ALA content by 17.68%–37.03%. These findings provided insights into the characteristics and functions of *PfFAD* genes in perilla.

Keywords: perilla, α-linolenic acid (C18:3), fatty acid desaturases, genome-wide analysis, plant transformation

OPEN ACCESS

Edited by:
Sreepriya Pramod, Altria, United States

Reviewed by:
Andrea Kaenel, University of Münster, Germany
Yourong Chai, Southwest University, China

***Correspondence:**
Shen Qi
shenqi-gz1983@163.com

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

Received: *03 July 2021*
Accepted: *28 October 2021*
Published: *24 November 2021*

Citation:
Duan W, Shi-Mei Y, Zhi-Wei S, Jing X, De-Gang Z, Hong-Bin W and Qi S (2021) Genome-Wide Analysis of the Fatty Acid Desaturase Gene Family Reveals the Key Role of PfFAD3 *in α-Linolenic Acid Biosynthesis in Perilla Seeds. Front. Genet. 12:735862. doi: 10.3389/fgene.2021.735862*

INTRODUCTION

Fatty acids (FAs), as a major energy source of plants and animals, are the main components of membrane lipids (e.g., phospholipids) and seed storage lipids (e.g., triacylglycerols) in plants (Manan et al., 2017). α-Linolenic acid (ALA, C18:3 $\Delta^{9,12,15}$) is primarily synthesized in oil-producing algae and terrestrial plants (Ángela et al., 2015). As one of ω-3 polyunsaturated fatty acids, ALA is crucial for maintaining the fluidity of membrane lipids in plants and a precursor of FA-derived signaling molecules such as jasmonic acid, which play important roles in plant development and stress responses (Weber, 2002). ALA is also the precursor for the biosynthesis of eicosapentaenoic acid (C20:5 $\Delta^{5,8,11,14,17}$) and docosahexaenoic acid (C22:6 $\Delta^{4,7,10,13,16,19}$), which have important health benefits for humans, such as regulating body development, promoting brain development, reducing blood pressure, and inhibiting senescence, with therapeutical effects on neurological, cardiovascular, and cerebrovascular diseases (Baker et al., 2016). Since most animals do not produce ALA, they must

obtain ALA from plants in their diet. Hence, oilseed crops rich in ALA have become an important research focus in recent years (Baker et al., 2016).

Perilla (*Perilla frutescens* var. *frutescens*, 2n = 40) is an annual herbaceous plant of the Lamiaceae family (Nitta et al., 2005). Perilla originated in China and is widely cultivated in Asian countries including Japan and Korea (Park et al., 2008). As a traditional medicinal and edible plant, perilla is widely used in the pharmaceutical, healthcare, functional food, and cosmetics industries (Igarash and Miyazaki, 2013; Mungmai et al., 2020). Perilla is also an important oilseed crop with high seed oil content (40%), of which ALA accounts for 60.9% (Ciftci et al., 2012).

Polyunsaturated fatty acid (PUFA) biosynthesis is catalyzed by a series of enzymes called fatty acid desaturases (FADs) (Alonso et al., 2003). In higher plants, the first-step desaturases are catalyzed primarily by Δ7/Δ9 desaturases (Smith et al., 2013). Δ9 desaturases, the only desaturases present in all organisms are soluble acyl-acyl carrier protein desaturases (Kachroo et al., 2007). Δ12 and ω-3 desaturases are widely reported in plants. They are used as secondary and tertiary desaturases catalyzing the conversion of oleic acid (C18:1) to linoleic acid (C18:2) and further produce ALA, respectively (Bhunia et al., 2016). Front-end FADs are functionally heterologous enzymes that produce long-chain PUFAs (Meesapyodsuk and Xiao, 2012). FAD4s, a novel class of FADs, produce Δ3-desaturated FAs in plants (Gao et al., 2009; Horn et al., 2020). Three highly conserved histidine motifs are present in most FAD proteins, which are thought to be pivotal for maintaining their catalytic activities (Sperling et al., 2003; Berestovoy et al., 2020).

ω-3 Desaturases (ω-3 FADs) are catalytic enzymes for ALA biosynthesis. Three genes encoding ω-3 FADs (*FAD3*, *FAD7*, and *FAD8*) in *Arabidopsis thaliana* share over 65% sequence identity (Iba et al., 1993; Watahiki and Yamamoto, 1994). *FAD3* encodes an endoplasmic reticulum membrane-bound desaturase that mainly functions in ALA biosynthesis in seeds (Zhou et al., 2010). *FAD7* and *FAD8* encode plastid membrane-bound desaturases with non-redundant roles in plant responses to low temperature and other environment stresses (Matsuda et al., 2005; Ángela et al., 2015). ω-3 FAD orthologous genes have been cloned and characterized in diverse plant species, such as flax (*Linum usitatissimum*) (Vrinten et al., 2005), soybean (*Glycine max*) (Anai et al., 2005), and rapeseed (*Brassica napus*) (Yang et al., 2012). These enzymes play important roles in ALA accumulation in plants.

PfFAD3 cDNA was isolated from developing perilla seeds and determined to be specifically expressed in seeds (Abdelreheem and Hildebrand, 2013). Transcriptome analysis suggested that *PfFAD3* encodes one major desaturase for seed ALA content in perilla (Zhang et al., 2017). However, the heterologous expression of *PfFAD3* in yeast resulted in limited production of ALA (1.3%) (Lee et al., 2016). Therefore, the specific function of *PfFAD3* remains unclear.

In this study, we performed genome-wide analysis and identified 42 *PfFAD* genes in the perilla genome. We also performed phylogenetic analysis of PUFA desaturase (FADs related to polyunsaturated FA synthesis) in plants. Transcriptomic analysis and quantitative reverse-transcription PCR (qRT-PCR) revealed the tissue-specific expression patterns of *PfFAD* genes. We cloned *PfFAD3.1*, which is highly expressed in perilla seeds, and heterologously overexpressed this gene in *A. thaliana*. The systematic identification and functional feature laid the foundation for functional research on other *PfFADs*.

MATERIALS AND METHODS

Plant Materials

A perilla variety (*Perilla frutescens* var. *frutescens*) with high seed oil content was selected for this study. The plants were grown in a greenhouse at the Institute of Medicinal Plant Physiology and Ecology, Guangzhou University of Chinese Medicine (23°06′N, 113°40′E). The hypocotyls and cotyledons of perilla were collected 10 days after seed germination; the roots, stems, and leaves of perilla were collected at the three-leaf period of seedling growth, and flowers were collected at the flowering period. Seeds of perilla were collected 5 days after flowering (DAF), 10DAF, 15DAF, and 20DAF. All tissues were immediately frozen in liquid nitrogen and stored at −80°C until use for RNA extraction and qRT-PCR analysis.

Columbia wild-type (WT) *Arabidopsis thaliana* was used for gene transformation. WT and transgenic *A. thaliana* were kept in the dark at 4°C for 3 days to break dormancy (stratification) and then moved into a climate chamber with a photoperiod of 16 h light/8 h dark at 22°C.

Identification of *PfFAD* Gene Family Members

Two methods were used to identify *PfFAD* gene family members. For hidden Markov method (HMM) searches, an HMM file (FA_desaturase domain: PF00487; FA_desaturase_2 domain: PF03405) was downloaded from the Pfam database (http://Pfam.xfam.org/search#tabview=tab3) (Finn et al., 2016), and HMMER 3.0 was used to retrieve the FAD sequences from the genome database (Potter et al., 2018). For BLAST alignment, the *A. thaliana* FAD protein sequences (**Supplementary Table S1**) were download from The Arabidopsis Information Resource (version 9.0: https://www.arabidopsis.org/) and used as queries to search for perilla FADs using the BLATSP programs with an E value of 1e−10. The sequences identified using the two methods were merged, and genes lacking the intact FAD domain were removed.

The whole-genome shogun contigs (WGS) of perilla (taxid: 48,385) was used to analyze the position of *PfFAD* genes (Zhang et al., 2021). Complete domain verification of the acquired perilla FADs was performed in Pfam (http://pfam.xfam.org/search#tabview=tab1). The structures of the perilla *FAD* genes were identified with the GSDS tool (http://gsds.cbi.pku.edu.cn/) based on genomic DNA sequences and coding sequences (CDSs) (Hu et al., 2014). The conserved domains and motifs of the perilla FADs were analyzed using the online software Pfam and MEME (http://meme-suite.org/) (Bailey et al., 2009). The chemical properties and subcellular localizations of the proteins were examined using ExPASy (http://web.expasy.org/protparam/)

and Cell-PLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant/) (Artimo et al., 2012; Chou and Shen, 2008). To analyze the promoters of the *FAD* genes, the upstream 2,000-bp sequences of the genes were extracted and the cis-acting regulatory elements were scanned in the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002).

Phylogenetic Analysis

FAD protein sequences of prokaryotic algae, eukaryotic algae, bryophytes, pteridophytes, and other plants (**Supplementary Table S1**) were downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/). A phylogenetic tree of PfFAD sequences was constructed using the neighbor-joining (NJ) method with MEGA-X (MEGA_X_11.0.1) software with a bootstrap value of 1,000. The online software tool iTOL (https://itol.embl.de/) was used for visualization (Kumar et al., 2018; Letunic and Bork, 2019).

Gene Expression Analysis

Transcriptome data from different perilla tissues were downloaded from NCBI (accession number: SRP111892) (Zhang et al., 2017), the fragments per kilobase of transcript per million mapped reads (FPKM) values of *FAD* gene families were extracted, and TBtools software was used for clustering and visual analysis (Chen et al., 2020).

qRT-PCR was used to analyze the expression of *FAD* genes. Total RNA was extracted from the plant tissues using a Quick RNA Isolation Kit (TaKaRa, Beijing, China), and cDNA was synthesized from the RNA using a PrimeScript RT Reagent Kit (TaKaRa, China). qRT-PCR reactions were performed using the 2x SYBR Green Pro Taq HS Premix (Accurate, Dongguan, China). Each 10-μl reaction mixture contained 5 μl 2x SYBR Green Pro Taq HS Premix, 0.3 μl of each primer, and 4.4 μl cDNA diluted 20 times. The qPCR cycling conditions were 1 cycle of pre-denaturation at 95°C, 30 s; 40 cycles of 95°C for 5 s and 60°C for 30 s; dissolution at 95°C for 5 s; and 60°C for 1 min. Each experiment was performed four times, and the data were processed by the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). *AtActin* and *PfActin* were used as endogenous control genes to normalize the semiquantitative RT-PCR analysis of plants. Primers used for qRT-PCR are listed in **Supplementary Table S2**.

Gene Cloning and Arabidopsis Transformation

Primers with *Bam*HI and *Sal*I enzyme loci were used for *PfFAD3.1* gene amplification. The amplified CDS of *PfFAD3.1* was digested with the restriction enzymes and inserted into the pCAMBIA2301-KY vector [under the control of the cauliflower mosaic virus (CaMV) 35S promoter] to generate the *35S::PfFAD3.1* construct. The construct was transformed into *Agrobacterium tumefaciens* strain EHA105 and introduced into *A. thaliana* using the floral dip method (Clough and Bent, 1998). The transformed seeds were germinated on half-strength Murashige and Skoog medium with 40 mg/l kanamycin. Seedlings with well-established roots and green leaves were transferred to moistened potting soil and were verified by the PCR method and GUS staining (**Supplementary Figure S1**). The transgenic T1 lines exhibiting a 3:1 segregation of resistance to 40 mg/l kanamycin were considered the T2 generation, and T2 transgenic lines were finally acquired. Mature seeds of the T2 generation overexpression *PfFAD3.1* Arabidopsis were collected for gene expression and FA content analyses.

FA Analysis

Dry Arabidopsis seeds (0.15 g), pulverized in liquid nitrogen and FA methyl esters (FAMEs), were prepared according to the method by Li et al. (Li et al., 2015). FAMEs were analyzed using a GC8890 instrument (Agilent, Santa Clara, CA, USA) and a 30 m × 0.25 mm DB-23 column with a nitrogen carrier. The gas chromatography (GC) procedure used the following parameters: split ratio 10:1, initial column temperature held at 100°C for 2 min, then raised to 230°C at the rate of 15°C min^{-1} with the final temperature held for 5 min; injection port temperature 230°C; transfer line temperature 250°C; ion source temperature 230 °C; speed of carrier gas 1.0 ml min^{-1}; and injection volume 1 μL. Methyl heptadecanoate was used as an internal standard, and a standard curve method was used as the quantitative approach to identify FAMEs (Yin et al., 2018).

RESULTS AND ANALYSIS

Identification and Analysis of *FADS* in Perilla

We identified 39 and 58 *PfFAD* candidate genes in the perilla genome using HMM searches and BLAST alignment, respectively. After merging the data, the pfam database was used to predict the conserved domains of the protein encoded by the candidate genes. Thirteen genes lacking the intact FAD domain and two genes with partial FAD domains were identified as pseudogenes and removed. A total of 42 protein-coding genes were confirmed finally, which are located on the other 18 perilla chromosomes except for the 14th and 18th chromosomes (**Supplementary Figure S1**). The protein sizes range from 178 to 1,111 amino acids, with molecular weights ranging from 20.91 to 126.81 kDa and theoretical isoelectric points ranging from 5.21 to 9.51 (**Supplementary Table S3**).

By comparison with the *A. thaliana FAD* family, the *PfFAD* genes were divided into five typical subfamilies based on phylogenetic analysis and sequence characteristics, including Δ7/Δ9 desaturases (Δ7/Δ9 FAD), Δ12 desaturases (Δ12 FAD), ω-3 desaturases (ω-3 FAD), front-end desaturases, and FAD4 desaturases (**Figure 1**). The Δ7/Δ9 desaturase subfamily contains 15 genes belonging to two types (Kazaz et al., 2020). Thirteen *PfFAB2* genes encode Δ9 desaturases containing an FA_desaturase2 domain (PF03405), and three conserved motifs (8, 9, 10), which were predicted to lack transmembrane domains (TMDs). Two *PfADS3* genes encode Δ7 desaturases with an FA_desaturase domain (PF00487), two motifs (4, 5), and one or two TMDs (**Figure 2A** and **Supplementary Table S3**). Δ7/Δ9 desaturases in perilla were predicted to localize in chloroplast. The novel PfFAD4 desaturase subfamily includes five genes,

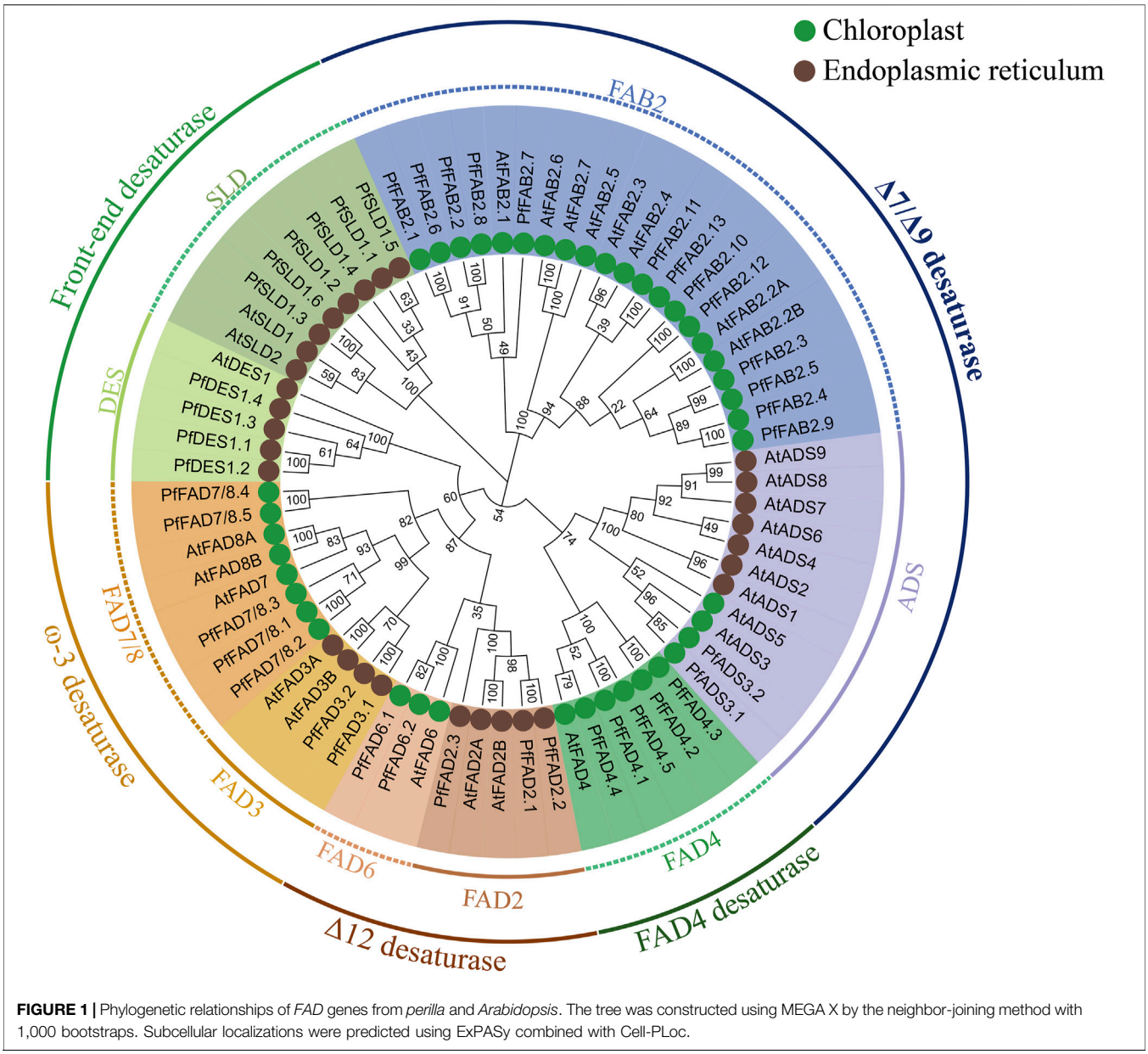


FIGURE 1 | Phylogenetic relationships of *FAD* genes from *perilla* and *Arabidopsis*. The tree was constructed using MEGA X by the neighbor-joining method with 1,000 bootstraps. Subcellular localizations were predicted using ExPASy combined with Cell-PLoc.

which contain the TMEM189_B_domain (PF10520) and two motifs (6 and 7).

Front-end desaturase, Δ12 desaturase, and ω3 desaturase proteins all contain FA_desaturase as the main domain. The front-end desaturase subfamily includes 10 genes, which encode endoplasmic reticulum-localized proteins and is divided into two clades. Four of these genes (*PfDES*s) encode sphingolipid Δ4 desaturases, which contain two exons and conserved motifs (1 and 3). Both PfDES1.1 and PfDES1.2 lack TMDs, and the remaining two genes encode proteins with five TMDs. Six *PfSLD* genes encode sphingolipid Δ8 desaturases, which contain one conserved motif (1, 2, or 3) but no introns.

The two types of Δ12 desaturase subfamily members contain three conserved motifs (1, 2, and 3). Among them, two *PfFAD6* genes contain 10 exons and encode proteins that only contain the FA_desaturase domain, which were predicted to have a chloroplast signal peptide at the N-termini and three or four TMDs (Xue et al., 2021). Three *PfFAD2* genes lack introns and encode proteins with an FA_desaturase domain and one short DUF3474 domain. PfFAD2 proteins were predicted to contain an endoplasmic reticulum retention signal peptide embedded within Φ-X-X-K-Φ (Φ is a hydrophobic amino acid residue) and three to six TMDs.

ω-3 desaturases are key enzymes for ALA formation (Lee et al., 2016). Here, seven ω-3 desaturase genes were identified in perilla, which were classified into two types (*PfFAD3* and *PfFAD7/8*) consistent with previous classification reports. Two *PfFAD3* genes encode proteins containing an endoplasmic reticulum retention signal (SKKI) at their C-termini and three *PfFAD7/8* genes contain a chloroplast transit peptide at their N-termini. Both

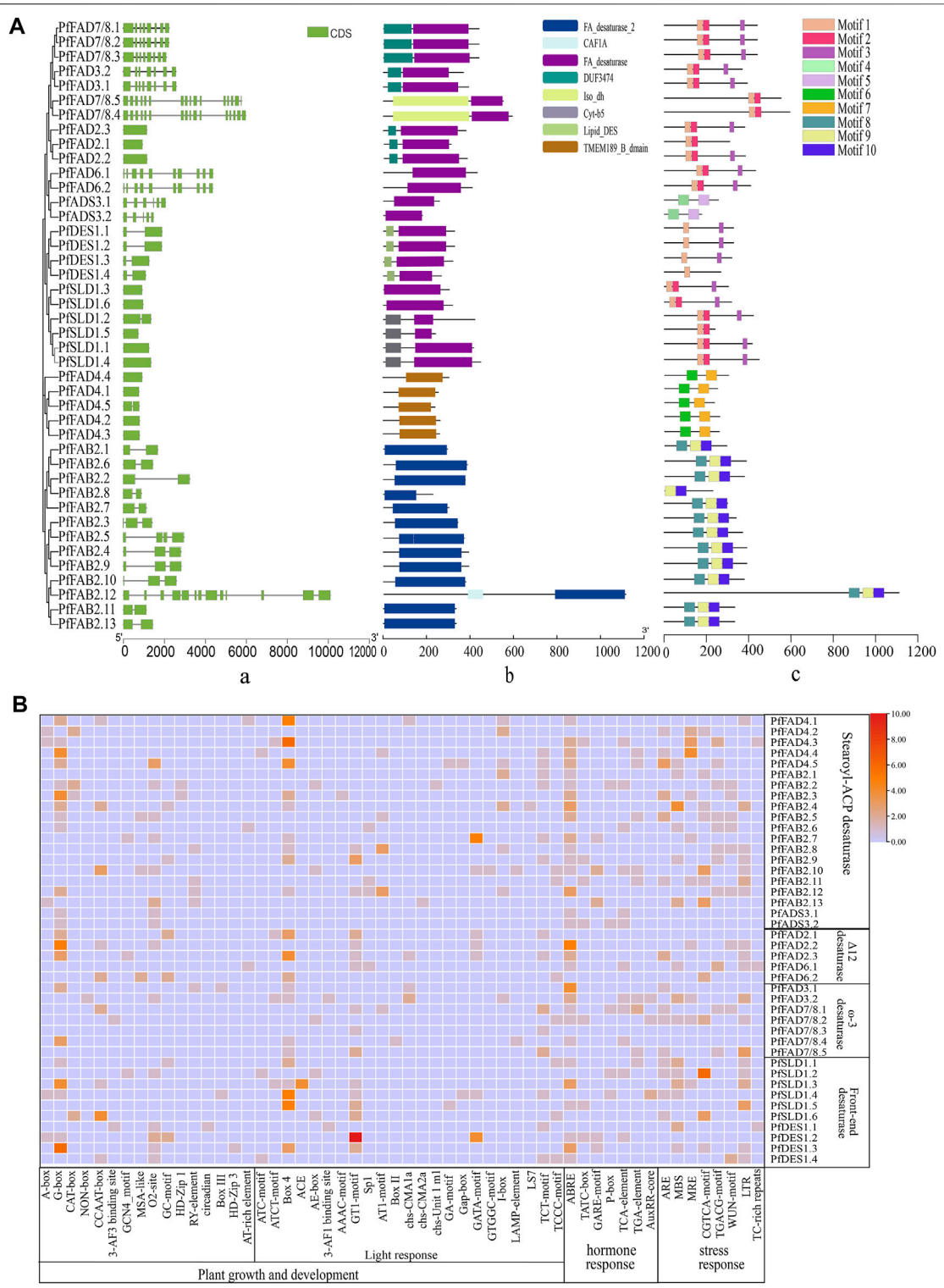


FIGURE 2 | Structure and cis-acting elements of *FADs* in perilla. **(A)** Analysis of gene structures **(i)**, domains **(ii)**, and motifs **(iii)** of the PfFADs. Introns/exons were identified using the GSDS tool. Conserved domains and motifs were analyzed using Pfam and MEME. **(B)** Analysis of the cis-acting elements in the promoters of *PfFADs*. The cis-acting regulatory elements in the upstream 2,000-bp sequences of the genes were identified using the PlantCARE website. Orange and red shading indicates the number of cis-acting elements present; light blue shading means no cis-acting elements were present.

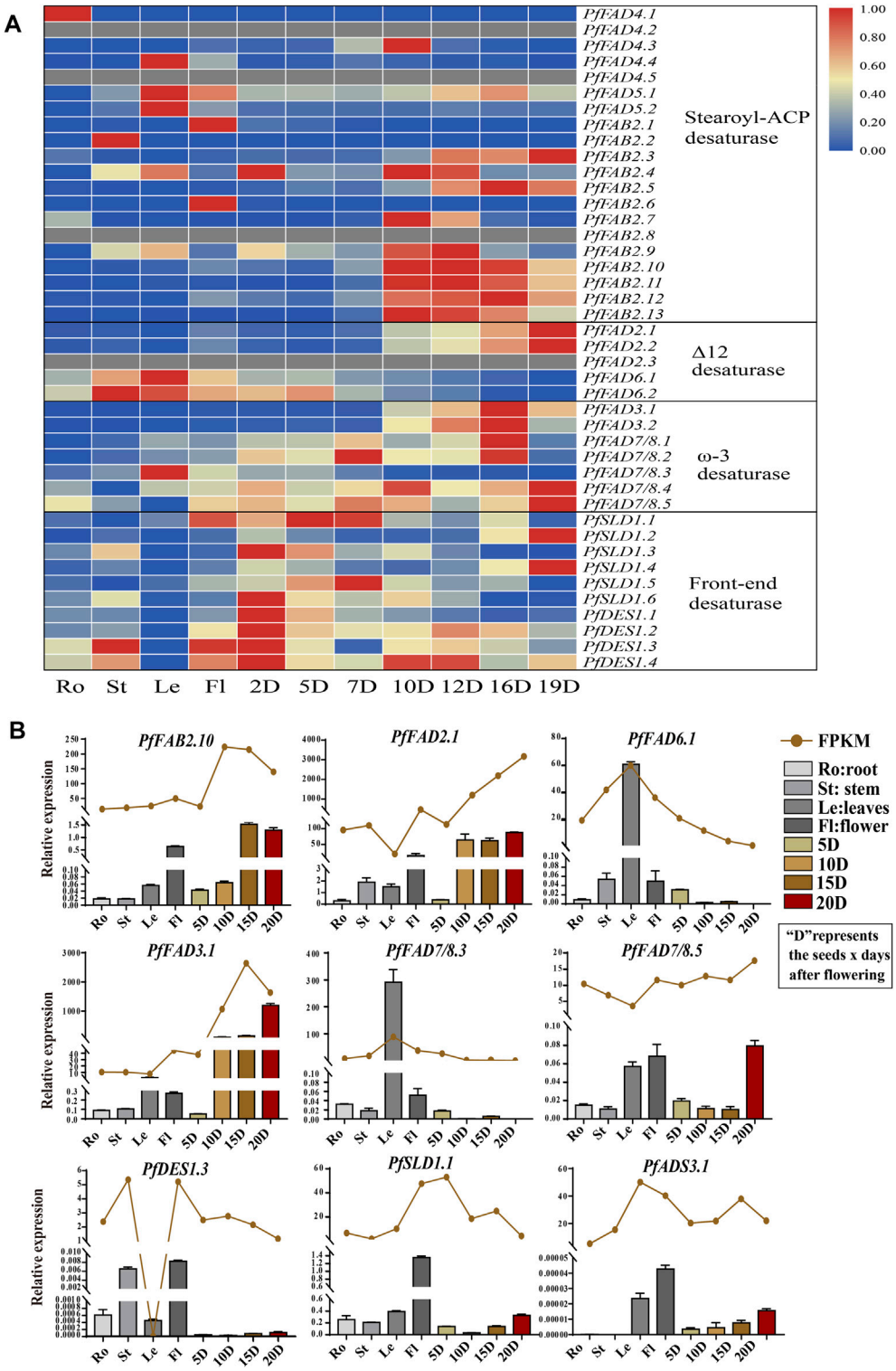


FIGURE 3 | Expression profiles of *FAD* genes in *P. frutescens*. **(A)** Expression patterns of *PfFADs* based on fragments per kilobase of transcript per million mapped reads (FPKM) from perilla. The color gradient blue–white–red indicates low to high levels of gene expression. **(B)** Relative expression levels of *PfFAD* genes determined by qRT-PCR and shown as FPKM. The perilla *Actin* gene was used as an internal control for normalization. Error bars represent SD from at least three biological replicates. Ro, roots; St, stems; Le, leaves; Fl, flowers; 2D–20D, seeds at 2–20 days after flowering.

types contain seven or eight exons and encode proteins with three TMDs, including an FA_desaturase domain and one longer DUF374 domain. Most of these genes contained four motifs (1, 2, and 3). However, there are two special *PfFAD7/8* genes (*PfFAD7/8.4* and *PfFAD7/8.5*) in perilla, which contain 16 exons, more than other *PfFAD7/8* genes in perilla, and other species (Xue et al., 2018). The proteins encoded by these genes contain chloroplast signals but no TMDs, and they also contain an Iso_dh domain in their N-termini in addition to the FA_desaturase domain.

Highly conserved histidine motifs are pivotal for maintaining the catalytic activity of FAD proteins. Based on our predictions, all three histidine boxes were present in the front-end desaturases, Δ12 desaturases, and ω-3 desaturases (**Supplementary Table S4**). Among the ω-3 desaturases, the histidine-rich motifs HDCGHG, HRTHH, and HVAHH were identified in PfFAD3 and HHDCGH/T(I)AVGHG, HRTHHQN, and HVAHH were identified in PfFAD7/8. The histidine-rich motifs HECG(A)HH, HR(D)RHH, and HVA (IP)HH were identified in Δ12 desaturases. PfFAD4 desaturase subfamily members also contain three histidine boxes: FQG (LD/Y)HH, HA/SWAH, and HAA (TK)HH. However, there are also exceptions. Among front-end desaturases, the histidine-rich motifs HDSGH, HNAHH, and QLEHHL were identified in all SLD1 proteins except PfSLD1.5, and the histidine-rich motifs HELSH, HLEHH, and HNEHH were identified in all DES1 proteins except PfDES1.4. Among Δ7/Δ9 desaturases, PfFAB2 of Δ9 desaturase subfamily members contain two highly conserved histidine-rich motifs, EENRHG and DEKRHE, but lack the third histidine-rich motif. PfADS3 of Δ7 desaturase subfamily members contain the HRHHH and HNNHH motifs but lack the first histidine-rich motif (Kazaz et al., 2020) (**Supplementary Table S4**).

The cis-regulatory elements in the promoter regions of *PfFAD* family genes were predicted using the online PlantCARE software (**Figure 2B**). Many light-responsive elements, hormone-responsive elements, and stress-responsive elements were identified in the promoters of *PfFAD* genes. Abscisic acid–responsive elements (ABREs) were widespread in the promoters of 30 *PfFAD* genes, and the ABRE motif was especially enriched in *PfFAD2.1* and *PfFAD3.1*. The GT1-motif and box4 light-responsive elements were widespread in the promoters of *PfFAD* genes. These results suggest that *PfFAD* genes might be involved in stress response and hormonal regulation.

Expression Analysis of *PfFAD* Family Genes

Based on transcriptome data from different tissues (**Figure 3A** and **Supplementary Table S5**), among Δ9 desaturases, most *PfFAB2* genes were highly expressed during the middle and later periods of seed development, except for *PfFAB2.1* and *PfFAB2.6*, which were expressed at very high levels in flowers, and *PfFAB2.2* was highly expressed in stems. Two *PfADS3* genes were highly expressed in leaves, and five *PfFAD4* genes were highly expressed in roots and seeds during the middle stages of development. Among genes in the front-end desaturases, *PfDES1* expression increased during early seed development, and *PfSLD1*

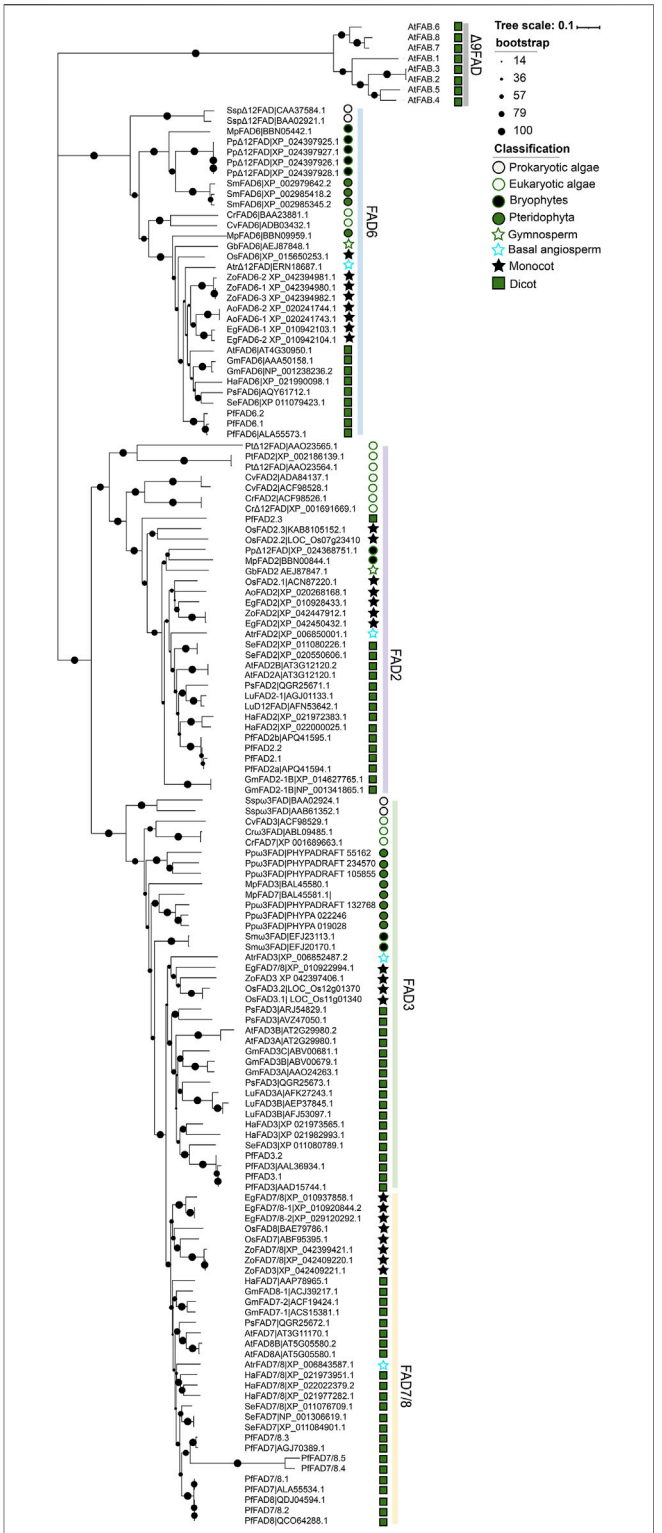


FIGURE 4 | Evolutionary analysis of PUFA desaturases. Δ9 desaturases from *A. thaliana* was used as the outgroup. The tree was constructed using MEGA X by the neighbor-joining method with 1,000 bootstraps. Detailed information on prokaryotic algae, eukaryotic algae, bryophytes, pteridophytes, and monocot and dicot plants is shown in **Supplementary Table S1**.

genes showed two major expression patterns during early and later seed development. Among genes in the Δ12 FAD subfamily, *PfFAD6.1* and *PfFAD6.2* were mainly expressed in leaves, stems, and flowers, and *PfFAD2.1* and *PfFAD2.2* expressions were increased during later seed development. *PfFAD2.3* was not expressed in all tissues. Among ω-3 FAD subfamily members, *PfFAD3.1* and *PfFAD3.2* were highly expressed during the middle and later stages of seed development, whereas *PfFAD7/8.3* expression increased in leaves. *PfFAD7/8.4* and *PfFAD7/8.5* expression increased during the early and later stages of seed development.

To verify the expression patterns, we examined tissue-specific gene expression patterns by qRT-PCR (**Figure 3B**). These patterns showed the same tendency as the FPKM. The expression of *PfFAD3.1* and *PfFAD2.1* increased during seed development, which were highly expressed during the later stages of seed development. *PfFAD7/8.3* and *PfFAD6.1* were highly expressed in leaves. *PfFAD7/8.5* was highly expressed in leaves, flower, and the later stage of seed development. *PfDES1.3*, *PfSLD1.1*, and *PfADS3.1* were expressed in flowers, and *PfFAB2.10* was expressed in flowers and during the later stages of seed development. The various expression patterns of *PfFAD* genes imply their functional differentiation.

Phylogenetic Analysis

To investigate the evolutionary relationships of PUFA desaturase genes in plants, we performed a phylogenetic analysis of Δ12 FAD and ω-3 FAD subfamily members in algae and plants using *A. thaliana* Δ9 desaturases (AtFAB2s) as the outgroup (**Figure 4** and **Supplementary Table S1**). FAD6 of the Δ12 desaturases differentiated earlier than FAD2 and ω-3 desaturases (FAD3 and FAD7/8). The ancestral Δ12 desaturases arose first in prokaryotic algae. Then, they independently differentiated to form Δ12 desaturases in bryophytes and pteridophytes and chloroplast-localized FAD6 in eukaryotic algae and higher plants. The endoplasmic reticulum-localized FAD2 derived from prokaryotic algae Δ12 desaturases. FAD2 independently evolved in eukaryotic algae and lost introns during evolution. ω-3

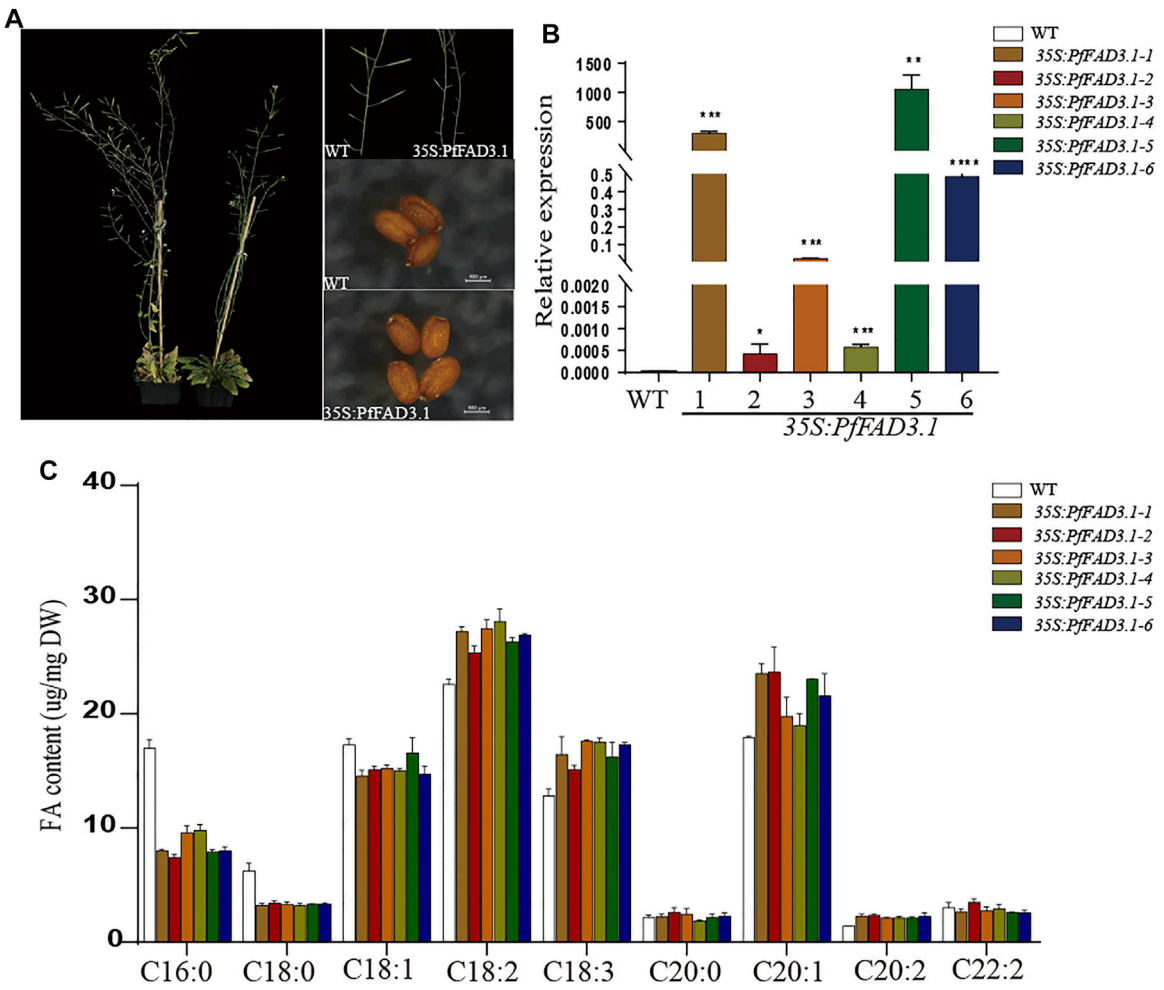


FIGURE 5 | Effect of transgenic *PfFAD3.1* expression on FA composition in *A. thaliana* seeds. **(A)** Transgenic CaMV35S::PfFAD3.1 Arabidopsis plants and seeds. **(B)** The expression of *PfFAD3.1* transgenic 35S::PfFAD3.1 Arabidopsis seeds determined by qRT-PCR. **(C)** FA composition in the seeds of transgenic 35S::PfFAD3.1 Arabidopsis. Error bars represent SD from at least three biological replicates. The asterisks indicate significant differences (*$p < 0.05$, Student's t-test).

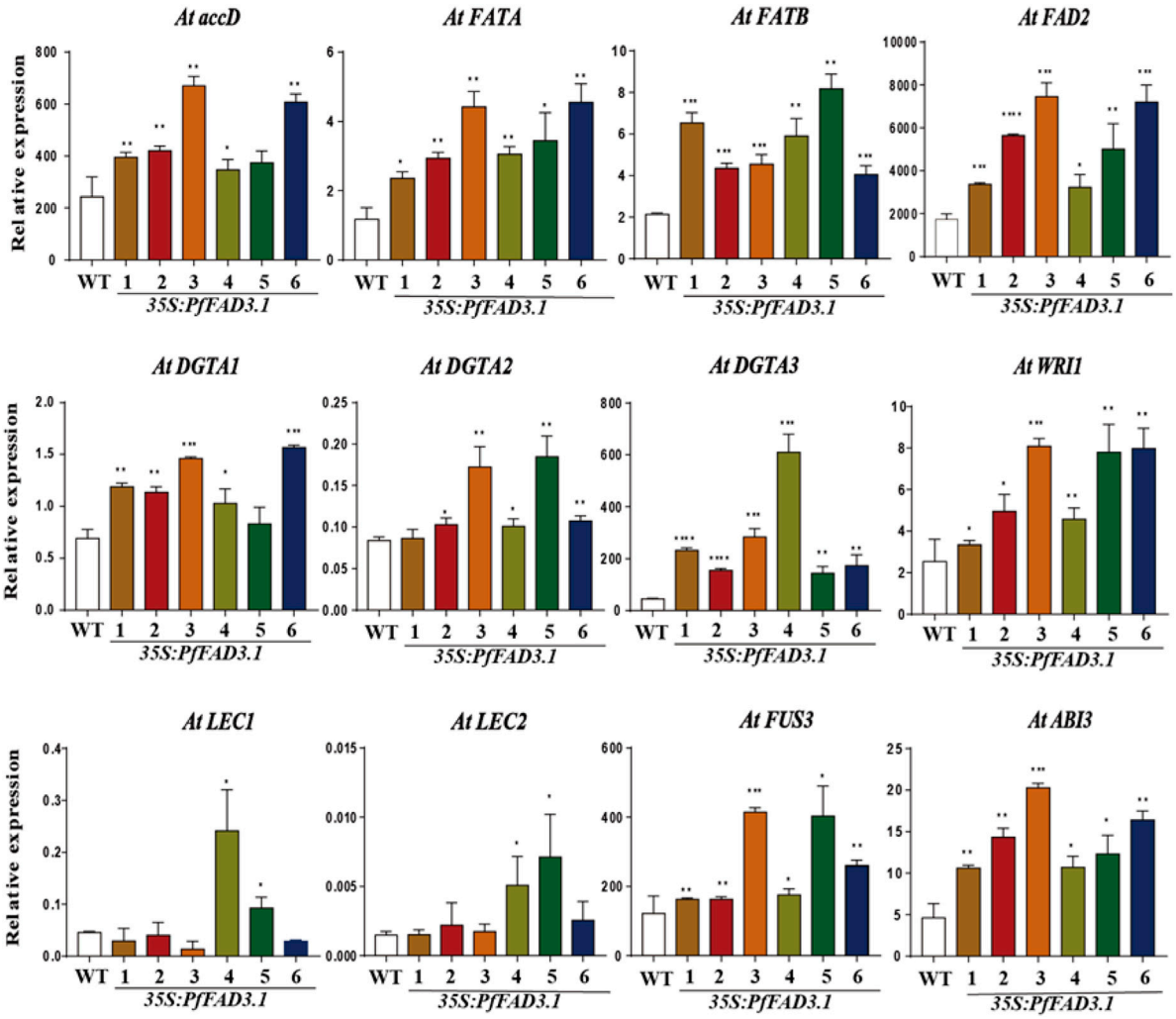


FIGURE 6 | The expression patterns of genes encoding key enzymes and transcription factors involved in α-linolenic acid biosynthesis in 35S:: PfFAD3.1 *Arabidopsis* seeds. Error bars represent SD from at least three biological replicates. The asterisks indicate significant differences (*$p < 0.05$, Student's t-test).

FAD also arose in prokaryotic algae. From prokaryotic algae to pteridophytes, the differentiation of ω-3 FAD was consistent with the evolution of terrestrial plants. A differentiation event subsequently occurred during the seed plant formation period to form the endoplasmic reticulum-localized FAD3 and the chloroplast-localized FAD7/8 groups in higher plants. Both Δ12 FAD and ω-3 FAD diverged earlier in monocotyledonous plants than dicotyledonous plants. Perilla FAD3 and FAD7/8 appear to be the latest diverged ω-3 FAD in dicotyledons. These findings shed light on the evolution and functional differentiation of FADs.

Heterologous Overexpression of *PfFAD3.1*

Based on previous reports and our genetic analysis, we speculated that *PfFAD3.1* is a key gene responsible for ALA biosynthesis in perilla seed. To further explore the function of *PfFAD3.1*, we amplified the full-length cDNA sequence and transformed it into *Arabidopsis* plants driven by the CaMV 35S promoter (**Supplementary Figure S2**) (Lee et al., 2016; Xue et al., 2018). After transgenic screening and identification, six T2-generation transgenic homozygous lines were obtained. Compared with the wild type, the phenotype of leaves and plants, seed size, and thousand seed weight of transgenic lines did not significantly vary (**Supplementary Figure S3**).

Then, we analyzed the composition of FAs in mature seeds of these plants using GC (**Figure 5C**). Compared with the wild type, the ratios of linoleic acid (C18:2) and ALA (C18:3) in transgenic CaMV35S::PfFAD3.1 *Arabidopsis* seeds were increased by 12.27%–24.36% and 17.68%–37.03%, respectively. Moreover, in the transgenic *Arabidopsis* seeds compared to the wild type, the ratios of the other long-chain unsaturated FAs (C20:1 and C20:2) were also higher (6.02%–32.13% and 49.50%–65.35%), and the ratios of palmitic acid (C16:0), stearic acid (C18:0), and oleic acid (C18:1) were decreased by 42.34%–56.29%, 45.58%–48.87%, and 3.97–15.85%, respectively.

To investigate the possible role of *PfFAD3.1* in regulating seed lipid biosynthesis, we examined the expression levels of genes involved in lipid biosynthesis and regulation in *Arabidopsis* seeds by qRT-PCR (**Figure 6**). Genes encoding key enzymes involved in lipid biosynthesis, including the acetyl-CoA carboxylase (ACCase) gene *AtaccD*, fatty acyl-ACP thioesterase (FAT) genes *AtFATA* and *AtFATB*, diacylglycerol acyltransferase (DGAT) genes *AtDGTA1-3*, and fatty acid desaturase 2 gene *AtFAD2*, were upregulated to different degrees in seeds of six transgenic lines. *WRINKLED1* (*WRI1*), *LEAFY COTYLEDON1* (*LEC1*) and *LEC2*, *ABSCISIC ACID INSENSITIVE3* (*ABI3*), and *FUSCA3* (*FUS3*) were reported as positive regulators of seed maturation and seed storage lipid accumulation in *Arabidopsis* (Mendoza et al., 2005; Mu et al., 2008; Elahi et al., 2016). The expression levels of *WRI1*, *ABI3*, and *FUS3* were upregulated in all transgenic *PfFAD3.1 Arabidopsis* lines, and *LEC1* and *LEC2* expressions were upregulated in multiple parts of the transgenic plants.

DISCUSSION

FADs catalyze PUFA biosynthesis, which are involved in important physiological processes in the plant kingdom (Alonso et al., 2003). In this study, we identified 42 FAD protein-coding genes in the perilla genome. Similar to FADs in other species, the PfFADs were classified into five subfamilies. Members of each subfamily generally had similar exon/intron structures, conserved domain sequences, and similar subcellular localizations. Δ9 desaturases are soluble proteins located in chloroplasts (Kazaz et al., 2020), which were predicted to possess no TMDs in perilla. Δ12 desaturases and ω-3 desaturases included both endoplasmic reticulum-localized and chloroplast-localized genes (Lee et al., 2016). In the perilla genome, two kinds of desaturases were also identified. In previous reports, the endoplasmic reticulum-localized FAD2 and FAD3 were highly expressed in seeds and catalyze ALA accumulation in triacylglycerol biosynthesis in plant seeds. Similarly, *PfFAD3* and *PfFAD2* expressions were increased during seed development in perilla (Rajwade et al., 2014; Xue et al., 2018). FAD6 and FAD7/8 encoded chloroplast-localized proteins that play non-redundant roles in plant responses to low temperature and drought stresses in previous studies (Ángela et al., 2015; Xue et al., 2021). Here, *PfFAD6* and *PfFAD7/8* were highly expressed in perilla leaves. The expression patterns of Δ12 and ω-3 desaturases were in accordance with their proposed functions. Interestingly, *PfFAD7/8.4* and *PfFAD7/8.5* were predicted as special ω-3 desaturases, which possess various gene structures (contained 16 exons and encoded proteins lacking TMD motifs) and expression patterns (upregulated expression in leaves and flowers and in the later stage of seed development) compared with other ω-3 desaturases.

Previous evolutionary analysis of FADs suggested that they share a common origin, and the Δ9 desaturases are the ancestor of membrane desaturase genes (Alonso et al., 2003; Nitta et al., 2005). In the present study, phylogenetic analysis showed that Δ12 and ω-3 desaturases share a common origin. The chloroplast-localized *FAD6* is previous PUFA desaturases, which diverged before *FAD2* and ω-3 desaturase genes. Most *ω-3FAD* genes are present in single copy in algae before evolution during land plant colonization (Xue et al., 2018). In this work, inner differentiation events of ω-3FAD appeared to occur during seed plants formation.

The phylogenetic relationships among PUFA desaturases are strongly associated with their sub-functionalization. Δ9 desaturases, which catalyze the conversion of stearic acid (C18:0) to oleic acid (C18:1), are the most ancestral group of FADs (Kachroo et al., 2007; Zhang et al., 2015). Δ12 desaturases, which differentiated from Δ9 desaturases, catalyze the second desaturase step from oleic acid (C18:1) to linoleic acid (C18:2) (Okuley et al., 1994). ω-3 desaturases, as latter differentiated from Δ12 desaturases, catalyze the desaturase step from linoleic acid (C18:2) to ALA (C18:3) (Shah et al., 1997). Δ9 desaturases are the only desaturases present in most animals and plants, whereas Δ12 and ω-3 desaturases are absent in animal evolutionary lineages.

FAD3 is a primary enzyme catalyzing ALA (C18:3) production in plant seeds. In previous reports, the expression levels of *AtFAD3* are connected with the seed C18:3 accumulation in *Arabidopsis* (Shah et al., 1997). In *Arabidopsis* seeds of overexpression *AtFAD3*, ALA contents decreased by approximately 20% (James and Dooner, 1990; Lemieux et al., 1990). Perilla seeds contain high levels of ALA. In this work, we observed high levels of expression of *PfFAD3.1* during seed development. The *PfFAD3.1* were cloned, which encoded proteins sharing 100% homology to *PfFAD3b* and *PfrFAD3-2* in previous reports (**Supplementary Figure S4**) (Lee et al., 2016; Xue et al., 2018). *PfFAD3b* was previously cloned and transformed into the budding yeast *Saccharomyces cerevisiae*. However, the transformed yeast produced limited amounts of ALA (1.3% ALA content) (Lee et al., 2016). These results showed that the yeast system does not explain well the function of *PfFAD3.1*. Hence, we constructed the 35S::PfFAD3.1 overexpression vector and transformed it into *Arabidopsis*. GC analysis showed that the ratios of ALA were obviously increased in 35S::PfFAD3.1-overexpressing *Arabidopsis* seeds. In a previous report, overexpression of tree peony *PsFAD3* significantly increased (0.6–1.5 times) the ALA content in seeds from five positive transgenic lines. The accumulation of ALA was accompanied with decreased palmitic acid and increased stearic acid and oleic acid (Yin et al., 2018). In the work, the obviously increased ALA contents were along with upregulated linoleic acid (C18:2) and other long-chain PUFA contents and downregulated palmitic acid (C16:0), stearic acid (C18:0), and oleic acid (C18:1) contents in transgenic *Arabidopsis* seeds. The same conclusion showed that *PfFAD3.1* possibly regulated FA metabolic flux, which not only increased ALA biosynthesis and PUFA content but also reduced saturated and monounsaturated FA contents in the heterologous plant system.

Seed maturation in higher plants is associated with the deposition of storage reserves, such as oil accumulation in oilseed plants (Bryant et al., 2019). Several important genes and transcription factors were involved in oil accumulation and seed maturation. ACCase catalyzes the formation of

malonyl-CoA, an essential component in FA biosynthesis (Madoka et al., 2002). FAT is a core enzyme in the FA biosynthesis process that determines FA chain length. DGAT catalyzes the covalent addition of a fatty acyl chain to diacylglycerol. In our work, those key genes involved in lipid biosynthesis were expressed at higher levels in transgenic 35S::PfFAD3.1 *Arabidopsis* plants. *WRI1* of the APETALA2/ethylene-responsive (AP2/EREB) family was the key transcription factor regulating seed maturity and FA biosynthesis (Baud et al., 2007). LEC1, an NF-Y family trimeric transcription factor, activates the B3-domain protein LEC2 during early seed development and targets *ABI3* and *FUS3* during the later stages of embryogenesis in *Arabidopsis* (Boulard et al., 2017). In previous reports, *LEC2* and *FUS3* can target *WRI1* to regulate FA biosynthesis (Wang and Perry, 2013; Na et al., 2019). In our work, *WRI1*, *ABI3*, *FUS3*, *LEC1*, and *LEC2* were upregulated in the transgenic seeds. These results suggest that the heterogeneous overexpression of PfFAD3.1 can influence the expression of key biosynthesis genes and transcription factors involved in FA biosynthesis in seed. These results suggest that the heterogeneous overexpression of PfFAD3.1 can influence the expression of key biosynthesis genes and transcription factors involved in FA biosynthesis in seed. However, the detailed molecular mechanisms remain unclean. We will prioritize the study of the mechanism by which *PfFAD3.1* may cause genetic changes related to seed development and will also explore the role of other *PfFADs* in the synthesis of ALA. It is expected to build a foundation for elucidating the functions of perilla FADs and ALA biosynthesis.

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

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

FUNDING

This work was funded by the National Natural Science Foundation for regional fund (31860391), Guizhou Province Science and Technology Plan Project (Qian Science Platform (2019)5656), and Qian Academy of Agricultural Sciences after the national foundation subsidy (2021) No. 29 and Guizhou Province High-level Innovative Talent Training Program Project ((2016) 4003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.735862/full#supplementary-material

REFERENCES

Abdel-reheem, M., and Hildebrand, D. (2013). Activity of Brassica Napus and Perilla Frutescens Microsomal ω-3 Desaturases Expressed in Yeast (*Saccharomyces cerevisiae*). *Turk J. Biol.* 37, 591–605. doi:10.3906/biy-1301-32

Anai, T., Yamada, T., Kinoshita, T., Rahman, S. M., and Takagi, Y. (2005). Identification of Corresponding Genes for Three Low-α-Linolenic Acid Mutants and Elucidation of Their Contribution to Fatty Acid Biosynthesis in Soybean Seedfication of Corresponding Genes for Three Low-A-Linolenic Acid Mutantsand Elucidation of Their Contribution to Fatty Acid Biosynthesis in Soybean Seed. *Plant Sci.* 168, 1615–1623. doi:10.1016/j.plantsci.2005.02.016

Ángela, R., María, L. H., Ángel, S.-G., Sara, L. G., Beatriz, L., Picorel, R., et al. (2015). Non-redundant Contribution of the Plastidial FAD8 ω-3 Desaturase to Glycerolipid Unsaturation at Different Temperatures in Arabidopsis. *Mol Plant* 8, 1599–1611. doi:10.1016/j.molp.2015.06.004

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., et al. (2012). ExPASy: SIB Bioinformatics Resource portal. *Nucleic Acids Res.* 40, W597–W603. doi:10.1093/nar/gks400

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

Baker, E. J., Miles, E. A., Burdge, G. C., Yaqoob, P., and Calder, P. C. (2016). Metabolism and Functional Effects of Plant-Derived omega-3 Fatty Acids in Humans. *Prog. Lipid Res.* 64, 30–56. doi:10.1016/j.plipres.2016.07.002

Baud, S., Mendoza, M. S., To, A., Harscoët, E., Lepiniec, L., and Dubreucq, B. (2007). WRINKLED1 Specifies the Regulatory Action of LEAFY COTYLEDON2 towards Fatty Acid Metabolism during Seed Maturation in Arabidopsis. *Plant J.* 50, 825–838. doi:10.1111/j.1365-313X.2007.03092.x

Berestovoy, M. A., Pavlenko, O. S., and Goldenkova-Pavlova, I. V. (2020). Plant Fatty Acid Desaturases: Role in the Life of Plants and Biotechnological Potential. *Biol. Bull. Rev.* 10, 127–139. doi:10.1134/S2079086420020024

Bhunia, R. K., Chakraborty, A., Kaur, R., Maiti, M. K., and Sen, S. K. (2016). Enhancement of α-linolenic Acid Content in Transgenic Tobacco Seeds by Targeting a Plastidial ω-3 Fatty Acid Desaturase (Fad7) Gene of Sesamum indicum to ER. *Plant Cel Rep* 35, 213–226. doi:10.1007/s00299-015-1880-z

Boulard, C., Fatihi, A., Lepiniec, L., and Dubreucq, B. (2017). Regulation and Evolution of the Interaction of the Seed B3 Transcription Factors with NF-Y Subunits. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1860, 1069–1078. doi:10.1016/j.bbagrm.2017.08.008

Bryant, F. M., Hughes, D., Hassani-Pak, K., and Eastmond, P. J. (2019). Basic LEUCINE ZIPPER TRANSCRIPTION FACTOR67 Transactivates DELAY of GERMINATION1 to Establish Primary Seed Dormancy in Arabidopsis. *Plant Cell* 31, 1276–1288. doi:10.1105/tpc.18.00892

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* 13, 1194–1202. doi:10.1016/j.molp.2020.06.009

Chou, K.-C., and Shen, H.-B. (2008). Cell-PLoc: a Package of Web Servers for Predicting Subcellular Localization of Proteins in Various Organisms. *Nat. Protoc.* 3, 153–162. doi:10.1038/nprot.2007.494

Ciftci, O. N., Przybylski, R., and Rudzińska, M. (2012). Lipid Components of Flax, Perilla, and Chia Seeds. *Eur. J. Lipid Sci. Technol.* 114, 794–800. doi:10.1002/ejlt.201100207

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

Elahi, N., Duncan, R. W., and Stasolla, C. (2016). Molecular Regulation of Seed Oil Accumulation. *J. Adv. Nutr. Hum. Metab.* 2, e1296. doi:10.14800/janhm

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam Protein Families Database: towards a More Sustainable Future. *Nucleic Acids Res.* 44, D279–D285. doi:10.1093/nar/gkv1344

Gao, J., AjjawiManoli, I., Manoli, A., Sawin, A., Xu, C., Froehlich, J. E., et al. (2009). FATTY ACID DESATURASE4 of *Arabidopsis* Encodes a Protein Distinct from Characterized Fatty Acid Desaturases. *Plant J.* 60, 832–839. doi:10.1111/j.1365-313X.2009.04001.x

Horn, P. J., Smith, M. D., Clark, T. R., Froehlich, J. E., and Benning, C. (2020). PEROXIREDOXIN Q Stimulates the Activity of the Chloroplast 16:1 Δ3trans FATTY ACID DESATURASE4. *Plant J.* 102, 718–729. doi:10.1111/tpj.14657

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2014). GSDS 2.0: an Upgraded Gene Feature Visualization Server. *Bioinformatics* 31, 1296–1297. doi:10.1093/bioinformatics/btu817

Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., et al. (1993). A Gene Encoding a Chloroplast omega-3 Fatty Acid Desaturase Complements Alterations in Fatty Acid Desaturation and Chloroplast Copy Number of the Fad7 Mutant of *Arabidopsis thaliana*. *J. Biol. Chem.* 268, 24099–24105. doi:10.1016/s0021-9258(20)80498-0

Igarashi, M., and Miyazaki, Y. (2013). A Review on Bioactivities of Perilla: Progress in Research on the Functions of Perilla as Medicine and Food. *Evidence-Based Complement. Altern. Med.* 2013, 1–7. doi:10.1155/2013/925342

James, D. W., and Dooner, H. K. (1990). Isolation of EMS-Induced Mutants in Arabidopsis Altered in Seed Fatty Acid Composition. *Theoret. Appl. Genet.* 80, 241–245. doi:10.1007/BF00224393

Kachroo, A., Shanklin, J., Whittle, E., Lapchyk, L., Hildebrand, D., and Kachroo, P. (2006). The Arabidopsis Stearoyl-Acyl Carrier Protein-Desaturase Family and the Contribution of Leaf Isoforms to Oleic Acid Synthesis. *Plant Mol. Biol.* 63, 257–271. doi:10.1007/s11103-006-9086-y

Kazaz, S., Barthole, G., Domergue, F., Ettaki, H., To, A., Vasselon, D., et al. (2020). Differential Activation of Partially Redundant Δ9 Stearoyl-ACP Desaturase Genes Is Critical for Omega-9 Monounsaturated Fatty Acid Biosynthesis during Seed Development in Arabidopsis. *Plant Cell* 32, 3613–3637. doi:10.1105/tpc.20.00554

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, K.-R., Lee, Y., Kim, E.-H., Lee, S.-B., Roh, K. H., Kim, J.-B., et al. (2016). Functional Identification of Oleate 12-desaturase and ω-3 Fatty Acid Desaturase Genes from Perilla Frutescens Var. Frutescens. *Plant Cel Rep* 35, 2523–2537. doi:10.1007/s00299-016-2053-4

Lemieux, B., Miquel, M., Somerville, C., and Browse, J. (1990). Mutants of Arabidopsis with Alterations in Seed Lipid Fatty Acid Composition. *Theoret. Appl. Genet.* 80, 234–240. doi:10.1007/BF00224392

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Peer, Y. V. d., et al. (2002). PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a portal to Tools for In Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* 30, 325–327. doi:10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) V4: Recent Updates and New Developments. *Nucleic Acids Res.* 47, W256–W259. doi:10.1093/nar/gkz239

Li, S.-S., Wang, L.-S., Shu, Q.-Y., Wu, J., Chen, L.-G., Shao, S., et al. (2015). Fatty Acid Composition of Developing Tree Peony (*Paeonia Section Moutan* DC.) Seeds and Transcriptome Analysis during Seed Development. *BMC Genomics* 16, 208. doi:10.1186/s12864-015-1429-0

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

López Alonso, D., García-Maroto, F., Rodríguez-Ruiz, J., Garrido, J. A., and Vilches, M. A. (2003). Evolution of the Membrane-Bound Fatty Acid Desaturases. *Biochem. Syst. Ecol.* 31, 1111–1124. doi:10.1016/S0305-1978(03)00041-3

Madoka, Y., Tomizawa, K.-I., Mizoi, J., Nishida, I., Nagano, Y., and Sasaki, Y. (2002). Chloroplast Transformation with Modified accD Operon Increases Acetyl-CoA Carboxylase and Causes Extension of Leaf Longevity and Increase in Seed Yield in Tobacco. *Plant Cel Physiol* 43, 1518–1525. doi:10.1093/pcp/pcf172

Manan, S., Chen, B., She, G., Wan, X., and Zhao, J. (2017). Transport and Transcriptional Regulation of Oil Production in Plants. *Crit. Rev. Biotechnol.* 37, 641–655. doi:10.1080/07388551.2016.1212185

Matsuda, O., Sakamoto, H., Hashimoto, T., and Iba, K. (2005). A Temperature-Sensitive Mechanism that Regulates Post-translational Stability of a Plastidial ω-3 Fatty Acid Desaturase (FAD8) in Arabidopsis Leaf Tissues. *J. Biol. Chem.* 280, 3597–3604. doi:10.1074/jbc.M407226200

Meesapyodsuk, D., and Qiu, X. (2012). The Front-End Desaturase: Structure, Function, Evolution and Biotechnological Use. *Lipids* 47, 227–237. doi:10.1007/s11745-011-3617-2

Mu, J., Tan, H., Zheng, Q., Fu, F., Liang, Y., Zhang, J., et al. (2008). LEAFY COTYLEDON1 Is a Key Regulator of Fatty Acid Biosynthesis in Arabidopsis. *Plant Physiol.* 148, 1042–1054. doi:10.1104/pp.108.126342

Mungmai, L., Preedalikit, W., Aunsri, N., and Amornlerdpison, D. (2020). Efficacy of Cosmetic Formulation Containing Perilla Frutescens Leaves Extract for Irritation and Aging Skin. *Biomed. Pharmacol. J.* 13, 779–787. doi:10.13005/bpj/1942

Na, G., Mu, X., Grabowski, P., Schmutz, J., and Lu, C. (2019). Enhancing Micro RNA 167A Expression in Seed Decreases the α-linolenic Acid Content and Increases Seed Size in Camelina Sativa. *Plant J.* 98, 346–358. doi:10.1111/tpj.14223

Nitta, M., Lee, J. K., Kang, C. W., Katsuta, M., Yasumoto, S., Liu, D., et al. (2005). The Distribution of Perilla Species. *Genet. Resour. Crop Evol.* 52, 797–804. doi:10.1007/s10722-003-6017-5

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., and Browse, J. (1994). Arabidopsis FAD2 Gene Encodes the Enzyme that Is Essential for Polyunsaturated Lipid Synthesis. *Plant Cell* 6, 147–158. doi:10.1105/tpc.6.1.147

Park, Y.-J., Dixit, A., Ma, K.-H., Lee, J.-K., Lee, M.-H., Chung, C.-S., et al. (2008). Evaluation of Genetic Diversity and Relationships within an On-Farm Collection of Perilla Frutescens (L.) Britt. Using Microsatellite Markers. *Genet. Resour. Crop Evol.* 55, 523–535. doi:10.1007/s10722-007-9258-x

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER Web Server: 2018 Update. *Nucleic Acids Res.* 46, W200–W204. doi:10.1093/nar/gky448

Rajwade, A. V., Kadoo, N. Y., Borikar, S. P., Harsulkar, A. M., Ghorpade, P. B., and Gupta, V. S. (2014). Differential Transcriptional Activity of SAD, FAD2 and FAD3 Desaturase Genes in Developing Seeds of Linseed Contributes to Varietal Variation in α-linolenic Acid Content. *Phytochemistry* 98, 41–53. doi:10.1016/j.phytochem.2013.12.002

Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L. (2005). LEAFY COTYLEDON 2 Activation Is Sufficient to Trigger the Accumulation of Oil and Seed Specific mRNAs inArabidopsisleaves. *FEBS Lett.* 579, 4666–4670. doi:10.1016/j.febslet.2005.07.037

Shah, S., Xin, Z., and Browse, J. (1997). Overexpression of the FAD3 Desaturase Gene in a Mutant of Arabidopsis. *Plant Physiol.* 114, 1533–1539. doi:10.1104/pp.114.4.1533

Smith, M. A., Dauk, M., Ramadan, H., Yang, H., Seamons, L. E., Haslam, R. P., et al. (2012). Involvement of Arabidopsis ACYL-COENZYME A DESATURASE-LIKE2 (At2g31360) in the Biosynthesis of the Very-Long-Chain Monounsaturated Fatty Acid Components of Membrane Lipids. *Plant Physiol.* 161, 81–96. doi:10.1104/pp.112.202325

Sperling, P., Ternes, P., Zank, T. K., and Heinz, E. (2003). The Evolution of Desaturases. *Prostaglandins, Leukot. Essent. Fatty Acids* 68, 73–95. doi:10.1016/s0952-3278(02)00258-2

Vrinten, P., Hu, Z., Munchinsky, M.-A., Rowland, G., and Qiu, X. (2005). Two FAD3 Desaturase Genes Control the Level of Linolenic Acid in Flax Seed. *Plant Physiol.* 139, 79–87. doi:10.1104/pp.105.064451

Wang, F., and Perry, S. E. (2013). Identification of Direct Targets of FUSCA3, a Key Regulator of Arabidopsis Seed Development. *Plant Physiol.* 161, 1251–1264. doi:10.1104/pp.112.212282

Watahiki, M. K., and Yamamoto, K. T. (1994). A New Isozyme of Plastid [omega]-3 Fatty Acid Desaturase in *Arabidopsis thaliana*. *Plant Physiol.* 105, 1451–1452. doi:10.1104/pp.105.4.1451

Weber, H. (2002). Fatty Acid-Derived Signals in Plants. *Trends Plant Sci.* 7, 217–224. doi:10.1016/s1360-1385(02)02250-1

Xue, Y., Chen, B., Win, A. N., Fu, C., Lian, J., Liu, X., et al. (2018). Omega-3 Fatty Acid Desaturase Gene Family from Two ω-3 Sources, Salvia Hispanica and Perilla Frutescens: Cloning, Characterization and Expression. *PLoS One* 13, e0191432. doi:10.1371/journal.pone.0191432

Xue, Y., Jiang, H., Chen, B., Jiang, J., Lin, N., Li, L., et al. (2021). Molecular Cloning and Characterization of Two Plastidial Δ12 Fatty Acid Desaturase cDNAs from Perilla Frutescens. *Acta Physiol. Plant* 43, 103. doi:10.1007/s11738-021-03277-3

Yang, Q., Fan, C., Guo, Z., Qin, J., Wu, J., Li, Q., et al. (2012). Identification of FAD2 and FAD3 Genes in Brassica Napus Genome and Development of Allele-specific Markers for High Oleic and Low Linolenic Acid Contents. *Theor. Appl. Genet.* 125, 715–729. doi:10.1007/s00122-012-1863-1

Yin, D.-D., Xu, W.-Z., Shu, Q.-Y., Li, S.-S., Wu, Q., Feng, C.-Y., et al. (2018). Fatty Acid Desaturase 3 (PsFAD3) from Paeonia Suffruticosa Reveals High α-linolenic Acid Accumulation. *Plant Sci.* 274, 212–222. doi:10.1016/j.plantsci.2018.05.027

Zhang, T., Song, C., Song, L., Shang, Z., Yang, S., Zhang, D., et al. (2017). RNA Sequencing and Coexpression Analysis Reveal Key Genes Involved in α-Linolenic Acid Biosynthesis in Perilla Frutescens Seed. *Int. J. Mol Sci.* 18, 2433. doi:10.3390/ijms18112433

Zhang, Y., Maximova, S. N., and Guiltinan, M. J. (2015). Characterization of a Stearoyl-Acyl Carrier Protein Desaturase Gene Family from Chocolate Tree, Theobroma Cacao L. *Front. Plant Sci.* 6, 239. doi:10.3389/fpls.2015.00239

Zhang, Y., Shen, Q., Leng, L., Zhang, D., Chen, S., Shi, Y., et al. (2021). Incipient Diploidization of the Medicinal Plant Perilla within 10,000 Years. *Nat. Commun.* 12, 5508. doi:10.1038/s41467-021-25681-6

Zhou, Z., Wang, M.-J., Hu, J.-J., Lu, M.-Z., and Wang, J. H. (2010). Improve Freezing Tolerance in Transgenic poplar by Overexpressing a ω-3 Fatty Acid Desaturase Gene. *Mol. Breed.* 25, 571–579. doi:10.1007/s11032-009-9352-1

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

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

Copyright © 2021 Duan, Shi-Mei, Zhi-Wei, Jing, De-Gang, Hong-Bin and Qi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fgene.2021.737293

Genome-Wide Identification and Co-Expression Analysis of *ARF* and *IAA* Family Genes in *Euscaphis konishii*: Potential Regulators of Triterpenoids and Anthocyanin Biosynthesis

Bobin Liu[1,2,3], Juanli Zhu[2,3], Lina Lin[2,3], Qixin Yang[2,3], Bangping Hu[2,3], Qingying Wang[2,3], Xiao-Xing Zou[2,3]* and Shuang-Quan Zou[2,3]*

[1]Jiangsu Key Laboratory for Bioresources of Saline Soils, Jiangsu Synthetic Innovation Center for Coastal Bio-agriculture, School of Wetlands, Yancheng Teachers University, Yancheng, China, [2]College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, China, [3]Fujian Colleges and Universities Engineering Research Institute for Conservation and Utilization of Natural Bioresources, Fuzhou, China

OPEN ACCESS

Edited by:
Sreepriya Pramod,
Altria, United States

Reviewed by:
Hui Song,
Qingdao Agricultural University, China
Yulong Ren,
Institute of Crop Sciences (CAAS), China
Yourong Chai,
Southwest University, China
Xiaojiao Han,
Chinese Academy of Forestry, China

***Correspondence:**
Bobin Liu
liubb@yctu.edu.cn
Xiao-Xing Zou
zxx0299@fafu.edu.cn

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

Received: 06 July 2021
Accepted: 18 November 2021
Published: 05 January 2022

Citation:
Liu B, Zhu J, Lin L, Yang Q, Hu B, Wang Q, Zou X-X and Zou S-Q (2022) Genome-Wide Identification and Co-Expression Analysis of ARF and IAA Family Genes in Euscaphis konishii: Potential Regulators of Triterpenoids and Anthocyanin Biosynthesis. Front. Genet. 12:737293. doi: 10.3389/fgene.2021.737293

Euscaphis konishii is an evergreen plant that is widely planted as an industrial crop in Southern China. It produces red fruits with abundant secondary metabolites, giving *E. konishii* high medicinal and ornamental value. Auxin signaling mediated by members of the AUXIN RESPONSE FACTOR (ARF) and auxin/indole-3-acetic acid (Aux/IAA) protein families plays important roles during plant growth and development. *Aux/IAA* and *ARF* genes have been described in many plants but have not yet been described in *E. konishii*. In this study, we identified 34 EkIAA and 29 EkARF proteins encoded by the *E. konishii* genome through database searching using HMMER. We also performed a bioinformatic characterization of *EkIAA* and *EkARF* genes, including their phylogenetic relationships, gene structures, chromosomal distribution, and *cis*-element analysis, as well as conserved motifs in the proteins. Our results suggest that *EkIAA* and *EkARF* genes have been relatively conserved over evolutionary history. Furthermore, we conducted expression and co-expression analyses of *EkIAA* and *EkARF* genes in leaves, branches, and fruits, which identified a subset of seven *EkARF* genes as potential regulators of triterpenoids and anthocyanin biosynthesis. RT-qPCR, yeast one-hybrid, and transient expression analyses showed that EkARF5.1 can directly interact with auxin response elements and regulate downstream gene expression. Our results may pave the way to elucidating the function of *EkIAA* and *EkARF* gene families in *E. konishii*, laying a foundation for further research on high-yielding industrial products and *E. konishii* breeding.

Keywords: *Euscaphis konishii*, Aux/IAA, ARF, triterpenoids, anthocyanin, fruit development

Abbreviations: AbA, antibiotic aureobasidin A; AD, activation domain; ARF, auxin response factor; Aux/IAA, auxin/indole-3-acetic acid; AuxRE, auxin response element; BD, binding domain; CTD, C-terminal dimerization domain; DBD, DNA-binding domain; GA, gibberellic acid; HMM, Hidden Markov model; LG, linkage group; Ka, number of nonsynonymous substitutions per nonsynonymous site; Ks, number of synonymous substitutions per synonymous site; MR, middle region; pI, isoelectric point; SD, synthetic defined.

INTRODUCTION

Plant secondary metabolites are not only important to plant development, but are also significant as nutritional resources for humans, as sources for color and odorants, and as potential materials for drug discovery (Huang et al., 2020a; Bing et al., 2021). Secondary metabolites include terpenes, phenolic compounds, and alkaloids, which play essential roles as food additives and in medicine, cosmetics, skincare, and industrial chemicals. The biosynthesis of plant secondary metabolites is very complex and is regulated by various environmental signals and development cues. Because of their critical roles, much attention is being paid to the biosynthesis and regulatory mechanism of secondary metabolites.

The phytohormone auxin regulates a wide range of processes in plant growth and development, including vascular differentiation, lateral root formation, apical dominance, and fruit development (Liscum and Reed, 2002; Aloni et al., 2006). The auxin/indole-3-acetic acid (Aux/IAA) and AUXIN RESPONSE FACTOR (ARF) proteins are critical players of auxin signal transduction (Hagen, 2015). ARFs are transcription factors consisting of a DNA-binding domain (DBD), a middle transcriptional regulatory region (MR), and a dimerization domain at their C termini (CTD), with the MR serving as an activation or repressor domain (Tiwari et al., 2003; Guilfoyle and Hagen, 2007). ARF DBDs bind to auxin response elements (AuxREs) located in the promoter region of auxin-responsive genes to activate or repress their transcription, depending on the type of MR (Tiwari et al., 2003; Guilfoyle and Hagen, 2007). Aux/IAA proteins comprise four domains, denoted domains I, II, III, and IV (Tiwari et al., 2001). Located at the N terminus, domain I is characterized by the LxLxLx motif and is mainly responsible for repression of gene expression. Domain II mediates the degradation of the protein via the degron sequence (GWPPV) by the 26S proteasome (Hagen and Guilfoyle, 2002; Tiwari et al., 2004). Domains III and IV at the C terminus interact with the CTD of ARFs to form homo- and heterodimers, resulting in the transcriptional induction or repression of downstream auxin-responsive genes (Ulmasov et al., 1999; Guilfoyle, 2015). The interaction between Aux/IAA and ARF proteins also depends on auxin concentration. When auxin levels are low, ARFs bind to AuxREs, but transcription is repressed through their interaction with Aux/IAAs. With increasing auxin levels, Aux/IAAs are degraded via the 26S proteasome, thus alleviating the repressive effect of Aux/IAAs on ARFs (Gray et al., 2001; Winkler et al., 2017; Roosjen et al., 2018). Both Aux/IAA and ARF proteins are encoded by large gene families with distinct expression patterns and regulatory mechanisms, contributing to the complexity of auxin signaling (Luo et al., 2018).

Aux/IAA and ARF family members play critical and extensive roles during the entire plant life cycle. During embryogenesis, ARF5 is involved in organ formation in Arabidopsis (*Arabidopsis thaliana*) (Hardtke and Berleth, 1998). ARF19 and ARF7 redundantly regulate lateral root initiation (Wilmoth et al., 2005), while ARF1 and ARF2 control floral organ senescence and abscission (Ellis et al., 2005). Most of the Aux/IAA family members described in Arabidopsis, such as IAA3, IAA14, IAA28, and IAA19, affect the growth and development of lateral roots (Goh et al., 2012; Luo et al., 2018). Aux/IAAs and ARFs also play important roles in regulating fruit development. ARF8 was reported to play a negative role during fertilization and fruit initiation in Arabidopsis (Goetz et al., 2006). SlARF10 and SlARF6A are involved in chlorophyll and sugar accumulation during tomato (*Solanum lycopersicum*) fruit development (Yuan et al., 2018a; Yuan et al., 2019). Auxin signaling also interacts with other phytohormone signaling pathways to control fruit development. For example, the tomato gibberellic acid (GA) repressor SlDELLA interacts with SlARF7/SlIAA9 to regulate fruit initiation (Hu et al., 2018). Likewise, SlIAA3 contributes to differential growth by integrating auxin and ethylene signaling (Chaabouni et al., 2009). The accumulation of secondary metabolites is an important biological process during fruit ripening, which includes fruit color formation and biosynthesis of volatile odorants. In apple (*Malus domestica*), MdARF13 interacts with MdIAA121 to regulate anthocyanin biosynthesis (Wang et al., 2018). Overexpression of *MdIAA26* boosts anthocyanin biosynthesis in apple and Arabidopsis seedlings (Wang et al., 2020). In addition to anthocyanins, auxin induces the accumulation of flavonols by promoting the expression of the gene encoding a key biosynthetic enzyme (Lewis et al., 2011). Auxin signaling also participates in the accumulation of glucosinolates, a class of important plant defense metabolites (Mitreiter and Gigolashvili, 2021). Although auxin plays critical roles in secondary metabolite biosynthesis, how auxin signaling synergistically regulates the contents of multiple secondary metabolites is unclear.

Euscaphis plants belong to the Staphyleaceae family and are deciduous shrubs or small trees widely distributed in East Asia, from Japan to Southern China (Zhang et al., 2012). These plants constitute an important industrial crop due to the medicinal compounds and other industrially desirable products extracted from their fruits, leaves, and roots, such as triterpenes, phenolic acid, and flavonoids (Liang et al., 2018). In China, *E. konishii* is cultivated as an ornamental plant due to its beautiful red-winged pericarp and also has a long history of use as a medicinal plant to cure colds and allergies (Yuan et al., 2018b). The potential medical applications of *Euscaphis* were recently supported by data indicating that total phenolic and methanolic extracts of *Euscaphis* plants can mitigate liver fibrosis in mice and inhibit hepatic stem cell proliferation (Lee et al., 2009; Huang et al., 2020b), prompting the expansion of *E. konishii* cultivation (Sun et al., 2019). Therefore, *E. konishii* is an economically useful crop for the production of medicinal compounds. However, the molecular mechanisms governing the biosynthesis of these compounds in this species are poorly understood, which limits the breeding of improved *E. konishii* varieties. Although the Aux/IAA and ARF families have been systematically identified and characterized in many plants, including Arabidopsis (Remington et al., 2004; Okushima et al., 2005), rice (*Oryza sativa*) (Jain et al., 2006; Wang et al., 2007), poplar (*Populus trichocarpa*) (Kalluri et al., 2007), maize (*Zea mays*) (Wang et al., 2010; Xing et al., 2011), soybean (*Glycine max*) (Van Ha et al., 2013; Singh and Jain, 2015), and pepper (*Capsicum*

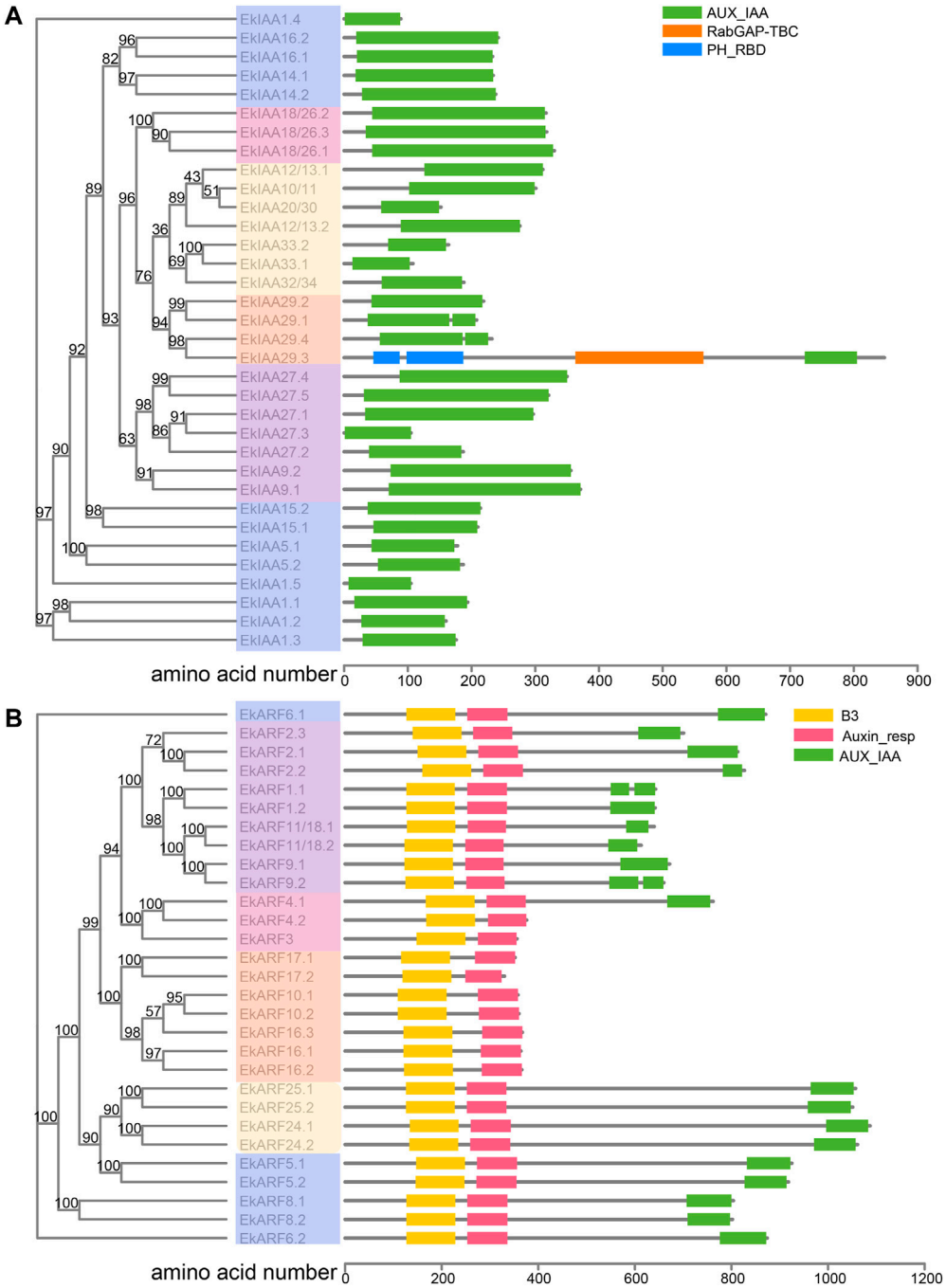


FIGURE 1 | Phylogenetic relationships and protein domain maps of *E. konishii* IAA **(A)** and ARF **(B)** proteins. Left, phylogenetic relationship of 34 *IAAs* and 29 *ARFs* from *E. konishii*. Different colors represent different groups. Right, IAA and ARF protein domain composition. Bootstrap values are shown close to branch nodes.

annuum) (Yu et al., 2017; Waseem et al., 2018), their functions in *E. konishii* are not clear. Given the importance of Aux/IAA and ARF proteins in plant development, we undertook a comprehensive survey of the *Aux/IAA* and *ARF* gene families in *E. konishii* to better understand the biosynthetic pathways of medicinal secondary metabolites. In this study, we identified 34 *Aux/IAA* and 29 *ARF* genes in *E. konishii* and analyzed their sequence features, phylogenetic relationships, *cis*-elements, and co-expression profiles. We also explored the function of EkARFs in the context of phenolic and anthocyanin biosynthesis. Our results may have uncovered a potential role for auxin in the biosynthesis of secondary products, which may provide useful information for breeding *E. konishii* with a high content of medicine compounds.

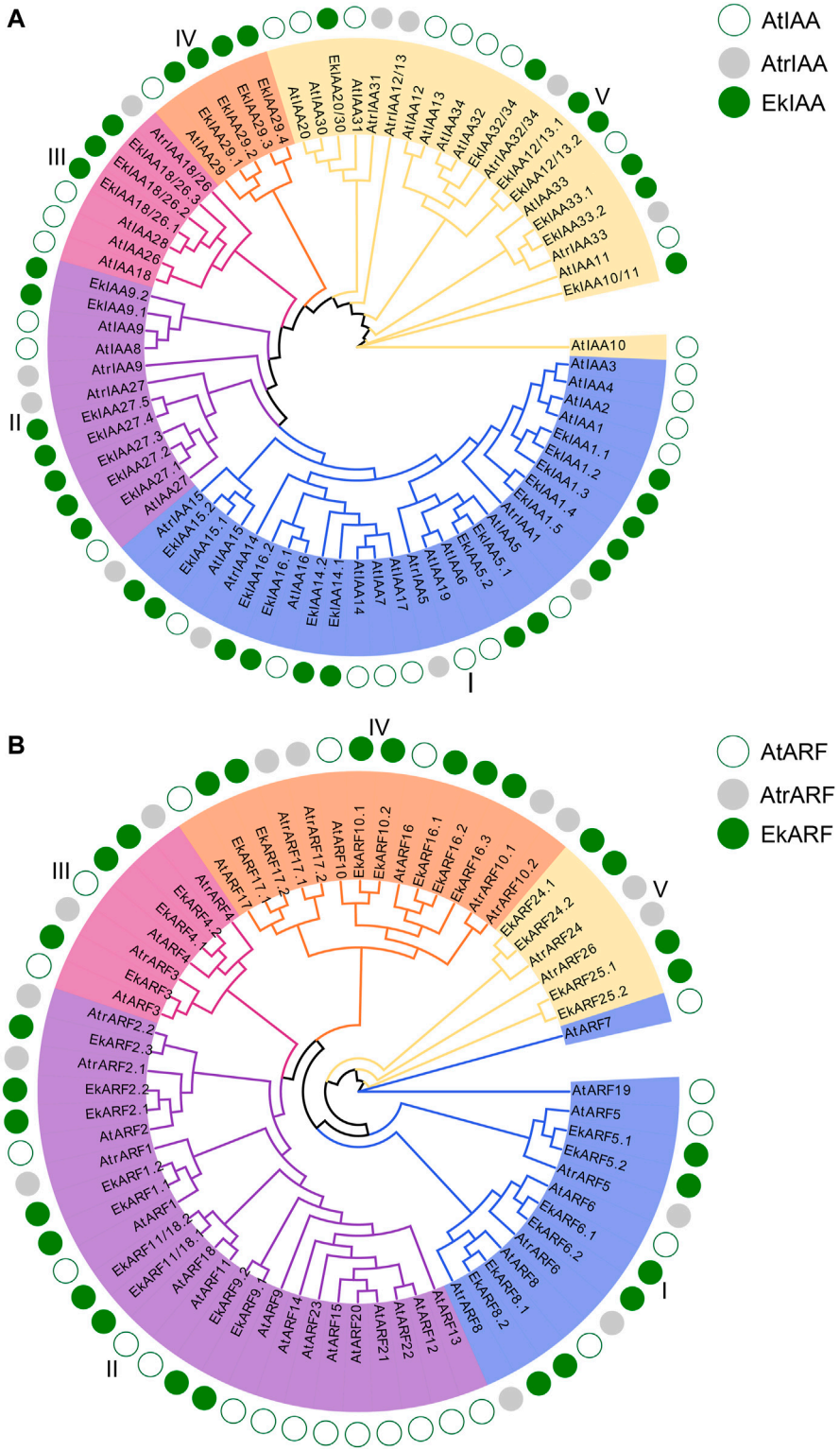


FIGURE 2 | Phylogenetic trees of EkIAAs **(A)** and EkARFs **(B)** from Arabidopsis, *Amborella trichopoda*, and *E. konishii*. Different groups have different colors. The green open circles, gray circles, and green solid circles represent Arabidopsis, *A. trichopoda*, and *E. konishii*, respectively.

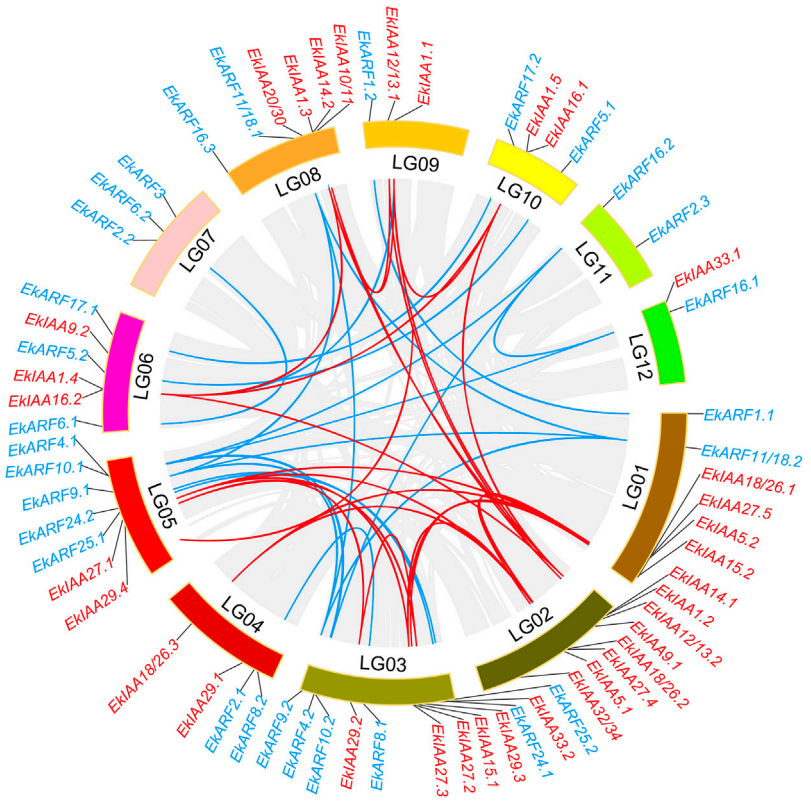


FIGURE 3 | Chromosomal locations and segmental duplicated genes for 34 *EkIAA* and 29 *EkARF* genes. Linkage group numbers (LG01–LG12) are shown at the bottom. Gray lines show all synteny blocks in the *E. konishii* genome, red lines show segmental duplication of *IAA* genes, and blue lines show segmental duplication of *ARF* genes.

MATERIALS AND METHODS

Identification of *IAA* and *ARF* Genes

Hidden Markov model (HMM) logos of Aux/IAA (PF02309) and ARF (PF06507) proteins were downloaded from the Pfam database (Finn et al., 2014) and used to scan the *E. konishii* predicted proteome (Sun et al., 2021) with the HMMER software package (Finn et al., 2011). The resulting Aux/IAA and ARF candidates were further used to generate HMM logos for EkIAAs and EkARFs using hmm-build from the HMMER suite (Finn et al., 2011), before scanning the *E. konishii* proteome again. Proteins with an *E*-value lower than 0.01 were retained, and the presence of conserved ARF or IAA domains was confirmed using the Conserved Domains Database (Marchler-Bauer et al., 2011), Pfam (Finn et al., 2014), and the Simple Modular Architecture Research Tool (Letunic and Bork, 2018). The proteins meeting all of the above criteria were used for further study. The number of amino acids, the predicted molecular weight, and the theoretical isoelectric point (pI) were determined using the ExPASy server (http://web.expasy.org/protparam/) (Gasteiger et al., 2003).

Gene Structure and Motif Analysis

TBtools (Chen et al., 2020) was employed to illustrate the exon/intron structures of all *EkIAA* and *EkARF* genes. Conserved protein motifs in their encoded proteins were predicted by the MEME program (parameters: number of maximum patterns, 10; maximum width, 50) (http://memesuite.org/tools/meme) (Bailey et al., 2006).

Multiple Sequence Alignment and Phylogenetic Analysis

The protein sequences for the 29 ARFs and 34 IAAs from Arabidopsis were obtained from published references (https://www.arabidopsis.org), and the protein sequences of 14 ARFs and 11 IAAs from *Amborella trichopoda* were identified following the same method described for *E. konishii*. Full-length protein sequences for all IAAs and ARFs identified in *E. konishii*, Arabidopsis, and *A. trichopoda* were used for phylogenetic analysis. The phylogenetic tree was built with the maximum likelihood method on the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) (Nguyen et al., 2015).

Analysis of *Cis*-acting Elements in the *EkIAA* and *EkARF* Promoters

The upstream sequences (2 kb) of *EkIAA* and *EkARF* genes were extracted via TBtools (Chen et al., 2020) and then submitted to the PlantCARE database (Lescot et al., 2002) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify *cis*-elements.

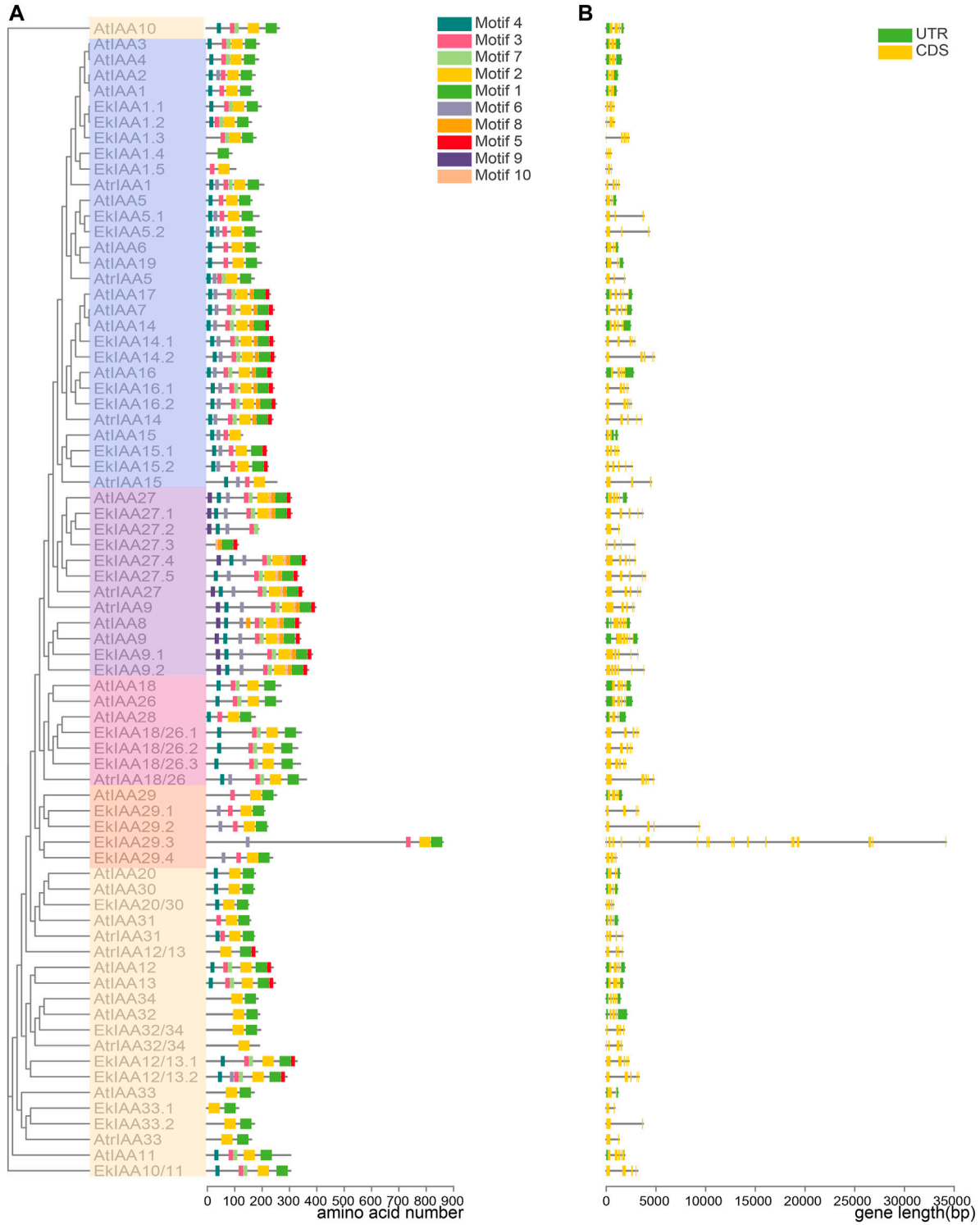


FIGURE 4 | Gene structures of *EklAA*s and conserved motifs in EklAAs. **(A)** EklAA phylogenetic tree and corresponding *EklAA* gene structures. **(B)** Conserved protein motifs within EklAAs. The phylogenetic tree was generated using the IQ-tree web server. Gene structures of *EklAA*s were predicted with TBtools. The conserved motifs were analyzed using the MEME web server.

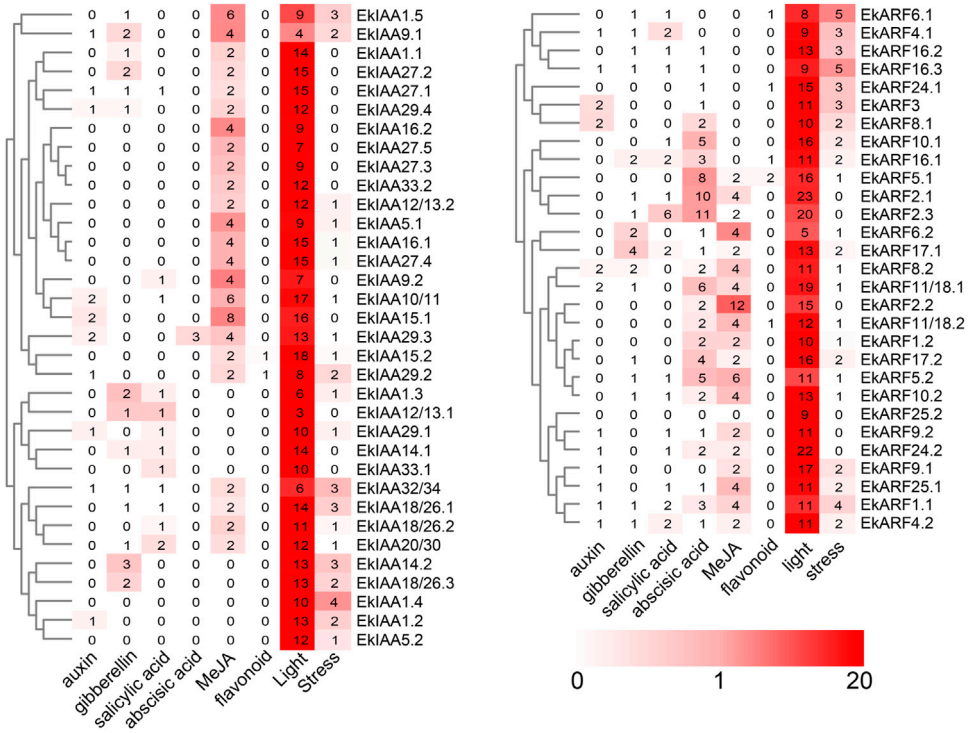


FIGURE 5 | Predicted *cis*-elements in the *EkIAA* and *EkARF* promoters. The promoter sequences (−2,000 bp) of 34 *EkIAA* and 29 *EkARF* genes were analyzed by PlantCARE. The color bar indicates the number of *cis*-elements.

EkIAA and *EkARF* Chromosomal Location and Duplication Event Analysis

The chromosomal distribution and location of all *EkIAA* and *EkARF* genes were acquired from the *E. konishii* genome annotation file. Colinear circles for *EkIAA* and *EkARF* genes were drawn with TBtools (Chen et al., 2020). Duplication events were confirmed on the basis of coverage (>70% of the entire gene body) and similarity (70%) of the two aligned sequences (Gu et al., 2002) and were considered tandem duplication pairs if they were located within 100 kb (Mehan et al., 2004). Genes located in duplicated regions with 70% similarity were identified as segmental duplications (Mehan et al., 2004). K_a/K_s values were calculated with TBtools (Chen et al., 2020).

Expression Analysis of *IAA* and *ARF* Genes by Transcriptome Deep Sequencing (RNA-Seq)

The RNA-seq data for three development stages (green, turning, and red fruit) and three tissues (red-winged pericarp, branch, and leaf) were downloaded from the National Center for Biotechnological Information (NCBI) Sequence Read Archive (SRA) under the accession numbers PRJNA548305 and PRJNA548305, respectively. The RNA-seq reads were mapped to the *E. konishii* reference genome via Salmon algorithm, and the transcripts per million reads (TPM) for *AUX/IAA* and *ARF* genes were extracted for further analysis. The heatmaps were drawn using TBtools (Chen et al., 2020).

Correlation Analysis

To study the effect of auxin on the regulation of anthocyanin and terpenoid biosynthesis in *E. konishii*, a comprehensive correlation analysis was first performed using the correlation test in R between anthocyanin contents and the expression levels of anthocyanin biosynthetic genes. Biosynthetic genes whose expression was positively correlated with anthocyanin contents were selected for further correlation analysis between their expression levels and those of *EkARF* genes. Pearson's correlation coefficients (r, p-value < 0.05) were used to define five correlation levels: no correlation ($|r| \leq 0.2$), weak correlation ($0.21 \leq |r| \leq 0.35$), moderate correlation ($0.36 \leq |r| \leq 0.67$), strong correlation ($0.68 \leq |r| \leq 0.90$), and very strong correlation ($0.91 \leq |r| \leq 1$), with $r > 0$ indicating positive correlations and $r < 0$ negative correlations (Prion and Haerling, 2014).

RT-qPCR Verification of *EkARF5.1* Gene Expression

Fruits at the green stage, turning stage, and red stage from *E. konishii* were harvested as materials for RT-qPCR. Total RNA was extracted using the RNAprep Pure kit (Tiangen, China), and then 1 μg of total RNA per sample was subjected to reverse transcription using the PrimeScript RT Reagent Kit

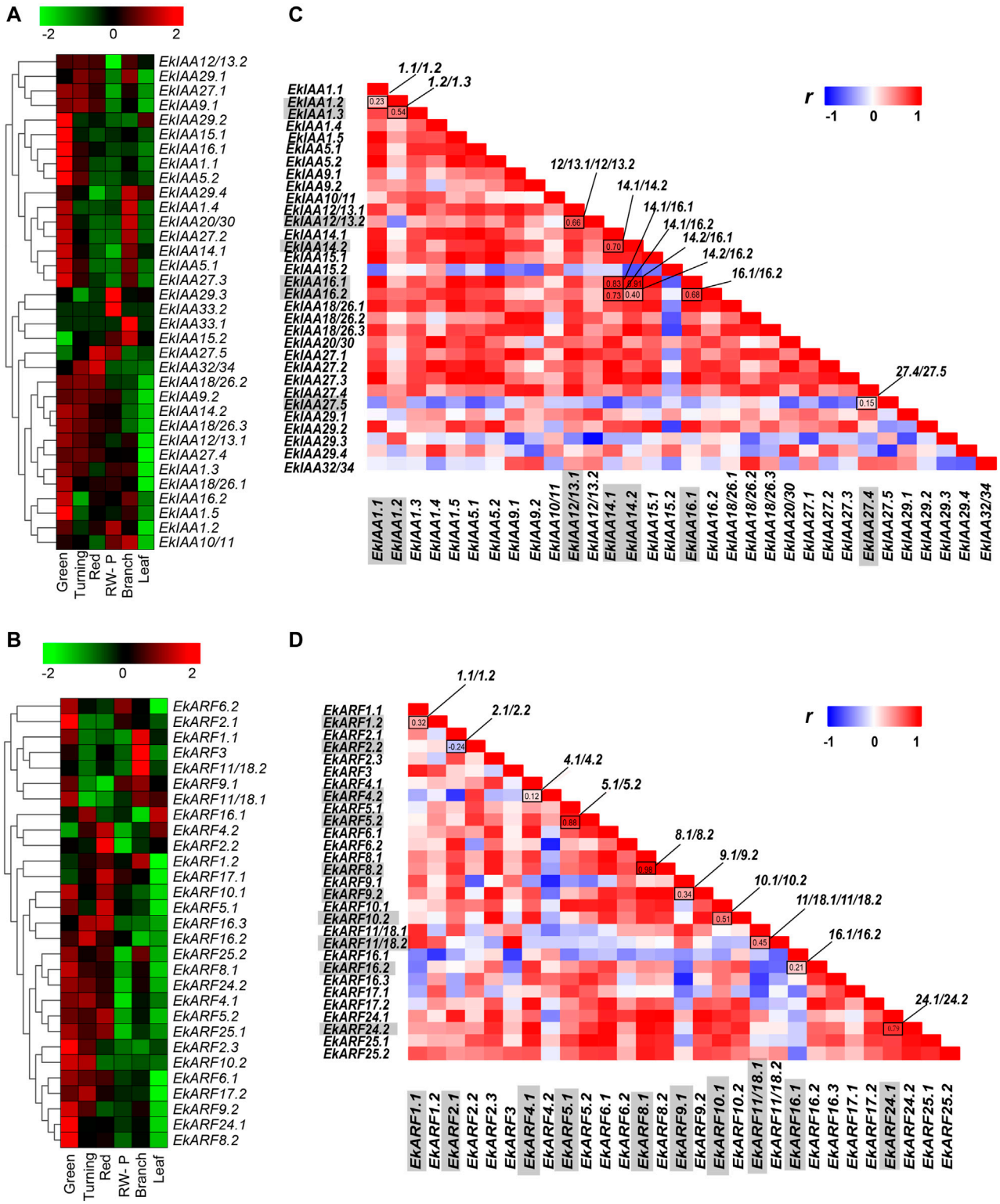


FIGURE 6 | Heatmap representation of the expression of *EkIAA* **(A)** and *EkARF* **(B)** genes during fruit developmental stages (green, turning, red fruit, and red-winged pericarp) in branches and leaves. The color bar indicates Log_2-normalized transcripts per million reads (TPM). RW-P, red-winged pericarp.

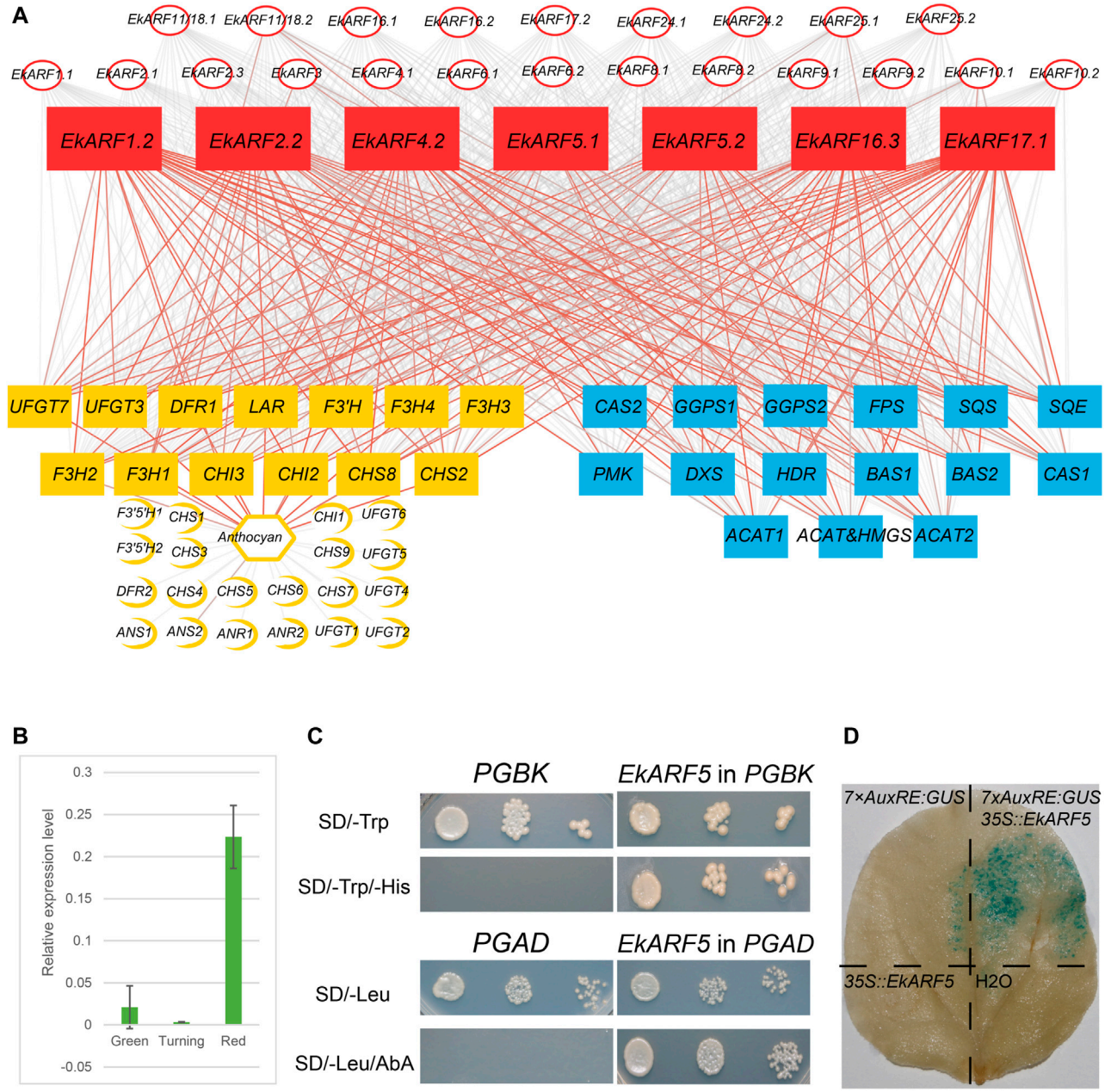


FIGURE 7 | Correlation analysis between anthocyanin contents and *EkARF* expression, and triterpenoid contents and *EkARF* expression. **(A)** Red, *EkARFs*; yellow, anthocyanins; blue, triterpenoids. The gray lines indicate weak correlation, and red lines indicate strong correlation. **(B)** RT-qPCR results. **(C)** Yeast one-hybrid analysis results. **(D)** Results from GUS staining.

(Takara, China) with gDNA Eraser (Takara, China). The specific primers for *EkARF5.1* (F: 5′-GCAACCTCCAACTCAAGAGC-3′, R: 5′-GACGCCTCACACCCACTAAT-3′) were designed by Primer 5 software and synthetized by Sangon Biotech (Shanghai, China). *UBC23* (F: 5′-AGCCACATAATCTCCGTGTAAG-3′, R: 5′-GCTGACCATGTTCGAGTAGTT-3′) was used as an internal reference (Yuan et al., 2018b). The reaction mixture consisted of 10 μl of 2× GoTaq qPCR Master Mix (Promega, United States), 0.4 μl of each gene-specific primer, 1 μl of cDNA (10× dilution), 0.4 μl of dye, and 7.8 μl of nuclease-free water. The reaction conditions were as follows: 95°C for 2 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Relative gene expression levels were calculated by the comparative ΔCt method. Three biological replications were assessed per sample.

DNA Binding and Transactivation Assay

The full-length *EkARF5.1* coding sequence was cloned into the pGBKT7 vector to generate BD-EkARF5.1, which was then

introduced into yeast strain Y2HGold. The resulting colonies were grown on synthetic defined (SD) medium lacking Trp and His for 2 days to observe the transcriptional activation activity of EkARF5.1, using empty pGBKT7 as a negative control.

To test the binding of EkARF5.1 to AuxREs, seven repeats of the AuxRE element (TGTCTC) were inserted into the multiple cloning site of the pAbAi vector to generate the 7×TGTCTC-pAbAi vector, which was integrated into the Y1HGold genome to construct the bait reporter strain. The AD-EkARF5.1 clone was generated by subcloning the full-length *EkARF5.1* coding sequence into the pGADT7 vector. AD-EkARF5.1 was then transformed into the bait reporter strain. The transformants were spotted onto SD medium lacking Leu or the same medium containing 100 ng/ml of the antibiotic aureobasidin A (AbA) and allowed to grow for 48 h to analyze binding activity.

Constructs consisting of the *β-GLUCURONIDASE* (*GUS*) reporter *7×TGTCTC:GUS* (cloned in pMDC164) and *35S:EkARF5.1* (cloned in pDMC32) were introduced into Agrobacterium (*Agrobacterium tumefaciens*) strain GV3101 for infiltration in *Nicotiana benthamiana* leaf epidermal cells. Agrobacterium harboring *7×TGTCTC:GUS* and *35S:EkARF5.1* were infiltrated into the abaxial side of *N. benthamiana* leaves with a syringe as described previously (Liu et al., 2014). *N. benthamiana* leaves were stained for GUS activity 3 days after infiltration.

Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: The E. konishii chromosome-level genome assembly and annotation data (Accession No. PRJCA005268/GWHBCHS00000000) were available from National Genomics Data Center at BioProject/GWH (https://bigd.big.ac.cn/gwh c).

RESULTS

Genome-wide Identification of *EkIAA* and *EkARF* Genes in the *E. konishii* Genome

To identify *EkIAA* and *EkARF* genes, we searched the *E. konishii* genome using HMMER v3 in two rounds (see Materials and Methods for details). We analyzed the resulting protein sequences using the Conserved Domains Database, Simple Modular Architecture Research Tool, and Pfam database, which resulted in 34 Aux/IAA and 29 ARF candidate proteins (**Figure 1**; **Supplementary Table S1**). We numbered the *E. konishii IAA* and *ARF* genes based on their homologs in Arabidopsis (Remington et al., 2004; Okushima et al., 2005); the full list is provided in **Supplementary Table S1**, along with their gene IDs, their coding sequences, genomic DNA and protein sequences, the lengths of the coding and protein sequences, and the predicted molecular weights and isoelectric points (pI) of the proteins.

We observed a broad variation in the lengths and biochemical properties of EkIAA and EkARF proteins. EkIAA proteins ranged from 91 (EkIAA1.4) to 849 (EkIAA29.3) amino acids, with predicted molecular weights from 10.5 to 97.2 kDa (**Supplementary Table S1**). The predicted pI values of EkIAA proteins varied from 4.7 (EkIAA32/34) to 9.3 (EkIAA33.2). Similarly, EkARF proteins ranged in length from 432 (EkARF17.2) to 1,110 (EkARF24.1) amino acids with predicted molecular weights from 48.2 to 122.7 kDa (**Supplementary Table S1**). The predicted pI values of EkARF proteins varied from 5.41 (EkARF5.1) to 8.42 (EkARF10.2) (**Supplementary Table S1**).

Phylogenetic Analysis of EkIAAs and EkARFs

To better understand their evolutionary history, we subjected all IAAs and ARFs identified in the model plant Arabidopsis, the early angiosperm *A. trichopoda*, and *E. konishii* to phylogenetic analysis with the MEGA-X software package (Kumar et al., 2018) and the IQ-TREE web server (Nguyen et al., 2015). In both protein families, individual members clustered into five branches, indicating that IAAs and ARFs are highly differentiated (**Figure 2A**), as previously reported in Arabidopsis (Remington et al., 2004) and poplar (Kalluri et al., 2007). EkIAA proteins were equally divided among groups Ⅰ, Ⅲ, and Ⅴ (**Figure 2A**). *EkIAA5* and *EkIAA15* from group I appeared to have undergone gene duplication, while group I had no clear *E. konishii* orthologs for Arabidopsis *IAA6* or *IAA19* (**Figure 2A**). As illustrated by the size of groups Ⅱ and Ⅳ, *EkIAA* genes have undergone gene duplication, especially *IAA27* and *IAA29* (**Figure 2A**). As with IAAs, the phylogenetic analysis of ARFs also divided the proteins into five groups (**Figure 2B**), as previously reported in Arabidopsis (Okushima et al., 2005). Group I consisted of six *EkARF* and five Arabidopsis *ARF* members, the latter having been reported to exhibit transcriptional activation activity (Guilfoyle and Hagen, 2007). Of note, *ARF5*, *ARF6*, and *ARF8* all showed gene duplication in the *E. konishii* genome, while orthologs for *ARF7* and *ARF19* appeared to be lacking. Some, but not all, group II members showed signs of gene duplication (**Figure 2B**), for example, *ARF1* and *ARF9*. Notably, IAAs and ARFs displayed the same distribution across groups in Arabidopsis and *E. konishii*, with the exception of group Ⅴ ARFs, which indicates that the *IAA* and *ARF* gene families in *E. konishii* are likely conserved. Within each group, several *EkIAA* and *EkARF* members had experienced duplication, with the exception of group Ⅴ IAA members (**Figure 2A**). In addition, we noted the absence of clear orthologs for several IAAs and ARFs in several groups (**Figure 2**), indicative of their independent evolution in *E. konishii* since the divergence from Arabidopsis.

Chromosomal Location and Gene Duplication Events of *EkIAA* and *EkARF* Genes

We determined the genomic positions of all *EkIAA* and *EkARF* genes along the linkage groups (LGs) of the *E. konishii* genome. Both groups of genes were unevenly distributed in the *E. konishii* LGs (**Figure 3**). For example, LG02 alone harbored eight *EkIAA* genes, whereas no *EkIAA* gene mapped to LG07 or LG11 (**Figure 3**). Several *EkIAA* genes clustered in close proximity

on LG01, LG2, LG03, and LG08 (**Figure 3**). The pattern for the *EkARF* genes was similar: six *EkARF* genes mapped to LG03, five to LG05, three each to LG06 and LG07, one each to LG09 and LG12, and none to LG02 (**Figure 3**).

During genome evolution, gene duplication and neofunctionalization are driven by tandem and segmental duplication (Cannon et al., 2004; Freeling, 2009). To elucidate the expansion of the *EkIAA* and *EkARF* gene families in *E. konishii*, we studied their segmental and tandem duplications. We identified 20 instances of segmental duplication events, involving 11 *EkIAA* and 20 *EkARF* genes, but no tandem duplication events in either gene family (**Figure 3**). For the 10 pairs of segmental duplicated *EkIAA* genes and the 10 pairs of segmental duplicated *EkARF* genes identified above, we calculated the ratios of nonsynonymous to synonymous substitutions (K_a/K_s) to evaluate their molecular evolutionary rates (**Supplementary Table S2**). The ratios for all duplicated pairs were less than 1 (**Supplementary Table S2**), suggesting that duplicated pairs of genes underwent purifying selection during evolution, thus raising the possibility that the biochemical characteristics of these EkIAAs and EkIAAs may not have changed very much since the initial duplication event.

Analysis of Conserved Motifs and Gene Structure

Protein motifs are critical for protein function and structure maintenance (Smith-Gill, 1991). Accordingly, we looked for conserved functional motifs in the predicted EkIAA and EkARF proteins with the MEME web server tool. We identified four domains conserved in EkIAA proteins (motifs 1–4), corresponding to IAA domains Ⅳ, Ⅲ, Ⅱ, and Ⅰ, respectively (**Figure 4**). Of the 34 EkIAA proteins, 23 (61.8%) contained all four conserved domains (domains I–IV) (**Figure 4**). Some EkIAA proteins lost one or more domains: For example, EkIAA1.3, EkIAA29.1, EkIAA29.2, EkIAA29.3, and EkIAA29.4 lack domain I; EkIAA20/30, EkIAA32/34, EkIAA33.1, and EkIAA33.2 lack domain II; EkIAA1.5 lacks domains I and IV; EkIAA27.2 lacks domains III and IV; and EkIAA1.4 and EkIAA27.3 have only domain I (**Figure 4**).

To understand the evolution of *EkIAA* genes, we examined their exon-intron structures (**Figure 4A**). Most *EkIAA* genes consisted of five exons and four introns, as in Arabidopsis, although introns in *EkIAA* genes were larger than those in their Arabidopsis counterparts (**Figure 4**). Three *EkIAA* genes (*EkIAA27.2*, *EkIAA33.1*, and *EkIAA33.2*) comprised only two exons and one intron, with another six EkIAA genes (*EkIAA1.1*, *EkIAA1.2*, *EkIAA1.4*, *EkIAA1.5*, *EkIAA5.1*, and *EkIAA5.2*) having three exons and two introns. Six *EkIAA* genes (*EkIAA20/30*, *EkIAA27.3*, *EkIAA29.1*, *EkIAA29.2*, *EkIAA29.4*, and *EkIAA32/34*) had four exons and three introns. *EkIAA9.2* had six exons and five introns, while *EkIAA29.3* had by far the most exons (20) and 19 introns (**Figure 4**). As the presence of conserved domains in the EkIAA proteins and the *EkIAA* gene structure are similar to those in their Arabidopsis orthologs (Remington et al., 2004), these results suggested that the *EkIAA* family in *E. konishii* is conserved.

We analyzed the conserved motifs and gene structure of EkARF proteins and *EkARF* genes, respectively (**Supplementary Figure S1**). EkARF proteins belonging to the same clade in the phylogenetic tree had the same functional motifs (**Supplementary Figure S1**). The DNA binding domain was represented by motifs 1, 9, and 10, while motifs 3, 6, and 8 matched the variable middle transcriptional regulatory region (MR). Motifs 7 and 5 formed part of the C-terminal dimerization domain (CTD) (**Supplementary Figure S1**). Of the 29 EkARFs, 21 (72.4%) contained all three functional domains, with only eight EkARFs (EkARF3, EkARF10.1, EkARF10.2, EkARF16.1, EkARF16.2, EkARF16.3, EkARF17.1, and EkARF17.2) lacking the CTD (**Supplementary Figure S1**). This variation in functional protein motifs may reflect mutations or deletions in the gene structure. Most *EkARF* genes contained 14 exons and 13 introns (**Supplementary Figure S1**). However, *EkARF17.1* consisted of only two exons and one intron, three *EkARF* genes (*EkARF16.1*, *EkARF16.2*, and *EkARF16.3*) were composed of three exons and two introns, *EkARF17.2* had five exons and four introns, and *EkARF3* contained 11 exons and 10 introns, which is consistent with the observed variation in protein domains (**Supplementary Figure S1**). Overall, *EkIAA* and *EkARF* genes appeared to be relatively conserved during evolution, but those derived from segmental duplication have experienced some structural divergence.

Cis-element Analysis of *EkIAA* and *EkARF* Promoters

To explore the transcriptional regulation of *EkARF* and *EkIAA* genes and predict their functions, we analyzed the *cis*-regulatory elements in their promoters. We extracted 2,000 bp of upstream sequence, which we submitted to the PlantCARE online tool (Lescot et al., 2002). We then counted the number of phytohormone-, environment-, and flavonoid-responsive elements and noted their locations (**Figure 5**). *cis*-elements in *EkIAA* and *EkARF* promoters exhibited a similar pattern (**Figure 5**). Indeed, phytohormone-responsive and environmental stress–related *cis*-elements were present in all promoters of the *EkIAA* and *EkARF* gene family in *E. konishii*.

Expression Patterns of *EkIAA* and *EkARF* Genes in *E. konishii*

We then used publicly available RNA-seq datasets to analyze the expression patterns of *EkIAA* and *EkARF* genes at four fruit developmental stages, including green, turning red fruit, and red-winged pericarp stages, as well as in branches and leaves (Yang et al., 2020); the results are summarized as heatmaps in **Figure 6**. *EkIAA* gene family members showed varying expression patterns. Most *EkIAA* genes were highly expressed in green fruits, with the exception of *EkIAA15.2*, *EkIAA33.1*, *EkIAA33.2*, and *EkIAA29.3*, of which the first two were expressed specifically in branches, whereas the latter two genes were specifically expressed in red-winged pericarp (**Figure 6A**). *EkIAA27.5* and *EkIAA32/34* were

highly expressed during the red fruit stage. *EkARF* genes were highly expressed during the fruit maturation stage (**Figure 6B**), of which *EkARF1.2*, *EkARF2.2*, *EkARF4.2*, *EkARF5.1*, *EkARF5.2*, *EkARF16.3*, and *EkARF17.1* showed high expression levels in the red fruit stage. Most *EkIAA* and *EkARF* genes were highly expressed in fruits, hinting at their potential involvement in fruit maturation, including the accumulation of associated secondary metabolites.

To assess the extent of functional diversification within these two families, we focused on duplicated gene pairs (10 *EkIAA* pairs and 10 *EkARF* pairs) and calculated the Pearson's correlation coefficients of their expression profiles. Several duplicated gene pairs did in fact exhibit differential expression across the samples tested. Six *EkIAA* gene pairs (*EkIAA12/13.1/-12/13.2*, *EkIAA14.1/-14.2*, *EkIAA14.1/-16.1*, *EkIAA14.1/-16.2*, *EkIAA14.2/-16.1*, and *EkIAA16.1/-16.2*) and three *EkARF* gene pairs (*EkARF5.1/-5.2*, *EkARF8.1/-8.2*, and *EkARF24.1/-24.2*) showed similar expression patterns within the pairs, as evidenced by their high correlation coefficients (**Figure 6B**). We also identified one *EkIAA* pair and one *EkARF* pair with distinct expression levels between duplicated copies: *EkIAA27.5* (expressed at high levels in all samples) and the duplicated copy *EkIAA27.4* (expressed at relatively low levels), with a correlation coefficient of 0.15 (**Figure 6C**); and *EkARF11/18.1* (expressed at low levels in all tissues) and *EkARF18.2* (highly expressed in branches), with a correlation coefficient of 0.45. (**Figure 6D**). These results suggested that the functions of duplicated genes may have diverged following the initial duplication event.

EkARFs May Regulate the Biosynthesis of Terpenoids and Anthocyanins

Medicinal compounds such as triterpenes, phenolic acids, and flavonoids have been isolated from *Euscaphis* fruits, leaves, and roots (Liang et al., 2018). The accumulation of anthocyanin and terpenoid secondary metabolites coincides with *E. konishii* fruit maturation (Yuan et al., 2018b; Liang et al., 2019). Generally, genes with similar expression patterns may have related roles as they belong to the same regulatory pathway or are regulated by the same upstream factors. Thus, co-expression detected from our correlation analyses may provide cues as to gene regulation or function. Because anthocyanin contents, and the key anthocyanin biosynthetic genes, were previously well characterized during *Euscaphis* fruit development (Yuan et al., 2018b), we performed a correlation analysis between anthocyanin levels and the expression estimates for anthocyanin biosynthesis genes, *EkIAA*, and *EkARF* genes. We determined that anthocyanin contents were positively and strongly correlated with the expression of five genes encoding key enzymes (*CHALCONE SYNTHASE8* [*CHS8*], *CHALCONE ISOMERASE2* [*CHI2*], *FLAVANONE 3-HYDROXYLASE1* [*F3H1*], *F3H2*, and *F3H3*) and very strongly correlated with that of another eight genes (*CHS2*, *CHI3*, *F3H4*, *FLAVONOID 3′-HYDROXYLASE* [*F3′H*], *DIHYDROFLAVONOL 4-REDUCTASE1* [*DFR1*], *LEUCOANTHOCYANIDIN REDUCTASE* [*LAR*], *FLAVONOL 3-O-GLUCOSYLTRANSFERASE3* [*UFGT3*], and *UFGT7*) (**Figure 7A**). Because ARF proteins regulate the expression of their target genes by binding to their cognate *cis*-elements in promoters, we looked for *EkARF* genes co-expressed with the anthocyanin biosynthetic genes listed above (**Figure 7A**), leading to the identification of seven such genes (*EkARF1.2*, *EkARF2.2*, *EkARF4.2*, *EkARF5.1*, *EkARF5.2*, *EkARF16.3*, and *EkARF17.1*).

Triterpenoids accumulate to high levels in *Euscaphis* fruits and are important raw materials for natural products, food additives, and chemical products. Genes encoding the enzymes involved in the biosynthesis of triterpenoids have been described in the *Euscaphis* genome (Huang et al., 2019; Liang et al., 2019), prompting us to test for correlations between their expression patterns and those of *ARFs* genes (**Figure 7A**). This analysis highlighted seven *EkARF* genes (*EkARF1.2*, *EkARF2.2*, *EkARF4.2*, *EkARF5.1*, *EkARF5.2*, *EkARF16.3*, and *EkARF17.1*) whose expression was positively and strongly correlated with genes involved in triterpenoid accumulation (**Figure 7**). These results suggested that *ARF* genes contribute to secondary metabolite biosynthesis in *Euscaphis*.

ARF proteins are transcription factors that can bind to AuxREs (TGTCTC) to regulate the expression of their target genes. We noticed at least one AuxRE either upstream, downstream, or within intronic regions of anthocyanin and triterpenoid biosynthetic genes (**Supplementary Table S3**). To assess the role of ARFs in anthocyanin and triterpenoid biosynthesis, we selected *EkARF5.1* for further characterization. RT-qPCR showed that *EkARF5.1* is highly expressed during the red fruit stage, which is consistent with the RNA-seq data (**Figure 7B**). We fused EkARF5.1 to the GAL4 activation domain (AD) (AD-EkARF5.1) and introduced the resulting construct into yeast strain Y1HGold carrying a reporter consisting of seven copies of the AuxRE sequence driving the expression of *Aureobasidin Resistance 1*, conferring resistance to the antibiotic aureobasidin A (AbA). Whereas yeast colonies carrying AD or AD-EkARF5.1 grew on synthetic defined medium (**Figure 7C**), only yeast cells harboring the *AD-EkARF5.1* construct survived growth on AbA-containing medium (**Figure 7C**), supporting the notion that EkARF5.1 binds to the AuxRE. We then tested the transactivation activity of EkARF5.1 by fusing full-length EkARF5.1 to the GAL4 DNA binding domain (BD) to generate the BD-EkARF5.1 fusion protein; the resulting encoding construct was introduced into yeast strain Y2HGold, which harbors the *His3* gene driven by a GAL4-responsive promoter (**Figure 7C**). Only yeast cells carrying the *BD-EkARF5.1* construct survived on synthetic medium lacking histidine, unlike yeast cells carrying the empty GAL4 DB vector (**Figure 7C**). These results indicated that EkARF5.1 has transactivation activity in yeast cells. Finally, we tested EkARF5.1 in *N. benthamiana* leaf epidermal cells by co-infiltrating a construct overexpressing *EkARF5.1* and a *β-GLUCURONIDASE* (*GUS*) reporter construct whose expression is driven by seven copies of the AuxRE. We detected GUS activity in plant cells only when *EkARF5* was co-expressed (**Figure 7D**). These results demonstrate that EkARF5.1 may play a role during anthocyanin and triterpenoid biosynthesis.

DISCUSSION

Plant secondary metabolites consist of various bioactive compounds with applications in medicine and industry. The biosynthesis of plant secondary metabolites is regulated not

only by plant growth and development signaling, but also by environmental stress cues. In China, *E. konishii* is widely planted as a medicinal and ornamental plant, but the regulatory mechanism of secondary metabolite biosynthesis is poorly understood, which limits genetic improvement and development of agronomic management techniques. In this study, we identified 34 *Aux/IAA* genes and 29 *ARF* coding sequences that map to some, but not all, linkage groups that constitute the *E. konishii* genome. Co-expression analysis suggested that seven *EkARF* genes may regulate anthocyanin and triterpenoid biosynthesis in *E. konishii*. Our data improve the understanding of the *EkIAA* and *EkARF* gene families and may provide valuable information on their biological functions in the context of secondary metabolite biosynthesis.

Auxin is an essential plant hormone, and Aux/IAAs and ARFs are key components of the signaling transduction process. Most of the current knowledge on Aux/IAA and ARF function, gene expression, and regulation has been obtained from studies in annual herbaceous plants such as Arabidopsis, rice, and tomato (Luo et al., 2018), while much less is known about IAAs and ARFs in longer-lived species, such as the evergreen shrub *E. konishii*. In this study, we identified 34 *Aux/IAA* and 29 *ARF* family members in the *E. konishii* genome (**Figure 1**; **Supplementary Table S1**), which was comparable to the numbers in other species, such as Arabidopsis (29 *Aux/IAAs* and 23 *ARFs*) (Remington et al., 2004), poplar (35 *Aux/IAAs* and 39 *ARFs*) (Kalluri et al., 2007), maize (31 *Aux/IAAs* and 31 *ARFs*) (Wang et al., 2010), and rice (31 *Aux/IAAs* and 25 *ARFs*) (Jain et al., 2006). *EkIAA* and *EkARF* genes clustered into five groups, as previously reported in Arabidopsis (Remington et al., 2004). Most *EkIAA* and *EkARF* genes within the same phylogenetic group shared similar exon-intron structures and the same arrangement of functional motifs in their encoded proteins (**Figure 4**), likely reflecting the gene duplication events that have shaped the expansion of the *EkIAA* and *EkARF* gene families in the *E. konishii* genome. K_a/K_s values of homologous genes further showed that duplicated genes underwent purifying selection (**Supplementary Table S1**). These results indicated that the two gene families are evolutionarily conserved with those from other plant species and may thus exhibit the same function and biochemical characteristics in *E. konishii*. However, we also identified six EkIAAs (EkIAA1.3, EkIAA29.1, EkIAA29.2, EkIAA29.3, and EkIAA29.4) that lack domain I (**Figure 4**), which are not expected to repress their downstream targets because domain I can repress the expression of target genes when in close proximity to the promoter (Hagen and Guilfoyle, 2002). Therefore, the *E. konishii* genome encodes conserved *EkIAA* and *EkARF* gene family members, although some members exhibit domain loss, possibly having arisen from unknown segmental duplication events, which will increase the complexity of auxin regulation.

The phytohormone auxin plays critical roles during plant growth. During fruit development, auxin also induces fruit set and growth, whereas it represses fruit maturation and ripening (Pattison et al., 2014). During tomato fruit maturation, the *SlARF2* expression level increases in response to stimulation by ethylene, suggesting that auxin may repress fruit ripening (Pattison et al., 2014). Although gene structure and their encoded protein motifs were conserved in *EkIAA* and *EkARF* gene families, a subset of *EkIAA* and *EkARF* genes showed high expression during fruit maturation and ripening (**Figure 6**). Seven *EkARF*s were highly expressed during fruit maturation, of which *EkARF1.2*, *EkARF2.2*, *EkARF4.2*, *EkARF16.3*, and *EkARF17.1* encode proteins containing a proline/serine/threonine-rich domain that acts as a transcriptional repressor (**Supplementary Figure S1**) (Tiwari et al., 2003). These results suggested that *E. konishii* fruit maturation and ripening may be similar to tomato. Interestingly, the two ARF5 homologs EkARF5.1 and EkARF5.2 contained a glutamate-rich domain that functions as a transcriptional activator domain; their encoding genes were highly expressed during *E. konishii* fruit maturation and ripening (**Supplementary Figure S1**) (Tiwari et al., 2003; Guilfoyle and Hagen, 2007), indicating that auxin signaling may play distinct roles during *E. konishii* fruit maturation and ripening processes. In Arabidopsis, ARF5 affects meristem development (Dastidar et al., 2019), while we showed here that *EkARF5* is highly expressed during fruit maturation and ripening, likely reflecting changes in the promoter region associated with *E. konishii* and suggesting that the EkARF family underwent subfunctionalization during its evolutionary history. The high accumulation of secondary metabolites is a main feature of *E. konishii* fruit maturation and ripening (Yuan et al., 2018b). Co-expression analysis further revealed that *EkARF5.1* and *EkARF5.2* expression is positively and strongly correlated with that of anthocyanin and triterpenoid biosynthetic genes (**Figure 7A**). We confirmed that *EkARF5.1* is highly expressed during fruit maturation and ripening stages and that EkARF5.1 can directly bind to AuxREs located within the promoter regions of anthocyanin and triterpenoid biosynthetic genes to activate their transcription in yeast and plant cells (**Figures 7C,D**). These results strongly suggest that EkARF5.1 and EkARF5.2 may be positive regulators of secondary metabolite biosynthesis, although the exact mechanisms by which they regulate anthocyanin biosynthesis require further study.

E. konishii fruit maturation and ripening involve pericarp splitting, pericarp overturn, and the accumulation of secondary metabolites (Yuan et al., 2018b; Liang et al., 2019; Huang et al., 2019; Sun et al., 2021). Pericarp overturn, the process in which the pericarp morphological changes after pericarp splitting (**Supplementary Figure S2**), is the main difference between *E. konishii* fruits and those of grape (*Vitis vinifera*) and tomato (Yuan et al., 2018b; Liang et al., 2019; Sun et al., 2021), which and may contribute to its survival and enable expansion to new environments (Sun et al., 2021). Pericarp overturn may be the result of uneven cell growth between epicarp and endocarp (Sun et al., 2021), which is associated with cell growth or differentiation (Ding et al., 2011). Auxin asymmetric distribution mediated by development and environmental cues results in uneven cell growth, thus regulating plant growth and response to environmental changes (Ding et al., 2011). Therefore, the highly expressed EkARF5s are probably involved in pericarp overturn in the last stage of *E. konishii* fruit maturation (**Figure 6B**; **Figure 7B**). It has been reported that auxin or its signaling is involved in biosynthesis of secondary metabolites such as anthocyanin, flavonols, and glucosinolates (Lewis et al., 2011; Wang et al.,

2018; Wang et al., 2020). Given that pericarp split, pericarp overturn, and secondary metabolites are coupled (Yuan et al., 2018b), they may be regulated by similar or identical molecular mechanisms. In our study, we identified AuxREs in the promoters of anthocyanin and triterpenoid biosynthetic genes (**Supplementary Table S3**), to which EkARF5 can bind in yeast and plant cells, suggesting that EkARF5-mediated auxin signaling may regulate multiple signaling pathways in *E. konishii* fruit maturation.

CONCLUSION

We comprehensively analyzed the *Aux/IAA* and *ARF* gene families in *E. konishii*, which are evolutionarily well conserved. Expression and co-expression analyses showed that EkARF5 may play critical roles during the regulation of secondary metabolite biosynthesis. This study provides the basis for uncovering the regulatory mechanisms necessary to boost the production of industrial products and breed new *E. konishii* varieties with high economic output.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The E. konishii chromosome-level genome assembly and annotation data (Accession No. GWHBCHS00000000) were available from National Genomics Data Center (https://ngdc.cncb.ac.cn/gwh/Assembly/reviewer/DsVzSvoOmVHEsqvwRYFCzmaHLYVsIsbpuBGJNDyDMGIWICZvIzpChREHnNqERbWc).

AUTHOR CONTRIBUTIONS

BL conceived the project. JZ carried out all analyses together with BL. JZ wrote the manuscript draft, and BL and X-XZ revised the manuscript. LL, QY, BH, and QW helped with the experiments. BL and S-QZ acquired the funding. All authors have read and approved the manuscript.

FUNDING

This work was supported by startup funds for scientific research of Yancheng Teachers University (72672166005C), Funds from Fujian Agriculture and Forestry University (KFA17295A, CXZX2019044G, 11899170128), and by the project in Fujian Province (2020N5004 and Min(2018)TG14).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.737293/full#supplementary-material

REFERENCES

Aloni, R., Aloni, E., Langhans, M., and Ullrich, C. I. (2006). Role of Cytokinin and Auxin in Shaping Root Architecture: Regulating Vascular Differentiation, Lateral Root Initiation, Root Apical Dominance and Root Gravitropism. *Ann. Bot.* 97 (5), 883–893. doi:10.1093/aob/mcl027

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Res.* 34 (Web Server issue), W369–W373. doi:10.1093/nar/gkl198

Bing, H., Liua, H., Han, X., Cui, P., and Xub, L-A. (2021). Multi-Omics Analysis of Ginkgo Biloba Preliminarily Reveals the Co-Regulatory Mechanism between Stilbenes and Flavonoids. *Ind. Crops Prod.* 167, 113434. doi:10.1016/j.indcrop.2021.113434

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The Roles of Segmental and Tandem Gene Duplication in the Evolution of Large Gene Families in *Arabidopsis T. BMC Plant Biol.* 4, 10. doi:10.1186/1471-2229-4-10

Chaabouni, S., Jones, B., Delalande, C., Wang, H., Li, Z., Mila, I., et al. (2009). Sl-IAA3, a Tomato Aux/IAA at the Crossroads of Auxin and Ethylene Signalling Involved in Differential Growth. *J. Exp. Bot.* 60 (4), 1349–1362. doi:10.1093/jxb/erp009

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 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009

Dastidar, M. G., Scarpa, A., Mägele, I., Ruiz-Duarte, P., von Born, P., Bald, L., et al. (2019). ARF5/MONOPTEROS Directly Regulates miR390 Expression in the *Arabidopsis T* Primary Root Meristem. *Plant Direct* 3 (2), e00116. doi:10.1002/pld3.116

Ding, Z., Galván-Ampudia, C. S., Demarsy, E., Łangowski, Ł., Kleine-Vehn, J., Fan, Y., et al. (2011). Light-Mediated Polarization of the PIN3 Auxin Transporter for the Phototropic Response in Arabidopsis. *Nat. Cel Biol* 13 (4), 447–452. doi:10.1038/ncb2208

Ellis, C. M., Nagpal, P., Young, J. C., Hagen, G., Guilfoyle, T. J., and Reed, J. W. (2005). AUXIN RESPONSE FACTOR1andAUXIN RESPONSE FACTOR2regulate Senescence and floral Organ Abscission inArabidopsis Thaliana. *Development* 132 (20), 4563–4574. doi:10.1242/dev.02012

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the Protein Families Database. *Nucl. Acids Res.* 42 (Database issue), D222–D230. doi:10.1093/nar/gkt1223

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* 39 (Web Server issue), W29–W37. doi:10.1093/nar/gkr367

Freeling, M. (2009). Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi:10.1146/annurev.arplant.043008.092122

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: The Proteomics Server for In-Depth Protein Knowledge and Analysis. *Nucleic Acids Res.* 31 (13), 3784–3788. doi:10.1093/nar/gkg563

Goetz, M., Vivian-Smith, A., Johnson, S. D., and Koltunow, A. M. (2006). AUXIN RESPONSE FACTOR8Is a Negative Regulator of Fruit Initiation inArabidopsis. *Plant Cell* 18 (8), 1873–1886. doi:10.1105/tpc.105.037192

Goh, T., Kasahara, H., Mimura, T., Kamiya, Y., and Fukaki, H. (2012). Multiple AUX/IAA-ARF Modules Regulate Lateral Root Formation: The Role of Arabidopsis SHY2/IAA3-Mediated Auxin Signalling. *Phil. Trans. R. Soc. B* 367 (1595), 1461–1468. doi:10.1098/rstb.2011.0232

Gray, W. M., Kepinski, S., Rouse, D., Leyser, O., and Estelle, M. (2001). Auxin Regulates SCFTIR1-Dependent Degradation of AUX/IAA Proteins. *Nature* 414 (6861), 271–276. doi:10.1038/35104500

Gu, Z., Cavalcanti, A., Chen, F.-C., Bouman, P., and Li, W.-H. (2002). Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast. *Mol. Biol. Evol.* 19 (3), 256–262. doi:10.1093/oxfordjournals.molbev.a004079

Guilfoyle, T. J., and Hagen, G. (2007). Auxin Response Factors. *Curr. Opin. Plant Biol.* 10 (5), 453–460. doi:10.1016/j.pbi.2007.08.014

Guilfoyle, T. J. (2015). The PB1 Domain in Auxin Response Factor and Aux/IAA Proteins: A Versatile Protein Interaction Module in the Auxin Response. *Plant Cell* 27 (1), 33–43. doi:10.1105/tpc.114.132753

Hagen, G. (2015). Auxin Signal Transduction. *Essays Biochem.* 58, 1–12. doi:10.1042/bse0580001

Hagen, G., and Guilfoyle, T. (2002). Auxin-Responsive Gene Expression: Genes, Promoters and Regulatory Factors. *Plant Mol. Biol.* 49 (3-4), 373–385. doi:10.1007/978-94-010-0377-3_9

Hardtke, C. S., and Berleth, T. (1998). The Arabidopsis Gene MONOPTEROS Encodes a Transcription Factor Mediating Embryo Axis Formation and Vascular Development. *Embo j* 17 (5), 1405–1411. doi:10.1093/emboj/17.5.1405

Hu, J., Israeli, A., Ori, N., and Sun, T.-P. (2018). The Interaction between DELLA and ARF/IAA Mediates Crosstalk between Gibberellin and Auxin Signaling to Control Fruit Initiation in Tomato. *Plant Cell* 30 (8), 1710–1728. doi:10.1105/tpc.18.00363

Huang, B., Rong, H., Ye, Y., Ni, Z., Xu, M., Zhang, W., et al. (2020). Transcriptomic Analysis of Flower Color Variation in the Ornamental Crabapple (Malus spp.) Half-Sib Family through Illumina and PacBio Sequel Sequencing. *Plant Physiol. Biochem.* 149, 27–35. doi:10.1016/j.plaphy.2020.01.033

Huang, W., Ding, H., Chen, L-Y., Ni, L., Ruan, Y-F., Zou, X-X., et al. (2019). Protective Effect of the Total Triterpenes of Euscaphis Konishii Hayata Pericarp on Bacillus Calmette-Guérin Plus Lipopolysaccharide-Induced Liver Injury. *Evid. Based Complement. Alternat Med.* 2019, 1806021. doi:10.1155/2019/1806021

Huang, W., Zheng, Y., Feng, H., Ni, L., Ruan, Y.-F., Zou, X.-X., et al. (2020). Total Phenolic Extract of Euscaphis Konishii Hayata Pericarp Attenuates Carbon Tetrachloride (CCl4)-Induced Liver Fibrosis in Mice. *Biomed. Pharmacother.* 125, 109932. doi:10.1016/j.biopha.2020.109932

Jain, M., Kaur, N., Garg, R., Thakur, J. K., Tyagi, A. K., and Khurana, J. P. (2006). Structure and Expression Analysis of Early Auxin-Responsive Aux/IAA Gene Family in Rice (Oryza Sativa). *Funct. Integr. Genomics* 6 (1), 47–59. doi:10.1007/s10142-005-0005-0

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.* 7, 59. doi:10.1186/1471-2229-7-59

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 (6), 1547–1549. doi:10.1093/molbev/msy096

Lee, M. K., Lee, K. Y., Jeon, H. Y., Sung, S. H., and Kim, Y. C. (2009). Antifibrotic Activity of Triterpenoids from the Aerial Parts ofEuscaphis Japonicaon Hepatic Stellate Cells. *J. Enzyme Inhib. Med. Chem.* 24 (6), 1276–1279. doi:10.3109/14756360902829709

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 (1), 325–327. doi:10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2018). 20 Years of the SMART Protein Domain Annotation Resource. *Nucleic Acids Res.* 46 (D1), D493–d496. doi:10.1093/nar/gkx922

Lewis, D. R., Ramirez, M. V., Miller, N. D., Vallabhaneni, P., Ray, W. K., Helm, R. F., et al. (2011). Auxin and Ethylene Induce Flavonol Accumulation through Distinct Transcriptional Networks. *Plant Physiol.* 156 (1), 144–164. doi:10.1104/pp.111.172502

Liang, W., Ni, L., Carballar-Lejarazú, R., Zou, X., Sun, W., Wu, L., et al. (2019). Comparative Transcriptome Among Euscaphis Konishii Hayata Tissues and Analysis of Genes Involved in Flavonoid Biosynthesis and Accumulation. *BMC Genomics* 20 (1), 24. doi:10.1186/s12864-018-5354-x

Liang, W., Zou, X., Carballar-Lejarazú, R., Wu, L., Sun, W., Yuan, X., et al. (2018). Selection and Evaluation of Reference Genes for qRT-PCR Analysis in Euscaphis Konishii Hayata Based on Transcriptome Data. *Plant Methods* 14, 42. doi:10.1186/s13007-018-0311-x

Liscum, E., and Reed, J. W. (2002). Genetics of Aux/IAA and ARF Action in Plant Growth and Development. *Plant Mol. Biol.* 49 (3-4), 387–400. doi:10.1007/978-94-010-0377-3_10

Liu, B., Wang, L., Zhang, J., Li, J., Zheng, H., Chen, J., et al. (2014). WUSCHEL-Related Homeobox Genes in Populus Tomentosa: Diversified Expression Patterns and a Functional Similarity in Adventitious Root Formation. *BMC Genomics* 15, 296. doi:10.1186/1471-2164-15-296

Luo, J., Zhou, J. J., and Zhang, J. Z. (2018). Aux/IAA Gene Family in Plants: Molecular Structure, Regulation, and Function. *Int. J. Mol. Sci.* 19 (1), 259. doi:10.3390/ijms19010259

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: A Conserved Domain Database for the Functional Annotation of Proteins. *Nucleic Acids Res.* 39 (Database issue), D225–D229. doi:10.1093/nar/gkq1189

Mehan, M. R., Freimer, N. B., and Ophoff, R. A. (2004). A Genome-Wide Survey of Segmental Duplications that Mediate Common Human Genetic Variation of Chromosomal Architecture. *Hum. Genomics* 1 (5), 335–344. doi:10.1186/1479-7364-1-5-335

Mitreiter, S., and Gigolashvili, T. (2021). Regulation of Glucosinolate Biosynthesis. *J. Exp. Bot.* 72 (1), 70–91. doi:10.1093/jxb/eraa479

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274. doi:10.1093/molbev/msu300

Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., et al. (2005). Functional Genomic Analysis of theAUXIN RESPONSE FACTORGene Family Members inArabidopsis Thaliana: Unique and Overlapping Functions ofARF7andARF19. *Plant Cell* 17 (2), 444–463. doi:10.1105/tpc.104.028316

Pattison, R. J., Csukasi, F., and Catalá, C. (2014). Mechanisms Regulating Auxin Action during Fruit Development. *Physiol. Plantarum* 151 (1), 62–72. doi:10.1111/ppl.12142

Prion, S., and Haerling, K. A. (2014). Making Sense of Methods and Measurement: Pearson Product-Moment Correlation Coefficient. *Clin. Simulation Nurs.* 10 (11), 587–588. doi:10.1016/j.ecns.2014.07.010

Remington, D. L., Vision, T. J., Guilfoyle, T. J., and Reed, J. W. (2004). Contrasting Modes of Diversification in the Aux/IAA and ARF Gene Families. *Plant Physiol.* 135 (3), 1738–1752. doi:10.1104/pp.104.039669

Roosjen, M., Paque, S., and Weijers, D. (2018). Auxin Response Factors: Output Control in Auxin Biology. *J. Exp. Bot.* 69 (2), 179–188. doi:10.1093/jxb/erx237

Singh, V. K., and Jain, M. (2015). Genome-Wide Survey and Comprehensive Expression Profiling of Aux/IAA Gene Family in Chickpea and Soybean. *Front. Plant Sci.* 6, 918. doi:10.3389/fpls.2015.00918

Smith-Gill, S. J. (1991). Protein-Protein Interactions: Structural Motifs and Molecular Recognition. *Curr. Opin. Biotechnol.* 2 (4), 568–575. doi:10.1016/0958-1669(91)90082-g

Sun, W. H., Li, Z., Xiang, S., Ni, L., Zhang, D., Chen, D-Q., et al. (2021). The Euscaphis Japonica Genome and the Evolution of Malvids. *Plant J.* doi:10.1111/tpj.15518

Sun, W., Yuan, X., Liu, Z.-J., Lan, S., Tsai, W.-C., and Zou, S.-Q. (2019). Multivariate Analysis Reveals Phenotypic Diversity of Euscaphis Japonica Population. *PLoS One* 14 (7), e0219046. doi:10.1371/journal.pone.0219046

Tiwari, S. B., Hagen, G., and Guilfoyle, T. J. (2004). Aux/IAA Proteins Contain a Potent Transcriptional Repression Domain. *Plant Cell* 16 (2), 533–543. doi:10.1105/tpc.017384

Tiwari, S. B., Hagen, G., and Guilfoyle, T. (2003). The Roles of Auxin Response Factor Domains in Auxin-Responsive Transcription. *Plant Cell* 15 (2), 533–543. doi:10.1105/tpc.008417

Tiwari, S. B., Wang, X. J., Hagen, G., and Guilfoyle, T. J. (2001). AUX/IAA Proteins Are Active Repressors, and Their Stability and Activity Are Modulated by Auxin. *Plant Cell* 13 (12), 2809–2822. doi:10.1105/tpc.13.12.2809

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999). Activation and Repression of Transcription by Auxin-Response Factors. *Proc. Natl. Acad. Sci.* 96 (10), 5844–5849. doi:10.1073/pnas.96.10.5844

Van Ha, C., Le, D. T., Nishiyama, R., Watanabe, Y., Sulieman, S., Tran, U. T., et al. (2013). The Auxin Response Factor Transcription Factor Family in Soybean: Genome-Wide Identification and Expression Analyses during Development and Water Stress. *DNA Res.* 20 (5), 511–524. doi:10.1093/dnares/dst027

Wang, C.-K., Han, P.-L., Zhao, Y.-W., Ji, X.-L., Yu, J.-Q., You, C.-X., et al. (2020). Auxin Regulates Anthocyanin Biosynthesis through the Auxin Repressor Protein MdIAA26. *Biochem. Biophysical Res. Commun.* 533 (4), 717–722. doi:10.1016/j.bbrc.2020.09.065

Wang, D., Pei, K., Fu, Y., Sun, Z., Li, S., Liu, H., et al. (2007). Genome-Wide Analysis of the Auxin Response Factors (ARF) Gene Family in rice (Oryza Sativa). *Gene* 394 (1-2), 13–24. doi:10.1016/j.gene.2007.01.006

Wang, Y.-C., Wang, N., Xu, H.-F., Jiang, S.-H., Fang, H.-C., Su, M.-Y., et al. (2018). Auxin Regulates Anthocyanin Biosynthesis through the Aux/IAA-ARF Signaling Pathway in Apple. *Hortic. Res.* 5, 59. doi:10.1038/s41438-018-0068-4

Wang, Y., Deng, D., Bian, Y., Lv, Y., and Xie, Q. (2010). Genome-Wide Analysis of Primary Auxin-Responsive Aux/IAA Gene Family in maize (Zea mays. L). *Mol. Biol. Rep.* 37 (8), 3991–4001. doi:10.1007/s11033-010-0058-6

Waseem, M., Ahmad, F., Habib, S., and Li, Z. (2018). Genome-Wide Identification of the Auxin/indole-3-Acetic Acid (Aux/IAA) Gene Family in Pepper, its Characterisation, and Comprehensive Expression Profiling under Environmental and Phytohormones Stress. *Sci. Rep.* 8 (1), 12008. doi:10.1038/s41598-018-30468-9

Wilmoth, J. C., Wang, S., Tiwari, S. B., Joshi, A. D., Hagen, G., Guilfoyle, T. J., et al. (2005). NPH4/ARF7 and ARF19 Promote Leaf Expansion and Auxin-Induced Lateral Root Formation. *Plant J.* 43 (1), 118–130. doi:10.1111/j.1365-313x.2005.02432.x

Winkler, M., Niemeyer, M., Hellmuth, A., Janitza, P., Christ, G., Samodelov, S. L., et al. (2017). Variation in Auxin Sensing Guides AUX/IAA Transcriptional Repressor Ubiquitylation and Destruction. *Nat. Commun.* 8, 15706. doi:10.1038/ncomms15706

Xing, H., Pudake, R. N., Guo, G., Xing, G., Hu, Z., Zhang, Y., et al. (2011). Genome-Wide Identification and Expression Profiling of Auxin Response Factor (ARF) Gene Family in maize. *BMC Genomics* 12, 178. doi:10.1186/1471-2164-12-178

Yang, C. L., Yuan, X. Y., Zhang, J., Sun, W. H., Liu, Z. J., and Zou, S. Q. (2020). Comprehensive Transcriptome Analysis of Reference Genes for Fruit Development of Euscaphis Konishii. *PeerJ* 8, e8474. doi:10.7717/peerj.8474

Yu, C., Zhan, Y., Feng, X., Huang, Z. A., and Sun, C. (2017). Identification and Expression Profiling of the Auxin Response Factors in Capsicum Annuum L. Under Abiotic Stress and Hormone Treatments. *Int. J. Mol. Sci.* 18 (12), 2719. doi:10.3390/ijms18122719

Yuan, X., Sun, W., Zou, X., Liu, B., Huang, W., Chen, Z., et al. (2018). Sequencing of Euscaphis Konishii Endocarp Transcriptome Points to Molecular Mechanisms of Endocarp Coloration. *Int. J. Mol. Sci.* 19 (10), 3209. doi:10.3390/ijms19103209

Yuan, Y., Mei, L., Wu, M., Wei, W., Shan, W., Gong, Z., et al. (2018). SlARF10, an Auxin Response Factor, Is Involved in Chlorophyll and Sugar Accumulation during Tomato Fruit Development. *J. Exp. Bot.* 69 (22), 5507–5518. doi:10.1093/jxb/ery328

Yuan, Y., Xu, X., Gong, Z., Tang, Y., Wu, M., Yan, F., et al. (2019). Auxin Response Factor 6A Regulates Photosynthesis, Sugar Accumulation, and Fruit Development in Tomato. *Hortic. Res.* 6, 85. doi:10.1038/s41438-019-0167-x

Zhang, L.-J., Cheng, J.-J., Liao, C.-C., Cheng, H.-L., Huang, H.-T., Kuo, L.-M., et al. (2012). Triterpene Acids from Euscaphis Japonica and Assessment of Their Cytotoxic and Anti-NO Activities. *Planta Med.* 78 (14), 1584–1590. doi:10.1055/s-0032-1315040

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 January 2022
doi: 10.3389/fgene.2021.818880

Genome-Wide Diversity of MADS-Box Genes in Bread Wheat is Associated with its Rapid Global Adaptability

Qasim Raza[1], Awais Riaz[1], Rana Muhammad Atif[2,3,4], Babar Hussain[5,6], Iqrar Ahmad Rana[3,7], Zulfiqar Ali[8], Hikmet Budak[9] and Ibrahim A. Alaraidh[10]**

[1]Molecular Breeding Laboratory, Rice Research Institute, Kala Shah Kaku, Pakistan, [2]Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Faisalabad, Pakistan, [3]Centre for Advanced Studies in Agriculture and Food Security, University of Agriculture Faisalabad, Faisalabad, Pakistan, [4]Precision Agriculture and Analytics Lab, National Centre for Big Data and Cloud Computing, University of Agriculture Faisalabad, Faisalabad, Pakistan, [5]Department of Biological Sciences, Middle East Technical University, Ankara, Turkey, [6]Department of Biotechnology, Faculty of Life Sciences, University of Central Punjab, Lahore, Pakistan, [7]Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture Faisalabad, Faisalabad, Pakistan, [8]Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Shareef University of Agriculture, Multan, Pakistan, [9]Montana BioAgriculture, Inc., Missoula, MT, United States, [10]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia

MADS-box gene family members play multifarious roles in regulating the growth and development of crop plants and hold enormous promise for bolstering grain yield potential under changing global environments. Bread wheat (*Triticum aestivum* L.) is a key stable food crop around the globe. Until now, the available information concerning MADS-box genes in the wheat genome has been insufficient. Here, a comprehensive genome-wide analysis identified 300 high confidence MADS-box genes from the publicly available reference genome of wheat. Comparative phylogenetic analyses with *Arabidopsis* and rice MADS-box genes classified the wheat genes into 16 distinct subfamilies. Gene duplications were mainly identified in subfamilies containing unbalanced homeologs, pointing towards a potential mechanism for gene family expansion. Moreover, a more rapid evolution was inferred for M-type genes, as compared with MIKC-type genes, indicating their significance in understanding the evolutionary history of the wheat genome. We speculate that subfamily-specific distal telomeric duplications in unbalanced homeologs facilitate the rapid adaptation of wheat to changing environments. Furthermore, our *in-silico* expression data strongly proposed MADS-box genes as active guardians of plants against pathogen insurgency and harsh environmental conditions. In conclusion, we provide an entire complement of MADS-box genes identified in the wheat genome that could accelerate functional genomics efforts and possibly facilitate bridging gaps between genotype-to-phenotype relationships through fine-tuning of agronomically important traits.

Keywords: genome-wide analysis, in-silico expression, TaMADS's, transcription factors, *Triticum aestivum*, wheat adaptability

OPEN ACCESS

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

Reviewed by:
Brijesh Kumar, Patanjali Research Institute, India
Soleyman Dayani, Imam Khomeini International University, Iran

****Correspondence:***
Qasim Raza
qasimnazami@gmail.com
Ibrahim A. Alaraidh
ialaraidh@ksu.edu.sa

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

Received: *20 November 2021*
Accepted: *21 December 2021*
Published: *17 January 2022*

Citation:
Raza Q, Riaz A, Atif RM, Hussain B, Rana IA, Ali Z, Budak H and Alaraidh IA (2022) Genome-Wide Diversity of MADS-Box Genes in Bread Wheat is Associated with its Rapid Global Adaptability. Front. Genet. 12:818880. doi: 10.3389/fgene.2021.818880

INTRODUCTION

In the fight for global food security and safety, bread wheat represents one of the largest contributing grain crops. It is one of the most produced, stored, and consumed food crops worldwide and a major source of energy and nutrients in the developing countries (Igrejas and Branlard, 2020). However, following the green revolution, improvement of wheat grain production has been hindered by various bottlenecks. This includes, but is not limited to, the non-availability of a reliable and fully annotated reference genome (Brenchley et al., 2012; Lukaszewski et al., 2014; Chapman et al., 2015; Clavijo et al., 2017; Zimin et al., 2017). The slow progress in harnessing a fully annotated reference genome was mainly due to the genome's allohexaploid nature, which comprises three closely related, but independently maintained, sub-genomes, known as A, B, and D. This, in combination with a high frequency of repetitive sequences, particularly hindered progress. Recently, an alliance of geneticists combined resources over 13 years, to produce a fully annotated sequence of the wheat genome, which had a resultant size of ~17 Gbps. This is the largest known genome among crop plants (Appels et al., 2018). To add to this, the high-quality of the reference genome, together with large scale RNA-seq data and expression repositories (Borrill et al., 2016; Ramírez-González et al., 2018), provide a rich resource for studying evolutionary dynamics and functional characterization of important gene families, which in turn could facilitate crop improvement efforts.

MADS-box transcription factors (TFs) are a well-documented group of genes known for playing vital roles in regulating the growth and development of several important plant species (Smaczniak et al., 2012; Ali et al., 2019). These genes influence diverse biological functions, including cell development, signal transduction, biotic and abiotic stress responses, vegetative organs development, control of flowering and anthesis time, formation of meristems and flower organs, ovule development, embryo development, dehiscence zone formation, and ripening of fruits and seeds (Colombo et al., 1995; Rounsley et al., 1995; Kuo et al., 1997; Riechmann and Meyerowitz, 1997; Alvarez-Buylla et al., 2000; Samach et al., 2000; Saedler et al., 2001; Moore et al., 2002; Messenguy and Dubois, 2003; Pařenicová et al., 2003; Smaczniak et al., 2012; Ali et al., 2019). Given their importance, identification, and characterization of MADS-box genes in agriculturally important species are critical for crop improvement and fine-tuning of specific traits through genetic exploitations.

All MADS-box genes contain a highly conserved MADS (M) domain of approximately 58–60 amino acids. The M domain enables DNA binding and is in the N-terminal region of the protein (Yanofsky et al., 1990). Members of this gene family are classified into M/type I and MIKC/type II super clades, which were generated after an ancient gene duplication event that occurred before the separation of animal and plant lineages (Becker and Theißen, 2003). M-type genes have simple intron-exon structures (zero or one intron), only contain a single M domain, and encode Serum Response Factor (SRF)-like proteins. These can be subcategorized into four clades (*Mα, Mβ, Mγ*, and *Mδ*). However, the *Mδ* clade genes closely resemble those within the *MIKC** group, as previously reported (De Bodt et al., 2003).

In comparison, MIKC-type genes have four domains: 1) a highly conserved DNA binding M domain, 2) a less conserved Intervening (I) domain of ~ 30 AA involved in dimer formation, 3) a moderately conserved Keratin (K) domain of ~ 70 AA which regulates heterodimerization of MADS proteins, and lastly 4) a highly inconstant C-terminal region which contributes to transcriptional regulation and higher-order protein complex formation (Henschel et al., 2002; Díaz-Riquelme et al., 2009). MIKC-type genes encode Myocyte Enhancer Factor 2 (MEF2)-like proteins and are categorized into *MIKC^c* and *MIKC** clades (Kaufmann et al., 2005). *MIKC**-type genes have duplicated K domains and relatively longer I domains than *MIKC^c* -types (Duan et al., 2014). *MIKC^c*-type genes can be further categorized into several subclades based on their phylogenetic relationships in flowering plants (Gramzow and Theißen, 2015).

Like what has been observed in other model and crop plant species, in wheat MADS-box genes are known to confer drought tolerance through regulation of drought tolerance genes and micro-RNAs (Budak et al., 2015). In addition to this, *TaMADS51, TaMADS4, TaMADS5, TaMADS6*, and *TaMADS18* all showed up-regulation, where *TaMADAGL17, TaMADAGL2, TaMADWM31C*, and *TaMADS14* were downregulated, under phosphorous (P) starvation, and further functional analysis confirmed their role in P-deficient stress responses (Shi et al., 2016). In another study, several MADS-box genes were activated and differentially expressed after inoculation of wheat spikes with fusarium head blight (FHB) (Kugler et al., 2013). Furthermore, a simple yet elegant ABCDE floral organ identity model explained the complex genetic interactions among MIKC-type wheat MADS-box genes for determining the fate of floral organs (Ali et al., 2019). These significant roles of MADS-box genes in biotic and abiotic stresses, fertilizer response, and flower development highlight the necessity of their comprehensive identification and characterization in the bread wheat genome.

Genome-wide identification and characterization of important gene families are becoming main stream research approaches during recent years (Ahmed et al., 2021; Aleem et al., 2022). Previously, comprehensive genome-wide analyses of MADS-box genes have been carried out in diverse plant species. However, the available information about wheat MADS-box genes is comparatively sparse. For example, Ma et al. (2017) reported the first genome-wide analysis of MADS-box genes using only a draft version of the wheat genome (TGACv1) which contains a lesser number of genes. Likewise, Schilling et al. (2020) studied only MIKC-type MADS-box genes from an updated genome version (Ref. Seq 1.0). Moreover, a significant number of low confidence and/or truncated protein-coding pseudogenes were also included in the before-mentioned studies.

To bridge the gaps in knowledge, in this study we comprehensively identify the entire complement of high confidence, full-length, protein-coding MADS-box genes present in the publicly available reference genome of bread wheat (IWGSC RefSeq v1.1). Through a combination of different search approaches, 300 high confidence, non-redundant, full-length MADS-box genes were identified.

Comparative phylogenetic analyses with model plant MADS-box genes further classified these into at least 16 *Arabidopsis* and/or grass specific subfamilies. Moreover, homeologous and duplicated genes were identified to study probable gene family expansion and evolution mechanisms. Furthermore, expression patterns of identified MADS-box genes were studied under several biotic and abiotic stress conditions. In this way, we provide a comprehensive resource of wheat MADS-box genes which have the potential to facilitate molecular breeders in fine-tuning important traits for further improvement.

MATERIAL AND METHODS

Identification of MADS-Box Genes

The publicly available genome version of wheat (IWGSC Ref. Seq v1.1) was accessed through the ensemble plants database (Bolser et al., 2017) and searched using protein family database (Pfam) identifiers of MADS (PF00319) and K (PF01486) domains. A total of 281 MADS-domain and 131 K-domain encoding genes were identified. Additionally, a query-based search using "MADS" yielded 300 genes. Together, these searches identified a total of 712 genes (**Supplementary Table S1**), among which 300 were found to be high confidence and non-redundant and considered for further analyses. Detailed information of these 300 genes was retrieved from the ensemble plants database and is provided in **Supplementary Table S2**. Moreover, amino acid sequences for the identified genes were uploaded into the NCBI conserved domain database (Marchler-Bauer et al., 2015) for validation of putative protein domains (**Supplementary Table S3**). Gene identifiers were used as gene names for subsequent analyses; however, alternative names used in previous studies (Schilling et al., 2020) were also provided in **Supplementary Table S4**.

Comparative Phylogenetic Analyses and Subfamily Classifications

Differentiation between type I (M-type) and type II (MIKC-type) MADS-box proteins were achieved by separate MAFFT alignments using only the MADS domain (L-INS-i algorithm) (Katoh and Standley, 2013; Katoh et al., 2018) between *Arabidopsis* (Pařenicová et al., 2003) and wheat, and rice (Arora et al., 2007) and wheat MADS-box proteins. Subsequently, maximum likelihood (ML) phylogenies were inferred using I_Q-T_{REE} (Nguyen et al., 2015) by choosing JTT + F + G4 best fit substitution model according to the Bayesian information criterion (BIC) (Kalyaanamoorthy et al., 2017). The consistency of the ML trees was validated by setting an Ultrafast bootstrap value of 1,000 (Minh et al., 2013; Hoang et al., 2018). The final phylogenetic trees were visualized with MEGA7 (Kumar et al., 2016).

Subfamily classifications were accomplished by separate MAFFT alignments within M-type (L-INS-i algorithm) and MIKC-type (E-INS-i algorithm) MADS-box protein sequences of *Arabidopsis*, rice, and wheat (Katoh and Standley, 2013; Katoh et al., 2018). The full-length alignments were subjected to the Gap Strip/Squeeze v2.1.0 tool (www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html) for masking the individual residues by removing the gaps with default parameters. Then, masked alignments of M-type and MIKC-type proteins were independently subjected to the I_Q-T_{REE} software (Nguyen et al., 2015) for generating ML phylogenetic trees as described above. Subfamily names were given by following subfamily classifications in *Arabidopsis* and/or major grass species (Gramzow and Theißen, 2015) (**Supplementary Table S2** and **Supplementary Table S4**).

Homeologs Identification

Putative homeologs were recognized based on strong phylogenetic relationships (Ultrafast bootstrap value >90) within different sub-families. Classifications reported in previous studies were also considered (Appels et al., 2018). The homeologs status of thirty-one genes could not be determined due to lower Ultrafast bootstrap values (**Supplementary Table S5**).

Gene Duplication and Evolution Analyses

Coding sequences (CDS) of all wheat MADS-box genes were retrieved from the ensemble plants database and blasted against each other using Sequence Demarcation Tool V1.2 (Muhire et al., 2014) for the identification of sequence identities. Gene pairs with ≥ 90% identity (E value < $1e^{-10}$) and non-homeologous status were considered as duplicated (Ning et al., 2017) (**Supplementary Table S6**). If the duplicated homologous gene pair was located on the same chromosome it was defined as tandem duplication. Otherwise, when homologous gene pairs were located on different chromosomes it was defined as segmental duplication. The CDS sequences of duplicated genes were MAFFT aligned, masked, and subjected to the Synonymous Non-Synonymous Analysis Program V2.1.1 (www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) to compute the synonymous (Ks) and non-synonymous (Ka) substitution rates. To find out which type of codon selection operated during evolution, the ratio of Ka/Ks was also calculated. The approximate divergence time between duplicated gene pairs was calculated by using formulae T = Ks/$2r \times 10^{-6}$ assuming a substitution rate (r) of 6.5×10^{-9} substitutions/synonymous site/year (El Baidouri et al., 2017) (**Supplementary Table S6**).

In-Silico Expression Analysis

Expression data under all available abiotic and biotic stress conditions were retrieved from the expVIP Wheat Expression Browser (Borrill et al., 2016; Ramírez-González et al., 2018) as Log_2 TPM (processed expression value in transcripts per million) obtained *via* RNA-seq analysis. Detailed information about expression levels of individual genes in tested tissues/growth stages and stresses/diseases are provided in the supplementary information (**Supplementary Table S7**). TBtools (Chen et al., 2020) was used to generate a heatmap from the obtained expression data. Expression based clustering of genes was achieved by following the K-means clustering method (K = 10, iterations = 1,000, runs = 5) (**Supplementary Table S8**).

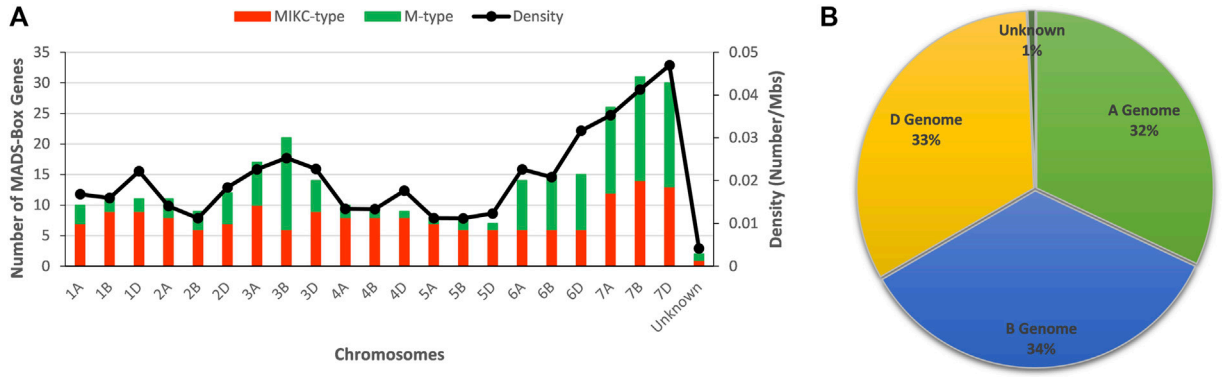


FIGURE 1 | Distribution pattern of MADS-box genes on wheat chromosomes and sub-genomes. **(A)** Number of MADS-box genes and their density on individual wheat chromosomes. **(B)** Percent contribution of each sub-genome to MADS-box genes galore.

RESULTS

MADS-Box Genes Galore in Wheat Genome

In this study, conserved domains and query search-based approaches identified a total of 300 high confidence and non-redundant MADS-box genes from the publicly available wheat genome (**Figure 1A**, **Supplementary Table S2**). The NCBI-CDD batch search further revealed that 167 (~55.7%) encode MADS-box proteins, 125 (~41.7%) encode MADS and K-box proteins, and only 8 (~2.6%) encode K-box domain-containing proteins (**Supplementary Table S3**). The protein lengths, molecular weight, and isoelectric points of MADS domain-containing proteins ranged from 58 to 450 amino acids, 6.483–48.304 kD, and 4.420–12.123 pI, respectively (**Supplementary Table S2**). This data suggests that different MADS-box genes may function within different environments.

Comparative phylogenetic analyses between *Arabidopsis* and wheat MADS-box genes (**Supplementary Figure S1**), as well as between rice and wheat (**Supplementary Figure S2**), distinguished wheat genes into M- and MIKC-types (**Supplementary Table S2**). In wheat, 128 (~43%) MADS-box genes exhibited high sequence similarity with *Arabidopsis* and rice M-type (type-I) genes, whereas 172 (~57%) revealed more homology with MIKC-type (type-II) genes. In general, MADS-box genes were equally distributed among the 21 wheat chromosomes, apart from the three homeologous chromosomes (7A, 7B and 7D), which harboured a significantly higher number of genes and displayed peak gene density values (**Figure 1A**). This observation could be explained by the higher prevalence of duplicated gene pairs on these chromosomes (Appels et al., 2018). However, M-type genes were randomly distributed among chromosomes and were predominantly located on three homeologous chromosomes of 3rd, 6th, and 7th linkage groups (**Figure 1A** and **Supplementary Figure S3**). In comparison, MIKC-type genes were equally distributed on all chromosomes. Furthermore, the percent contributions of A, B, and D genomes were also comparable (**Figure 1B**). As expected, MIKC-type genes were the largest group of MADS-box genes in wheat.

Subfamily Diversity in MADS-Box Genes

Separate ML phylogenies among *Arabidopsis*, rice, and wheat MADS-box genes exposed 14 MIKC-type and 3 M-type major subfamilies. The 172 MIKC-type genes were unevenly dispersed into *AP1* (9), *AP3* (6), *PI* (6), *AG/STK* (12), *SEP* (28), *AGL6* (3), *AGL12* (6), *AGL17* (31), *Bsister* (19), *MIKC** (27), *OsMADS32* (3), *SOC1* (13), and *SVP* (9) subfamilies (**Figure 2A**, **Supplementary Table S2**). As expected, monocot (*OsMADS32*) and eudicot (*FLC*) specific gene subfamilies were also witnessed. Likewise, 128 M-type genes were randomly distributed into *Mα* (53), *Mβ* (28), and *My* (47) subfamilies (**Figure 2B**).

In general, *Arabidopsis*, rice and wheat MADS-box gene subfamilies roughly followed the species-specific phylogenetic clades. Triads of wheat homeologous genes exhibited close relationships with one or more rice genes, with *Arabidopsis* genes representing a sister group association to grass genes (e.g., the *AP1*, *AP3*, *PI*, *AG/STK*, *AGL6*, *AGL12*, *OsMADS32*, *SOC1,* and *SVP* subfamilies; **Figure 2A**). Whereas the subfamily phylogenies were more complex in the case of *AGL17*, *Bsister*, *MIKC**, *SEP,* and all M-type MADS-box genes, probably due to multiple duplication events during the polyploidization of wheat genome.

Previously Classified *FLC*-Like Genes Grouped With *MIKC**-Like Genes

FLOWERING LOCUS C (FLC)-like genes have been confirmed to regulate flowering time and vernalization responses in plants (Distelfeld et al., 2009; Andrés and Coupland, 2012) and their wheat and rice counterparts have been reported based on sequence homology and phylogenetic tree reconstructions (Ruelens et al., 2013; Schilling et al., 2020). However, in this study, despite employing updated resources and tools (see *Materials and Method* section), none of the wheat and rice MADS-box genes fell into the *Arabidopsis* specific *FLC*-clade

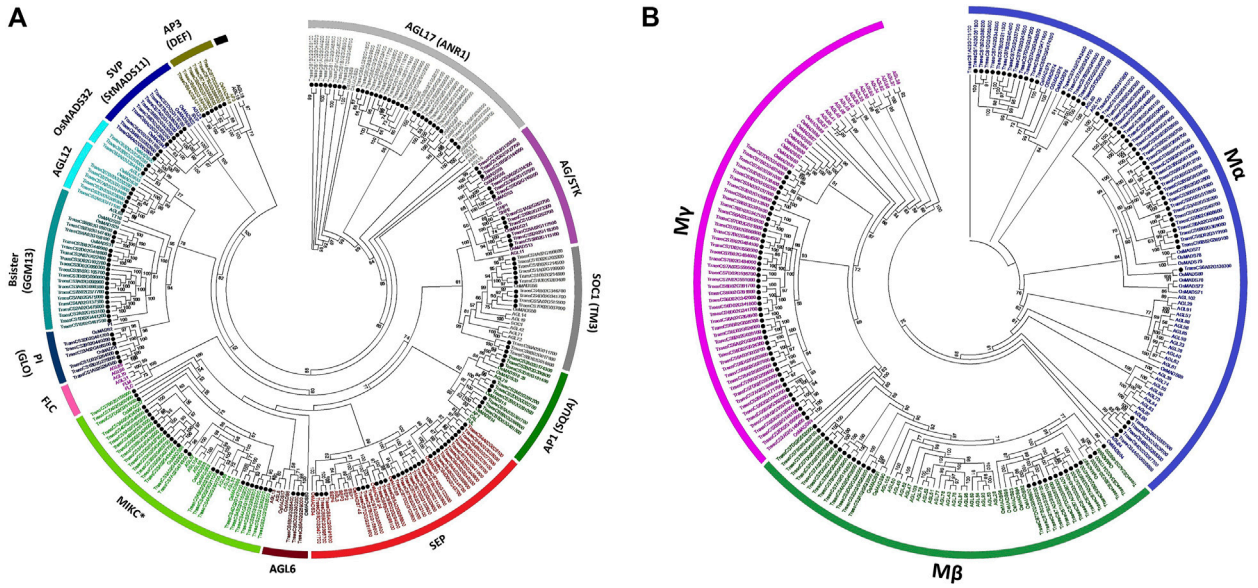


FIGURE 2 | Comparative phylogenetic analysis-based subfamily classifications of wheat MADS-box genes. *Arabidopsis*, rice, and wheat MADS-box proteins were MAFFT aligned (Katoh and Standley, 2013; Katoh et al., 2018), maximum likelihood phylogenies inferred using $IQ\text{-}T_{REE}$ software (Nguyen et al., 2015) and final trees visualized with MEGA7 (Kumar et al., 2016). Separate phylogenetic trees were inferred among *Arabidopsis*, rice, and wheat MIKC-type **(A)** and M-type **(B)** proteins. Wheat MADS-box genes are highlighted with solid black circles. Only bootstrap values ≥ 50%, as calculated from 1,000 replicates, could be displayed on the tree nodes. Subfamily-specific colouring was adopted for differentiating the different subfamilies. Subfamily names (outer bands) were given by following the subfamily classifications in *Arabidopsis* and/or major grass species (Gramzow and Theißen, 2015). Alternate gene names used in previous studies are also provided in **Supplementary Table S4**.

(**Figure 2A**). All the wheat genes which were previously classified as *FLC*-like (**Supplementary Table S4**) were grouped with *MIKC**-like MADS-box genes. Furthermore, we compared the amino acid sequences of *Arabidopsis* specific *FLC* genes and found that these were significantly different from *MIKC**-like genes (**Figure 3**). The most significant differences were detected in MADS domain region at 30th, 34th, and 50th positions where glutamic acid (E), glutamine (Q), and alanine/glycine/serine (A/G/S) residues of *FLC* genes were substituted with lysine (K), glutamic acid (E), and proline (P), respectively. Collectively, these results indicate that wheat and rice genomes might lack *FLC* clade.

Subfamily-Specific Gene Duplications in MADS-Box Genes

Overall, MADS-box genes were equally located in the interstitial and proximal regions (R2a, R2b and C) and distal telomeric (R1 and R3) ends of chromosomes (49 and 51%, respectively) (**Supplementary Table S2**). However, substantial differences were observed among gene locations of M-type and MIKC-type genes, as well as among subfamilies. Most of the M-type genes were in distal telomeric segments (62%), whereas MIKC-type genes were more prevalent in central chromosomal segments (57%). Generally, a larger portion of genes belonging to significantly expended subfamilies tend to be in distal telomeric ends, whereas genes of smaller subfamilies were more clustered in central chromosomal segments (**Supplementary Table S2**).

Gene duplications were identified through sequence similarities in coding sequences of all MADS-box genes. A total of 201 duplicated gene pairs with ≥90% sequence homology were identified, which corresponded to 123 non-redundant genes (**Figure 4**, **Supplementary Table S6**). Two genes (*TraesCSU02G209900*, *TraesCSU02G235300*) with unknown chromosomal location information were also recognized to be duplicated. However, *TraesCSU02G235300* showed duplications with genes located on chromosome 3B only, strongly suggesting that it was also located on the 3B chromosome. The MIKC-type MADS-box genes were also found to be more duplicated than M-type genes (60 vs 40% of duplicated gene pairs, respectively), particularly due to expended subfamilies (e.g., *AGL17*, *Bsister*, *MIKC**, *SEP*, and *SOC1*). Remarkably, duplicated gene pairs were subfamily-specific and particularly recognized in subfamilies containing unbalanced homeologs, except for the *AP3* subfamily (**Table 1**, **Supplementary Table S5** and **Supplementary Table S6**). Among subfamilies, *Mα*, *AGL17*, *MIKC**, and *SEP* contained 25, 21, 14, and 12% of the total duplicated gene pairs, respectively, and the majority of these were in distal telomeric and sub-telomeric (one gene located on the telomeric segment and other on the central segment) chromosomal regions. We also observed that the majority of the duplicated gene pairs (~51%) were in distal telomeric segments, whereas only 26 and 23% of the duplicated gene pairs were in proximal and sub-telomeric segments of chromosomes. These results could be explained by higher gene density in distal vs central chromosomal

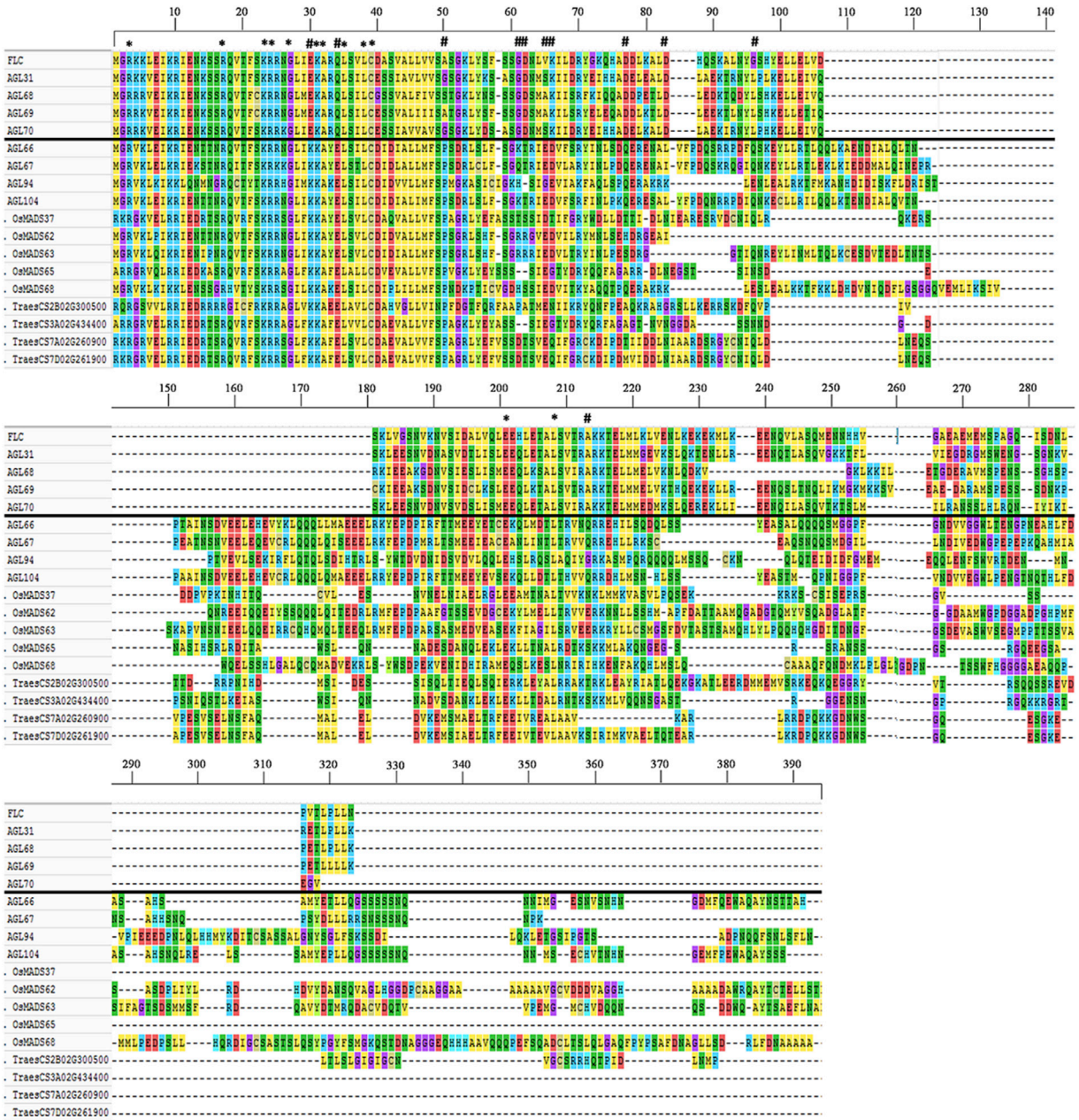


FIGURE 3 | Structural differentiation between *FLC*- and *MIKC**-like MADS-box genes. Amino acid sequences of representative genes from both subfamilies were aligned using Clustal Omega (Sievers et al., 2011) with default parameters and multiple sequence alignment visualized with MEGA7 (Kumar et al., 2016). *FLC*-like genes (above solid black line) were separated from *MIKC**-like genes (below solid black line). The conserved and diverged residues were indicated with star (*) and asterisk (#) symbols, respectively. Multiple sequence alignment demonstrated conservation of amino acid residues/motifs within both subfamilies, whereas diversification between subfamilies.

segments (Appels et al., 2018). Furthermore, segmental duplications were more prevalent than tandem duplications (61 and 37%) in 197 of the duplicated gene pairs with available chromosomal information (**Supplementary Table S6**). Interestingly, >37% of all tandem duplications were identified on chromosome 3B, consistent with IWGSC findings (Appels et al., 2018). Collectively, these results strongly suggest that unbalanced homeologs in distal telomeric regions derive MADS-box subfamilies expansion through segmental duplications.

Rapid Evolution of M-Type MADS-Box Genes

To investigate evolution rates, we estimated substitutions in coding sequences and computed approximate divergence time

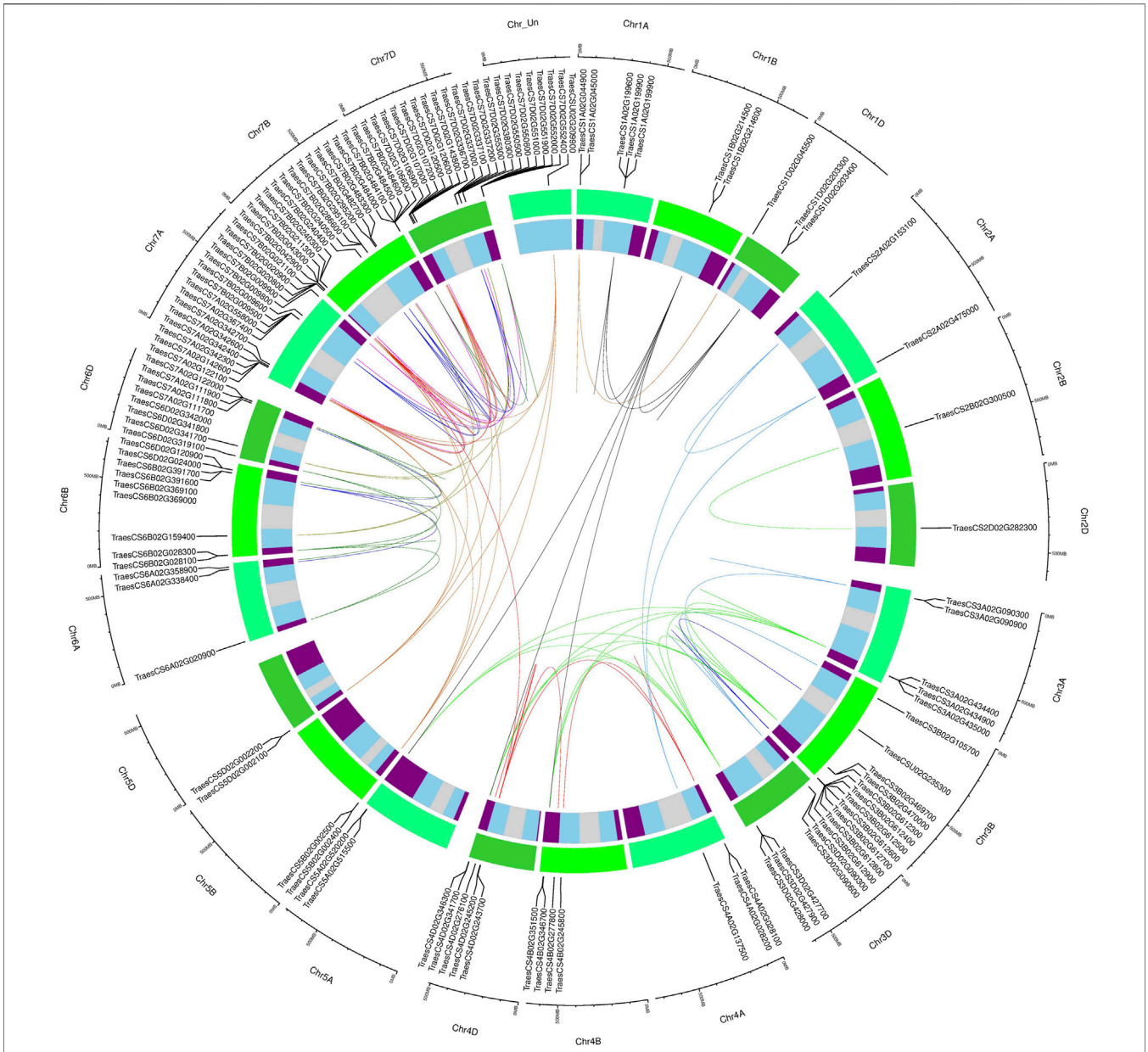


FIGURE 4 | Subfamily-specific gene duplications among wheat MADS-box genes. Duplicated genes were plotted in a circular diagram using respective physical positions with shinyCircos (Yu et al., 2018). The outer track indicates three different sub-genomes (shades of green), and the inner track represents chromosomal segments (R1 and R3, purple; R2a and R2b, sky blue; C, light grey) (Appels et al., 2018). Duplicated genes were identified through sequence similarity (**Supplementary Table S6**; see material and method section) and linked with subfamily-specific colours as in **Figures 2A,B**, except for *AGL17*-like genes which were linked using chocolate colour. The linked duplicated genes positioned on different wheat chromosomes represent segmental duplications, whereas tandem duplications were indicated by incomplete links within the same chromosomes.

between duplicated gene pairs (**Supplementary Table S6**). In M-type genes, 1st and 3rd quartiles of synonymous substitutions (Ks) were considerably narrower than MIKC-type genes, whereas non-synonymous to synonymous substitution ratios (Ka/Ks) of M-type genes were significantly higher than 1st and 3rd quartiles of MIKC-type genes (**Figure 5**). Furthermore, the percentage of duplication events because of positive/Darwinian selection (Ka/Ks ratio >1) was almost doubled (23%) in M-type genes as compared with MIKC-type genes (12%) (**Supplementary Table S6**). Additionally, the estimated divergence time of M-type genes was significantly narrower than the MIKC-type genes. The mean divergence time of M-type genes was nearly half of the mean of MIKC-type genes (**Figure 5**). Taken together, these

TABLE 1 | A potential relationship between unbalanced homeologs and gene duplications in wheat MADS-box gene family.

Subfamily	Total homeologs[a]	Balanced homeologs[b]	Unbalanced homeologs[c]	Gene duplications
Mα	53	33	20	Present
Mβ	28	10	18	Present
Mγ	47	26	21	Present
AG/STK	12	12	0	Absent
AGL12	6	6	0	Absent
AGL17	31	22	9	Present
AGL6	3	3	0	Absent
AP1	9	9	0	Absent
AP3	6	6	0	Present[d]
Bsister	19	10	9	Present
*MIKC**	27	21	6	Present
OsMADS32	3	3	0	Absent
PI	6	6	0	Absent
SEP	28	25	3	Present
SOC1	13	12	1	Present
SVP	9	9	0	Absent

[a]*Total number of gene copies from three different wheat genomes (A, B, D).*
[b]*At least one gene copy is present in all diploid genomes.*
[c]*Gene copy is missing in one or two genomes.*
[d]*Exception-gene duplications identified despite holding balanced homeologs.*

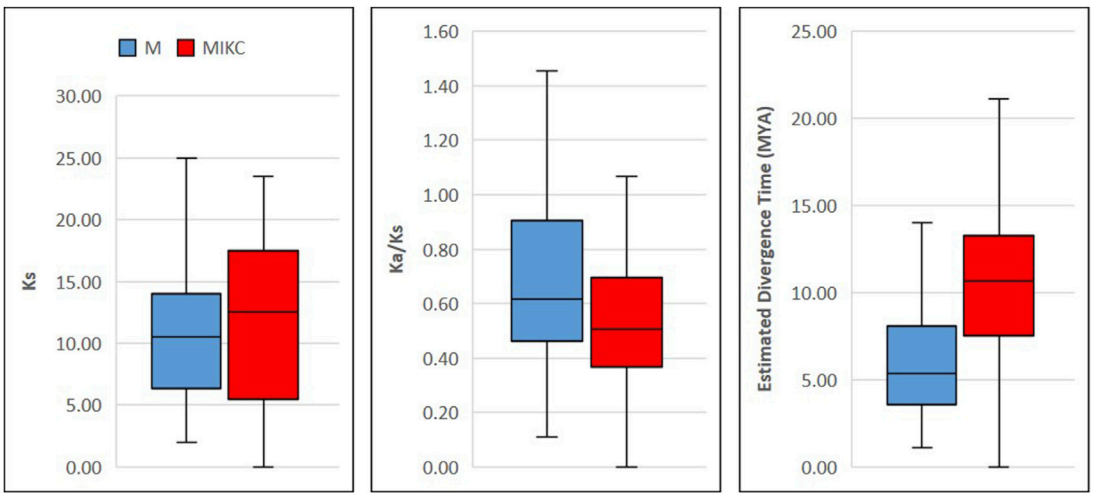


FIGURE 5 | Boxplots showing evolution patterns in M- and MIKC-type wheat MADS-box genes. The coding sequences of the duplicated genes were MAFFT aligned (Katoh and Standley, 2013; Katoh et al., 2018), masked and subjected to SNAP V2.1.1 program for computing the synonymous (Ks) and non-synonymous (Ka) substitution rates. The divergence time between duplicated genes pairs was estimated in million years ago (MYA) by following El Baidouri et al. (2017) and boxplots were generated using Microsoft Excel 2019.

data strongly indicate a rapid evolution of M-type MADS-box genes.

Expression Patterns of MADS-Box Genes Under Abiotic and Biotic Stresses

Extensive investigations have been carried out to study the expression patterns of MADS-box genes during the growth and development of crop plants. However, their transcriptional regulation under stressful conditions is somewhat obscure. Therefore, we analysed RNA-seq based expression data of 300 MADS-box genes under all available biotic and abiotic stress conditions in the exVIP wheat expression browser (Borrill et al., 2016; Ramírez-González et al., 2018) (**Figure 6**, **Supplementary Table S7**). Out of the total 300 genes, nearly 57% were expressed (log_2 TPM 0.20–7.89) during at least one developmental stage of one or more of the stresses included in this study. Whereas the remaining 43% of genes showed no or very low expression (log_2 TPM <0.0) and were subsequently considered as not expressed.

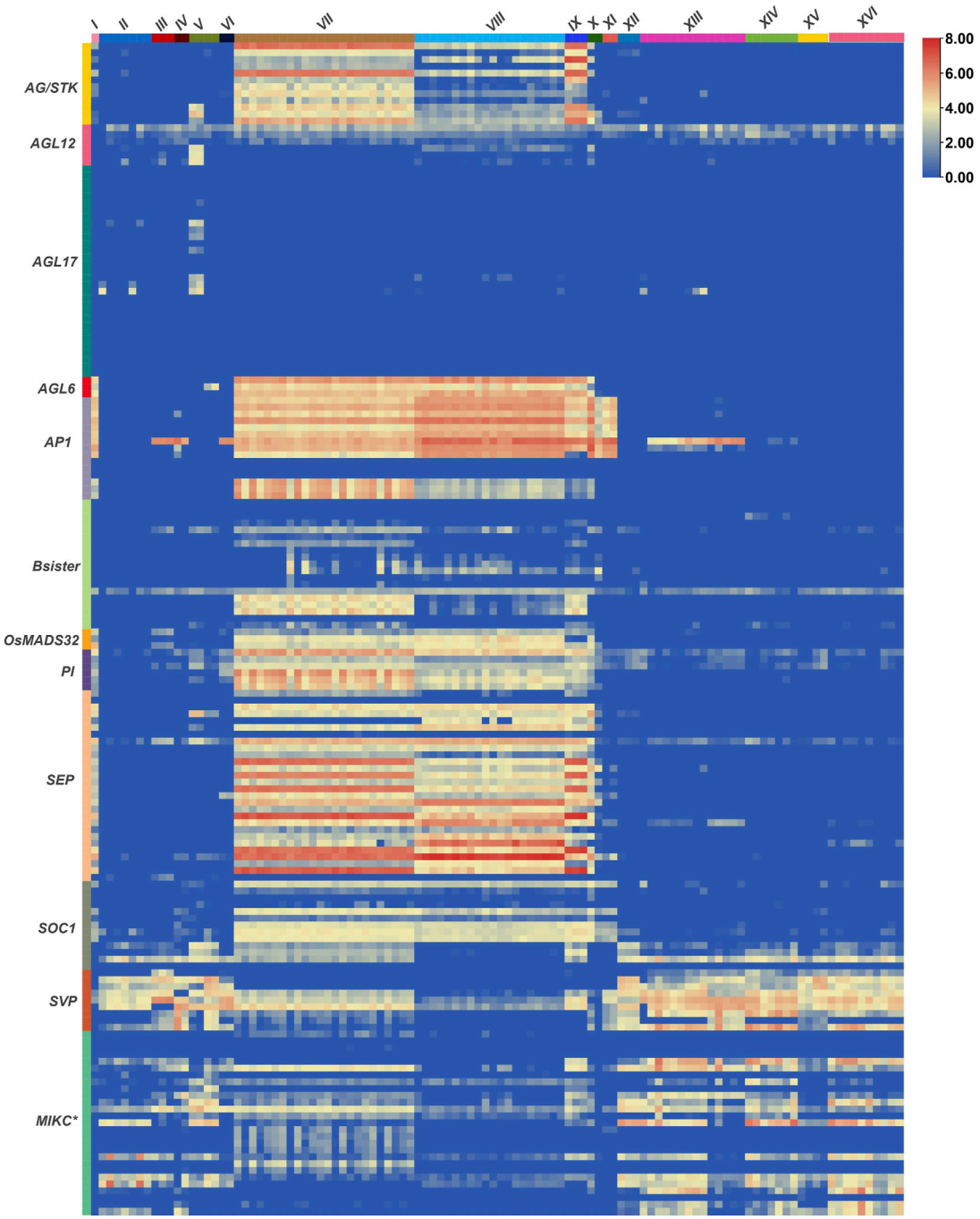


FIGURE 6 | Expression patterns of wheat MADS-box genes under stressful conditions. RNA-seq based expression data of all MADS-box genes were retrieved from the exVIP wheat expression browser (Borrill et al., 2016; Ramírez-González et al., 2018) and a heatmap was generated with TBtools (Chen et al., 2020). Processed expression levels of all genes under different abiotic (I, spikes with water stress; II, drought and heat-stressed seedlings; III, seedling treated with PEG to induce drought; IV, shoots after 2 weeks of cold stress; V, phosphate starvation in roots/shoots/leaves) and biotic stresses (VI, coleoptile infection with *Fusarium pseudograminearum*/crown root; VII, FHB infected spikelets (0–48 h); VIII, FHB infected spikelets (30–50 h); IX, spikes inoculated with FHB and ABA/GA; X, CS spikes inoculated with FHB; XI, leaves naturally infected with *Magnaporthe oryzae*; XII, PAMP inoculation of seedlings; XIII, stripe rust infected seedlings; XIV, stripe rust and powdery mildew infection in seedlings; XV, *Septoria tritici* infected seedlings; XVI, *Zymoseptoria tritici* infected seedlings) (columns) and in different subfamilies (rows) are presented as Log_2 transcripts per million (Log_2 TPM). Detailed information about expression levels of individual genes in tested tissues/growth stages and stresses/diseases are provided in **Supplementary Table S7**.

Overall, MIKC-type MADS-box genes were highly expressed under all studied stresses, whereas nearly 75% of the non-expressed genes belonged to the three M-type subfamilies. Interestingly, the remaining 25% of the non-expressed genes belonged to MIKC-subfamilies containing duplicated genes, e.g., *AGL17, MIKC*,* and *Bsister*. As most of the duplicated genes were in distal telomeric and sub-telomeric segments of chromosomes, it could be expected that promoters of these genes had undergone H3K27me3 (trimethylated histone H3 lysine 27) hypermethylation which resulted in their lower or non-existent expression (Ramírez-González et al., 2018).

In general, genes of all MIKC-type subfamilies (except *AGL17*) were ubiquitously expressed in fusarium head blight infected spikelets/spikes (**Figure 6**). Nevertheless, floral homeotic genes (*AP1*, *AP3*, *PI*, *AG/STK*, *AGL6,* and *SEP*) showed the highest expression patterns. *SVP*, *MIKC**, *AGL12,* and *SOC1*-like genes were also ubiquitously expressed in all studied abiotic and biotic stresses, strongly indicating that genes belonging to these subfamilies are active guardians of the wheat genome under stressful environments. Remarkably, all *AP1*-like genes demonstrated high expression in *Magnaporthe oryzae* infected leaves, suggesting disease-specific expression, as well as the involvement of these important subfamily genes in diverse aspects of wheat growth and development. Similarly, *AGL17*-like genes were expressed only in seedlings under drought, heat, and phosphate starvation conditions (**Figure 6**).

We also calculated expression based hierarchical clusters to analyze diversification in expression patterns present within subfamilies (**Supplementary Figure S4**, **Supplementary Table S8**). The *MIKC** and *SOC1*-like genes were grouped into seven and six random clusters, respectively. Similarly, *Bsister* and *SEP*-like genes were separately grouped into five random clusters. By contrast, *AP1* and *Mγ*-like genes exhibited no variation in their expression patterns, as all genes were grouped into a single random cluster. Likewise, *AGL6*, *AGL17*, *AP3*, *OsMADS32*, *Mα,* and *Mβ*-like genes displayed little diversification in expression, and genes of each subfamily were grouped into two random clusters. In comparison, many of the genes with no or very low expression (~58%) were grouped into cluster 10 and all belonged to subfamilies with duplicated genes (**Supplementary Table S8**).

DISCUSSION

Recent Significant Advancements in Identification and Characterization of MADS-Box Family Members in Wheat

MADS-box gene family members have been identified across all groups of eukaryotes and are known to confer diverse biological functions. In plants, these play a central role during growth and development and are thus very important targets for crop improvement (Ali et al., 2019). Recently, significant advancements have been made in the identification and characterization of MADS-box family members in wheat, which could facilitate crop breeding efforts and aid in the development of more resilient and higher-yielding cultivars. Ma et al. (2017) reported the first genome-wide analysis of MADS-box family members and identified a total of 180 MADS-box genes in an earlier genome version (Lukaszewski et al., 2014). Their *in-silico* expression data provided insights into the stress associated functions of MADS-box genes. To add to this, Schilling et al. (2020) conducted a comprehensive analysis of MIKC-type MADS-box genes and identified a total of 201 genes from an updated genome version (Ref. Seq 1.0) (Appels et al., 2018). They speculated that pervasive duplications, functional conservation, and putative neofunctionalization may have contributed to the adaptation of wheat to diverse environments. These advancements, along with results presented in this study, have begun to elucidate the broad landscape of wheat MADS-box genes and provide new insights into the phylogenomics, evolution, and stress associated functions of MADS-box family members.

MADS-Box Genes Are Underestimated in the Wheat Genome

Bread wheat is a hexaploid species (AABBDD, $2n = 6x = 42$), that originated from a series of naturally occurring hybridization events among three closely related and independently maintained sub-genomes. Its full genome size is ~17 Gbps, among which ~14.5 Gbps (85%) is sequenced and contains ~107,891 coding genes (Appels et al., 2018). In this study, through a comprehensive genome-wide analysis, we identified a total of 300 high confidence MADS-box genes (**Figure 1**, **Supplementary Table S2**). To date, this is the second-highest number of MADS-box genes identified in a crop plant, closely following *Brassica napus* (Wu et al., 2018), which contains 307 full-lengths and/or incomplete (pseudo) MADS-box genes. Possible explanations for the abundance found in wheat may encompass the larger genome size, higher gene number, high rate of homeolog retention, and hexaploid nature of bread wheat (Appels et al., 2018). Moreover, at least 35 low confidence MIKC-type MADS-box genes were recently reported in the IWGSC reference genome (Schilling et al., 2020). If these low confidence genes are also included with the 300 MADS-box genes identified in this study, then the final number will be the highest identified in a plant genome. Furthermore, a more recent chromosome-scale assembly of the bread wheat genome revealed hundreds of new genes and thousands of additional gene copies (Alonge et al., 2020) which were substantially missing from the IWGSC reference genome explored in the current study. Collectively, all these observations strongly suggest that MADS-box genes are underestimated in wheat and their exploration might help in understanding the rapid global adaptability of wheat.

Cereal Genomes Probably Lack *FLC*-Like Genes

*FLC*s are key genes conferring vernalization requirement, which act as flowering repressors in *Arabidopsis* (Whittaker and Dean, 2017). High expression of *FLCs* repress other flowering activator genes and causes delayed flowering. In winter varieties of

temperate cereals, including barley, *Brachypodium,* and wheat, vernalization is regulated through *VERNALIZATION* genes (*VRN1*, *VRN2*, *VRN3*) (Greenup et al., 2009). To date, strong disagreement exists in the published literature regarding the presence or absence of *FLC*-like genes in grasses (Paolacci et al., 2007; Zhao et al., 2011; Ruelens et al., 2013; Wei et al., 2014; Fatima et al., 2020; Schilling et al., 2020). Previously, grass genes were classified into *FLC*-like based on synteny and phylogenetic reconstruction methods (Ruelens et al., 2013; Schilling et al., 2020). However, despite using a more recent reference genome sequence and sophisticated phylogenetic tree reconstruction methods, we could not find *FLC*-clade in wheat and rice genomes (**Figure 2A**). Interestingly, all those genes which were previously classified as *FLC*-like (Ruelens et al., 2013; Schilling et al., 2020), as well as *OsMADS37* and *OsMADS51/65,* were grouped with *MIKC**-like genes (**Figure 2A**, **Supplementary Table S4**). In comparison, all other *FLC* homologs were grouped into a separate distinct *Arabidopsis*-specific *FLC* clade. Moreover, we also compared the multiple sequence amino acid alignments of *FLC*- and *MIKC**-like genes and observed that these were significantly dissimilar and none of the grass genes shared sequence homology with *Arabidopsis FLC* paralogs (**Figure 3**). Altogether, these results indicate that *FLC*-like genes might have been lost in cereal genomes after they diverged from eudicots. Several other studies are also in agreement with current results and were unable to identify *FLC*-like genes in grass genomes (Arora et al., 2007; Paolacci et al., 2007; Zhao et al., 2011; Wei et al., 2014; Fatima et al., 2020). These observations suggest that vernalization pathways in monocots and dicots evolved independently and are regulated through a completely different set of genes. However, earlier reports on the identification of *FLC*-like genes in grasses (Ruelens et al., 2013; Sharma et al., 2017; Schilling et al., 2020) and functional conservation of *ODDSOC2* with *FLC* in the regulation of vernalization (Greenup et al., 2010), suggest that evolution of vernalization pathways between monocots and dicots may not be fully independent. Thus, the biological question of the presence or absence of *FLC*-like genes in grass genomes still needs to be explored. Further research efforts in this direction might address this fundamental question, and propitious outcomes could help in breeding efforts to develop winter varieties of temperate cereals which are adapted to changing environments.

Subfamily-Specific Distal Telomeric Duplications in Unbalanced Homeologs Facilitate Rapid Adaptation to Changing Environments

Distal telomeric chromosomal segments are evolutionary hotspots for frequent recombination events and give rise to fast-evolving genes (Glover et al., 2015; Chen N. W. G. et al., 2018). Many adaptability traits related genes that are induced in response to external stimuli are found to be predominately located in distal chromosomal segments. In comparison, genes related to housekeeping and conserved developmental functions are positioned in central chromosomal segments (Appels et al., 2018; Ramírez-González et al., 2018). In this study, the majority of the identified duplicated gene pairs were in distal telomeric (~51%) and sub-telomeric (23%) chromosomal segments (**Figure 4**, **Supplementary Table S6**). Remarkably, all these duplications were subfamily-specific and primarily identified in larger subfamilies containing unbalanced homeologs (*Mα*, *Mβ*, *Mγ*, *AGL17*, *Bsister*, *MIKC**, *SEP*, and *SOC1*) (**Table 1**). Interestingly, genes belonging to these subfamilies have been reported to regulate plants adaptability in changing environments. For example, *OsMADS57* (*AGL17*-like gene in rice) is induced by abscisic acid, chilling, drought, and salinity stresses (Arora et al., 2007), promotes cold tolerance by interacting with a defence gene (*OsWRKY94*) (Chen L. et al., 2018), and modulates root to shoot nitrate translocation under deprived nitrate conditions (Huang et al., 2019). Similarly, downregulation of a *SEP* clade gene in pepper (*CaMADS*) caused more sensitivity to cold, salt, and osmotic stresses, whereas its overexpression in *Arabidopsis* conferred higher tolerance against these stresses (Chen et al., 2019). By contrast, almost negligible duplications were identified in a single smaller subfamily (*AP3*) with completely balanced homeologs (**Table 1**). These results strongly indicate that unbalanced homeologs of expended subfamilies have undergone subfamily-specific distal telomeric duplications, thus facilitating the rapid adaptation of bread wheat to diverse global environments. Contrarily, completely balanced homeologs of functionally conserved smaller subfamilies might have some evolutionary advantage in minimizing the developmentally detrimental gene copy number variations, due to their localization in central chromosomal segments.

M-Type MADS-Box Genes Are Evolving at a Faster Rate

Unlike MIKC-type MADS-box genes, the M-type genes have not been extensively studied and little information is available about their evolutionary origin in crop plants (Masiero et al., 2011). In this study, we observed that M-type genes were predominately located in distal telomeric chromosomal segments (**Supplementary Table S2**) and are evolving at a faster rate, probably due to the higher frequency of tandem gene duplications and stronger purifying selections (**Figure 5**, **Supplementary Table S6**). Moreover, lesser synonymous substitutions and higher non-synonymous to synonymous substitution ratios of M-type duplicated genes exposed their more rapid evolution as compared with MIKC-type genes (**Figure 5**, **Supplementary Table S6**). These results agree with Nam et al. (2004), who also reported faster birth-and-death evolution of M-type genes in angiosperms. This data might indicate the significance of M-type MADS-box genes in understanding the evolutionary history of the bread wheat genome.

MADS-Box Genes as Active Guardians of Plants Against Pathogen Insurgency and Harsh Environments

MADS-box genes regulate diverse developmental processes, and their functions are well studied in plant morpho- and

organogenesis (Smaczniak et al., 2012; Ali et al., 2019). However, several members of the MADS-box gene family are reported to be involved in the regulation of biotic and abiotic stress responses (Zhang et al., 2016; Wang Q. et al., 2018; Castelán-Muñoz et al., 2019), which point towards possible dynamic roles in stress response. In this study, we analysed the complete atlas of their expression profiles in all publicly available wheat transcriptomic data under biotic and abiotic stresses (**Figure 6** and **Supplementary Figure S4**, **Supplementary Table S7** and **Supplementary Table S8**). Except for *AGL17*, genes of all other MIKC-type subfamilies were ubiquitously expressed in FHB infected spikes/spikelets (**Figure 6**) mimicking their defensive roles against *Fusarium* infection. Previously, Yang et al. (2015) functionally characterized *FgMcm1* in the causal agent (*Fusarium graminearum*) of barley and wheat head blight disease and demonstrated that *FgMcm1* played crucial roles in cell identity and fungal development. More recently, Xu et al. (2020) also reported a close relationship between anther extrusion and field FHB resistance and pointed out that *Rht* and *Vrn* genes might have pleiotropic effects on these traits. Since FHB is a floral disease and floral organ identity is controlled by MIKC-type MADS-box genes (Ali et al., 2019), this indicates a cooperate relationship for their involvement in the pathogen response network. In future, it would be interesting to investigate how MADS-box genes are involved in fighting FHB infection.

Similarly, *SVP*, *MIKC**, *AGL12,* and *SOC1* subfamily genes were ubiquitously expressed under all studied stresses (**Figure 6**, **Supplementary Table S7**), pointing towards their more critical role in stress associated functions. In comparison, members of the largest MIKC group (*AGL17*) were expressed under phosphate starvation, and heat and drought stress conditions. Several functionally characterized genes belonging to all these subfamilies have also been reported to regulate different abiotic and/or biotic stresses in plants (Khong et al., 2015; Shi et al., 2016; Wang Z. et al., 2018; Hwang et al., 2019; Li P. et al., 2020). Interestingly, few M-type genes, especially *Mα*-like, were weakly expressed under different biotic stresses (**Supplementary Table S7**) and these expression patterns agree with Guo et al. (2013), who observed differential expression of an M-type gene in response to stripe rust infection in wheat. Collectively, these results strongly highlight MADS-box genes, especially MIKC-type, as key members of gene regulatory networks (Castelán-Muñoz et al., 2019) and their involvement in wheat response to possible pathogen insurgency and harsh environments.

Towards Bridging the Gaps Between Genotype and Phenotype Using Genetic and Genomic Resources

Understanding the functions of candidate genes controlling agronomically important traits is critical for the acceleration of crop improvement efforts. Following the unveiling of recent wheat genetic and genomic resources (Avni et al., 2017; Luo et al., 2017; Maccaferri et al., 2019; Adamski et al., 2020), rapid testing of previous model plant discoveries and their cross-application in staple crops is paving the way towards bridging the gap between genotype and phenotype. During recent years, several wheat genes related to flowering, yield, quality, disease resistance, male sterility, and nutrient use efficiency traits have been cloned and functionally characterized [briefly reviewed by Li J. et al. (2020)], which exposed the unprecedented potential of these genetic and functional genomic resources in establishing genotype-to-phenotype relationships. These resources could also facilitate in elucidating the regulatory roles of neglected MADS-box subfamilies, such as *Mα*, *Mβ*, *Mγ,* and *MIKC**, during wheat growth and development, and subsequently help in deciphering whether MADS-box genes confer pathogen resistance and/or abiotic stress tolerance. Furthermore, many other fundamental biological questions about their phylogenomics, evolutionary origin, and stress associated functions, could be addressed by exploiting these paramount resources.

CONCLUSION

MADS-box genes are critically important for wheat growth and development and hold enormous promise for bolstering yield potential under changing global environments. Our results speculate that the abundance of MADS-box genes in the wheat genome might be associated with its global adaptability and that subfamily-specific distal telomeric duplications in unbalanced homeologs facilitate its rapid adaptation. In addition, through comprehensive genome-wide and comparative analyses, we demonstrated that wheat and rice genomes might lack *FLC*-like genes, which could help molecular breeders in the identification of alternative target genes for fine-tuning of winter wheat varieties. Moreover, our *in-silico* expression data strongly indicated the possibility of protective roles of MADS-box genes against pathogen attacks and harsh climatic conditions. In this way, we provided an entire complement of MADS-box genes identified in the wheat genome that could accelerate functional genomics efforts and possibly facilitate in bridging genotype-to-phenotype relationships through fine-tuning of agronomically important traits.

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

QR conceived and designed the study. QR, AR, RA, BH and IR retrieved the data. QR, AR and RA curated the data and performed final analyses. QR, AR and BH drafted the manuscript. QR, IR, ZA, HB and IA revised and finalized the manuscript. IA secured funding for research publication. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

All authors are grateful to Dr Yuling Jiao, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences and the reviewers for their comments and suggestions on the manuscript. The inspiration and support from Precision Agriculture and Analytics Lab (PAAL), an affiliate of National Centre in Big Data and Cloud Computing (NCBC) is also gratefully acknowledged. A preprint version of the manuscript is publicly available at BioRxiv (Raza et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.818880/full#supplementary-material

REFERENCES

Adamski, N. M., Borrill, P., Brinton, J., Harrington, S. A., Marchal, C., Bentley, A. R., et al. (2020). A Roadmap for Gene Functional Characterisation in Crops with Large Genomes: Lessons from Polyploid Wheat. *Elife* 9, e55646. doi:10.7554/eLife.55646

Ahmed, S., Rashid, M. A. R., Zafar, S. A., Azhar, M. T., Waqas, M., Uzair, M., et al. (2021). Genome-wide Investigation and Expression Analysis of APETALA-2 Transcription Factor Subfamily Reveals its Evolution, Expansion and Regulatory Role in Abiotic Stress Responses in Indica Rice (Oryza Sativa L. Ssp. Indica). *Genomics* 113, 1029–1043. doi:10.1016/J.YGENO.2020.10.037

Aleem, M., Riaz, A., Raza, Q., Aleem, M., Aslam, M., Kong, K., et al. (2022). Genome-wide Characterization and Functional Analysis of Class III Peroxidase Gene Family in Soybean Reveal Regulatory Roles of GsPOD40 in Drought Tolerance. *Genomics* 114, 45–60. doi:10.1016/J.YGENO.2021.11.016

Ali, Z., Raza, Q., Atif, R. M., Aslam, U., Ajmal, M., and Chung, G. (2019). Genetic and Molecular Control of floral Organ Identity in Cereals. *Int. J. Mol. Sci.* 20, 2743. doi:10.3390/ijms20112743

Alonge, M., Shumate, A., Puiu, D., Zimin, A. V., and Salzberg, S. L. (2020). Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. *Genetics* 216, 599–608. doi:10.1534/genetics.120.303501

Alvarez-Buylla, E. R., Liljegren, S. J., Pelaz, S., Gold, S. E., Burgeff, C., Ditta, G. S., et al. (2000). MADS-box Gene Evolution beyond Flowers: Expression in Pollen, Endosperm, Guard Cells, Roots and Trichomes. *Plant J.* 24, 457–466. doi:10.1046/j.1365-313X.2000.00891.x

Andrés, F., and Coupland, G. (2012). The Genetic Basis of Flowering Responses to Seasonal Cues. *Nat. Rev. Genet.* 13, 627–639. doi:10.1038/nrg3291

Appels, R., Eversole, K., Appels, R., Eversole, K., Feuillet, C., Keller, B., et al. (2018). Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome. *Science* 361, 361. doi:10.1126/science.aar7191

Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., et al. (2007). MADS-box Gene Family in rice: Genome-wide Identification, Organization and Expression Profiling during Reproductive Development and Stress. *BMC Genomics* 8, 242. doi:10.1186/1471-2164-8-242

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., et al. (2017). Wild Emmer Genome Architecture and Diversity Elucidate Wheat Evolution and Domestication. *Science* 357, 93–97. doi:10.1126/science.aan0032

Becker, A., and Theißen, G. (2003). The Major Clades of MADS-Box Genes and Their Role in the Development and Evolution of Flowering Plants. *Mol. Phylogenet. Evol.* 29 (3), 464–489. doi:10.1016/S1055-7903(03)00207-0

Bolser, D. M., Staines, D. M., Perry, E., and Kersey, P. J. (2017). Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomic Data. *Methods Mol. Biol.* 1533, 1–31. doi:10.1007/978-1-4939-6658-5_1

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: A Customizable RNA-Seq Data Analysis and Visualization Platform. *Plant Physiol.* 170, 2172–2186. doi:10.1104/pp.15.01667

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'Amore, R., Allen, A. M., et al. (2012). Analysis of the Bread Wheat Genome Using Whole-Genome Shotgun Sequencing. *Nature* 491, 705–710. doi:10.1038/nature11650

Budak, H., Hussain, B., Khan, Z., Ozturk, N. Z., and Ullah, N. (2015). From Genetics to Functional Genomics: Improvement in Drought Signaling and Tolerance in Wheat. *Front. Plant Sci.* 6, 1012. doi:10.3389/fpls.2015.01012

Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-Ponce, B., et al. (2019). MADS-box Genes Are Key Components of Genetic Regulatory Networks Involved in Abiotic Stress and Plastic Developmental Responses in Plants. *Front. Plant Sci.* 10, 853. doi:10.3389/fpls.2019.00853

Chapman, J. A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., et al. (2015). A Whole-Genome Shotgun Approach for Assembling and Anchoring the Hexaploid Bread Wheat Genome. *Genome Biol.* 16, 26. doi:10.1186/s13059-015-0582-8

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

Chen, L., Zhao, Y., Xu, S., Zhang, Z., Xu, Y., Zhang, J., et al. (2018). OsMADS57 Together with OsTB1 Coordinates Transcription of its Target OsWRKY94 and D14 to Switch its Organogenesis to Defense for Cold Adaptation in rice. *New Phytol.* 218, 219–231. doi:10.1111/nph.14977

Chen, N. W. G., Thareau, V., Ribeiro, T., Magdelenat, G., Ashfield, T., Innes, R. W., et al. (2018). Common Bean Subtelomeres Are Hot Spots of Recombination and Favor Resistance Gene Evolution. *Front. Plant Sci.* 9, 1185. doi:10.3389/fpls.2018.01185

Chen, R., Ma, J., Luo, D., Hou, X., Ma, F., Zhang, Y., et al. (2019). CaMADS, a MADS-Box Transcription Factor from Pepper, Plays an Important Role in the Response to Cold, Salt, and Osmotic Stress. *Plant Sci.* 280, 164–174. doi:10.1016/j.plantsci.2018.11.020

Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. (2017). An Improved Assembly and Annotation of the Allohexaploid Wheat Genome Identifies Complete Families of Agronomic Genes and Provides Genomic Evidence for Chromosomal Translocations. *Genome Res.* 27, 885–896. doi:10.1101/gr.217117.116

Colombo, L., Franken, J., Koetje, E., Van Went, J., Dons, H. J., Angenent, G. C., et al. (1995). The Petunia MADS Box Gene FBP11 Determines Ovule Identity. *Plant Cell* 7, 1859–1868. doi:10.1105/tpc.7.11.1859

De Bodt, S., Raes, J., Van De Peer, Y., and Theißen, G. (2003). And Then There Were many: MADS Goes Genomic. *Trends Plant Sci.* 8, 475–483. doi:10.1016/j.tplants.2003.09.006

Díaz-Riquelme, J., Lijavetzky, D., Martínez-Zapater, J. M., and Carmona, M. J. (2009). Genome-wide Analysis of MIKCC-type MADS Box Genes in grapevine. *Plant Physiol.* 149, 354–369. doi:10.1104/pp.108.131052

Distelfeld, A., Li, C., and Dubcovsky, J. (2009). Regulation of Flowering in Temperate Cereals. *Curr. Opin. Plant Biol.* 12, 178–184. doi:10.1016/j.pbi.2008.12.010

Duan, W., Song, X., Liu, T., Huang, Z., Ren, J., Hou, X., et al. (2014). Genome-wide Analysis of the MADS-Box Gene Family in Brassica Rapa (Chinese Cabbage). *Mol. Genet. Genomics* 290, 239–255. doi:10.1007/s00438-014-0912-7

El Baidouri, M., Murat, F., Veyssiere, M., Molinier, M., Flores, R., Burlot, L., et al. (2017). Reconciling the Evolutionary Origin of Bread Wheat (*Triticum aestivum*). *New Phytol.* 213, 1477–1486. doi:10.1111/nph.14113

Fatima, M., Zhang, X., Lin, J., Zhou, P., Zhou, D., and Ming, R. (2020). Expression Profiling of MADS-Box Gene Family Revealed its Role in Vegetative Development and Stem Ripening in S. Spontaneum. *Sci. Rep.* 10, 20536. doi:10.1038/s41598-020-77375-6

Glover, N. M., Daron, J., Pingault, L., Vandepoele, K., Paux, E., Feuillet, C., et al. (2015). Small-scale Gene Duplications Played a Major Role in the Recent

Evolution of Wheat Chromosome 3B. *Genome Biol.* 16, 188. doi:10.1186/s13059-015-0754-6

Gramzow, L., and Theißen, G. (2015). Phylogenomics Reveals Surprising Sets of Essential and Dispensable Clades of MIKCc-Group MADS-Box Genes in Flowering Plants. *J. Exp. Zool. (Mol. Dev. Evol.* 324, 353–362. doi:10.1002/jez.b.22598

Greenup, A. G., Sasani, S., Oliver, S. N., Talbot, M. J., Dennis, E. S., Hemming, M. N., et al. (2010). ODDSOC2 Is a MADS Box floral Repressor that Is Down-Regulated by Vernalization in Temperate Cereals. *Plant Physiol.* 153 (3), 1062–1073. doi:10.1104/pp.109.152488

Greenup, A., Peacock, W. J., Dennis, E. S., and Trevaskis, B. (2009). The Molecular Biology of Seasonal Flowering-Responses in Arabidopsis and the Cereals. *Ann. Bot.* 103, 1165–1172. doi:10.1093/aob/mcp063

Guo, J., Shi, X.-X., Zhang, J.-S., Duan, Y.-H., Bai, P.-F., Guan, X.-N., et al. (2013). A Type I MADS-Box Gene Is Differentially Expressed in Wheat in Response to Infection by the Stripe Rust Fungus. *Biol. Plant* 57, 540–546. doi:10.1007/s10535-012-0297-6

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T., and Theißen, G. (2002). Two Ancient Classes of MIKC-type MADS-Box Genes Are Present in the moss Physcomitrella Patens. *Mol. Biol. Evol.* 19, 801–814. doi:10.1093/oxfordjournals.molbev.a004137

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522. doi:10.1093/molbev/msx281

Huang, S., Liang, Z., Chen, S., Sun, H., Fan, X., Wang, C., et al. (2019). A Transcription Factor, OsMADS57, Regulates Long-Distance Nitrate Transport and Root Elongation. *Plant Physiol.* 180 (2), 882–895. doi:10.1104/pp.19.00142

Hwang, K., Susila, H., Nasim, Z., Jung, J.-Y., and Ahn, J. H. (2019). Arabidopsis ABF3 and ABF4 Transcription Factors Act with the NF-YC Complex to Regulate SOC1 Expression and Mediate Drought-Accelerated Flowering. *Mol. Plant* 12, 489–505. doi:10.1016/j.molp.2019.01.002

Igrejas, G., and Branlard, G. (2020). "The Importance of Wheat," in *Wheat Quality for Improving Processing and Human Health*. Editors G. Igrejas, T. Ikeda, and C. Guzmán (Cham, Switzerland: Springer, Cham), 1–7.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* 14, 587–589. doi:10.1038/nmeth.4285

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief. Bioinform* 20, 1160–1166. doi:10.1093/bib/bbx108

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010

Kaufmann, K., Melzer, R., and Theißen, G. (2005). MIKC-type MADS-Domain Proteins: Structural Modularity, Protein Interactions and Network Evolution in Land Plants. *Gene* 347, 183–198. doi:10.1016/j.gene.2004.12.014

Khong, G. N., Pati, P. K., Richaud, F., Parizot, B., Bidzinski, P., Mai, C. D., et al. (2015). OsMADS26 Negatively Regulates Resistance to Pathogens and Drought Tolerance in rice. *Plant Physiol.* 169 (4), 2935–2949. doi:10.1104/pp.15.01192

Kugler, K. G., Siegwart, G., Nussbaumer, T., Ametz, C., Spannagl, M., Steiner, B., et al. (2013). Quantitative Trait Loci-dependent Analysis of a Gene Co-expression Network Associated with Fusarium Head Blight Resistance in Bread Wheat (Triticum aestivum L.). *BMC Genomics* 14, 728. doi:10.1186/1471-2164-14-728

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Kuo, M. H., Nadeau, E. T., and Grayhack, E. J. (1997). Multiple Phosphorylated Forms of the *Saccharomyces cerevisiae* Mcm1 Protein Include an Isoform Induced in Response to High Salt Concentrations. *Mol. Cel. Biol.* 17, 819–832. doi:10.1128/mcb.17.2.819

Li, J., Yang, J., Li, Y., and Ma, L. (2020). Current Strategies and Advances in Wheat Biology. *Crop J.* 8, 879–891. doi:10.1016/j.cj.2020.03.004

Li, P., Zhang, Q., He, D., Zhou, Y., Ni, H., Tian, D., et al. (2020). AGAMOUS-LIKE67 Cooperates with the Histone Mark Reader EBS to Modulate Seed Germination under High Temperature. *Plant Physiol.* 184 (1), 529–545. doi:10.1104/pp.20.00056

Lukaszewski, A. J., Alberti, A., Sharpe, A., Kilian, A., Stanca, A. M., Keller, B., et al. (2014). A Chromosome-Based Draft Sequence of the Hexaploid Bread Wheat (*Triticum aestivum*) Genome. *Science* 345, 1251788. doi:10.1126/science.1251788

Luo, M.-C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome Sequence of the Progenitor of the Wheat D Genome Aegilops Tauschii. *Nature* 551, 498–502. doi:10.1038/nature24486

Ma, J., Yang, Y., Luo, W., Yang, C., Ding, P., Liu, Y., et al. (2017). Genome-wide Identification and Analysis of the MADS-Box Gene Family in Bread Wheat (*Triticum aestivum* L.). *PLoS One* 12, e0181443. doi:10.1371/journal.pone.0181443

Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., et al. (2019). Durum Wheat Genome Highlights Past Domestication Signatures and Future Improvement Targets. *Nat. Genet.* 51, 885–895. doi:10.1038/s41588-019-0381-3

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's Conserved Domain Database. *Nucleic Acids Res.* 43, D222–D226. doi:10.1093/nar/gku1221

Masiero, S., Colombo, L., Grini, P. E., Schnittger, A., and Kater, M. M. (2011). The Emerging Importance of Type I MADS Box Transcription Factors for Plant Reproduction. *Plant Cell* 23, 865–872. doi:10.1105/tpc.110.081737

Messenguy, F., and Dubois, E. (2003). Role of MADS Box Proteins and Their Cofactors in Combinatorial Control of Gene Expression and Cell Development. *Gene* 316, 1–21. doi:10.1016/S0378-1119(03)00747-9

Minh, B. Q., Nguyen, M. A. T., and Von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi:10.1093/molbev/mst024

Moore, S., Vrebalov, J., Payton, P., and Giovannoni, J. (2002). Use of Genomics Tools to Isolate Key Ripening Genes and Analyse Fruit Maturation in Tomato. *J. Exp. Bot.* 53, 2023–2030. doi:10.1093/jxb/erf057

Muhire, B. M., Varsani, A., and Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS One* 9, e108277. doi:10.1371/journal.pone.0108277

Nam, J., Kim, J., Lee, S., An, G., Ma, H., and Nei, M. (2004). Type I MADS-Box Genes Have Experienced Faster Birth-And-Death Evolution Than Type II MADS-Box Genes in Angiosperms. *Pnas* 101, 1910–1915. doi:10.1073/pnas.0308430100

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi:10.1093/molbev/msu300

Ning, P., Liu, C., Kang, J., and Lv, J. (2017). Genome-wide Analysis of WRKY Transcription Factors in Wheat (*Triticum aestivum* L.) and Differential Expression under Water Deficit Condition. *PeerJ* 5, e3232. doi:10.7717/peerj.3232

Paolacci, A. R., Tanzarella, O. A., Porceddu, E., Varotto, S., and Ciaffi, M. (2007). Molecular and Phylogenetic Analysis of MADS-Box Genes of MIKC Type and Chromosome Location of SEP-like Genes in Wheat (*Triticum aestivum* L.). *Mol. Genet. Genomics* 278, 689–708. doi:10.1007/s00438-007-0285-2

Pařenicova´, L., De Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and Phylogenetic Analyses of the Complete MADS-Box Transcription Factor Family in Arabidopsis. *Plant Cell* 15, 1538–1551. doi:10.1105/tpc.011544

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The Transcriptional Landscape of Polyploid Wheat. *Science* 361, 361. doi:10.1126/science.aar6089

Raza, Q., Riaz, A., Atif, R. M., Hussain, B., Ali, Z., and Budak, H. (2020). MADS-box Genes Galore in Wheat Genome: Phylogenomics, Evolution and Stress Associated Functions. *bioRxiv*. doi:10.1101/2020.10.23.351635

Riechmann, J. L., and Meyerowitz, E. M. (1997). MADS Domain Proteins in Plant Development. *Biol. Chem.* 378 (10), 1079–1101. doi:10.1515/bchm.1997.378.10.1079

Rounsley, S. D., Ditta, G. S., and Yanofsky, M. F. (1995). Diverse Roles for MADS Box Genes in Arabidopsis Development. *Plant Cell* 7, 1259–1269. doi:10.1105/tpc.7.8.1259

Ruelens, P., De Maagd, R. A., Proost, S., Theißen, G., Geuten, K., and Kaufmann, K. (2013). FLOWERING LOCUS C in Monocots and the Tandem Origin of Angiosperm-specific MADS-Box Genes. *Nat. Commun.* 4, 2280. doi:10.1038/ncomms3280

Saedler, H., Becker, A., Winter, K. U., Kirchner, C., and Theissen, G. (2001). MADS-box Genes Are Involved in floral Development and Evolution. *Acta Biochim. Pol.* 48, 351–358. doi:10.18388/abp.2001_3920

Samach, A., Onouchi, H., Gold, S. E., Ditta, G. S., Schwarz-Sommer, Z., Yanofsky, M. F., et al. (2000). Distinct Roles of Constans Target Genes in Reproductive Development of Arabidopsis. *Science* 288, 1613–1616. doi:10.1126/science.288.5471.1613

Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., and Melzer, R. (2020). Genome-wide Analysis of MIKC -type MADS -box Genes in Wheat: Pervasive Duplications, Functional Conservation and Putative Neofunctionalization. *New Phytol.* 225, 511–529. doi:10.1111/nph.16122

Sharma, N., Ruelens, P., D'Hauw, M., Maggen, T., Dochy, N., Torfs, S., et al. (2017). A Flowering Locus C Homolog Is a Vernalization-Regulated Repressor in Brachypodium and Is Cold Regulated in Wheat. *Plant Physiol.* 173 (2), 1301–1315. doi:10.1104/pp.16.01161

Shi, S. ya., Zhang, F. fei., Gao, S., and Xiao, K. (2016). Expression Pattern and Function Analyses of the MADS Thranscription Factor Genes in Wheat (*Triticum aestivum* L.) under Phosphorus-Starvation Condition. *J. Integr. Agric.* 15 (8), 1703–1715. doi:10.1016/S2095-3119(15)61167-4

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75

Smaczniak, C., Immink, R. G. H., Angenent, G. C., and Kaufmann, K. (2012). Developmental and Evolutionary Diversity of Plant MADS-Domain Factors: Insights from Recent Studies. *Dev* 139, 3081–3098. doi:10.1242/dev.074674

Wang, Q., Du, M., Wang, S., Liu, L., Xiao, L., Wang, L., et al. (2018). MADS-box Transcription Factor madsA Regulates Dimorphic Transition, Conidiation, and Germination of Talaromyces marneffei. *Front. Microbiol.* 9, 1781. doi:10.3389/fmicb.2018.01781

Wang, Z., Wang, F., Hong, Y., Yao, J., Ren, Z., Shi, H., et al. (2018). The Flowering Repressor SVP Confers Drought Resistance in Arabidopsis by Regulating Abscisic Acid Catabolism. *Mol. Plant* 11, 1184–1197. doi:10.1016/j.molp.2018.06.009

Wei, B., Zhang, R.-Z., Guo, J.-J., Liu, D.-M., Li, A.-L., Fan, R.-C., et al. (2014). Genome-wide Analysis of the MADS-Box Gene Family in Brachypodium Distachyon. *PLoS One* 9, e84781. doi:10.1371/journal.pone.0084781

Whittaker, C., and Dean, C. (2017). The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation. *Annu. Rev. Cel Dev. Biol.* 33, 555–575. doi:10.1146/annurev-cellbio-100616-060546

Wu, Y., Ke, Y., Wen, J., Guo, P., Ran, F., Wang, M., et al. (2018). Evolution and Expression Analyses of the MADS-Box Gene Family in Brassica Napus. *PLoS One* 13, e0200762. doi:10.1371/journal.pone.0200762

Xu, K., He, X., Dreisigacker, S., He, Z., and Singh, P. K. (2020). Anther Extrusion and its Association with Fusarium Head Blight in CIMMYT Wheat Germplasm. *Agronomy* 10, 47. doi:10.3390/agronomy10010047

Yang, C., Liu, H., Li, G., Liu, M., Yun, Y., Wang, C., et al. (2015). The MADS-Box Transcription Factor FgMcm1 Regulates Cell Identity and Fungal Development inFusarium Graminearum. *Environ. Microbiol.* 17, 2762–2776. doi:10.1111/1462-2920.12747

Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The Protein Encoded by the Arabidopsis Homeotic Gene Agamous Resembles Transcription Factors. *Nature* 346, 35–39. doi:10.1038/346035a0

Yu, Y., Ouyang, Y., and Yao, W. (2018). ShinyCircos: An R/Shiny Application for Interactive Creation of Circos Plot. *Bioinformatics* 34, 1229–1231. doi:10.1093/bioinformatics/btx763

Zhang, Z., Li, H., Qin, G., He, C., Li, B., and Tian, S. (2016). The MADS-Box Transcription Factor Bcmads1 Is Required for Growth, Sclerotia Production and Pathogenicity of Botrytis Cinerea. *Sci. Rep.* 6, 33901. doi:10.1038/srep33901

Zhao, Y., Li, X., Chen, W., Peng, X., Cheng, X., Zhu, S., et al. (2011). Whole-genome Survey and Characterization of MADS-Box Gene Family in maize and Sorghum. *Plant Cel Tiss Organ. Cult* 105, 159–173. doi:10.1007/s11240-010-9848-8

Zimin, A. V., Puiu, D., Hall, R., Kingan, S., Clavijo, B. J., and Salzberg, S. L. (2017). The First Near-Complete Assembly of the Hexaploid Bread Wheat Genome, *Triticum aestivum*. *Gigascience* 6, 1–7. doi:10.1093/gigascience/gix097

Conflict of Interest: Author HB is employed by Montana BioAgriculture Incorporation.

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

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 Raza, Riaz, Atif, Hussain, Rana, Ali, Budak and Alaraidh. 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: 24 January 2022
doi: 10.3389/fgene.2021.784883

Comparative Transcriptome Analysis in *Taraxacum koksaghyz* to Identify Genes that Determine Root Volume and Root Length

Annika Wieghaus[1†], *Kai-Uwe Roelfs*[2†], *Richard M. Twyman*[3], *Dirk Prüfer*[1,2] *and Christian Schulze Gronover*[2*]

[1]Institute of Plant Biology and Biotechnology, University of Münster, Münster, Germany, [2]Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Münster, Germany, [3]TRM Ltd., Scarborough, United Kingdom

OPEN ACCESS

Edited by:
Himabindu Kudapa, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Riza Putranto, Indonesian Research Institute for Biotechnology and Bioindustry (IRIBB), Indonesia
Yourong Chai, Southwest University, China
Chathurani Ranathunge, Eastern Virginia Medical School, United States

***Correspondence:**
Christian Schulze Gronover christian.schulze.gronover@ime.fraunhofer.de

[†]*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: *28 September 2021*
Accepted: *16 December 2021*
Published: *24 January 2022*

Citation:
Wieghaus A, Roelfs K-U, Twyman RM, Prüfer D and Schulze Gronover C (2022) Comparative Transcriptome Analysis in Taraxacum koksaghyz *to Identify Genes that Determine Root Volume and Root Length. Front. Genet. 12:784883. doi: 10.3389/fgene.2021.784883*

The Russian dandelion (*Taraxacum koksaghyz*, family Asteraceae) produces large amounts of natural rubber in the laticifers of its roots. This species has been proposed as an alternative source of natural rubber to augment or partly replace the rubber tree (*Hevea brasiliensis*) but domestication would require genetic improvement to increase rubber yields and agronomic optimization to facilitate harvesting and processing. Optimization has focused thus far on the size and shape of the roots, the primary storage organ for natural rubber and inulin. However, the corresponding genetic factors are poorly understood. Here we describe the comparative transcriptomic analysis of root tissues from *T. koksaghyz* plant sets featuring different root sizes and shapes, aiming to identify differentially expressed genes correlating with root length or root diameter in the upper root and root tip. The resulting datasets revealed multiple candidate genes for each trait and root part, including a glucan endo-1,3-β-D-glucosidase, an allene oxide synthase 3, and a TIFY10A/JAZ1 homolog. These three genes were tested by qRT-PCR in outdoor-grown plants with diverse root morphology, and the expression of two genes correlated with the appropriate root morphotype, confirming the effectiveness of our method. We evaluated the candidate genes to gain insight into their potential functions in root development. Such candidate genes could be suitable for marker-assisted breeding programs in the future.

Keywords: ***Taraxacum koksaghyz*, root, natural rubber, transcriptomics, breeding**

INTRODUCTION

The development of new crop varieties and the improvement of existing ones requires the identification and subsequent introduction of new traits such as pest and disease resistance, tolerance of extreme climatic conditions, and higher yields. Although research has focused on food and feed crops to ensure food security, the same objectives mentioned previously apply to crops that serve as feedstock for industrially relevant raw materials. One example is the Russian dandelion (*Taraxacum koksaghyz*), a member of the family Asteraceae originally native to Kazakhstan, which produces large amounts of natural rubber and inulin in its roots and has therefore been proposed as an alternative source of both materials (Krotkov, 1945; Whaley and Bown, 1947; van Beilen and Poirier, 2007a, 2007b).

The Russian dandelion was first cultivated as a crop during World War II when imports of natural rubber produced from the primary source (the rubber tree *Hevea brasiliensis*) were interrupted. But these attempts were abandoned in the early 1950s when the reinstatement of rubber imports made the alternative source economically less attractive (van Beilen and Poirier, 2007b; Schulze Gronover et al., 2011). More recently, interest in alternative sources of natural rubber has returned due to increasing demand and the disadvantages of the rubber tree (Cornish, 2017). *H. brasiliensis* is a slow-growing species with a long breeding cycle, is restricted to specific climate regions, and mainly grows in monoculture plantations with narrow genetic diversity and a high susceptibility to pathogens (Lieberei, 2007). The Russian dandelion produces natural rubber of comparable quality to the rubber tree, with yields of up to 15% of the root dry weight (van Beilen and Poirier, 2007a). Furthermore, up to 50% of the root dry weight is the storage carbohydrate inulin (Uhlmann, 1951). This is used as a sweetener and additive in the food industry and serves as feedstock for bioethanol production (Ohta et al., 1993; Roberfroid, 2005). The current main industrial source of inulin is chicory (Van Den Ende et al., 2004).

The Russian dandelion is a fast-developing annual plant that can produce large amounts of root biomass when cultivated in temperate climate zones (van Beilen and Poirier, 2007b). If large-scale regional cultivation were established, this species could therefore meet a large proportion of the current demand for natural rubber, reducing the vulnerability of supply chains and the carbon footprint of rubber imports from South and Southeast Asia. Several studies have already assessed the performance of *T. koksaghyz* plants in terms of biomass accumulation, rubber and inulin content, and cultivation time and conditions (Arias et al., 2016a, 2016b; Kreuzberger et al., 2016; Eggert et al., 2018). However, the agronomic performance requires further optimization to increase yields and facilitate harvesting and processing in order to make the crop more competitive. One way to achieve these aims is to optimize the root volume (increasing the capacity for rubber and inulin accumulation) and root shape (medium length with more volume in the upper part, to facilitate harvesting and extraction).

Root growth is influenced by internal and external factors, particularly the interactions between phytohormones such as brassinosteroids (Roddick et al., 1993; Clouse et al., 1996), jasmonate (Dathe et al., 1981; Staswick et al., 1992) and auxins (Chadwick and Burg, 1967), as well as various signaling peptides (Yamada and Hsiao, 2021). In *T. koksaghyz*, only the influence of the small signaling peptide TkRALF1 has been investigated thus far (Wieghaus et al., 2019). Little is known about the hormonal response to environmental conditions such as phosphorous and nitrogen availability in the rhizosphere, the root microbiome, or abiotic stresses such as water deprivation. The same applies to the corresponding physiological and morphological adaptations in the roots, thus limiting the value of population-wide analysis of root growth to identify correlations between single genes and phenotypes.

To address this knowledge gap, we analyzed the transcriptome of Russian dandelion plant sets differing in terms of root length and root volume. Massive analysis of cDNA ends (MACE) was chosen for transcriptome generation because this technique is able to detect rare transcripts at significantly lower sequencing depth than classical RNAseq and is not susceptible to bias that originates from different transcript lengths (Zawada et al., 2014; Bojahr et al., 2016).

We used a forward genetics approach to identify genes associated with beneficial root phenotypes suitable for future breeding programs. Here we describe the first candidate genes, and more importantly provide sets of candidate genes related specifically to *T. koksaghyz* root volume or root length.

METHODS

Plant Cultivation and Processing

The plants used in this study were provided by Dr. Fred Eickmeyer, ESKUSA GmbH, and originate from a wild-type population of highly heterozygous individuals which is due to the fact that *T. koksaghyz* is a facultative outcrossing, sexually reproducing species. Seeds were collected from a large set of *T. koksaghyz* plants grown under the greenhouse conditions described below after open pollination. Seeds were sown in rows in large buckets containing a mixture of 70% topsoil (Botanical Garden Münster, Germany) and 30% sand. Seeds started to germinate after 3 days and only seedlings that had emerged on days three or four were used further. The plants were grown in the greenhouse at 22–25°C (day) and 14–18°C (night) with a 16-h photoperiod (20 klx light intensity). A first set of plants was grown for transcriptomic analysis. We harvested the plants after 12 weeks and assigned them to three pools of 21–36 plants based on root morphology: thick roots of medium length (pool A), short and thin roots (pool B), or long and thin roots (pool C) (detailed classification in **Supplementary Table S1**). The roots were measured lengthwise using a ruler and the diameter was determined using digital calipers. Each root was then divided into two parts: proximal (upper root including lateral roots) and distal (root tip up to the first emerging lateral root). The parts were separately frozen in liquid nitrogen, lyophilized and ground to powder for transcriptome analysis.

We then cultivated a second set of wild-type *T. koksaghyz* plants as described above, but the pots were placed outside so that the plants were exposed to natural weather and day length. This second set of plants was used to validate first possible candidate genes gained from the transcriptomic analysis of the first set of wild-type plants. Quantitative real-time PCR (qRT-PCR) was used for the gene expression analysis of individual plants harvested from this second set. The roots were harvested after 12 weeks and measured as described above. We then processed whole roots without division into proximal and distal segments, using the same process of freezing, lyophilization and grinding as described above. The whole-root material was used to provide sufficient amounts of RNA for purification and cDNA synthesis followed by qRT-PCR.

For spatial gene expression analysis, *T. koksaghyz* plants were cultivated for 12 weeks as described above in the greenhouse. The plants were harvested individually and subsequently separated

into flower, peduncle, leaf, root, and latex. The tissue material was then processed as described above.

MACE

Massive analysis of cDNA ends (MACE) libraries (one library per pool) were prepared and sequenced by GenXPro (Frankfurt, Germany) as previously described (Zawada et al., 2014; Bojahr et al., 2016). The sequencing data generated by MACE are deposited in the SRA under the bioproject accession PRJNA783530.

Determination of the Natural Rubber Content

The natural rubber content of the root material used to prepare MACE libraries was quantified by accelerated solvent extraction as previously described (Benninghaus et al., 2020). The natural rubber content was determined gravimetrically by weighing the empty collection vials and those containing dried hexane extracts, then calculating the weight percentage as a proportion of root dry weight using the exact quantity of root material.

Quality Control, Read Mapping and Annotation

To increase mapping accuracy, low-quality bases (Phred score <20) were trimmed from the 3′ end of all reads using *cutadapt* v1.16 (Martin, 2011). Given the nature of the MACE technique, reads originate from 50 to 800 bp upstream of the 3′ ends of the transcripts, which often results in reads containing poly(A) tail sequences. Therefore, following the previous step, a *de novo* script was used to detect and remove poly(A) tail sequences, which would otherwise interfere with mapping. Finally, reads <25 bp were discarded using *cutadapt* to prevent nonspecific mapping. Quality-trimmed reads were mapped to the reference genome (Lin et al., 2017) using *segemehl* (Hoffmann et al., 2009) with split reads included.

Differential Expression Analysis

Raw read counts were generated using *bamutils* v0.5.9 (Breese and Liu, 2013) with partially counted multiple hits, which represented the basis for differential expression analysis using the MA plot method with random sampling (MARS) in the *DEGSeq* R package v1.32.0 (Wang et al., 2009). Transcripts were annotated by using the NCBI BLAST suite v2.7.1 (Altschul et al., 1990) to screen the NCBI non-redundant proteins (O'Leary et al., 2016) and UniProtKB/SWISS-PROT (Bateman et al., 2017) databases. Transcripts with a normalized one-fold or greater differential expression (log-2-fold change ≥ 1 or ≤ -1, respectively) combined with a p-value less than 0.001 ($p < 0.001$) were considered significant and the corresponding genes were defined as differentially expressed genes (DEGs). Candidate genes for root length and diameter were selected according to following conditions: transcripts that were differentially expressed in pools A vs B and B vs C but not A vs C were considered as candidates for root length, whereas those differentially expressed in pools A vs B and A vs C but not B vs C were considered as candidates for root diameter.

Data Analysis

Multi-dimensional scaling (MDS) plot analysis was performed using limma R package v3.46.0 (Ritchie et al., 2015). Venn diagrams were prepared using Vennplex (https://www.nia.nih.gov/research/labs/vennplex; Cai et al., 2013). Heatmaps showing differential gene expression were generated using heatmapper (http://www.heatmapper.ca; Babicki et al., 2016). Gene Ontology (GO) enrichment was analyzed using WEGO (http://wego.genomics.org.cn; Ye et al., 2018, 2006) with all transcripts in the corresponding root part used as the background. Glycosylphosphatidylinositol (GPI) anchors, protein domains and signal peptides were predicted using PredGPI (Pierleoni et al., 2008), CD search (Lu et al., 2020) and SignalP v5.0 (Almagro Armenteros et al., 2019), respectively. Phylogenetic trees were constructed using MEGA X (Kumar et al., 2018) and graphically rendered using iTOL v6.3.1 (Letunic and Bork, 2021).

Extraction of Total RNA and cDNA Synthesis

Total RNA was extracted from *T. koksaghyz* root material using the innuPREP RNA Mini Kit (Analytik Jena, Jena, Germany) according to the manufacturer's instructions, and cDNA was synthesized using PrimeScript RT Master Mix (TAKARA, Clontech, Sain-Germain-en-Laye, France) according to the manufacturer's instructions.

Gene Expression Analysis by qRT-PCR

Quantitative real-time PCR was carried out as previously described (Laibach et al., 2015) with slight modifications. The elongation factor 1α (*TkEF1α*) and ribosomal protein L27 (*TkRP*) reference genes were used for the normalization of gene expression as described before (Niephaus et al., 2019; Benninghaus et al., 2020). The qRT-PCR data were analyzed, and normalized expression levels were calculated using Bio-Rad CFX Manager v3.1 (Bio-Rad Laboratories, Hercules, CA, United States). All oligonucleotides used for qRT-PCR analysis are listed in **Supplementary Table S2** along with corresponding primer efficiencies. For determination of primer efficiencies, a standard curve using a serial dilution of a representative cDNA sample was generated and subsequently analyzed using Bio-Rad CFX Manager v3.1. Efficiencies are calculated by the software as described by Pfaffl (2001) and Vandesompele et al. (2002), converted to percentage efficiency.

Statistical Analysis

Data analysis using the two-tailed *t*-test, Kolmogorov-Smirnov normal distribution analysis, Grubbs's test for the identification of outliers, and correlation studies (linear fit with calculation of Pearson's correlation coefficient) was carried out using OriginPro 2019b (OriginLab Corporation, Northampton, MA, United States).

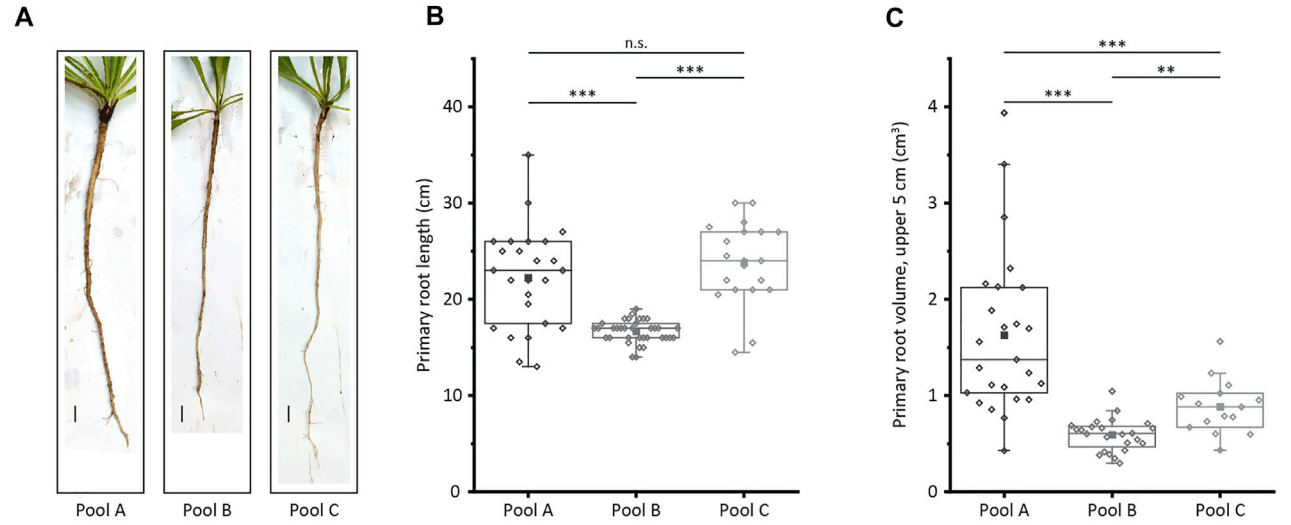


FIGURE 1 | Characterization of the three plant pools selected for MACE. The primary root length and the volume of the upper 5 cm of the root in 12-week-old *T. koksaghyz* wild-type plants were determined after harvesting. The plants were then assigned to three plant pools **(A)** Pictures of plants representing the three pools. Scale bar = 1 cm **(B)** Primary root length **(C)** Volume of the upper 5 cm of the primary root, determined by measuring the root diameter immediately below the leaf rosette and 5 cm further down. The boxes delimit values from the 25th to the 75th percentile of each dataset, the horizontal line in the box represents the median, the filled square represents the mean, and the lower and upper whiskers represent values that differ least from 25th percentile –1.5*IQR or 75th percentile +1.5*IQR, respectively (n = 15–35 plants, normal distribution confirmed using the Kolmogorov-Smirnov test, statistical significance confirmed using a two-tailed *t*-test; n. s = not significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$).

RESULTS

Transcriptome Data Generation and Processing

To identify genes that potentially affect *T. koksaghyz* root morphology, we generated a set of wild-type plants and subdivided it into three pools of plants with different root characteristics. The root morphology of the pools differed significantly in terms of primary root length, diameter directly below the leaf rosette, and the volume of the upper 5 cm, which was calculated by determining the root volume at two defined points (**Figure 1**, **Supplementary Table S1**). However, all three pools produced comparable amounts of natural rubber (**Supplementary Table S1**). Pool A comprised plants with medium to long roots that were also thick (large volume), representing the most desirable root phenotype for biomass accumulation as well as harvesting and handling. Pool B comprised plants with short and thin roots, whereas pool C comprised plants with long and thin roots.

We compared gene expression profiles between pools in two segments of the root. Each complete root was divided into the proximal segment (upper root including lateral roots) and distal segment (root tip up to the first emerging lateral root), resulting in a total of six pooled samples. *T. koksaghyz* is a sexually reproducing and self-incompatible species, so wild-type plants show a high degree of genetic diversity. By pooling plants with similar root morphology prior to transcriptome analysis, we aimed to reduce the background noise and emphasize transcripts that are differentially expressed between pools, thus capturing those representing genes related to the development of specific root characteristics.

We prepared and sequenced cDNA libraries from all six samples using MACE technology (Kahl et al., 2012). The six libraries contained between 4.65×10^6 and 6.68×10^6 single-end raw reads of 16–145 bp in length, which were further processed using a custom Python script to remove poly(A) tails. Following the removal of adapters and the trimming of low-quality bases, the libraries contained between 4.55×10^6 and 6.48×10^6 clean reads, 86.3–87.3% of which mapped to the reference genome (Lin et al., 2017), and 74.3–78.5% of which were uniquely matched (**Table 1**). We obtained 30,955 unique transcripts from the six libraries, which were then annotated using the NCBI and UniProt databases. To confirm sufficient reduction of background noise by pooling, we constructed an MDS plot, which showed the clear separation of all samples (**Supplementary Figure S1**).

Identification Root Volume and Root Length Candidate Genes by Differential Gene Expression Analysis

Next, we identified the genes with significant differential expression between pools, defined as those with a log2 fold change (l2fc) ≥ 1 or ≤ –1 ($p < 0.001$) in any pairwise comparison. We then constructed a Venn diagram and used this to find genes related to either root volume or root length (**Figure 2A**).

Genes with significant differential expression when comparing pool A vs B and pool A vs C were considered as candidates influencing root volume (RV in **Figure 2A**) because the volume of

TABLE 1 | Summary of MACE alignment statistics in six libraries mapped to the reference genome.

	Pool A		Pool B		Pool C	
	Upper root	**Root tip**	**Upper root**	**Root tip**	**Upper root**	**Root tip**
Total base pairs	535,992,378	444,400,123	600,607,725	754,363,862	508,730,146	777,887,360
Total clean reads	5,288,641	4,551,524	4,941,227	6,265,938	4,807,513	6,479,601
Mapped reads	4,619,116 (87.34%)	3,956,571 (86.93%)	4,293,960 (86.90%)	5,415,759 (86.43%)	4,179,649 (86.94%)	5,593,472 (86.32%)
Unique matches among mapped reads	3,608,474 (78.12%)	2,941,045 (74.33%)	3,370,356 (78.49%)	4,221,179 (77.94%)	3,202,090 (76.61%)	4,302,574 (76.92%)
Multi-position among mapped reads	750,024 (16.24%)	798,720 (20.19%)	654,546 (15.24%)	856,771 (15.82%)	745,256 (17.83%)	928,985 (16.61%)
Split-reads among mapped reads	260,618 (5.64%)	216,806 (5.48%)	269,058 (6.27%)	337,809 (6.24%)	232,303 (5.56%)	361,913 (6.47%)
Unmapped reads	669,525 (12.66%)	594,953 (13.07%)	647,267 (13.10%)	850,179 (13.57%)	627,864 (13.06%)	886,129 (13.68%)

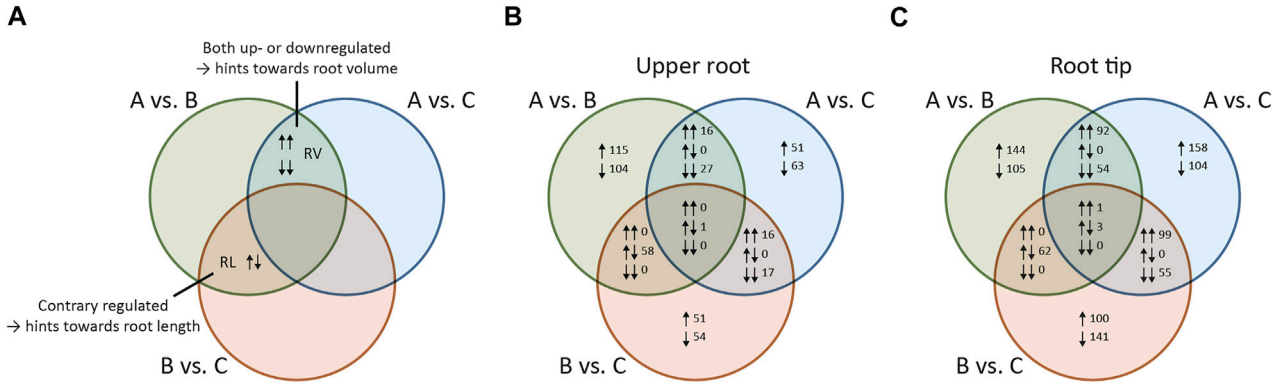


FIGURE 2 | Venn diagram of differentially expressed genes (DEGs) identified by pairwise comparison of plant pools differing in root morphology. All pairwise comparisons between the pools were set in relation to each other based on log2 fold change values in order to identify shared DEGs **(A)** Schematic Venn diagram illustrating the identification of DEGs related to primary root volume or primary root length **(B)** Venn diagram showing DEGs identified by comparison among the pools in the upper root **(C)** Venn diagram showing DEGs identified by comparison among the pools in the root tip. Pool A = plants with medium to long and thick roots (high value for primary root length and volume in the upper 5 cm of the primary root). Pool B = plants with short, thin roots. Pool C = Plants with long, thin roots. RV = DEGs influencing primary root volume; R = DEGs influencing primary root length; ↑ = DEGs with positive log2 fold changes (more abundant in the first pool in each comparison); ↓ = DEGs with negative log2 fold changes (more abundant in the second pool in each comparison); ↑↑ = DEGs with positive log2 fold changes for all comparisons (more abundant in the first pool of each comparison); ↓↓ = DEGs with negative log2 fold changes for all comparisons (more abundant in the second pool of each comparison); ↑↓ = DEGs with opposing log2 fold change values (positive in one comparison but negative in the other).

the roots in pool A was significantly higher compared to both pool B and pool C. The initial list of candidates was reduced by excluding those also showing differential expression when comparing pool B vs C because the roots in these pools were similar in volume. By applying this last criterion, we arrived at a final set of RV candidate genes (schematic description in **Supplementary Figure S2**). Genes with significant differential expression when comparing pool A vs B and pool B vs C were considered as candidates influencing root length (RL in **Figure 2A**) because the roots in pools A and C were significantly longer than the roots in pool B. As above, we narrowed the initial list of candidates by excluding those also showing differential expression between pools A and C because the roots in these pools were similar in length. By applying this last criterion, we arrived at a final set of RL candidate genes (schematic description in **Supplementary Figure S2**). The Venn diagrams showed that DEGs defining the RV intersection were either upregulated or downregulated in both comparisons (A vs B *and* A vs C), but no counter-regulated genes were observed. In contrast, all DEGs in the RL intersection were counter-regulated (either upregulated in A vs B but downregulated in B vs C or vice versa).

We identified 573 DEGs for the upper root segment. We extracted 40 RV candidate genes from the 43 DEGs in the RV intersection by manually excluding genes showing differential expression when comparing pool B vs C with $p > 0.001$, because we considered these l2fc values as not statistically robust (**Figure 2B**). The corresponding heat map revealed that 33% of the RV candidate genes were expressed more strongly in roots with a higher volume (pool A) compared to those with a lower volume (pools B and C) as indicated by the positive l2fc values (**Figure 3A**). We also extracted 55 RL candidate genes from the 58 DEGs in the RL intersection (**Figure 2B**), 67% of which were expressed at higher levels in long roots (pools A and C) compared to short roots (pool B) (**Figure 3C**).

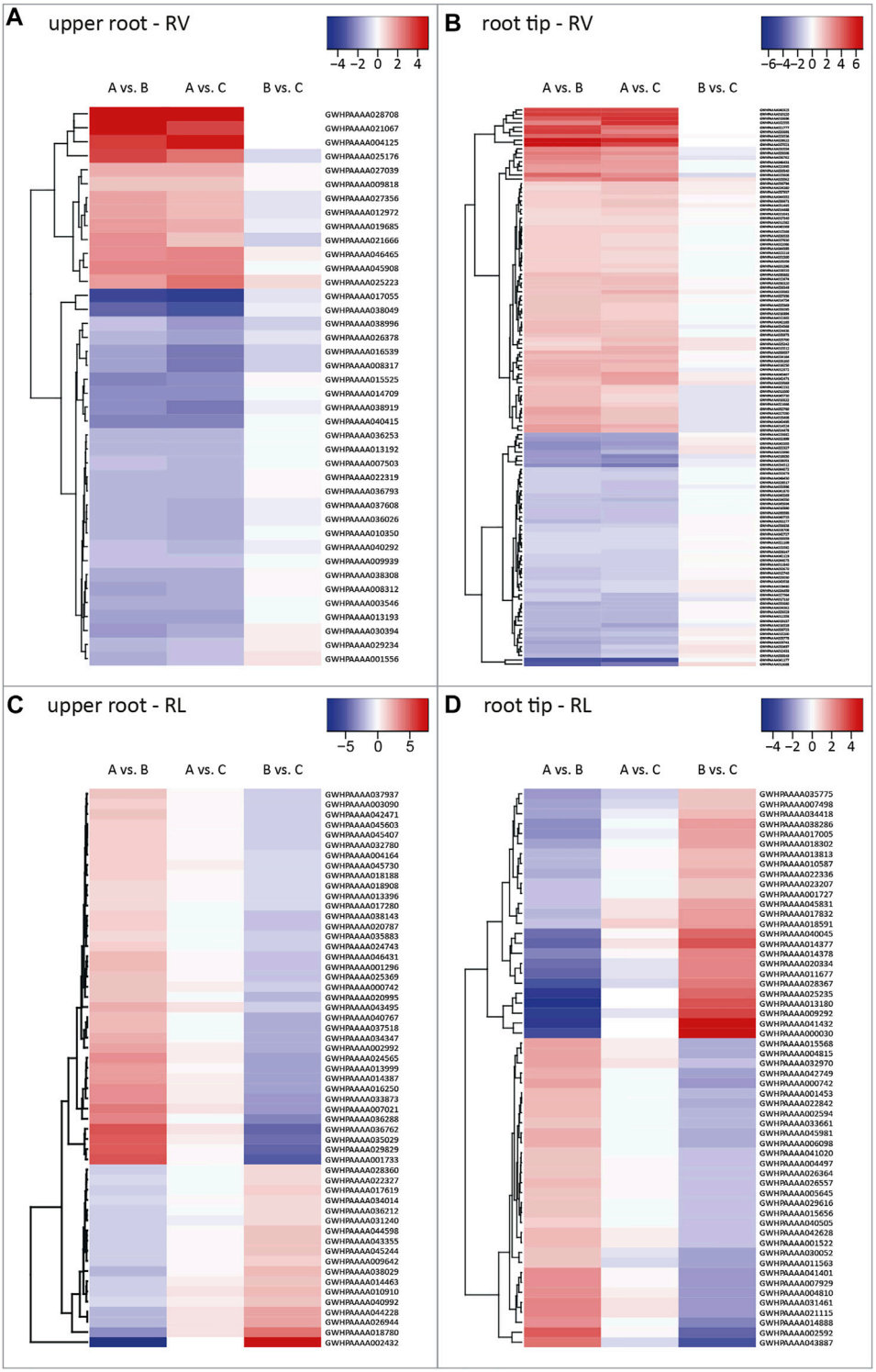


FIGURE 3 | Heat maps of root volume (RV) and root length (RL) candidate genes. Heat maps visualizing all RV and RL candidate genes identified in different root parts with their corresponding log2 fold change values for pairwise comparisons among the three plant pools. Each row depicts the log2 fold change values of the comparisons with blue, white and red boxes representing negative, close-to-zero and positive values, as indicated by the corresponding scale bars. Shown are RV candidate genes identified in **(A)** the upper root and **(B)** the root tip, as well as RL candidate genes identified in **(C)** the upper root, and **(D)** the root tip. Each row shows log2 fold change values of pairwise comparisons between the plant pools. Pool A = plants with medium to long and thick roots (high value for primary root length and volume in the upper 5 cm of the primary root). Pool B = plants with short, thin roots. Pool C = Plants with long, thin roots. RV = DEGs influencing primary root volume. RL = DEGs influencing primary root length.

TABLE 2 | Summary of the GO enrichment analysis at deeper GO levels.

Candidate gene set	GO term (↑/↓)	GO class and level
upper root - RV candidate genes	hydrolase activity (↑)	molecular function, level 3
root tip - RV candidate genes	intrinsic component of membrane (↓)	cellular component, level 3/4
	integral component of membrane (↓)	cellular component, level 4
	nuclear part (↓)	cellular component, level 4
	response to endogenous stimulus (↑)	biological process, level 3
	cellular component organization (↓)	biological process, level 3
	regulation of metabolic process (↓)	biological process, level 3/4
	response to chemical (↑)	biological process, level 3
	cellular metabolic process (↓)	biological process, level 3
	primary metabolic process (↓)	biological process, level 3
	response to acid chemical (↑)	biological process, level 4
	response to hormone (↑)	biological process, level 4
	drug metabolic process (↑)	biological process, level 4
	cellular macromolecule metabolic process (↓)	biological process, level 4
	response to oxygen-containing compound (↑)	biological process, level 4
upper root - RL candidate genes	non-membrane-bounded organelle (↑)	cellular component, level 3
	membrane-bounded organelle (↓)	cellular component, level 3
	intracellular non-membrane-bounded organelle (↑)	cellular component, level 4
	intracellular membrane-bounded organelle (↓)	cellular component, level 4
	small molecule metabolic process (↑)	biological process, level 3
root tip - RL candidate genes	cell periphery (↑)	cellular component, level 3/4

↑, *enriched in candidate genes compared to all transcripts;* ↓, *underrepresented in candidate genes compared to all transcripts.*

We identified 1,118 DEGs for the root tip, including 129 RV candidate genes from 146 DEGs assigned to the RV intersection and 56 RL candidate genes from 62 DEGs assigned to the RL intersection (**Figure 2C**). We found that 58% of the RV candidate genes were expressed at higher levels in roots with a larger volume (**Figure 3B**), and 55% of the RL candidate genes were expressed at higher levels in longer roots (**Figure 3D**). All DEGs and candidate genes are listed in **Supplementary Table S3**.

GO Term Enrichment of the RV and RL Candidate Genes

GO term enrichment analysis was carried out to determine the putative functions of the RV and RL candidate genes. Overall, we observed a similar distribution of second-level GO terms for all sets of candidate genes in the three main GO categories: cellular component, molecular function and biological process (**Supplementary Figure S3**). The most abundant GO terms assigned in the cellular component category were *cell*, *cell part*, *organelle*, *organelle part*, and *membrane*. For the molecular function category, the most represented GO terms were *catalytic activity* and *binding*. For the biological process category, the most represented GO terms were *cellular process*, *metabolic process*, *response to stimulus*, *biological regulation*, and *regulation of biological process*. Several GO terms were significantly enriched or underrepresented at level 2 (**Supplementary Figure S3**). Among the RV candidate genes in the root tip dataset, the cellular component *extracellular region* was enriched compared to all transcripts, whereas multiple terms (*cell*, *cell part*, *membrane*, and *membrane part* in the cellular component category; *cellular process*, *metabolic process*, *biological regulation*, and *cellular component organization or biogenesis* in the biological process category) were underrepresented compared to all transcripts. In addition to this general characterization, more specific GO terms at deeper levels were also evaluated by enrichment analysis (**Table 2**). Compared to all transcripts, *hydrolase activity* was enriched in the RV candidate genes of the upper root. Several GO terms describing response to different stimuli (chemical, hormone, oxygen-containing compound) were enriched in the RV candidate gene set in the root volume. Among others, *intrinsic component of membrane*, *integral component of membrane*, *regulation of metabolic process*, *cellular metabolic process*, and *cellular macromolecule metabolic process* were underrepresented in this candidate gene set. For the GO term distribution in the RL candidate genes of the upper root we found components not associated with membrane-bound organelles to be enriched (*non-membrane-bounded organelle*, *intracellular non-membrane-bounded organelle*) and vice versa (*membrane-bounded organelle*, *intracellular membrane-bounded organelle*). *Cell periphery* was enriched in the RL candidate genes of the root tip.

Transcriptome Characterization

To characterize our transcriptome dataset further, we searched for typical factors affecting root growth and development like transcription factors, hormone pathway genes and cell wall/glucan biosynthesis/degradation genes based on the UniProt annotation (**Supplementary Table S4–S6**).

We identified 10 differentially expressed transcription factors which mostly originate from the volume and length trait groups in the root tip. Members of this set include two basic helix-loop-helix transcription factors (Transcription factor bHLH62, GWHPAAAA007503; Transcription factor bHLH145, GWHPAAAA042185), a transcription factor of the TCP

family (Transcription factor TCP8, GWHPAAAA044352), an AP2 domain containing or RAP2 (related to AP2) family protein (Ethylene-responsive transcription factor RAP2-4, GWHPAAAA002086) and Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10 (GWHPAAAA017340) of the ALOG family. Other members of the ALOG family were proposed to be developmental regulators because they promote cell growth in *Arabidopsis thaliana* (Zhao et al., 2004; Cho and Zambryski, 2011). This might indicate that the member we identified in our dataset plays a role during root development in *T. koksagyhz*.

We also found 16 genes associated with or induced by certain hormone pathways, including brassinosteroids (Protein EXORDIUM, GWHPAAAA001556 and GWHPAAAA007090; Shaggy-related protein kinase eta, GWHPAAAA008535), auxin (Auxin response factor 2A, GWHPAAAA014478; Remorin, GWHPAAAA014754) and jasmonates (Protein TIFY 10A, GWHPAAAA022842; Protein SGT1 homolog B, GWHPAAAA001300; Allene oxide synthase 3, GWHPAAAA036026). Most genes were found in the volume trait groups of the upper root and root tip. Protein EXORDIUM was shown to promote growth in the root and shoot when overexpressed in *A. thaliana* (Coll-Garcia et al., 2004) and Auxin response factor 2A downregulation increased root branching in tomato (Hao et al., 2015).

The last group comprised 11 genes encoding cell wall/glucan biosynthesis/degradation factors. Members of this group included not only glucan endo-1,3-β-D-glucosidases (β-1,3-glucanases; GWHPAAAA003778, GWHPAAAA042727, GWHPAAAA038049, GWHPAAAA027039) and xyloglucan endotransglucosylases (GWHPAAAA046431, GWHPAAAA046431, GWHPAAAA013192, GWHPAAAA013193), but also a cellulose synthase A catalytic subunit 6 (UDP-forming)-like gene (GWHPAAAA015656). The latter was previously found to play a crucial role in root and hypocotyl elongation in *A. thaliana* (Hauser et al., 1995; Fagard et al., 2000) and is also associated with the root length trait of the root tip in our dataset.

Verification of Candidate Genes by qRT-PCR

To confirm the suitability of our approach for root transcriptome analysis, we generated an independent set of wild-type *T. koksaghyz* plants (**Supplementary Table S7**). These plants were grown outside under natural climate conditions to approximate field growth. The roots were harvested individually after 12 weeks. Subsequently, qRT-PCR analysis was performed using primers specific for a small set of genes from the previous step, which were not obviously associated with root growth, as our goal was to identify novel factors influencing root morphology. They were also selected to represent various types of developmental and cellular mechanisms.

GWHPAAAA042727 was differentially expressed in the root tip and was annotated as a glucan endo-1,3-β-D-glucosidase (β-1,3-glucanase; UniProt ID Q94G86). This gene was strongly expressed in thin roots (pools B and C) but only weakly expressed in thick roots (pool A) and therefore represented a putative negative regulator of root volume. This was confirmed by qRT-PCR (**Figure 4A**). Additionally, spatial gene expression analysis showed the highest expression in roots (**Supplementary Figure S4A**). Protein domain analysis revealed a glycoside hydrolase family 17 (GH-17) domain spanning amino acids 23–343 and an X8/CBM43 superfamily domain spanning amino acids 372–456. The first part of the N-terminus corresponded to a signal peptide with a cleavage site between residues 21 and 22, but no GPI anchor site was predicted, suggesting the protein is not attached to the membrane but probably binds to carbohydrates through its X8/CBM43 domain (Barral et al., 2005). Phylogenetic analysis (**Supplementary Figure S5**) revealed that this β-1,3-glucanase belongs to the α-clade characterized by a diverse set of protein domain structures and expression patterns (Doxey et al., 2007). Like approximately one third of the β-1,3-glucanases in this clade, the *T. koksaghyz* protein represents protein domain architecture type II, consisting of a signal peptide, a GH-17 domain and an X8/CBM43 domain.

GWHPAAAA036026 was differentially expressed in the proximal root and was annotated as allene oxide synthase 3 (TkAOS3; UniProt ID Q8GZP5). This gene was also expressed at higher levels in pools B and C compared to pool A and thus represented another putative negative regulator of root volume. Amino acid sequence analysis showed that it shares the same characteristics as other CYP74 family P450 cytochromes, lacking the threonine residue required for the binding and activation of oxygen in the I-helix of P450 monooxygenases (Chapple, 1998). Instead, the small hydrophobic residue valine is found at this position, which is common for CYP74 proteins (Itoh et al., 2002). TkAOS3 also fulfils the CYP74 consensus sequence for cysteinyl heme ligand residues near the C-terminus with the sequence NKQC (A/**P**)(G/**A**)K(**D**/N)XV (Itoh et al., 2002). An unrooted maximum-likelihood tree (**Supplementary Figure S6**) containing several members of the CYP74 family from different species was constructed to assign TkAOS3 to one of the four CYP74 subfamilies. TkAOS3 clustered with other members of CYP74 subfamily C, which show high affinity for 9-hydroperoxides of linolenic or linoleic acid (9-HPOT/9-HPOD) derived from 9-lipoxygenase (9-LOX) activity, yielding the corresponding allene oxides (9,10-EOD/9,10-EOT) (Itoh et al., 2002; Stumpe et al., 2006). These unstable intermediates are likely to undergo spontaneous cyclization (forming 10-OPDA/10-OPEA) or hydrolysis (forming α-ketol/γ-ketol). However, the tomato AOS3 gene (*LeAOS3*) is thought to encode a multifunctional enzyme that can also catalyze the cyclization or hydrolysis of the intermediate to form *rac-cis*-10-OPEA or *9R*-α-ketol, respectively (Grechkin et al., 2008). The expression of AOS3 is restricted to the roots in tomato (Itoh et al., 2002) and to the sprouting eyes, stolons, tubers and roots in potato (Stumpe et al., 2006). Our qRT-PCR analysis did not find any correlation because the expression of *GWHPAAAA036026* did not negatively correlate with root volume (**Figure 4B**) as initially suggested by the MACE dataset. Among the different plant tissues, we observed the highest gene expression in latex, which is highly abundant in

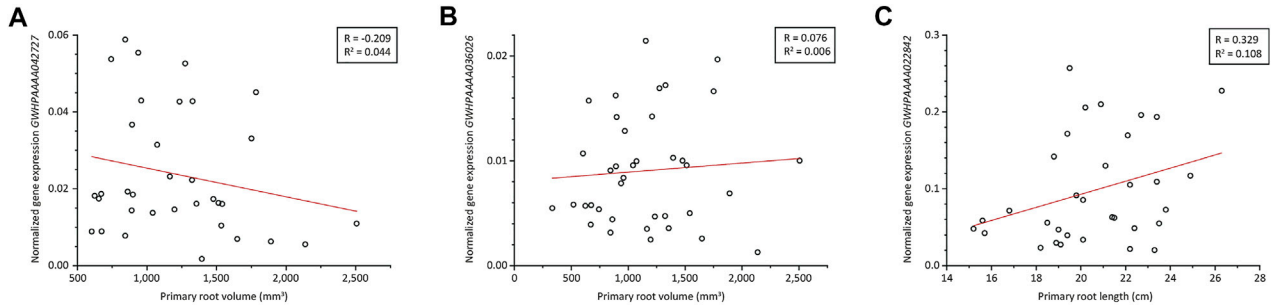


FIGURE 4 | Expression analysis of three candidate genes identified in the *T. koksaghyz* root transcriptome. Gene expression in single *T. koksaghyz* wild-type plants was determined by qRT-PCR. Normalized gene expression of *GWHPAAAA042727* **(A)** and *GWHPAAAA036026* **(B)** was plotted against the primary root volume in the upper 5 cm, and normalized gene expression of *GWHPAAAA022842* **(C)** was plotted against the primary root length. Normal distribution of all parameters was confirmed using the Kolmogorov-Smirnoff-test, and outliers were identified using Grubbs's test. Red line = regression line; R = Pearson's R correlation coefficient; R^2 = coefficient of determination; n = 33–38.

root tissue. In all other examined tissues, *GWHPAAAA036026* is only weakly expressed in comparison to latex (**Supplementary Figure S4B**).

GWHPAAAA022842 was differentially expressed in the root tip and was annotated as a homolog of TIFY10A/JAZ1 (UniProt ID Q9LMA8). This gene was expressed at significantly higher levels in pools A and C compared to pool B, and was thus considered a candidate for a positive regulator of root length. Our qRT-PCR analysis in single plants revealed a positive correlation with the primary root length (**Figure 4C**), consistent with the MACE dataset. *GWHPAAAA022842* is expressed in almost all tissues except flowers at a moderate level, and the highest expression was observed in the peduncle (**Supplementary Figure S4C**). The protein contains a TIFY domain (spanning residues 106–138) with a TIFLNG motif differing slightly from the consensus sequence [TIF (F/Y)XG] (Vanholme et al., 2007). It also features a Jas/CCT domain (spanning residues 182–207). A maximum likelihood tree was constructed from *A. thaliana* TIFY/JAZ proteins, revealing that our protein is closely related to AtTIFY10A and AtTIFY10B (**Supplementary Figure S7**). AtTIFY10A/JAZ1 belongs to the JAZ family of repressors that target transcription factors such as MYC2, MYC3 and MYC4 (Chini et al., 2007; Fernández-Calvo et al., 2011). These transcription factors promote the expression of several early jasmonate response genes and are involved in defense responses as well as developmental processes such as lateral root formation (Yadav et al., 2005; Grunewald et al., 2009; Kazan and Manners, 2013). The degradation of JAZ proteins depends on jasmonyl-isoleucine (JA-Ile)-mediated interactions with COI1 through its Jas domain, which is part of the SCF(COI1) E3 ubiquitin ligase complex, and subsequent processing by the 26S proteasome (Thines et al., 2007). The TIFY domain of JAZ1 interacts with NINJA/AFPH2, which recruits TOPLESS to repress MYC2 (Pauwels et al., 2010). In turn, the expression of multiple JAZ genes (including *JAZ1*) is promoted by MYC2 as well as the closely related transcription factors MYC3 and MYC4, resulting in a negative feedback loop to attenuate the jasmonate signal (Chini et al., 2007; Niu et al., 2011).

DISCUSSION

Domestication of the Russian dandelion (*T. koksagyhz*) would provide a sustainable source of natural rubber. However, it is first necessary to develop breeding programs to optimize rubber yields as well as the ease of harvesting and processing. Marker-assisted selection is a powerful method to improve precision and efficiency during plant breeding. Numerous quantitative trait loci (QTLs) and/or associated genes for a variety of plant species and traits have been identified to assist breeders in plant selection as early as the seedling stage, for example in rice (Yonemaru et al., 2010). Suitable markers have not yet been identified in *T. koksagyhz*. Most studies thus far have focused on dandelion metabolism (Schmidt et al., 2010; Epping et al., 2015; Pütter et al., 2017; Stolze et al., 2017; Niephaus et al., 2019) whereas few have considered growth characteristics (Munt et al., 2012; Wieghaus et al., 2019) or field trials (Arias et al., 2016a; Kreuzberger et al., 2016; Eggert et al., 2018; Keener et al., 2018). Most of the relevant studies have aimed to improve yields or gain knowledge about agronomic performance.

To facilitate the breeding of new dandelion varieties with improved traits, we developed a MACE-based transcriptomic approach to identify candidate genes involved in root growth. Three sets of plants with diverse root morphotypes were harvested and the roots were divided into upper (proximal) and root tip (distal) segments. The primary root volume and root length differed significantly between the morphotypes, whereas the rubber content was consistent. By pooling the plants, we increased the likelihood of identifying specific transcripts associated with root morphology in a heterogenous set of wild-type plants while minimizing differences in the genetic background. This reduced the number of DEGs we recovered, but ensured that they were associated with a given root phenotype. The analysis of differential gene expression in the resulting datasets produced four sets of candidate genes influencing primary root volume or length in the upper root or root tip.

Our transcriptomic pipeline allowed the identification of multiple genes for specific root traits. Several of these were assigned to transcription factors, hormone pathway/responsive

factors and cell wall/glucan biosynthesis/degradation proteins. We found that genes in these groups, encoding products such as Protein EXORDIUM and cellulose synthase A catalytic subunit 6 (UDP-forming) are already known to play a role during root development, but also highlighted several interesting genes which are yet to be evaluated for a putative role in root growth and morphogenesis.

We selected three candidate genes for verification in single *T. koksagyhz* plants from a new set of wild-type plants grown under simulated field conditions. Because of the heterogeneity of the plants used, we did not expect the correlation to be very high. Many factors determine the root phenotype and our candidate genes might contribute to it. Therefore, we think that a weak correlation might still be reasonable when combined with other identified DEGs. Nevertheless, two of the candidate genes, encoding proteins resembling β-1,3-glucanase 11 (TkBG11) and jasmonate ZIM-domain protein 1 (TkTIFY10A/JAZ1), were confirmed by qRT-PCR as more strongly expressed in thinner and longer roots, respectively. In contrast, there was no correlation in the expression of TkAOS3, encoding a putative allene oxide synthase.

The TkBG11 protein may affect the architecture of plasmodesmata during root development, influencing the way all plant cells connect to their neighbors. The permeability of these membrane-lined microscopic channels is controlled by interactions between callose synthases and β-1,3-glucanases, which regulate the accumulation of callose in the surrounding cell walls and thus determine the aperture size (Wolf et al., 1991; Levy et al., 2007; Vatén et al., 2011; De Storme and Geelen, 2014). Plasmodesmata play an important role in cell–cell communication via the symplastic transport of auxin (Liu et al., 2017; Mellor et al., 2020), transcription factors such as SHORT ROOT (Nakajima et al., 2001; Kim et al., 2002, 2005; Carlsbecker et al., 2010; Vatén et al., 2011), LEAFY (Wu et al., 2003), KNOTTED1 (Lucas et al., 1995) and PLETHORA 2 (Mähönen et al., 2014), and microRNAs such as miRNA165/6 (Carlsbecker et al., 2010; Vatén et al., 2011). For example, the quiescent center in the meristematic zone of the root tip maintains a single layer of immediately adjacent stem cells by signaling through the plasmodesmata (Barclay et al., 1982; Van Den Berg et al., 1997; Stahl et al., 2013; Liu et al., 2017). Likewise, lateral root development in *A. thaliana* depends on symplastic connectivity (Benitez-Alfonso et al., 2013). Several pathogen-related β-1,3-glucanases of the PR-2 family not only play a role in defense by degrading fungal cell walls (Thanseem et al., 2005; Liu et al., 2009) but also by inhibiting the spread of viruses through plasmodesmata (Zavaliev et al., 2013). Following the hypothesis that TkBG11 plays a role in callose homeostasis in root cell plasmodesmata, its impact on root development may reflect the depletion of callose in the plasmodesmata to increase the trafficking of hormones and/or miRNA.

TkTIFY10A/JAZ1 is likely to play a role during root elongation because it was strongly expressed in long roots and this was confirmed by qRT-PCR in plants grown under field conditions. In *A. thaliana*, the knockdown of JAZ1 sensitized the plants to methyl jasmonate (MeJA), which inhibited primary root growth but increased the number of lateral roots (Grunewald et al., 2009). In contrast, the overexpression of JAZ1 or knockout of COI1 increased the hypocotyl length under far-red light (Robson et al., 2010; Liu and Wang, 2020). Furthermore, JAZ1 is upregulated by auxin (Grunewald et al., 2009). These findings indicate a similar role for JAZ1 during root development in *T. koksaghyz*, because stronger expression was observed in plants with significantly longer roots. However, it is unclear whether a jasmonate-mediated response to pathogens resulted in shorter roots in pool B.

Although qRT-PCR did not confirm the correlation between primary root length and the expression of TkAOS3 as suggested by the transcriptomic data, the gene may still influence root development in the context of defense against pathogens, and this may explain the discrepancy between our two sets of plants (one grown in the greenhouse and the other exposed to pathogens in the environment). This hypothesis is also supported by its high expression in latex. The products of the 9-LOX pathway have been proposed to mediate defense responses in plants (Vellosillo et al., 2007; Gao et al., 2008; León Morcillo et al., 2012; Vicente et al., 2012; Marcos et al., 2015; Christensen et al., 2016; Morcillo et al., 2016). The strong expression of TkAOS3 therefore hints at a thin-root phenotype resulting from interactions with pathogens during root development. Interestingly, *LeAOS3* is not detected in the roots of *jai1/myc2* tomato mutants (Itoh et al., 2002). MYC2 regulates several early jasmonate response genes as a component of plant defenses (Yadav et al., 2005; Grunewald et al., 2009; Kazan and Manners, 2013), which indicates a link between the 9-LOX pathway and jasmonate defense system (Itoh et al., 2002). Given that our TIFY10A/JAZ1 homolog, which is more strongly expressed in long roots, may repress MYC2 (Chini et al., 2007; Fernández-Calvo et al., 2011) and that our AOS3 homolog is more strongly expressed in thin roots, the expression profile is at least confirmed for pools A and B. But because pool C is characterized by the stronger expression of both the AOS3 and TIFY10A/JAZ1 homologs, there appears to be a more complex relationship between these pathways.

CONCLUSION

We have provided initial insight into the expression of genes in the upper root and root tip of *T. koksaghyz* plants by identifying genes that are differentially expressed between diverse root phenotypes. Several candidate genes were identified and the expression of three of them was tested by qRT-PCR in a new set of wild-type plants grown under field conditions, correlating with the transcriptomic data in two cases and confirming the validity of our method. We also generated a set of candidate genes suitable for further research and subsequent use in marker-assisted breeding. The results strengthen our strategy of first identifying bulk DEGs by comparing pools of plants with similar phenotypes, followed by a verification on a separate set of plants. The latter step reduces the overall number of candidate genes but homes in on the most promising candidates. In this case, the candidate genes TkBG11 and TkTIFY10A may be useful for the optimization of root phenotypes in *T. koksaghyz*, which entails roots of medium length with a high biomass preferably in the

upper root part. Varieties showing this root phenotype are likely to be more suitable for harvesting and processing, thereby optimizing not only the yield of rubber and inulin but also the recovery of these valuable products.

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 SRA BioProject, accession no: PRJNA783530.

AUTHOR CONTRIBUTIONS

AW, K-UR, DP and CG conceived and designed the experiments. AW and K-UR performed the experiments. AW, K-UR, CG and RT analyzed the data. AW, K-UR, CG, RT and DP wrote the paper. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge the technical assistance of Daniela Ahlert (Institute of Plant Biology and Biotechnology, University of Münster, Germany), Denise Weinberg (Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Münster, Germany) and Sascha Ahrens (Institute of Plant Biology and Biotechnology, University of Münster, Germany).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.784883/full#supplementary-material

REFERENCES

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* 37, 420–423. doi:10.1038/s41587-019-0036-z

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Arias, M., Hernández, M., and Ritter, E. (2016a). How Does Water Supply Affect *Taraxacum koksaghyz* Rod. Rubber, Inulin and Biomass Production? *Ind. Crops Prod.* 91, 310–314. doi:10.1016/j.indcrop.2016.07.024

Arias, M., Herrero, J., Ricobaraza, M., Hernández, M., and Ritter, E. (2016b). Evaluation of Root Biomass, Rubber and Inulin Contents in Nine *Taraxacum koksaghyz* Rodin Populations. *Ind. Crops Prod.* 83, 316–321. doi:10.1016/j.indcrop.2016.01.023

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., et al. (2016). Heatmapper: Web-Enabled Heat Mapping for All. *Nucleic Acids Res.* 44, W147–W153. doi:10.1093/nar/gkw419

Barclay, G. F., Peterson, C. A., and Tyree, M. T. (1982). Transport of Fluorescein in Trichomes of *Lycopersicon esculentum*. *Can. J. Bot.* 60, 397–402. doi:10.1139/b82-055

Barral, P., Suárez, C., Batanero, E., Alfonso, C., Alché, J. D. D., Rodríguez-García, M. I., et al. (2005). An Olive Pollen Protein with Allergenic Activity, Ole e 10, Defines a Novel Family of Carbohydrate-Binding Modules and is Potentially Implicated in Pollen Germination. *Biochem. J.* 390, 77–84. doi:10.1042/BJ20050456

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., et al. (2017). UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* 45, D158–D169. doi:10.1093/nar/gkw1099

Benitez-Alfonso, Y., Faulkner, C., Pendle, A., Miyashima, S., Helariutta, Y., and Maule, A. (2013). Symplastic Intercellular Connectivity Regulates Lateral Root Patterning. *Developmental Cell* 26, 136–147. doi:10.1016/j.devcel.2013.06.010

Benninghaus, V. A., Van Deenen, N., Müller, B., Roelfs, K.-U., Lassowskat, I., Finkemeier, I., et al. (2020). Comparative Proteome and Metabolome Analyses of Latex-Exuding and Non-Exuding *Taraxacum koksaghyz* Roots Provide Insights into Laticifer Biology. *J. Exp. Bot.* 71, 1278–1293. doi:10.1093/jxb/erz512

Bojahr, J., Nhengiwa, O., Krezdorn, N., Rotter, B., Saal, B., Ruge-Wehling, B., et al. (2016). Massive Analysis of cDNA Ends (MACE) Reveals a Co-Segregating Candidate Gene for LpPg1 Stem Rust Resistance in Perennial Ryegrass (*Lolium perenne*). *Theor. Appl. Genet.* 129, 1915–1932. doi:10.1007/s00122-016-2749-4

Breese, M. R., and Liu, Y. (2013). NGSUtils: A Software Suite for Analyzing and Manipulating Next-Generation Sequencing Datasets. *Bioinformatics* 29, 494–496. doi:10.1093/bioinformatics/bts731

Cai, H., Chen, H., Yi, T., Daimon, C. M., Boyle, J. P., Peers, C., et al. (2013). VennPlex–A Novel Venn Diagram Program for Comparing and Visualizing Datasets with Differentially Regulated Datapoints. *PLoS One* 8, e53388. doi:10.1371/journal.pone.0053388

Carlsbecker, A., Lee, J.-Y., Roberts, C. J., Dettmer, J., Lehesranta, S., Zhou, J., et al. (2010). Cell Signalling by microRNA165/6 Directs Gene Dose-Dependent Root Cell Fate. *Nature* 465, 316–321. doi:10.1038/nature08977

Chadwick, A. V., and Burg, S. P. (1967). An Explanation of the Inhibition of Root Growth Caused by Indole-3-Acetic Acid. *Plant Physiol.* 42, 415–420. doi:10.1104/pp.42.3.415

Chapple, C. (1998). Molecular-Genetic Analysis of Plant Cytochrome P450-Dependent Monooxygenases. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 311–343. doi:10.1146/annurev.arplant.49.1.311

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., et al. (2007). The JAZ Family of Repressors is the Missing Link in Jasmonate Signalling. *Nature* 448, 666–671. doi:10.1038/nature06006

Cho, E., and Zambryski, P. C. (2011). Organ Boundary1 defines a Gene Expressed at the Junction between the Shoot Apical Meristem and Lateral Organs. *Proc. Natl. Acad. Sci. USA* 108, 2154–2159. doi:10.1073/pnas.1018542108

Christensen, S. A., Huffaker, A., Hunter, C. T., Alborn, H. T., and Schmelz, E. A. (2016). A Maize Death Acid, 10-Oxo-11-Phytoenoic Acid, is the Predominant Cyclopentenone Signal Present during Multiple Stress and Developmental Conditions. *Plant Signaling Behav.* 11, e1120395. doi:10.1080/15592324.2015.1120395

Clouse, S. D., Langford, M., and McMorris, T. C. (1996). A Brassinosteroid-Insensitive Mutant in *Arabidopsis thaliana* Exhibits Multiple Defects in Growth and Development. *Plant Physiol.* 111, 671–678. doi:10.1104/pp.111.3.671

Coll-Garcia, D., Mazuch, J., Altmann, T., and Müssig, C. (2004). EXORDIUM Regulates Brassinosteroid-Responsive Genes. *FEBS Lett.* 563, 82–86. doi:10.1016/S0014-5793(04)00255-8

Cornish, K. (2017). Alternative Natural Rubber Crops: Why Should We Care? *Technology Innovation* 18, 245–256. doi:10.21300/18.4.2017.245

Dathe, W., Rönsch, H., Preiss, A., Schade, W., Sembdner, G., and Schreiber, K. (1981). Endogenous Plant Hormones of the Broad Bean, *Vicia faba* L. (-)-Jasmonic Acid, a Plant Growth Inhibitor in Pericarp. *Planta* 153, 530–535. doi:10.1007/BF00385537

De Storme, N., and Geelen, D. (2014). Callose Homeostasis at Plasmodesmata: Molecular Regulators and Developmental Relevance. *Front. Plant Sci.* 5, 1–23. doi:10.3389/fpls.2014.00138

Doxey, A. C., Yaish, M. W. F., Moffatt, B. A., Griffith, M., and McConkey, B. J. (2007). Functional Divergence in the *Arabidopsis* β-1,3-Glucanase Gene Family Inferred by Phylogenetic Reconstruction of Expression States. *Mol. Biol. Evol.* 24, 1045–1055. doi:10.1093/molbev/msm024

Eggert, M., Schiemann, J., and Thiele, K. (2018). Yield Performance of Russian Dandelion Transplants (*Taraxacum koksaghyz* L. Rodin) in Flat Bed and Ridge Cultivation with Different Planting Densities. *Eur. J. Agron.* 93, 126–134. doi:10.1016/j.eja.2017.12.003

Epping, J., van Deenen, N., Niephaus, E., Stolze, A., Fricke, J., Huber, C., et al. (2015). A Rubber Transferase Activator is Necessary for Natural Rubber Biosynthesis in Dandelion. *Nat. Plants* 1, 15048. doi:10.1038/nplants.2015.48

Fagard, M., Desnos, T., Desprez, T., Goubet, F., Refregier, G., Mouille, G., et al. (2000). *Procuste1* Encodes a Cellulose Synthase Required for normal Cell Elongation Specifically in Roots and Dark-Grown Hypocotyls of Arabidopsis. *Plant Cell* 12, 2409–2423. doi:10.1105/tpc.12.12.2409

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.-M., Gimenez-Ibanez, S., Geerinck, J., et al. (2011). The *Arabidopsis* bHLH Transcription Factors MYC3 and MYC4 are Targets of JAZ Repressors and Act Additively with MYC2 in the Activation of Jasmonate Responses. *Plant Cell* 23, 701–715. doi:10.1105/tpc.110.080788

Gao, X., Starr, J., Göbel, C., Engelberth, J., Feussner, I., Tumlinson, J., et al. (2008). Maize 9-Lipoxygenase ZmLOX3 Controls Development, Root-Specific Expression of Defense Genes, and Resistance to Root-Knot Nematodes. *Mol. Plant Microbe Interact.* 21, 98–109. doi:10.1094/MPMI-21-1-0098

Grechkin, A. N., Mukhtarova, L. S., Latypova, L. R., Gogolev, Y., Toporkova, Y. Y., and Hamberg, M. (2008). Tomato CYP74C3 is a Multifunctional Enzyme Not Only Synthesizing Allene Oxide but Also Catalyzing its Hydrolysis and Cyclization. *ChemBioChem* 9, 2498–2505. doi:10.1002/cbic.200800331

Grunewald, W., Vanholme, B., Pauwels, L., Plovie, E., Inzé, D., Gheysen, G., et al. (2009). Expression of the *Arabidopsis* Jasmonate Signalling Repressor JAZ1/TIFY10A is Stimulated by Auxin. *EMBO Rep.* 10, 923–928. doi:10.1038/embor.2009.103

Hao, Y., Hu, G., Breitel, D., Liu, M., Mila, I., Frasse, P., et al. (2015). Auxin Response Factor SlARF2 is an Essential Component of the Regulatory Mechanism Controlling Fruit Ripening in Tomato. *PLoS Genet.* 11, e1005649. doi:10.1371/journal.pgen.1005649

Hauser, M. T., Morikami, A., and Benfey, P. N. (1995). Conditional Root Expansion Mutants of *Arabidopsis*. *Development* 121, 1237–1252. doi:10.1242/dev.121.4.1237

Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., et al. (2009). Fast Mapping of Short Sequences with Mismatches, Insertions and Deletions Using Index Structures. *PLoS Comput. Biol.* 5, e1000502. doi:10.1371/journal.pcbi.1000502

Hsiao, Y.-C., and Yamada, M. (2021). The Roles of Peptide Hormones and their Receptors during Plant Root Development. *Genes* 12, 22–13. doi:10.3390/genes12010022

Itoh, A., Schilmiller, A. L., McCaig, B. C., and Howe, G. A. (2002). Identification of a Jasmonate-Regulated Allene Oxide Synthase that Metabolizes 9-Hydroperoxides of Linoleic and Linolenic Acids. *J. Biol. Chem.* 277, 46051–46058. doi:10.1074/jbc.M207234200

Kahl, G., Molina, C., Rotter, B., Jüngling, R., Frank, A., Krezdorn, N., et al. (2012). Reduced Representation Sequencing of Plant Stress Transcriptomes. *J. Plant Biochem. Biotechnol.* 21, 119–127. doi:10.1007/s13562-012-0129-y

Kazan, K., and Manners, J. M. (2013). MYC2: The Master in Action. *Mol. Plant* 6, 686–703. doi:10.1093/mp/sss128

Keener, H., Shah, A., Klingman, M., Wolfe, S., Pote, D., and Fioritto, R. (2018). Progress in Direct Seeding of an Alternative Natural Rubber Plant, *Taraxacum kok-saghyz* (L.E. Rodin). *Agronomy* 8, 182. doi:10.3390/agronomy8090182

Kim, J.-Y., Rim, Y., Wang, J., and Jackson, D. (2005). A Novel Cell-To-Cell Trafficking Assay Indicates that the KNOX Homeodomain is Necessary and Sufficient for Intercellular Protein and mRNA Trafficking. *Genes Dev.* 19, 788–793. doi:10.1101/gad.332805

Kim, J. Y., Yuan, Z., Cilia, M., Khalfan-Jagani, Z., and Jackson, D. (2002). Intercellular Trafficking of a KNOTTED1 Green Fluorescent Protein Fusion in the Leaf and Shoot Meristem of *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 99, 4103–4108. doi:10.1073/pnas.052484099

Kreuzberger, M., Hahn, T., Zibek, S., Schiemann, J., and Thiele, K. (2016). Seasonal Pattern of Biomass and Rubber and Inulin of Wild Russian Dandelion (*Taraxacum koksaghyz* L. Rodin) under Experimental Field Conditions. *Eur. J. Agron.* 80, 66–77. doi:10.1016/j.eja.2016.06.011

Krotkov, G. (1945). A Review of Literature on *Taraxacum koksaghyz* Rod. *Bot. Rev.* 11, 417–461. doi:10.1007/BF02861139

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

Laibach, N., Hillebrand, A., Twyman, R. M., Prüfer, D., and Schulze Gronover, C. (2015). Identification of a *Taraxacum brevicorniculatum* Rubber Elongation Factor Protein that is Localized on Rubber Particles and Promotes Rubber Biosynthesis. *Plant J.* 82, 609–620. doi:10.1111/tpj.12836

León Morcillo, R. J., Ocampo, J. A., and García Garrido, J. M. (2012). Plant 9-Lox Oxylipin Metabolism in Response to Arbuscular Mycorrhiza. *Plant Signaling Behav.* 7, 1584–1588. doi:10.4161/psb.22098

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

Levy, A., Erlanger, M., Rosenthal, M., and Epel, B. L. (2007). A Plasmodesmata-Associated β-1,3-Glucanase in *Arabidopsis*. *Plant J.* 49, 669–682. doi:10.1111/j.1365-313X.2006.02986.x

Lieberei, R. (2007). South American Leaf Blight of the Rubber Tree (*Hevea* spp.): New Steps in Plant Domestication Using Physiological Features and Molecular Markers. *Ann. Bot.* 100, 1125–1142. doi:10.1093/aob/mcm133

Lin, T., Xu, X., Ruan, J., Liu, S., Wu, S., Shao, X., et al. (2017). Genome Analysis of *Taraxacum kok-saghyz* Rodin Provides New Insights into Rubber Biosynthesis. *Natl. Sci. Rev.* 5, 78–87. doi:10.1093/nsr/nwx101

Liu, B., Lu, Y., Xin, Z., and Zhang, Z. (2009). Identification and Antifungal Assay of a Wheat β-1,3-Glucanase. *Biotechnol. Lett.* 31, 1005–1010. doi:10.1007/s10529-009-9958-8

Liu, Y., and Wang, H. (2020). JA Modulates Phytochrome a Signaling via Repressing FHY3 Activity by JAZ Proteins. *Plant Signaling Behav.* 15, 1726636–6. doi:10.1080/15592324.2020.1726636

Liu, Y., Xu, M., Liang, N., Zheng, Y., Yu, Q., and Wu, S. (2017). Symplastic Communication Spatially Directs Local Auxin Biosynthesis to Maintain Root Stem Cell Niche in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 114, 4005–4010. doi:10.1073/pnas.1616387114

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi:10.1093/nar/gkz991

Lucas, W. J., Bouché-Pillon, S., Jackson, D. P., Nguyen, L., Baker, L., Ding, B., et al. (1995). Selective Trafficking of KNOTTED1 Homeodomain Protein and its mRNA through Plasmodesmata. *Science* 270, 1980–1983. doi:10.1126/science.270.5244.1980

Mähönen, A. P., ten Tusscher, K., Siligato, R., Smetana, O., Díaz-Triviño, S., Salojärvi, J., et al. (2014). PLETHORA Gradient Formation Mechanism Separates Auxin Responses. *Nature* 515, 125–129. doi:10.1038/nature13663

Marcos, R., Izquierdo, Y., Vellosillo, T., Kulasekaran, S., Cascón, T., Hamberg, M., et al. (2015). 9-Lipoxygenase-Derived Oxylipins Activate Brassinosteroid Signaling to Promote Cell Wall-Based Defense and Limit Pathogen Infection. *Plant Physiol.* 169, 00992. doi:10.1104/pp.15.00992

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* 17, 10. doi:10.14806/ej.17.1.200

Mellor, N. L., Voß, U., Janes, G., Bennett, M. J., Wells, D. M., and Band, L. R. (2020). Auxin Fluxes through Plasmodesmata Modify Root-Tip Auxin Distribution. *Development* 147, dev181669. doi:10.1242/dev.181669

Morcillo, R. J. L., Navarrete, M. I. T., Bote, J. A. O., Monguio, S. P., and García-Garrido, J. M. (2016). Suppression of Allene Oxide Synthase 3 in Potato Increases Degree of Arbuscular Mycorrhizal Fungal Colonization. *J. Plant Physiol.* 190, 15–25. doi:10.1016/j.jplph.2015.11.003

Munt, O., Arias, M., Hernandez, M., Ritter, E., Schulze Gronover, C., and Prüfer, D. (2012). Fertilizer and Planting Strategies to Increase Biomass and Improve Root Morphology in the Natural Rubber Producer *Taraxacum brevicorniculatum*. *Ind. Crops Prod.* 36, 289–293. doi:10.1016/j.indcrop.2011.10.014

Nakajima, K., Sena, G., Nawy, T., and Benfey, P. N. (2001). Intercellular Movement of the Putative Transcription Factor SHR in Root Patterning. *Nature* 413, 307–311. doi:10.1038/35095061

Niephaus, E., Müller, B., Deenen, N., Lassowskat, I., Bonin, M., Finkemeier, I., et al. (2019). Uncovering Mechanisms of Rubber Biosynthesis in *Taraxacum koksaghyz* – Role of cis-Prenyltransferase-Like 1 Protein. *Plant J.* 100, 591–609. doi:10.1111/tpj.14471

Niu, Y., Figueroa, P., and Browse, J. (2011). Characterization of JAZ-Interacting bHLH Transcription Factors that Regulate Jasmonate Responses in *Arabidopsis*. *J. Exp. Bot.* 62, 2143–2154. doi:10.1093/jxb/erq408

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* 44, D733–D745. doi:10.1093/nar/gkv1189

Ohta, K., Hamada, S., and Nakamura, T. (1993). Production of High Concentrations of Ethanol from Inulin by Simultaneous Saccharification and Fermentation Using *Aspergillus niger* and *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 59, 729–733. doi:10.1128/aem.59.3.729-733.1993

Pauwels, L., Barbero, G. F., Geerinck, J., Tilleman, S., Grunewald, W., Pérez, A. C., et al. (2010). NINJA Connects the Co-Repressor TOPLESS to Jasmonate Signalling. *Nature* 464, 788–791. doi:10.1038/nature08854

Pfaffl, M. W. (2001). A New Mathematical Model for Relative Quantification in Real-Time RT-PCR. *Nucleic Acids Res.* 29, 45e–45. doi:10.1093/nar/29.9.e45

Pierleoni, A., Martelli, P. L., and Casadio, R. (2008). PredGPI: A GPI-Anchor Predictor. *BMC Bioinformatics* 9, 1–11. doi:10.1186/1471-2105-9-392

Pütter, K. M., van Deenen, N., Unland, K., Prüfer, D., and Schulze Gronover, C. (2017). Isoprenoid Biosynthesis in Dandelion Latex is Enhanced by the Overexpression of Three Key Enzymes Involved in the Mevalonate Pathway. *BMC Plant Biol.* 17, 88. doi:10.1186/s12870-017-1036-0

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007

Roberfroid, M. B. (2005). Introducing Inulin-Type Fructans. *Br. J. Nutr.* 93, S13–S25. doi:10.1079/BJN20041350

Robson, F., Okamoto, H., Patrick, E., Harris, S.-R., Wasternack, C., Brearley, C., et al. (2010). Jasmonate and Phytochrome A Signaling in *Arabidopsis* Wound and Shade Responses are Integrated through JAZ1 Stability. *Plant Cell* 22, 1143–1160. doi:10.1105/tpc.109.067728

Roddick, J. G., Rijnenberg, A. L., and Ikekawa, N. (1993). Developmental Effects of 24-Epibrassinolide in Excised Roots of Tomato Grown *In Vitro*. *Physiol. Plant* 87, 453–458. doi:10.1111/j.1399-3054.1993.tb02493.x

Schmidt, T., Lenders, M., Hillebrand, A., van Deenen, N., Munt, O., Reichelt, R., et al. (2010). Characterization of Rubber Particles and Rubber Chain Elongation in *Taraxacum koksaghyz*. *BMC Biochem.* 11, 11. doi:10.1186/1471-2091-11-11

Schulze Gronover, C., Wahler, D., and Prüfer, D. (2011). "Natural Rubber Biosynthesis and Physic-Chemical Studies on Plant Derived Latex," in *Biotechnology of Biopolymers* (London, United Kingdom: InTech), 75–88. doi:10.5772/17144

Stahl, Y., Grabowski, S., Bleckmann, A., Kühnemuth, R., Weidtkamp-Peters, S., Pinto, K. G., et al. (2013). Moderation of *Arabidopsis* Root Stemness by CLAVATA1 and ARABIDOPSIS CRINKLY4 Receptor Kinase Complexes. *Curr. Biol.* 23, 362–371. doi:10.1016/j.cub.2013.01.045

Staswick, P. E., Su, W., and Howell, S. H. (1992). Methyl Jasmonate Inhibition of Root Growth and Induction of a Leaf Protein are Decreased in an *Arabidopsis thaliana* Mutant. *Proc. Natl. Acad. Sci. USA* 89, 6837–6840. doi:10.1073/pnas.89.15.6837

Stolze, A., Wanke, A., van Deenen, N., Geyer, R., Prüfer, D., and Schulze Gronover, C. (2017). Development of Rubber-Enriched Dandelion Varieties by Metabolic Engineering of the Inulin Pathway. *Plant Biotechnol. J.* 15, 740–753. doi:10.1111/pbi.12672

Stumpe, M., Göbel, C., Demchenko, K., Hoffmann, M., Klösgen, R. B., Pawlowski, K., et al. (2006). Identification of an Allene Oxide Synthase (CYP74C) that Leads to Formation of α-Ketols from 9-Hydroperoxides of Linoleic and Linolenic Acid in Below-Ground Organs of Potato. *Plant J.* 47, 883–896. doi:10.1111/j.1365-313X.2006.02843.x

Thanseem, I., Joseph, A., and Thulaseedharan, A. (2005). Induction and Differential Expression of β-1,3-glucanase mRNAs in Tolerant and Susceptible *Hevea* Clones in Response to Infection by *Phytophthora meadii*. *Tree Physiol.* 25, 1361–1368. doi:10.1093/treephys/25.11.1361

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., et al. (2007). JAZ Repressor Proteins are Targets of the SCFCOI1 Complex during Jasmonate Signalling. *Nature* 448, 661–665. doi:10.1038/nature05960

Uhlmann, M. (1951). *Wertvolle Kautschukpflanzen des gemässigten Klimas. Dargestellt auf Grund von Sowjetischer Forschungsarbeiten*. Verlag: Berlin: Akademie Verlag.

van Beilen, J. B., and Poirier, Y. (2007a). Establishment of New Crops for the Production of Natural Rubber. *Trends Biotechnol.* 25, 522–529. doi:10.1016/j.tibtech.2007.08.009

van Beilen, J. B., and Poirier, Y. (2007b). Guayule and Russian Dandelion as Alternative Sources of Natural Rubber. *Crit. Rev. Biotechnol.* 27, 217–231. doi:10.1080/07388550701775927

Van Den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P., and Scheres, B. (1997). Short-Range Control of Cell Differentiation in the *Arabidopsis* Root Meristem. *Nature* 390, 287–289. doi:10.1038/36856

Vandenende, W., Deconinck, B., and Vanlaere, A. (2004). Plant Fructan Exohydrolases: A Role in Signaling and Defense? *Trends Plant Sci.* 9, 523–528. doi:10.1016/j.tplants.2004.09.008

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate Normalization of Real-Time Quantitative RT-PCR Data by Geometric Averaging of Multiple Internal Control Genes. *Genome Biol.* 3, research0034.1. doi:10.1186/gb-2002-3-7-research0034

Vanholme, B., Grunewald, W., Bateman, A., Kohchi, T., and Gheysen, G. (2007). The Tify Family Previously Known as ZIM. *Trends Plant Sci.* 12, 239–244. doi:10.1016/j.tplants.2007.04.004

Vatén, A., Dettmer, J., Wu, S., Stierhof, Y.-D., Miyashima, S., Yadav, S. R., et al. (2011). Callose Biosynthesis Regulates Symplastic Trafficking during Root Development. *Developmental Cell* 21, 1144–1155. doi:10.1016/j.devcel.2011.10.006

Vellosillo, T., Marti´nez, M., Lo´pez, M. A., Vicente, J., Casco´n, T., Dolan, L., et al. (2007). Oxylipins Produced by the 9-Lipoxygenase Pathway in Arabidopsis Regulate Lateral Root Development and Defense Responses through a Specific Signaling Cascade. *Plant Cell Online* 19, 831–846. doi:10.1105/tpc.106.046052

Vicente, J., Cascón, T., Vicedo, B., García-Agustín, P., Hamberg, M., and Castresana, C. (2012). Role of 9-Lipoxygenase and α-Dioxygenase Oxylipin Pathways as Modulators of Local and Systemic Defense. *Mol. Plant* 5, 914–928. doi:10.1093/mp/ssr105

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2009). DEGseq: An R Package for Identifying Differentially Expressed Genes from RNA-Seq Data. *Bioinformatics* 26, 136–138. doi:10.1093/bioinformatics/btp612

Whaley, W. G., and Bown, J. S. (1947). *Russian Dandelion (kok-saghyz) - An Emergency Source of Natural Rubber*. Washington, D.C.: U.S. Dept. of Agriculture.

Wieghaus, A., Prüfer, D., and Schulze Gronover, C. (2019). Loss of Function Mutation of the Rapid Alkalinization Factor (RALF1)-Like Peptide in the Dandelion *Taraxacum koksaghyz* Entails a High-Biomass Taproot Phenotype. *PLoS One* 14, e0217454–23. doi:10.1371/journal.pone.0217454

Wolf, S., Deom, C. M., Beachy, R., and Lucas, W. J. (1991). Plasmodesmatal Function is Probed Using Transgenic Tobacco Plants that Express a Virus Movement Protein. *Plant Cell* 3, 593–604. doi:10.1105/tpc.3.6.593

Wu, X., Dinneny, J. R., Crawford, K. M., Rhee, Y., Citovsky, V., Zambryski, P. C., et al. (2003). Modes of Intercellular Transcription Factor Movement in the *Arabidopsis* Apex. *Development* 130, 3735–3745. doi:10.1242/dev.00577

Yadav, V., Mallappa, C., Gangappa, S. N., Bhatia, S., and Chattopadhyay, S. (2005). A Basic Helix-Loop-Helix Transcription Factor in *Arabidopsis*, MYC2, Acts as a Repressor of Blue Light-Mediated Photomorphogenic Growth. *Plant Cell* 17, 1953–1966. doi:10.1105/tpc.105.032060

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

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., et al. (2018). WEGO 2.0: A Web Tool for Analyzing and Plotting GO Annotations, 2018 Update. *Nucleic Acids Res.* 46, W71–W75. doi:10.1093/nar/gky400

Yonemaru, J.-i., Yamamoto, T., Fukuoka, S., Uga, Y., Hori, K., and Yano, M. (2010). Q-TARO: QTL Annotation Rice Online Database. *Rice* 3, 194–203. doi:10.1007/s12284-010-9041-z

Zavaliev, R., Levy, A., Gera, A., and Epel, B. L. (2013). Subcellular Dynamics and Role of *Arabidopsis* β-1,3-Glucanases in Cell-To-Cell Movement of Tobamoviruses. *Mol. Plant Microbe Interact.* 26, 1016–1030. doi:10.1094/MPMI-03-13-0062-R

Zawada, A. M., Rogacev, K. S., Müller, S., Rotter, B., Winter, P., Fliser, D., et al. (2014). Massive Analysis of cDNA Ends (MACE) and miRNA Expression

Profiling Identifies Proatherogenic Pathways in Chronic Kidney Disease. *Epigenetics* 9, 161–172. doi:10.4161/epi.26931

Zhao, L., Nakazawa, M., Takase, T., Manabe, K., Kobayashi, M., Seki, M., et al. (2004). Overexpression of *LSH1*, a Member of an Uncharacterised Gene Family, Causes Enhanced Light Regulation of Seedling Development. *Plant J.* 37, 694–706. doi:10.1111/j.1365-313X.2003.01993.x

Conflict of Interest: Author RT was employed by the company TRM Ltd.

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 February 2022
doi: 10.3389/fgene.2021.784878

Genome-Wide Evolution and Comparative Analysis of Superoxide Dismutase Gene Family in Cucurbitaceae and Expression Analysis of *Lagenaria siceraria* Under Multiple Abiotic Stresses

*Shamsur Rehman[1], Arif Rashid[2], Muhammad Aamir Manzoor[3], Lingling Li[1], Weibo Sun[1], Muhammad Waheed Riaz[4,5], Dawei Li[1] and Qiang Zhuge[1]**

[1]Co-Innovation Center for Sustainable Forestry in Southern China, Key Laboratory of Forest Genetics and Biotechnology, College of Biology and the Environment, Nanjing Forestry University, Ministry of Education, Nanjing, China, [2]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China, [3]School of Life Sciences, Anhui Agricultural University, Hefei, China, [4]State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, China, [5]Zhejiang Provincial Key Laboratory of Resources Protection and Innovation of Traditional Chinese Medicine, Zhejiang A&F University, Hangzhou, China

OPEN ACCESS

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

Reviewed by:
Ali Raza,
Fujian Agriculture and Forestry University, China
Irfan Ali Sabir,
Shanghai Jiao Tong University, China
Haidong Yan,
University of Georgia, United States

***Correspondence:**
Qiang Zhuge
qzhuge@njfu.edu.cn

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

Received: 28 September 2021
Accepted: 30 December 2021
Published: 08 February 2022

Citation:
Rehman S, Rashid A, Manzoor MA, Li L, Sun W, Riaz MW, Li D and Zhuge Q (2022) Genome-Wide Evolution and Comparative Analysis of Superoxide Dismutase Gene Family in Cucurbitaceae and Expression Analysis of Lagenaria siceraria Under Multiple Abiotic Stresses. Front. Genet. 12:784878. doi: 10.3389/fgene.2021.784878

Superoxide dismutase (SOD) is an important enzyme that serves as the first line of defense in the plant antioxidant system and removes reactive oxygen species (ROS) under adverse conditions. The SOD protein family is widely distributed in the plant kingdom and plays a significant role in plant growth and development. However, the comprehensive analysis of the *SOD* gene family has not been conducted in Cucurbitaceae. Subsequently, 43 *SOD* genes were identified from Cucurbitaceae species [*Citrullus lanatus* (watermelon), *Cucurbita pepo* (zucchini), *Cucumis sativus* (cucumber), *Lagenaria siceraria* (bottle gourd), *Cucumis melo* (melon)]. According to evolutionary analysis, *SOD* genes were divided into eight subfamilies (I, II, III, IV, V, VI, VII, VIII). The gene structure analysis exhibited that the *SOD* gene family had comparatively preserved exon/intron assembly and motif as well. Phylogenetic and structural analysis revealed the functional divergence of Cucurbitaceae *SOD* gene family. Furthermore, microRNAs 6 miRNAs were predicted targeting 3 *LsiSOD* genes. Gene ontology annotation outcomes confirm the role of *LsiSODs* under different stress stimuli, cellular oxidant detoxification processes, metal ion binding activities, SOD activity, and different cellular components. Promoter regions of the *SOD* family revealed that most *cis*-elements were involved in plant development, stress response, and plant hormones. Evaluation of the gene expression showed that most *SOD* genes were expressed in different tissues (root, flower, fruit, stem, and leaf). Finally, the expression profiles of eight *LsiSOD* genes analyzed by qRT-PCR suggested that these genetic reserves responded to drought, saline, heat, and cold stress. These findings laid the foundation for further study of the role of the *SOD* gene family in Cucurbitaceae. Also, they provided the potential for its use in the genetic improvement of Cucurbitaceae.

Keywords: *LsiSOD* genes, phylogenetic analysis, stress response, duplication events, expression pattern

INTRODUCTION

In natural conditions, plants are often susceptible to drought, salt, extreme temperatures, heavy metals, and other stresses that have a significant impact on the growth, development, and production of plants (Mittler and Blumwald, 2010; Cramer et al., 2011). When a plant is stressed, it adjusts its homeostatic apparatus by producing more reactive oxygen species (ROS) in its cells and ROS, toxic-free radicals produced under stress by plant cells that can oxidize the protein, destroy the cell membrane, and cause DNA damage (Lee et al., 2007; Karuppanapandian et al., 2011; Feng et al., 2015). These stresses inevitably accompany the development of ROS, including hydroxyl radicals (OH), superoxide anion radicals (O_2^-), peroxide radicals (HOO^-), hydrogen peroxide (H_2O_2), and singlet oxygen ($^1O_2$), which cause damage to the cell membrane, peroxidization and deterioration of macromolecules, and ultimately lead to the death of cells. ROS are also considered signaling molecules in different organisms and can affect various physiological processes in plants, For example, some prominent active oxygen scavengers can resist environmental stresses by regulating the expression of enzyme reaction family genes such as superoxide dismutase (SOD), catalase (CAT), peroxidase (POD), glutathione peroxidase (GPX), and peroxidase (PrxR) (Mittler et al., 2004; Ahmad et al., 2010; Bafana et al., 2011; Filiz and Tombuloğlu, 2015). Mainly, ROS are formed in the apoplast, mitochondria, plasma membrane, chloroplast, peroxisomes, endoplasmic reticulum, and cell walls (Mittler, 2017; Hasanuzzaman et al., 2020). Therefore, to manage ROS noxiousness, plants have established well-organized and composite antioxidant defense systems, including numerous non-enzymatic and enzymatic antioxidants.

SOD is a type of metal enzyme that is first found in bovine red blood cells and then characterized in bacteria, vertebrates, and higher plants (Mann and Keilin, 1938; Rabinowitch and Sklan, 1980; Tepperman and Dunsmuir, 1990; Kim et al., 1996; Zelko et al., 2002). Researchers of different studies have found that SODs can catalyze superoxide O_2 to disproportionate into O_2 and H_2O_2 (McCord and Fridovich, 1969; Tepperman and Dunsmuir, 1990). SODs are detected in plants in roots, fruits, leaves, and seeds, which provide necessary protection for cells against oxidative stress (Tepperman and Dunsmuir, 1990). As per the binding pattern of metallic cofactors that cooperate with vigorous sites, *SOD* genes are categorized further into four groups: 1) iron *FeSODs*, 2) copper/zinc *Cu/ZnSODs*, 3) manganese *MnSODs*, and 4) nickel *NiSODs* (Mittler et al., 2004; Feng et al., 2016; Yan et al., 2016; Su et al., 2021). The different subtypes of *SODs* have almost comparable functions. However, they have different metallic cofactors and amino acid sequencing, *in vitro* subcellular location and crystal structure, and different hydrogen peroxide sensitivity (Abreu and Cabelli, 2010; Yan et al., 2016). These SODs are distributed in cell compartments individually and play an essential role in oxidative stress (Alscher et al., 2002). Among the *SODs*, *Cu/ZnSODs* are predominantly distributed in the chloroplast, extracellular space, and cytoplasm, and are present in certain bacteria and all eukaryotes, while *MnSODs* mostly present in plant mitochondria (Pilon et al., 2011; Feng et al., 2015; Song et al., 2018; Wang et al., 2018; Hu et al., 2019; Verma et al., 2019; Lu et al., 2020; Silva et al., 2020). *MnSODs* in the plant genome play a role in eliminating ROS from mitochondria (Møller, 2001). *FeSODs* are primarily distributed in protozoa, prokaryotes, cytoplasms, and plant chloroplasts, while *NiSODs* are found in *Streptomyces* and also in some cyanobacteria, but not in plants (Youn et al., 1996; Wuerges et al., 2004; Miller, 2012).

Recent studies have shown that SODs can protect plants from abiotic stress such as heat, drought, cold, abscisic acid, salt, and ethylene (Wang et al., 2004; Pilon et al., 2011; Asensio et al., 2012; Feng et al., 2015). Several studies have shown that the *SOD* genes can be induced and transcribed in different plants under various stress conditions, such as heat, drought, cold, salt, osmotic stress, oxidative stress, and hormonal signal transduction (Wang et al., 2004; Pilon et al., 2011; Feng et al., 2015; Feng et al., 2016; Yan et al., 2016). *SOD* gene family under different hormones and abiotic stress conditions in rapeseed (Su et al., 2021), *Salvia miltiorrhiza* (Han et al., 2020), *Zostera marina* (Zang et al., 2020), and *Hordeum vulgare* (Zhang et al., 2021) were recently published articles. Furthermore, diverse forms of *SOD* genes display different expression patterns under different stress conditions. In tomatoes, for example, *SlSOD1* is the only significantly upregulated gene in the nine *SlSOD* genes, while *SlSOD*2, *SlSOD*5, *SlSOD*6, and *SlSOD*8 are regulated by salt stress. However, the expression levels of four "*SlSOD*2, *SlSOD*5, *SlSOD*6, and *SlSOD*8" genes are found high during imposed drought environment (Feng et al., 2016). Furthermore, the expression patterns of the same type of *SOD* gene were different under stress. For example, the studies found that there was no change in the expression of *MnSODs* in *Arabidopsis* under oxidative stress, and the researchers found that there was a significant change in the expression of *MnSODs* in peas, wheat, and *Zostera marina* under salt stress (Gómez et al., 1999; Wu et al., 1999; Baek and Skinner, 2003; Zang et al., 2020). These results suggest that different *SOD* genes have different expression patterns under different environmental stresses. Furthermore, researchers have also discovered that alternative splicing and miRNAs may be involved in the regulation of *SOD* expression (Srivastava et al., 2009; Lu et al., 2011). Until now, the *SOD* gene family has been described in many plant species, including *Arabidopsis*, *Sorghum*, *Musa acuminata*, *Phaseolus vulgaris*, and *Populus* (Kliebenstein et al., 1998; Zelko et al., 2002; Ballal and Manna, 2009; Fernández-Ocaña et al., 2011; Lin and Lai, 2013; Molina-Rueda et al., 2013; Filiz et al., 2014; Cui et al., 2015; Feng et al., 2015; Feng et al., 2016; Ho et al., 2017; Verma et al., 2019). Cucurbitaceae is a significant economic and nutritional crop in the world and *SOD* gene exists as a superfamily. The phylogeny line, genome circulation, gene assembly, preserved patterns, and expression profiles of these genes in different tissues have been thoroughly analyzed, laying the foundation for functional identification of SOD genes in Cucurbitaceae.

MATERIALS AND METHODS

Retrieval of *SOD* Gene Family in Five Cucurbitaceae Species

To study the *SOD* gene family of cucurbit plants, the whole genome of individual species of Cucurbitaceae was searched by

the Blastp search method and *Arabidopsis* sequence of SOD used as the query (Kliebenstein et al., 1998). Genomic, protein, and CDS (coding DNA sequence) sequences of *SOD* gene family have been identified and downloaded from the Cucurbitaceae database (CuGenDB) (http://cucurbitgenetics.org/) (Zheng et al., 2019) Subsequently, we used two methods to identify SOD genes in five Cucurbitaceae species, i.e., BLASTP (protein blast) and the hidden Markov model (HMM). For BLASTP, we used eight *A. thaliana SODs* (AT1G08830.1/*AtCSD1*, AT2G28190.1/*AtCSD2*, AT5G18100.1/*AtCSD3*, AT4G25100.1/*AtFSD1*, AT5G51100.1/*AtFSD2*, AT5G23310.1/*AtFSD3*, AT3G10920.1/*AtMSD1*, and AT3G56350.1/*At00MSD2*) amino acid sequences as a query with an e-value set to 1e−5. The amino acid sequences of eight *AtSODs* were obtained from the TAIR *Arabidopsis* genome database (http://www.arabidopsis.org/). The database web resources SMART (http:/smart.embl-heidelberg.de/) (Schultz et al., 1998) and the Pfam protein domain database (http://pfam.xfam.org/) were used to scan certain amino acid sequences conserved domain SOD_Cu (PF00080.21) and SOD_Fe_C (PF02777.19). The ProtParam software (http:/web.expasy.org/ProtParam) (Gasteiger et al., 2005) was used to evaluate the physiochemical properties (molecular weight, amino acid length, and isoelectric point) of each SOD protein.

Phylogenetic Analyses of *SOD* Gene Family

All SOD full-length protein sequences of the five Cucurbitaceae species [*Citrullus lanatus* (watermelon), *Cucurbita pepo* (zucchini), *Cucumis sativus* (cucumber), *Lagenaria siceraria* (bottle gourd), *Cucumis melo* (melon)] including seven other plant species were aligned using clustalX software with default parameters (1,000 bootstraps, pairwise deletion) (Thompson et al., 1997). A phylogenetic tree was constructed with a maximum likelihood method (MLM) by using the online version of the IQ-TREE software (http:/iqtree.cibiv.univie.ac.at). Finally, the phylogenetic tree was visualized through online iTOL software (https://itol.embl.de/) (Letunic and Bork, 2019; Manzoor et al., 2021a).

Chromosomal Distribution of *SOD* Genes

The chromosomal localizations of Cucurbitaceae were obtained from CuGenDB database (http://cucurbitgenetics.org/). For chromosome mapping of *SOD* gene, map inspect tool (http://mapinspect.software.com) was used. The aforementioned description is used to map the distribution of *SOD* gene throughout the Cucurbitaceae family individual species used in this study (Hu et al., 2016).

Intron/Exon Structure and Conserved Motif Analysis of Protein

The genome sequence and the coding sequence of the *SOD* gene were downloaded from the genome of the individual species of the Cucurbitaceae family from CuGenDB database. The Gene Structure and Display Server (GSDS) (http:/gsds.cbi.pku.edu.cn/) (Hu et al., 2015) was used to evaluate intron distribution dynamics and the splicing mechanism of each *SOD* gene. The Multiple Expectation Maximization for Motif Elicitation (MEME) software (Bailey et al., 2009) identified the conserved SOD protein motif (http:/meme-suite.org/tools/meme). Finally, the subcellular localization was predicted using WoLF PSORT (http://wolfpsort.org) (Horton et al., 2007).

Prediction of the *SOD Cis*-Regulatory Elements in the Promoter

For the prediction of regulatory elements on the promoter regions of *CuSODs*, 1,500 kb upstream DNA sequence of each *SOD* gene was collected from all *SOD* genes using the online web server PlantCARE (http:/bioinformatics.psb.ugent.be/webtools/PlantCARE/html/) (Higo et al., 1999; Lescot et al., 2002).

Gene Duplications and Collinearity Relationship Analysis

Multiple collinearity scan toolkit (MCScanX) was used for collinearity analysis with BLASTP (E < 1e−5) against five Cucurbitaceae species (Wang et al., 2012; Qiao et al., 2018). Different modes of duplications (WGD/segmental, dispersed, proximal, and tandem duplications) among five Cucurbitaceae species (*Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativus*, *Lagenaria siceraria*, *Cucumis melo*) were used. Gene duplications and collinearity relationships were visualized by using TBtools (Chen et al., 2020) and circos software.

Gene Ontology Annotation and MicroRNA Target Site Analysis

We used the CELLO v.2.5 software functional annotation platform to determine the Cucurbitaceae *SOD* gene's functional Classification. CELLO (http://cello.life.nctu.edu.tw/site) platform connects the genes with GO terms through hierarchical vocabularies (Conesa and Götz, 2008). Functional enrichment analysis of *SOD* genes was performed using DAVID online tools (DAVID 6.8; http://david.ncifcrf.gov/) (Huang et al., 2009). The GO terms were classified into three categories: biological process (BP), cellular component (CC), and molecular function (MF). The upregulated *SOD* genes and downregulated *SOD* genes were entered separately and p <0.01 was considered to indicate a statistically significant difference.

To determine the miRNA-mediated posttranscriptional regulation of *SOD*s, we searched the 5′ and 3′ untranslated regions (UTRs), and the coding regions, of the *SODs* for target sites of *Lagenaria siceraria* miRNAs obtained from various databases and published articles on the psRNA Target server using default parameters (Dai et al., 2018).

Plant Materials and Abiotic Stresses

The *L. siceraria* plants were planted in the growth chamber; seeds (variety: Winall 808) were provided by The National Engineering Laboratory of Crop Resistance Breeding, School of Life Sciences, Anhui Agricultural University, Hefei, China. To examine the behavior of the *SOD* genes under abiotic stress, seedlings of 2-week-old plants were carefully watered and grown under 24 h/18 h light and 16 h/8 h dark conditions in a growth chamber

before different stress inductions. However, seedlings were later transferred to the growth chamber at 50°C under normal lighting conditions during the heat treatment. The plant was treated for salt and drought resistance on the Murashige and Skoog (MS) liquid media containing 300 mM of PEG-6000. For the cold treatment, seedlings in the growth chamber were transferred to 4°C under light conditions (Hu et al., 2016). After treatment, the leaves were collected at 0, 4, 8, 16, and 24 h, and instantly frozen in liquid nitrogen immediately, and the sample was obtained to determine the transcription level of the *SOD* gene family in treated plant seedlings. Finally, all samples were instantly frozen in liquid nitrogen and stored at −80°C until further use.

RNA Isolation and Quantitative Real-Time PCR Reaction

RNA extraction reagent kit (Trizol) was purchased from Hua and Maike Biotechnology Co. Ltd., Beijing, China. Total RNA was extracted following the manufacturer's instructions. The RNA reverse transcription kit (TaKaRa Company, Japan) and Fluorescent Quantitative Reagent SYBR Green Master (Roche, United States) were used, and qRT-PCR was performed using a previously described method (Zhou et al., 2014). Gene-specific primers of *LsiSOD*s and *Lsi*-actin for the qRT-PCR system were designed by using GenScript server (www.genscript.com) and synthesized by Sangon Biotech (Shanghai) (**Supplementary Table S1**).

Actin gene of *L. siceraria* was used as an endogenous control to detect the relative expression of *LsiSOD* genes based on previous studies (Zhang et al., 2018). Primers used are given in **Supplementary Table S1**. qRT-PCR reactions were performed in biological triplicates. The relative expression level was calculated by $2^{-\Delta Ct}$ method and statistical analyses were measured using Microsoft Office 2010.

RESULTS

Genome-Wide Characterization of *SOD* Genes in Five Cucurbitaceae Species

In total, 43 *SODs* were identified in five Cucurbitaceae species from CuGenDB, and 8–10 genes were retrieved for individual species such as *Citrullus lanatus* (8 *SODs*), *Cucurbita pepo* (10 *SODs*), *Cucumis sativus* (9 *SODs*), *Lagenaria siceraria* (8 *SODs*), and *Cucumis melo* (8 *SODs*) (**Supplementary Table S2**). The detailed information of Cucurbitaceae species (molecular weight, starting and ending point, and theoretical p*I*) of the protein are listed in **Table 1**, and all protein sequences are shown in **Supplementary Table S10**.

In *Citrullus lanatus*, we detected eight *SOD* genes, containing five copper/zinc *SODs* and three iron/manganese *SODs*. The molecular weight, length, and p*I* of SOD protein were 15.12–35.01 kDa, 152–330 amino acids, and 5.31–7.95, respectively: five *Cu/ZnSODs* (*ClaSOD*1, *ClaSOD*2, *ClaSOD*3, *ClaSOD*6, *ClaSOD*8) were acidic in nature and one *MnSOD* (*ClaSOD*5) was basic, and in the two *FeSODs*, *ClaSOD7* was acidic and *ClaSOD*4 was slightly basic. Subcellular localization prediction results showed that six SODs (*ClaSOD1*, *ClaSOD2*, *ClaSOD3*, *ClaSOD6*, *ClaSOD7*, *ClaSOD8*) of *Cu/Zn SODs* were in the chloroplast, and *Cla-SOD*4 and *Cla-SOD*5 were localized to the mitochondria.

In *Cucurbita pepo*, ten *SOD* genes were detected, including seven *Cu/ZnSODs* and three Fe/Mn-SODs; the molecular weight, length, and p*I* values of SOD proteins were within the ranges of 15.28–59.17 kDa, 52–530 amino acids, and 5.18–9.73, respectively. The subcellular localization prediction showed that *Cu/ZnSODs CpSOD1*, *CpSOD3*, *CpSOD4*, *CpSOD6*, and *CpSOD7* were localized to the chloroplast, and two *Cu/ZnSODs*, *CpSOD*2 and *CpSOD*9, were localized to the cytoplasm. Furthermore, *CpSOD*5 and *CpSOD*10 were localized to the mitochondrion.

In *Cucumis sativus*, nine *SODs* were retrieved—six were *Cu/ZnSODs* and three were *Fe/MnSODs*. The molecular weight, length, and p*I* values of SOD proteins were within the range of 15.26–86.82, while amino acids range from 73 to 318 and 4.97–9.83 kDa, respectively. *CsaSOD2*, *CsaSOD3*, *CsaSOD4*, *CsaSOD6*, *CsaSOD7*, *CsaSOD8,* and *CsaSOD9* were acidic and *CsaSOD1* and *CsaSOD5* were basic. In the prediction of subcellular localization, *CsaSOD*1 and *CsaSOD*2 were localized to the mitochondrion. *CsaSOD3*, *CsaSOD4*, *CsaSOD6*, and *CsaSOD8* are localized to the chloroplast; *CsaSOD5* and *CsaSOD9* are in the nucleus; and *CsaSOD7* is localized to the cytoplasm.

In *Lagenaria siceraria*, eight *SODs* were detected collectively, including five copper/zinc SODs and three iron/manganese *SODs*. The molecular weight, length, and p*I* of SOD protein were 15.07–42.32 kDa, 152–397 amino acids, and 5.35–8.65, respectively. *LsiSOD1* and *LsiSOD3* of Cu/Zn-SODs are basic and *LsiSOD6* is slightly basic; *LsiSOD2*, *LsiSOD4*, *LsiSOD5*, *LsiSOD7*, and *LsiSOD9* are acidic. Subcellular localizations of *LsiSOD1*, *LsiSOD4*, *LsiSOD5*, *LsiSOD6*, *LsiSOD7*, and *LsiSOD8* are found in the chloroplast while *LsiSOD*2 and *LsiSOD*3 are in the mitochondrion.

In *Cucumis melo*, a total of eight *SODs* were identified, including four *Cu/ZnSODs* and four *Fe/MnSODs*, and the SOD protein's molecular weight, length, and p*I* values were observed within the ranges of 15.90–36.95 kDa, 134–347 amino acids, and 5.16–8.49, respectively; three SODs such as *MelSOD1*, *MelSOD4*, and *MelSOD8* are basic, and *MelSOD2*, *MelSOD3*, *MelSOD5*, *MelSOD6*, and *MelSOD7* are acidic. The prediction of subcellular localizations showed that *MelSOD1* and *MelSOD8* are in the mitochondrion, and *MelSOD3* and *MelSOD7* are in the cytoplasm. *MelSOD2*, *MelSOD4*, and *MelSOD5* are in the chloroplast (**Table 1**).

Structural Phylogenetic Analysis of *SOD* Genes Family in Five Cucurbitaceae Species

To define the evolutionary relationship of SODs in different plant species, we constructed a phylogenetic tree of SODs based on the full length of protein sequences and divided them into eight groups. A total of 106 SOD proteins were obtained from 13 plant

TABLE 1 | Physio-chemical characteristics of 43 *SOD* identified genes and sequence information in five species of Cucurbitaceae family.

Entry ID	Pfam domain		Exon	Intron	Protein						DNA	
					Localization by PC	MW kDa[1]	AA[2]	pI[3]	pH[4]	Gray[5]	MW Da[6]	Length[7]
ClaSOD7	IMA, IMC	PF00081, PF02777	8	7	Chloroplast	34.56	304	6.17	6.19	−0.757	1,415,651.18	4,585
ClaSOD8	HMA.CZ	PF00080	7	6	Chloroplast	35.01	330	5.26	5.26	−0.083	1,453,499.90	4,704
CpSOD1	CZ	PF00080	5	4	Chloroplast	12.45	122	6.71	7.32	−0.298	82,357.07	960
CpSOD2	CZ	PF00080	6	5	Cytoplasm	17.14	162	5.18	5.18	−0.255	1,096,062.63	3,657
CpSOD3	CZ	PF00080	2	1	Chloroplast	53.46	52	8.40	9.54	−0.010	20,704.82	246
CpSOD4	IMA, IMC	PF00081, PF02777	8	7	Chloroplast	33.36	295	5.51	5.51	−0.603	842,465.78	2,730
CpSOD5	IMA, IMC	PF00081, PF02777	6	5	Mitochondrion	26.23	235	7.94	8.07	−0.274	1,150,425.35	3,733
CpSOD6	CZ	PF00080	8	7	Chloroplast	22.50	221	6.02	6.03	0.108	1,126,708.80	3,645
CpSOD7	CZ	PF00080	6	5	Chloroplast	34.22	323	5.37	5.24	−0.053	347,046.97	4,122
CpSOD8	CZ	PF00080	7	6	Chloroplast	15.28	152	5.28	5.28	−0.117	848,531.63	2,741
CpSOD9	CZ, Thio	PF00080	5	4	Cytoplasm	59.17	530	9.73	9.73	−0.168	995,717.59	3,226
CpSOD10	IMA, IMC	PF00081, PF02777	6	5	Mitochondrion	23.79	216	7.98	8.74	−0.197	229,375.36	2,730
CsaSOD1	IMA, IMC	PF00081, PF02777	6	5	Mitochondrion	26.88	242	7.88	8.01	−0.245	1,145,395.55	3,707
CsaSOD2	IMA, IMC	PF00081, PF02777	8	7	Mitochondrion	30.85	267	6.79	6.86	−0.390	2,029,893.05	6,567
CsaSOD3	CZ	PF00080	7	6	Chloroplast	15.26	152	5.44	5.44	−0.140	1,369,045.72	4,434
CsaSOD4	CZ	PF00080	7	6	Chloroplast	15.40	152	4.97	4.97	−0.121	1,446,146.27	4,677
CsaSOD5	CZ, HMA	PF00080	5	4	Nucleus	22.09	223	9.83	10.44	−0.417	315,861.52	3,708
CsaSOD6	IMA, IMC	PF00081, PF02777	8	7	Chloroplast	36.38	318	5.78	5.78	−0.825	1,242,734.18	4,017
CsaSOD7	CZ	PF00080	7	6	Cytoplasm	15.87	157	6.53	6.57	−0.154	1,315,300.39	4,267
CsaSOD8	CZ	PF00080	8	7	Chloroplast	22.62	223	5.87	5.88	0.178	1,851,704.38	5,987
CsaSOD9	CZ	PF00080	3	2	Nucleus	86.82	73	5.80	5.85	−0.203	26,119.11	316
LsiSOD1	AP2, CZ	PF00080	8	7	Chloroplast	42.32	397	8.51	8.52	−0.376	3,503,807.65	11,347
LsiSOD2	IMA, IMC	PF00081, PF02777	8	7	Mitochondrion	32.45	292	6.67	5.36	−0.314	1,577,617.15	5,112
LsiSOD3	IMC, IMA	PF00081, PF02777	6	5	Mitochondrion	31.25	274	8.65	8.66	−0.369	1,709,879.76	5,529
LsiSOD4	CZ	PF00080	7	6	Chloroplast	15.07	152	5.59	5.59	−0.030	1,095,314.26	3,557
LsiSOD5	IMA, IMC	PF00081, PF02777	6	5	Chloroplast	32.29	286	6.11	5.90	−0.753	1,752,081.95	5,669
LsiSOD6	HMA, CZ	PF00080	6	5	Chloroplast	36.61	344	6.84	7.01	−0.125	1,422,371.22	4,647
LsiSOD7	CZ	PF00080	10	9	Chloroplast	32.46	239	5.90	5.90	0.030	1,636,024.86	5,283
LsiSOD8	CZ	PF00080	7	6	Chloroplast	17.93	171	5.35	6.12	−0.264	1,418,129.87	4,591
MelSOD1	IMA, IMC	PF00081, PF02777	8	7	Mitochondrion	30.82	267	8.24	8.50	−0.378	1,216,363.52	3,937
MelSOD2	CZ	PF00080	7	6	Chloroplast	17.59	172	5.61	6.16	−0.091	1,452,368.96	4,704
MelSOD3	CZ	PF00080	6	5	Cytoplasm	13.73	134	5.48	8.28	−0.171	1,817,515.13	5,882
MelSOD4	IMA, IMC	PF00081, PF02777	6	5	Chloroplast	36.94	347	8.49	5.48	−0.061	1,364,456.23	4,424
MelSOD5	IMA, IMC	PF00081, PF02777	7	6	Chloroplast	33.07	291	6.15	6.57	−0.829	1,331,923.50	4,313
MelSOD6	CZ	PF00080	7	6	Cytoplasm	15.90	157	6.53	5.61	−0.116	960,111.72	3,110
MelSOD7	CZ	PF00080	7	6	Chloroplast	22.01	217	5.78	8.65	0.070	996,665.45	3,222
MelSOD8	IMA, IMC	PF00081, PF02777	5	4	Mitochondrion	30.89	278	8.65	5.88	0.003	802,622.21	2,607

1, molecular weight of protein; 2, amino acid; 3, protein isoelectric point; 4, pH; 5, GRAVY; 6, molecular weight of DNA; 7, length. CZ, copper/zinc superoxide dismutase (PF00080) (SODC); IMA, iron/manganese superoxide dismutase alpha-hairpin domain (PF00081); IMC, iron/manganese superoxide dismutase (PF02777), C-terminal domain; PC, ProtComp9.0 server.

species including *Arabidopsis thaliana*, *Lagenaria siceraria*, *Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativa*, *Cucumis melo*, *Gossypium arboretum*, *Solanum lycopersicum*, *Zea mays*, *Solanum tuberosum*, *Sorghum bicolor*, *Populus trichocarpa*, and *Oryza sativa*. Three methods, maximum likelihood (ML), minimal evolution (ME), and maximum parsimony (MP), yielded nearly identical phylogenetic trees; therefore, only NJ tree was used for further analyses (Kumar et al., 2016) (**Figure 1**; **Supplementary Figure S2**). Based on the phylogenetic tree, the *SOD* genes were divided into five subgroups according to *Arabidopsis* (Silva et al., 2020). However, three unique clades in a phylogenetic tree may have independent evolutionary trajectories from other clades. It is also suggested that these clades may have individual evolution that is different from *Arabidopsis* (Nijhawan et al., 2008; Wei et al., 2012; Manzoor et al., 2021a). According to the bootstrap values quoted on the nodes, topography, and sequence similarities, all identified *SODs* from the Cucurbitaceae species were categorized into eight subfamilies (I–VIII). SOD protein sequences from all Cucurbitaceae species contributed in all subfamilies (**Figure 1**). These results suggested that there may have been gene loss or gain events that occurred throughout the evolutionary process. The gain and loss of specific *SOD* gene members caused functional divergence.

Conserved Motif Analysis of *SOD* Gene Family in Five Cucurbitaceae Species

The conserved motif analysis of the *SOD* family supported the classification and evolutionary relationships of the five Cucurbitaceae species *SOD* genes. In total, 20 motifs were detected from 43 *SOD* genes in *Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativus*, *Lagenaria siceraria*, and *Cucumis melo*. All *SOD* genes contained at least two motifs except *CpSOD*3 and *CsaSOD*5, which contain only one motif (**Figure 2B**): *Cu/ZnSOD* and *Fe/MnSODs*. Besides, *FeSODs* and *MnSODs* were clustered in the same group and belonged to a similar subcluster, while the *Cu/ZnSOD* was clustered in a different group. A similar cluster distribution was detected in the SOD proteins of each species, indicating that these *SOD* genes are highly conserved in different plants (**Figure 2A**). In Cucurbitaceae species, 5–9 exons were detected in *SOD* genes by comparing the genomic DNA and CDS sequences using the GSDS 2.0 utility (**Figure 2C**). However, differences were observed in the size and number of exons/introns in Cucurbitaceae. In the organization of exons/introns, a high degree of conservation has been observed, which is consistent with the high degree of similarity found by multiple alignments between protein sequences, which gives high similarities between them. The evolutionary analysis suggested that structural gene diversity is the primary source for the evolution of multigene families (Xu et al., 2012; Manzoor et al., 2021b).

C. lanatus contains 6–8 exons and *C. pepo* contains 3–9 exons; interestingly, *CpSOD3* contains 2 exons with 1 intron. *C. sativus* contains 5–8 exons while *CsaSOD9* has 3 exons and 2 introns. Both *CpSOD3* and *CsaSOD9* genes are the smallest genes among all 43 SOD genes identified in five Cucurbitaceae species. *C. melo* consists of 5–8 exons in which *MelSOD8* has 5 exons and 4 introns. *L. siceraria* contains 6–10 exons; interestingly, *LsiSOD1* and *LsiSOD*2 comprise 8/7 exons/introns. On the other hand, *LsiSOD3*, *LsiSOD5*, and *LsiSOD6* comprise 6/5 exons/introns that have the upstream and downstream sequence. *LsiSOD4* contains exons/introns (7/6) with upstream and downstream sequences and *LsiSOD7* contains 10 exons with the only downstream sequence. The remaining *LsiSOD8* contains 7 exons and 6 introns with both upstream and downstream sequences.

Chromosomal Distribution and Promoter Analysis of *SOD* Gene Family

The chromosomadal distribution of the *SOD* gene family of the Cucurbitaceae species was determined, as shown in **Figure 3** (**Supplementary Table S3**), and detailed data are given in **Supplementary Table S4**. *C. lanatus* chromosome 2 contains two genes, and chromosome 3 contains three genes, while chromosomes 4, 7, and 10 contain only one gene. In *C. pepo*, chromosome 0 contains three genes; 1, 2, 5, 11, and 20 contain only one gene while chromosome 8 contains 2 genes. In *C. sativus*, chromosome 1 contains three genes, and chromosomes 4 and 6 contain two genes each, while only one gene was found on chromosomes 2 and 3. In *L. siceraria*, chromosomes 1, 2, 10, and 11 contain one gene, while chromosomes 6 and 7 contain two genes, respectively. *C. melo* contains two genes on chromosome 2, while chromosomes 5, 6, 7, 8, 11, and 12 contain only one *SOD* gene.

To clearly understand the function and regulation of SOD proteins, *cis*-elements in the promoter sequence of *SOD* genes in Cucurbitaceae were retrieved. The 1,500-bp upstream region of the start codon of each *SOD* gene was analyzed by using the PlantCARE database. According to the obtained results, the *cis*-elements can be divided into three classes: stress related, light related, and hormone response elements (**Figure 4A**). Five *cis*-elements of stress response were determined, including LTR, TC rich, MBS, ARE, and box-w1; these elements reflected the response of plants to drought, low temperature, anaerobic induction, stress defense, and fungal inducer. Four hormone-sensitive *cis*-elements (SA, MeJA, GA, and ethylene) were identified, which are associated with ABA, SA, ethylene, and MeJA responses (**Figure 4B**). On the promoter region of *SOD* gene family of Cucurbitaceae, a considerable number of phyto-response *cis*-elements were detected (**Supplementary Table S4**). The results suggested that the *cis*-elements of *SOD* promoter had a positive response to abiotic stress and regulation mechanism of plant growth and development.

Gene Duplication Events and Collinearity Relationships in Five Cucurbitaceae Genomes

To further understand the evolution and new function of genes, gene duplication events of the *SOD* gene family were identified in five Cucurbitaceae species (*Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativus*, *Lagenaria siceraria*, *Cucumis melo*). We analyzed three modes of gene duplications in all *SOD* gene



FIGURE 1 | Phylogenetic tree of 106 SOD proteins from *Lagenaria siceraria* and other plants including 13 plant species—*Arabidopsis thaliana*, *Lagenaria siceraria*, *Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativa*, *Cucumis melo*, *Gossypium arboretum*, *Solanum lycopersicum*, *Zea mays*, *Solanum tuberosum*, *Sorghum bicolor*, *Populus trichocarpa*, and *Oryza sativa*. Protein sequences were aligned using ClustalW2 sequence alignment program and the phylogenetic tree was constructed by software MEGA7.

families, including transposed duplication (TRD), dispersed duplication (DSD), and whole-genome duplication (WGD). We identified many kinds of gene duplications and their contributions to the expansion of the *SOD* genes. Interestingly, 25 pairs of the duplicated gene were identified in five Cucurbitaceae species, with a maximal number of duplicated gene pairs derived from dispersed duplications (15 genes pair out of 25), followed by transposed duplications (8 genes pair out of 25) and whole-genome duplications (2 genes pair out of 25) showing that the expression of the *SOD* gene family was mainly

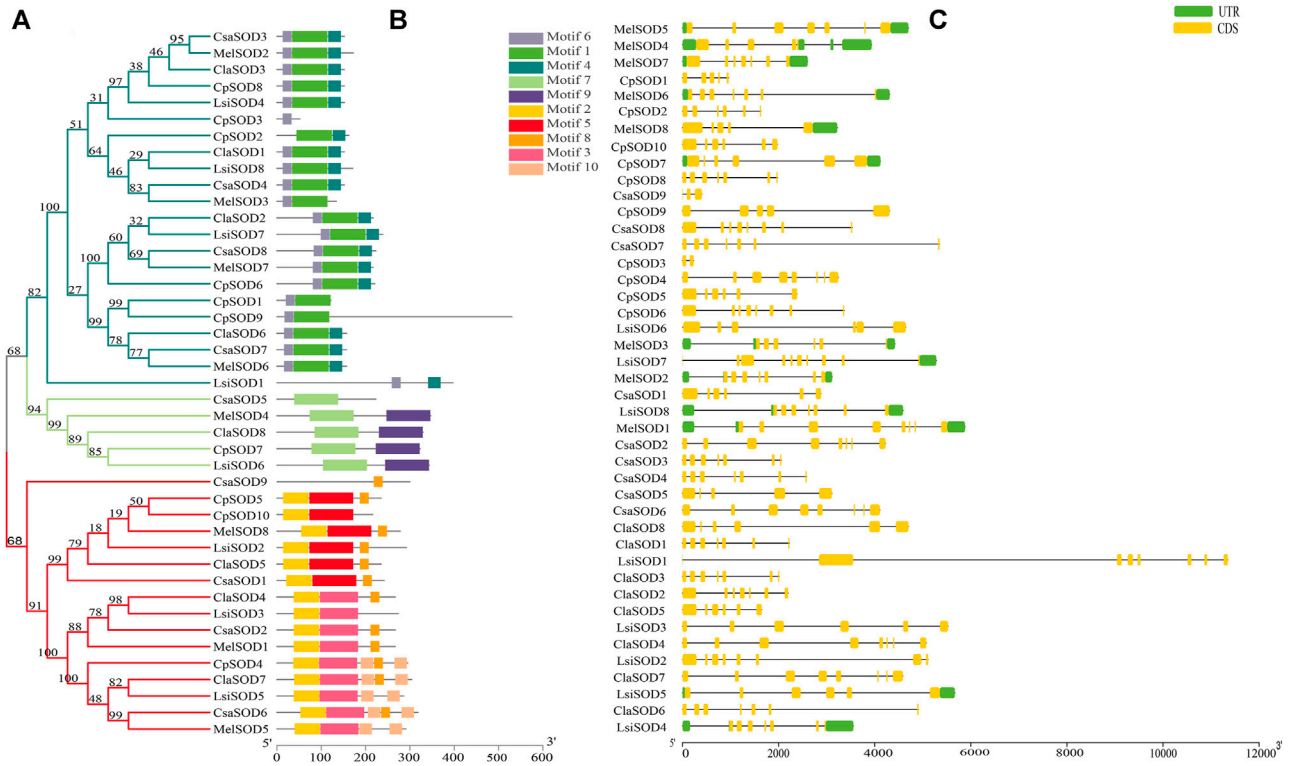


FIGURE 2 | (A) The protein structures of SODs based on the presence of conserved motifs were arranged corresponding to the phylogenetic tree. **(B)** All conserved motifs of the SOD proteins were identified by the MEME program. Different motifs are highlighted with different colored boxes with numbers 1 to 10. **(C)** Gene structures of SODs exons are shown as yellow boxes, introns are shown as thin black lines, and UTRs are shown as blue boxes.

associated with WGD, TRD, and DSD events (**Supplementary Table S5**) (**Figure 5**). These results indicate that DSDs play a vital role in *SOD* gene expansion in *Cucumis sativus*, *Lagenaria siceraria*, and *Cucumis melo*. WGDs might contribute to the expansion of the *SOD* gene family (**Supplementary Table S5**). Our study showed that duplication events play an important role in *SOD* gene expansion, and TRDs and WGD occurred at high frequency in Cucurbitaceae species (**Figure 6**).

We further analyzed the collinearity relationships of *SOD* genes between five Cucurbitaceae species (*Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativus*, *Lagenaria siceraria*, *Cucumis melo*) as these five plants belong to the Cucurbitaceae family and shared a similar antique (**Supplementary Table S6**) (**Figure 7**). A total of 35 collinear gene pairs were found between the five Cucurbitaceae genomes, including 5 orthologous gene pairs between *Arabidopsis* and *Cucurbita pepo*, 4 orthologous gene pairs between *Arabidopsis* and *Lagenaria siceraria*, 4 orthologous gene pairs between *Arabidopsis* and *Cucumis melo*, 8 orthologous gene pairs between *Citrullus lanatus* and *Cucumis sativus*, 7 orthologous gene pairs between *Citrullus lanatus* and *Lagenaria siceraria*, and 7 orthologous gene pairs between *Cucurbita pepo* and *Lagenaria siceraria*, suggesting a close relationship among the five Cucurbitaceae genomes. The results show that the genetic relationship between *SOD* gene pairs in *C. lanatus*, *C. pepo*, *C. sativus*, *C. melo*, and *L. siceraria* is close. No pairs of collinear *SODs* are shared between *Arabidopsis* and five Cucurbitaceae genomes, indicating the long distance phylogenetic relationship between two species.

Gene Ontology Annotation Study of *SOD* Gene Family

Functional enrichment analysis of *SOD* genes was performed using DAVID. The *SOD* genes were categorized into three functional groups—biological processes, molecular functions, and cellular components—that are characteristics of genes or gene products, which enable us to understand the diverse molecular functions of proteins (Altschul et al., 1990). These results may be related to protein sequence similarities caused by genomic events (8, **Supplementary Table S7**). Evaluation of the biological processes mediated by *SODs* indicated that the same percentage (~27%) of proteins was involved in oxidoreductase activity and ion binding. Among the SOD proteins, ~14.58% of members showed potential involvement in protein binding, enzyme binding, and DNA binding, respectively, while nucleic acid binding transcription factor activity involvement is ~2.08% during the 8 Lsi*SOD* genes of *Lagenaria siceraria* life cycle (**Figure 8**). At the same time, in *SOD* genes of five Cucurbitaceae species, "Cellular Components" includes 17.40% response to stress and reproduction, respectively. Meanwhile, 15.83% of genes were involved in aging, and 10.73% of genes were involved in homeostatic process and transport, respectively (**Figure 8**). Furthermore, the *SOD* genes of 8 Lsi*SOD* genes of

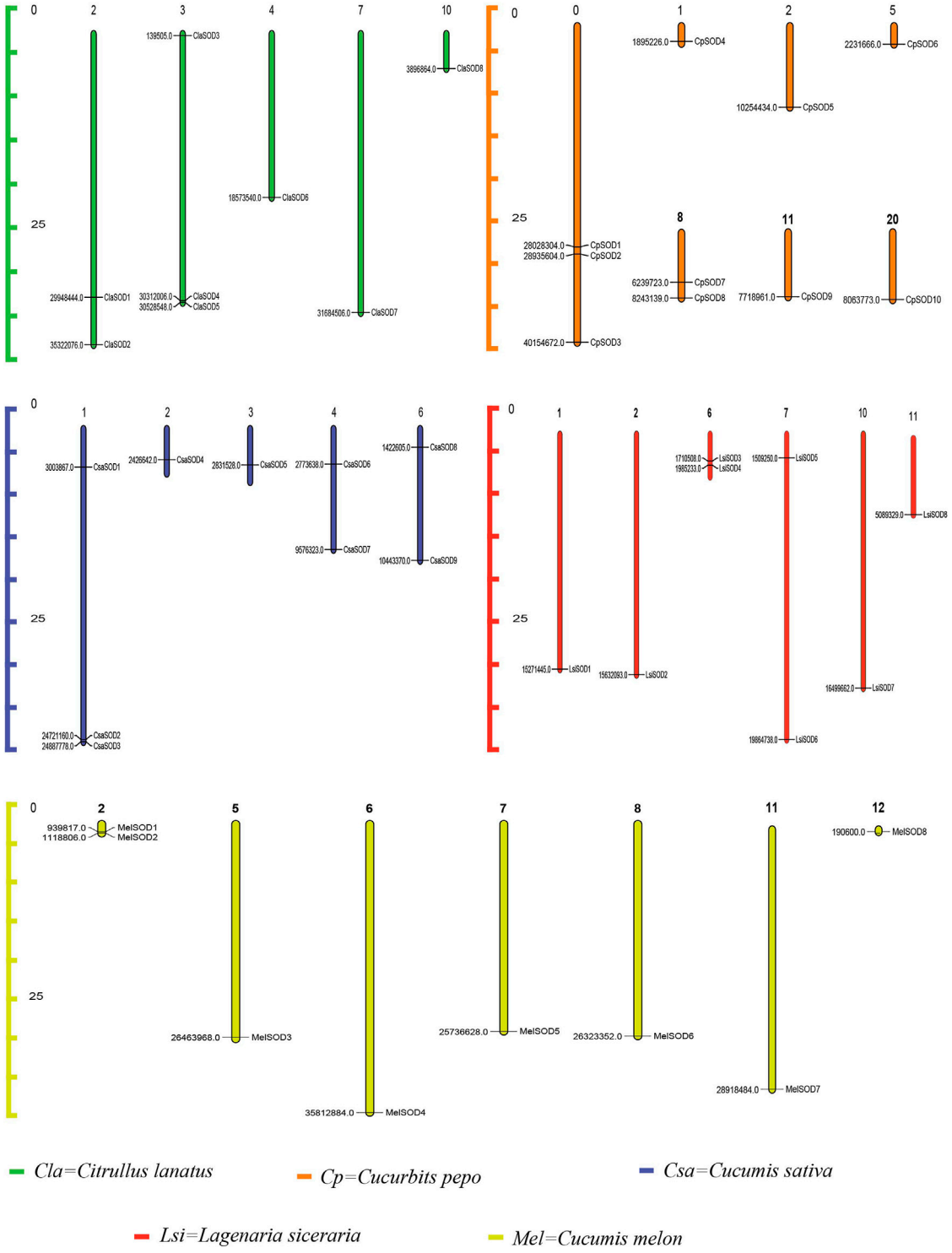


FIGURE 3 | Chromosomal locations of 43 SOD genes in the genome of five species of the Cucurbitaceae family. Watermelon: Cla; melon: Mel; cucumber: Csa; zucchini: Cp; bottle gourd: Lsi. The chromosome numbers are indicated at the top of chromosomes, and the size of the chromosome is represented with a vertical scale. The gene number is located on each chromosome and the left scale is in megabases (Mb).

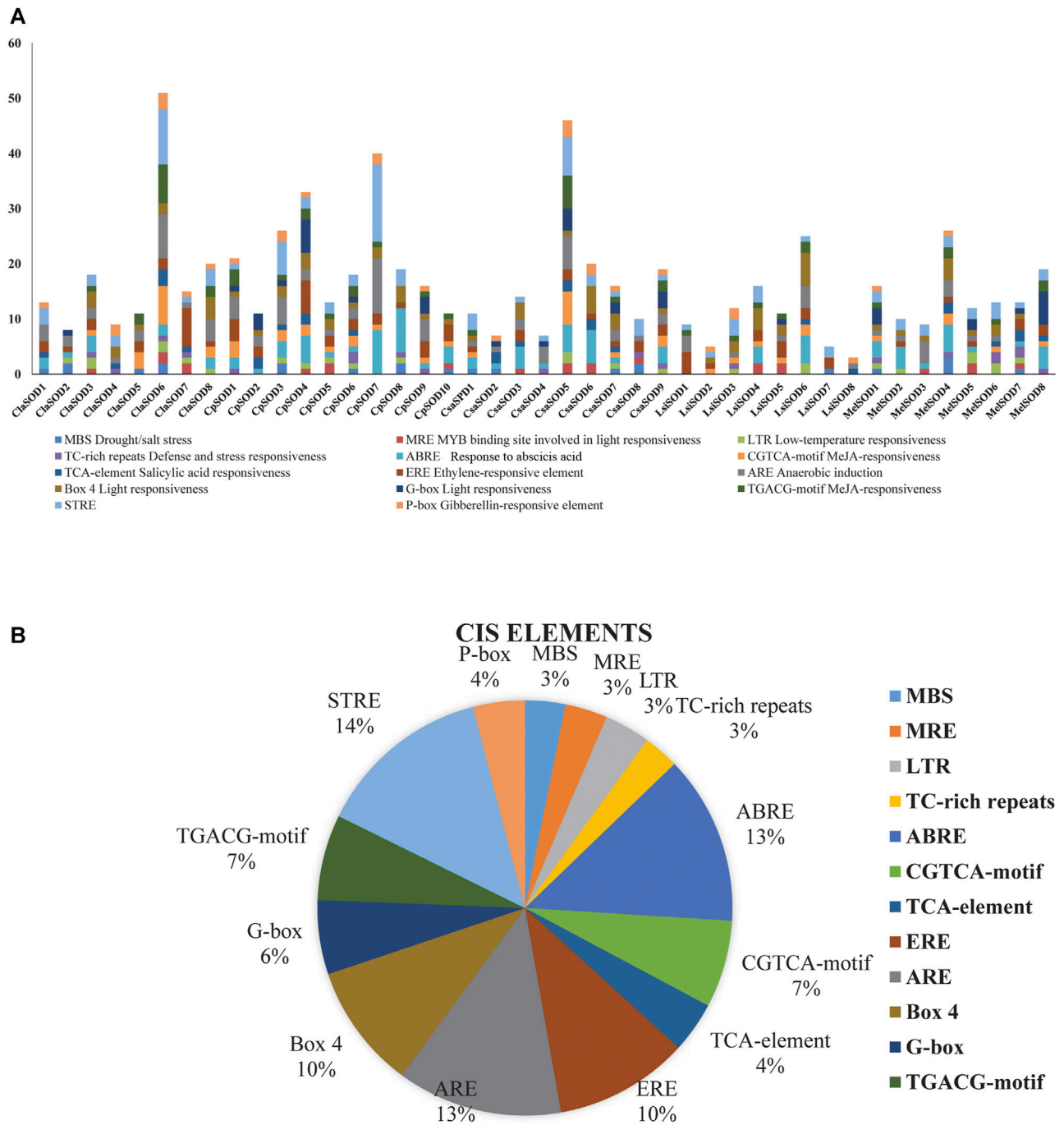


FIGURE 4 | Analysis of *cis*-elements in putative promoters of SOD gene family of five species of Cucurbitaceae family. 1,500 kb upstream of the transcription initiation site, sequences of SOD genes were retrieved and analyzed by PlantCARE. **(A)** Different stress-responsive and hormone-associated elements are identified, and their numbers are plotted on the graph. Various elements in each promoter are coded in different colors according to the legend at the top. **(B)** The size of the pie charts showed the percentage of promoter element in each category.

Lagenaria siceraria were involved in the molecular functions, 11.9% of the genes were involved in intracellular, cell, cytoplasm, organelle, and mitochondrion while 10.51% of the genes were involved in plastid and 5.65% were involved in extracellular region, cytosol, nucleus, and extracellular space, respectively (**Figure 8**). The results corroborated the putative *SOD* promoter analysis (Jagadeeswaran et al., 2009; Jia et al., 2009; Guan et al., 2013; Manzoor et al., 2021a). Furthermore, analysis of the molecular function annotations revealed that all of the *LsiSOD* protein functions were enriched in SOD activity and metal ion binding.

Genome-Wide Analysis of miRNA Targeting *LsiSOD* Genes

MicroRNAs (miRNAs) are endogenous microRNAs that play an important role in plant growth/development and stress responses. Thus, for a deep understanding of miRNA-mediated post-transcriptional regulation of *LsiSOD*, we identified 6 miRNAs (Lsi-MIR164a, Lsi-MIR164b, Lsi-MIR164c, Lsi-MIR164f, Lsi-miRN1740, and Lsi-miRN1741) with different classes of miRNA families (**Figure 9**; **Supplementary Table S8**). The analysis indicated that 3 *LsiSOD* genes (LsiSOD3, LsiSOD6, and LsiSOD7) were targeted on different miRNA. The expression profiles of these miRNAs and their targets are needed to detect and verify in further experiments to determine their biological functions in bottle gourd. The regulation of *SOD* genes' expression by miRNA needs to be studied further.

Tissue-Specific Expression of *LsiSOD* Genes

A strong link between gene expression and function has been suggested and that the SOD gene family is primarily involved in plant growth, development, and stress responses. To determine the biological functions in *Lagenaria siceraria*, the expression profiles of the 8 *LsiSOD* genes were analyzed in tissues (root, flower, fruit, stem, and leaf) using RNA-seq data downloaded from CuGenDB (http://cucurbitgenetics.org/) (Zheng et al., 2019). These results indicated that *LsiSOD* genes have tissue-specific expression patterns. However, *LsiSOD*4, *LsiSOD*5, and *LsiSOD*6 exhibited low transcript abundance in root and flower while their highest expression was detected in stems and leaves. This suggested that they play an important role in the development of plants. In addition, *LsiSOD1*, *LsiSOD2*, and *LsiSOD3* are highly expressed in all parts of plants especially in the early stages of plant development (**Figure 10**). All *LsiSOD* genes were widely expressed and showed tissue-specific expression patterns while *LsiSOD8* shows less expression in all different developmental parts.

Gene Expression Analysis of *LsiSOD* Gene Under Abiotic Stresses

To understand the role of the *SODs*, qRT-PCR was used to analyze the expression patterns of the *SOD* gene in response to stress from heat, cold, drought, and NaCl. Significant differences were observed in the expression levels of the *LsiSOD* genes under different treatments, and complexity was detected in their expression patterns (**Figure 11**). The 8 *LsiSOD* genes were induced by heat treatment, and their expression profiles were significantly enhanced. Among them, at 4 h of treatment one gene *LsiSOD7* (2.62-fold), at 8 h of treatment two genes *LsiSOD3* (20.35-fold) and *LsiSOD4* (14.62-fold), at 16 h of treatment two genes *LsiSOD*2 (13.22-fold) and *LsiSOD*3 (16.41-fold), and at 24 h 2 genes *LsiSOD2* (18.24-fold) and *LsiSOD3* (24.16-fold) reached the highest expression level respectively. Interestingly, the transcription of *LsiSOD*3 and *LsiSOD*4 was gradually induced to 8 h, then decreased at 16 h, and finally reached a high level at 24 h, as described in **Figure 11**. Besides, in the expression of *LsiSOD*7, a dramatic increase was observed, which reached the maximum level at 4 h and decreased significantly at 8 h, then increased gradually with the treatment (**Figure 11A**). During cold treatment at 4 h, many *LsiSOD* genes were downregulated, *LsiSOD*3 (1.17-fold) remained unchanged, *LsiSOD*7 (1.27-fold) slightly increased, and then *LsiSOD*3 (7.27-fold) increased significantly at 8 h (**Figure 11B**). After 8 h, except for the increased expression of *LsiSOD*2 (1.71-fold) at 16 h, the other *LsiSOD* genes were significantly downregulated at 16 h and continued to decline at 24 h as compared with 0 h. After 8 h, the expression of *LsiSOD*7 (4.53-fold) and *LsiSOD*8 (4.39-fold) was progressively induced, and the induction peak appeared at 16 and 24 h, respectively. Under the PEG stress treatment, almost all *LsiSOD* gene expressions increased significantly at 4 h except *LsiSOD1*, and then the level of transcription decreased. At the same time, *LsiSOD1* decreased somewhat at 4 h, and its transcription remained unchanged (**Figure 11C**). At 4 h of treatment with PEG, the expression of *LsiSOD8* (4.27-fold) was maximum, whereas the expression of *LsiSOD8* (1.36-fold) was highest during 8 h of treatment, while at 12 h of treatment *LsiSOD4* (2.17-fold) shows maximum expression. During NaCl stress treatment, mostly the expression pattern of all *LsiSOD* genes increased tremendously at 8 h, decreased at 16 h, and finally increased at 24 h (**Figure 11D**). At 4 h of treatment, expression of *LsiSOD*3 increased (3.71-fold), and at 8 h of treatment with NaCl, expression of *LsiSOD2* increased (8.59-fold), while at 12 h of treatment, the expression of *LsiSOD*8 increased (3.01-fold) and the expression of *LsiSOD4* increased (4.53-fold). It should be worth mentioning that *LsiSOD*2 transcription level increased significantly in 24 h, which was significantly higher than other *LsiSOD* genes.

DISCUSSION

Environmental pressure poses a considerable challenge to crop production. Superoxide dismutase activity plays an important role in plant resistance to different stresses, such as salt, drought, and metal toxicity (Alscher et al., 2002; Schützendübel and Polle, 2002; Fernández-Ocaña et al., 2011; Atkinson et al., 2013; Lin and Lai, 2013; Wang et al., 2018). *SOD* gene family is widely distributed among many plant species, including *Arabidopsis* (Kliebenstein et al., 1998), rice (Dehury et al., 2013), longan (Lin and Lai, 2013), banana (Molina-Rueda et al., 2013), poplar (Gosavi et al., 2014), sorghum (Nath et al., 2014), cotton (Feng et al., 2015), and tomato (Filiz and Tombuloğlu, 2015; Zhang et al., 2016; Wang et al., 2016; Wang et al., 2017). Our study provides a systematic and comprehensive whole-genome evolutionary analysis of SOD members obtained from five Cucurbitaceae genomes while the focus of our study is on *Lagenaria siceraria*, which is one of the excellent vegetables that contain all the essential ingredients necessary for human health and quality of life (Division and Jambheshwar, 2011). *SODs* are identified as a crucial enzyme involved in many oxidation processes and protect the plant against ROS

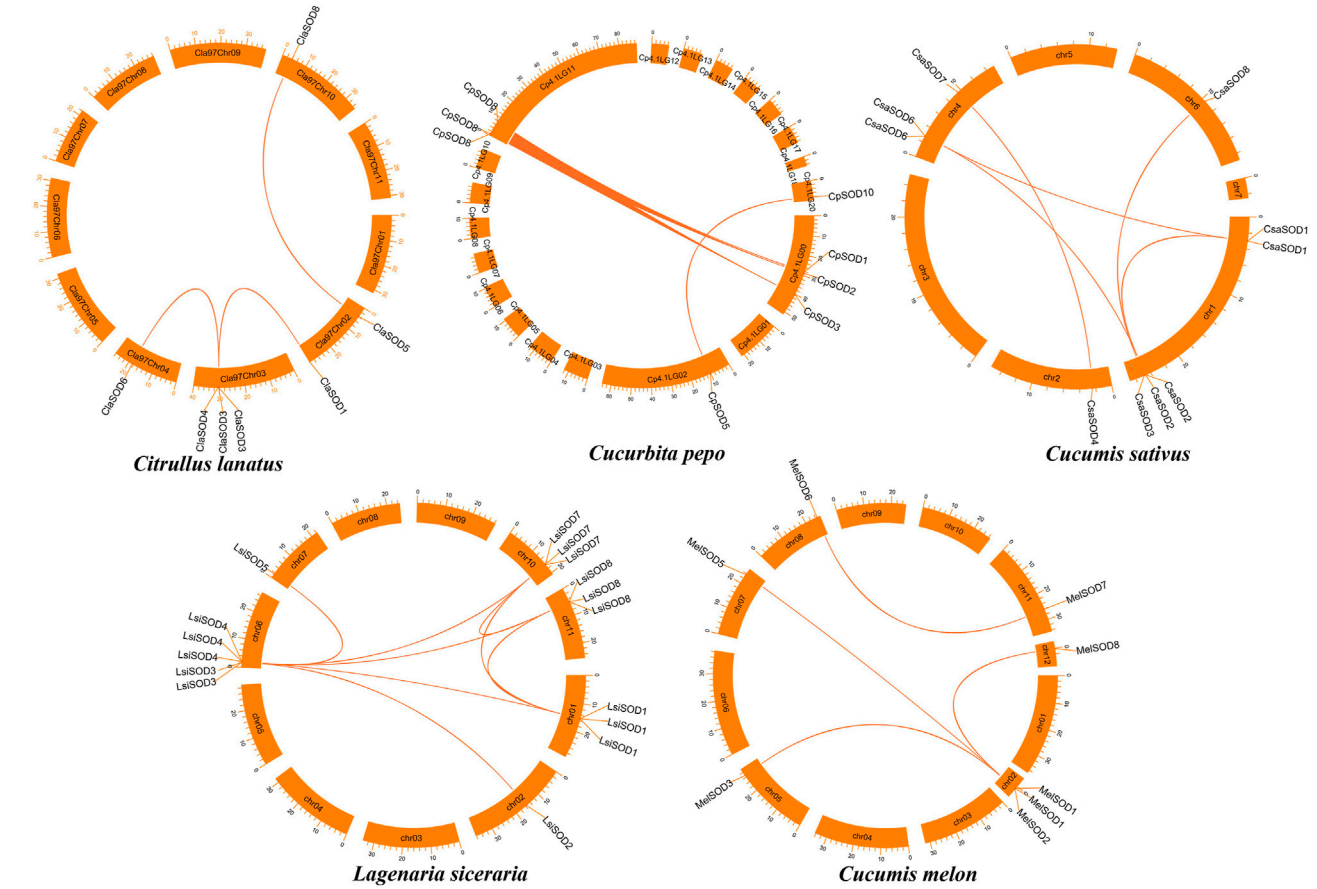


FIGURE 5 | Chromosomal localization and duplication events of five Cucurbitaceae genomes and gene pairs are joined with a colored line.

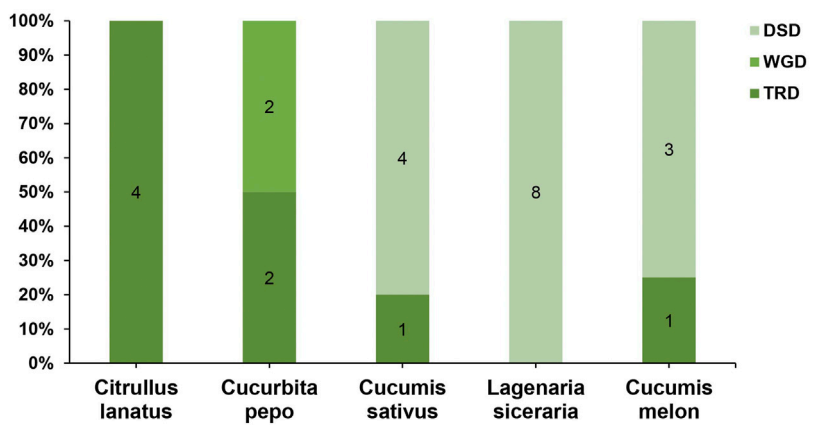


FIGURE 6 | Comparison of gene duplication events in five Cucurbitaceae genomes. WGD, whole-genome duplicates; TRD, transposed duplicates; DSD, dispersed duplicates.

(Alscher et al., 2002). Therefore, we systematically analyzed the *SOD* gene family of Cucurbitaceae and determined the gene expression patterns in *L. siceraria* under different abiotic stresses (drought, heat, salt, and cold).

A phylogenetic study was conducted on SOD proteins in *Lagenaria siceraria* and four other cucurbit plants, namely, *Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativus*, and *Cucumis melo*. *SODs* could be classified into eight groups with

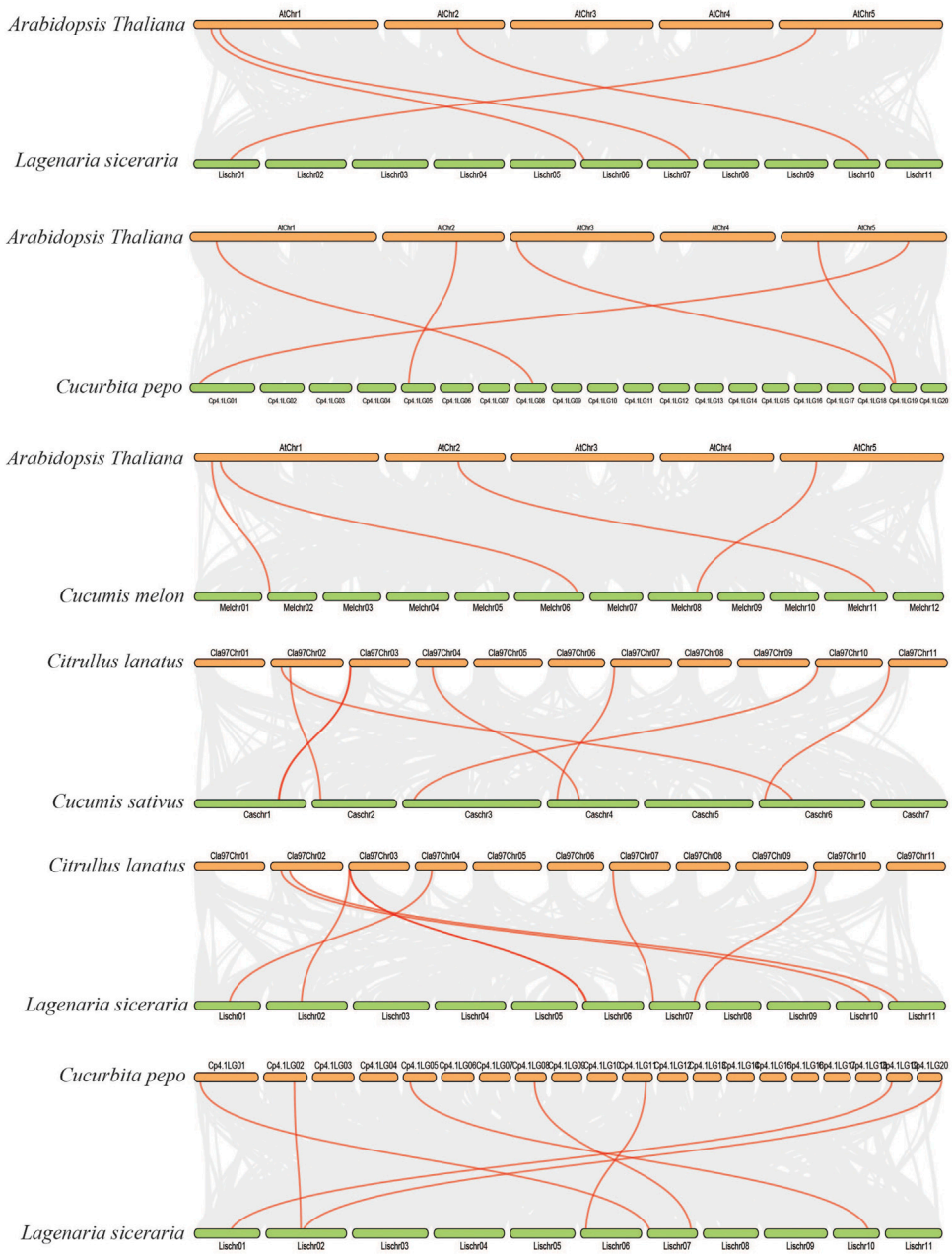


FIGURE 7 | Collinearity relationships in five Cucurbitaceae genomes and *Arabidopsis thaliana*; different line color represents synteny relationships in the five Cucurbitaceae genomes.

good statistical support in line with previous studies (Alscher et al., 2002; Feng et al., 2015; Filiz and Tombuloğlu, 2015; Feng et al., 2016; Wang et al., 2017; Verma et al., 2019). *SOD* genes are localized to chloroplasts, and mitochondria were classified into the same subgroups. Furthermore, nearly all *LsiSOD*s are systematically grouped with at least one member of all plant species under consideration, highlighting the functional conservation of *LsiSOD*s with other plant species *SODs* (**Figure 1**).

In this study, *SOD* gene families were found in Cucurbitaceae genome, including *Citrullus lanatus*, *Cucurbita pepo*, *Cucumis sativus*, *Lagenaria siceraria,* and *Cucumis melo*. There are considerable differences in the size of genomes and the number of *SODs* in these plant organisms, but there is no substantial dependence on the size of genomes. The number of amino acids of SOD proteins ranged from 52 to 530 kDa, showing a significant variation. The difference in the number of *SOD* genes in different plant species may be due to gene duplication. In addition, the difference between clades might be related to different functions and diversity of exons/introns and conserved motif structure. Intron and exon variations play a

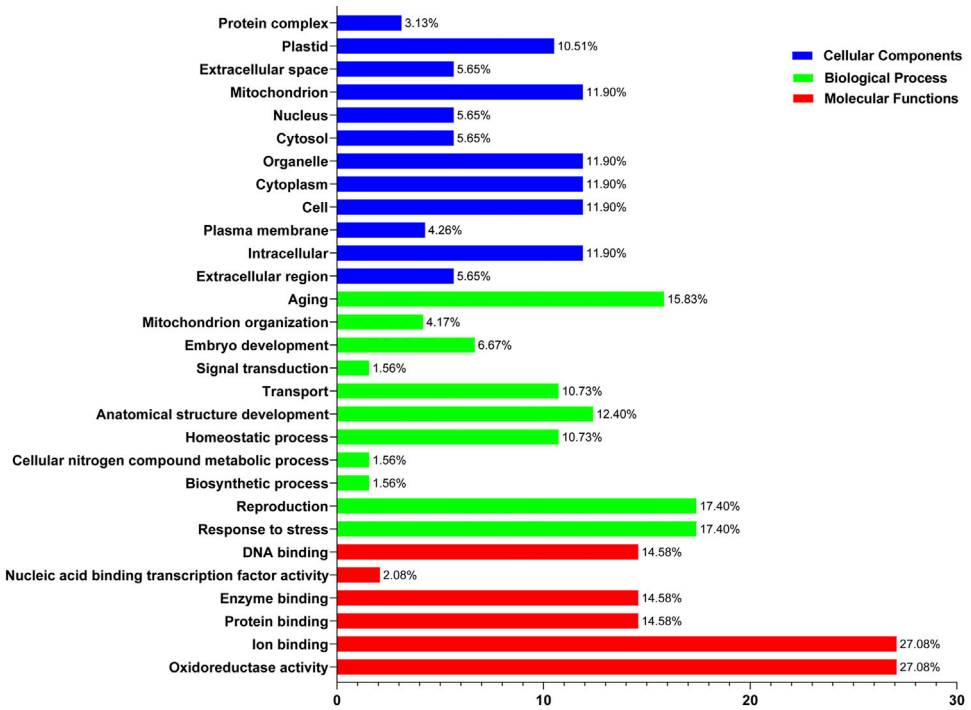


FIGURE 8 | Gene Ontology (GO) annotation results of SOD genes of five species in the Cucurbitaceae family. GO analysis of SOD protein sequences is analyzed for their involvement or function in three important categories: biological process, molecular function, and cellular component.

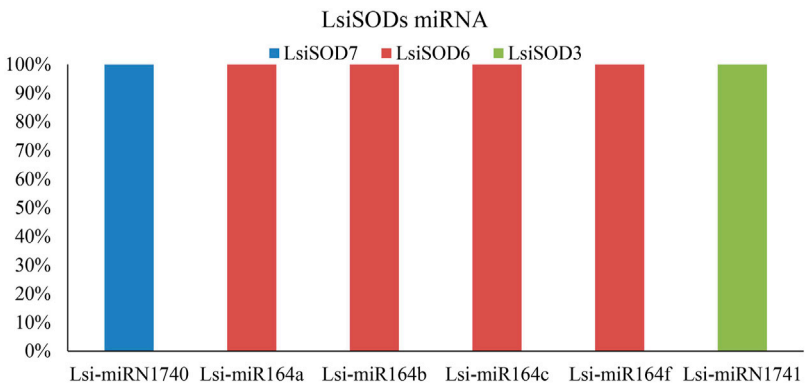


FIGURE 9 | SODs targeted by miRNAs of *Lagenaria siceraria* (bottle gourd). The RNA sequence of each complementary site from 5′ to 3′ and the predicted miRNA sequence from 3′ to 5′ are shown in the expanded regions.

major role in the evolution of different genes. The variation in introns/exons and motif structure of *SOD* genes suggests a high level of complexity between Cucurbitaceae species.

We further investigated the diversity between *SODs* by analyzing their subcellular localization as localization plays a vital role in their functions. The cytoplasm (34%), chloroplast (33%), mitochondria (22%), and nucleus (11%) *SOD* gene localizations were predicted in five Cucurbitaceae species (**Supplementary Table S8**), and further study is needed to confirm those localizations. Increasing evidence points to an important role of miRNAs in stress tolerance. Several studies have reported that miRNAs regulate the expression of stress-responsive protein-coding genes at post-transcriptional level, showing a reverse correlation between the miRNA and the target expression (Sunkar et al., 2007; Lomakina, 2018; Manzoor et al., 2020). These miRNAs resulted from computational predictions and deep sequencing, and they are involved in some biological processes reported in plants, including responses to environmental stresses and regulating cell growth, development, and metabolism (Qiu et al., 2007; Li

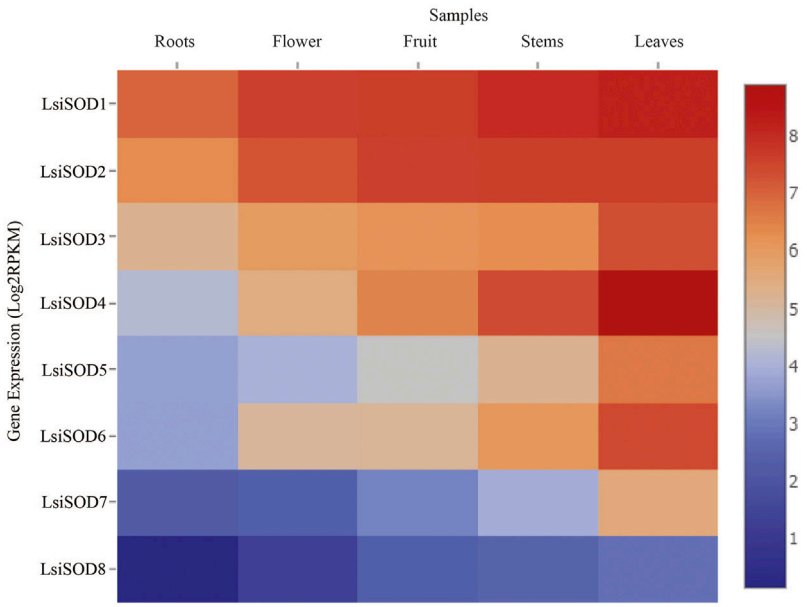


FIGURE 10 | Relative expression profiles of SOD genes of five species in Cucurbitaceae family during different developmental stages. The FPKMs calculated by RNA-seq data are shown as a heat map. The color scale is shown at the side of the heat map reflecting the higher and lower relative abundance in each tissue.

et al., 2012). In the present study, we identified 3 miRNAs targeting 8 *LsiSOD* genes (**Figure 9**; **Supplementary Table S8**). In *Lagenaria siceraria*, the miR164 family comprises 3 members that generate four mature products, miR164a/b/c and miR164f, which could target at least three LsiSOD genes. The miR164 family is a highly conserved miRNA that has been found in many plant species. miRNA164a thereby increased the tolerance of the plant to the abiotic stress or increasing the biomass, vigor, or yield of the plant. The miRNA miR164 plays a central role during the development of serrated leaf margins in *Arabidopsis*. In this study, transcripts of three miRNA families were identified because miRNA expression is highly regulated under different abiotic stress conditions (**Figure 9**). Further studies are needed to determine the role of *SODs* in *Lagenaria siceraria.*

Gene duplications are an essential mechanism for creating genetic novelty in all plants, which could help organisms to adapt to environmental change (Alvarez-Buylla et al., 2000; Kondrashov, 2012; Manzoor et al., 2020). There are different kinds of gene duplication, including dispersed duplication (DSD), transposed duplication (TRD), and whole-genome duplication (WGD), which differentially contribute to the expansion of plant-specific genes (Qiao et al., 2018, Qiao et al., 2019). Gene replication can cause variation in the number of *SOD* genes in different plant species, and involves tandem and segmented replication, which plays an important role in *SOD* gene diversity, expansion, and duplication (Filiz and Tombuloğlu, 2015; Wang et al., 2016; Zhang et al., 2016; Wang et al., 2017). Thus, tandem repetition is likely to play a vital function in the amplification of the *LsiSOD*s, for example, the two neighboring genes *LsiSOD*3 and *LsiSOD*6 on Chr6 (**Figure 3**). Characterization of the gene structure describes that the number of introns from the *SOD* gene family differed between the observed Cucurbitaceae species (**Figure 2**); in addition, the number of introns in *L. siceraria* was 6–8. Previous studies have shown that the plant's *SOD* gene has a strongly conserved intron pattern and that seven introns are present in most cytoplasmic and chloroplast SODs (Fink and Scandalios, 2002). In our study, only the *LsiSOD* members were predicted to contain seven introns, as shown in **Figure 2**.

Analysis of the *cis*-elements in SOD gene promoters resulted in the detection of three major types of *cis*-elements associated with light, abiotic stress, and hormone response as well as *cis*-elements related to developmental processes and tissue-specific expression. A relatively large number of light-responsive *cis*-elements were detected in *SOD* gene promoters, suggesting that SODs might participate in the abiotic response. Some studies have shown that SOD genes are involved in the response to abiotic stress in different plants like *Dendrobium catenatum*, *Pennisetum glaucum*, maize, and *Arabidopsis* (Divya et al., 2019; Huang et al., 2020; Kliebenstein et al., 2020; Liu et al., 2021). In addition, a series of *cis*-elements related to abiotic stress responses were identified in SOD gene promoters, such as MBS, ERE, TC-rich repeats, ARE, ABRE, and Box-4, which may regulate gene expression under various stresses. Most of the SOD genes in *Arabidopsis*, banana, rice, tomato, poplar, cotton, and other different plants can be induced in response to various abiotic stresses such as heat, cold, drought, and salinity (Kliebenstein et al., 1998; Lee et al., 2007; Dehury et al., 2013; Nath et al., 2014; Feng et al., 2016; Wang et al., 2016; Zhang et al., 2016; Wang et al., 2017; Verma et al., 2019). GO annotation analysis further verified these results (**Supplementary Table S7**). In addition, some researchers have reported similar findings in different crops and found that *SOD* gene plays an important role

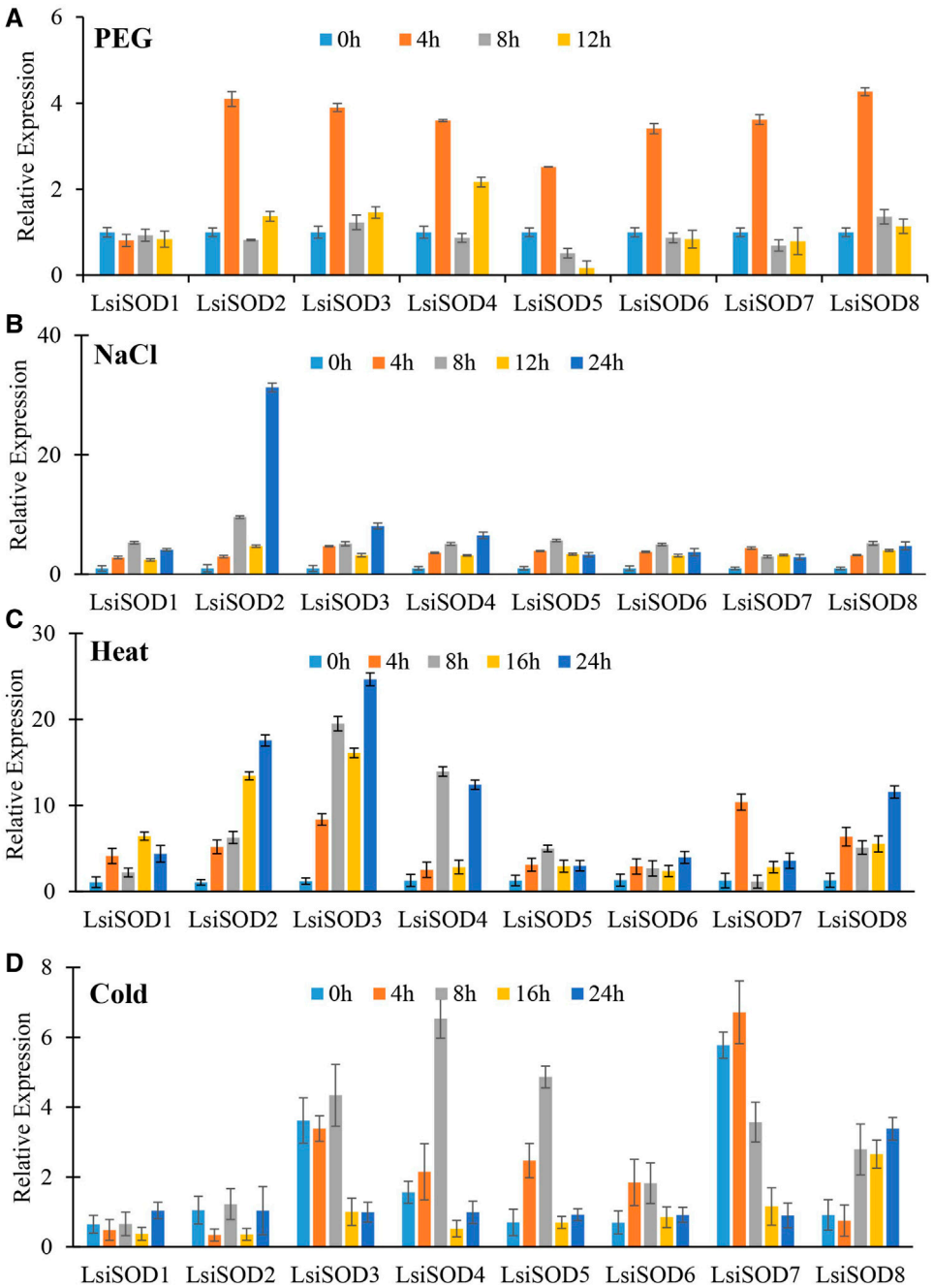


FIGURE 11 | qRT-PCR study of SOD genes of bottle gourd in response to different abiotic stresses including **(A)** heat, **(B)** cold, **(C)** PEG, and **(D)** NaCl at different time points (0, 4, 8, 16, 24 h). Error bars indicate ±SD based on three biological replicates.

under different stress conditions. GO annotation results confirmed the *LsiSOD*'s role in response to different stress stimuli, cellular oxidant detoxification processes, metal ion binding activities, SOD activity, and different cellular components. These results can promote our understanding of *LsiSOD* genes under different environmental conditions. We also found protein similarities (**Supplementary Table S9**) by using protein–protein interactions (PPIs) handle a wide range of biological processes, including metabolic and developmental control and cell-to-cell interactions (Mittler, 2017). Regarding the possible role of *SODs* from Cucurbitaceae, the expression profile of *SODs* in different tissues was analyzed based on sequence evidence from RNA-seq data. Data analysis suggests that the gene family of *SOD* was expressed in all tissues; some *SODs* had tissue-specific expression patterns (**Figure 10**).

Expression Patterns of *SOD* Gene Family of *Lagenaria siceraria* Under Heat, Cold, Drought, and Salt Stress

Excess ROS resulting from abiotic stress can pose threats to *L. siceraria* yield. The SODs play an active role in ROS removal from plants caused by various abiotic stresses (Gill and Tuteja, 2010). However, *LsiSOD*s' specific response to heat, cold, drought, and salt is not well understood. Therefore, the analyses of qRT-PCR provide important clues for understanding the possible role of *LsiSOD*s under various stresses. Based on the evaluation of *cis*-elements of *LsiSOD* gene promoter, three main *cis*-element types related to light, abiotic stress, and hormone response, as well as *cis*-elements related to development process, were determined. The *cis*-elements in *LsiSODs* promoter include TC-rich motif, LTR motif, MBS motif, ARE motif, and ABRE motif, which is the evidence of abiotic stress response (**Supplementary Table S4**). These motifs were previously observed in different plant species to cope with abiotic stresses, such as bananas, tomatoes, Brassica, tobacco, and millet (Oppenheim et al., 1996; Kurepa et al., 1997; Kliebenstein et al., 1998; Pilon et al., 2011; Feng et al., 2015; Wang et al., 2018; Hu et al., 2019; Verma et al., 2019). In this study, the expression level of eight *LsiSOD* genes also changed significantly under different stresses, indicating that these genes play an important regulatory role in stress response and may have a certain functional relationship. Almost all *LsiSOD* genes were upregulated during heat treatment, and some displays have similar expression (**Figure 11C**).

Under cold stress, the expression level of all *LsiSOD* genes is unregulated significantly, and their patterns of expression are different from each other, as shown in **Figure 11D**, which means that *LsiSOD* genes under cold stress may have functional diversity. Similar expression patterns were observed in *LsiSOD1* with *MaCSD1B* and *MaCSD2B* (Feng et al., 2015). In PEG treatment, almost all *Lsi-SOD1* genes have identical expression patterns, reach the highest level at 4 h, and then decline, as shown in **Figure 11A**, meaning that the function of the *LsiSOD* gene is related to drought.

The expression of most of the *LsiSOD* genes changed during NaCl treatment; *LsiSOD*2 is the only one of the eight *LsiSOD* genes with a significant increase, as shown in **Figure 11B**, indicating that *LsiSOD*2 plays an active role in detoxification of ROS during salt stress. Similar results were observed during salt stress in tomato (Feng et al., 2016). *SlSOD1* and *LsiSOD*2 were grouped in the same subgroup, showing approximately high amino acid sequence homology. Altogether, we conclude that the *LsiSOD* gene plays a specific role in ROS removal induced by abiotic stress, which enhances plant adaptability to stress. In addition, some *LsiSOD* genes were correlated to abiotic stress exhibiting different expression patterns. For example, cold treatment reduced the expression of *LsiSOD1*, heat treatment and NaCl treatment increased the expression of *LsiSOD1*, while under PEG stress, the expression level remained the same without significant change, as shown in **Figure 11**, which means that the function of *LsiSOD1* in different signaling pathways was different. It needs to be further explored to elucidate the role of the *LsiSOD* gene in *L. siceraria* under various abiotic stresses.

CONCLUSION

In this study, we identified *SOD* genes from the Cucurbitaceae family and analyzed their genomic structure, GO annotation analysis (molecular processes, biological functions, and cellular components), miRNA, gene duplication events (TD, PD, DSD, WGD, and TD), conserved motif patterns, phylogenetic relationships, mode of gene duplications, subcellular localization, and RNA-seq data analysis. qRT-PCR was used to evaluate the *LsiSOD* gene regulatory response under a variety of abiotic stresses, such as heat, cold, PEG, and NaCl. This research provided insight and further functional identification of the Cucurbitaceae family of *SOD* genes and laid the framework for understanding the molecular mechanism of the *SOD* gene in response to stress and plant growth. Genome-wide study of *SOD* genes provides insights into the evolutionary history and has laid a foundation for gene role, functional characteristics, and molecular mechanism in the plant development process and stress response (Mistry et al., 2021).

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: CuGenDB, PRJNA387615.

AUTHOR CONTRIBUTIONS

SR performed the main experiments. SR and AR prepared the article draft. MAM helped in bioinformatic tools and methodology. LL, WS, MWR, and DL reviewed the article.

FUNDING

This research was supported by the National Natural Science Foundation of China (NSFC) (31570650) and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

ACKNOWLEDGMENTS

Thank full for this grant from National Natural Science Foundation of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.784878/full#supplementary-material

REFERENCES

Abreu, I. A., and Cabelli, D. E. (2010). Superoxide Dismutases-A Review of the Metal-Associated Mechanistic Variations. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1804, 263–274. doi:10.1016/j.bbapap.2009.11.005

Ahmad, P., Umar, S., and Sharma, S. (2010). Mechanism of Free Radical Scavenging and Role of Phytohormones in Plants under Abiotic Stresses. *Plant Adaptation and Phytoremediation*, 99–118. doi:10.1007/978-90-481-9370-7_5

Alscher, R. G., Erturk, N., and Heath, L. S. (2002). Role of Superoxide Dismutases (SODs) in Controlling Oxidative Stress in Plants. *J. Exp. Bot.* 53, 1331–1341. doi:10.1093/jexbot/53.372.1331

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., et al. (2000). An Ancestral MADS-Box Gene Duplication Occurred before the Divergence of Plants and Animals. *Proc. Natl. Acad. Sci.* 97, 5328–5333. doi:10.1073/pnas.97.10.5328

Asensio, A. C., Gil-Monreal, M., Pires, L., Gogorcena, Y., Aparicio-Tejo, P. M., and Moran, J. F. (2012). Two Fe-Superoxide Dismutase Families Respond Differently to Stress and Senescence in Legumes. *J. Plant Physiol.* 169, 1253–1260. doi:10.1016/j.jplph.2012.04.019

Atkinson, N. J., Lilley, C. J., and Urwin, P. E. (2013). Identification of Genes Involved in the Response of Arabidopsis to Simultaneous Biotic and Abiotic Stresses. *Plant Physiol.* 162, 2028–2041. doi:10.1104/pp.113.222372

Baek, K.-H., and Skinner, D. Z. (2003). Alteration of Antioxidant Enzyme Gene Expression during Cold Acclimation of Near-Isogenic Wheat Lines. *Plant Sci.* 165, 1221–1227. doi:10.1016/S0168-9452(03)00329-7

Bafana, A., Dutt, S., Kumar, S., and Ahuja, P. S. (2011). Superoxide Dismutase: An Industrial Perspective. *Crit. Rev. Biotechnol.* 31, 65–76. doi:10.3109/07388551.2010.490937

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

Ballal, A., and Manna, A. C. (2009). Regulation of Superoxide Dismutase (Sod) Genes by SarA in *Staphylococcus aureus*. *J. Bacteriol.* 191, 3301–3310. doi:10.1128/JB.01496-08

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

Conesa, A., and Götz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int. J. Plant Genomics* 2008, 1–12. doi:10.1155/2008/619832

Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of Abiotic Stress on Plants: a Systems Biology perspective.Pdf (Application/pdf Object), *BMC Plant Biol.*, 11, 1–14. Available at: http://www.biomedcentral.com/content/pdf/1471-2229-11-163.pdf.doi:10.1186/1471-2229-11-163

Cui, L., Huang, Q., Yan, B., Wang, Y., Qian, Z., Pan, J., et al. (2015). Molecular Cloning and Expression Analysis of a Cu/Zn SOD Gene (BcCSD1) from Brassica Campestris Ssp. Chinensis. *Food Chem.* 186, 306–311. doi:10.1016/j.foodchem.2014.07.121

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. doi:10.1093/nar/gky316

Dehury, B., Sarma, K., Sarmah, R., Sahu, J., Sahoo, S., Sahu, M., et al. (2013). In Silico analyses of Superoxide Dismutases (SODs) of rice (Oryza Sativa L). *J. Plant Biochem. Biotechnol.* 22, 150–156. doi:10.1007/s13562-012-0121-6

Division, P., and Jambheshwar, G. (2011). Is Bottle Gourd a Natural Guard. *Int. Res. J. Pharm.* 2, 13–17.

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

Feng, K., Yu, J., Cheng, Y., Ruan, M., Wang, R., Ye, Q., et al. (2016). The SOD Gene Family in Tomato: Identification, Phylogenetic Relationships, and Expression Patterns. *Front. Plant Sci.* 7, 1–12. doi:10.3389/fpls.2016.01279

Feng, X., Lai, Z., Lin, Y., Lai, G., and Lian, C. (2015). Genome-wide Identification and Characterization of the Superoxide Dismutase Gene Family in Musa Acuminata Cv. Tianbaojiao (AAA Group). *BMC Genomics* 16, 1–16. doi:10.1186/s12864-015-2046-7

Fernández-Ocaña, A., Chaki, M., Luque, F., Gómez-Rodríguez, M. V., Carreras, A., Valderrama, R., et al. (2011). Functional Analysis of Superoxide Dismutases (SODs) in sunflower under Biotic and Abiotic Stress Conditions. Identification of Two New Genes of Mitochondrial Mn-SOD. *J. Plant Physiol.* 168, 1303–1308. doi:10.1016/j.jplph.2011.01.020

Fink, R. C., and Scandalios, J. G. (2002). Molecular Evolution and Structure-Function Relationships of the Superoxide Dismutase Gene Families in Angiosperms and Their Relationship to Other Eukaryotic and Prokaryotic Superoxide Dismutases. *Arch. Biochem. Biophys.* 399, 19–36. doi:10.1006/abbi.2001.2739

Fïlïz, E., Koç, İ., and Tombuloğlu, H. (2014). Genome-wide Identification and Analysis of Growth Regulating Factor Genes in Brachypodium Distachyon: In Silico Approaches. *Turk J. Biol.* 38, 296–306. doi:10.3906/biy-1308-57

Fïlïz, E., and Tombuloğlu, H. (2015). Genome-wide Distribution of Superoxide Dismutase (SOD) Gene Families in Sorghum Bicolor. *Turk J. Biol.* 39, 49–59. doi:10.3906/biy-1403-9

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. *The Proteomics Protoc. Handbook*, 571–607. doi:10.1385/1-59259-890-0:571

Gill, S. S., and Tuteja, N. (2010). Reactive Oxygen Species and Antioxidant Machinery in Abiotic Stress Tolerance in Crop Plants. *Plant Physiol. Biochem.* 48, 909–930. doi:10.1016/j.plaphy.2010.08.016

Gómez, J. M., Hernández, J. A., Jiménez, A., Del Río, L. A., and Sevilla, F. (1999). Differential Response of Antioxidative Enzymes of Chloroplasts and Mitochondria to Long-Term NaCl Stress of Pea Plants. *Free Radic. Res.* 31, 11–18. doi:10.1080/10715769900301261

Gosavi, G. U., Jadhav, A. S., Kale, A. A., Gadakh, S. R., Pawar, B. D., and Chimote, V. P. (2014). Effect of Heat Stress on Proline, Chlorophyll Content, Heat Shock Proteins and Antioxidant Enzyme Activity in Sorghum (Sorghum Bicolor) at Seedlings Stage. *Indian J. Biotechnol.* 13, 356–363.

Guan, Q., Lu, X., Zeng, H., Zhang, Y., and Zhu, J. (2013). Heat Stress Induction ofmiR398triggers a Regulatory Loop that Is Critical for Thermotolerance in Arabidopsis. *Plant J.* 74, 840–851. doi:10.1111/tpj.12169

Han, L.-m., Hua, W.-p., Cao, X.-y., Yan, J.-a., Chen, C., and Wang, Z.-z. (2020). Genome-wide Identification and Expression Analysis of the Superoxide Dismutase (SOD) Gene Family in Salvia Miltiorrhiza. *Gene* 742, 144603. doi:10.1016/j.gene.2020.144603

Hasanuzzaman, M., Bhuyan, M. H. M. B., Zulfiqar, F., Raza, A., Mahmud, J., Fujita, M., et al. (2020). Reactive Oxygen Species and Antioxidant Defense in Plants under Abiotic Stress : Revisiting the Crucial Role of a Universal Defense Regulator. *Antioxidants* 9, 681. doi:10.3390/antiox9080681

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

Ho, K.-H., Chen, P.-H., Hsi, E., Shih, C.-M., Chang, W.-C., Cheng, C.-H., et al. (2017). Identification of IGF-1-Enhanced Cytokine Expressions Targeted by miR-181d in Glioblastomas via an Integrative miRNA/mRNA Regulatory Network Analysis. *Sci. Rep.* 7, 1–13. doi:10.1038/s41598-017-00826-0

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* 35, W585–W587. doi:10.1093/nar/gkm259

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

Hu, L., Yang, Y., Jiang, L., and Liu, S. (2016). The Catalase Gene Family in Cucumber: Genome-wide Identification and Organization. *Genet. Mol. Biol.* 39, 408–415. doi:10.1590/1678-4685-gmb-2015-0192

Hu, X., Hao, C., Cheng, Z.-M., and Zhong, Y. (2019). Genome-wide Identification, Characterization, and Expression Analysis of the grapevine Superoxide Dismutase (SOD) Family. *Int. J. Genomics* 2019, 1–13. doi:10.1155/2019/7350414

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* 37, 1–13. doi:10.1093/nar/gkn923

Huang, H., Wang, H., Tong, Y., and Wang, Y. (2020). Insights into the Superoxide Dismutase Gene Family and its Roles in Dendrobium Catenatum under Abiotic Stresses. *Plants* 9 (11), 1452. doi:10.3390/plants9111452

Jagadeeswaran, G., Saini, A., and Sunkar, R. (2009). Biotic and Abiotic Stress Down-Regulate miR398 Expression in Arabidopsis. *Planta* 229, 1009–1014. doi:10.1007/s00425-009-0889-3

Jia, X., Wang, W.-X., Ren, L., Chen, Q.-J., Mendu, V., Willcut, B., et al. (2009). Differential and Dynamic Regulation of miR398 in Response to ABA and Salt Stress in Populus Tremula and *Arabidopsis thaliana*. *Plant Mol. Biol.* 71, 51–59. doi:10.1007/s11103-009-9508-8

Karuppanapandian, T., Wang, H. W., Prabakaran, N., Jeyalakshmi, K., Kwon, M., Manoharan, K., et al. (2011). 2,4-dichlorophenoxyacetic Acid-Induced Leaf Senescence in Mung Bean (Vigna Radiata L. Wilczek) and Senescence Inhibition by Co-treatment with Silver Nanoparticles. *Plant Physiol. Biochem.* 49, 168–177. doi:10.1016/j.plaphy.2010.11.007

Kim, E.-J., Kim, H.-P., Hah, Y. C., and Roe, J.-H. (1996). Differential Expression of Superoxide Dismutases Containing Ni and Fe/Zn in Streptomyces Coelicolor. *Eur. J. Biochem.* 241, 178–185. doi:10.1111/j.1432-1033.1996.0178t.x

Kliebenstein, D. J., Monde, R.-A., and Last, R. L. (1998). Superoxide Dismutase in Arabidopsis: An eclectic Enzyme Family with Disparate Regulation and Protein Localization. *Plant Physiol.* 118, 637–650. doi:10.1104/pp.118.2.637

Kliebenstein, D. J., Monde, R., and Last, R. L. (2020). Superoxide Dismutase in Arabidopsis : An Eclectic Enzyme Family with Disparate Regulation and Protein Localization 1. *Plant Physiol.* 118, 637–650. doi:10.1104/pp.118.2.637

Kondrashov, F. A. (2012). Gene Duplication as a Mechanism of Genomic Adaptation to a Changing Environment. *Proc. R. Soc. B.* 279, 5048–5057. doi:10.1098/rspb.2012.1108

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Kurepa, J., Van Montagu, M., and Inzé, D. (1997). Expression of sodCp and sodB Genes in Nicotiana Tabacum: Effects of Light and Copper Excess. *J. Exp. Bot.* 48, 2007–2014. doi:10.1093/jexbot/48.317.2007

Lee, S.-H., Ahsan, N., Lee, K.-W., Kim, D.-H., Lee, D.-G., Kwak, S.-S., et al. (2007). Simultaneous Overexpression of Both CuZn Superoxide Dismutase and Ascorbate Peroxidase in Transgenic Tall Fescue Plants Confers Increased Tolerance to a Wide Range of Abiotic Stresses. *J. Plant Physiol.* 164, 1626–1638. doi:10.1016/j.jplph.2007.01.003

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., et al. (2002). PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a portal to Tools for In Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* 30, 325–327. doi:10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) V4: Recent Updates and New Developments. *Nucleic Acids Res.* 47, W256–W259. doi:10.1093/nar/gkz239

Li, Q., Jin, X., and Zhu, Y.-X. (2012). Identification and Analyses of miRNA Genes in Allotetraploid Gossypium Hirsutum Fiber Cells Based on the Sequenced Diploid G. Raimondii Genome. *J. Genet. Genomics* 39, 351–360. doi:10.1016/j.jgg.2012.04.008

Lin, Y.-L., and Lai, Z.-X. (2013). Superoxide Dismutase Multigene Family in Longan Somatic Embryos: A Comparison of CuZn-SOD, Fe-SOD, and Mn-SOD Gene Structure, Splicing, Phylogeny, and Expression. *Mol. Breed.* 32, 595–615. doi:10.1007/s11032-013-9892-2

Liu, J., Xu, L., Shang, J., Hu, X., Yu, H., Wu, H., et al. (2021). Gene Family Reveals Important Roles in Drought and Salt Responses. *Genet. Mol. Biol.* 44, 20210035. doi:10.1590/1678-4685-gmb-2021-0035

Lomakina, N. (2018). Forming Long-Term Competitiveness of a mineral Sector as Part of Economy Innovative Development Strategy. *E3s Web Conf.* 56, 04010–04148. doi:10.1051/e3sconf/20185604010

Lu, W., Duanmu, H., Qiao, Y., Jin, X., Yu, Y., Yu, L., et al. (2020). Genome-wide Identification and Characterization of the Soybean SOD Family during Alkaline Stress. *PeerJ* 8, e8457–19. doi:10.7717/peerj.8457

Lu, Y., Feng, Z., Bian, L., Xie, H., and Liang, J. (2011). miR398 Regulation in rice of the Responses to Abiotic and Biotic Stresses Depends on CSD1 and CSD2 Expression. *Funct. Plant Biol.* 38, 44. doi:10.1071/fp10178

Mann, T., and Keilin, D. (1938). Haemocuprein and Hepatocuprein, Copper-Protein Compounds of Blood and Liver in Mammals. *Proc. R. Soc. Lond. Ser. B - Biol. Sci.* 126, 303–315. doi:10.1098/rspb.1938.0058

Manzoor, M. A., Cheng, X., Li, G., Su, X., Abdullah, M., and Cai, Y. (2020). Gene Structure, Evolution and Expression Analysis of the P-ATPase Gene Family in Chinese Pear (Pyrus Bretschneideri). *Comput. Biol. Chem.* 88, 107346. doi:10.1016/j.compbiolchem.2020.107346

Manzoor, M. A., Li, G., Abdullah, M., Han, W., Wenlong, H., Yang, Z., et al. (2021a). Genome-wide Investigation and Comparative Analysis of MATE Gene Family in Rosaceae Species and Their Regulatory Role in Abiotic Stress Responses in Chinese Pear (Pyrus Bretschneideri). *Physiologia Plantarum* 173, 1163–1178. doi:10.1111/ppl.13511

Manzoor, M. A., Manzoor, M. M., Li, G., Abdullah, M., Han, W., Wenlong, H., et al. (2021b). Genome-wide Identification and Characterization of bZIP Transcription Factors and Their Expression Profile under Abiotic Stresses in Chinese Pear (Pyrus Bretschneideri). *BMC Plant Biol.* 21, 1–16. doi:10.1186/s12870-021-03191-3

McCord, J. M., and Fridovich, I. (1969). Superoxide Dismutase. An Enzymic Function for Erythrocuprein (Hemocuprein). *J. Biol. Chem.* 244, 6049–6055.

Miller, A.-F. (2012). Superoxide Dismutases: Ancient Enzymes and New Insights. *FEBS Lett.* 586, 585–595. doi:10.1016/j.febslet.2011.10.048

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi:10.1093/nar/gkaa913

Mittler, R., and Blumwald, E. (2010). Genetic Engineering for Modern Agriculture: Challenges and Perspectives. *Annu. Rev. Plant Biol.* 61, 443–462. doi:10.1146/annurev-arplant-042809-112116

Mittler, R. (2017). ROS Are Good. *Trends Plant Sci.* 22, 11–19. doi:10.1016/j.tplants.2016.08.002

Mittler, R., Vanderauwera, S., Gollery, M., and Van Breusegem, F. (2004). Reactive Oxygen Gene Network of Plants. *Trends Plant Sci.* 9, 490–498. doi:10.1016/j.tplants.2004.08.009

Molina-Rueda, J. J., Tsai, C. J., and Kirby, E. G. (2013). The Populus Superoxide Dismutase Gene Family and its Responses to Drought Stress in Transgenic Poplar Overexpressing a Pine Cytosolic Glutamine Synthetase (GS1a). *PLoS One* 8, e56421. doi:10.1371/journal.pone.0056421

Møller, I. M. (2001). PLANT MITOCHONDRIA and OXIDATIVE STRESS: Electron Transport, NADPH Turnover, and Metabolism of Reactive Oxygen Species. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 561–591. doi:10.1146/annurev.arplant.52.1.561

Nath, K., Kumar, S., Poudyal, R. S., Yang, Y. N., Timilsina, R., Park, Y. S., et al. (2014). Developmental Stage-dependent Differential Gene Expression of Superoxide Dismutase Isoenzymes and Their Localization and Physical Interaction Network in rice (Oryza Sativa L). *Genes Genom* 36, 45–55. doi:10.1007/s13258-013-0138-9

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

Oppenheim, A. V., Willsky, A. S., and Hamid, S. (1996). *Signals and Systems*. 2nd Edition, 18, 957. doi:10.1105/tpc.106.041673.1

Pilon, M., Ravet, K., and Tapken, W. (2011). The Biogenesis and Physiological Function of Chloroplast Superoxide Dismutases. *Biochim. Biophys. Acta - Bioenerg.* 1807, 989–998. doi:10.1016/j.bbabio.2010.11.002

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene Duplication and Evolution in Recurring Polyploidization-Diploidization Cycles in Plants. *Genome Biol.* 20, 1–23. doi:10.1186/s13059-019-1650-2

Qiao, X., Yin, H., Li, L., Wang, R., Wu, J., Wu, J., et al. (2018). Different Modes of Gene Duplication Show Divergent Evolutionary Patterns and Contribute Differently to the Expansion of Gene Families Involved in Important Fruit Traits in Pear (Pyrus Bretschneideri). *Front. Plant Sci.* 9, 1–16. doi:10.3389/fpls.2018.00161

Qiu, C. X., Xie, F. L., Zhu, Y. Y., Guo, K., Huang, S. Q., Nie, L., et al. (2007). Computational Identification of microRNAs and Their Targets in Gossypium Hirsutum Expressed Sequence Tags. *Gene* 395, 49–61. doi:10.1016/j.gene.2007.01.034

Rabinowitch, H. D., and Sklan, D. (1980). Superoxide Dismutase: A Possible Protective Agent against Sunscald in Tomatoes (Lycopersicon Esculentum mill). *Planta* 148, 162–167. doi:10.1007/BF00386417

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a Simple Modular Architecture Research Tool: Identification of Signaling Domains. *Proc. Natl. Acad. Sci.* 95, 5857–5864. doi:10.1073/pnas.95.11.5857

Schützendübel, A., and Polle, A. (2002). Plant Responses to Abiotic Stresses: Heavy Metal-Induced Oxidative Stress and protection by Mycorrhization. *J. Exp. Bot.* 53, 1351–1365. doi:10.1093/jxb/53.372.1351

Silva, S. A. F., Silva, F. L. B., Ribas, A. F., de Souza, S. G. H., and dos Santos, T. B. (2020). Genome-wide In Silico Analysis of SOD Genes in Common Bean (Phaseolus vulgaris L). *J. Crop Sci. Biotechnol.* 23, 241–251. doi:10.1007/s12892-020-00030-z

Song, J., Zeng, L., Chen, R., Wang, Y., and Zhou, Y. (2018). In Silico identification and Expression Analysis of Superoxide Dismutase (SOD) Gene Family in Medicago Truncatula. *3 Biotech.* 8, 1–12. doi:10.1007/s13205-018-1373-1

Srivastava, V., Srivastava, M. K., Chibani, K., Nilsson, R., Rouhier, N., Melzer, M., et al. (2009). Alternative Splicing Studies of the Reactive Oxygen Species Gene Network in Populus Reveal Two Isoforms of High-Isoelectric-Point Superoxide Dismutase. *Plant Physiol.* 149, 1848–1859. doi:10.1104/pp.108.133371

Su, W., Raza, A., Gao, A., Jia, Z., Zhang, Y., Hussain, M. A., et al. (2021). Genome-wide Analysis and Expression Profile of Superoxide Dismutase (Sod) Gene Family in Rapeseed (brassica Napus l.) under Different Hormones and Abiotic Stress Conditions. *Antioxidants* 10, 1182. doi:10.3390/antiox10081182

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J.-K. (2007). Small RNAs as Big Players in Plant Abiotic Stress Responses and Nutrient Deprivation. *Trends Plant Sci.* 12, 301–309. doi:10.1016/j.tplants.2007.05.001

Tepperman, J. M., and Dunsmuir, P. (1990). Transformed Plants with Elevated Levels of Chloroplastic SOD Are Not More Resistant to Superoxide Toxicity. *Plant Mol. Biol.* 14, 501–511. doi:10.1007/BF00027496

Thompson, J., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Res.* 25, 4876–4882. doi:10.1093/nar/25.24.4876

Verma, D., Lakhanpal, N., and Singh, K. (2019). Genome-wide Identification and Characterization of Abiotic-Stress Responsive SOD (Superoxide Dismutase) Gene Family in Brassica Juncea and B. Rapa. *BMC Genomics* 20, 1–18. doi:10.1186/s12864-019-5593-5

Wang, B., Lüttge, U., and Ratajczak, R. (2004). Specific Regulation of SOD Isoforms by NaCl and Osmotic Stress in Leaves of the C3 Halophyte Suaeda Salsa L. *J. Plant Physiol.* 161, 285–293. doi:10.1078/0176-1617-01123

Wang, T., Song, H., Zhang, B., Lu, Q., Liu, Z., Zhang, S., et al. (2018). Genome-wide Identification, Characterization, and Expression Analysis of Superoxide Dismutase (SOD) Genes in Foxtail Millet (Setaria Italica L). *3 Biotech.* 8, 1–11. doi:10.1007/s13205-018-1502-x

Wang, W., Xia, M., Chen, J., Deng, F., Yuan, R., Zhang, X., et al. (2016). Genome-wide Analysis of Superoxide Dismutase Gene Family in Gossypium Raimondii and G. Arboreum. *Plant Gene* 6, 18–29. doi:10.1016/j.plgene.2016.02.002

Wang, W., Zhang, X., Deng, F., Yuan, R., and Shen, F. (2017). Genome-wide Characterization and Expression Analyses of Superoxide Dismutase (SOD) Genes in Gossypium Hirsutum. *BMC Genomics* 18, 1–25. doi:10.1186/s12864-017-3768-5

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* 40, e49. doi:10.1093/nar/gkr1293

Wei, K., Chen, J., Wang, Y., Chen, Y., Chen, S., Lin, Y., et al. (2012). Genome-wide Analysis of bZIP-Encoding Genes in maize. *DNA Res.* 19, 463–476. doi:10.1093/dnares/dss026

Wu, G., Wilen, R. W., Robertson, A. J., and Gusta, L. V. (1999). Isolation, Chromosomal Localization, and Differential Expression of Mitochondrial Manganese Superoxide Dismutase and Chloroplastic Copper/Zinc Superoxide Dismutase Genes in Wheat1. *Plant Physiol.* 120, 513–520. doi:10.1104/pp.120.2.513

Wuerges, J., Lee, J.-W., Yim, Y.-I., Yim, H.-S., Kang, S.-O., and Djinovic Carugo, K. (2004). Crystal Structure of Nickel-Containing Superoxide Dismutase Reveals Another Type of Active Site. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8569–8574. doi:10.1073/pnas.0308514101

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of Duplicate Genes in Exon-Intron Structure. *Proc. Natl. Acad. Sci.* 109, 1187–1192. doi:10.1073/pnas.1109047109

Yan, J.-J., Zhang, L., Wang, R.-Q., Xie, B., Li, X., Chen, R.-L., et al. (2016). The Sequence Characteristics and Expression Models Reveal Superoxide Dismutase Involved in Cold Response and Fruiting Body Development in Volvariella Volvacea. *Int. J. Mol. Sci.* 17, 34. doi:10.3390/ijms17010034

Youn, H.-D., Kim, E.-J., Roe, J.-H., Hah, Y. C., and Kang, S.-O. (1996). A Novel Nickel-Containing Superoxide Dismutase from Streptomyces Spp. *Biochem. J.* 318, 889–896. doi:10.1042/bj3180889

Zang, Y., Chen, J., Li, R., Shang, S., and Tang, X. (2020). Genome-wide Analysis of the Superoxide Dismutase (SOD) Gene Family in Zostera marina and Expression Profile Analysis under Temperature Stress. *PeerJ* 8, e9063–18. doi:10.7717/peerj.9063

Zelko, I. N., Mariani, T. J., and Folz, R. J. (2002). Superoxide Dismutase Multigene Family: A Comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) Gene Structures, Evolution, and Expression. *Free Radic. Biol. Med.* 33, 337–349. doi:10.1016/S0891-5849(02)00905-X

Zhang, C., Zheng, H., Wu, X., Xu, H., Han, K., Peng, J., et al. (2018). Genome-wide Identification of New Reference Genes for RT-qPCR Normalization in CGMMV-Infected Lagenaria Siceraria. *PeerJ* 6, e5642. doi:10.7717/peerj.5642

Zhang, J., Li, B., Yang, Y., Hu, W., Chen, F., Xie, L., et al. (2016). Genome-Wide Characterization and Expression Profiles of the Superoxide Dismutase Gene Family in Gossypium. *Int. J. Genomics* 2016, 1–11. doi:10.1155/2016/8740901

Zhang, X., Zhang, L., Chen, Y., Wang, S., Fang, Y., Zhang, X., et al. (2021). Genome-wide Identification of the SOD Gene Family and Expression Analysis under Drought and Salt Stress in Barley. *Plant Growth Regul.* 94, 49–60. doi:10.1007/s10725-021-00695-8

Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., et al. (2019). Cucurbit Genomics Database (CuGenDB): A central portal for Comparative and Functional Genomics of Cucurbit Crops. *Nucleic Acids Res.* 47, D1128–D1136. doi:10.1093/nar/gky944

Zhou, Y., Liu, L., Huang, W., Yuan, M., Zhou, F., Li, X., et al. (2014). Overexpression of OsSWEET5 in Rice Causes Growth Retardation and Precocious Senescence. *PLoS ONE* 9, e94210–9. doi:10.1371/journal.pone.0094210

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 Rehman, Rashid, Manzoor, Li, Sun, Riaz, Li and Zhuge. 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: 24 February 2022
doi: 10.3389/fgene.2022.842868

CerealESTDb: A Comprehensive Resource for Abiotic Stress-Responsive Annotated ESTs With Predicted Genes, Gene Ontology, and Metabolic Pathways in Major Cereal Crops

Sanjeev Kumar[*†], *Jyotika Bhati*[†], *Arijit Saha*[†], *Shashi Bhushan Lal, Pankaj Kumar Pandey, Dwijesh Chandra Mishra, Mohammad Samir Farooqi, Anuj Kumar, Krishna Kumar Chaturvedi and Anil Rai*

Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India

Cereals are the most important food crops and are considered key contributors to global food security. Loss due to abiotic stresses in cereal crops is limiting potential productivity in a significant manner. The primary reasons for abiotic stresses are abrupt temperature, variable rainfall, and declining nutrient status of the soil. Varietal development is the key to sustaining productivity under influence of multiple abiotic stresses and must be studied in context with genomics and molecular breeding. Recently, advances in a plethora of Next Generation Sequencing (NGS) based methods have accelerated the enormous genomic data generation associated with stress-induced transcripts such as microarray, RNAseq, Expressed Sequenced Tag (ESTs), etc. Many databases related to microarray and RNA-seq based transcripts have been developed and profusely utilized. However, an abundant amount of transcripts related to abiotic stresses in various cereal crops arising from EST technology are available but still remain underutilized in absence of a consolidated database. In this study, an attempt has been made with a primary goal to integrate, analyse, and characterise the available resources of ESTs responsive to abiotic stresses in major cereals. The developed CerealESTdb presents a customisable search in two different ways in the form of searchable content for easy access and potential use. This database comprises ESTs from four major cereal crops, namely rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), sorghum (*Sorghum bicolour* L.), and maize (*Zea mays* L.), under a set of abiotic stresses. The current statistics of this cohesive database consists of 55,826 assembled EST sequences, 51,791 predicted genes models, and their 254,609 gene ontology terms including extensive information on 1,746 associated metabolic pathways. We anticipate that developed CerealESTdb will be helpful in deciphering the knowledge of complex biological phenomena under abiotic stresses to accelerate the molecular breeding programs towards the development of crop cultivars resilient to abiotic stresses. The CerealESTdb is publically available with the URL http://cabgrid.res.in/CerealESTDb.

Keywords: cereals, abiotic stress, ests, genes, pathways, ontology

OPEN ACCESS

Edited by:
Ramsey Steven Lewis, North Carolina State University, United States

Reviewed by:
Upendra Kumar, Chaudhary Charan Singh Haryana Agricultural University, India
Romesh Kumar Salgotra, Sher-e-Kashmir University of Agricultural Sciences and Technology, India
Shumayla, Panjab University, India

***Correspondence:**
Sanjeev Kumar
sanjeevbunty@gmail.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: *24 December 2021*
Accepted: *21 January 2022*
Published: *24 February 2022*

Citation:
Kumar S, Bhati J, Saha A, Lal SB, Pandey PK, Mishra DC, Farooqi MS, Kumar A, Chaturvedi KK and Rai A (2022) CerealESTDb: A Comprehensive Resource for Abiotic Stress-Responsive Annotated ESTs With Predicted Genes, Gene Ontology, and Metabolic Pathways in Major Cereal Crops. Front. Genet. 13:842868. doi: 10.3389/fgene.2022.842868

INTRODUCTION

To provide more than nine billion people by 2050 with essential nutrients and energy in their everyday diet, food production must be increased by 70% with reduced fertilizer inputs (Wani et al., 2021; Kumar et al., 2018; Balyan et al., 2016). Crop productivity suffers due to multiple abiotic stresses during the crop life cycle. The important four cereal crops, namely rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), sorghum (*Sorghum bicolour* L.), and maize (*Zea mays* L.) show sensitivity to a wide range of abiotic stress viz., abscisic acid (ABA), cold, drought, heat, and salt (Askari-Khorasgani et al., 2021; Cramer et al., 2011). In crop plants, mitigation mechanisms against these environmental stresses are complex biological processes. Although, there is a surge of biological information on plant responses to abiotic stresses, the understanding of the molecular mechanisms involved is still enigmatic (Zhang et al., 2021; Wani et al., 2018; Gahlaut et al., 2016). The mining of candidate genes and gene families responsive to multiple abiotic stresses from whole-genome sequencing data facilitates the molecular understanding of the complex cellular and molecular processes in a significant manner. Also, functional and structural annotations of genes and gene families will provide insight into the complex interactions regulating stress responses (Takahashi et al., 2020; Atkinson et al., 2013). Recently, advances in high throughput omics technologies have accelerated the identification and characterization of novel genes and their biological functions in a tissue-specific manner. At the genome and transcriptome level, the pathway reconstruction will play a principal role in quantifying and characterising the genotype to phenotype relationship (Gupta et al., 2020; Zhang et al., 2020; Feist and Palsson, 2008). In this ongoing genomic era, a huge amount of genomic data have been generated and analysed to provide key information leading to the development of crop cultivars resilient to abiotic stresses (Alter et al., 2015). This includes various transcriptomic data of gene expression such as Expressed Sequence Tags (EST), microarray, RNAseq etc. (Vincent et al., 2013; Zeng and Extavour, 2012). Whlie the era of Next Generation Sequencing (NGS) generates a huge amount of transcriptome data, previously generated EST resources still hold an important position. EST sequence has been defined as a short genomic sub-sequence and is important for the identification of gene transcripts under different environmental conditions. It can provide a means of accessing the gene space of almost any organism (Bouck and Vision, 2007; Nagaraj et al., 2007). EST libraries are a cost-effective tool and can serve as a starting point for the development of effective molecular genetic markers, such as gene-linked microsatellites, single nucleotide polymorphisms, etc. However, these EST resources are too scattered to provide insight into the biological process including the involvement of genes/regulatory elements and their functional role in responding to different abiotic stresses. Though the process of generation of EST data is historical, simple, cost-effective, and abundantly available it largely remains underutilised due to the scarce availability of proper databases. Efforts were made in the past by several research groups related to the development of EST resources in cereals such as Gramene (Jaiswal, 2011), cerealsDB (Wilkinson et al., 2012), Wheat Genome (Lai et al., 2012), MaizeGDB (Lawrence et al., 2004), Plant Stress Gene Database (Prabha et al., 2011), and Uniprot (Apweiler et al., 2004) summarized in **Table 1**.

Therefore, an attempt has been made to develop a user friendly, searchable, and interactive database of assembled and annotated ESTs for major cereal crops i.e., maize, rice, sorghum, and wheat along with associated genes and metabolic pathways involved in abiotic stresses *viz.*, ABA, cold, drought, heat and salt. In order to develop this database, these abiotic stresses related ESTs of considered cereal crops were collected from various sources and assembled. Subsequently, genes were identified and putative genes were predicted and further, these genes were functionally annotated using bioinformatics tools (Bhati et al., 2016; Chaduvula et al., 2015; Bhati et al., 2014). Partial amplification of nine identified candidate genes related to salinity stress response genes was standardised and carried out in *Oryza sativa* var. *indica* IR64 (Bhati et al., 2016), and 12 candidate genes were selected for salinity stress and amplified in nine indigenous *Sorghum bicolor* genotypes (Chaduvula et al., 2015). Moreover, information associated with various metabolic pathways involving these genes was identified and collected. This extracted and derived information was stored in relational database management systems (RDBMS) and made available through a searchable web interface as CerealESTdb which is unique in terms of its contents and features. This database provides various search options to retrieve the desired information in a few clicks and selections. This database contains 51,791 genes, 254,609 gene ontology terms, and 1,746 pathways. The primary goal of this database is to serve as a central data repository of processed and annotated EST resources of major cereal crops under abiotic stress conditions. The database is accessible from http://cabgrid.res.in/CerealESTDb.

MATERIALS AND METHODS

Source of Data and Functional Annotation

The EST sequences (ESTs) were searched and downloaded from National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov) by using various combinations of terms crop name, the scientific name of crop, and stress name for four important cereal crops *viz*., maize, rice, sorghum, and wheat and five abiotic stresses *viz*., abscisic acid (ABA), cold, drought, heat, and salt stresses (**Figure 1**). These EST sequences were re-checked with the writing of suitable Perl scripts and filtered out all non-related EST sequences. The remaining EST sequences were also masked to eliminate the incorrect part in these sequences and then finally assembled using EGAssembler (https://www.genome.jp/tools/egassembler/) by considering the default parameter settings and the similarity-based clustering was set to 80 percent overlap identity (Masoudi-Nejad et al., 2006). These assembled EST-contigs were then translated to reading frames. The BLASTx algorithm was applied against the protein sequence database (NCBI nr) for the identification of potential translated products (Altschul et al., 1990). The homologous EST-contigs with annotated proteins that exists in the nr database were selected for functional annotations based on maximum

TABLE 1 | Summary of the comparison with other database resources.

Database	Developed	Updated	Web address	Advantages	Limitations
Gramene (Jaiswal, 2011)	2010	2019	http://www.gramene.org/	It is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species	The information flow in this database is not very user-friendly as this is a generic data resource. End-users will have to carry out a rigorous search to extracting the required information
cerealsDB (Wilkinson et al., 2012)	2012	2020	https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php	This database contains data related to all SNPs including flanking sequences as well as links to various other genomics tools for further analysis	This database contains SNP markers related to wheat crops only
Wheat Genome (Lai et al., 2012)	2012	No latest update	http://www.wheatgenome.info/	This repository stores Wheat genome sequences generated from different assembly processes	Information only on the genomic sequences fails to provide other useful information such as genes, proteins, pathwaysetc.
MaizeGDB (Lawrence et al., 2004)	2004	2020	http://www.maizegdb.org/	This service provides basic cultivation-related information on Maize crops. It also includes SNP, mutants, phenotypes, and other useful tools	This resource is lacking in providing the information related to genes, proteins, and pathways along with ontologies related to maize genomes
Plant Stress Gene Database (Prabha et al., 2011)	2011	No latest update	http://ccbb.jnu.ac.in/stressgenes/frontpage.html	This is a static database with 259 stress-related genes of 11 important species and these genes are linked to NCBI resources for further information	The limitation of this resource is in providing the proteins, pathways, and GO terms with respect to stress-specific genes
Uniprot (Apweiler et al., 2004)	2002	2020	https://www.uniprot.org/	This resource acts as an encyclopaedia of proteins and their related information	The stress-specific information for important cereal crops is not included in this data resource

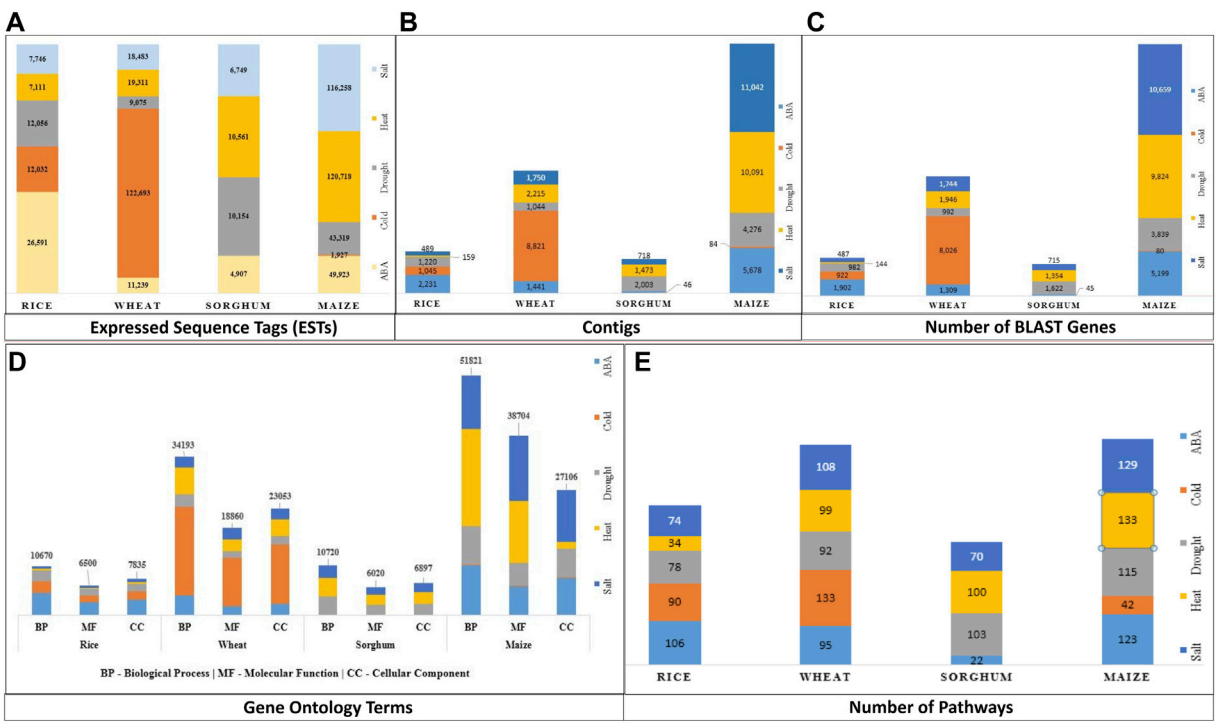


FIGURE 1 | Data Summary—**(A)** ESTs, **(B)** Annotated contigs along with their corresponding **(C)** BLAST genes, **(D)** GO term, and **(E)** Pathway (GO: Gene Ontology, BP, Biological process; MF, Molecular Function; CC, Cellular Components).

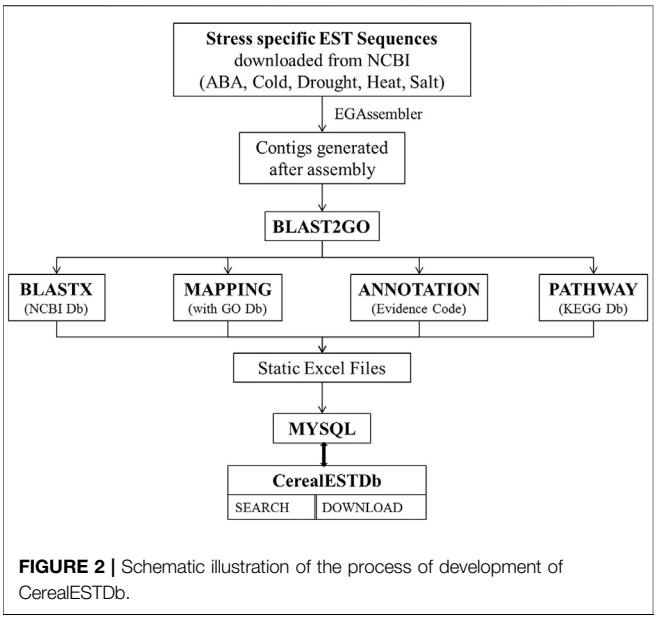


FIGURE 2 | Schematic illustration of the process of development of CerealESTDb.

E-value ($1E^{-3}$) and the minimum alignment size of high scoring pair (HSP) length of 33. Blast2GO v 2.5 was employed to perform the functional analysis of the EST-contigs (Conesa and Gotz, 2008; Gotz et al., 2008). Further, the EST-contigs were categorized according to the GO vocabularies into three categories *i.e.* molecular function, biological process, and cellular component. These EST-contigs were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) and attained the corresponding metabolic pathways. A flowchart depicting the pipeline for above-described process is presented in **Figure 2**.

Software Architecture and Database Design

The CerealESTDb has been implemented using standard 3-tier web architecture (Lee et al., 2009) followed to develop this web application. Different components of CerealESTDb architecture are presented in **Figure 3**. Access to this database has been provided by developing an end-user interface using HyperText Markup Language (HTML), cascading style sheets (CSS), and javascript. The search request will be submitted to a web server i.e., apache-tomcat, and processed by corresponding servlets created using Java Server Pages (JSP) and javascript. These servlets fetched the data from the developed RDBMS in MySQL using Java Database Connectivity (JDBC) technology to process and send the response to the end user's web browser.

The database has been implemented using MySQL and contained different information related to crop, stress, gene name, gene ontology, enzyme, and pathway. The design of the database has followed the normalization procedure of a relational database management system to maintain data consistency and avoid data redundancy in the database. The database comprises tables for crop, stress, contigs, gene ontology, enzyme, pathway, contigs_pathway, and pathway_enzyme. The entity-relationship

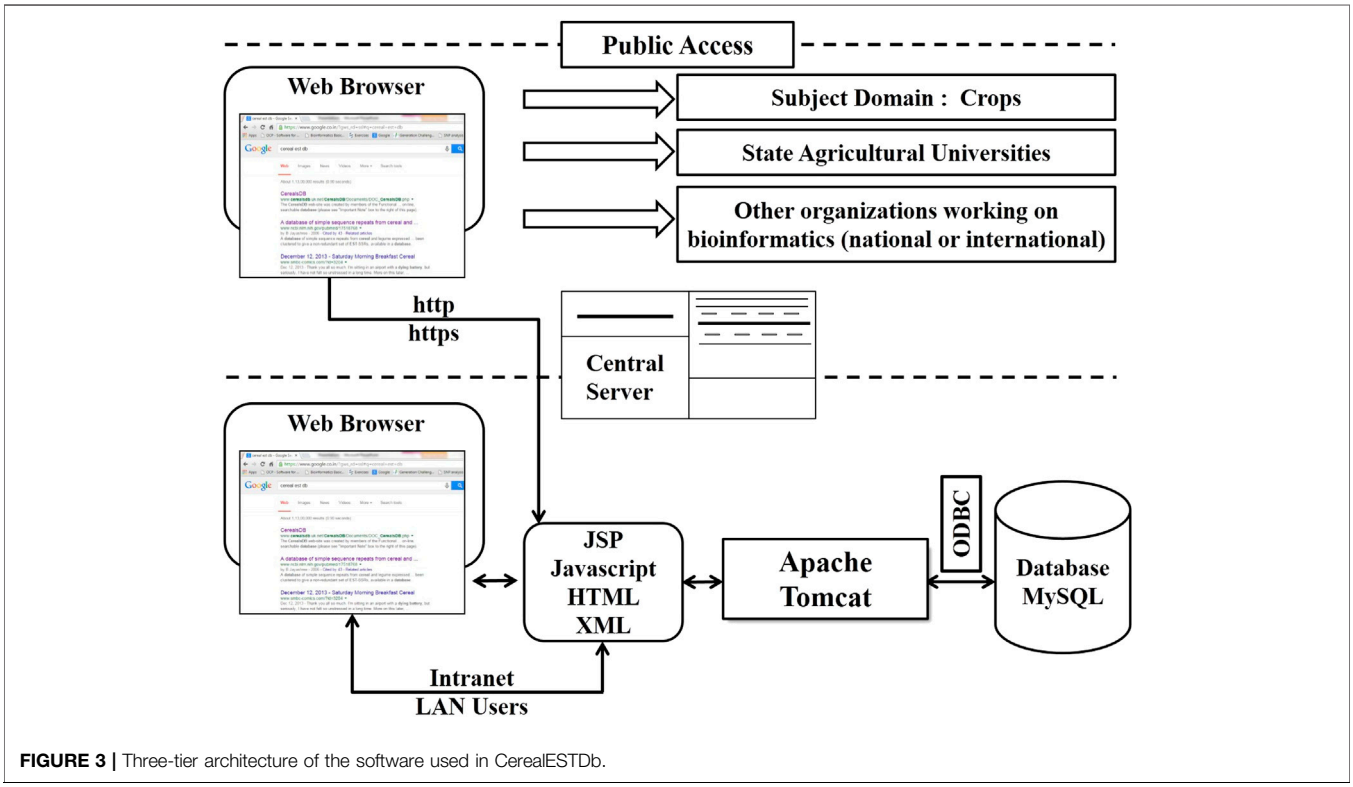


FIGURE 3 | Three-tier architecture of the software used in CerealESTDb.

(ER) diagram of this database is mentioned in **Supplementary Figure S1**.

User Interface of CerealESTDb

The CerealESTDb comprises of Home, About, Search, Advance Search, Important Link, Contact Us, and Help. The "Home" page describes the database and its possible use in a molecular breeding program. The "About" page describes the process workflow used for extracting and processing the data, starting from collection of data to annotation of the genes and finally populating into the database. Various search options provide the facility to extract the stored information in two ways namely, "Search" (**Supplementary Figure S2A**) and "Advance Search" (**Supplementary Figure S2B**).

Several options have been provided in the web interface to enable a fast and efficient search for the end-user to retrieve the desired results with and meet the varied requirements of end users in terms of selecting various attributes. The users have the option to enter the free text in the search box or select the required values from dropdown lists. After selecting or entering the values for various input boxes, the users will click on the "Submit" button to obtain the required results. After clicking on the "Submit" button, the selected and entered values are converted into suitable Structured Query Language (SQL) format and fetched the results from the database and delivered the results in easy to use format to the end users' browser window.

RESULTS AND DISCUSSION

Out of 610,893 ESTs downloaded for abiotic stresses viz., ABA, cold, drought, heat, and salt across all four crops viz., maize, rice, sorghum, and wheat, only 9% were assembled into contigs (55,826 in number). ESTs for cold stress in sorghum were not assembled because EGAssembler requires a minimum of 64 EST sequences and there were only 40 ESTs for cold stress in sorghum. There were a total of 51,791 genes that were annotated by BLAST out of 55,826 contigs across all the abiotic stresses. The total number of the assembled contigs and corresponding annotated BLAST genes against the *nr* database of NCBI, GO terms, and pathways are shown in **Figure 1**. The options to select the desired attributes have also been provided to the users and avoided the unnecessary attributes from the database that save the query response time too. Some of the important attribute names are selected by default and users have the option to either select or ignored other attributes, if needed. These results can further be combined with identified GO terms and associated pathways if any. The identified pathways will be helpful for retrieving the enzymatic information and hyperlink to pathways that can be further visualized for additional information or relationships. The users can also download the result in.xls format to the user's desktop for later use. The above description is summarized in **Figure 4** for a better understating of the retrieval process.

Stress-wise annotated genes are also represented using pie charts for the cereal crops under study (**Figure 5A**). **Figure 5B** shows the percentage of GO terms in different crops across these five abiotic stresses. The GO terms were further classified into three categories, biological process (BP), molecular function (MF), and cellular component (CC). The number of biological process-related GO terms are largest for each stress in all crops, followed by cellular component and molecular function, except in salt stress of maize, rice, and wheat.

Many plant genes have their expression modulated by stress conditions. The common genes in all crops and conditions were auxin-binding protein/auxin response factor (Guo et al., 2018), CCT motif transcription factor (Li and Xu, 2017), heat-shock proteins (Jacob et al., 2017; Shanker el at., 2012), Ser/Thr protein kinase (Islam et al., 2013), ribosomal proteins (Moin et al., 2017), and zinc-finger proteins (Zang et al., 2016). The small auxin upregulated RNA (SAUR) is a plant-specific gene family of auxin-responsive genes. As evident from previous reports, the overexpression of these SAUR genes contributed to the tolerance to drought and salt in *Arabidopsis* (Guo et al., 2018). Auxin-response factor proteins have been well studied in crop plants for their direct involvement in response to multiple stresses (Stortenbeker and Bemer, 2019). The heat-shock proteins play key roles in protecting plant cells or tissues from various abiotic stresses (Wen et al., 2017). Other common genes such as 14-3-3 genes, cytochrome c/b/p450, ubiquitin-like protein, etc., have been identified in the present study. The 14-3-3 genes are involved in various abiotic stresses of different plant species such as rice (Chen et al., 2006; Chen et al., 2019), *Populus* (Tian et al., 2015), Phaseolus vulgaris L. (Li et al., 2015), and *Hevea brasiliensis* (Yang et al., 2014), as reported by previous studies.

The maximum number of GO terms is pertaining to response to stress/defense response in all abiotic stresses for all considered plant species. Moreover, some GO terms are specific to the abiotic stress like Abscisic acid stimulus/signalling/recovery, drought response, response to heat, salt, and cold.

Further, there were 1,746 KEGG pathways associated with the annotated genes in all crops for these five stresses. The sugar metabolism pathway is the most common pathway in all cereal crops under study. Sugars are chemically active biomolecules and are involved in crucial Physico-chemical mechanisms such as photosynthesis, respiration, seed germination, flowering, senescence, and so forth. Therefore, modulating sugar composition or concentration in plants may improve their responses or adaptation to abiotic stress (Gangola and Ramadoss, 2018). The other common pathways were carbon fixation in photosynthetic organisms, purine metabolism, and metabolic pathways. Evidence suggests that certain purine metabolites like allantoin and allantoate, products of purine metabolism contribute to stress tolerance in plants (Casartelli et al., 2019; Watanabe et al., 2014).

The user interface of this database has been developed to provide access to the compiled resources in the form of CerealESTDb and important links have been made for Home,

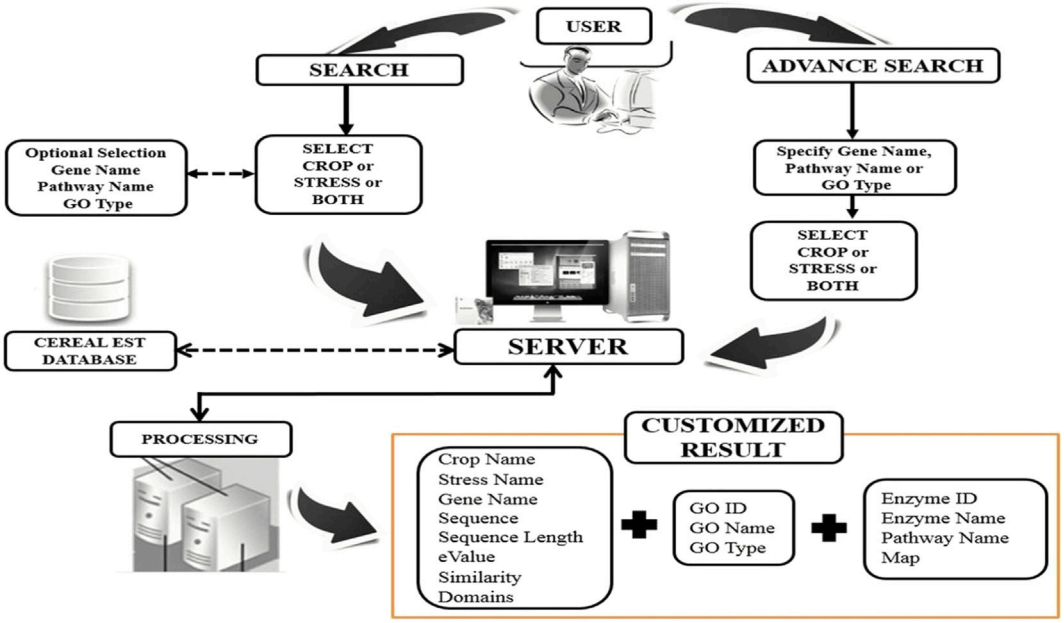


FIGURE 4 | Browsing CerealESTDb features diagram.

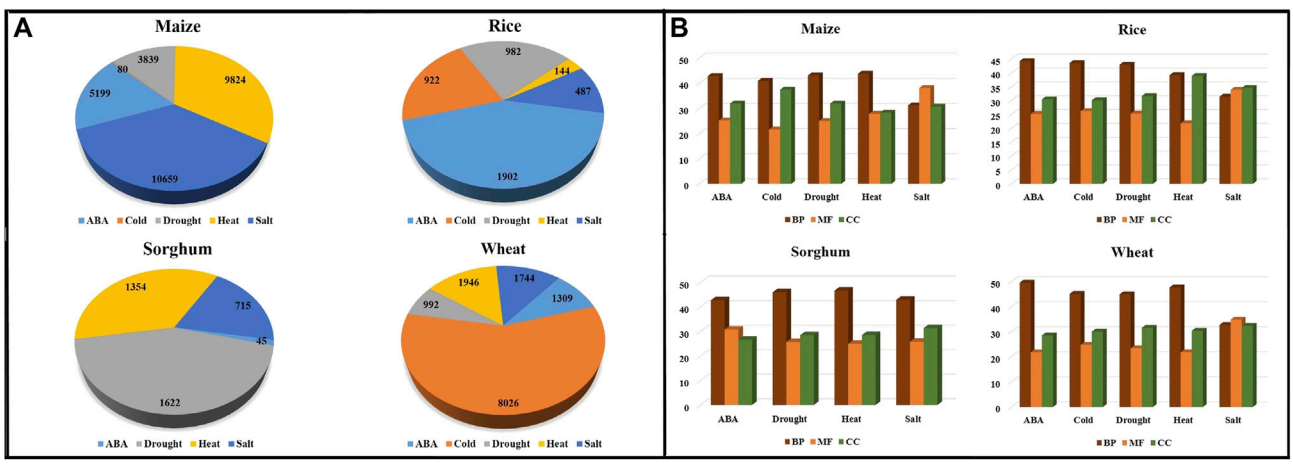


FIGURE 5 | Distribution of **(A)** annotated contigs across different abiotic stresses in the crops. **(B)** GO terms—Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) across abiotic stress in crops.

About CerealESTDb, Search, Advance Search, Important Links, Contact Us, and Help. There are two types of search options developed in the user interface in the form of "Search" and "Advanced Search". The first option allows the user to select crop or stress or a combination of crop and associated stress (**Figure 6A**).

The users have the option to either click on "Submit" or "Genes/Pathway". If a user clicks on "Submit," it will further ask users to select the list of attributes or parameters from available attributes (i.e., Crop Name, Stress Name, Gene Name, Sequence, Sequence length, e-value, similarity, and domains). After selecting the desired attributes, the required information will be displayed to the users and this information can also be downloaded and exported in.xls format for further analysis. Another option is to select genes and corresponding pathways for selected Crop/Stress/Crop-Stress and further choose the required attributes for additional information of the selected combination (**Figure 6B**). After selecting the gene/pathway option, it is compulsory to enter the gene name to proceed to the next level of search to retrieve the desired results.

Further, the users have the option to search by specifying "Gene Ontology" terms related to biological process, molecular function, and cellular component (**Figure 6C**). The results under

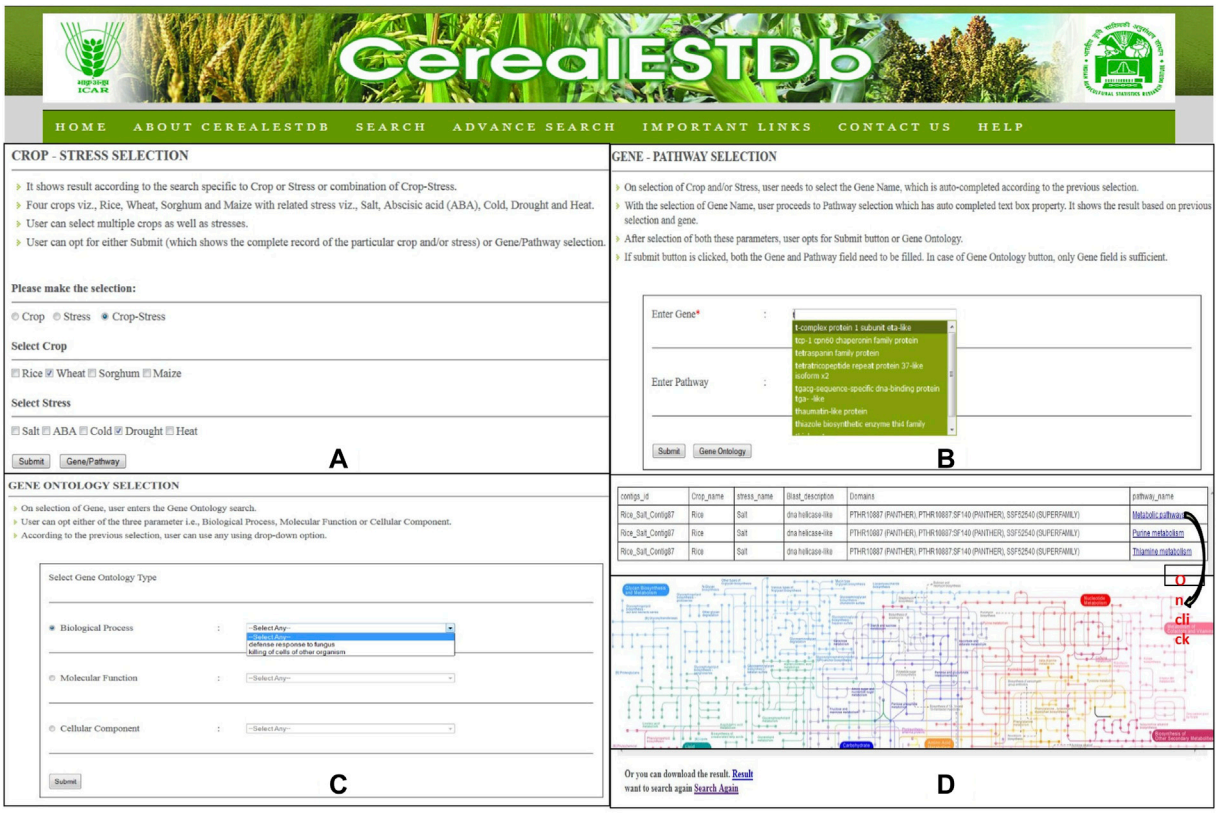


FIGURE 6 | Details of Search Options in CerealESTDb **(A)** Crop-Stress Selection page with options of select either crop or stress or combination of both **(B)** Gene-Pathway Selection page with mandatory field Gene and optional field Pathway **(C)** Gene Ontology Selection page, to select either of the three options of Biological Process or Molecular Function or Cellular Component **(D)** Final Result page after going through all the three selection pages with pathway image (on click).

each selection will be displayed to the end-user along with a pathway map (**Figure 6D**). The user can download these results in.xls format and the corresponding pathway map can also be saved as.gif.

The "Advanced Search" option provides users the ability to directly search information from this searchable database by selecting a specific gene name (**Figure 7A**), pathway name (**Figure 7B**), or gene ontology type (**Figure 7C**). The user is then asked to select the combination of crop name and stress names (**Figure 7D**). The relevant tabular information is displayed to the user which can be downloaded in.xls format (**Figure 7E**). In order to make the search user-friendly, the gene name, pathway name, and ontology type will be selected through a drop-down menu displayed from the database to the user after typing the first few letters of these in the search options.

"Important Link" helps the user find the web pages of relevant organizations and databases. "Contact Us" gives contact details of research team members and "Help" provides steps and flow of various search routes of this database.

The features of the CerealESTDb are:

a. User-friendly, searchable, and interactive data resource for searching a set of genomic information associated with response to abiotic stresses in cereal crops.
b. Provides a unique platform for interactive retrieval of annotated genes, GO terms, and corresponding generated metabolic pathways.
c. Provides alternate ways to explore 51,791 genes, 254,609 GO terms, and 1,746 generated metabolic pathways
d. Provides a facility to users to export results in.xls format for further annotations.

The few advantages of the present database are given below:

a. The developed database will provide the desired information related to search queries even without knowing anything specific to search query criteria and also facilitate the users through selection.
b. In this database, an effort has been made to collect various ESTs related to four important cereal crops across five specific abiotic stresses. A pipeline has been developed to mine the genes/proteins and that can also be applied in other plant species to annotate the ESTs and gene models.
c. The present database provides extensive information on metabolic pathways and extracted GO information with the help of BLAST2Go tool specific to particular abiotic stress i.e. ABA, cold, drought, heat, or salt-related to considered cereal crops.

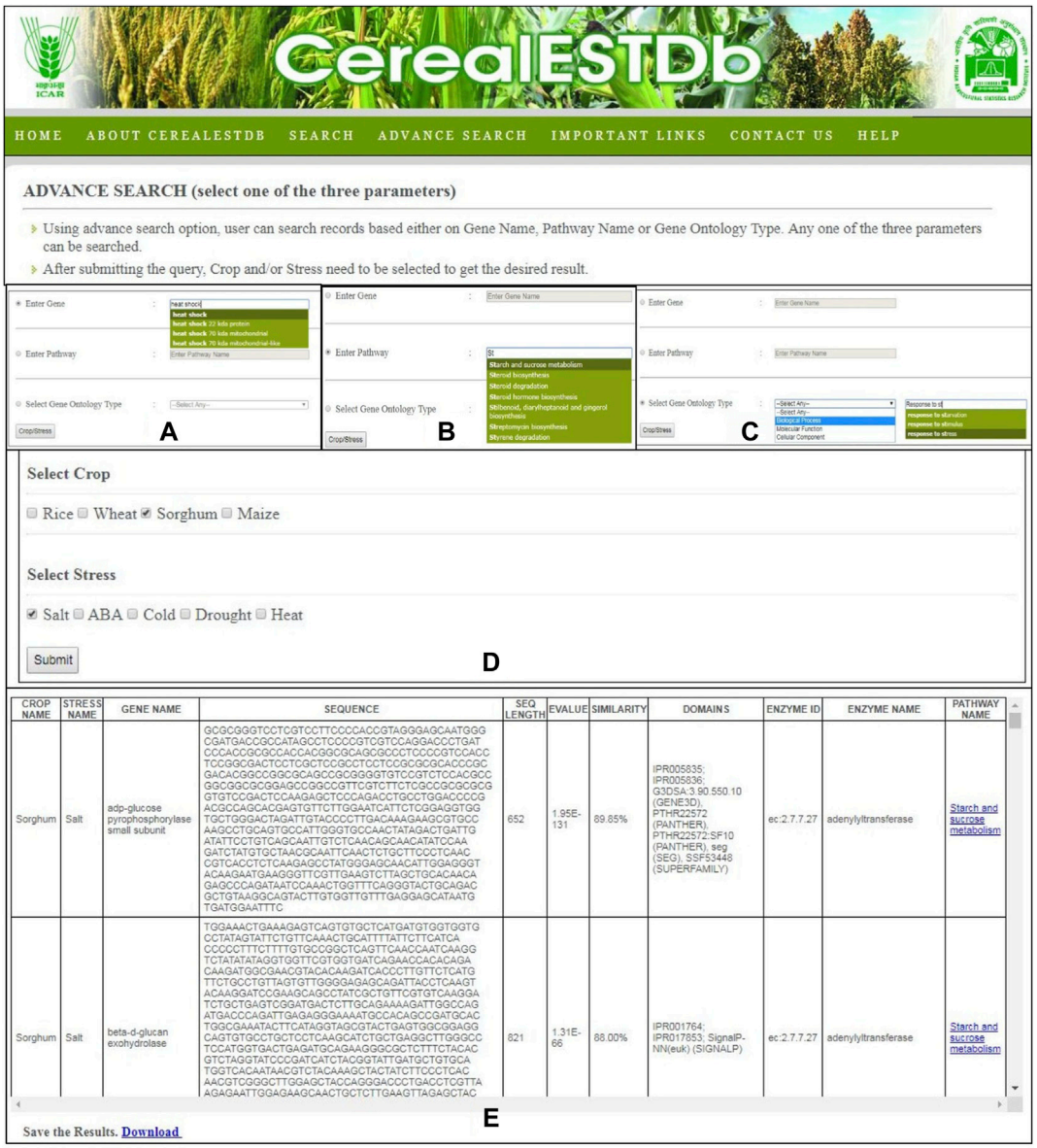


FIGURE 7 | Details of Advance Search options of CerealESTDb. The search could be done through three options **(A)** Gene name search or **(B)** Pathway name search or **(C)** Gene ontology term search, to select either of the three options of Biological Process or Molecular Function or Cellular Component and their corresponding values **(D)** Common Crop-Stress Selection page with options of select either crop or stress or combination of both **(E)** Final Result page after selecting either of the three options.

CONCLUSION

The CerealESTdb, a user-friendly, searchable, and interactive database has been developed with a primary goal to provide information on assembled and annotated ESTs from four major crop plants, namely wheat, rice, maize, and sorghum under multiple environmental stresses including cold, heat, drought, and salt stress as well as on the application of ABA. This cohesive database will also provide extensive information about genes, ontology, and corresponding metabolic pathways related to these ESTs. The CerealESTdb also facilitates users to access a description of the biological functions of genes involved in more than one related abiotic stress. This database will help in providing new solutions to molecular biologists and plant breeders for accelerating their efforts to develop abiotic stress-resistant cultivars without compromising the nutritional quality through introgression and gene pyramiding. With further progress in NGS methods and the availability of genome sequences data in the public domain, we hope that statistics of CerealESTdb will eventually grow by two-to-three folds in upcoming years.

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

All authors contributed to study design. JB did the data collection, data analysis, and data interpretation. KC, SL, JB, AS, and PP designed the database. SK, KC and JB wrote the manuscript draft. All authors contributed to discussing and writing the final text.

FUNDING

This research was funded by the project "Network Project for Agricultural Bioinformatics and Computational Biology under CABin Scheme," ICAR and project "Modeling Network of Gene Responses to Abiotic Stress in Rice," funded by NASF, ICAR.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.842868/full#supplementary-material

Supplementary Figure S1 | Entity-Relationship Diagram

Supplementary Figure S2 | Step-wise options followed in **(A)** simple Search page and **(B)** Advance Search page

REFERENCES

Alter, S., Bader, K. C., Spannagl, M., Wang, Y., Bauer, E., Schön, C.-C., et al. (2015). DroughtDB: an Expert-Curated Compilation of Plant Drought Stress Genes and Their Homologs in Nine Species. *Database* 2015. doi:10.1093/database/bav046

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215 (3), 403–410. doi:10.1016/S0022-2836(05)80360-2

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the Universal Protein Knowledgebase. *Nucleic Acids Res.* 32 (Suppl. l_1), 115D–119D. doi:10.1093/nar/gkh131

Askari-Khorasgani, O., Rehmani, M. I. A., Wani, S. H., and Kumar, A. (2021). *Osmotic Stress. In Handbook of Plant and Crop Physiology*, 4. New York: CRC Press, 445–464. doi:10.1201/9781003093640

Atkinson, N. J., Lilley, C. J., and Urwin, P. E. (2013). Identification of Genes Involved in the Response of Arabidopsis to Simultaneous Biotic and Abiotic Stresses. *Plant Physiol.* 162 (4), 2028–2041. doi:10.1104/pp.113.222372

Balyan, H. S., Gahlaut, V., Kumar, A., Jaiswal, V., Dhariwal, R., Tyagi, S., et al. (2016). Nitrogen and Phosphorus Use Efficiencies in Wheat: Physiology, Phenotyping, Genetics, and Breeding. *Plant Breed. Rev.* 40, 167–234. doi:10.1002/9781119279723.ch4

Bhati, J., Chaduvula K, P., Rai, A., Gaikwad, K., Soma Marla, S., and Kumar, S. (2016). In-silico Prediction and Functional Analysis of Salt Stress Responsive Genes in rice (*Oryza Sativa*). *J. Rice Res.* 4 (164), 2. doi:10.4172/2375-4338.1000164

Bhati, J., Chaduvula, P. K., Rani, R., Kumar, S., and Rai, A. (2014). In-silico Prediction and Functional Analysis of Salt Stress Responsive Genes in maize (*Zea mays*). *Eur. J. Mol. Biol. Biochem.* 1, 151–157.

Bouck, A., and Vision, T. (2007). The Molecular Ecologist's Guide to Expressed Sequence Tags. *Mol. Ecol.* 16 (5), 907–924. doi:10.1111/j.1365-294x.2006.03195.x

Casartelli, A., Melino, V. J., Baumann, U., Riboni, M., Suchecki, R., Jayasinghe, N. S., et al. (2019). Opposite Fates of the Purine Metabolite Allantoin under Water and Nitrogen Limitations in Bread Wheat. *Plant Mol. Biol.* 99 (4), 477–497. doi:10.1007/s11103-019-00831-z

Chaduvula, P. K., Bhati, J., Rai, A., Gaikwad, K., Marla, S. S., Elangovan, M., et al. (2015). In-silico' Expressed Sequence Tag Analysis in Identification and Characterization of Salinity Stress Responsible Genes in '*Sorghum Bicolor*. *Aust. J. Crop Sci.* 9 (9), 799–806.

Chen, F., Li, Q., Sun, L., and He, Z. (2006). The rice 14-3-3 Gene Family and its Involvement in Responses to Biotic and Abiotic Stress. *DNA Res.* 13 (2), 53–63. doi:10.1093/dnares/dsl001

Chen, Y.-S., David Ho, T.-H., Liu, L., Lee, D. H., Lee, C.-H., Chen, Y.-R., et al. (2019). Sugar Starvation-Regulated MYBS2 and 14-3-3 Protein Interactions Enhance Plant Growth, Stress Tolerance, and Grain Weight in rice. *Proc. Natl. Acad. Sci. USA* 116 (43), 21925–21935. doi:10.1073/pnas.1904818116

Conesa, A., and Götz, S. (2008). Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int. J. Plant Genomics* 2008, 1–12. doi:10.1155/2008/619832

Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of Abiotic Stress on Plants: a Systems Biology Perspective. *BMC Plant Biol.* 11, 163. doi:10.1186/1471-2229-11-163

Feist, A. M., and Palsson, B. Ø. (2008). The Growing Scope of Applications of Genome-Scale Metabolic Reconstructions Using *Escherichia coli*. *Nat. Biotechnol.* 26 (6), 659–667. doi:10.1038/nbt1401

Gahlaut, V., Jaiswal, V., Kumar, A., and Gupta, P. K. (2016). Transcription Factors Involved in Drought Tolerance and Their Possible Role in Developing Drought Tolerant Cultivars with Emphasis on Wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 129 (11), 2019–2042. doi:10.1007/s00122-016-2794-z

Gangola, M. P., and Ramadoss, B. R. (2018). "Sugars Play a Critical Role in Abiotic Stress Tolerance in Plants," in *Biochemical, Physiological and Molecular Avenues for Combating Abiotic Stress Tolerance in Plants* (Academic Press), 17–38. doi:10.1016/b978-0-12-813066-7.00002-4

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

Guo, Y., Jiang, Q., Hu, Z., Sun, X., Fan, S., and Zhang, H. (2018). Function of the Auxin-Responsive Gene TaSAUR75 under Salt and Drought Stress. *Crop J.* 6 (2), 181–190. doi:10.1016/j.cj.2017.08.005

Gupta, A., Rico-Medina, A., and Caño-Delgado, A. I. (2020). The Physiology of Plant Responses to Drought. *Science* 368, 266–269. doi:10.1126/science.aaz7614

Islam, M. R., Hosen, M. I., Zaman, A., and Islam, M. O. (2013). Structural, Functional and Molecular Docking Study to Characterize GMI1 from *Arabidopsis thaliana*. *Interdiscip. Sci. Comput. Life Sci.* 5 (1), 13–22. doi:10.1007/s12539-013-0153-1

Jacob, P., Hirt, H., and Bendahmane, A. (2017). The Heat-Shock Protein/chaperone Network and Multiple Stress Resistance. *Plant Biotechnol. J.* 15 (4), 405–414. doi:10.1111/pbi.12659

Jaiswal, P. (2011). Gramene Database: a Hub for Comparative Plant Genomics. *Methods Mol. Biol.* 678, 247–275. doi:10.1007/978-1-60761-682-5_18

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Kumar, A., Pandeya, A., Malik, G., Sharma, M., P., H. K., S., A. K., et al. (2018). A Web Resource for Nutrient Use Efficiency-Related Genes, Quantitative Trait Loci and microRNAs in Important Cereals and Model Plants. *F1000Res* 7, 673. doi:10.12688/f1000research.14561.1

Lai, K., Berkman, P. J., Lorenc, M. T., Duran, C., Smits, L., Manoli, S., et al. (2012). WheatGenome.info: An Integrated Database and Portal for Wheat Genome Information. *Plant Cel Physiol.* 53 (2), e2. doi:10.1093/pcp/pcr141

Lawrence, C. J., Dong, Q., Polacco, M. L., Seigfried, T. E., and Brendel, V. (2004). MaizeGDB, the Community Database for maize Genetics and Genomics. *Nucleic Acids Res.* 32 (Suppl. l_1), 393D–397D. doi:10.1093/nar/gkh011

Lee, D.-Y., Saha, R., Yusufi, F. N. K., Park, W., and Karimi, I. A. (2009). Web-based Applications for Building, Managing and Analysing Kinetic Models of Biological Systems. *Brief. Bioinform.* 10 (1), 65–74. doi:10.1093/bib/bbn039

Li, R., Jiang, X., Jin, D., Dhaubhadel, S., Bian, S., and Li, X. (2015). Identification of 14-3-3 Family in Common Bean and Their Response to Abiotic Stress. *PLoS One* 10 (11), e0143280. doi:10.1371/journal.pone.0143280

Li, Y., and Xu, M. (2017). CCT Family Genes in Cereal Crops: A Current Overview. *Crop J.* 5 (6), 449–458. doi:10.1016/j.cj.2017.07.001

Masoudi-Nejad, A., Tonomura, K., Kawashima, S., Moriya, Y., Suzuki, M., Itoh, M., et al. (2006). EGassembler: Online Bioinformatics Service for Large-Scale Processing, Clustering and Assembling ESTs and Genomic DNA Fragments. *Nucleic Acids Res.* 34 (Suppl. l_2), W459–W462. doi:10.1093/nar/gkl066

Moin, M., Bakshi, A., Madhav, M. S., and Kirti, P. B. (2017). Expression Profiling of Ribosomal Protein Gene Family in Dehydration Stress Responses and Characterization of Transgenic rice Plants Overexpressing RPL23A for Water-Use Efficiency and Tolerance to Drought and Salt Stresses. *Front. Chem.* 5, 97. doi:10.3389/fchem.2017.00097

Nagaraj, S. H., Deshpande, N., Gasser, R. B., and Ranganathan, S. (2007). ESTExplorer: an Expressed Sequence Tag (EST) Assembly and Annotation Platform. *Nucleic Acids Res.* 35 (Suppl. l_2), W143–W147. doi:10.1093/nar/gkm378

Prabha, R., Ghosh, I., and Singh, D. P. (2011). Plant Stress Gene Database: A Collection of Plant Genes Responding to Stress Condition. *ARPN J. Sci. Tech.* 1, 28–31.

Shanker, A. K., Maddaala, A., Kumar, M. A., Yadav, S. K., Maheswari, M., and Venkateswarlu, B. (2012). In Silico targeted Genome Mining and Comparative Modelling Reveals a Putative Protein Similar to an *Arabidopsis* Drought Tolerance DNA Binding Transcription Factor in Chromosome 6 of *Sorghum Bicolor* Genome. *Interdiscip. Sci. Comput. Life Sci.* 4 (2), 133–141. doi:10.1007/s12539-012-0121-1

Stortenbeker, N., and Bemer, M. (2019). TheSAURgene Family: the Plant's Toolbox for Adaptation of Growth and Development. *J. Exp. Bot.* 70 (1), 17–27. doi:10.1093/jxb/ery332

Takahashi, F., Kuromori, T., Urano, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2020). Drought Stress Responses and Resistance in Plants: from Cellular Responses to Long-Distance Intercellular Communication. *Front. Plant Sci.* 11, 556972. doi:10.3389/fpls.2020.556972

Tian, F., Wang, T., Xie, Y., Zhang, J., and Hu, J. (2015). Genome-wide Identification, Classification, and Expression Analysis of 14-3-3 Gene Family in *Populus*. *PLoS One* 10 (4), e0123225. doi:10.1371/journal.pone.0123225

Vincent, J., Dai, Z., Ravel, C., Choulet, F., Mouzeyar, S., Bouzidi, M. F., et al. (2013)dbWFA: a Web-Based Database for Functional Annotation of *Triticum aestivum* Transcripts. *Database* 2013. doi:10.1093/database/bat014

Wani, S. H., Khan, H., Riaz, A., Joshi, D. C., Hussain, W., Rana, M., et al. (2022). "Genetic Diversity for Developing Climate-Resilient Wheats to Achieve Food Security Goals," in *Advances in Agronomy* (Academic Press), 255–303. doi:10.1016/bs.agron.2021.08.006

Wani, S. H., Tripathi, P., Zaid, A., Challa, G. S., Kumar, A., Kumar, V., et al. (2018). Transcriptional Regulation of Osmotic Stress Tolerance in Wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* 97 (6), 469–487. doi:10.1007/s11103-018-0761-6

Watanabe, S., Matsumoto, M., Hakomori, Y., Takagi, H., Shimada, H., and Sakamoto, A. (2014). The Purine Metabolite Allantoin Enhances Abiotic Stress Tolerance through Synergistic Activation of Abscisic Acid Metabolism. *Plant Cel Environ* 37 (4), 1022–1036. doi:10.1111/pce.12218

Wen, F., Wu, X., Li, T., Jia, M., Liu, X., Li, P., et al. (2017). Genome-wide Survey of Heat Shock Factors and Heat Shock Protein 70s and Their Regulatory Network under Abiotic Stresses in *Brachypodium Distachyon*. *PLoS One* 12 (7), e0180352. doi:10.1371/journal.pone.0180352

Wilkinson, P. A., Winfield, M. O., Barker, G. L., Allen, A. M., Burridge, A., Coghill, J. A., et al. (2012). CerealsDB 2.0: an Integrated Resource for Plant Breeders and Scientists. *BMC Bioinformatics* 13 (1), 1–6. doi:10.1186/1471-2105-13-219

Yang, Z.-P., Li, H.-L., Guo, D., Tang, X., and Peng, S.-Q. (2014). Identification and Characterization of the 14-3-3 Gene Family in *Hevea Brasiliensis*. *Plant Physiol. Biochem.* 80, 121–127. doi:10.1016/j.plaphy.2014.03.034

Zang, D., Li, H., Xu, H., Zhang, W., Zhang, Y., Shi, X., et al. (2016). An *Arabidopsis* Zinc finger Protein Increases Abiotic Stress Tolerance by Regulating Sodium and Potassium Homeostasis, Reactive Oxygen Species Scavenging and Osmotic Potential. *Front. Plant Sci.* 7, 1272. doi:10.3389/fpls.2016.01272

Zeng, V., and Extavour, C. G. (20122012). ASGARD: an Open-Access Database of Annotated Transcriptomes for Emerging Model Arthropod Species. *Database* 2012, bas048. doi:10.1093/database/bas048

Zhang, H., Zhao, Y., and Zhu, J.-K. (2020). Thriving under Stress: How Plants Balance Growth and the Stress Response. *Develop. Cel* 55, 529–543. doi:10.1016/j.devcel.2020.10.012

Zhang, H., Zhu, J., Gong, Z., and Zhu, J.-K. (2021). Abiotic Stress Responses in Plants. *Nat. Rev. Genet.* 23, 104–119. doi:10.1038/s41576-021-00413-0

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

The reviewer UK declared a past co-authorship with the author AK 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 Kumar, Bhati, Saha, Lal, Pandey, Mishra, Farooqi, Kumar, Chaturvedi and Rai. 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: 09 March 2022
doi: 10.3389/fgene.2022.821163

Global Survey, Expressions and Association Analysis of *CBLL* Genes in Peanut

Weifang Ren[1,2,3], *Zhaocong Zeng*[1,2,3], *Sijian Wang*[1,2,3], *Juncheng Zhang*[4], *Jiahai Fang*[1,2,3]* *and Liyun Wan*[1,2,3]*

[1]Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang, China, [2]Southern Regional Collaborative Innovation Center for Grain and Oil Crops in China, Nanchang, China, [3]College of Agronomy, Jiangxi Agricultural University, Nanchang, China, [4]Huazhong Agricultural University, Wuhan, China

Cystathionine γ-synthase (CGS), methionine γ-lyase (MGL), cystathionine β-lyase (CBL) and cystathionine γ-lyase (CGL) share the Cys_Met_Meta_PP domain and play important roles in plant stress response and development. In this study, we defined the genes containing the Cys_Met_Meta_PP domain (PF01053.20) as *CBL*-like genes (*CBLL*). Twenty-nine *CBLL* genes were identified in the peanut genome, including 12 from cultivated peanut and 17 from wild species. These genes were distributed unevenly at the ends of different chromosomes. Evolution, gene structure, and motif analysis revealed that *CBLL* proteins were composed of five different evolutionary branches. Chromosome distribution pattern and synteny analysis strongly indicated that whole-genome duplication (allopolyploidization) contributed to the expansion of *CBLL* genes. Comparative genomics analysis showed that there were three common collinear *CBLL* gene pairs among peanut, Arabidopsis, grape, and soybean, but no collinear *CBLL* gene pairs between peanut and rice. The prediction results of *cis*-acting elements showed that *AhCBLLs*, *AdCBLLs, and AiCBLLs* contained different proportions of plant growth, abiotic stress, plant hormones, and light response elements. Spatial expression profiles revealed that almost all *AhCBLLs* had significantly higher expression in pods and seeds. All *AhCBLL*s could respond to heat stress, and some of them could be rapidly induced by cold, salt, submergence, heat and drought stress. Furthermore, one polymorphic site in *AiCBLL7* was identified by association analysis which was closely associated with pod length (PL), pod width (PW), hundred pod weight (HPW) and hundred seed weight (HSW). The results of this study provide a foundation for further research on the function of the *CBLL* gene family in peanut.

Keywords: peanut, cis-acting element, genome-wide association analysis, CBLL gene family, expression pattern

OPEN ACCESS

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

Reviewed by:
Swarup Roy Choudhury, Indian Institute of Science Education and Research, India
Santosh Kumar Upadhyay, Panjab University, India
Zulfiqar Ali, Muhammad Nawaz Shareef University of Agriculture, Pakistan

***Correspondence:**
Jiahai Fang
fangjiahai2019@126.com
Liyun Wan
wanliyun2019@163.com

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

Received: *23 November 2021*
Accepted: *28 January 2022*
Published: *09 March 2022*

Citation:
Ren W, Zeng Z, Wang S, Zhang J, Fang J and Wan L (2022) Global Survey, Expressions and Association Analysis of CBLL Genes in Peanut. Front. Genet. 13:821163. doi: 10.3389/fgene.2022.821163

INTRODUCTION

Sulfur-containing amino acids play an important role in the growth and development of plants and animals (Lorraine et al., 1986). When plants lack sulfur-containing amino acid, their metabolic processes will be abnormal and growth will be affected (Kery et al., 1994). Methionine (Met) is a sulfur-containing amino acid that is essential to all organisms and indirectly regulates a variety of cellular processes through S-adenosine methionine (SAM) (Amir, 2010). SAM is also a methyl donor

for the methylation of proteins, lipids, DNA, and RNA, and is a precursor to biosynthesis of the plant hormones ethylene, polyamines, and biotin (Benkova et al., 2003). A recent study showed that Met activates GLR (glutamate receptor), thereby activating Ca^{2+} channels that regulate stomatal movement and plant growth (Galili et al., 2016). Met also promotes root morphogenesis and enhances chlorophyll content to enhance photosynthesis (Sarropoulou et al., 2013).

Met biosynthesis requires cystathionine γ-synthase (CGS), cystathionine β-lyase (CBL), and Met synthase enzymes (Messerschmidt et al., 2003). In plants, CGS first converts cystine (Cys) and activates the conversion of homoserine to cystathionine. Cystathionine is then cleaved by CBL to yield homoCys, pyruvate, and ammonia. Met synthase finally methylates homo-Cys to produce Met (Ravanel et al., 1998). Animals and some microorganisms can metabolize Met back to Cys by a reverse trans-sulfuration pathway that involves cystathionine β-synthase (CBS) and cystathionine γ-lyase (CGL). Plants and certain bacteria are proposed to have an alternative route (or routes) to convert methionine to cysteine, of which the first step is mediated by methionine γ-lyase (MGL). The subsequent steps in this pathway(s) have not been definitively established (Goyer et al., 2007). Arabidopsis mutants perturbed in Met metabolism have been described,*mto* alleles carried single base-pair mutation in a conserved domain led to over-accumulation of Met (Goto et al., 2002). Over-expression of *AtCGS* caused an increased level of Met, which induced the up-regulation of genes involved in ethylene and abscisic acid homeostasis and light, sucrose, salt and osmotic stresses regulation (Hacham et al., 2013; Cohen et al., 2014, 2017; Whitcomb et al., 2018). Pharmacological treatment of plants with inhibitors against *CGS* or *CBL* induced a deficiency in Met biosynthesis and caused growth inhibition (Ravanel et al., 1998). Knockdown of *CGS* and *CBL* in Arabidopsis led to abnormal leaf development stunted (Kim and Leustek., 2000; Levin et al., 2000). CBL is crucial for embryo patterning and the maintenance of the root stem cell niche in Arabidopsis (Liu et al., 2019). The Met homeostasis gene *METHIONINE GAMMA LYASE* (*AtMGL*) is up-regulated by dual stress in leaves, conferring resistance to nematodes when overexpressed, *AtMGL* regulates Met metabolism under conditions of multiple stressors, Met degradation, and plays a subordinate role to threonine deaminase (Goyer et al., 2007; Joshi and , 2009; Atkinson et al., 2013). A similar strategy has been applied to produce transgenic plants with reduced *CBL* levels in potato. These *CBL* antisense plants exhibited a short bushy stature, altered leaf morphology, and small tuber size (Maimann et al., 2000). In soybean, the overexpress of *AtD-CGS* notably increased the level of soluble Met in developing green seeds (3.8–7-fold), and these soybean seeds also showed high levels of other amino acids; furthermore, the total Met content, which included Met incorporated into proteins, notably increased in the mature dry seeds of these two transgenic lines by 1.8- and 2.3-fold, respectively (Song et al., 2013).

Peanut is an annual leguminous herb, is widely cultivated around the world as an important source of oil and protein for humans (Yang et al., 2014; Liao et al., 2018; Bertioli et al., 2019). Peanut kernels are rich in natural nutrients such as proteins, fatty acids (FAs), vitamins, minerals, and fiber (Chen et al., 2010). Studies have demostrated that the protein content of peanut seeds is around 23–33%; however, the content of sulfur-containing Met and Cy is low, especially for the former (Yang et al., 2001). Met and the enzyme coding genes involved in the Met/Cys interconversion pathway play inhibitory roles in both amino acid composition and content of storage protein, and also in tissue development and stress response regulation. Therefore, it is of great significance to further explore the *CBLL* gene family in peanut and identifythe function of family members in improving peanut yield, quality, and stress tolerance. In this study, we defined the genes containing the Cys_Met_Meta_PP domain (PF01053.20) as *CBL*-like genes (*CBLL*). Twenty-nine *CBLL* genes were identified from the peanut genome, whose conserved domains, phylogenetic tree, chromosome distribution, gene structure and expression pattern were analyzed. The results provided a basis for the role of peanut *CBLL* in the development and formation of peanut pods, and stress response regulation. This research laid a foundation for the identification and utilization of peanut *CBLL* genes, which is of great significance for the molecular based breeding of cultivars with multi-resistance, high yield, and good quality.

MATERIALS AND METHODS

Identification of the *CBLL* Genes in Peanut

To identify *CBLL* genes in peanut, the predicted protein sequences were downloaded from PeanutBase (https://peanutbase.org) (*Arachis Duranensis* V14167: A-genome; *Arachis Ipaensis* K30076: B-genome). The Cys_Met_Meta_PP domain (PF01053.20) was identified from the protein sequences by using the HMMER 3.0 program at a standard E-value < 1×10^{-5} (Deng et al., 2019; Misra et al., 2019). Conserved domain searches were performed against the conserved domain database in NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Members with incomplete conserved functional domains were removed. The *CBLL* genes were named as *AdCBLL1* to *AdCBLL9*, *AiCBLL1* to *AiCBLL8*, and *AhCBLL1* to *AhCBLL12* according to their positions on peanut chromosomes. Physicochemical parameters of peanut CBLL proteins were then generated by ProtParam Tools, including theoretical isoelectric points (pI) and molecular weights (MW) (Gasteiger et al., 2005).

Analyses of Phylogeny, Gene Structure, and Conserved Motifs

Phylogenetic analysis was carried out based on the protein sequences of the *CBLL* genes in peanut (*A.hypogaea* L., *A.duranensis, A.ipaensis*) and Arabidopsis. The protein sequences were aligned by ClustalW, and the unrooted Neighbor-Joining (NJ) phylogenetic tree was constructed by MEGA 5.2 software with 1,000 bootstrap replicates. Gene annotation information was downloaded from PeanutBase

(http://www.peanutbase.org/) and GSDS 2.0 (http://gsds.gao-lab.org/) was used to visualize the gene structure. The composition of conserved motifs was searched by the Multiple EM for Motif Elicitation (MEME) online tool by setting a maximum number as 20 (http://meme-suite.org/tools/meme).

Gene Duplication and Synteny Analysis of the Peanut *CBLL* Genes

MCScan (http://chibba.agtec.uga.edu/duplication/mcscan) was used to identify *AhCBLs* duplications and the synteny block of *CBLL* genes of peanut and the other four species (Arabidopsis, rice, grape and soybean). BLASTP was applied to find homologous sequences of *CBLL* genes between peanut and Arabidopsis, after that literatures were reviewed to explore the function of published *CBLL* genes in Arabidopsis. Tandem duplications were defined as adjacent homologous genes on the same chromosome with a distance of <50 kb (Cannon et al., 2004). If they were paralogs located on duplicated chromosomal blocks, they were defined as a segmental duplication event (Guo et al., 2016). Non-synonymous (Ka) and synonymous (Ks) substitution of each duplicated genes were calculated using the PAL2NAL program (Suyama et al., 2006), which was based on the codon model program in PAML (Yang, 2007).

Cis-Acting Element Analysis in the Promoters of Peanut *CBLLs*

Regulatory elements of promoter sequences can control gene expression. The 2-kb promoter sequences of 29 *CBLL* genes were downloaded from PeanutBase (http://www.peanutbase.org/) and used to predict the *cis*-regulatory element through the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot, 2002). The radar figures were manually generated by R3.5.1 scripts.

Expression Profiles of *AhCBLL* Genes in Different Tissues

RNA-seq data sets of 22 peanut tissues were downloaded from PeanutBase and the NCBI SRA database to explore the expression profiles of *AhCBLL* genes in different tissues, which were submitted by Clevenger et al. (Clevenger et al., 2016).

Plant Materials, Growth Conditions and Treatments

Different experimental treatments were carried out for 10-day-old seedlings of a Chinese elite peanut cultivar Changhua18 which were planted in vermiculite and irrigated with sterilized water (26°C, 16-h light/8-h dark). For hormone treatments, seedlings were sprayed with solutions containing 6-benzyl amino purine (6-BA) (25 μM), indole-3-acetic acid (IAA) (50 μM), gibberellic acid (GA) (100 μM), salicylic acid (SA) (100 μM), abscisic acid (ABA) (100 μM), ethylene (ACC) (500 μM) and methyl jasmonate (MJ) (100 μM)(Jain et al., 2006; Wang et al., 2016). For heat and cold stresses, the 10-day-old seedlings were transferred to two environmental temperatures of 40°C (H40) and 4°C (Cold 4) respectively (Jain et al., 2006; Wang et al., 2016). Seedling samples were collected at 0, 1, 3, 6, 9 and 12 h after the above treatment. For submergence (Sub), the seedlings were soak in water to a depth of 5 cm from tip to surface, and samples were collected at 0, 6, 12, 24, 48, and 72 h after treatment (Wang et al., 2016). For the NaCl and polyethylene glycol (PEG) treatments, the 10-day-old seedlings were immersed in NaCl solution (200 mM) and PEG6000 (20%, w/v) (Jain et al., 2006; Song et al., 2009), and samples were collected after 0, 0.5, 1, 3, 6, 12, and 24 h after seedling treatment. Three biological replicates were performed; each sample included around eight seedlings. All samples were placed in liquid nitrogen during sampling and stored at −80°C to preserve RNA integrity.

The genotype data of the *CBLL* genes used here were obtained from transcriptome sequencing data of a peanut germplasm population with 146 accessions (unpublished data). Each line of the peanut populations was planted in five different environments (Wuhan 2016, Wuhan 2017, Yangluo 2016, Yangluo 2017, and Zhanjiang 2016). All seedlings were planted within the experimental plot with 12 plants in a line in each environment.

RNA Isolation and qRT-PCR Analysis

Total RNA of samples were isolated using TRIzol reagent (Invitrogen) according to the manufacture's requirements. M-MLV reverse transcriptase (Promega) was used to synthesize the first chain of cDNA from 5 μg total RNA. Quantitative Real-Time PCR (qRT-PCR) was performed using 2×SYBR Green Master Mix (Bio-Rad) on a 96-well plate with a gene-specific primer (**Supplementary Table S1**). The thermal cycle was as follows: 95°C for 5 min; 40 cycles of 95°C for 10 s, primer-specific annealing temperature of, 72°C for 10 s, for 15 cycles; then the melt curve was from 65 to 95°C.

RESULTS

Identification of the *CBLL* Gene Family in Peanut

In order to identify *CBLL* gene families in peanut, we downloaded the published peanut genome sequence from PeanutBase (https://peanutbase.org/). The Cys_Met_Meta_PP domain (PF01053.20) containing proteins were identified by HMMER 3.0 with a standard E-value < 1×10^{-5}; further, we removed the incomplete sequences, and identified 29 *AhCBLL* members (**Table 1**) from the cultivated peanut (*A. hypogaea* L.) and its diploid progenitors (*A. duranensis*, *A. ipaensis*). **Table 1** summarizes genes with complete sequences, among which 17 *CBLLs* were from A-genome, and 12 *CBLLs* were from B-genome; they were distributed unevenly across chromosomes. The 12 members from *A. hypogaea* L. were named as *AhCBLL1~AhCBLL12*, the nine *AhCBLLs* from *A. duranensis* were named as *AdCBLL1~AdCBLL9*, and the eight

TABLE 1 | *CBLL* genes identified in peanuts.

Gene name	Gene locus	CDS length (bp)	AA[a]	MW (kDa)[b]	pI[c]	TMD[d]	Chr
AdCBLL1	*Aradu.02IGP*	1,520	426	46.17	5.93	0	Aradu.A04
AdCBLL2	*Aradu.M0JX8*	1,158	248	27.97	8.77	0	Aradu.A04
AdCBLL3	*Aradu.V9ADB*	618	77	8.44	7.82	0	Aradu.A04
AdCBLL4	*Aradu.N0LVM*	772	182	20.62	5.01	2	Aradu.A04
AdCBLL5	*Aradu.W013I*	1,545	497	53.51	6.02	0	Aradu.A04
AdCBLL6	*Aradu.JQ7JG*	1,843	464	50.42	6.18	0	Aradu.A06
AdCBLL7	*Aradu.93LBV*	1,020	208	22.83	4.96	0	Aradu.A06
AdCBLL8	*Aradu.UE7BN*	1,758	308	33.45	6.30	0	Aradu.A10
AdCBLL9	*Aradu.FE0Z7*	1,441	313	35.17	6.76	0	None
AiCBLL1	*Araip.3V2A7*	946	204	22.36	9.07	0	Araip.B01
AiCBLL2	*Araip.V3B0A*	1,654	426	46.07	6.01	0	Araip.B04
AiCBLL3	*Araip.I7QC8*	1,891	137	15.40	6.09	0	Araip.B04
AiCBLL4	*Araip.26T6F*	1,635	471	51.31	7.64	0	Araip.B04
AiCBLL5	*Araip.P8SRT*	1,858	464	50.39	6.06	0	Araip.B06
AiCBLL6	*Araip.B0SMD*	2,090	141	15.03	5.28	0	Araip.B09
AiCBLL7	*Araip.H1TPN*	1,724	442	48.10	6.25	0	Araip.B09
AiCBLL8	*Araip.KUG1C*	2,237	564	60.63	5.90	0	Araip.B10
AhCBLL1	*Arahy.AA870A*	1,876	250	26.57	5.30	0	Arahy.04
AhCBLL2	*Arahy.C0PD3X*	462	153	16.85	8.45	0	Arahy.04
AhCBLL3	*Arahy.2E7M4N*	623	204	22.79	4.92	0	Arahy.04
AhCBLL4	*Arahy.0JH0K6*	1,822	531	56.85	6.40	0	Arahy.04
AhCBLL5	*Arahy.F5KNV4*	5,885	1,855	205.59	6.00	0	Arahy.06
AhCBLL6	*Arahy.AGM2GS*	2,887	560	60.28	5.85	0	Arahy.10
AhCBLL7	*Arahy.V9A8GQ*	1,852	426	46.07	6.01	0	Arahy.14
AhCBLL8	*Arahy.Q4SW2C*	915	139	15.69	8.13	1	Arahy.14
AhCBLL9	*Arahy.I2BVW6*	1,815	531	56.81	6.40	0	Arahy.14
AhCBLL10	*Arahy.0AC10P*	5,944	1,830	202.91	5.98	0	Arahy.16
AhCBLL11	*Arahy.1FPN7V*	1,646	442	48.11	6.25	0	Arahy.19
AhCBLL12	*Arahy.S7AN2D*	2,691	564	60.63	5.90	0	Arahy.20

[a]*Length of the amino acid sequence.*
[b]*Molecular weight of the amino acid sequence.*
[c]*Isoelectric point of the AhCBLL.*
[d]*Number of transmembrane domains, as predicted by the TMHMM Server v2.0.*

AhCBLLs from *A.ipaensis* were named as *AiCBLL2~AiCBLL8*, respectively, according to their chromosomal order. We then determined the chromosome location, amino acid number (AA), mRNA length, theoretical isoelectric points (pI) and other information of peanut CBLLs (**Table 1**). The open reading frame (ORF) lengths of the *CBLL* genes ranged from 462 bps to 5,944 bps (**Table 1**). The protein sequences of the peanut *CBLL* genes were significantly different; sequence lengths ranged from 77 to 1855 aa. The molecular weights (MWs) of AdCBLLs varied from 8.44 kDa (AdCBLL3) to 53. 51 kDa (AdCBLL5); AiCBLLs ranged from 15.03 kDa (AiCBLL6) to 60.63 kDa (AiCBLL8); and AhCBLLs varied from 15.69 kDa (AhCBLL8) to 205.59 kDa (AhCBLL5). The pI was small for the overwhelming majority of CBLLs, ranging from 4.92 (AhCBLL3) to 9.07 (AiCBLL1). AhCBLL8 carried one conservative transmembrane domain (TMDs), AdCBLL4 carried two TMDs, and other peanut CBLLs did not contain TMDs.

Phylogeny, Gene Structure, and Conserved Motifs of the *CBLL* Gene Family in Peanut

Arabidopsis thaliana emerged is the model organism of choice for in plant biology research, we screened CBLL members through the *A. thaliana* genome, and four members were detected. To investigate the evolutionary relationships of the *CBLL* family genes in peanut, we conducted an NJ-phylogenetic tree, and analyzed the gene exon/intron structural and conserved motifs. The results demonstrated that the 29 peanut and four Arabidopsis CBLL proteins could be integrated into five clades (**Figure 1A**). Eight, six, one, four, and 10 CBLLs pertained to clade I (CGS clade), clade II (MGL clade), clade III (new clade), clade IV (CBL clade) and clade V (new clade), respectively (**Figure 1A**). All the four large clades included CBLLs from the cultivated peanut and its two diploid progenitors. Interestingly, these five distinct groups have different gene structure and motif arrangement. The CGS clade had approximately 11 exons, and only *AiCBLL3* had 10 exons. In sharp contrast to those in clade I, the MGL clade possessed two-three exons with one exception (*AiCBLL1*, 10 exon); the CBL clade had the greatest exon number, ranging from 12 to 29. For the two new clades, the only gene in clade III had 13 exons and clade V had 2–13 exons (**Figure 1B**). Further, we identified 20 different conserved motifs (**Figure 1C**). In general, clade I, clade II, and clade IV had more motifs than clade III and clade V. Motif 1 was the most common, present in all *CBLL* genes except *AdCBLL8*, *AiCBLL5,* and *AhdCBLL3*. Otherwise, the vast majority of CBLLs included motif 2, 3, 4, 6,

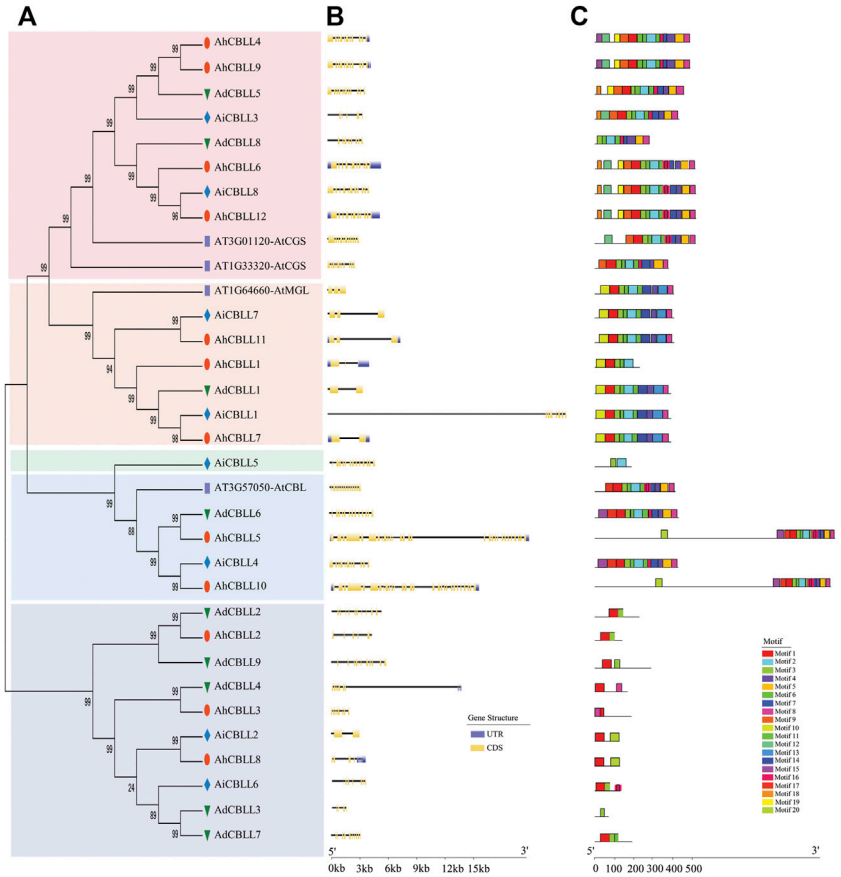


FIGURE 1 | Comparison of the gene structure and motif of 29 *CBLL* genes in peanut. **(A)** Unrooted phylogenetic tree with over 50% bootstrap value above the branch. The clade I, II, III, IV and V were displayed in pink, orange, green, blue and dark blue colors, separately. The names of species were abbreviated to two letters, named as *Arachis duranensis* (Ad), *Arachis ipaensis* (Ai), *Arachis hypogaea* L. *Tifrunner* (Ah). **(B)** Exon/introns and untranslated regions (UTRs) of *CBLL*s. Green boxes denote UTR (untranslated region); yellow boxes denote CDS (coding sequence); black lines denote introns. The length of protein can be estimated using the scale at the bottom. **(C)** Motif architectures of all *CBLL* genes. Each motif is illustrated with a specific color, and the distribution of identified motifs corresponds to their positions.

8, and 11. Motif 18, and 19 were clade-specific elements in clade I, motif 13 and 14 only existed in clade II, and motif 15 only existed in clade IV.

Gene Duplication and Synteny Analyses of Peanut *CBLLs*

Chromosomal location analyses revealed that the 12 *AhCBLLs* were distributed unevenly on seven chromosomes (chromosomes 04, 06, 10, 14, 16, 19 and 20); nine *AdCBLLs* were present on chromosomes A04, A06 and A10; and eight *AiCBLLs* were distributed on chromosomes B01, B04, B06, B09, and B10 (**Figure 2**). A total of 17 chromosomal fragment repeat gene pairs were identified without tandem repeats (**Figure 3** and **Supplementary Tables S2–S5**). Genomic synteny analyses between cultivated and wild peanut species uncovered five (*AiCBLL1, AiCBLL6, AdCBLL3, AdCBLL7 and AdCBLL9*) wild species-specific *CBLL* members (**Figures 2**, **3** and **Supplementary Tables S2–S5**). Further, we calculated the Ks (synonymous) and Ka (non-synonymous) values of the duplicated gene pairs and found that the Ka/Ks ratio for duplicated *AhCBLL* gene pairs ranged from 0.03 to 0.95 with an average of 0.25 (**Supplementary Table S6**). The ω values of all duplicated gene pairs were less than one, domostrated that purifying selection occurred on these duplicated gene pairs. The whole genome-wide collinear analysis identified that 41.38% (six *AhCBLL* genes, six pairs, **Supplementary Table S7**; three *AdCBLL* genes, three pairs; three *AiCBLL* genes, three pairs, **Supplementary Table S8**), 41.38% (six *AhCBLL* genes, eight pairs, **Supplementary Table S9**; three *AdCBLL* genes, three pairs; three *AiCBLL* genes, three pairs, **Supplementary Table S10**), and 48.28% (six *AhCBLL* genes, six pairs, **Supplementary Table S11**; four *AdCBLL* genes, seven pairs; four *AiCBLL* genes, seven pairs, **Supplementary Table S12**) of the *AhCBLLs* were orthologous with Arabidopsis, grape, and soybean *CBLLs*, but none with rice, respectively (**Figure 4**). Synteny analysis with soybean, Arabidopsis, and grape revealed three conserved *CBLL* genes (*AhCBLL4, AhCBLL6,* and *AhCBLL9*) in these species. Both collinear and BLAST methods were used to

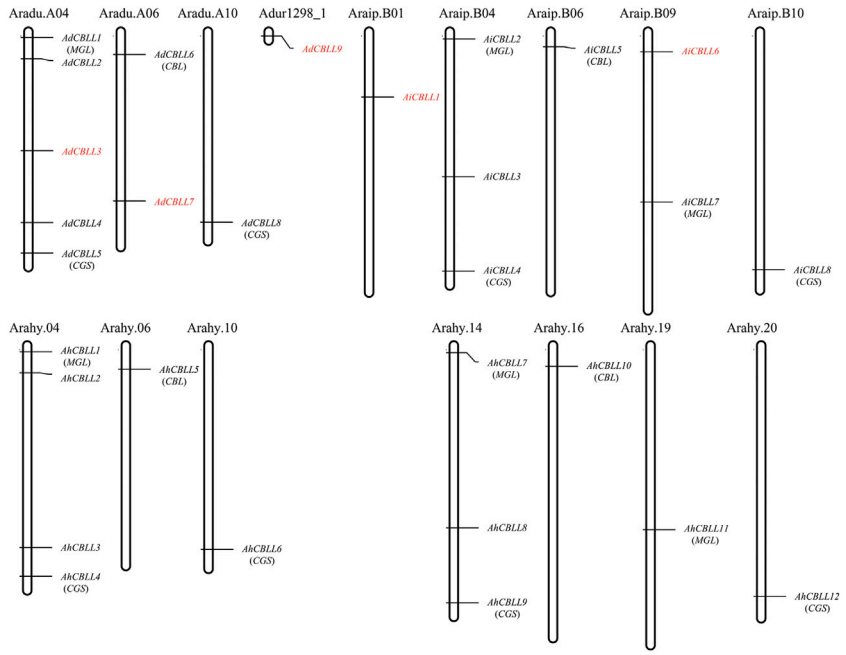


FIGURE 2 | Chromosomal locations of peanut *CBLL* genes. Chromosomal positions of the peanut *CBLL* genes were mapped based on data from PeanutBase. The chromosome number was indicated above each chromosome. Genes in red indicated wild species specific.

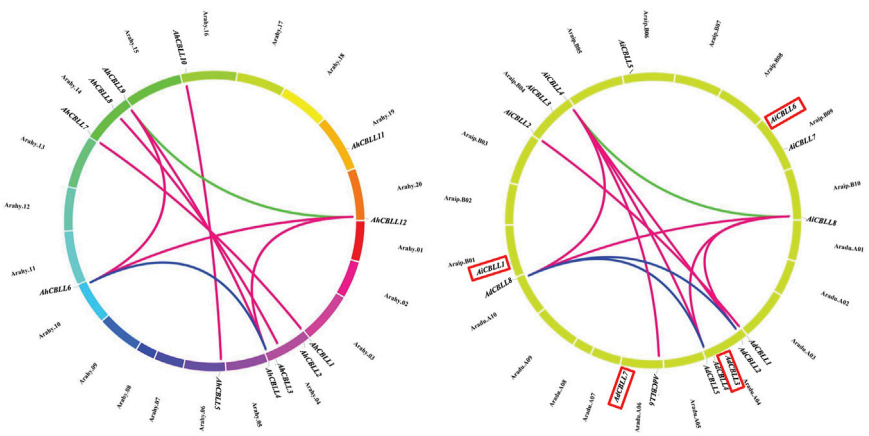


FIGURE 3 | Chromosomal distribution and gene duplications of the *AdCBLL*, *AiCBLL*, and AhCBLL genes. The scales on the circle were in Megabases. Each colored bar represented a chromosome as indicated. Gene IDs were labeled on the basis of their positions on the chromosomes. Red frames indicated wild species specific CBLL genes.

identify *AhCBLL* gene orthologs between peanut and Arabidopsis; 13 orthologous pairs were found (**Table 2**). The orthologs in Arabidopsis included *AtMGL* participating in Met degradation and plant defense (Ricarda et al., 2005; Goyer et al., 2007; Joshi and Jander, 2009; Atkinson et al., 2013), *AtMOT1/CGS* relating to the regulation of seed growth and metabolism in Arabidopsis (Goto et al., 2002; Hacham et al., 2013; Cohen et al., 2014, 2017; Whitcomb et al., 2018), and *AtCBL* relating to the synthesis of plant hormones (Kim and Leustek., 2000; Levin et al., 2000; Liu et al., 2019). Therefore, we speculated that these *AhCBLL* homologous genes might play multiple roles in peanut growth, development, and stress resistance.

Cis-Acting Element Prediction of Peanut *CBLLs*

We analyzed and predicted *cis*-acting elements in the 2-kb upstream sequences of *CBLL* genes *via* the PlantCARE database. In total, 51 *cis*-regulatory elements were detected

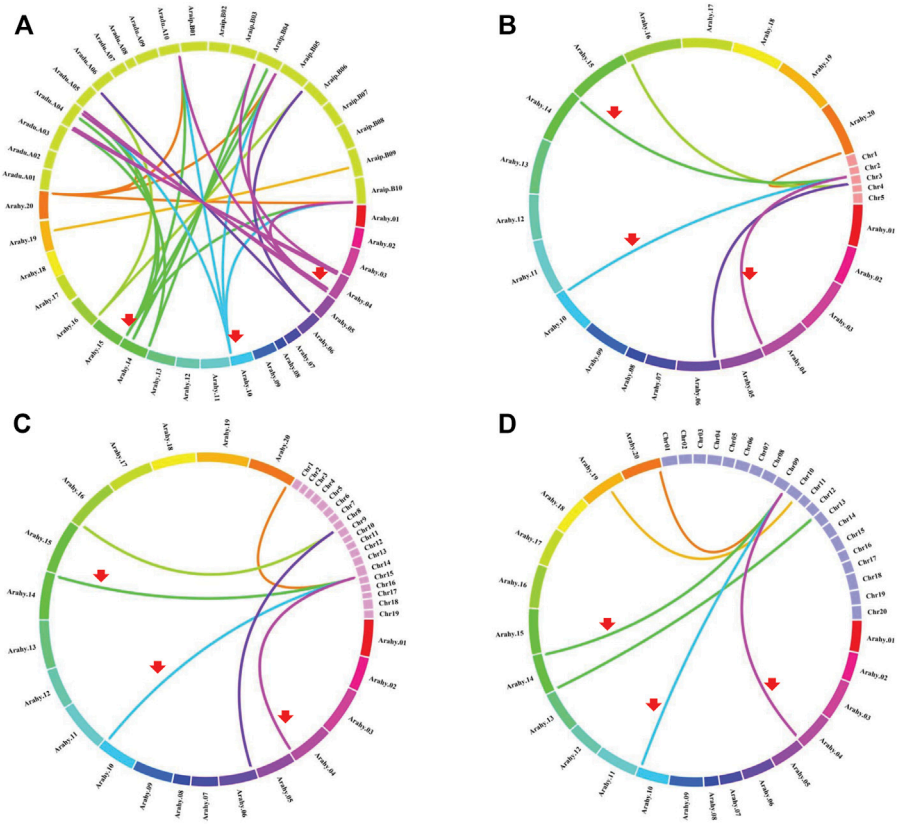


FIGURE 4 | Comparative physical mapping showed the degree of orthologous relationships of *AhCBLL* genes with **(A)** peanut, **(B)** Arabidopsis, **(C)** grape and **(D)** soybean. Red arrows indicated common collinear *CBLL* gene pairs.

TABLE 2 | The function of genes homologous to *AhCBLL* genes in Arabidopsis.

Peanut	Arabidopsis	Function	Reference
AdCBLL1/AiCBLL2/AiCBLL7/AhCBLL1/ AhCBLL7/AhCBLL11	*AtMGL* (*AT1G64660*)	Regulate methionine degradation, involved in the response to simultaneous biotic and abiotic stresses	Ricarda et al. (2005) Goyer et al. (2007) Joshi and Jander, (2009) Atkinson et al. (2013)
AdCBLL5/AdCBLL8/AiCBLL4/AhCBLL9	AtMTO1/AtCGS (AT3G01120)	Influences Met metabolism in seeds, related to light, sucrose and salt and osmotic stresses regulation	Hacham et al. (2013) Cohen et al. (2014) Cohen et al. (2017) Whitcomb et al. (2018)
AdCBLL6/AiCBLL5/AhCBLL5	AtCBL (AT3G57050)	Associated with MET biosynthesis, crucial for embryo patterning and the maintenance of root stem cell niche	Kim and Leustek, 2000 Levin et al. (2000) Liu et al. (2019)

(**Figures 5A–C**), 11 subclasses and four main categories were defined as plant growth, abiotic stress, phytohormone responsiveness, and light responsiveness element groups (**Figures 5D–F**). In the promoter region of the *AdCBLLs*, the largest subdivision was the light responsiveness group, containing 44.1% predicted *cis*-elements, phytohormone responsiveness elements ranked second (34.1%); abiotic stress response elements were 14.7%, and elements involved in plant growth

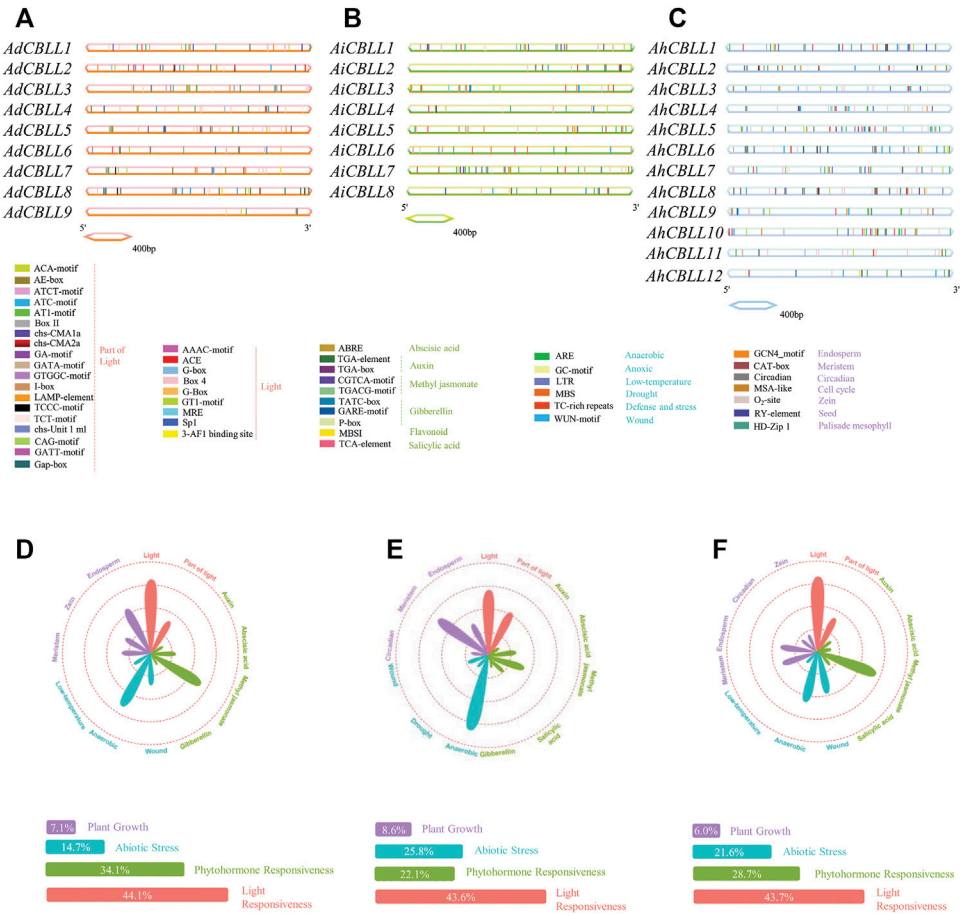


FIGURE 5 | Identification of *cis*-acting elements in all the peanut *CBLL* genes. **(A)** Distribution of *cis*-acting elements in 2 kb upstream of each *AdCBLLs*. **(B)** Distribution of *cis*-acting elements in 2 kb upstream of each *AiCBLLs*. **(C)** Distribution of *cis*-acting elements in 2 kb upstream of each *AhCBLLs*. The different colored boxes indicate distinct promoter elements. **(D)** Assessment *AdCBLL* different subclass and category proportions in a radar chart. **(E)** Assessment *AiCBLL* different subclass and category proportions in a radar chart. **(F)** Assessment *AhCBLL* different subclass and category proportions in a radar chart. The lengths of the petals are proportional to the number of elements in each subclass qualitatively. Purple, blue, green and red petals represent plant growth regulation, abiotic stress responses, phytohormone and light responsiveness, respectively.

accounted for 7.1%. *AdCBLL5* had the greatest number of elements with 33 in total (**Figure 5D**). For *AiCBLLs*, the percentage of light, phytohormone, abiotic stress, and plant growth responsiveness *cis*-elements was 43.6, 25.8, 22.1, and 8.6% (**Figure 5E**). *AiCBLL7* had the greatest number of elements at 35 in total. In *AhCBLLs*, the similar proportions were 43.7, 28.7, 21.6, and 6.0% (**Figure 5F**). In the light response category, Box 4tbox4 (light-responsive element) and GT1-motif (part of a module for light response) were the most dominant. Meanwhile, *cis*-acting elements responding to auxin, abscisic acid, gibberellin, flavonoids, methyl jasmonate and salicylic acid were involved in the phytohormone responsiveness group. The CGTCA-motif and TGACG-motif (methyl jasmonate response elements) were followed by ABRE (related to the abscisic acid response). In the abiotic stress response category, ARE (elements regarding oxygen-deficient induction) covered the largest portion, and both WUN-motif (wound-responsive element) and LTR (relating to low-temperature responsiveness) exited in the promoters of *AdCBLLs*, *AiCBLLs*, and *AhCBLLs*. GCN4 motif (elements related to endosperm expression) and CAT-box (referred to meristem expression) were the largest motifs in plant growth regulation. Intriguingly, all types of *cis*-regulatory elements were distributed widely throughout the promoter regions of *CBLL* genes.

Expression Profile of *AhCBLLs* in Different Tissues of Peanut

Previous studies showed the involvement of *CBLL* family genes during different development stages (Kim and Leustek., 2000; Levin et al., 2000; Maimann et al., 2000; Goto et al., 2002; Joshi and Jander, 2009; Atkinson et al., 2013; Hacham et al., 2013; Song et al., 2013; Cohen et al., 2014, 2017; Whitcomb et al., 2018; Liu et al., 2019). Thus, the holistic expression patterns of peanut *CBLLs* in different tissues are needed to provide more insight into their roles during plant growth and development. Tissue analyses

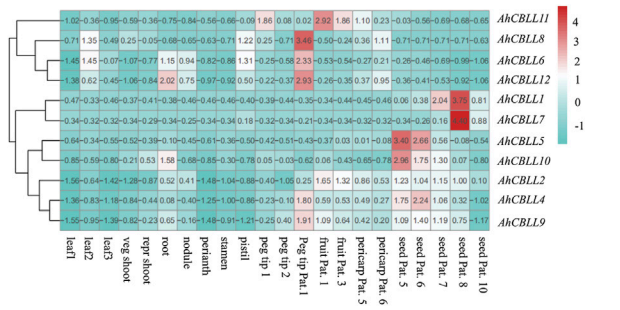


FIGURE 6 | The expression profiles of *AhCBLL* genes. The heat map of the *AhCBLL* gene expression levels was hierarchically clustered using the R package "pheatmap" with the data normalization method of Z-score standardization. The color scale bar ranging from blue to red represents low and high expression, respectively. Abbreviation of the 22 tissues used in the expression profiles of *AhCBLL* genes were as following: seedling leaf 10 days post emergence (leaf 1), main stem leaf (leaf 2), lateral stem leaf (leaf 3), vegetative shoot tip from main stem (veg shoot), reproductive shoot tip from first lateral (repr shoot), 10-day roots (root), 25-day nodules (nodule), perianth, stamen, pistil, aerial gynophore tip (peg tip 1), subterranean peg tip (peg tip 2), Pattee 1 stalk (peg tip Pat. 1), Patte 1 pod (fruit Pat. 1), Pattee 3 pod (fruit Pat.3), Pattee 5 pericarp (pericarp Pat.5), Pattee 6 pericarp (pericarp Pat.6), Pattee 5 seed (seed Pat. 5), Pattee 6 seed (seed Pat. 6), Pattee 7 seed (seed Pat.7), Pattee 8 seed (seed Pat. 8), Pattee 10 seed (seed Pat.10).

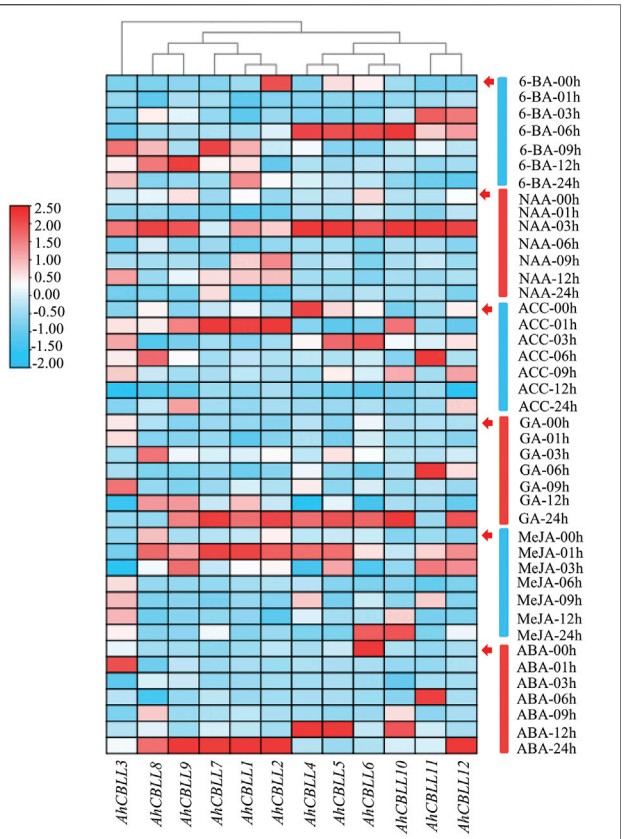


FIGURE 7 | Relative expression of *AhCBLLs* under 6-BA, NAA, ACC, GA, MeJA and ABA treatments. Expression characteristics of *AhCBLLs* in response to different phytohormone at six-time points (00, 01, 03, 06, 09 and 12 h) were normalized to 00 h treatment. The fold changes values were calculated by the $2^{-\Delta\Delta Ct}$ method and log2 and represented in color scale legend at the left of the heatmap: red indicated up-regulation and blue showed down-regulated expression.

of 12 *CBLL* genes showed distinct tissue-specific expression patterns across the 22 tissues (leaf, stem, root, flower, pod and seed) (**Figure 6**). *AhCBLL2, 4,* and *9* showed higher expression level in almost all the sink tissues; *AhCBLL5* and *10* were expressed mostly in the early seed developmental stages (seed pattee 5 and 6); *AhCBLL1* and *7* had strong expression during the relatively later seed developmental stages (seed pattee 7, 8, and 10), especially in seed pattee 8. *AhCBLL6* and *12* were highly expressed in root, nodule, and peg tip; Additionally, *AhCBLL6* was enriched in peg tip pat 1 and *AhCBLL11* was expressed highly in the peg and fruit tissues.

Expression Pattern of the *AhCBLLs* in Plant Hormone Response

To uncover the possible functions of *AhCBLLs* in response to hormone stress, we conducted qRT-PCR to analyze their relative expressions under 6-BA, NAA, ACC, GA, MeJA, and ABA treatments (**Figure 7**). In this study, a two-fold change (|log2| > 1) was considered as significantly different for gene expression under each treatment. Of all the 12 *AhCBLL*s, nine genes showed increased expression under all treatments, while the remaining three (*AhCBLL2*, *4* and *6*) had reduced expression at least one treatment (ABA, ACC, or 6-BA). For instance, *AhCBLL6*, *AhCBLL4*, and *AhCBLL2* were down-regulated under ABA, ACC, and 6-BA treatments, respectively. Interestingly, the responses of the *AhCBLLs* to these plant hormones were different; *AhCBLLs* responded to NAA, ACC, and MeJA in the early time series, later to 6-BA, and were to slow responded to GA and ABA (mostly at 24 h after treatment). The results indicated that these *AhCBLL* genes might regulate relevant hormone signaling pathways.

Expression Pattern of the *AhCBLLs* Under Abiotic Stresses

Plants suffer from a wide variety of environmental stressors under natural conditions. We investigated the expression of *AhCBLL* genes, responding to five abiotic stressors (**Figure 8**). Results showed that the accumulation of *AhCBLL1*, *AhCBLL2*, *AhCBLL3*, *AhCBLL8*, *AhCBLL9,* and *AhCBLL11* transcripts occurred at different time points after the submergence treatment, while the remaining six showed obviously down or up regulated curves. Expression levels of *AhCBLL3*, *AhCBLL8*, and *AhCBLL9* were elevated starting at 6 h under cold stress and the other nine members reached their highest expression at 24 h. Interestingly, all analyzed *AhCBLL* genes showed a positive response to heat stress. *AhCBLL 1*, 5, *7*, and *8* were up-regulated rapidly after 1 h of treatment, while *AhCBLL 3*, *4*, *6*, *10*, *11*, and *12* were strongly upregulated after 3 h of treatment. Moreover, three *AhCBLL* genes, *AhCBLL1*, *-2*, and *3* were inhibited under NaCl stress, and the remaining nine were up-regulated after NaCl treatment at different time points. Surprisingly, all *AhCBLLs* were induced

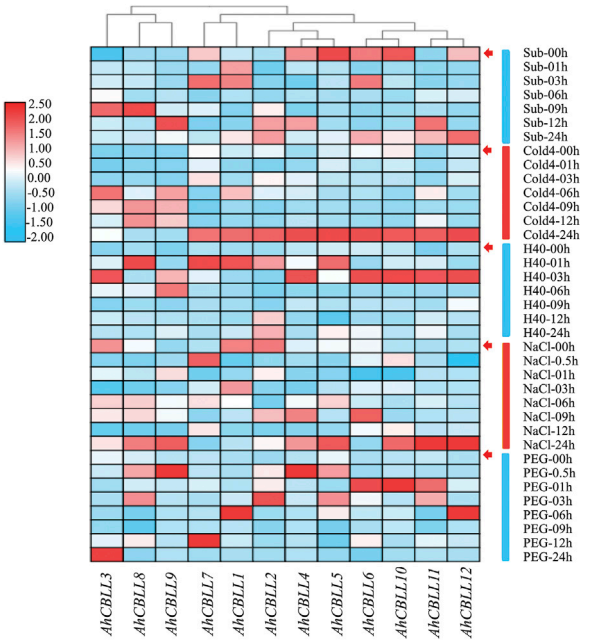


FIGURE 8 | Expression changes of *AhCBLLs* under different abiotic stresses. The abbreviation represented as following, Sub: submergence, Cold4: cold of 4°C, H40: the heat of 40°C, NaCl: 200 mM NaCl, and PEG, polyethylene glycol.

after PEG stress. In summary, *AhCBLL* genes might play important roles in various types of environmental stress regulation.

Association Analysis of Peanut *CBLLs* with 104 Traits of Peanut

To uncover the roles of *CBLL* genes in peanut development and stress response, we performed candidate gene association analysis using five SNPs (single nucleotide polymorphisms) in *AhCBLLs* from transcriptome data of 146 peanut lines, and 104 phenotypes related to peanut development and stress response were collected from the five environments. The results indicated that one polymorphic site [B09_90283818$^{(C/M/A)}$] was significantly associated with PL, PW, HPW, and HSW traits (**Figure 9A** and **Supplementary Tables S13–S15**). The site B09_90283818 mainly formed three haplotypes [B09_90283818$^{(C/M/A)}$) (**Figure 9B**) in the population and were located in the predicted exon region of *AiCBLL7* (**Figure 9C**). Results showed that PL, PW, HPW, and HSW in haplotype A were significantly higher than those in haplotype C (**Figure 9D**).

DISCUSSION

Duplication Contributed to *AhCBLLs* Expansion

Gene duplication contributes significantly to the proliferation of genes in plant species (Davidson et al., 2013; Hou et al., 2014), and leads to gene diversification or drives the evolution of genes. Based on synteny analyses, nine of the 12 *AhCBLLs* had syntenic relationships, all of them were inherited from A. duranensis and A. ipaensis genomes (**Figures 2**, **3** and **Supplementary Table S2**); the main expansion mechanism of *AhCBLLs* was whole-genome duplication (allopolyploidization), while none-segmental and tandem duplication events occurred in the tetraploid stage.

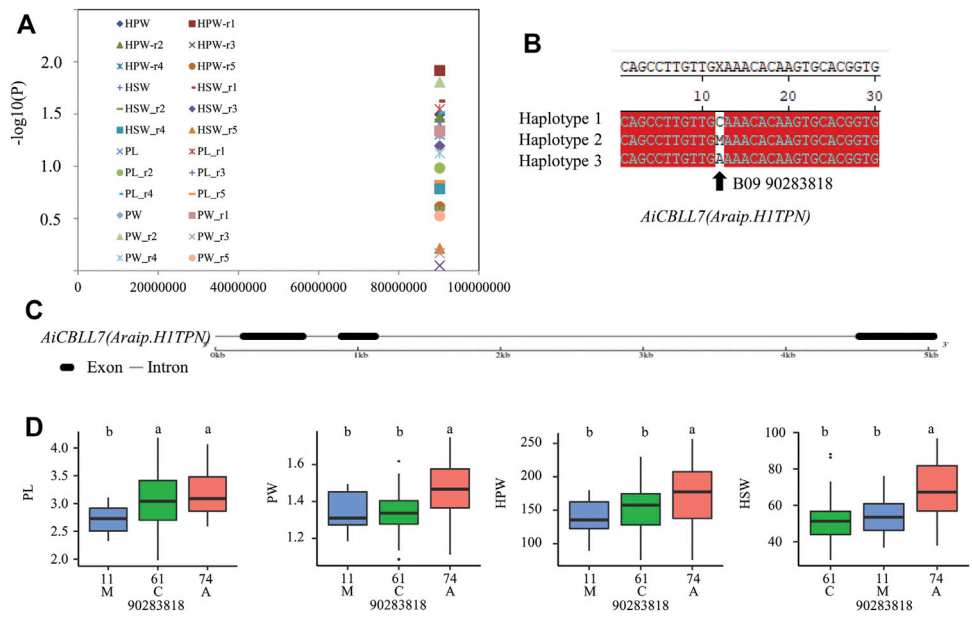


FIGURE 9 | Association mapping results and the phenotypes of the polymorphic sites of the peanut *CBLLs* associated with HPW/HSW/PL/PW variation. **(A)** Association results between HPW/HSW/PL/PW and the polymorphisms in peanut *CBLLs*. **(B)** Sequences of site significantly associated with HPW/HSW/PL/PW variation. **(C)** The gene structures of *AiCBLL7*. **(D)** Phenotypic comparison of haplotypes of the associated site with HPW/HSW/PL/PW in five environments of the population.

Duplicated genes often evolve to lose their original functions and/or obtain new functions to enhance the adaptability of plants (Dias et al., 2003). Previous research demonstrates that a diversified expression pattern might be a significant reason for retaining duplicated genes in the genome (Gu et al., 2002). The duplicated genes lead to gene-dose effects or result in functional diversity. We found that all the *AhCBLL* gene pairs from allopolyploid species displayed a similar tissues expression pattern. For example, gene pair *AhCBLL1* (*Arahy.04*)-*AhCBLL7* (*Arahy.14*) was mainly expressed in the later seed development stage. Pat.8; *AhCBLL4* (*Arahy.04*)-*AhCBLL9* (*Arahy.14*) showed a higher expression level in peg. tip. Pat.1, *AhCBLL5* (Arahy.06)–*AhCBLL10* (Arahy.16) were enriched in the earlier seed development stage. Pat.5 and 6, and *Ah CBLL6* (Arahy.10)-*AhCBLL12* (Arahy.20) were highly expressed in root, nodule, and peg tip. Pat.1 (**Figure 6**). Most of the four allopolyploidy duplicated gene pairs demonstrated similar expression patterns to hormone treatments and abiotic stresses that might result in the way of gene-dose-effect; however, the expression of *AhCBLL4* was inhibited under submergence and ACC treatment while *AhCBLL9* was active ated; *AhCBLL6* was down-regulated under ABA treatment while *AhCBLL12* was up-regulated (**Figures 7**, **8**); this opposite expression profile of the allopolyploidy duplicated gene pairs implied functional diversity. These results suggested that gene duplication of the *CBLL* family in peanut provided intricate regulation of signal transduction.

Subfunctionalization of the *AhCBLL* Genes in Peanut

The phylogenetic tree divided all the 29 peanut CBLL proteins into four clades (**Figure 1A**). Eight, six, one, four, and 10 CBLLs pertained to clade I (CGS clade), clade II (MGL clade), clade III (new clade), clade IV (CBL clade), clade V (new clade), respectively. The members in the CGS clade, MGL clade and CBL clade might have similar functions to the corresponding Arabidopsis genes; however, the function of the members in clade III and clade V was uncertain. The annotation of the one gene of clade III and most genes of clade V were CBLs; however, the gene structure and motif arrangement were quite different from those in the CBL clade, indicating divergence in gene function.

The prediction of *cis*-acting elements can provide important clues for the study of gene expression regulation (Zhu et al., 2017). In this study, analysis of the 2-kb upstream sequences of the initial codon of *AhCBLL* genes showed that the *AhCBLLs* contained multiple stress and hormone response elements, but the types and numbers were different. The percentage of *cis*-acting elements to phytohormone in A. duranensis was approximately 12%, which was more than that in A. ipaensis, indicating that the transcriptomic regulation of *AdCBLLs* might be more complex than in *AiCBLLs*. Most *AhCBLL* genes not only had *cis*-acting elements respond to abiotic adversity, but also had elements that responded to hormonal signals such as gibberellin, abscisic acid, salicylic acid and methyl jasmonate (**Figure 5**). For example, TCA-element, LTR, circadian and TC-rich repeats were exited in the promotor region of *AhCBLL1,* implying that *AhCBLL1* might respond to salicylic acid signals and participate in the regulation of cold stress, circadian rhythm and plant defense response. The positive response of most *AhCBLLs* genes to hormones was the mainstream regardless of the *AhCBL2* and *AhCBL6* were down-regulated in 6-BA and ABA, respectively (**Figure 7**), implying their relevant roles in these hormone signal pathways. Surprising finding was that all analyzed *AhCBLL* genes showed positive response to heat and PEG stresses (**Figure 8**), suggesting that the *AhCBLL* may be important for peanut resistance to heat stress (**Figure 8**). Similarly, *AhCBLL1*, *AhCBLL2*, *AhCBLL3*, *AhCBLL8*, *AhCBLL9,* and *AhCBLL11* were found to up-regulated by submergence (**Figure 8**), indicating the physiological function of these genes in peanut water logging stress tolerance mechanisms. Additinally, the expression of the nine *AhCBLL* genes except *AhCBLL1*, *2*, and *3* were obviously enhanced after NaCl treatment at different time points (**Figure 8**). Considering the common positive response of *AhCBLL8*, *AhCBLL9,* and *AhCBLL11* under heat, drought, submergence, and salt stresses, overexpression of these three genes in peanut may be an effective method to improve the peanut comprehensive abiotic stress resistance. Many important genes were selectively expressed in specific tissues during various physiological and developmental processes (Wan et al., 2014). Tissue expression of the *AhCBLLs* showed multiple tissue expression patterns suggesting subfunctionalization of this family.

Among the 13 *CBLL* orthologous pairs between peanut and Arabidopsis (**Table 2**), the functions of the corresponding ortholog genes in Arabidopsis have been determined; they functioned in influencing flowering, cell elongation, pollen tube growth, and played an important role in seed germination (**Table 2**). *AtMOT1/AtCGS* have been found to affect the physiological and behavioral processes of seeds, related to the biosynthesis of ethylene and polyamines (Goto et al., 2002; Hacham et al., 2013; Cohen et al., 2014, 2017; Whitcomb et al., 2018). Furthermore, *AtMGL*, regulates Met degradation, involved in the response to simultaneous biotic and abiotic stresses (Ricarda et al., 2005; Goyer et al., 2007; Joshi and Jander, 2009; Atkinson et al., 2013). Therefore, these *AhCBLL* orthologous genes may also play multiple roles in peanut development, and plant hormone synthesis or response.

AhCBLL Gene Plays an Important Role in Peanut Pod and Seed Development

The single-nucleotide polymorphic sites in *AiCBLL7* (corresponding to *AhCBLL11*), were significantly associated with PL, PW, HPW and HSW variation. The polymorphic site in *AiCBLL7*, [B09_90283818(C/M/A)], located in the predicted exon region of the gene, B09_90283818(C/M/A) led a 268E to 268 K amino acid transition in the peanut population. These results indicated that B09_90283818(C/M/A) sequence polymorphisms might be the actual functional sites. Further, *AhCBLL11* was mainly expressed in the peg and fruit, especially in early fruit development stages, which provided additional evidence for its function in peanut pod development. Most *AhCBLLs* exhibited tissue-specific expression patterns, almost all the *AhCBLLs* were expressed in

higher levels in the peg, fruit, or seed (**Figure 6**). Further investigation was needed to confirm the roles of *AiCBLL7* (*AhCBLL11*) in the pod and seed development of peanut.

CONCLUSION

In summary, this genome-wide identification, characterization and expression analysis of peanut *CBLL* genes provides valuable information for understanding the evolution and molecular functions of the peanut *CBLL* gene family, and highlights potential *CBLL* genes involved in peanut pod (seed) development and abiotic stress responses. The results of this study provide a foundation for further research regarding the function of the peanut *CBLL* gene family.

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

WR and LW carried out all the experiments and data analyses. WR and ZZ prepared the figures and tables. SW performed the qRT-PCR experiments. LW, JF, and JZ made modifications to the article. All authors read and approved the final manuscript.

FUNDING

This research was funded by grants from the National Key Research and Development Program of China (No. 2018YFD1000900) and the Jiangxi Agriculture Research System (No. JCARS-18). Funders did not participate in the design of the study, analysis of the results and writing of the manuscript, but provided financial support for the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.821163/full#supplementary-material

REFERENCES

Amir, R. (2010). Current Understanding of the Factors Regulating Methionine Content in Vegetative Tissues of Higher Plants. *Amino Acids* 39, 917–931. doi:10.1007/s00726-010-0482-x

Atkinson, N. J., Lilley, C. J., and Urwin, P. E. (2013). Identification of Genes Involved in the Response of Arabidopsis to Simultaneous Biotic and Abiotic Stresses. *Plant Physiol.* 162 (4), 2028–2041. doi:10.1104/pp.113.222372

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

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., et al. (2019). The Genome Sequence of Segmental Allotetraploid Peanut *Arachis hypogaea*. *Nat. Genet.* 51 (5), 877–884. doi:10.1038/s41588-019-0405-z

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The Roles of Segmental and Tandem Gene Duplication in the Evolution of Large Gene Families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4, 10–21. doi:10.1186/1471-2229-4-10

Chen, Z., Zou, X., Song, L., Zou, X., Zhang, J., Chen, L., et al. (2010). Present Situation and Development Countermeasures of Peanut Production in Jiangxi Province. *Bull. Agric. Sci. Technol.* 06, 18–21.

Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A Developmental Transcriptome Map for Allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* 7, 1446. doi:10.3389/fpls.2016.01446

Cohen, H., Hacham, Y., Panizel, I., Rogachev, I., Aharoni, A., and Amir, R. (2017). Repression of CYSTATHIONINE γ-SYNTHASE in Seeds Recruits the S-Methylmethionine Cycle. *Plant Physiol.* 174 (3), 1322–1333. doi:10.1104/pp.17.00579

Cohen, H., Israeli, H., Matityahu, I., and Amir, R. (2014). Seed-Specific Expression of a Feedback-Insensitive Form of CYSTATHIONINE-γ-SYNTHASE in Arabidopsis Stimulates Metabolic and Transcriptomic Responses Associated with Desiccation Stress. *Plant Physiol.* 166 (3), 1575–1592. doi:10.1104/pp.114.246058

Davidson, C. J., Guthrie, E. E., and Lipsick, J. S. (2013). Duplication and Maintenance of the Myb Genes of Vertebrate Animals. *Biol. Open* 2 (2), 101–110. doi:10.1242/bio.20123152

Deng, X., An, B., Zhong, H., Yang, J., Kong, W., and Li, Y. (2019). A Novel Insight into Functional Divergence of the MST Gene Family in rice Based on Comprehensive Expression Patterns. *Genes* 10 (3), 239. doi:10.3390/genes10030239

Dias, A. P., Braun, E. L., McMullen, M. D., and Grotewold, E. (2003). Recently Duplicated maize R2R3 Myb Genes Provide Evidence for Distinct Mechanisms of Evolutionary Divergence after Duplication. *Plant Physiol.* 131 (2), 610–620. doi:10.1104/pp.012047

Galili, G., Amir, R., and Fernie, A. R. (2016). The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. *Annu. Rev. Plant Biol.* 67, 153–178. doi:10.1146/annurev-arplant-043015-112213

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. *2-D Proteome Anal. Protoc.*pp, 571–607. doi:10.1385/1-59259-890-0:571

Goto, D. B., Ogi, M., Kijima, F., Kumagai, T., Werven, F. v., Onouchi, H., et al. (2002). A Single-Nucleotide Mutation in a Gene Encoding S-Adenosylmethionine Synthetase Is Associated with Methionine Over-accumulation Phenotype in *Arabidopsis thaliana*. *Genes Genet. Syst.* 77 (2), 89–95. doi:10.1266/ggs.77.89

Goyer, A., Collakova, E., Shachar-Hill, Y., and Hanson, A. D. (2007). Functional Characterization of a Methionine -Lyase in Arabidopsis and its Implication in an Alternative to the Reverse Trans-sulfuration Pathway. *Plant Cel Physiol.* 48 (2), 232–242. doi:10.1093/pcp/pcl055

Gu, Z., Cavalcanti, A., Chen, F.-C., Bouman, P., and Li, W.-H. (2002). Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast. *Mol. Biol. Evol.* 19, 256–262. doi:10.1093/oxfordjournals.molbev.a004079

Guo, M., Liu, J.-H., Ma, X., Zhai, Y.-F., Gong, Z.-H., and Lu, M.-H. (2016). Genome-wide Analysis of the Hsp70 Family Genes in Pepper (*Capsicum Annuum* L.) and Functional Identification of *CaHsp70-2* Involvement in Heat Stress. *Plant Sci.* 252, 246–256. doi:10.1016/j.plantsci.2016.07.001

Hacham, Y., Matityahu, I., and Amir, R. (2013). Light and Sucrose Up-Regulate the Expression Level of Arabidopsis Cystathionine γ-synthase, the Key Enzyme of Methionine Biosynthesis Pathway. *Amino Acids* 45 (5), 1179–1190. doi:10.1007/s00726-013-1576-z

Holowach, L. P., Madison, J. T., and Thompson, J. F. (1986). Studies on the Mechanism of Regulation of the mRNA Level for a Soybean Storage Protein Subunit by Exogenous L-Methionine. *Plant Physiol.* 80, 561–567. doi:10.1104/pp.80.2.561

Hou, X.-J., Li, S.-B., Liu, S.-R., Hu, C.-G., and Zhang, J.-Z. (2014). Genome-wide Classification and Evolutionary and Expression Analyses of Citrus MYB Transcription Factor Families in Sweet orange. *PLoS One* 9 (911), e112375. doi:10.1371/journal.pone.0112375

Jain, M., Nijhawan, A., Tyagi, A. K., and Khurana, J. P. (2006). Validation of Housekeeping Genes as Internal Control for Studying Gene Expression in rice by Quantitative Real-Time PCR. *Biochem. Biophysical Res. Commun.* 345, 646–651. doi:10.1016/j.bbrc.2006.04.140

Joshi, V., and Jander, G. (2009). Arabidopsis Methionine γ-Lyase Is Regulated According to Isoleucine Biosynthesis Needs but Plays a Subordinate Role to Threonine Deaminase. *Plant Physiol.* 151 (1), 367–378. doi:10.1104/pp.109.138651

Jost, R., Altschmied, L., Bloem, E., Bogs, J., Gershenzon, J., Hähnel, U., et al. (2005). Expression Profiling of Metabolic Genes in Response to Methyl Jasmonate Reveals Regulation of Genes of Primary and Secondary Sulfur-Related Pathways in *Arabidopsis thaliana*. *Photosynth Res.* 86, 491–508. doi:10.1007/s11120-005-7386-8

Kery, V., Bukovska, G., and Kraus, J. P. (1994). Transsulfuration Depends on Heme in Addition to Pyridoxal 5'-phosphate. Cystathionine Beta-Synthase Is a Heme Protein. *J. Biol. Chem.* 269 (41), 25283–25288. doi:10.1016/s0021-9258(18)47244-4

Kim, J., and Leustek, T. (2000). Repression of Cystathionine γ-synthase in *Arabidopsis thaliana* Produces Partial Methionine Auxotrophy and Developmental Abnormalities. *Plant Sci.* 151, 9–18. doi:10.1016/s0168-9452(99)00188-0

Lescot, M. (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

Levin, J. Z., Framond, A. J. D., Tuttle, A., Bauer, M. W., and Heifetz, P. B. (2000). Methods of Double-Stranded RNA-Mediated Gene Inactivation in Arabidopsis and Their Use to Define an Essential Gene in Methionine Biosynthesis. *Plant Mol. Biol.* 44, 759–775. doi:10.1023/a:1026584607941

Liao, B., Yin, Y., and Ma, N. (2018). Review and prospect of Development of Oil Crop Industry in China. *J. Agric.* 8 (1), 115–120.

Liu, G., Yang, W., Zhang, X., Peng, T., Zou, Y., Zhang, T., et al. (2019). Cystathionine Beta-lyase Is Crucial for Embryo Patterning and the Maintenance of Root Stem Cell Niche in Arabidopsis. *Plant J.* 99 (3), 536–555. doi:10.1111/tpj.14343

Maimann, S., Wagner, C., Kreft, O., Zeh, M., Willmitzer, L., Hofgen, R., et al. (2000). Transgenic Potato Plants Reveal the Indispensable Role of Cystathionine Beta-Lyase in Plant Growth and Development. *Plant J.* 23, 747–758. doi:10.1046/j.1365-313x.2000.00842.x

Messerschmidt, A., Worbs, M., Steegborn, C., Wahl, M. C., Huber, R., Laber, B., et al. (2003). Determinants of Enzymatic Specificity in the Cys-Met-Metabolism PLP-dependent Enzyme Family: Crystal Structure of Cystathionine γ-Lyase from Yeast and Intrafamiliar Structure Comparison. *Biol. Chem.* 384 (3), 373–386. doi:10.1515/BC.2003.043

Misra, V. A., Wafula, E. K., Wang, Y., dePamphilis, C. W., and Timko, M. P. (2019). Genome-wide Identification of MST, SUT and SWEET Family Sugar Transporters in Root Parasitic Angiosperms and Analysis of Their Expression during Host Parasitism. *BMC Plant Biol.* 19 (1), 196. doi:10.1186/s12870-019-1786-y

Ravanel, S., Gakiere, B., Job, D., and Douce, R. (1998). The Specific Features of Methionine Biosynthesis and Metabolism in Plants. *Proc. Natl. Acad. Sci.* 95, 7805–7812. doi:10.1073/pnas.95.13.7805

Song, S., Hou, W., Godo, I., Wu, C., Yu, Y., Matityahu, I., et al. (2013). Soybean Seeds Expressing Feedback-Insensitive Cystathionine γ-synthase Exhibit a Higher Content of Methionine. *J. Exp. Bot.* 64 (7), 1917–1926. doi:10.1093/jxb/ert053

Song, Y., Wang, L., and Xiong, L. (2009). Comprehensive Expression Profiling Analysis of *OsIAA* Gene Family in Developmental Processes and in Response to Phytohormone and Stress Treatments. *Planta* 229, 577–591. doi:10.1007/s00425-008-0853-7

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust Conversion of Protein Sequence Alignments into the Corresponding Codon Alignments. *Nucleic Acids Res.* 34 (Web Server), W609–W612. doi:10.1093/nar/gkl315

Wan, L., Wu, Y., Huang, J., Dai, X., Lei, Y., Yan, L., et al. (2014). Identification of ERF Genes in Peanuts and Functional Analysis of *AhERF008* and *AhERF019* in Abiotic Stress Response. *Funct. Integr. Genomics* 14 (3), 467–477. doi:10.1007/s10142-014-0381-4

Wang, Z., Wang, Y., Yang, J., Hu, K., An, B., Deng, X., et al. (2016). Reliable Selection and Holistic Stability Evaluation of Reference Genes for Rice under 22 Different Experimental Conditions. *Appl. Biochem. Biotechnol.* 179, 753–775. doi:10.1007/s12010-016-2029-4

Whitcomb, S. J., Nguyen, H. C., Brückner, F., Hesse, H., and Hoefgen, R. (2018). CYSTATHIONINE GAMMA-SYNTHASE Activity in rice Is Developmentally Regulated and Strongly Correlated with Sulfate. *Plant Sci.* 270, 234–244. doi:10.1016/j.plantsci.2018.02.016

Yang, J., Zhang, Y., Yang, P., and Ji, M. (2014). Analysis of Peanut Industry Development in Shandong Province. *Jiangsu Agric. Sci.* 42 (9), 433–436.

Yang, X., Chen, Z., and Zhao, M. (2001). Isolation and Some Properties of Peanut Protein. *J. Chin. Cereals Oils Assoc.* 16 (05), 25–28. doi:10.3901/jme.2001.03.025

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24 (8), 1586–1591. doi:10.1093/molbev/msm088

Zhu, Q., Zhao, Y., Lv, H., Luan, F., and Gao, P. (2017). Identification and Characterization of CBL Family Genes in Watermelon. *North. Hortic.* (15), 18–24.

Sarropoulou, V., Dimassi-Theriou, K., and Therios, L. (2013). L-Methionine Influences *in vitro* Root Regeneration, Total Chlorophyll, Total Carbohydrate and Endogenous Proline Content in the Sweet Cherry Rootstock MxM 14 (*Prunus avium* × *Prunus mahaleb*). *Sci. Horti.* 161, 88–94. doi:10.1016/j.scienta.2013.06.012

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 Ren, Zeng, Wang, Zhang, Fang and Wan. 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.2022.836030

Genome-Wide Identification and Expression Analysis of the Thioredoxin (*Trx*) Gene Family Reveals Its Role in Leaf Rust Resistance in Wheat (*Triticum aestivum* L.)

Ramesh Bhurta[1†], *Deepak T. Hurali*[1†], *Sandhya Tyagi*[2], *Lekshmy Sathee*[2*], *Sandeep Adavi B*[2], *Dalveer Singh*[2], *Niharika Mallick*[1], *Viswanathan Chinnusamy*[2], *Vinod*[1] *and Shailendra K. Jha*[1*]

[1]Division of Genetics, ICAR-IARI, New Delhi, India, [2]Division of Plant Physiology, ICAR-IARI, New Delhi, India

OPEN ACCESS

Edited by:
Himabindu Kudapa,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

Reviewed by:
Raman Dhariwal,
Lethbridge Research Center, Canada
Gautam Saripalli,
University of Maryland, United States

***Correspondence:**
Lekshmy Sathee
lekshmyrnair@gmail.com
Shailendra K. Jha
jhashail78@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 December 2021
Accepted: 28 February 2022
Published: 25 March 2022

Citation:
Bhurta R, Hurali DT, Tyagi S, Sathee L, Adavi B S, Singh D, Mallick N, Chinnusamy V, Vinod and Jha SK (2022) Genome-Wide Identification and Expression Analysis of the Thioredoxin (Trx) Gene Family Reveals Its Role in Leaf Rust Resistance in Wheat (Triticum aestivum L.). Front. Genet. 13:836030. doi: 10.3389/fgene.2022.836030

Bread wheat (*Triticum aestivum* L.; *Ta*) is the staple cereal crop for the majority of the world's population. Leaf rust disease caused by the obligate fungal pathogen, *Puccinia triticina* L., is a biotrophic pathogen causing significant economic yield damage. The alteration in the redox homeostasis of the cell caused by various kinds of reactive oxygen species (ROS) and reactive nitrogen species (RNS) in response to pathogenic infections is controlled by redox regulators. Thioredoxin (Trx) is one of the redox regulators with low molecular weight and is thermostable. Through a genome-wide approach, forty-two (42) wheat *Trx* genes (*TaTrx*) were identified across the wheat chromosome groups A, B, and D genomes containing 12, 16, and 14 *Trx* genes, respectively. Based on *in silico* expression analysis, 15 *TaTrx* genes were selected and utilized for further experimentation. These 15 genes were clustered into six groups by phylogenetic analysis. MicroRNA (miRNA) target analysis revealed eight different miRNA-targeted *TaTrx* genes. Protein–protein interaction (PPI) analysis showed *TaTrx* proteins interact with thioredoxin reductase, peroxiredoxin, and uncharacterized proteins. Expression profiles resulting from quantitative real-time PCR (qRT-PCR) revealed four *TaTrx* genes (*TaTrx11-5A*, *TaTrx13-5B*, *TaTrx14-5D*, and *TaTrx15-3B*) were significantly induced in response to leaf rust infection. Localization of ROS and its content estimation and an assay of antioxidant enzymes and expression analysis suggested that *Trx* have been involved in ROS homeostasis at span 24HAI-72HAI during the leaf rust resistance.

Keywords: wheat, leaf rust, thioredoxin, reactive oxygen species (ROS), genome-wide analysis

INTRODUCTION

Wheat (*Triticum aestivum* L.) leaf rust disease caused by *Puccinia triticina* L. is the most prevalent disease among the three rust-related diseases. About 50% yield reduction has been reported when conditions are favorable for leaf rust infection and depending on the plant's developmental stage at disease (Huerta-Espino et al., 2011; Dakouri et al., 2013). Though the damage caused by rust can be minimized through chemicals, host resistance is the most economical and sustainable way to cope with this disease. So far, eighty (80) leaf rust

resistance genes have been identified and designated, and different *Lr* genes (APR and SR) are being used in different wheat breeding programs to develop rust-resistant varieties (Tomar et al., 2014; Kumar et al., 2021).

A strategy for manipulating resistance can be developed by understanding the pathways involved in defense signaling. Plants develop and execute various multifaceted sensory mechanisms to recognize a particular pathogen attack and work against it to maximize plant survival (Lamers et al., 2020). These molecular mechanisms are well elucidated and explained to a great depth (Wang et al., 2019). The impact of thioredoxin (Trx) and reactive oxygen species (ROS) as a critical factor in the plant system against biotic stress responses is one area to understand and exploit its role in resistance. Regulations of cysteine modification depend on Trx activity. Thiol-based redox modification of Trxs alters the activity of target proteins involved in the interaction network (Rouhier et al., 2015). Reactive oxygen species and oxidative stress produced by metabolism, respiration, and defense responses are sensed by thioredoxins, and the downstream target signaling proteins are modulated to regain redox balance (He et al., 2017).

Thioredoxins (Trxs) are heat-stable multifunctional ubiquitous proteins that act as oxidoreductase and are involved in various metabolic processes such as photosynthesis and response to abiotic and biotic stresses, depending on their inherent biochemical properties (Holmgren, 1968; Wong, 2002). Thioredoxins are also confirmed to be involved in the orchestration of Systemic Acquired Resistance (SAR) in plants by its redox activity that converts the oligomeric form of the nonexpressor of pathogenesis-related gene1 (NPR1), a master regulator of SAR, into a monomeric form in *Arabidopsis* (Fu et al., 2012; Moreau et al., 2012). Trxs possess a characteristic conserved active site consisting of WCGPC residues that catalyze thiodisulfide bridges (Collet et al., 2003). Trx proteins exhibit numerous variants, and in plants, they are classified as Trxf, Trxm, Trxh, Trxo, Trxx, and Trxy isoforms (Gelhaye et al., 2004). Apart from their prominent role in the redox-based modification of proteins that are involved in resistance response, thioredoxins are also associated with various biochemical processes, viz., enzyme regulation (Holmgren, 1989), modulation of transcription factor (Schenk et al., 1994), as the donor of hydrogen (Holmgren, 1989), protection against oxidative stress (Chae et al., 1994), and in-plant resistance (Mata-Pérez and Spoel, 2019).

The role of ROS/RNS is well known as an essential mediator of redox signaling in defense pathways (Apel and Hirt, 2004; Baxter et al., 2014; Mittler, 2002; Noctor et al., 2018; Mittler et al., 2004; Torres et al., 2006). The activation of the plant resistance responses against a pathogen with the rapid production of an enormous amount of ROS and reactive nitrogen species (RNS) that results in alteration of redox homeostasis of the cell (Mata-Pérez and Spoel, 2019). Plants induce a defense response against pathogen attack via inducing pathogenesis-related (PR) genes and localized cell death at the infection site, collectively known as the hypersensitive response (HR). Further ROS is generated after pathogen recognition in various subcellular compartments responsible for the orchestration of HR (Zubriggen et al., 2010). Wang et al. (2020) demonstrated that ROS (superoxide radicals and H_2O_2) is generated during the interaction of leaf rust, *Puccinia triticina* (Pt) and wheat. When a pathogen attacks the plant, the plant–microbe interaction is sensed by membrane receptors, and an oxidative (ROS) burst occurs in the apoplast by activating NADPH oxidases (NOX/RHOH) (Rahikainen et al., 2016). Furthermore, ROS signaling is also required to induce immunity-related genes (de Torres Zabala et al., 2015).

Cellular ROS homeostasis involves several enzymatic and non-enzymatic antioxidant systems in plants (Lu and Holmgren, 2014; Considine and Foyer, 2020). Among these antioxidant systems, Trx modulates ROS scavenging and actively regulates cellular redox homeostasis. It was reported that functional loss of Trx leads to altered levels of ROS (Vieira Dos Santos and Rey, 2006; Schmidt et al., 2016). For instance, thioredoxin genes *Trxh8* and *Trxh5* are highly induced in response to abiotic and biotic stresses in rice (Tsukamoto et al., 2005). A novel cis-element regulates three antioxidant genes *Trxh*, *Grx*, and *SOD* in rice (Tsukamoto et al., 2005). These studies advocate the coordinated participation of thioredoxin and other antioxidants in cellular defense against high levels of ROS. In the recent past, significant advancements have been made in understanding the pathway of defense responses against rust pathogen, and new dimensions are expected to be added in the future. However, to our best knowledge, change in the expression profile of *Trx* genes against leaf rust resistance in wheat and its relation with *ROS* have not been studied yet. Therefore, based on the importance of Trx proteins in biotic stress response and ROS balance during plant–pathogen interaction, we conducted 1) genome-wide identification and *in silico* analysis of *Trx* family genes in wheat and 2) expression profiling of selected *TaTrx* family genes in response to leaf rust infection in wheat.

MATERIALS AND METHODS

In Silico Chromosome Localization of Wheat *Trx* Genes

The amino acid sequences of Trx proteins from *Arabidopsis (AtTrxh3; At5G42980)*) and rice (*OsTrxh5*; *Os07T0190800)* were extracted from TAIR (https://www.arabidopsis.org/index.jsp), and The Rice Annotation Project database rap-db (https://rapdb.dna.affrc.go.jp). The hidden Markov model (HMM) profile of the Trx domain (PF00085.22) and (PF00085.22) was used as a query to identify the wheat Trx domain in the wheat genome database using the software HMMER (http://pfam.sanger.ac.uk. A BLAST algorithm was also used to identify the putative *Trx* genes present across the wheat genome. The amino acid sequences of *Trx* candidate genes from *Arabidopsis* and rice were used *OsTrxh5*, *Os07T0190800* as a query against the fully annotated genome sequence of wheat accessible at Ensembl

Plants release 47 (https://plants.ensembl.org/index.html). All non-redundant sequences with E-values < 1.0^{E-05} were identified and selected. Protein, CDS, cDNA, and genomic sequences of all the chosen IDs were downloaded for further analysis. All the identified sequences were further analyzed to search conserved domains using secondary databases, including InterPro (https://www.ebi.ac.uk/interpro) and PROSITE (https://prosite.expasy.org). To determine the conserved or diverse *Trx* genes among the different wheat species, a BlastP search was also conducted using the genome databases. This analysis used the sequence release 52 of three wheat species *T. Urartu* (A genome), *T. turgidum* (AB genome), and *A. tauschii* (D Genome) available on Ensembl Plants. All the identified wheat *Trx* genes (*TaTrx*) IWGSC IDs screened against information available in public repositories, including IWGSC-URGI (https://wheat-urgi.versailles.inra.fr/) and EnsemblPlants (Bolser et al., 2016) and were *in silico* mapped on seven homeologous chromosome groups accordingly.

Physiochemical Properties and Subcellular Localization of Trx Protein

The amino acid sequences of all the identified Trx proteins retrieved from Ensembl Plants were analyzed using the ProtParam tool (Expasy website: https://web.expasy.org/protparam/) (Gasteiger et al., 2005) to obtain different physiochemical properties like the grand average of hydropathicity (GRAVY), isoelectric point (IP), aliphatic index (Ai), molecular weight (Mw), and instability index (Ii). Further, BUSCA (http://busca.biocomp.unibo.it/) (Savojardo et al., 2018), a web server, was used to predict the subcellular location of candidate Trx proteins.

Gene Structure, Conserved Motif Identification, and Evolutionary Analysis

Both cDNA and genomic sequences were analyzed in GSDS v2.0 (Gene Structure Display Server) (Hu et al., 2015). The regulatory motif variation in identified *TaTrx* genes was analyzed using the amino acid sequence of respective *TaTrx* genes in MEME 5.0.5 (Bailey et al., 2009) with selection criteria set at 15 AA, minimum length of 6, and a maximum length of 50 amino acids (Bailey et al., 1994). Phylogenetic analysis was performed using the identified amino acid sequences of *TaTrxs* and the *Trxs* earlier reported in *Arabidopsis* and rice. For this purpose, protein sequences of *Arabidopsis* (22 Trxs) and rice (28 Trxs) were downloaded from TAIR (https://www.arabidopsis.org/) and rap-db (https://rapdb.dna.affrc.go.jp/), respectively. All the collected amino acid sequences of identified *Trx* genes were aligned with ClustalW (http://ebi.ac.uk/Tools/msa/clustalW2). The aligned output file was used by the software MEGA7.0 (Kumar et al., 2016) to establish evolutionary relationship among extracted all *Trx* proteins. MEGA 7.0 constructed a phylogenetic tree using the maximum likelihood method (Kumar et al., 2016) with the substitution model, uniform rates, and pair-wise deletion with bootstrap values for 1,000 iterations calculated and expressed as percentages (Felsenstein, 1985).

Putative Cis-Regulatory Element Identification in the Promoter Region of the *TaTrx* Genes

The 1,000 bp *TaTrx* gene sequences upstream to the start codon were retrieved from Ensembl Plants. *TaTrx* sequences were analyzed in the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to scan cis-regulatory elements (Lescot et al., 2002). These 1,000 bp upstream sequences were investigated to identify the cis-regulatory elements to hormone responsiveness (involved in stress) and defense and stress responsiveness.

Potential miRNA Target for *TaTrx* Genes

Prediction of potential miRNA target(s) for all the identified *TaTrx* genes were made by aligning the mRNA sequence against the miRNAs in wheat with perfect or near-perfect complementarity to mRNA using the psRNATarget server with default parameters (Dai X. et al., 2018). The interaction network of potential miRNA and *Trx* genes in wheat was predicted using Cytoscape (Dai L. et al., 2018).

Homology Modeling and 3D Structure Analysis of TaTrx Proteins

To compare the three-dimensional (3D) structure of TaTrx proteins, Homology modeling (comparative modeling) was done using the SWISS server (https://swissmodel.expasy.org/) to compare the three-dimensional (3D) structure of TaTrx proteins. Position-specific iterated BLAST (PSI-BLAST) was done to align template and target proteins; homolog template was selected based on similarity from protein database (PDB). The 3D structure of TaTrx proteins was predicted by an automated SWISS-MODEL server (https://swissmodel.expasy.org) (Waterhouse et al., 2018) and was further visualized in UCSF CHIMERA (Pettersen et al., 2004). TaTrx proteins were compared against templates 3d22.1.A, 2iwt.1.A, 2vlt.1.A, 2vm1.3.A, 1fb0.1.A, and 1faa.1.A from PDB. Ramachandran plot, per cent sequence identity of the target TaTrx proteins with template protein, template protein ID, QMEAN, template description, oligo state, and Ramachandran favored per cent were analyzed using the SWISS server.

Analysis of Protein–Protein Interaction Network

Interaction of TaTrx proteins with other proteins was analyzed using amino acid sequences of respective TaTrx proteins in STRING v11.0 (https://string-db.org) (Szklarczyk et al., 2019), the protein with the highest bit score was considered for analysis. GeneMANIA server (Warde-Farley et al., 2010) was also used for protein–protein interaction analysis by utilizing *Arabidopsis* interactome data.

In Silico Expression Analysis

Wheat RNA transcriptome data [WheatEXP (http://www.wheat-expression.com/)] available under different biotic stress (Powdery mildew pathogen E09 and stripe rust pathogen CYR31) treatments were used to compare the relative expression of all the identified *TaTrx* genes. Extracted transcript IDs were used for generating a heat map using the Wheat Expression browser expVIP (http://www.wheat-expression.com).

Plant Materials and Pathogen

The seeds of two contrasting wheat genotypes, Chinese spring (CS; susceptible genotype) and Transfer (TR; resistant genotype), an Introgression line of leaf rust resistance gene (*Lr9*) in CS background available at the Division of Genetics, IARI New Delhi, were used for the current experiment. The uredospores of leaf rust pathogen *Puccinia triticina* Eriks Pathotype 77-5 (121-R-63) were used to inoculate two wheat genotypes.

Inoculation and Sampling

Seedlings of CS and TR were grown in the National Phytotron Facility at ICAR-Indian Agricultural Research Institute, New Delhi, following the protocol by Prabhu et al. (2012). Ten to fifteen seeds of each genotype were sown, raised, and seedlings at the two-leaf stage were inoculated with the pathotype 77-5 using the protocol by Dhariwal et al. (2011). After inoculation, seedlings were kept in a humidity chamber for 24 h. Sampling was done by collecting the leaves at 0 HAI (Hours After Inoculation, uninoculated), 24 HAI, 72 HAI, and 144 HAI for analysis.

Isolation of Total RNA and cDNA Preparation

Total RNA was extracted from sampled leaf tissues (3 biological and 3 technical replicates) of control and treated plants at different intervals after leaf rust infection using RNAeasy kit (Qiagen Inc., Chatsworth CA 91311, United States, Cat No: 749040) as per the protocol and instruction described in the kit. RNA was quantified using a Thermo nanodrop spectrophotometer, and the purity was analyzed by checking the values of the ratio of A_{260}/A_{280} and using gel electrophoresis. The cDNA was prepared using Invitrogen, Life Technologies, United States, reverse transcriptase kit.

Primer Designing and qRT-PCR Analysis

cDNA sequences of all the identified *TaTrx* genes were used to design primers (**Supplementary Table S1**). The qRT-PCR assay was performed to determine the expression level of *Trx* genes in wheat. The qRT-PCR assay was done using a mixture of cDNA, forward primer, reverse primer, and SYBR* Green Master Mix (Applied Biosystems, United States) on a real-time detection system (CFX96 Touch real-time PCR detection, BIO-RAD life sciences). To set up qRT-PCR, all components of cDNA, forward and reverse primer, and SYBR* Green were carefully added in PCR plates and centrifuged for 1–2 min. All components were thawed properly to avoid differences in concentration. All the above steps were performed on ice. Care was taken to prevent the exposure of SYBR* Green to light and high temperatures. A master mix includes 4.5 µl RNase-free water, 0.75 µl forward primer, 0.75 µl reverse primer, and 7.5 µl Power SYBR Green Master Mix. 1.5 µl cDNA template was pipetted out into the PCR plate followed by the master mix. The master mix was kept in the dark to avoid light exposure. Wheat gene *TaActin* (housekeeping gene) was used as an internal control for normalization of the data for each transcript (Sathee et al., 2018) and level of expression or fold change in expression of *TaTrx* genes were analyzed using the 2^{-ddCt} method (Livak and Schmittgen., 2001).

Estimation of Reactive Oxygen Species

Based on previous reports on ROS and RNS abundance (Qiao et al., 2015), 72 HAI samples were selected to study ROS localization and the activity of antioxidant enzymes. After inoculation with pathotype 77-5, the leaves of the two genotypes were sampled 72 h after inoculation for ROS staining, ROS determination, and for an assay of antioxidant enzymes. The second leaves of inoculated and uninoculated wheat seedlings were used to analyze tissue localization of superoxide ions (O^{2-}) anions and H_2O_2. For determining O^{2-} radicals, 1 cm long leaf cuttings were incubated in a 6 mM NBT solution prepared in sodium citrate buffer (pH 7.5). The samples were vacuum infiltrated at 60 KPa for 10 min. After 10 min at room temperature, the samples were boiled in 80% ethanol in boiling water until the tissue turned translucent. Samples were mounted on glass slides after dipping in 20% glycerol. The formation of the dark blue color specified the localization of O^{2-} radicals. For detecting the formation of H_2O_2, the leaf cuttings were incubated in 1 mg/ml (p^H 3.8) DAB (3,3′-Diaminobenzidine) solution. Vacuum infiltration and chlorophyll removal steps were followed as described previously. The formation of the brown color specified the localization of H_2O_2. The slides were observed under stereomicroscope (EVOS XL Core), and the images were captured following Kumar et al. (2014).

Assay of Oxidative Stress

Spectrophotometric assay quantitatively estimated superoxide radicals at 72HAI in fresh leaf tissue (Chaitanya and Naithani, 1994). The amount of NBT reduced by O^{2-} radicals was assayed. Phosphate buffer of 0.2 M (pre-cooled, *p*H 7.2) was used to homogenize the leaf sample (1 g). The homogenized sample was centrifuged at 10000 g for 30 min at 4°C. The supernatant was collected and stored in a −20°C freezer until further analysis. The assay mixture comprised 0.1 mM EDTA, 0.075 mM NBT, 13.33 mM L-methionine, 25 mM Na_2CO_3, 250 µl of supernatant in a final volume of 3 ml, and the absorbance was measured at 540 nm.

For estimation of H_2O_2, 1 g of leaf sample was crushed in liquid nitrogen, homogenized in 10 ml cooled acetone, and filtered (Whatman no. 1). The filtrate was mixed with 5 ml ammonium solution and 4 ml titanium reagent for precipitation of titanium-hydrogen peroxide complex. The precipitate obtained after centrifugation at 10000 g for 10 min

was dissolved in 10 ml of 2M H_2SO_4. This dissolved precipitate was re-centrifuged, and the supernatant was used for spectroscopic absorbance measurement at 415 nm (Rao et al., 1997). The membrane injury was estimated by Evans blue staining followed by spectrophotometric estimation of tissue-bound dye (Kato et al., 2007).

Estimation of Antioxidant Enzymes

One gram of fresh leaf sample was sampled for preparation of extract to measure the activity of ascorbate peroxidase (APX), superoxide dismutase (SOD), peroxidase (POX) and catalase (CAT) enzymes. The sample was homogenized in 10 ml of 0.1 M phosphate extraction buffer (pH 7.5, comprising 1 mM ascorbic acid and 0.5 mM EDTA). The extract was filtered using a four-layered cheesecloth. The filtrate obtained was centrifuged at 15,000 g for 20 min. The activity of enzymes was estimated using the supernatant (Dhindsa et al., 1981).

Inhibition of the photochemical reduction of NBT was the basis for SOD activity estimation (Dhindsa et al., 1981). From 2 mM riboflavin stock, 0.1 ml was mixed into the reaction mixture containing 0.1 ml enzyme extract and incubated for 15 min under 15W lamps. A set of reaction mixture kept in the dark (transparent) served as the blank. After 15 min of incubation, absorbance was recorded at 560 nm.

Ascorbic acid-induced reduction in absorbance (290 nm) was served as the basis for estimation of Ascorbate peroxidase activity (Nakano and Asada, 1981). The reaction mixture comprised of 1.5 ml of 100 mM potassium phosphate buffer, 0.5 ml 3.0 mM ascorbic acid, 0.1 ml of 3.0 mM of EDTA, 0.1 ml of 3.0 mM Hydrogen peroxide, 0.1 ml enzyme of enzyme extract, and final volume made to 3 ml using double distilled water. Before measuring the optical density (OD), 0.2 ml of hydrogen peroxide was added to the reaction mixture. The reduction in the OD was recorded for 1 min at 290 nm wavelength in a UV–visible spectrophotometer.

Reduction in absorbance (240 nm) due to hydrogen peroxide decomposition served as the basis for estimating catalase (CAT) activity. The reaction mixture (3 ml) includes 1.5 ml of 0.1 M phosphate buffer and 0.5 ml of 75 mM H_2O_2. To which diluted enzyme extract of 50 µl was added and UV–visible spectrophotometer was used to measure the decline in the absorbance for 1 min at 240 nm.

The rise in absorbance (470 nm) due to the development of tetra-guaiacol served as the basis for estimating peroxidase enzyme activity. The reaction mixture includes enzyme extract (0.1 ml), guaiacol (16 mM), phosphate buffer (50 mM, pH 6.1), and 2 mM H_2O_2. The reaction started with the addition of guaiacol (16 mM). The rise in absorbance at 470 nm was recorded in a UV–visible spectrophotometer for 1 min after the addition of guaiacol (16 mM). Tetra-guaiacol extinction coefficient (ε = 26.6 mM^{-1} cm^{-1}) was used to calculate the enzyme activity (Castillo et al., 1984).

Statistical Analysis

GraphPad Prism version 8 (La Jolla, California, United States) was used to compute analysis of variance (ANOVA), correlation, and prepare graphs. A model showing thioredoxins (Trx) mediated regulation of incompatible interaction between wheat and leaf rust pathogen was created using **Inkscape** 0.92.3 (freely available at: https://inkscape.org/).

RESULTS

Genome-Wide Identification of Trx Genes

Genome-wide analysis of orthologous *Trx* genes in wheat (*TaTrx*) lead to the identification of 42 *TaTrx* genes from IWGSC RefSeq assembly v1.0 (**Supplementary Table S2**). Out of 42 *TaTrx* genes, 10 *TaTrx* genes were mapped on genome A, 21 *TaTrx* genes were located on genome B, and 11 *TaTrx* genes were positioned on genome D. The identified *TaTrx* genes were named *TaTrx*-1 to *TaTrx*-42 based on their chromosomal positions. InterPro and PROSITE analysis resulted in a conserved Trx domain in all the identified TaTrx and Trx proteins sequences of Rice and Arabidopsis. Blast search against four different wheat species led to identifying a set of 10, 20, and 32, *Trx* genes in *A. tauschii (DD), T. urartu (AA)*, and *T. turgidum (AABB),* respectively. Ten *Trx* genes in *A. tauschii* are located on six different chromosomes (1D: 3, 2D:2, 3D:1, 4D: 1, 5D: 2, and 6D: 1). *T. turgidum* contains a set of 32 *Trx* genes over six different homeologous groups involving wheat genome A (1A: 3, 2A:3, 3A:2, 4A:3, and 7A:1) and B (1B:6, 2B:3, 3B:4, and 5B:5) (**Supplementary Table S3**).

In Silico Expression Analysis of *TaTrx* Genes

Expression of thioredoxins was retrieved from expVIP from experiments in wheat inoculated with stripe rust pathogen (Pathotype: CYR31) at the duration of 24 h, 48 h, and 72 h after inoculation (HAI) and powdery mildew pathogen (Pathotype: E09) at the span of 24 h, 48 h, and 72 h after inoculation. As leaf rust is also a biotrophic pathogen, we considered stripe rust's *in silico* expression and powdery mildew (**Figure 1**). Based on the *in-silico* expression analysis in response to perturbance to the two biotrophic pathogens in the wheat expression browser, 15 selected *Trx* transcripts/genes showed a differential expression level to biotic perturbation. The Ensemble gene ID, gene length (bp), protein length (amino acid residues), number of splice variants and splice variant selected, coordinates, chromosomal location, and subcellular localization of *TaTrx* genes are listed in **Table 1**. Gene sequence investigation revealed that the length of coding sequence (CDS) of identified *TaTrx* genes range from 411 (*TaTrx10-2D*) to bp (*TaTrx13-5B*), and amino acids of corresponding protein range from 118 (*TaTrx9-2A*) to 189 (*TaTrx15-3B*).

Chromosomal Distribution

Based on the wheat genome database available at IWGSC-URGI and EnsemblPlants, positions of the selected *TaTrx* genes on corresponding chromosomes of A, B, and D genomes are shown in **Figure 2**. Selected 15 *TaTrx* genes were located on nine different wheat chromosomes (1A, 1B, 1D, 2A, 2D, 3B, 5A, 5B, and 5D). Out of 15 *TaTrx* genes, four genes were present

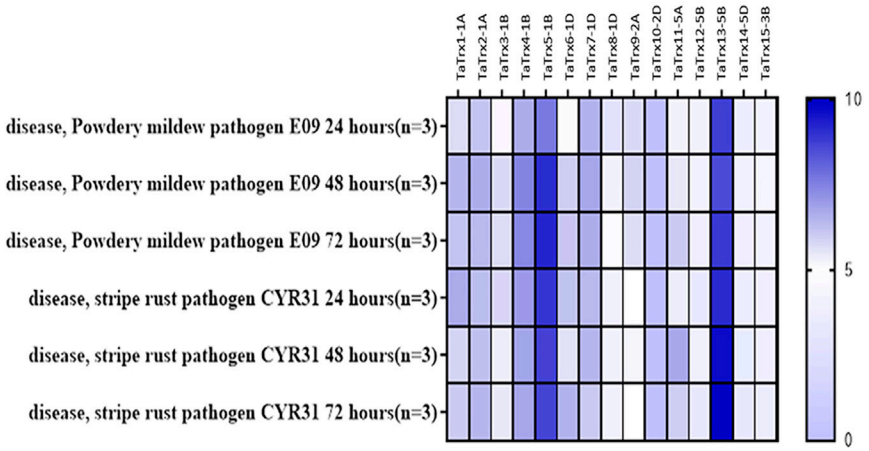


FIGURE 1 | *In silico* expression analysis of TaTrx genes against biotic stresses using expVIP (http://wheat-expression.com/).

TABLE 1 | Details of selected TaTrx genes in wheat genome along with their subcellular locations.

S. no.	Gene	Ensemble ID	Splice variant	Coordinates	Length bp	aa	Genome location	Subcellular location
1	*TaTrx1-1A*	TraesCS1A02G112400	1	114,060,494–114,063,211	858	131	1A: 114060494	Cytoplasm
2	*TaTrx2-1A*	TraesCS1A02G325600	1	516,430,393–516,432,298	845	130	1A: 516430393	Extracellular space
3	*TaTrx3-1B*	TraesCS1B02G132600	1	167,108,459–167,111,433	1,023	131	1B: 167108459	Cytoplasm
4	*TaTrx4-1B*	TraesCS1B02G338800	1	566,685,128–566,687,824	788	127	1B: 566685128	Chloroplast
5	*TaTrx5-1B*	TraesCS1B02G339000	1	566,823,400–566,831,663	742	188	1B: 566823400	Nucleus
6	*TaTrx6-1D*	TraesCS1D02G114000	1	110,052,054–110,054,117	664	131	1D: 110052054	Cytoplasm
7	*TaTrx7-1D*	TraesCS1D02G327200	1	419,471,916–419,473,737	757	126	1D: 419471916	Extracellular space
8	*TaTrx8-1D*	TraesCS1D02G327500	1	419,717,234–419,723,448	637	119	1D: 419717234	Nucleus
9	*TaTrx9-2A*	TraesCS2A02G243200	1	341,625,205–341,628,292	734	118	2A: 341625205	Cytoplasm
10	*TaTrx10-2D*	TraesCS2D02G244400	1	274,405,306–274,408,352	411	136	2D: 274405306	Cytoplasm
11	*TaTrx11-5A*	TraesCS5A02G448200	1	629,709,454–629,713,182	899	131	5A: 629709454	Nucleus
12	*TaTrx12-5B*	TraesCS5B02G111800	1	169,679,181–169,680,177	909	175	5B: 169679181	Mitochondrion
13	*TaTrx13-5B*	TraesCS5B02G452700	1	625,521,633–625,524,237	1,268	131	5B: 625521633	Nucleus
14	*TaTrx14-5D*	TraesCS5D02G454800	1	501,772,372–501,774,724	1,006	131	5D: 501772372	Nucleus
15	*TaTrx15-3B*	TraesCS3B02G453600	1	694,324,532–694,325,779	971	189	3B: 694324532	Chloroplast outer membrane

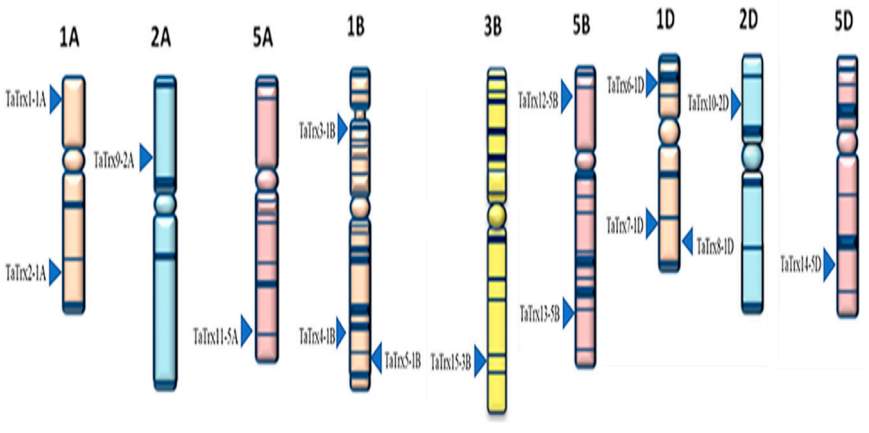


FIGURE 2 | Distribution of the selected 15 TaTrx genes across the nine chromosomes of groups A, B, and D sub-genomes of wheat. Figure taken from the IWGSC website (https://www.wheatgenome.org/).

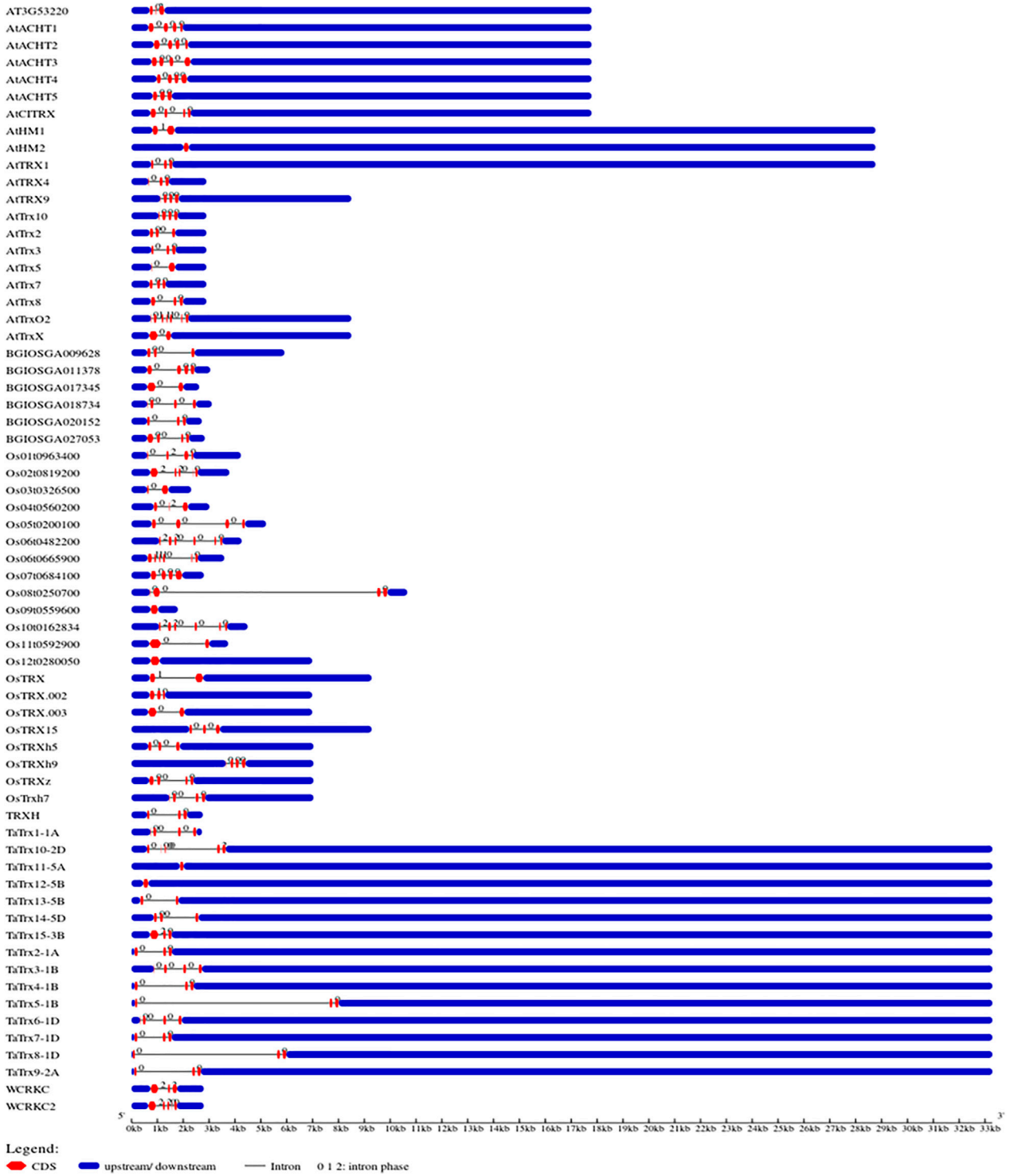


FIGURE 3 | Illustration of gene structure of identified wheat TaTrx genes along with Trxs from *Arabidopsis* and rice, showing the distribution of exons/introns, and exon phase created in the Gene Structure Display Server (GSDS v2.0) (http://gsds.cbi.pku.edu.cn/).

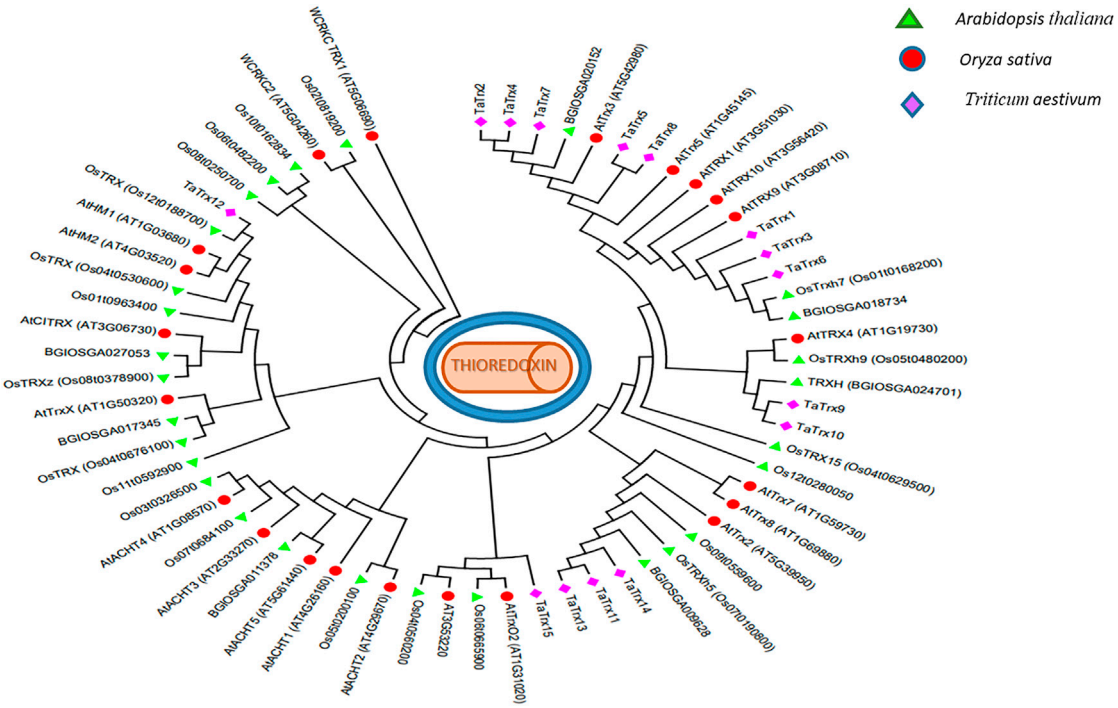


FIGURE 4 | Phylogenetic tree constructed using MEGA X (v7) software (https://www.megasoftware.net/) of the identified wheat TaTrx protein along with *Arabidopsis* and rice Trxs.

on three chromosomes of the genome A (1A:2 genes, 2A:1gene and 5A:1gene), six *TaTrx* genes were located on three B genome chromosomes (1B:3 genes, 3B:1 gene, and 5B:2 genes), and five *TaTrx* genes were mapped on three D genome chromosomes (1D: 3 genes, 2D:1 gene, and 5D: 1 gene).

Predicted Gene Structure, Phylogenetic Relationships, and Conserved Motif

In the present study, genomic sequences and cDNA sequences of 15 individual selected *TaTrx* genes were compared, and gene structure was obtained, as shown in **Figure 3**. Analysis of gene structure shows the presence of introns in all *TaTrx* genes, with the number of introns ranging from 0 to 5. *TaTrx11-5A* and *TaTrx12-5B* lack intron while *TaTrx10-2D* contains maximum of five introns. *TaTrx13-5B* possesses a single intron. *TaTrx3-1B* and *TaTrx6-1D* contain three introns, while the rest of the identified *TaTrx* genes contain two introns. *TaTrx5-1B* and *TaTrx8-1D* have the longest intron among all the identified *TaTrxs*, whereas all others have approximately similar lengths of introns.

Phylogeny analysis led to the clustering of a set of 65 protein sequences (**Figure 4**). The evolutionary relationship of identified 15 TaTrx proteins and Trx proteins in Rice (indica and japonica group) and *Arabidopsis* are represented by a phylogenetic tree constructed using the maximum likelihood method, suggesting Trx proteins mainly clustered into six major groups. Group I, the largest group, includes TaTrx2, TaTrx4, TaTrx5, TaTrx7, and TaTrx8 and two Trx from *Arabidopsis* one from rice. Group II contains TaTrx1, TaTrx3, and TaTrx6, three Arabidopsis, and two from Rice. TaTrx9 and TaTrx10 form the III group, TaTrx11, TaTrx13, and TaTrx14 form the IV group. Group V and Group VI contain only one Trx, TaTrx15, and TaTrx12. (**Figure 4**). All identified TaTrxs and Trx of *Arabidopsis* and rice shared a highly conserved Trx domain.

MEME software was used to analyze the conserved motifs of TaTrx proteins, and five distinct conserved regulatory motifs were predicted in *TaTrx* genes. All the five predicted regulatory motifs were conserved across all 15 TaTrx proteins schematically represented in **Supplementary Figure S1**. The sequences of conserved motifs with further information are provided in **Supplementary Table S4**.

Putative Cis-Regulatory Element Identification in the Promoter Region of the *TaTrx* Genes

TaTrx cis-regulatory elements investigation revealed potential cis-acting regulatory elements (CAREs) responsive to various hormones and defense and stress-responsive elements. Hormone-related elements include MeJA-responsiveness (MeJARE), abscisic acid responsiveness (ABRE), and salicylic acid responsiveness (SARE). Some other detected elements were related to development responsiveness (GARE, ARE, Meristem expression) and abiotic stress (DRE, MYB, MYC, LTRE, LRE) (**Supplementary Table S5**).

TABLE 2 | Predicted physicochemical properties of TaTrx proteins.

S. no.	Protein	aa	Mol. weight g/mol	Theoretical PI	Instability index	Aliphatic index	GRAVY	Stability
1	TaTrx1-1A	131	14,506.60	5.15	24.09	78.09	–0.273	YES
2	TaTrx2-1A	130	13,750.04	5.12	17.02	94	0.377	YES
3	TaTrx3-1B	131	14,506.60	5.15	24.09	78.09	–0.273	YES
4	TaTrx4-1B	127	13,523.76	5.12	16.52	94.65	0.347	YES
5	TaTrx5-1B	188	12,858.97	5.2	32.7	90.08	0.064	YES
6	TaTrx6-1D	131	14,506.60	5.15	24.09	78.09	–0.273	YES
7	TaTrx7-1D	126	13,446.62	5.12	14.79	96.11	0.308	YES
8	TaTrx8-1D	119	12,977.22	5.06	31.09	102.44	0.224	YES
9	TaTrx9-2A	118	12,693.68	5.29	18.71	86.86	0.156	YES
10	TaTrx10–2D	136	14,963.27	5.32	13.89	90.37	0.065	YES
11	TaTrx11-5A	131	14,538.72	6.29	31.29	78.17	–0.101	YES
12	TaTrx12-5B	175	19,132.37	8.67	47.13	87.49	0.025	NO
13	TaTrx13-5B	131	14,582.77	5.9	30.84	77.4	–0.136	YES
14	TaTrx14-5D	131	14,556.74	5.9	29.36	78.17	–0.11	YES
15	TaTrx15-3B	189	19,932.04	8.46	45.72	80.53	–0.02	NO

Of the 15 *TaTrx* genes, nine contained elements related to defense and stress, while all *TaTrx* genes contained conserved cis-elements for various plant hormones. Among these 9 *TaTrx* genes, each *TaTrx1-1B*, *TaTrx3-1B*, *TaTrx5-1B*, *TaTrx6-1D*, *TaTrx11-5A*, *TaTrx15-3B* gene holding only 1, *TaTrx13-5B*, *TaTrx14-5D* possess 2, and *TaTrx8-1D* possesses 4 cis-acting elements involved in defense and stress responsiveness. All identified cis-acting elements involved in defense and stress responsiveness possess TC-rich repeats.

Identification of miRNA Targets

Analysis of putative miRNA targets for *TaTrx* genes identified eight miRNAs that include targets for different *TaTrx* genes, as shown in **Supplementary Table S6**. The length of miRNA ranges from 19 nucleotides to 23 nucleotides. **Supplementary Table S6** also includes information on the sequence of miRNA aligned with the target sequence. The mode of action of miRNA miR9673-5p, tae-miR9674a-5p, and tae-miR9676-5p is translational inhibition, while the rest of miRNA causes cleavage of mRNA. Among 15 *TaTrx* genes, *TaTrx*10-2D is putative target for tae-miR9674a-5p and tae-miR9658-3p miRNA. *TaTrx*2-1A gene is putative target for tae-miR396-5p and tae-miR9673-5p. Putative targets for *TaTrx*5-1B, *TaTrx*13-5B, *TaTrx*7-1D and *TaTrx*4-1B are tae-miR1136, tae-miR9666a-3p, tae-miR9674b-5p, and tae-miR9676-5p miRNA, respectively. TaTrx- miRNA interaction network model was constructed using Cytoscape that shows putative miRNAs for targeting *TaTrx* genes (**Supplementary Figure S2**).

Physiochemical Properties of TaTrx Proteins

Physiochemical properties such as average residue weight (g/mol), charge, isoelectric point, molecular weight, theoretical PI, instability index, and grand average of hydropathicity (GRAVY) of 15 selected TaTrx proteins are listed in **Table 2**. Molecular weight (Mw) ranges from 12,858.97 g/mol (TaTrx5-1B) to 19,932.04 g/mol (TaTrx15-3B). Average residue weight ranges from 105.461 g/mol (TaTrx15-3B) to 111.319 g/mol (TaTrx14-5B), isoelectric point ranges from 4.8781 (TaTrx*2*-1A) to 8.3220 (TaTrx12-5B).

Out of 15 Trx proteins, 13 are stable with an instability index of less than 40, while the remaining two proteins are unstable with index more than 40. The aliphatic index of identified Trx proteins ranges from 77.40 (TaTrx13-5B) to 102.44 (TaTrx8-1D). A set of seven (7) TaTrx proteins: TaTrx1-1A, TaTrx3-1B, TaTrx6-1D, TaTrx11-5A, TaTrx13-5B, TaTrx14-5B, and TaTrx15-3B showed the negative value of the grand average of hydropathicity (GRAVY) while the remaining eight (8) TaTrx proteins: TaTrx2-1A, TaTrx4-1B, TaTrx5-1B, TaTrx7-1D, TaTrx8-1D, TaTrx9-2A, TaTrx10-2D, and TaTrx12-5B were recorded with a positive value.

Prediction of subcellular localization of Trx proteins (**Table 2**; **Supplementary Table S7**) shows that one TaTrx is located each in chloroplast (TaTrx4-1B), chloroplast outer membrane (TaTrx15-3B), and in mitochondria (TaTrx12-5B), two TaTrx (TaTrx2-1A and TaTrx7-1D) are extracellular while that five TaTrx each are located in cytosol (TaTrx1-1A, TaTrx3-1B, TaTrx6-1D, TaTrx9-2A, TaTrx10-2D) and in the nucleus (TaTrx5-1B, TaTrx8-1D, TaTrx11-5A, TaTrx13-5B, TaTrx14-5D).

Homology Modeling and 3D Structure Analysis of TaTrx Proteins

Three-dimensional (3D) structure of TaTrx proteins homology modeling (comparative modeling) is given in **Supplementary Table S8**, **Figure 5**. Homology modeling and 3D structure analysis showed that four TaTrx exist as dimers and 11 TaTrx as monomeric. TaTrx proteins were compared against templates

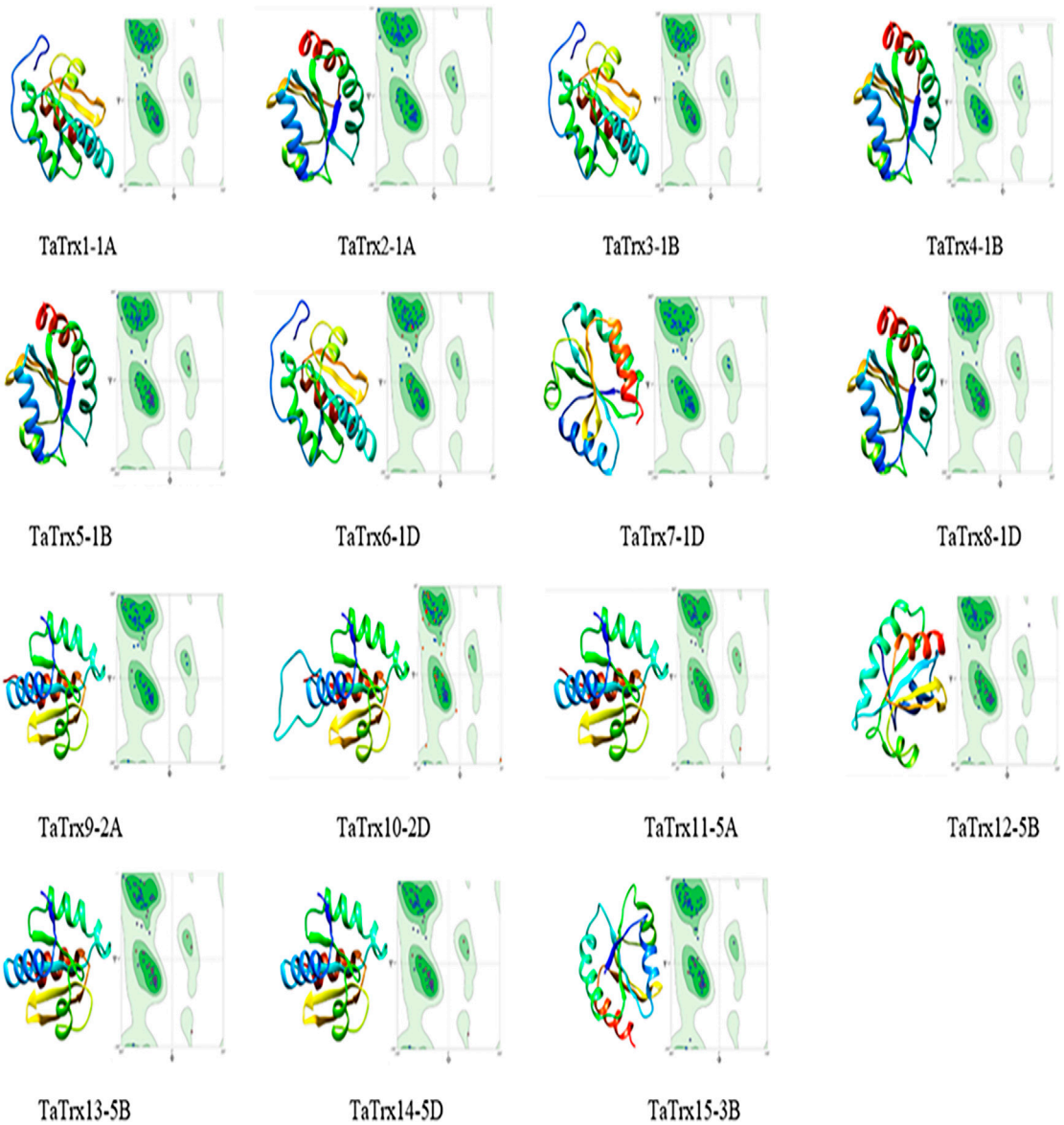


FIGURE 5 | Automatic prediction of 3D structure and homology modeling of TaTrx proteins generated using UCSF chimera (http://www.cgl.ucsf.edu/chimera). Ramachandran plot of corresponding TaTrx protein showing most favored region and allowed region created using SWISS-MODEL server.

3d22.1.A, 2iwt.1.A, 2vlt.1.A, 2vm1.3.A, 1fb0.1.A, and 1faa.1.A from PDB, which unveils the structure of Poplar thioredoxin h, thioredoxin h2, thioredoxin h isoform 2, barley thioredoxin h isoform 1, spinach thioredoxin M, and spinach thioredoxin F, respectively.

The per cent sequence identity of the target TaTrx proteins with template protein, template protein ID, QMEAN, template description, oligo state, and Ramachandran favored per cent were shown in the **Supplementary Table S8**. The predicted 3D structure of TaTrx proteins is shown in **Figure 5**, along with the Ramachandran plot of the corresponding TaTrx protein. Ramachandran plot shows that more than 95% (in some proteins more than 99%) of the amino acid residues of thioredoxin proteins lies in the most favored area.

Molecular Interaction Networks

The results of protein–protein interaction analysis done by STRING v11.0 for different TaTrx proteins are given in **Supplementary Figures S3A–I**. The PPI network shows interaction with thioredoxin reductases and several other uncharacterized proteins. Apart from thioredoxin reductase and a few uncharacterized proteins, TaTrx12-5B also interacted with chloroplastic peroxiredoxin. Protein–protein

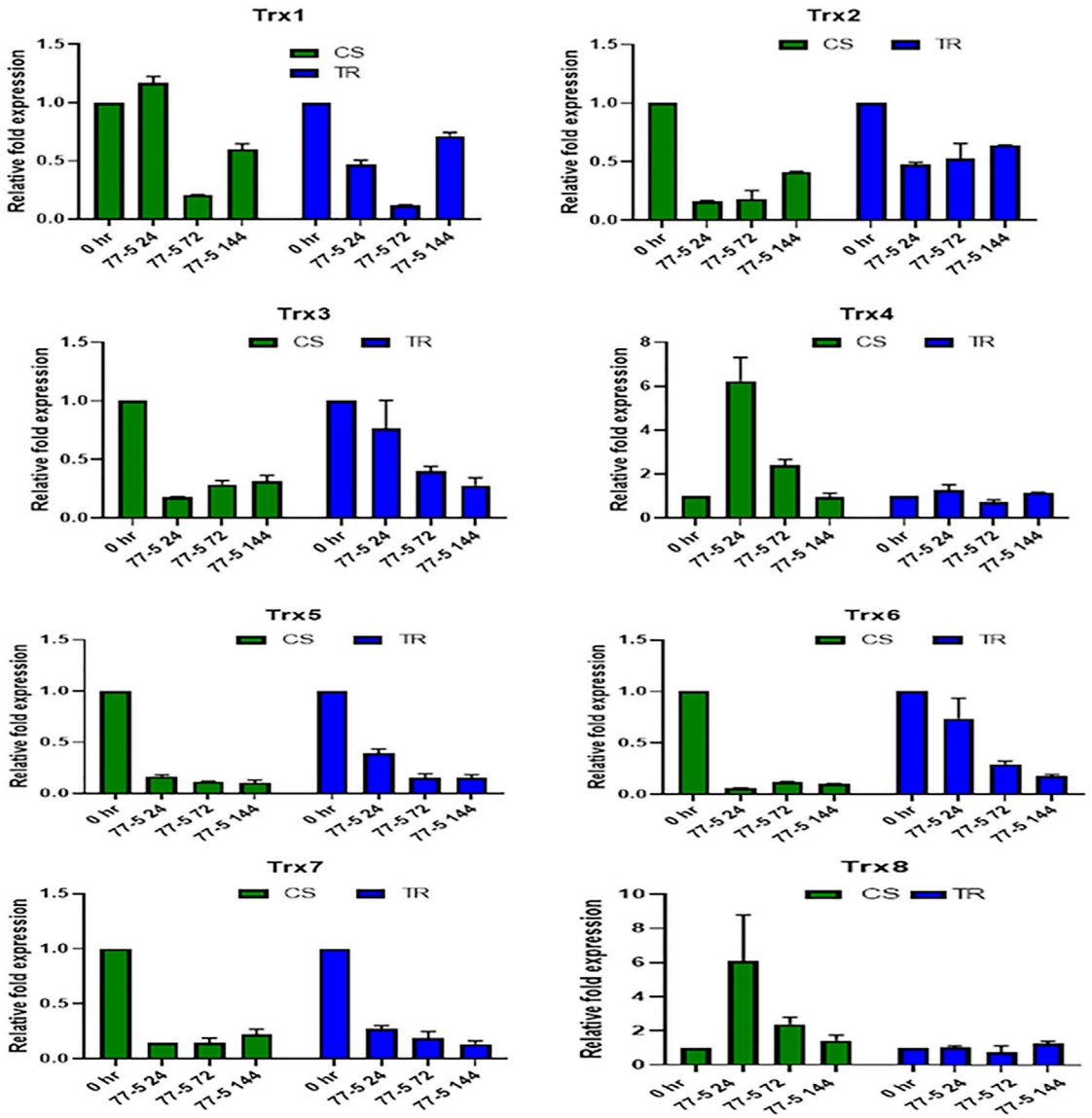


FIGURE 6 | Expression analysis of Trx family genes (Trx 1, Trx 2, Trx 3, Trx 4, Trx 5, Trx 6, Trx 7, and Trx 8) in response to leaf rust inoculation. Chinese spring (CS) and introgression line of leaf rust resistance gene *(Lr9)* in CS, Transfer (TR). Leaf rust pathogen *Puccinia triticina* Eriks Pathotype 77-5 (121-R-63) uredospores were inoculated on 10 days old wheat seedlings and kept in a humidity chamber for 24 h. Leaf samples for qRT-PCR analysis were collected at 0 h after inoculation (HAI), 24 HAI, 72 HAI, and 144 HAI. Values are mean (±SE) of three biological replicates.

interaction analysis of TaTrx protein revealed interaction of Trx5 with NPR1 and NPR3 (**Supplementary Figures S4, S5**).

Expression Analysis of Thioredoxin Genes in Wheat Seedlings

The infection recorded a total fold change in the genes expression of selected TaTrx genes profiling during different time intervals (0 HAI, 24 HAI, 72 HAI, and 144 HAI). The expression level of selected *Trx* genes is depicted in **Figures 6**, **7**. Among 15 *Trx* genes, the expression of *TaTrx4-1B* and *TaTrx8-1D* showed no or little change in expression in respect to control. *TaTrx3-1B*, *TaTrx5-1B, TaTrx6-1D*, *TaTrx7-1D*, and *TaTrx9-2A* showed a regular reduction in expression span of 24 HAI, 72 HAI, and 144 HAI in response to leaf rust infection after the inoculation in TR with respect to control. Whereas, *TaTrx11-5A* (3 fold, 72 HAI), *TaTrx13-5B* (3.36 fold, 72 HAI), *TaTrx14-5D* (3.86 fold, 72 HAI), and *TaTrx15-3B* (3 fold, 72 HAI) genes were significantly upregulated in expression in a span of 72 HAI and then declined sharply at 144 HAI (0.2 fold) in response to leaf rust infection in respect to control. Among the downregulated *TaTrx* genes, the expression of *TaTrx2-1A*, *TaTrx3-1B*,

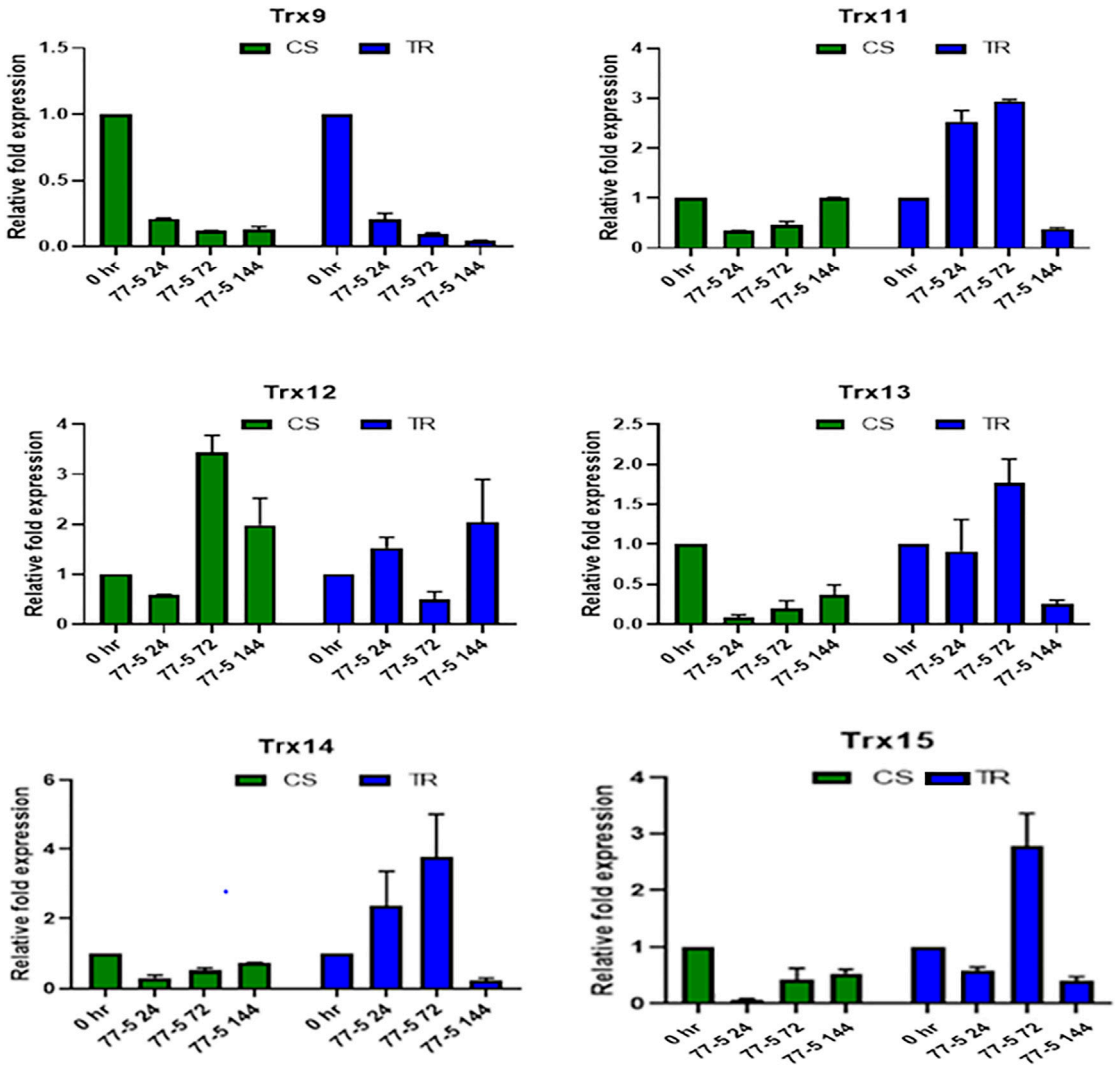


FIGURE 7 | Expression analysis of Trx family genes (Trx 9, Trx 11, Trx 12, Trx 13, Trx 14, and Trx 15) in response to leaf rust inoculation. Chinese spring (CS) and Introgression line of leaf rust resistance gene *(Lr9)* in CS, Transfer (TR). Leaf rust pathogen *Puccinia triticina* Eriks Pathotype 77-5 (112-R-63) uredospores were inoculated on 10 days old wheat seedlings and kept in a humidity chamber for 24 h. Leaf samples for qRT-PCR analysis were collected at 0 hafter inoculation HAI, 24 HAI, 72 HAI, and 144 HAI. Values are mean (±SE) of three biological replicates.

TaTrx5-1B, and *TaTrx6-1D* transcripts abundance was higher in resistance/incompatible interaction (TR) in comparison to susceptible/compatible interaction (CS) at different expression span.

Localization and Accumulation of ROS (Superoxide Radical (SOR) and Hydrogen Peroxide) and Activity of Antioxidant Enzymes in Wheat Seedlings

Compatible interaction of leaf rust resulted in ROS burst as indicated by localization and content of SOR and hydrogen peroxide. As presented in **Figures 8A,C**, the occurrence of dark blue color and the appearance of the brown-colored product confirmed the presence of SOR and hydrogen peroxide. Spectrophotometric assay of SOR and hydrogen peroxide also showed higher ROS levels in the compatible interaction of leaf rust (**Figures 8B,D**). Leaf rust inoculation triggered the activity of antioxidant enzymes Peroxidase (POX), Superoxide dismutase (SOD), and Ascorbate peroxidase (APX), while catalase activity was reduced after inoculation with the pathogen. The magnitude of upregulation in SOD and APX activity was higher (approximately three-fold) in incompatible interaction as compared to compatible interaction (approximately two-fold) that could have arrested the ROS burst as depicted as a lower abundance of both the ROS considered in the study (**Figure 9**).

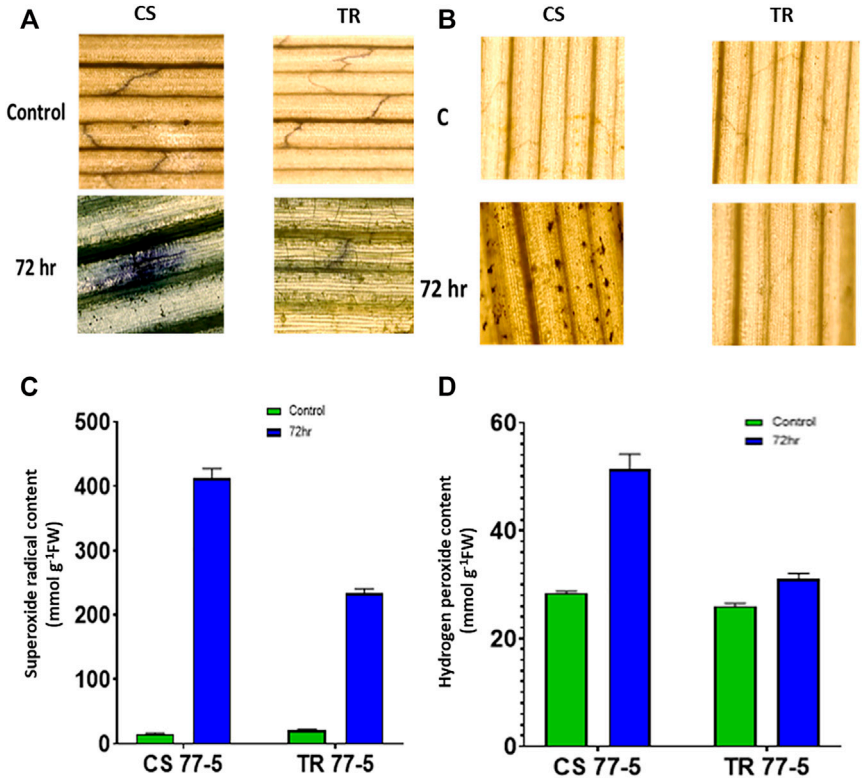


FIGURE 8 | Effects of leaf rust pathogen on the localisation and quantification of reactive oxygen species, superoxide radical **(A,B)** and hydrogen peroxide **(C,D)**. Chinese spring (CS) and introgression line of leaf rust resistance gene *(Lr9)* in CS, Transfer (TR). Leaf rust pathogen *Puccinia triticina* Eriks Pathotype 77-5 (121-R-63) uredospores were inoculated on 10 days old wheat seedlings and kept in a humidity chamber for 24 h. Leaf samples for analysis were collected at 0 h after inoculation (HAI), and 72 HAI. Values are mean (±SE) of three biological replicates.

Analysis of Variance

Analysis of variance (ANOVA) showed that out of 15 identified TaTrx, 13 TaTrx were significant at the expression level compared to control. At the same time, the remaining two genes, *TaTrx4-1B* and *TaTrx8-1D,* were non-significant (**Supplementary Table S9**). A heat map showing the Pearson correlation matrix relationship between expression of TRX genes, ROS accumulation and antioxidant enzyme activities in leaves of two contrasting wheat genotypes, CS and TR is given in **Supplementary Figure S6**. The correlation heat map showed that the accumulation of APX, SOD, POX, SOR, and HP (H_2O_2) positively correlated with the *TaTrx* genes exhibiting upregulation against infection. On the other hand, the accumulation of CAT showed a positive correlation with the *TaTrx* genes showed downregulation during the span of infection (**Supplementary Figure S6**).

DISCUSSION

Thioredoxin is a crucial factor that plays an essential role in host resistance and several other plant constituents. ROS production and scavenging are associated with the modulation of plant immune signaling pathways (Vieira Dos Santos and Rey, 2006). Trx acts as a critical regulator element in defense mechanisms and is essential for modulating the redox status of components of defense pathways. The study was based on genome-wide analysis of *Thioredoxin* genes in bread wheat that revealed one set of 15 *TaTrx* genes based on *in silico* expression analysis for biotic stress responses of wheat. Genome-wide analyses of *Thioredoxin* genes have also been performed in *Arabidopsis* and rice (Meyer et al., 2008; Nuruzzaman et al., 2012). Several genes included the Trx domain in various crop plants such as 74 genes in Arabidopsis, 61 in Rice, 11 in sorghum, and 36 in maize (Nuruzzaman et al., 2012).

A comparative analysis with the conserved or diverse *Trx* genes among the different wheat species; *T. Urartu* (A genome), *T. turgidum* (AB genome), and *A. tauschii* (D Genome), led to the identification of conserved genes on different three genomes A, B, and D. Both *T. aestivum* and progenitor's species shared a similar maximum number of genes on three different genomes, i.e., genome A with ~10 Trx genes, B carried ~21 Trx genes, and Genome D contributed a set of ~10 Trx genes. This is evidence that the Trx gene is conserved in the wheat genome throughout evolution.

Out of 15 TaTrxs, *TaTrx15-3B* (3 fold at 72 HAI) putatively localized in chloroplast was found to be upregulated in

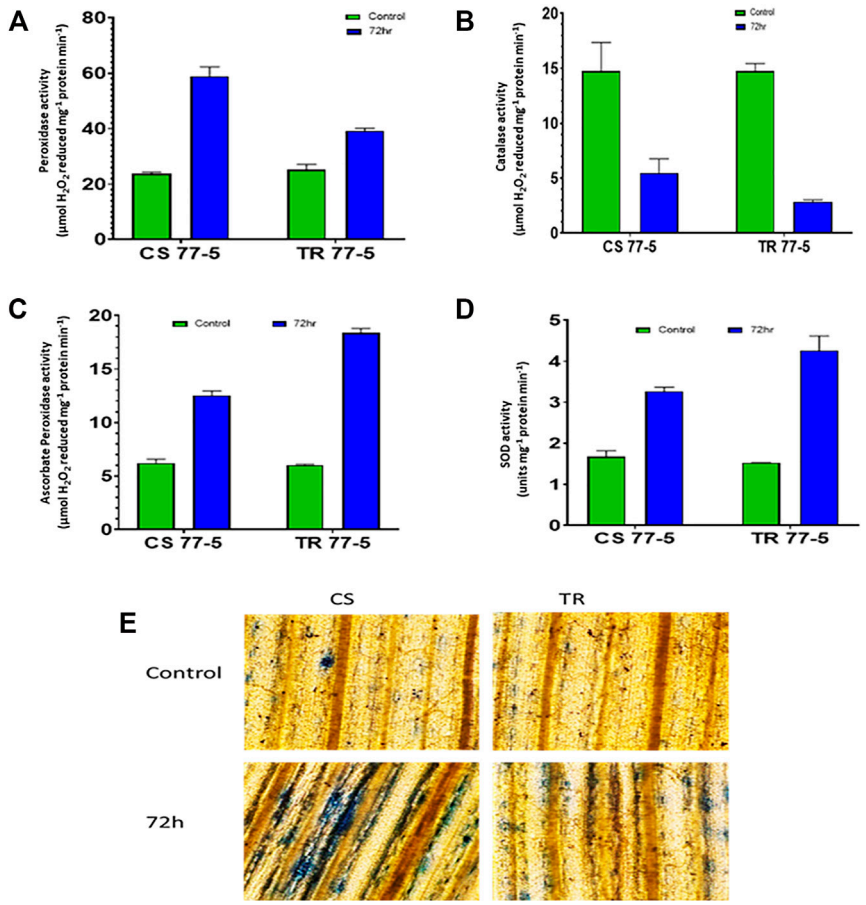


FIGURE 9 | Effect of leaf rust pathogen on the activity of antioxidant enzymes **(A)** peroxidase, **(B)** catalase, **(C)** ascorbate peroxidase, **(D)** superoxide dismutase, and **(E)** membrane injury visualized by Evans blue staining of leaves of wheat genotypes Chinese spring (CS) and introgression line of leaf rust resistance gene (Lr9) in CS, Transfer (TR). Leaf rust pathogen *Puccinia triticina* Eriks pathotype 77-5 (121-R-63) uredospores were inoculated on 10-day-old wheat seedlings and kept in a humidity chamber for 24 h. Leaf samples for analysis were collected at 0 h after inoculation (HAI) and 72 HAI. Values are means (±SE) of three biological replicates.

incompatible interaction with leaf rust pathogen. Most chloroplastic Trxs are involved in the photosynthetic process and respond to stress and hormone signaling (Fernández-Trijueque et al., 2019). Overexpression of a chloroplastic thioredoxin NtTRXh3 in tobacco showed to be involved in enhanced resistance against cucumber mosaic virus (CMV) and tobacco mosaic virus (TMV) (Sun et al., 2010). This study showed that of 15, 5 TaTrx genes (TaTrx1-1A, TaTrx3-1B, TaTrx6-1D, TaTrx9-2A, and TaTrx10-2D) located in cytoplasm either showed no or little change in expression with respect to treatments; however, the previous studies reported that in most of the cases cytosolic Trxs such as AtTrxh5 against *Pseudomonas syringae* and *Cochliobolus victoriae*, CITRX in response to *C. fulvum* were upregulated and involved in defense responses. Based on the location of three highly upregulated TaTrx genes, we found that nuclear Trxs were upregulated in response to leaf rust likely function in plant immunity.

In this study, we predicted the physiochemical properties, including theoretical PI, molecular weight, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) of TaTrx proteins. The p^H-related characteristics of a protein depend upon theoretical PI; at this point, protein has no charge and is less soluble, which facilitates the isolation of protein (Righetti, 2004; Audain et al., 2016). *In vivo* stability of a protein is inferred from the instability index calculated by the dipeptide composition-based method; a value less than 40 signifies that protein is stable, while a value more than 40 means that protein is unstable (Gamage et al., 2019). In the study, 13 proteins have an instability index value of less than 40, while two have more than 40, indicating the stable nature of the TaTrx protein. The aliphatic index (AI) is the relative volume occupied by side aliphatic amino acid residues that signify the thermostability of a protein (Ikai, 1980). All selected TaTrx proteins in wheat have a wide range of the aliphatic index. This suggests the wide range of thermostability of the Trx protein. In the present study, TaTrx proteins showed both negative (7 proteins) and positive (8 proteins) GRAVY values. GRAVY measures solubility and hydrophobicity or hydrophilicity; negative value specifies hydrophobic nature

while a positive value specifies the hydrophilic nature of a protein (Fernández-Fernández and Corpas, 2016).

Phylogeny analysis of *TaTrx* and *Trx* of *Arabidopsis* and rice indicated that these belong to six different groups having similarities in exon-intron position and presence of domain. Genomic and CDS sequences were retrieved and used to analyze gene structure, indicating that the number of exons ranges from 1 to 6 while introns range from 0 to 5. Highly upregulated *TaTrx* genes (*TaTrx*11, *TaTrx* 13, *TaTrx*14, and *TaTrx*15) have 0–2 introns and are shorter than other identified Trx proteins. It is well known that introns play a key regulator role in alternative splicing (AS) and non-sense-mediated decay (NMD) (Kalyna et al., 2012), and it has been reported that genes with the fewer number of introns allow plants to respond more quickly against stress (Jaffares et al., 2008; Zheng et al., 2020). Therefore, identified *TaTrx* genes having fewer introns may express highly and quickly against leaf rust pathogen. However, the role of intron number on the function of TaTrx proteins needs further validation. Conserved motif analysis revealed that Trx contains five conserved regulatory motifs. All other Trx proteins, except four highly upregulated TaTrxs, have a similar pattern of conserved motifs, representing their involvement in gene regulation. Out of four highly expressed *TaTrxs*, *TaTrx*15-3B possesses only one motif identical to the other three upregulated Trx proteins. We can predict that the dissimilarity of TaTrx15 protein from others may provide its specificity during resistance. The *TaTrx15-3B* gene is located in chloroplast, while the remaining three highly upregulated genes are present in the nucleus showed significant upregulation during infection at different cellular compartment levels.

In the present study, homology modeling and 3D structure analysis showed that proteins shared more than 95% (in some proteins more than 99%) similarity. According to the PROCHECK algorithm, a quality protein model is expected to possess more than 90% amino acid residues in the most favored region (Laskowski et al., 1993), which advocates that our predicted models are of good quality. In our study, promoter regions of *TaTrx* genes possess CAREs as follows: MeJARE, ABRE, MYB, G-box, and W-Box. G-box and ABRE provide the suitable binding site for bZIP TFs that regulates and plays a vital role in NPR1 (NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1) mediated SAR (Heinekamp et al., 2004). Cis-acting regulatory elements are essential in response to biotic and abiotic stress (Sheshadri et al., 2016). Jasmonic acid and salicylic acid plays a vital role in orchestring signaling for SAR (Fu et al., 2012; Heil et al., 2012). As the SA level increases after pathogen infection during basal resistance, SA binds to NPR4 and releases more NPR1, which activates the SAR (Fu et al., 2012). MeJA act as an activator of the antioxidant system in plants during the production of the damaging level of ROS (Yu et al., 2018). Along with Trxs, MeJA may play a crucial role in ROS homeostasis during pathogen infection. Also, the protein–protein interaction showed Trx is a putative partner in interaction with NPR1.

Along with regulatory elements, some miRNAs were also predicted to be associated with the modulation of gene expression. Eight putative miRNAs were discovered that seven different target *TaTrx* genes. Differential expression of miRNAs has been observed during defense responses (Ramachandran et al., 2020). Identified *Trx* gene, *TaTrx13-5B* (upregulated during leaf rust infection) is a putative target for tae-miR9666a miRNA while the Trx gene *TaTrx5-1B* gene (downregulation during infection) is a putative target of MIR1136. The expression of miRNA tae-miR9666a and MIR1136 was found to be downregulated and upregulated, respectively, during treatment in stripe rust infected spring wheat cultivar (Ramachandran et al., 2020). This supports our study that downregulation of tae-miR9666a during the resistance to leaf rust might be involved in the upregulation of the expression of *TaTrx13-5B* in wheat. In contrast, the change in the expression level of MIR1136 could be associated with the downregulation of its target *TaTrx5-1B* gene.

Although most of the proteins in wheat are still uncharacterized, protein–protein interaction analysis using STRING recognizes that all TaTrxs interact with thioredoxin reductase. However, protein–protein interaction analysis in *Arabidopsis* revealed Trx5, thioredoxin H-type 5 (cytosolic thioredoxin) acts as a reducer of disulphide bridges and interacts with NPR1, NPR3 involved in systemic acquired resistance (SAR) and other thioredoxins (**Supplementary Figures S4, S5**). From the previous and present studies, it is clear that Trxs have a diverse role in plant immunity. Recently, a study provided insights into wheat's molecular interaction network during the defense response pathway against stripe rust (*P. striiformis* f.sp. *tritici*). A thioredoxin gene has been identified and cloned in wheat (*TaTrxh1)* showed interaction with *TaCP1* (RD-19 like Cysteine protease) that enhances resistance against wheat stripe rust (*P. striiformis f.* sp. *tritici*) through modulation of PCD *via* SA dependent defense signaling pathway. Overexpression of *TaTrxh1* induces cell death, and *TaTrxh1* expression is induced significantly in response to *P. striiformis f.* sp. *tritici* infection (Shi et al., 2021). *TaTrx13-5B* (3.36 fold) showed upregulation in response to leaf rust infection with maximum expression at 72 HAI. This result aligned well with the data obtained in the *in silico* analysis experiment regarding stripe rust and powdery mildew disease (**Figures 1**, **7**), as the expression of *TaTrx13-5B* transcript was also upregulated under both the biotic stresses. Therefore, it indicates that *TaTrx13-5B* could be the Trx involved in response to most biotrophic pathogens.

Superoxide production, a ROS, was first reported to be associated with potato's hypersensitive response (HR) to infection of an incompatible race of *Phytophthora infestans*. Out of three inoculation spans (24 HAI, 72HAI, and 144HAI), maximum accumulation of superoxide and SOD and CAT activity was observed at 72 HAI. Localization of superoxide was observed in the case of the incompatible one but not in the compatible one (Doke, 1983). Similar results were observed from our experiment of superoxide localization in response to

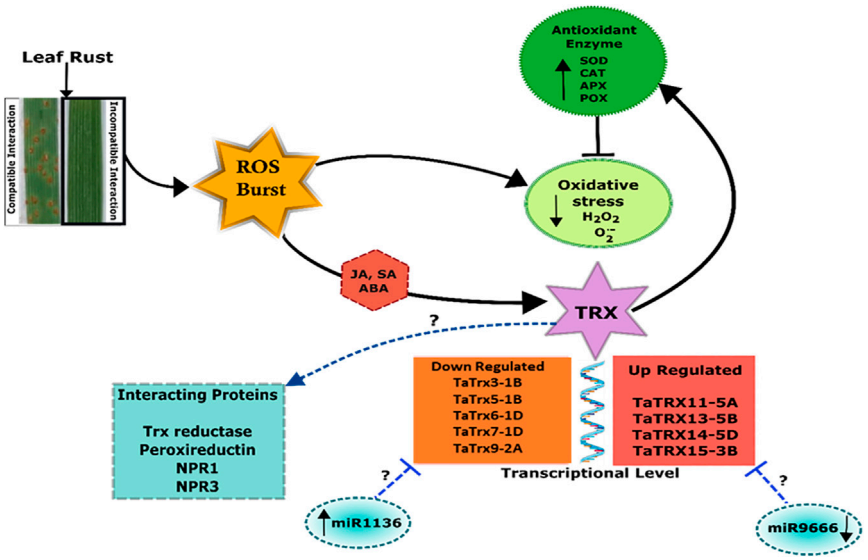


FIGURE 10 | Schematic model showing thioredoxin (Trx)-mediated regulation of incompatible interaction between wheat and leaf rust pathogen. Figure is made based on the results and *in silico* prediction (enclosed in dashed lines).

leaf rust that showed maximum accumulation at 72 HAI. In barley, H_2O_2 accumulation started at 6 HAI in response to powdery mildew infection and increased with time, HR was observed during 24 HAI (Thordal-Christensen et al., 1997). In our case, maximum accumulation of H_2O_2 was observed during 72 HAI (**Figure 8**). In uninoculated lettuce plants, H_2O_2 was localized in secondary thickening of xylem vessels while inoculation with *Pseudomonas syringae pv phaseolicola* causes highly localized accumulation of H_2O_2 resulting in HR (Bestwick et al., 1997). We found a similar accumulation of H_2O_2 around the leaf rust spores (**Figure 8**). Study of root apoplast secretome in wilt resistance *Gossypium barbadense* in response to *Verticillium dahliae* showed upregulation of ROS and defense-related proteins. Silencing a thioredoxin GbNRX1 showed that Trx is associated with ROS metabolism and is required for resistance against *Verticillium dahliae* in cotton (Li et al., 2016). Three types of Trx-m and a Trx-x in *Arabidopsis thaliana* were found to be upregulated in response to H_2O_2 accumulation and induce H_2O_2 tolerance (Issakidis-Bourguet et al., 2008). Thioredoxin, *PsTrx-h1* in pea was reported to be significantly induced in response to oxidative stress and showed that PsTrx-h1 was involved in redox homeostasis (Traverso et al., 2007). A recent study in wheat clearly showed that silencing of a wheat thioredoxin *TaTrx1* reduces the level of accumulation of ROS in response to *P. striiformis* f. sp. *tritici* infection.

Further, estimation of H_2O_2 in TaTrx1 silenced plants following inoculation with *P. striiformis* f. sp. *Tritici* race CYR-23 showed a significantly reduced area of H_2O_2 production at 120 HAI as compared to control plants. Parallel to this, the expression of antioxidants TaCAT and TaSOD were significantly induced as expected (Shi et al., 2021), which supports our study of the involvement of TaTrx in ROS homeostasis and ROS-dependent defense. We found that compatible interaction of leaf rust resulted in ROS burst as indicated by localization and content of SOR and hydrogen peroxide. Leaf rust inoculation triggered the activity of antioxidant enzymes peroxidase (POX), superoxide dismutase (SOD), and ascorbate peroxidase (APX), while catalase activity was reduced after inoculation with the pathogen (**Figure 9**). The magnitude of upregulation in SOD and APX activity was higher in incompatible interaction, which could have arrested the ROS abundance. Based on our results, we can say that Trxs in wheat also act as an antioxidant and are involved in the scavenging of ROS and reduce oxidative stress during incompatible interaction with leaf rust. Based on superoxide and H_2O_2 quantification and localization, it is inferred that ROS and antioxidant enzymes attained a maximum level at 72 HAI. Upregulated *Trx* genes also showed maximum expression during 72 HAI. The correlation of *Trx* gene expression, accumulation of ROS, and antioxidant enzymes at 72 HAI also indicates a significant association. For example, as depicted in the correlation heat map (**Supplementary Figure S6**), *TaTrx1-1A* and *TaTrx4-1B* expression showed a positive correlation with catalase activity and negative correlation with ROS accumulation *TaTrx2-1A* and *TaTrx6-1D* showed a positive correlation with most of the antioxidant enzymes. The perturbation in ROS homeostasis and gene expression indicates the involvement of thioredoxins (which has a regulatory role as antioxidant enzyme activity) in ROS homeostasis and leaf rust resistance in wheat (**Figure 10**).

CONCLUSION

The wheat genome contains 42 members of the *Trx* gene family unevenly distributed on the seven chromosomes of A, B, and D genomes. Expression profiles of selected 15 Trxs, estimation, and localization of ROS demonstrate that the *Trx* gene family is associated

with ROS homeostasis during biotic stress. Although previous studies indicated the participation of Trxs in various metabolic processes, their role in plant immunity is so far not entirely understood. Overall, this study provides insights into the function of the *Trx* gene family in response to leaf rust.

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.

ETHICS STATEMEMT

The experimental research on plants complied with relevant institutional, national, and international guidelines and legislation.

AUTHOR CONTRIBUTIONS

RB conducted the investigation and wrote the original draft. DH conducted the investigation. ST reviewed the manuscript and supervised the work. LS helped in conceptualization, data curation, and writing the original draft, reviewing, and editing. SA conducted data curation. DS conducted supervision. NM collected the resources and reviewed the manuscript. V collected the resources and helped in conceptualization. VC helped in conceptualization. SJ helped in conceptualization, obtained resources, and reviewed and edited the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to the ICAR-Indian Agricultural Research Institute for funding and providing the necessary facilities. They also acknowledge the financial support received from NAHEP-CAAST, ICAR-IARI. DH and RB acknowledge the ICAR- Junior research fellowship support received during the course of the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.836030/full#supplementary-material

REFERENCES

Apel, K., and Hirt, H. (2004). Reactive Oxygen Species: Metabolism, Oxidative Stress, and Signal Transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi:10.1146/annurev.arplant.55.031903.141701

Audain, E., Ramos, Y., Hermjakob, H., Flower, D. R., and Perez-Riverol, Y. (2016). Accurate Estimation of Isoelectric point of Protein and Peptide Based on Amino Acid Sequences. *Bioinformatics* 32, 821–827. doi:10.1093/bioinformatics/btv674

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 (Suppl. l_2), W202–W208. doi:10.1093/nar/gkp335

Bailey, T. L., Elkan, C., et al. (1994). *Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers*. UCSD Technical Report CS94-351.

Baxter, A., Mittler, R., and Suzuki, N. (2014). ROS as Key Players in Plant Stress Signalling. *J. Exp. Bot.* 65, 1229–1240. doi:10.1093/jxb/ert375

Bestwick, C. S., Brown, I. R., Bennett, M. H., and Mansfield, J. W. (1997). Localization of Hydrogen Peroxide Accumulation during the Hypersensitive Reaction of Lettuce Cells to *Pseudomonas syringae* Pv Phaseolicola. *Plant Cell* 9, 209–221. doi:10.1105/tpc.9.2.209

Bolser, D., Staines, D. M., Pritchard, E., and Kersey, P. (2016). Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods Mol. Biol.* 1374, 115–140. doi:10.1007/978-1-4939-3167-5_6

Castillo, F. J., Penel, C., and Greppin, H. (1984). Peroxidase Release Induced by Ozone in Sedum Album Leaves. *Plant Physiol.* 74, 846–851. doi:10.1104/pp.74.4.846

Chae, H. Z., Chung, S. J., and Rhee, S. G. (1994). Thioredoxin-dependent Peroxide Reductase from Yeast. *J. Biol. Chem.* 269, 27670–27678. doi:10.1016/s0021-9258(18)47038-x

Chaitanya, K. S. K., and Naithani, S. C. (1994). Role of Superoxide, Lipid Peroxidation and Superoxide Dismutase in Membrane Perturbation during Loss of Viability in Seeds of Shorea Robusta Gaertn.F. *New Phytol.* 126, 623–627. doi:10.1111/j.1469-8137.1994.tb02957.x

Collet, J.-F., D'Souza, J. C., Jakob, U., and Bardwell, J. C. A. (2003). Thioredoxin 2, an Oxidative Stress-Induced Protein, Contains a High Affinity Zinc Binding Site. *J. Biol. Chem.* 278 (46), 45325–45332. doi:10.1074/jbc.m307818200

Considine, M. J., and Foyer, C. H. (2020). Oxygen and Reactive Oxygen Species-dependent Regulation of Plant Growth and Development. *Plant Physiol.* 186, 1. doi:10.1093/plphys/kiaa077

Dai, L., Zhao, T., Bisteau, X., Sun, W., Prabhu, N., Lim, Y. T., et al. (2018a). Modulation of Protein-Interaction States through the Cell Cycle. *Cell* 173, 1481–1494. doi:10.1016/j.cell.2018.03.065

Dai, X., Zhuang, Z., and Zhao, P. X. (2018b). psRNATarget: a Plant Small RNA Target Analysis Server (2017 Release). *Nucleic Acids Res.* 46, W49–W54. doi:10.1093/nar/gky316

Dakouri, A., McCallum, B. D., Radovanovic, N., and Cloutier, S. (2013). Molecular and Phenotypic Characterization of Seedling and Adult Plant Leaf Rust Resistance in a World Wheat Collection. *Mol. Breed.* 32 (3), 663–677. doi:10.1007/s11032-013-9899-8

de Torres Zabala, M., Littlejohn, G., Jayaraman, S., Studholme, D., Bailey, T., Lawson, T., et al. (2015). Chloroplasts Play a central Role in Plant Defence and Are Targeted by Pathogen Effectors. *Nat. Plants* 1, 1–10. doi:10.1038/nplants.2015.74

Dhariwal, R., Vyas, S., Bhaganagare, G. R., Jha, S. K., Khurana, J. P., Tyagi, A. K., et al. (2011). Analysis of Differentially Expressed Genes in Leaf Rust Infected Bread Wheat Involving Seedling Resistance Gene Lr28. *Funct. Plant Biol.* 38, 479–492. doi:10.1071/fp10246

Dhindsa, R. S., Plumb-Dhindsa, P., and Thorpe, T. A. (1981). Leaf Senescence: Correlated with Increased Levels of Membrane Permeability and Lipid Peroxidation, and Decreased Levels of Superoxide Dismutase and Catalase. *J. Exp. Bot.* 32, 93–101. doi:10.1093/jxb/32.1.93

Doke, N. (1983). Involvement of Superoxide Anion Generation in the Hypersensitive Response of Potato Tuber Tissues to Infection with an Incompatible Race of Phytophthora Infestans and to the Hyphal wall Components. *Physiol. Plant Pathol.* 23, 345–357. doi:10.1016/0048-4059(83)90019-x

Felsenstein, J. (1985). Confidence Limits on Phylogenies: an Approach Using the Bootstrap. *Evolution* 39, 783–791. doi:10.1111/j.1558-5646.1985.tb00420.x

Fernández-Fernández, Á. D., and Corpas, F. J. (2016). In Silico analysis of *Arabidopsis thaliana* Peroxisomal 6-phosphogluconate Dehydrogenase. *Scientifica (Cairo)* 2016, 3482760. doi:10.1155/2016/3482760

Fernández-Trijueque, J., Serrato, A.-J., and Sahrawy, M. (2019). Proteomic Analyses of Thioredoxins F and M *Arabidopsis thaliana* Mutants Indicate Specific Functions for These Proteins in Plants. *Antioxidants* 8, 54. doi:10.3390/antiox8030054

Fu, Z. Q., Yan, S., Saleh, A., Wang, W., Ruble, J., Oka, N., et al. (2012). NPR3 and NPR4 Are Receptors for the Immune Signal Salicylic Acid in Plants. *Nature* 486, 228–232. doi:10.1038/nature11162

Gamage, D. G., Gunaratne, A., Periyannan, G. R., and Russell, T. G. (2019). Applicability of Instability index for *In Vitro* Protein Stability Prediction. *Ppl* 26, 339–347. doi:10.2174/0929866526666190228144219

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. *Proteomics Protoc. Handb.* 112, 571–607. doi:10.1385/1-59259-890-0:571

Gelhaye, E., Rouhier, N., Gérard, J., Jolivet, Y., Gualberto, J., Navrot, N., et al. (2004). A Specific Form of Thioredoxin H Occurs in Plant Mitochondria and Regulates the Alternative Oxidase. *Proc. Natl. Acad. Sci.* 101, 14545–14550. doi:10.1073/pnas.0405282101

He, L., He, T., Farrar, S., Ji, L., Liu, T., and Ma, X. (2017). Antioxidants Maintain Cellular Redox Homeostasis by Elimination of Reactive Oxygen Species. *Cell. Physiol. Biochem.* 44, 532–553. doi:10.1159/000485089

Heil, M., Ibarra-Laclette, E., Adame-Álvarez, R. M., Martínez, O., Ramirez-Chávez, E., Molina-Torres, J., et al. (2012). How Plants Sense Wounds: Damaged-self Recognition Is Based on Plant-Derived Elicitors and Induces Octadecanoid Signaling. *PLoS One* 7, e30537. doi:10.1371/journal.pone.0030537

Heinekamp, T., Strathmann, A., Kuhlmann, M., Froissard, M., Müller, A., Perrot-Rechenmann, C., et al. (2004). The Tobacco bZIP Transcription Factor BZI-1 Binds theGH3promoterin Vivoand Modulates Auxin-Induced Transcription. *Plant J.* 38, 298–309. doi:10.1111/j.1365-313x.2004.02043.x

Holmgren, A. (1989). Thioredoxin and Glutaredoxin Systems. *J. Biol. Chem.* 264, 13963–13966. doi:10.1016/s0021-9258(18)71625-6

Holmgren, A. (1968). Thioredoxin. 6. The Amino Acid Sequence of the Protein from *Escherichia coli* B. *Eur. J. Biochem.* 6, 475–484. doi:10.1111/j.1432-1033.1968.tb00470.x

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

Huerta-Espino, J., Singh, R. P., Germán, S., McCallum, B. D., Park, R. F., Chen, W. Q., et al. (2011). Global Status of Wheat Leaf Rust Caused by Puccinia Triticina. *Euphytica* 179, 143–160. doi:10.1007/s10681-011-0361-x

Ikai, A. (1980). Thermostability and Aliphatic index of Globular Proteins. *J. Biochem.* 88, 1895–1898. doi:10.1093/oxfordjournals.jbchem.a133104

Issakidis-Bourguet, E., Mouaheb, N., Meyer, Y., and Miginiac-Maslow, M. (2008). Heterologous Complementation of Yeast Reveals a New Putative Function for Chloroplast M-type Thioredoxin. *Plant J.* 25, 127–135. doi:10.1046/j.0960-7412.2000.00943.x

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly Regulated Genes Are Intron Poor. *Trends Genetics* 24 (8), 375–378. doi:10.1016/j.tig.2008.05.006

Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative Splicing and Nonsense-Mediated Decay Modulate Expression of Important Regulatory Genes in *Arabidopsis*. *Nucleic Acids Res.* 40, 2454–2469. doi:10.1093/nar/gkr932

Kato, Y., Miura, E., Matsushima, R., and Sakamoto, W. (2007). White Leaf Sectors Inyellow variegated2Are Formed by Viable Cells with Undifferentiated Plastids. *Plant Physiol.* 144, 952–960. doi:10.1104/pp.107.099002

Kumar, D., Yusuf, M. A., Singh, P., Sardar, M., and Sarin, N. B. (2014). Histochemical Detection of Superoxide and H2O2 Accumulation in Brassica Juncea Seedlings. *Bio-protocol* 4, e1108. doi:10.21769/bioprotoc.1108

Kumar, S., Bhardwaj, S. C., Gangwar, O. P., Sharma, A., Qureshi, N., Kumaran, V. V., et al. (2021). Lr80: A New and Widely Effective Source of Leaf Rust Resistance of Wheat for Enhancing Diversity of Resistance Among Modern Cultivars. *Theor. Appl. Genet.* 134 (3), 849–858. doi:10.1007/s00122-020-03735-5

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054

Lamers, J., Van Der Meer, T., and Testerink, C. (2020). How Plants Sense and Respond to Stressful Environments. *Plant Physiol.* 182, 1624–1635. doi:10.1104/pp.19.01464

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* 26, 283–291. doi:10.1107/s0021889892009944

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., 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, Y.-B., Han, L.-B., Wang, H.-Y., Zhang, J., Sun, S.-T., Feng, D.-Q., et al. (2016). The Thioredoxin GbNRX1 Plays a Crucial Role in Homeostasis of Apoplastic Reactive Oxygen Species in Response to Verticillium dahliae Infection in Cotton. *Plant Physiol.* 170, 2392–2406. doi:10.1104/pp.15.01930

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *methods* 25, 402–408. doi:10.1006/meth.2001.1262

Lu, J., and Holmgren, A. (2014). The Thioredoxin Antioxidant System. *Free Radic. Biol. Med.* 66, 75–87. doi:10.1016/j.freeradbiomed.2013.07.036

Mata-Pérez, C., and Spoel, S. H. (2019). Thioredoxin-mediated Redox Signalling in Plant Immunity. *Plant Sci.* 279, 27–33. doi:10.1016/j.plantsci.2018.05.001

Meyer, Y., Siala, W., Bashandy, T., Riondet, C., Vignols, F., and Reichheld, J. P. (2008). Glutaredoxins and Thioredoxins in Plants. *Biochim. Biophys. Acta (Bba) - Mol. Cell Res.* 1783, 589–600. doi:10.1016/j.bbamcr.2007.10.017

Mittler, R. (2002). Oxidative Stress, Antioxidants and Stress Tolerance. *Trends Plant Sci.* 7, 405–410. doi:10.1016/s1360-1385(02)02312-9

Mittler, R., Vanderauwera, S., Gollery, M., and Van Breusegem, F. (2004). Reactive Oxygen Gene Network of Plants. *Trends Plant Sci.* 9, 490–498. doi:10.1016/j.tplants.2004.08.009

Moreau, M., Tian, M., and Klessig, D. F. (2012). Salicylic Acid Binds NPR3 and NPR4 to Regulate NPR1-dependent Defense Responses. *Cell Res* 22, 1631–1633. doi:10.1038/cr.2012.100

Nakano, Y., and Asada, K. (1981). Hydrogen Peroxide Is Scavenged by Ascorbate-specific Peroxidase in Spinach Chloroplasts. *Plant Cell Physiol.* 22, 867–880.

Noctor, G., Reichheld, J.-P., and Foyer, C. H. (2018). ROS-related Redox Regulation and Signaling in Plants. *Semin. Cell Dev. Biol.* 80, 3–12. doi:10.1016/j.semcdb.2017.07.013

Nuruzzaman, M., Sharoni, A. M., Satoh, K., Al-Shammari, T., Shimizu, T., Sasaya, T., et al. (2012). The Thioredoxin Gene Family in Rice: Genome-wide Identification and Expression Profiling under Different Biotic and Abiotic Treatments. *Biochem. Biophysical Res. Commun.* 423, 417–423. doi:10.1016/j.bbrc.2012.05.142

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25, 1605–1612. doi:10.1002/jcc.20084

Prabhu, K. V., Gupta, S. K., Charpe, A., Koul, S., Cherukuri, D. P., Dhaliwal, H. S., et al. (2012). Molecular Markers Detect Redundancy and Miss-Identity in Genetic Stocks with Alien Leaf Rust Resistance Genes Lr32 and Lr28 in Bread Wheat. *J. Plant Biochem. Biotechnol.* 12, 123–129. doi:10.1007/bf03263172

Qiao, M., Sun, J., Liu, N., Sun, T., Liu, G., Han, S., et al. (2015). Changes of Nitric Oxide and its Relationship with H_2O_2 and Ca^{2+} in Defense Interactions between Wheat and *Puccinia Triticina*. *PLoS One* 10, e0132265. doi:10.1371/journal.pone.0132265

Rahikainen, M., Pascual, J., Alegre, S., Durian, G., and Kangasjärvi, S. (2016). PP2A Phosphatase as a Regulator of ROS Signaling in Plants. *Antioxidants* 5, 8. doi:10.3390/antiox5010008

Ramachandran, S. R., Mueth, N. A., Zheng, P., and Hulbert, S. H. (2020). Analysis of miRNAs in Two Wheat Cultivars Infected with *Puccinia Striiformis* F. Sp. Tritici. *Front. Plant Sci.* 10, 1574. doi:10.3389/fpls.2019.01574

Rao, M. V., Paliyath, G., Ormrod, D. P., Murr, D. P., and Watkins, C. B. (1997). Influence of Salicylic Acid on H2O2 Production, Oxidative Stress, and H_2O_2-Metabolizing Enzymes (Salicylic Acid-Mediated Oxidative Damage Requires H2O2). *Plant Physiol.* 115, 137–149. doi:10.1104/pp.115.1.137

Righetti, P. G. (2004). Determination of the Isoelectric point of Proteins by Capillary Isoelectric Focusing. *J. Chromatogr. A* 1037, 491–499. doi:10.1016/j.chroma.2003.11.025

Rouhier, N., Cerveau, D., Couturier, J., Reichheld, J.-P., and Rey, P. (2015). Involvement of Thiol-Based Mechanisms in Plant Development. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1850, 1479–1496. doi:10.1016/j.bbagen.2015.01.023

Sathee, L., Adavi, S. B., Jain, V., Pandey, R., Khetarpal, S., Meena, H. S., et al. (2018). Influence of Elevated CO2 on Kinetics and Expression of High Affinity Nitrate Transport Systems in Wheat. *Ind. J. Plant Physiol.* 23, 111–117. doi:10.1007/s40502-018-0355-y

Savojardo, C., Martelli, P. L., Fariselli, P., Profiti, G., and Casadio, R. (2018). BUSCA: an Integrative Web Server to Predict Subcellular Localization of Proteins. *Nucleic Acids Res.* 46, W459–W466. doi:10.1093/nar/gky320

Schenk, H., Klein, M., Erdbrügger, W., Dröge, W., and Schulze-Osthoff, K. (1994). Distinct Effects of Thioredoxin and Antioxidants on the Activation of Transcription Factors NF-Kappa B and AP-1. *Proc. Natl. Acad. Sci.* 91, 1672–1676. doi:10.1073/pnas.91.5.1672

Schmidt, R., Kunkowska, A. B., and Schippers, J. H. M. (2016). Role of Reactive Oxygen Species during Cell Expansion in Leaves. *Plant Physiol.* 172, 2098–2106. doi:10.1104/pp.16.00426

Sheshadri, S. A., Nishanth, M. J., and Simon, B. (2016). Stress-mediated Cis-Element Transcription Factor Interactions Interconnecting Primary and Specialized Metabolism in Planta. *Front. Plant Sci.* 7, 1725. doi:10.3389/fpls.2016.01725

Shi, B., Zhao, X., Li, M., Dong, Z., Yang, Q., Wang, Y., et al. (2021). Wheat Thioredoxin (*TaTrxh*1) Associates with RD19-like Cysteine Protease TaCP1 to Defend against Stripe Rust Fungus through Modulation of Programmed Cell Death. *Mpmi* 34, 426–438. doi:10.1094/mpmi-11-20-0304-r

Sun, L., Ren, H., Liu, R., Li, B., Wu, T., Sun, F., et al. (2010). An H-type Thioredoxin Functions in Tobacco Defense Responses to Two Species of Viruses and an Abiotic Oxidative Stress. *Mpmi* 23, 1470–1485. doi:10.1094/mpmi-01-10-0029

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Thordal-Christensen, H., Zhang, Z., Wei, Y., and Collinge, D. B. (1997). Subcellular Localization of H2O2 in Plants. H2O2 Accumulation in Papillae and Hypersensitive Response during the Barley—Powdery Mildew Interaction. *Plant J.* 11 (6), 1187–1194.

Tomar, S. M. S., Singh, S. K., Sivasamy, M., and Vinod (2014). Wheat Rusts in India: Resistance Breeding and Gene Deployment -A Review. *Ind. Jrnl. Gen. Plnt. Bree.* 74, 129–156. doi:10.5958/0975-6906.2014.00150.3

Torres, M. A., Jones, J. D. G., and Dangl, J. L. (2006). Reactive Oxygen Species Signaling in Response to Pathogens. *Plant Physiol.* 141, 373–378. doi:10.1104/pp.106.079467

Traverso, J. A., Vignols, F., and Chueca, A. (2007). Thioredoxin and Redox Control within the New Concept of Oxidative Signaling. *Plant Signaling Behav.* 2, 426–427. doi:10.4161/psb.2.5.4419

Tsukamoto, S., Morita, S., Hirano, E., Yokoi, H., Masumura, T., and Tanaka, K. (2005). A Novel Cis-Element that Is Responsive to Oxidative Stress Regulates Three Antioxidant Defense Genes in Rice. *Plant Physiol.* 137, 317–327. doi:10.1104/pp.104.045658

Vieira Dos Santos, C., and Rey, P. (2006). Plant Thioredoxins Are Key Actors in the Oxidative Stress Response. *Trends Plant Sci.* 11, 329–334. doi:10.1016/j.tplants.2006.05.005

Wang, X., Che, M. Z., Khalil, H. B., McCallum, B. D., Bakkeren, G., Rampitsch, C., et al. (2020). The Role of Reactive Oxygen Species in the Virulence of Wheat Leaf Rust Fungus Puccinia Triticina. *Environ. Microbiol.* 22 (7), 2956–2967. doi:10.1111/1462-2920.15063

Wang, Z., Ma, L.-Y., Cao, J., Li, Y.-L., Ding, L.-N., Zhu, K.-M., et al. (2019). Recent Advances in Mechanisms of Plant Defense to *Sclerotinia sclerotiorum*. *Front. Plant Sci.* 10, 1314. doi:10.3389/fpls.2019.01314

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427

Wong, J. H., Yano, H., Lee, Y.-M., Cho, M.-J., and Buchanan, B. B. (2002). Identification of Thioredoxin-Linked Proteins by Fluorescence Labeling Combined with Isoelectric Focusing/Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. *Methods Enzym* 347, 339–349. doi:10.1016/s0076-6879(02)47034-6

Yu, X., Zhang, W., Zhang, Y., Zhang, X., Lang, D., and Zhang, X. (2018). The Roles of Methyl Jasmonate to Stress in Plants. *Funct. Plant Biol.* 46, 197–212. doi:10.1071/FP18106

Zheng, Y., Chen, X., Wang, P., Sun, Y., Yue, C., and Ye, N. (2020). Genome-wide and Expression Pattern Analysis of JAZ Family Involved in Stress Responses and Postharvest Processing Treatments in *Camellia Sinensis*. *Sci. Rep.* 10, 2792. doi:10.1038/s41598-020-59675-z

Zurbriggen, M. D., Carrillo, N., and Hajirezaei, M.-R. (2010). ROS Signaling in the Hypersensitive Response. *Plant Signaling Behav.* 5 (4), 393–396. doi:10.4161/psb.5.4.10793

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 Bhurta, Hurali, Tyagi, Sathee, Adavi B, Singh, Mallick, Chinnusamy, Vinod and Jha. 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

MINI REVIEW
published: 31 March 2022
doi: 10.3389/fgene.2022.816057

An Update on Resistance Genes and Their Use in the Development of Leaf Rust Resistant Cultivars in Wheat

Kuldeep Kumar[1†], *Irfat Jan*[1,2†], *Gautam Saripalli*[1,3], *P. K. Sharma*[1], *Reyazul Rouf Mir*[2], *H. S. Balyan*[1] *and P. K. Gupta*[1*]

[1]Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, India, [2]Division of Genetics and Plant Breeding, Faculty of Agriculture, Sher-e-Kashmir University of Agricultural Sciences and Technology, Wadura, India, [3]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD, United States

Wheat is one of the most important cereal crops in the world. The production and productivity of wheat is adversely affected by several diseases including leaf rust, which can cause yield losses, sometimes approaching >50%. In the present mini-review, we provide updated information on (i) all Lr genes including those derived from alien sources and 14 other novel resistance genes; (ii) a list of QTLs identified using interval mapping and MTAs identified using GWAS (particular those reported recently i.e., after 2018) and their association with known Lr genes; (iii) introgression/pyramiding of individual Lr genes in commercial/prominent cultivars from 18 different countries including India. Challenges and future perspectives of breeding for leaf rust resistance are also provided at the end of this mini-review. We believe that the information in this review will prove useful for wheat geneticists/breeders, not only in the development of leaf rust-resistant wheat cultivars, but also in the study of molecular mechanism of leaf rust resistance in wheat.

Keywords: bread wheat, leaf rust, genes, QTLs, markers, molecular breeding

OPEN ACCESS

Edited by:
Karthikeyan Adhimoolam, Jeju National University, South Korea

Reviewed by:
Suraj Sapkota, Crop Genetics and Breeding Research (USDA-ARS), United States
Prashant Vikram, International Center for Biosaline Agriculture (ICBA), United Arab Emirates

***Correspondence:**
P. K. Gupta
pkgupta36@gmail.com
pkgupta@ccsuniversity.ac.in

†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: *16 November 2021*
Accepted: *28 February 2022*
Published: *31 March 2022*

Citation:
Kumar K, Jan I, Saripalli G, Sharma PK, Mir RR, Balyan HS and Gupta PK (2022) An Update on Resistance Genes and Their Use in the Development of Leaf Rust Resistant Cultivars in Wheat. Front. Genet. 13:816057. doi: 10.3389/fgene.2022.816057

INTRODUCTION

Leaf rust caused by the fungal pathogen *Puccinia triticina* Eriks. & E. Henn is an important disease in wheat, which causes significant yield losses, sometimes approaching up to >50% (Riaz and Wong 2017). The study of the genetic basis of this disease and breeding for leaf rust resistance in wheat has been an important area of research (Dyck 1993; Kolmer and Liu, 2002; Oelke and Kolmer 2005; Datta et al., 2008; Rosa et al., 2016; Aoun et al., 2017). Each individual Lr gene apparently shows resistance against a specific race of *P. triticina* (*Pt*), which must carry the corresponding avirulence (Avr) gene, such that a specific Lr gene in the host and the corresponding specific Avr gene in the pathogen always follow a "gene-for-gene" relationship (Flor 1946). The pathogen Pt keeps on developing new virulent races through mutations or recombination involving Avr genes; new strains may also migrate from other geographical areas, and may carry one or more new Avr genes for which the corresponding R gene may be absent in the host (Samborski 1985; Bolton et al., 2008). Therefore, the host resistance breaks down and is short-lived. It is thus obvious that a majority of race specific Lr genes individually do not provide durable resistance (Johnson 1984).

Lr genes provide either seedling resistance (SR), also described as all stage resistance (ASR), or adult plant resistance (APR genes), the latter expressed only at the adult plant stage, particularly after booting. It is also known that ASR genes provide resistance, which breaks down within a few years, while APR provides long-term durable resistance (Ellis et al., 2014). Some of the APR genes like *Lr34* and *Lr67* have also been cloned and were found to be complex loci including *Lr34*/*Sr57*/*Yr18*/*Pm38*

TABLE 1 | Details of leaf rust (Lr) resistant genes including novel Lr genes identified in bread wheat.

Gene	Chr	Marker	References
Lr1	5DL	psr567	Sylvie Cloutier et al. (2007)
Lr2a	2DS	Xwmc453 - XwPt0330	Tsilo et al. (2014)
Lr3	6BL	Xmwg798	Sacco et al. (1998)
Lr3a	6BL	UBC 840	Khan et al. (2005)
Lr9	6BL	SCS5	Gupta et al. (2005)
Lr10	1AS	Lrk10D1	Schachermayr et al. (1997)
Lr11	2DS	SCAR32/35	Darino et al. (2015)
Lr12	4BL	Xgwm251 - Xgwm149	Singh and Bowden (2011)
Lr13	2BS	Xbarc55-2B	Seyfarth et al. (1998); Seyfarth et al. (1999)
Lr14a	7BL	wPt-4038-HRM	Terracciano et al. (2013)
Lr15	2DS	Xgwm4562 - Xgwm102	Dholakia et al. (2013)
Lr16	2BS	Xwmc764, Xgwm210, and Xwmc661	McCartney et al. (2005)
Lr17	2AS	Xgwm614 - gwm407	Bremenkamp-Barrett et al. (2008)
Lr18	5BL	IWB41960 - gwm547	Carpenter et al. (2017)
Lr19	7DL	SCS265 and SCS253	Gupta et al. (2006)
Lr20	7AL	STS638	Neu et al. (2002)
Lr21	1DS	Lr21_GQ5044819_2175_G/A KASPar assay and Lr21_GQ5044819_3146_C/T KASPar assay	Neelam et al. (2013)
Lr22a	2DS	gwm296	Hiebert et al. (2007)
Lr23	2BS	Xtam72	Nelson et al. (1997)
Lr24	3DL	SCS1302	Prabhu et al. (2004)
Lr25	4BS	Xgwm251	Singh et al. (2012)
Lr26	1BL	P6M12-P	Mago et al. (2002), Zhou et al. (2014)
Lr27	3BS	cdo460	Nelson et al. (1997)
Lr28	4AL	SCS421	Naik et al. (1998), Cherukuri et al. (2005)
Lr29	7DS	ubc219	Procunier et al. (1995)
Lr30	4AL	IWA4359 - IWA2585	Aoun et al. (2019)
Lr31	4BL	XksuG10	Nelson et al. (1997)
Lr32	3DS	Xbcd1278	Autrique et al. (1995)
Lr34	7DS	csLV34	Lagudah et al. (2006, 2009)
Lr35	2BL	Xbcd260	Seyfarth et al. (1999)
Lr36	6BS	cfd1, gwm508	Dadkhodaie et al. (2011)
Lr37	2AS	VENTRIUP/LN2	Helguera et al. (2003)
Lr38	6DL	wmc773 - barc273	Mebrate et al. (2008)
Lr39	2DS	Xgwm210	Raupp et al. (2001)
Lr41	2DS	Xbarc124	Sun et al. (2009)
Lr42	1DS	Xwmc432	Sun et al. (2010)
Lr45	2AS	cfd168, $G372_{94}$ and $G372_{185}$	Naik et al. (2015)
Lr46	1BL	XSTS1BL9	Mateos-Hernandez et al. (2006)
Lr47	7AL	PS10	Helguera et al. (2000)
Lr48	2BL	Xksm58 - Xstm773-2	Bansal et al. (2008)
Lr49	4BL	Xbarc163 - Xwmc349	Bansal et al. (2008)
Lr50	2BL	Xgwm382	Brown-Guedira et al. (2003)
Lr51	1BL	e XAga7	Helguera et al. (2005)
Lr52	5BS	Xwmc149, Xtxw200	Tar et al. (2008)
Lr53	6BS	cfd1, gwm508	Dadkhodaie et al. (2011)
Lr57	5DS	Lr57/Yr40-MAS-CAPS16	Kuraparthy et al. (2009)
Lr58	2BL	Xcfd50	Kuraparthy et al. (2007)
Lr59	6BS	IWA1495, IWA6704	Poudel (2015)
Lr60	1DS	Xbarc149	Hiebert et al. (2008)
Lr61	6BS	P81/M70	Herrera-Foessel et al. (2008)
Lr62	6AS	Xgwm334	Marais et al. (2009)
Lr63	3AS	barc 57 and barc 321	Kolmer et al. (2010)
Lr64	6AS	K-IWB59855	Kolmer (2019)
Lr65	2AS	barc124, barc212, gwm614	Mohler et al. (2012)
Lr66	3AS	S13-R16	Marais et al. (2010)
Lr67	4DL	cfd71	Hiebert et al. (2010)
Lr68	7BL	Psy1-1 - gwm146	Herrera-Foessel et al. (2012)
Lr70	5DS	barc130	Hiebert et al. (2014)
Lr71	1BS	gwm18 - barc187	Singh et al. (2012)
Lr72	7BS	wmc606	Herrera-Foessel et al. (2014)
Lr73	2BL	wPt8760 - wPt-8235	Park et al. (2014)
Lr74	3BS	Xcfb5006 - Xgwm533	Li et al. (2017)
Lr75	1BS	gwm604 - swm271	Singla et al. (2017)

(Continued on following page)

TABLE 1 | (*Continued*) Details of leaf rust (Lr) resistant genes including novel Lr genes identified in bread wheat.

Gene	Chr	Marker	References
Lr76	5DL	*Lr57/Yr40*-MAS-CAPS16	Kuraparthy et al. (2009)
Lr77	3BL	IWB10344	Kolmer et al. (2018a)
Lr78	5DS	IWA6289	Kolmer et al. (2018b)
Lr79	3BL	sun786 - sun770	Qureshi et al. (2018)
Lr80	2DS	KASP_17425, KASP_17148	Kumar et al. (2021)
LrX	1DS	K-IWB38437	Kolmer et al. (2019)
LrTs276-2	1DS	Xcfd15 - Xcfd61	Dinkar et al. (2020)
Lr2K38	1AL	IWB20487	Sapkota et al. (2020)
Novel Lr genes			
LrZH84	1BL	Xgwm582 - Xbarc8	Zhao et al. (2008)
LrBi16	7BL	Xcfa2257	Zhang et al. (2011)
LrSV1	2DS	Xgwm261	Ingala et al. (2012)
LrSV2	3BL	Xgwm389, Xgwm533, Xgwm493	Ingala et al. (2012)
LrG6	2BL	Xgwm526	Ingala et al. (2012)
LrFun	7BL	Xgwm344	Xing et al. (2014)
LrNJ97	1BL	Xwmc317 - Xbarc159	Zhao et al. (2013)
Lr5R	3DL	Xbarc71 - OPJ-09	Wang et al. (2014)
LrAc	5DS	Ta5DS_2737450	Toor et al. (2016)
LrZH22	2BS	Xgwm374	Wang et al. (2016)
LrE1	7BL	Xgwm131	
LrP	5DS	BS00163889	Narang et al. (2019)
Lr.ace-4A	4AS	IWA232, IWA1793	Aoun et al. (2019)
LrM	2AS	SNP_AX948171722AS	Rani et al. (2020)

and *Lr67*/*Sr55*/*Yr46*/*Pm46* (Lagudah et al., 2006; Moore et al., 2015). These gene complexes confer durable resistance not only against leaf rust, but also against stripe rust, stem rust, powdery mildew, and barley yellow dwarf virus (Singh and Rajaram, 1993). The use of APR genes along with 4–5 Lr genes is a strategy that provides durable resistance.

A number of reviews on leaf rust resistance in wheat have already been published (Kolmer 1996; Kolmer 2013; McCallum et al., 2016; Pinto da Silva et al., 2018; Dinh et al., 2020; Figlan et al., 2020; Ghimire et al., 2020; Prasad et al., 2020). Information about QTLs for leaf rust resistance has also been recently reviewed (Pinto da Silva et al., 2018). However, considerable literature has appeared during the last 3–4 years, where many more QTLs and as many as 600 new MTAs have been added thus warranting a fresh look on the subject, hence this minireview.

According to some recent reports, currently more than 80 Lr genes and 14 other genes for leaf rust resistance are known in wheat (McIntosh et al., 2017, 2020). The above 14 genes have not been assigned a new number in Lr series, perhaps because these genes have not been subjected to test of allelism with the known Lr genes to ascertain their novelty. Since literature on Lr genes keep on appearing on a regular basis, any review published soon becomes out of date thus creating the need for a fresh review. The present mini-review caters to this need and provides an updated information on all Lr genes and other genes including genes derived from alien species. The mini-review includes information about chromosomal location of all these genes (including 14 other resistance genes, which could not be assigned to any of the known Lr genes; modified names were used for these 14 genes based on the cultivar in which they were identified). We also provide information about the wild relatives of wheat as a source of Lr genes and the molecular markers associated with most of these genes (wherever known). Information about cloning and characterisation of Lr genes has also been included, wherever available. The wheat varieties carrying different Lr genes developed in 18 different countries are also listed.

LR GENES/NOVEL LR GENES CATALOGUED SO FAR

More than 80 Lr genes (~50% derived from alien species) are already known to be distributed on all the 21 wheat chromosomes, with majority of genes located on the short arms of individual chromosomes (**Table 1**; **Supplementary Tables S1, S2**). Most of the Lr genes are located on the B sub-genome, relative to those located on either A sub-genome or D sub-genome. Maximum number of ten Lr genes (including two novel genes *LrZH22* and *LrG6*) are located on chromosome 2B. At least two of these genes, namely *Lr18* and *LrZH22*, are known to be temperature sensitive; *Lr18* exhibits resistance at 15–18°C, ineffective at >18°C (Carpenter et al., 2017). The other gene *LrZH22* confers resistance at higher temperatures (22–25°C; Wang et al., 2016). Lr genes conferring APR include the following: *Lr34, Lr46, Lr67, Lr68, Lr74, Lr75, Lr77* and *Lr78*. Information on QTLs/MTAs was also included in an earlier review (Pinto da Silva et al., 2018) and has been compiled by us also in this mini-review (**Supplementary Tables S1, S2**). A set of 14 novel resistance genes (including three genes from alien species) are known, which differ from other available Lr genes, since they show seedling reaction pattern, which was different from reaction patterns known for different Lr genes studied so

far. These 14 genes along with associated markers are also listed in **Table 1**. These genes were mapped on 10 out of the 21 wheat chromosomes with maximum number of these genes available on B sub-genome (8) followed by sub-genome D (4) and sub-genome A (2).

QTLs/MTAs LINKED TO LR GENES

In recent years, a number of newer approaches (based on DNA markers) led to the discovery of a large number of QTLs/QRLs and marker-trait associations (MTAs) for resistance against plant diseases including leaf rust. Qualitative resistance provided by Lr genes is generally compromised within a short period of time (Goyeau et al., 2006; Goyeau and Lannou, 2011), but quantitative disease resistance (QDR) provides effective and durable resistance involving major reduction in the level of disease (Mundt et al., 2002; Parlevliet, 2002; Stuthman et al., 2007). The QDR generally depends upon the presence of few major QTLs/genes and a fairly large number of minor QTLs (Ballini et al., 2008; Clair, 2010). Only a solitary example, where QDR for leaf rust resistance has been utilized is the French wheat cultivar Apache, which carried sustained resistance against leaf rust for a fairly long time (Papaïx et al., 2011). The availability of large number of QTLs/MTAs in wheat, as demonstrated in several studies, suggests that QDR against leaf rust is common in this crop, but has not been fully exploited.

A large number of QTLs, mostly associated with Lr genes were listed in some earlier reviews. For instance, in one report, 250 QTLs (reported till 2017) were listed, which were reported in 70 different studies (Pinto da silva et al., 2018). In second study, 35 meta-QTLs (MQTLs) were listed, which were identified using QTLs reported in several studies (Soriano and Royo, 2015). During the last 4 years (after 2017), additional 103 QTLs were reported in 18 studies; 29 of these QTLs were shown to be associated with Lr genes and Lr/Yr genes (**Supplementary Table S1**).

In addition to QTLs, ~200 MTAs based on GWAS involving seven association panels (AM) were also reported earlier (Pinto da silva et al., 2018). As mentioned earlier, after publication of this review, ~600 MTAs were reported in eight genome-wide association studies (GWAS); 42 of these MTAs were found to be linked to Lr genes (**Supplementary Table S2**). The maximum number of QTLs and Lr genes for leaf rust resistance are present in the B sub-genome. The PVE of the individual QTLs ranged from 4.63% to 75.3%; 29 of these QTLs had a PVE >20% suggesting their utility in MAS for breeding (**Supplementary Table S1**).

WILD RELATIVES AS A SOURCE OF LR GENES

At least 50% of Lr genes are derived from wild relatives (alien resources). One of the important alien sources from Fertile Crescent region is Sharon goatgrass (*Aegilops sharonensis*), which is a very valuable source of unique genes/QTLs for resistance to several wheat diseases including leaf rust (for reviews see Ghimire et al., 2020; Figlan et al., 2020). Following other important wild relatives of wheat have also been identified as sources of Lr genes/QTLs: (i) Tausch's goatgrass (*Ae. tauschii*) (*Lr21*, *Lr22a*, and *Lr39*), (ii) wheatgrass (*Thinopyrum ponticum*) (*Lr24*), (iii) *Ae. geniculate* (*Lr57*), (iv) *Ae. ventricosa* (*Lr37*/*Yr17*), (v) *Ae. umbellulata* (*Lr9*), (vi) *Thinopyrum elongatum* Zhuk. (*Lr 19*), (vii) *Agropyron elongatum* (*Lr24*), (viii) *Secale cereale* L. (*Lr26*), (ix) *Ae. peregrina* (*Lr59*), (x) *Ae. kotschyi (Lr54)*, (xi) *Ae. sharonensis (Lr56)*, (xii) *Ae. triuncialis (Lr58)*, and (xiii) *Ae. neglecta (Lr62)*; however this list is not exclusive (McIntosh, 1975; Autrique et al., 1995; McIntosh et al., 1995; Marais and Botes, 2003; Kumar et al., 2022).

MAS FOR PRE-BREEDING

There are ~700 cultivars/varieties from 18 different countries (including India), each cultivar carrying one to six resistance genes for leaf rust including both ASR and APR genes (the details of varieties and their country of origin, are available in **Supplementary Tables S2, S4**). Two different approaches (including conventional breeding and marker assisted breeding, including pre-breeding) are available for developing resistant cultivars (**Figure 1**). Since markers associated with each of a number of Lr genes and QTLs including MTAs are available, MAS has become routine for supplementing conventional breeding (**Supplementary Table S5**). These markers are particularly useful for pyramiding of resistance genes, since introgression of additional resistance genes in the presence of existing resistance genes using phenotypic selection is rather difficult. There are at least a dozen examples (seven from India involving PBW343 and HD2329), where associated markers have been used to supplement conventional breeding including pre-breeding. A number of wheat varieties belonging to hard red winter or soft red winter wheats from United States were also developed using MAS (USDA website; https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1124692/1/Doc188-online-Sandra-Brammer.pdf. Using MAS, up to 10 Lr genes could be pyramided into the same wheat cultivar.

CONCLUSION AND FUTURE PERSPECTIVES

The present mini-review is yet another effort to provide a summary of updated published literature on resistance against leaf rust in wheat, including known R genes (~80 Lr genes and 14 novel genes) (**Supplementary Table S1**), known QTLs/MTAs (**Supplementary Tables S1, S2**) and details of varieties containing one or more of these important leaf rust genes/QTLs/MTAs (**Supplementary Tables S3, S4**). Some details about the use of MAS for introgression of Lr genes into wheat varieties are also included (**Supplementary Table S5**).

One of the major challenges for wheat breeders is the regular development of new cultivars or improvement of old cultivars using new resistance genes, since new virulence pathotypes and

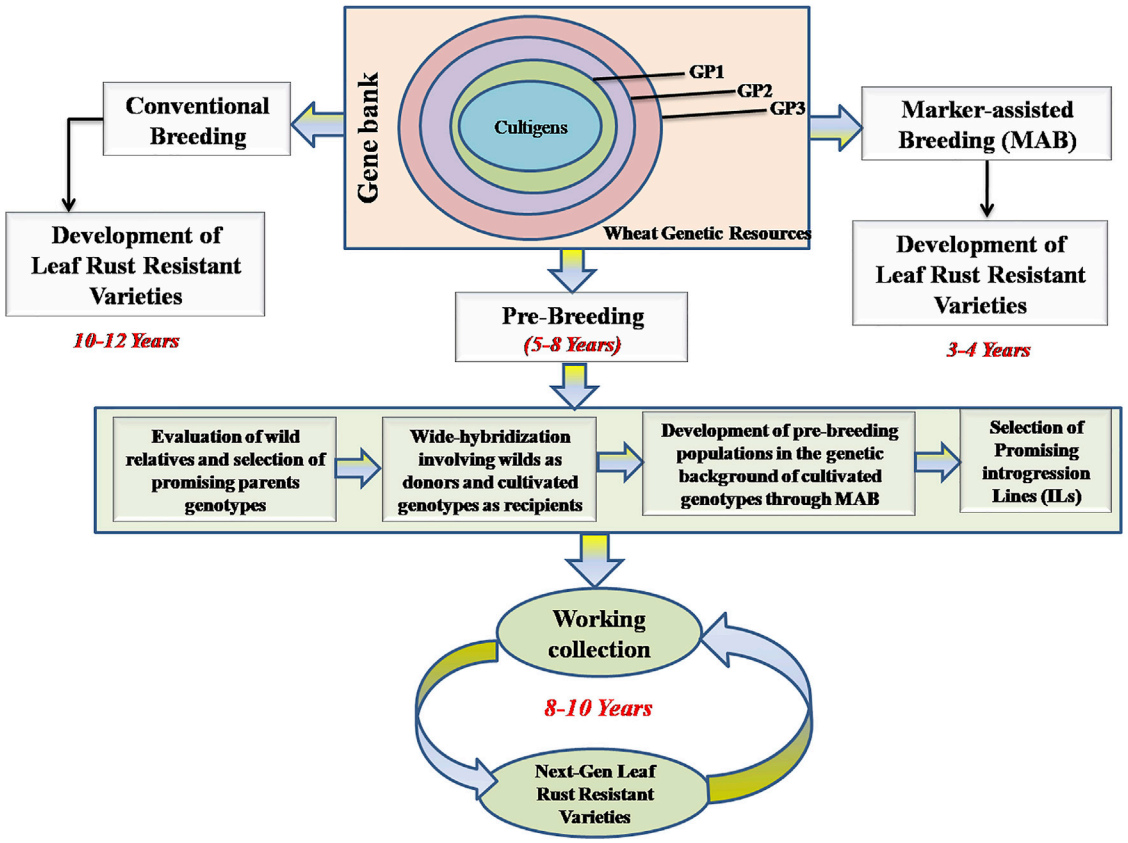


FIGURE 1 | Various pre—breeding steps involved in use of wild relatives in the development of leaf rust resistant wheat varieties. The panels show how wheat genetic resources including wild relatives belonging to primary (GP1), secondary (GP2) and tertiary (GP3) gene pools conserved in different gene banks can be used in pre-breeding programs. The panels also shows the advantages of use of marker—assisted breeding (only 3–4 years in developing new cultivars) over conventional breeding (taking 10–12 years in cultivar development).

races keep on appearing (Figlan et al., 2020; Ghimire et al., 2020). Therefore, continuous rigorous efforts are needed to locate sources for novel genes/QTLs to overcome new emerging races of pathogen and gain long-term resistance in the field. There are several other areas, which need attention and will certainly be the subject for future research. These will be briefly discussed as the future perspectives.

Although most R genes encode NLR proteins (with NBS-LRR domain), there are several other mechanisms involved as shown in a recent review, where work done during last 25 years involving >300 cloned R genes is reviewed. At least 60% of these R genes were shown to encode NLR proteins, the remaining 40% encoding RLKs/RLPs (Kourelis and van der Hoorn, 2018). Based on the study of these cloned R genes and the corresponding Avr genes of the pathogens, nine different mechanisms for the function of R genes have also been identified and summarised (Kourelis and van der Hoorn, 2018). However, the resistance mechanism of reported Lr genes is not clear and therefore can be a subject for future research.

The most common product of R genes, the NLRs have recently been shown to function through an assembly of a high-resolution structure called 'resistosome' which was first resolved in Arabidopsis and is responsible for providing resistance (Wang et al., 2019). Two additional examples of the high-resolution structures of interaction between NLRs and the effector molecules, through formation of resistosome also became available, thus suggesting that formation of the resistosome may be of wide occurrence (Ma et al., 2020; Martin et al., 2020). These three recent studies improved our understanding of the action of NLR at the molecular level. However, no Lr gene has been subjected to such studies involving formation of a resistosome. Therefore this is also an important area of future research.

Another important challenge in breeding for leaf rust resistance is the limited number of Lr genes that have been cloned (*Lr1, Lr10, Lr21, Lr22a, Lr34, Lr67*) and therefore cloning more genes is needed to understand the variation between such a large number of Lr genes and the mechanism used for their operation for providing resistance (Dinh et al., 2020; Prasad et al., 2020). According to some optimistic views, it will be possible in the next 15 years to clone most of the ~460 known wheat resistance genes and their corresponding effectors, making it possible to design suitable strategies for resistance

breeding in wheat (Wulff and Krattinger 2022). We, however feel that cloning of so many genes in a short period of 15 years may not be immediately possible. Therefore, closely linked markers may be used to identify which of the Lr genes encode NLR proteins and which other proteins may be encoded by other Lr genes. Bioinformatics may be used for this purpose and the results of this exercise may then be verified using suitable experiments.

Genomics of the pathogen is another important area, since genomes of a number of races of the pathogen have already been sequenced (Kiran et al., 2016; Wu et al., 2020; Fellers et al., 2021). This should facilitate use of bioinformatics for identification of effectors, using knowledge about conserved domains that have been discovered to be present in effector molecules. The genome sequences of different races of Pt have been worked out and many more genomes from the pathogen will also allow us to know the pangenome of Pt, which includes core genome, dispensible genome and unique genome. This knowledge will also allow to identify effectors and in planning suitable strategies for wheat breeding involving resistance against leaf rust.

It may also be necessary to study the effect of environment on expression of many resistance genes in the host since expression of genes has been found to vary with changing temperature (Figlan et al., 2020). This will involve study of the mode of action of resistance genes in the host, their interactions with other host genes, interactions with *Avr* gene while providing stable and durable resistance across environments. The recent advances in genomics tools and techniques including whole genome sequencing, genome annotation and high-throughput genomics tools like pathogenomics, gene cloning, genome editing are expected to offer deeper insights into host-pathogen interactions. This should eventually help in achieving durable rust resistance (Dinh et al., 2020). Molecular biology tools including HIGS/VIGS have also become very important for understanding and analyzing different facets of host and pathogen biology that includes secretome analysis, transcriptional profiling, putative virulence gene identification, structural gene annotation, and alternative transcript splicing. Another important area of future research is identification of vir genes, and effectors, which together make the subject of effectoromics and effector based breeding. This will allow the use of knowledge about effectors to screen the germplasm for resistance.

Epigenomics is another area, which has started attracting the attention of wheat geneticists. This will allow us to understand the role of DNA methylation, histone modifications, noncoding RNAs (e.g., miRNAs, lncRNAs) and chromatin states, thus further resolving the mechanism of resistance at the molecular level (Saripalli et al., 2020a; Saripalli et al., 2020b; Jain et al., 2020; Prasad et al., 2020).

AUTHOR CONTRIBUTIONS

PG, HB, and PS conceived and outlined the review. KK, IJ, and GS collected the literature and wrote the first draft of the review. PG, HB, PS, and RM edited and finalized the manuscript with the help of KK, IJ, and GS.

FUNDING

Department of Biotechnology (DBT), Govt. of India provided funds in the form of a research project (Award number: BT/PR21024/AGIII/103/925/2016). No grant has been received for publishing this review.

ACKNOWLEDGMENTS

Thanks are due to the Department of Biotechnology (DBT), Govt of India for providing funds in the form of a research project (Award number: BT/PR21024/AGIII/103/925/2016) and fellowships to KK and IJ for undertaking the research work related to this review, to National Agricultural Science Fund (NASF)-Indian Council of Agricultural Research (ICAR) for providing fellowship to GS and to Indian National Science Academy (INSA) for the award of position of INSA Honorary Scientist to HB.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.816057/full#supplementary-material

REFERENCES

Aoun, M., Kolmer, J. A., Rouse, M. N., Chao, S., Bulbula, W. D., and Elias, E. M. (2017). Inheritance and Bulked Segregant Analysis of Leaf Rust and Stem Rust Resistance in Durum Wheat Genotypes. *Phytopathol* 107, 1496–1506. doi:10.1094/phyto-12-16-0444-r

Aoun, M., Kolmer, J. A., Rouse, M. N., Elias, E. M., Breiland, M., Bulbula, W. D., et al. (2019). Mapping of Novel Leaf Rust and Stem Rust Resistance Genes in the Portuguese Durum Wheat Landrace. *G* 9, 2535–2547. doi:10.1534/g3.119.400292

Autrique, E., Tanksley, S. D., Sorrells, M. E., and Singh, R. P. (1995). Molecular Markers for Four Leaf Rust Resistance Genes Introgressed into Wheat from Wild Relatives. *Genome* 38 (1), 75–83. doi:10.1139/g95-009

Ballini, E., Morel, J. B., Droc, G., Price, A., Courtois, B., Notteghem, J. L., et al. (2008). A Genome-wide Meta-Analysis of rice Blast Resistance Genes and Quantitative Trait Loci Provides New Insights into Partial and Complete Resistance. *Mol. Plant Microbe Interact* 21, 859–868. doi:10.1094/mpmi-21-7-0859

Bansal, U. K., Hayden, M. J., Venkata, B. P., Khanna, R., Saini, R. G., and Bariana, H. S. (2008). Genetic Mapping of Adult Plant Leaf Rust Resistance Genes *Lr48* and *Lr49* in Common Wheat. *Theor. Appl. Genet.* 117 (3), 307–312. doi:10.1007/s00122-008-0775-6

Bhardwaj, S. C., Prashar, M., Jain, S. K., Kumar, S., and Datta, D. (2010). Adult Plant Resistance in Some Indian Wheat Genotypes and Postulation of Leaf Rust Resistance Genes. *Indian Phytopathol* 63, 174–180.

Bhardwaj, S. C., Singh, G. P., Gangwar, O. P., Prasad, P., and Kumar, S. (2019). Status of Wheat Rust Research and Progress in Rust Management-Indian Context. *J. Agron.* 9, 892. doi:10.3390/agronomy9120892

Bokore, F. E., Knox, R. E., Cuthbert, R. D., Pozniak, C. J., McCallum, B. D., N'Diaye, A., et al. (2020). Mapping Quantitative Trait Loci Associated with Leaf Rust Resistance in Five spring Wheat Populations Using Single Nucleotide Polymorphism Markers. *Plos One* 15, e0230855. doi:10.1371/journal.pone.0230855

Bolton, M. D., Kolmer, J. A., and Garvin, D. F. (2008). Wheat Leaf Rust Caused by Pucciniatriticina. *Mol. Plant Pathol.* 9, 563–575. doi:10.1111/j.1364-3703.2008.00487.x

Bremenkamp-Barrett, B., Faris, J. D., and Fellers, J. P. (2008). Molecular Mapping of the Leaf Rust Resistance Gene *Lr17a* in Wheat. *Crop Sci.* 48 (3), 1124–1128.

Brown-Guedira, G. L., Singh, S., and Fritz, A. 2. (2003). Performance and Mapping of Leaf Rust Resistance Transferred to Wheat from Triticum Timopheevii Subsp. *Armeniacum*. *Phytopathology* 93 (7), 784–789. doi:10.1094/phyto.2003.93.7.784

Carpenter, N. R., Griffey, C. A., Malla, S., Chao, S., and Brown-Guedira, G. L. (2017). Mapping *Lr18*: a Leaf Rust Resistant Gene Widely Deployed in Soft Red winter Wheat. *J. Plant Dis. Biomark* 1, 4–10.

Charpe, A., Koul, S., Gupta, S. K., Singh, A., Pallavi, S. K., and andPrabhu, K. V. (2012). Marker Assisted Gene Pyramiding of Leaf Rust Resistance Genes *Lr24, Lr28* and *Lr9* in a Bread Wheat Cultivar HD2329. *J. Wheat Res.* 4, 20–28.

Cherukuri, D. P., Gupta, S. K., Charpe, A., Koul, S., Prabhu, K. V., Singh, R. B., et al. (2005). Molecular Mapping of Aegilops Speltoides Derived Leaf Rust Resistance Gene *Lr28* in Wheat. *Euphytica* 143 (1), 19–26. doi:10.1007/s10681-005-1680-6

Chhuneja, P., Vikal, Y., Kaur, S., Singh, R., Juneja, N. S., Bains, S., et al. (2011). Marker-assisted Pyramiding of Leaf Rust Resistance Genes *Lr24* and *Lr28* in Wheat (*Triticumaestivum*). *Ind. J. Agrisci* 81, 214–218.

Clair, St. D. A. (2010). Quantitative Disease Resistance and Quantitative Resistance Loci in Breeding. *Ann. Rev. Phytopathol* 48, 247–268. doi:10.1146/annurev-phyto-080508-081904

Cloutier, S., McCallum, B. D., Loutre, C., Banks, T. W., Wicker, T., Feuillet, C., et al. (2007). Leaf Rust Resistance Gene *Lr1*, Isolated from Bread Wheat (*Triticum aestivum* L.) Is a Member of the Large Psr567 Gene Family. *Plant Mol. Biol.* 65 (1), 93–106. doi:10.1007/s11103-007-9201-8

Cuddy, W., Park, R., Bariana, H., Bansal, U., Singh, D., Roake, J., et al. (2016). Expected Responses of Australian Wheat, Triticale and Barley Varieties to the Cereal Rust Diseases and Genotypic Data for Oat Varieties. *Cereal Rust Rep.* 14, 1–8.

Dadkhodaie, N. A., Karaoglou, H., Wellings, C. R., and Park, R. F. (2011). Mapping Genes *Lr53* and Yr35 on the Short Arm of Chromosome 6B of Common Wheat with Microsatellite Markers and Studies of Their Association with *Lr36*. *Theor. Appl. Genet.* 122 (3), 479–487. doi:10.1007/s00122-010-1462-y

Darino, M. A., Dieguez, M. J., Singh, D., Ingala, L. R., Pergolesi, M. F., Park, R. F., et al. (2015). Detection and Location of *Lr11* and Other Leaf Rust Resistance Genes in the Durably Resistant Wheat Cultivar Buck Poncho. *Euphytica* 206 (1), 135–147. doi:10.1007/s10681-015-1486-0

Datta, D., Nayar, S. K., Bhardwaj, S. C., Prashar, M., and Kumar, S. (2008). Detection and Inheritance of Leaf Rust Resistance in Common Wheat Lines Agra Local and IWP94. *Euphytica* 159, 343–351. doi:10.1007/s10681-007-9522-3

Dholakia, B. B., Rajwade, A. V., Hosmani, P., Khan, R. R., Chavan, S., Reddy, D. M. R., and Gupta, V. S. (2013). Molecular Mapping of Leaf Rust Resistance Gene *Lr15* in Hexaploid Wheat. *Mol. Breed.* 31 (3), 743–747. doi:10.1007/s11032-012-9813-9

Dinh, H. X., Singh, D., Periyannan, S., Park, R. F., and Pourkheirandish, M. (2020). Molecular Genetics of Leaf Rust Resistance in Wheat and Barley. *Theor. Appl. Genet.* 133, 2035–2050. doi:10.1007/s00122-020-03570-8

Dinkar, V., Jha, S. K., Mallick, N., Niranjana, M., Agarwal, P., and Sharma, J. B. (2020). Molecular Mapping of a New Recessive Wheat Leaf Rust Resistance Gene Originating from Triticumspelta. *Sci. Rep.* 10, 1–9. doi:10.1038/s41598-020-78679-3

Draz, I. S., Abou-Elseoud, M. S., Kamara, A. E., Alaa-Eldein, O. A., and El-Bebany, A. F. (2015). Screening of Wheat Genotypes for Leaf Rust Resistance along with Grain Yield. *Ann. Agric Sci* 60, 29–39. doi:10.1016/j.aoas.2015.01.001

Dyck, P. L. (1993). The Inheritance of Leaf Rust Resistance in the Wheat Cultivar Pasqua. *Canad J. Plant Sci.* 73, 903–906. doi:10.4141/cjps93-118

El-Orabey, W. M., and Nagaty, H. H. (2013). Detection of the Leaf Rust Resistance Gene *Lr9* in Some Egyptian Wheat Varieties. *Minufiya J. Agric. Res.* 38, 895–907.

Ellis, J. G., Lagudah, E. S., Spielmeyer, W., and Dodds, P. N. (2014). The Past, Present and Future of Breeding Rust Resistant Wheat. *Front. Plant Sci.* 5, 641. doi:10.3389/fpls.2014.00641

Fellers, J. P., Sakthikumar, S., He, F., McRell, K., Bakkeren, G., Cuomo, C. A., et al. (2021). Whole-genome Sequencing of Multiple Isolates of Puccinia Triticina Reveals Asexual Lineages Evolving by Recurrent Mutations. *G* 11 (9), jkab219. doi:10.1093/g3journal/jkab219

Figlan, S., Ntushelo, K., Mwadzingeni, L., Terefe, T., Tsilo, T. J., and Shimelis, H. (2020). Breeding Wheat for Durable Leaf Rust Resistance in Southern Africa: Variability, Distribution, Current Control Strategies, Challenges and Future Prospects. *Front. Plant Sci.* 11, 549. doi:10.3389/fpls.2020.00549

Flor, H. H. (1946). Genetics of Pathogenicity in Melampsoralini. *J. Agric. Res.* 73, 335–357.

Gautam, T., Dhillon, G. S., Saripalli, G., Singh, V. P., Prasad, P., Kaur, S., et al. (2020). Marker-assisted Pyramiding of Genes/QTL for Grain Quality and Rust Resistance in Wheat (*Triticumaestivum* L.). *Mol. Breed.* 40, 49. doi:10.1007/s11032-020-01125-9

Genievskaya, Y., Abugalieva, S., Rsaliyev, A., Yskakova, G., Turuspekov, Y., Griffiths, S., et al. (2020a). QTL Mapping for Seedling and Adult Plant Resistance to Leaf and Stem Rusts in Pamyati Azieva× Paragon Mapping Population of Bread Wheat. *J. Agron.* 10 (9), 1285. doi:10.3390/agronomy10091285

Genievskaya, Y., Fedorenko, Y., Sarbayev, A., Amalova, A., Abugalieva, S., et al. (2019). Identification of OTLs for Resistance to Leaf and Stem Rusts in Bread Wheat (*Triticumaestivum* L.) Using a Mapping Population of 'PamyatiAzieva X Paragon'Vavilov. *J. Genet. Breed.* 23, 887–895. doi:10.18699/vj19.563

Genievskaya, Y., Turuspekov, Y., Rsaliyev, A., and Abugalieva, S. (2020b). Genome-wide Association Mapping for Resistance to Leaf, Stem, and Yellow Rusts of Common Wheat under Field Conditions of South Kazakhstan. *Peer J.* 8, 9820. doi:10.7717/peerj.9820

Ghimire, B., Sapkota, S., Bahri, B. A., Martinez-Espinoza, A. D., Buck, J. W., and Mergoum, M. (2020). Fusarium Head Blight and Rust Diseases in Soft Red winter Wheat in the Southeast United States: State of the Art, Challenges and Future Perspective for Breeding. *Front. Plant Sci.* 11, 1080. doi:10.3389/fpls.2020.01080

Goyeau, H., and Lannou, C. (2011). Specific Resistance to Leaf Rust Expressed at the Seedling Stage in Cultivars Grown in France from 1983 to 2007. *Euphytica* 178, 45–62. doi:10.1007/s10681-010-0261-5

Goyeau, H., Park, R., Schaeffer, B., and andLannou, C. (2006). Distribution of Pathotypes with Regard to Host Cultivars in French Wheat Leaf Rust Populations. *Phytopathology* 96, 264–273. doi:10.1094/phyto-96-0264

Gul'tyaeva, E. I., Kanyuka, I. A., Alpat'eva, N. V., Baranova, O. A., Dmitriev, A. P., and Pavlyushin, V. A. (2009). Molecular Approaches in Identifying Leaf Rust Resistance Genes in Russian Wheat Varieties. *Russ. Agricsci* 35, 316–319.

Gupta, A., Singh, C., Kumar, V., Tyagi, B. S., Tiwari, V., Chatrath, R., et al. (2018). *Wheat Varieties Notified in india since 1965*. Karnal: ICAR-Indian Institute of Wheat and Barley Research, 101.

Gupta, S. K., Charpe, A., Koul, S., Prabhu, K. V., and Haq, Q. M. R. (2005). Development and Validation of Molecular Markers Linked to an Aegilops Umbellulata–Derived Leaf-Rust-Resistance Gene, *Lr9*, for Marker-Assisted Selection in Bread Wheat. *Genome* 48 (5), 823–830. doi:10.1139/g05-051

Gupta, S. K., Charpe, A., Prabhu, K. V., and Haque, Q. M. R. (2006). Identification and Validation of Molecular Markers Linked to the Leaf Rust Resistance Gene *Lr19* in Wheat. *Theor. Appl. Genet.* 113 (6), 1027–1036. doi:10.1007/s00122-006-0362-7

Helguera, M., Khan, I. A., and Dubcovsky, J. (2000). Development of PCR Markers for the Wheat Leaf Rust Resistance Gene *Lr47*. *Theor. Appl. Genet.* 100 (7), 1137–1143. doi:10.1007/s001220051397

Helguera, M., Khan, I. A., Kolmer, J., Lijavetzky, D., Zhong-Qi, L., and Dubcovsky, J. (2003). PCR Assays for the *Lr37-Yr17-Sr38* Cluster of Rust Resistance Genes and Their Use to Develop Isogenic Hard Red spring Wheat Lines. *Crop Sci.* 43 (5), 1839–1847. doi:10.2135/cropsci2003.1839

Helguera, M., Vanzetti, L., Soria, M., Khan, I. A., Kolmer, J., and Dubcovsky, J. (2005). PCR Markers for Triticum Speltoides Leaf Rust Resistance Gene *Lr51* and Their Use to Develop Isogenic Hard Red spring Wheat Lines. *Crop Sci.* 45 (2), 728–734. doi:10.2135/cropsci2005.0728

Herrera-Foessel, S. A., Huerta-Espino, J., Calvo-Salazar, V., Lan, C. X., and Singh, R. P. (2014). *Lr72* Confers Resistance to Leaf Rust in Durum Wheat Cultivar Atil C2000. *Plant Dis.* 98 (5), 631–635. doi:10.1094/pdis-07-13-0741-re

Herrera-Foessel, S. A., Singh, R. P., Huerta-Espino, J., Rosewarne, G. M., Periyannan, S. K., Viccars, L., et al. (2012). *Lr68*: a New Gene Conferring Slow Rusting Resistance to Leaf Rust in Wheat. *Theor. Appl. Genet.* 124 (8), 1475–1486. doi:10.1007/s00122-012-1802-1

Herrera-Foessel, S. A., Singh, R. P., Huerta-Espino, J., William, H. M., Djurle, A., and Yuen, J. (2008). Molecular Mapping of a Leaf Rust Resistance Gene on the Short Arm of Chromosome 6B of Durum Wheat. *Plant Dis.* 92 (12), 1650–1654. doi:10.1094/pdis-92-12-1650

Hiebert, C. W., McCallum, B. D., and Thomas, J. B. (2014). *Lr70*, a New Gene for Leaf Rust Resistance Mapped in Common Wheat Accession KU3198. *Theor. Appl. Genet.* 127 (9), 2005–2009. doi:10.1007/s00122-014-2356-1

Hiebert, C. W., Thomas, J. B., McCallum, B. D., Humphreys, D. G., DePauw, R. M., Hayden, M. J., et al. (2010). An Introgression on Wheat Chromosome 4DL in RL6077 (Thatcher* 6/PI 250413) Confers Adult Plant Resistance to Stripe Rust and Leaf Rust (*Lr67*). *Theor. Appl. Genet.* 121 (6), 1083–1091. doi:10.1007/s00122-010-1373-y

Hiebert, C. W., Thomas, J. B., McCallum, B. D., and Somers, D. J. (2008). Genetic Mapping of the Wheat Leaf Rust Resistance Gene *Lr60* (*LrW2*). *Crop Sci.* 48 (3), 1020–1026. doi:10.2135/cropsci2007.08.0480

Hiebert, C. W., Thomas, J. B., Somers, D. J., McCallum, B. D., and Fox, S. L. (2007). Microsatellite Mapping of Adult-Plant Leaf Rust Resistance Gene *Lr22a* in Wheat. *Theor. Appl. Genet.* 115 (6), 877–884. doi:10.1007/s00122-007-0604-3

Ingala, L., López, M., Darino, M., Pergolesi, M. F., Diéguez, M. J., and Sacco, F. (2012). Genetic Analysis of Leaf Rust Resistance Genes and Associated Markers in the Durable Resistant Wheat Cultivar Sinvalocho MA. *Theor. Appl. Genet.* 124, 1305–1314. doi:10.1007/s00122-012-1788-8

Jain, N., Sinha, N., Krishna, H., Singh, P. K., Gautam, T., Prasad, P., et al. (2020). A Study of miRNAs and lncRNAs during *Lr28*-Mediated Resistance against Leaf Rust in Wheat (*Triticum aestivum* L.). *Physiol. Mol. Plant Pathol.* 112, 101552. doi:10.1016/j.pmpp.2020.101552

Johnson, R. (1984). A Critical Analysis of Durable Resistance. *Ann. Rev. Phtopathol* 22, 309–330. doi:10.1146/annurev.py.22.090184.001521

Joukhadar, R., Hollaway, G., Shi, F., Kant, S., Forrest, K., Wong, D., et al. (2020). Genome-wide Association Reveals a Complex Architecture for Rust Resistance in 2300 Worldwide Bread Wheat Accessions Screened under Various Australian Conditions. *Theor. Appl. Genet* 133, 2695–2712. doi:10.1007/s00122-020-03626-9

Kandiah, P., Chhetri, M., Hayden, M., Ayliffe, M., Bariana, H., and Bansal, U. (2020). Mapping of Adult Plant Leaf Rust Resistance in Aus27506 and Validation of Underlying Loci by In-Planta Fungal Biomass Accumulation. *Agronomy* 10, 943. doi:10.3390/agronomy10070943

Khan, H., Bhardwaj, S. C., Gangwar, O. P., Prasad, P., Kashyap, P. L., Savadi, S., et al. (2017). Identifying Some Additional Rust Resistance Genes in Indian Wheat Varieties Using Robust Markers. *Cereal Res. Commun.* 45, 633–646. doi:10.1556/0806.45.2017.041

Khan, R. R., Bariana, H. S., Dholakia, B. B., Naik, S. V., Lagu, M. D., Rathjen, A. J., et al. (2005). Molecular Mapping of Stem and Leaf Rust Resistance in Wheat. *Theor. Appl. Genet.* 111 (5), 846–850. doi:10.1007/s00122-005-0005-4

Kiran, K., Rawal, H. C., Dubey, H., Jaswal, R., Devanna, B. N., Gupta, D. K., et al. (2016). Draft Genome of the Wheat Rust Pathogen (Puccinia Triticina) Unravels Genome-wide Structural Variations during Evolution. *Genome Biol. Evol.* 8 (9), 2702–2721. doi:10.1093/gbe/evw197

Kolmer, J. A., Anderson, J. A., and Flor, J. M. (2010). Chromosome Location, Linkage with Simple Sequence Repeat Markers, and Leaf Rust Resistance Conditioned by Gene *Lr63* in Wheat. *Crop Sci.* 50 (6), 2392–2395. doi:10.2135/cropsci2010.01.0005

Kolmer, J. A., Bernardo, A., Bai, G., Hayden, M. J., and Anderson, J. A. (2019). Thatcher Wheat Line RL6149 Carries *Lr64* and a Second Leaf Rust Resistance Gene on Chromosome 1DS. *Theor. Appl. Genet.* 132 (10), 2809–2814. doi:10.1007/s00122-019-03389-y

Kolmer, J. A., Bernardo, A., Bai, G., Hayden, M. J., and Chao, S. (2018b). Adult Plant Leaf Rust Resistance Derived from Toropi Wheat Is Conditioned by *Lr78* and Three Minor QTL. *Phytopathology* 108 (2), 246–253. doi:10.1094/phyto-07-17-0254-r

Kolmer, J. A. (1996). Genetics of Resistance to Wheat Leaf Rust. *Ann. Rev. Phytopathol* 34, 435–455. doi:10.1146/annurev.phyto.34.1.435

Kolmer, J. A., and Liu, J. Q. (2002). Inheritance of Leaf Rust Resistance in the Wheat Cultivars AC Majestic, AC Splendor, and AC Karma. *Canad J. Plant Pathol.* 24, 327–331. doi:10.1080/07060660209507017

Kolmer, J. A., Su, Z., Bernardo, A., Bai, G., and Chao, S. (2018a). Mapping and Characterization of the New Adult Plant Leaf Rust Resistance Gene *Lr77* Derived from Santa Fe winter Wheat. *Theor. Appl. Genet.* 131 (7), 1553–1560. doi:10.1007/s00122-018-3097-3

Kolmer, J. A. (2013). Leaf Rust of Wheat: Pathogen Biology, Variation and Host Resistance. *Forests* 4, 70–84. doi:10.3390/f4010070

Kolmer (2019). Virulence of *Puccinia Triticina*, the Wheat Leaf Rust Fungus in the United States in 2017. *Plant Dis.* 103, 2113–2120. doi:10.1094/pdis-09-18-1638-sr

Kourelis, J., and van der Hoorn, R. A. (2018). Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *The Plant Cell* 30 (2), 285–299. doi:10.1105/tpc.17.00579

Kthiri, D., Loladze, A., N'Diaye, A., Nilsen, K. T., Walkowiak, S., Dreisigacker, S., et al. (2019). Mapping of Genetic Loci Conferring Resistance to Leaf Rust from Three Globally Resistant Durum Wheat Sources. *Front. Plant Sci.* 10, 1247. doi:10.3389/fpls.2019.01247

Kumar, A., Choudhary, A., Kaur, H., and Mehta, S. (2022). A Walk towards Wild Grasses to Unlock the Clandestine of Gene Pools for Wheat Improvement: A Review. *Plant Stress* 3. doi:10.1016/j.stress.2021.100048

Kumar, D., Kumar, A., Chhokar, V., Gangwar, O. P., Bhardwaj, S. C., Sivasamy, M., et al. (2020). Genome-wide Association Studies in Diverse spring Wheat Panel for Stripe, Stem, and Leaf Rust Resistance. *Front. Plant Sci.* 11, 748. doi:10.3389/fpls.2020.00748

Kumar, S., Bhardwaj, S. C., Gangwar, O. P., Sharma, A., Qureshi, N., Kumaran, V. V., et al. (2021). *Lr80*: A New and Widely Effective Source of Leaf Rust Resistance of Wheat for Enhancing Diversity of Resistance Among Modern Cultivars. *Theor. Appl. Genet.* 134, 849–858. doi:10.1007/s00122-020-03735-5

Kuraparthy, V., Sood, S., Chhuneja, P., Dhaliwal, H. S., Kaur, S., Bowden, R. L., et al. (2007). A Cryptic Wheat–Aegilops Triuncialis Translocation with Leaf Rust Resistance Gene *Lr58*. *Crop Sci.* 47 (5), 1995–2003. doi:10.2135/cropsci2007.01.0038

Kuraparthy, V., Sood, S., See, D. R., and Gill, B. S. (2009). Development of a PCR Assay and Marker-assisted Transfer of Leaf Rust and Stripe Rust Resistance Genes *Lr57* and *Yr40* into Hard Red winter Wheats. *Crop Sci.* 49 (1), 120–126. doi:10.2135/cropsci2008.03.0143

Lagudah, E. S., Krattinger, S. G., Herrera-Foessel, S., Singh, R. P., Huerta-Espino, J., Spielmeyer, W., et al. (2009). Gene-specific Markers for the Wheat Gene *Lr34/Yr18/Pm38* Which Confers Resistance to Multiple Fungal Pathogens. *Theor. Appl. Genet.* 119 (5), 889–898. doi:10.1007/s00122-009-1097-z

Lagudah, E. S., McFadden, H., Singh, R. P., Huerta-Espino, J., Bariana, H. S., and Spielmeyer, W. (2006). Molecular Genetic Characterization of the *Lr34*/Yr18 Slow Rusting Resistance Gene Region in Wheat. *Theor. Appl. Genet.* 114 (1), 21–30. doi:10.1007/s00122-006-0406-z

Li, C., Wang, Z., Li, C., Bowden, R., Bai, G., Li, C., et al. (2017). Mapping of Quantitative Trait Loci for Leaf Rust Resistance in the Wheat Population Ning7840× Clark. *Plant Dis.* 101 (12), 1974–1979. doi:10.1094/pdis-12-16-1743-re

Li, J., Shi, L., Wang, X., Zhang, N., Wei, X., Zhang, L., et al. (2018). Leaf Rust Resistance of 35 Wheat Cultivars (Lines). *J. Plant Pathol. Microbiol.* 9, 429.

Li, Z., Yuan, C., Herrera-Foessel, S. A., Randhawa, M. S., Huerta-Espino, J., Liu, D., et al. (2020). Four Consistent Loci Confer Adult Plant Resistance to Leaf Rust in the Durum Wheat Lines Heller# 1 and Dunkler. *Phytopathology* 110, 892–899. doi:10.1094/phyto-09-19-0348-r

Ma, S., Lapin, D., Liu, L., Sun, Y., Song, W., Zhang, X., et al. (2020). Direct Pathogen-Induced Assembly of an NLR Immune Receptor Complex to Form a Holoenzyme. *Science* 370 (6521), eabe3069. doi:10.1126/science.abe3069

Mago, R., Spielmeyer, W., Lawrence, G., Lagudah, E., Ellis, J., and Pryor, A. (2002). Identification and Mapping of Molecular Markers Linked to Rust Resistance Genes Located on Chromosome 1RS of rye Using Wheat-rye Translocation Lines. *Theor. Appl. Genet.* 104 (8), 1317–1324. doi:10.1007/s00122-002-0879-3

Malaker, P. K., and Reza, M. M. (2011). Resistance to Rusts in Bangladeshi Wheat. *Czech J. Genet. Plant Breed.* 47, S155–S159. doi:10.17221/3271-cjgpb

Mallick, N., Sharma, J. B., Tomar, R. S., Sivasamy, M., and Prabhu, K. V. (2015). Marker Assisted Back-Cross Breeding to Combine Multiple Rust Resistance in Wheat. *Plant Breed* 134, 172–177. doi:10.1111/pbr.12242

Marais, G. F., and Botes, W. C. (2003). "Recurrent Mass Selection as a Means to Pyramid Major Genes for Pest Resistance in Spring Wheat," in Proceedings of the 10th International Wheat Genetics Symposium, September, 1–6.

Marais, F., Marais, A., McCallum, B., and Pretorius, Z. (2009). Transfer of Leaf Rust and Stripe Rust Resistance Genes *Lr62* and *Yr42* from Aegilops Neglecta Req. Ex Bertol. To Common Wheat. *Crop Sci.* 49 (3), 871–879. doi:10.2135/cropsci2008.06.0317

Marais, G. F., Bekker, T. A., Eksteen, A., McCallum, B., Fetch, T., and Marais, A. S. (2010). Attempts to Remove Gametocidal Genes Co-transferred to Common Wheat with Rust Resistance from Aegilops Speltoides. *Euphytica* 171 (1), 71–85. doi:10.1007/s10681-009-9996-2

Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., and Staskawicz, B. J. (2020). Structure of the Activated ROQ1 Resistosome Directly Recognizing the Pathogen Effector XopQ. *Science* 370 (6521), eabd9993. doi:10.1126/science.abd9993

Mateos-Hernandez, M., Singh, R. P., Hulbert, S. H., Bowden, R. L., Huerta-Espino, J., Gill, B. S., et al. (2006). Targeted Mapping of ESTs Linked to the Adult Plant Resistance Gene *Lr46* in Wheat Using Synteny with rice. *Funct. Integr. Genomic* 6 (2), 122–131. doi:10.1007/s10142-005-0017-9

McCallum, B. D., Hiebert, C. W., Cloutier, S., Bakkeren, G., Rosa, S. B., Humphreys, D. G., et al. (2016). A Review of Wheat Leaf Rust Research and the Development of Resistant Cultivars in Canada. *Canad J. Plant Pathol.* 38, 1–8. doi:10.1080/07060661.2016.1145598

McCartney, C. A., Somers, D. J., McCallum, B. D., Thomas, J., Humphreys, D. G., Menzies, J. G., et al. (2005). Microsatellite Tagging of the Leaf Rust Resistance Gene *Lr16* on Wheat Chromosome 2BSc. *Mol. Breed* 15 (4), 329–337. doi:10.1007/s11032-004-5948-7

McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Morris, C., and Xia, X. C. (2017). *Catalog of Gene Symbols for Wheat: 2017*. Aberdeen, USA: supplement.

McIntosh, R. A. (1975). Cytogenetical Studies in Wheat VIII. Telocentric Mapping and Linkage Studies Involving Sr22 and Other Genes in Chromosome 7AL. *Aust. J. Biol. Sci.* 28 (6), 531–8.

McIntosh, R. A., Friebe, B., Jiang, J., and Gill, B. S. (1995). Cytogenetical Studies in Wheat XVI. Chromosome Location of a New Gene for Resistance to Leaf Rust in a Japanese Wheat-Rye Translocation Line. *Euphytica* 83 (2), 141–7.

McIntosh, R. A., Dubcovsky, J., Rogers, W. J., Xia, X. C., and Raupp, W. J. (2020). Catalogue of Gene Symbols for Wheat: 2020 Supplement. *Ann. Wheat Newslett* 66.

Mebrate, S. A., Oerke, E. C., Dehne, H., and Pillen, K. (2008). Mapping of the Leaf Rust Resistance Gene Lr38 on Wheat Chromosome Arm 6DL Using SSR Markers. *Euphytica* 162 (3), 457–466. doi:10.1007/s10681-007-9615-z

Mohler, V., Singh, D., Singrün, C., and Park, R. F. (2012). Characterization and Mapping of *Lr65* in Spelt Wheat 'Altgold Rotkorn. *Plant breed* 131 (2), 252–257. doi:10.1111/j.1439-0523.2011.01934.x

Moore, J. W., Herrera-Foessel, S., Lan, C., Schnippenkoetter, W., Ayliffe, M., Huerta-Espino, J., et al. (2015). A Recently Evolved Hexose Transporter Variant Confers Resistance to Multiple Pathogens in Wheat. *Nat. Genet.* 47, 1494–1498. doi:10.1038/ng.3439

Moullet, O., Fossati, D., Mascher, F., Guadangnolo, R., and andSchori, A. (2009). "Use of Marker-Assisted Selection (MAS) for Pyramiding Two Leaf Rust Resistance Genes, (*Lr9,Lr24* and *Lr22a*) in Wheat," in *Int. Conf. Conventional Mol. Breed. Field Veg. Crops* (Novi Sad, Serbia: Springer).

Muhammad, S., Ahmad, A., Awan, F. S., Khan, A. I., Qasim, M., Rehman, A., et al. (2018). Genome Wide Association Analysis for Leaf Rust Resistance in spring Wheat (Triticumaestivum L.) Germplasm. *Int. J. Agric. Biol.* 20, 2387–2394.

Mundt, C. C., Cowger, C., and Garrett, K. A. (2002). Relevance of Integrated Disease Management to Resistance Durability. *Euphytica* 124, 245–252. doi:10.1023/a:1015642819151

Naik, B. K., Sharma, J. B., Sivasamy, M., Prabhu, K. V., Tomar, R. S., and Tomar, S. M. S. (2015). Molecular Mapping and Validation of the Microsatellite Markers Linked to the Secale Cereale-Derived Leaf Rust Resistance Gene *Lr45* in Wheat. *Mol. Breed* 35 (2), 1–10. doi:10.1007/s11032-015-0234-4

Naik, S., Gill, K. S., Prakasa Rao, V. S., Gupta, V. S., Tamhankar, S. A., Pujar, S., et al. (1998). Identification of a STS Marker Linked to the Aegilops Speltoides-Derived Leaf Rust Resistance Gene *Lr28* in Wheat. *Theor. Appl. Genet.* 97 (4), 535–540. doi:10.1007/s001220050928

Narang, D., Kaur, S., Steuernagel, B., Ghosh, S., Dhillon, R., Bansal, M., et al. (2019). Fine Mapping of Aegilopsperegrina Co-segregating Leaf and Stripe Rust Resistance Genes to Distal-Most End of 5DS. *Theor. Appl. Genet.* 132, 1473–1485. doi:10.1007/s00122-019-03293-5

Neelam, K., Brown-Guedira, G., and Huang, L. (2013). Development and Validation of a Breeder-Friendly KASPar Marker for Wheat Leaf Rust Resistance Locus *Lr21*. *Mol. Breed* 31 (1), 233–237. doi:10.1007/s11032-012-9773-0

Nelson, J. C., Singh, R. P., Autrique, J. E., and Sorrells, M. E. (1997). Mapping Genes Conferring and Suppressing Leaf Rust Resistance in Wheat. *Crop Sci.* 37 (6), 1928–1935. doi:10.2135/cropsci1997.0011183x003700060043x

Neu, C., Stein, N., and Keller, B. (2002). Genetic Mapping of the *Lr20 Pm1* Resistance Locus Reveals Suppressed Recombination on Chromosome Arm 7AL in Hexaploid Wheat. *Genome* 45 (4), 737–744. doi:10.1139/g02-040

Nocente, F., Gazza, L., and Pasquini, M. (2007). Evaluation of Leaf Rust Resistance Genes *Lr1, Lr9, Lr24, Lr47* and Their Introgression into Common Wheat Cultivars by Marker-Assisted Selection. *Euphytica* 155, 329–336. doi:10.1007/s10681-006-9334-x

Oelke, L. M., and Kolmer, J. A. (2005). Genetics of Leaf Rust Resistance in spring Wheat Cultivars Alsen and Norm. *Phytopathol* 95, 773–778. doi:10.1094/phyto-95-0773

Papaïx, J., Goyeau, H., DuCheyron, P., Monod, H., and Lannou, C. (2011). Influence of Cultivated Landscape Composition on Variety Resistance: an Assessment Based on Wheat Leaf Rust Epidemics. *New Phytol.* 191, 1095–1107. doi:10.1111/j.1469-8137.2011.03764.x

Park, R. F., Mohler, V., Nazari, K., and Singh, D. (2014). Characterisation and Mapping of Gene *Lr73* Conferring Seedling Resistance to Puccinia Triticina in Common Wheat. *Theor. Appl. Genet.* 127 (9), 2041–2049. doi:10.1007/s00122-014-2359-y

Parlevliet, J. E. (2002). Durability of Resistance against Fungal, Bacterial and Viral Pathogens; Present Situation. *Euphytica* 124, 147–156. doi:10.1023/a:1015601731446

Pathan, A. K., and Park, R. F. (2006). Evaluation of Seedling and Adult Plant Resistance to Leaf Rust in European Wheat Cultivars. *Euphytica* 149, 327–342. doi:10.1007/s10681-005-9081-4

Pietrusińska, A., Czembor, J. H., and Czembor, P. C. (2011). Pyramiding Two Genes for Leaf Rust and Powdery Mildew Resistance in Common Wheat. *Cereal Res. Commun.* 39, 577–588.

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. *Phytopathol* 108, 1344–1354. doi:10.1094/phyto-06-18-0208-rvw

Ponce-Molina, L. J., Huerta-Espino, J., Singh, R. P., Basnet, B. R., Alvarado, G., Randhawa, M. S., et al. (2018). Characterization of Leaf Rust and Stripe Rust Resistance in spring Wheat 'Chilero. *Plant Dis.* 102, 421–427. doi:10.1094/pdis-11-16-1545-re

Poudel, R. S. (2015). *The Acquisition of Useful Disease Resistance Genes for Hard Red winter Wheat Improvement*. Fargo, North Dakota: North Dakota State University. Doctoral dissertation.

Prabhu, K. V., Gupta, S. K., Charpe, A., and Koul, S. (2004). SCAR Marker Tagged to the Alien Leaf Rust Resistance Gene *Lr19* Uniquely Marking the Agropyron Elongatum-derived Gene *Lr24* in Wheat: a Revision. *Plant Breed* 123 (5), 417–420. doi:10.1111/j.1439-0523.2004.00971.x

Prasad, P., Savadi, S., Bhardwaj, S. C., and Gupta, P. K. (2020). The Progress of Leaf Rust Research in Wheat. *Fungal Biol.* 124 (6), 537–550. doi:10.1016/j.funbio.2020.02.013

Procunier, J. D., Townley-Smith, T. F., Fox, S., Prashar, S., Gray, M., Kim, W. K., et al. (1995). PCR-based RAPD/DGGE Markers Linked to Leaf Rust Resistance Genes *Lr29* and *Lr25* in Wheat (*Triticum aestivum* L.). *J. Genet. Breed.* 49 (1), 87–91.

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 (5), 1091–1098. doi:10.1007/s00122-018-3060-3

Rani, K., Raghu, B. R., Jha, S. K., Agarwal, P., Mallick, N., Niranjana, M., et al. (2020). A Novel Leaf Rust Resistance Gene Introgressed from

Aegilopsmarkgrafii Maps on Chromosome Arm 2AS of Wheat. *Theor. Appl. Genet.* 133, 2685–2694. doi:10.1007/s00122-020-03625-w

Raupp, W. J., Brown-Guedira, G. L., and Gill, B. S. (2001). Cytogenetic and Molecular Mapping of the Leaf Rust Resistance Gene *Lr39* in Wheat. *Theor. Appl. Genet.* 102 (2), 347–352. doi:10.1007/s001220051652

Revathi, P., Tomar, S. M. S., and Singh, N. K. (2010). Marker Assisted Gene Pyramiding of Leaf Rust Resistance Genes *Lr24, Lr28* along with Stripe Rust Resistance Gene *Yr15* in Wheat (*Triticumaestivum* L.). *Ind. J. Genet.* 70, 349–354.

Riaz, M., and Wong, Y. (2017). Estimation of Yield Losses Due to Leaf Rust and Late Seeding on Wheat (*Triticum aestivum* L) Variety Seher-06 in District Faisalabad, Punjab, Pakistan. *Adv. Biotech. Micro* 4. doi:10.19080/AIBM.2017.04.555657

Roelfs, A. P. (1992). *Rust Diseases of Wheat: Concepts and Methods of Disease Management*. Mexico: Cimmyt.

Rosa, S. B., McCallum, B., Brûlé-Babel, A., Hiebert, C., Shorter, S., Randhawa, H. S., et al. (2016). Inheritance of Leaf Rust and Stripe Rust Resistance in the Brazilian Wheat Cultivar 'Toropi. *Plant Dis.* 100, 1132–1137. doi:10.1094/pdis-10-15-1128-re

Rosa, S. B., Zanella, C. M., Hiebert, C. W., Brûlé-Babel, A. L., Randhawa, H. S., Shorter, S., et al. (2019). Genetic Characterization of Leaf and Stripe Rust Resistance in the Brazilian Wheat Cultivar Toropi. *Phytopathol* 109, 1760–1768. doi:10.1094/phyto-05-19-0159-r

Sacco, F., Suarez, E. Y., and Naranjo, T. (1998). Mapping of the Leaf Rust Resistance Gene *Lr3* on Chromosome 6B of Sinvalocho MA Wheat. *Genome* 41 (5), 686–690. doi:10.1139/g98-067

Samborski, D. J. (1985). "Wheat Leaf Rust," in *Diseases, Distribution, Epidemiology, and Control* (Academic Press), 39–59. doi:10.1016/b978-0-12-148402-6.50010-9

Samsampour, D., Zanjani, B. M., Singh, A., Pallavi, J. K., and Prabhu, K. V. (2009). Marker Assisted Selection to Pyramid Seedling Resistance Gene *Lr24* and Adult Plant Resistance Gene *Lr48* for Leaf Rust Resistance in Wheat. *Ind. J. Genet.* 69, 1–9.

Sapkota, S., Hao, Y., Johnson, J., Buck, J., Aoun, M., and andMergoum, M. (2019). Genome-wide Association Study of a Worldwide Collection of Wheat Genotypes Reveals Novel Quantitative Trait Loci for Leaf Rust Resistance. *Plant Genome* 12, 190033. doi:10.3835/plantgenome2019.05.0033

Sapkota, S., Mergoum, M., Kumar, A., Fiedler, J. D., Johnson, J., Bland, D., et al. (2020). A Novel Adult Plant Leaf Rust Resistance Gene *Lr2K38* Mapped on Wheat Chromosome 1AL. *Plant Genome* 13 (3), e20061. doi:10.1002/tpg2.20061

Saripalli, G., Sharma, C., Gautam, T., Singh, K., and Jain, N. (2020a). Complex Relationship between DNA Methylation and Gene Expression Due to *Lr28* in Wheat-Leaf Rust Pathosystem. *Mol. Bio Rep.* 47, 1339–1360. doi:10.1007/s11033-019-05236-1

Saripalli, G., Singh, K., Gautam, T., Santosh, K., Saurabh, R., Prasad, P., et al. (2020b). Genome-wide Analysis of H3K4me3 and H3K27me3 Modifications Due to *Lr28* for Leaf Rust Resistance in Bread Wheat (*Triticum aestivum*). *Plant Mol. Biol.* 104, 113–136. doi:10.1007/s11103-020-01029-4

Schachermayr, G., Feuillet, C., and Keller, B. (1997). Molecular Markers for the Detection of the Wheat Leaf Rust Resistance Gene *Lr10* in Diverse Genetic Backgrounds. *Mol. Breed.* 3 (1), 65–74. doi:10.1023/a:1009619905909

Seyfarth, R., Feuille, C., and Keller, B. (1998). "Development and Characterization of Molecular Markers for the Adult Plant Leaf Rust Resistance Genes Lr13 and Lr35 in Wheat," in *Proc 9th Int Wheat Genet Symp*. Editors A. E. Slinkard (Saskatoon, Canada: University Extension Press), Vol. 3, 154.

Seyfarth, R., Feuillet, C., Schachermayr, G., Winzeler, M., and Keller, B. (1999). Development of a Molecular Marker for the Adult Plant Leaf Rust Resistance Gene *Lr35* in Wheat. *Theor. Appl. Genet.* 99 (3), 554–560. doi:10.1007/s001220051268

Singh, A., Pallavi, J. K., Gupta, P., and Prabhu, K. V. (2012). Identification of Microsatellite Markers Linked to Leaf Rust Resistance Gene *Lr25* in Wheat. *J. Appl. Genet.* 53 (1), 19–25. doi:10.1007/s13353-011-0070-0

Singh, M., Mallick, N., Chand, S., Kumari, P., Sharma, J. B., Sivasamy, M., et al. (2017). Marker-assisted Pyramiding of Thinopyrum-Derived Leaf Rust Resistance Genes *Lr19* and *Lr24* in Bread Wheat Variety HD2733. *J. Genet.* 96, 951–957. doi:10.1007/s12041-017-0859-7

Singh, R. P., and Rajaram, S. (2002). *Breeding for Disease Resistance in Wheat*. Rome, Italy: FAO plant production and protection series, 30.

Singh, R. P., and Rajaram, S. (1993). Genetics of Adult Plant Resistance to Stripe Rust in Ten spring Bread Wheats. *Euphytica* 72, 1–7. doi:10.1007/bf00023766

Singh, S., and Bowden, R. L. (2011). Molecular Mapping of Adult-Plant Race-specific Leaf Rust Resistance Gene *Lr12* in Bread Wheat. *Mol. Breed.* 28 (2), 137–142. doi:10.1007/s11032-010-9467-4

Singla, J., Lüthi, L., Wicker, T., Bansal, U., Krattinger, S. G., and Keller, B. (2017). Characterization of *Lr75*: a Partial, Broad-Spectrum Leaf Rust Resistance Gene in Wheat. *Theor. Appl. Genet.* 130 (1), 1–12. doi:10.1007/s00122-016-2784-1

Sivasamy, M. (2014). *Deployment of Rust Resistance Genes in Wheat Varities Increased Wheat Production in India in National Seminar on Challenges and Innovative Approaches in Crop Improvement*. Madurai, Tamilnadu, India: Agricultural college and research institute.

Slikova, S., Gregova, E., Bartos, P., and andKraic, J. (2003). Marker-assisted Selection for Leaf Rust Resistance in Wheat by Transfer of Gene *Lr19*. *Plant Protect Sci.* 39, 13–17.

Soriano, J. M., and Royo, C. (2015). Dissecting the Genetic Architecture of Leaf Rust Resistance in Wheat by QTL Meta-Analysis. *Phytopathol* 105, 1585–1593. doi:10.1094/phyto-05-15-0130-r

Stuthman, D. D., Leonard, K. J., and Miller-Garvin, J. (2007). Breeding Crops for Durable Resistance to Disease. *AdvAgron* 95, 319–367. doi:10.1016/s0065-2113(07)95004-x

Sun, X., Bai, G., Carver, B. F., and Bowden, R. (2010). Molecular Mapping of Wheat Leaf Rust Resistance Gene *Lr42*. *Crop Sci.* 50 (1), 59–66. doi:10.2135/cropsci2009.01.0049

Sun, X., Bai, G., and Carver, B. F. (2009). Molecular Markers for Wheat Leaf Rust Resistance Gene *Lr41*. *Mol. Breed* 23 (2), 311–321. doi:10.1007/s11032-008-9237-8

Tar, M., Purnhauser, L., and Csősz, M. (2008). Identification and Localization of Molecular Markers Linked to the *Lr52* Leaf Rust Resistance Gene of Wheat. *Cereal Res. Commun.* 36 (3), 409–415. doi:10.1556/crc.36.2008.3.5

Terracciano, I., Maccaferri, M., Bassi, F., Mantovani, P., Sanguineti, M. C., Salvi, S., et al. (2013). Development of COS-SNP and HRM Markers for High-Throughput and Reliable Haplotype-Based Detection of *Lr14a* in Durum Wheat (Triticum Durum Desf.). *Theor. Appl. Genet.* 126 (4), 1077–1101. doi:10.1007/s00122-012-2038-9

Tomar, S. M. S., Singh, S. K., and Sivasamy, M. (2014). Wheat Rusts in India: Resistance Breeding and Gene Deployment-A Review. *Ind. J. Genet. Plant Breed.* 74, 129–156. doi:10.5958/0975-6906.2014.00150.3

Toor, P. I., Kaur, S., Bansal, M., Yadav, B., and andChhuneja, P. (2016). Mapping of Stripe Rust Resistance Gene in an Aegilops Caudate Introgression Line in Wheat and its Genetic Association with Leaf Rust Resistance. *J. Genet.* 95, 933–938. doi:10.1007/s12041-016-0718-y

Tsilo, T. J., Kolmer, J. A., and Anderson, J. A. (2014). Molecular Mapping and Improvement of Leaf Rust Resistance in Wheat Breeding Lines. *Phytopathol* 104 (8), 865–870. doi:10.1094/phyto-10-13-0276-r

Wang, C., Yin, G., Xia, X., He, Z., Zhang, P., Yao, Z., et al. (2016). Molecular Mapping of a New Temperature-Sensitive Gene *LrZH22* for Leaf Rust Resistance in Chinese Wheat Cultivar Zhoumai 22. *Mol. Breed.* 36, 18. doi:10.1007/s11032-016-0437-3

Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., and Chai, J. (2019). Reconstitution and Structure of a Plant NLR Resistosome Conferring Immunity. *Science* 364 (6435), eaav5870. doi:10.1126/science.aav5870

Wang, J., Shi, L., Li, X., and Liu, D. (2014). Genetic Analysis and Molecular Mapping of Leaf Rust Resistance Genes in the Wheat Line 5R618. *Czech J. Genet. Plant Breed.* 50, 262–267. doi:10.17221/164/2014-cjgpb

Wu, J. Q., Dong, C., Song, L., and Park, R. F. (2020). Long-Read–Based De Novo Genome Assembly and Comparative Genomics of the Wheat Leaf Rust Pathogen Puccinia Triticina Identifies Candidates for Three Avirulence Genes. *Front. Genet.* 11, 521. doi:10.3389/fgene.2020.00521

Wulff, B. B., and Krattinger, S. G. (2022). The Long Road to Engineering Durable Disease Resistance in Wheat. *Curr. Opin. Biotechnol.* 73, 270–275. doi:10.1016/j.copbio.2021.09.002

Xing, L., Wang, C., Xia, X., He, Z., Chen, W., Liu, T., et al. (2014). Molecular Mapping of Leaf Rust Resistance Gene *LrFun* in Romanian Wheat Line Fundulea 900. *Mol. Breed.* 33, 931–937. doi:10.1007/s11032-013-0007-x

Yuan, C., Singh, R. P., Liu, D., Randhawa, M. S., Huerta-Espino, J., and Lan, C. (2020). Genome-wide Mapping of Adult Plant Resistance to Leaf Rust and

Stripe Rust in CIMMYT Wheat Line Arableu# 1. *Plant Dis.* 104, 1455–1464. doi:10.1094/pdis-10-19-2198-re

Zhang, H., Xia, X., He, Z., Li, X., Li, Z., and Liu, D. (2011). Molecular Mapping of Leaf Rust Resistance Gene *LrBi16* in Chinese Wheat Cultivar Bimai 16. *Mol. Breed.* 28, 527–534. doi:10.1007/s11032-010-9501-6

Zhang, P., Yin, G., Zhou, Y., Qi, A., Gao, F., Xia, X., et al. (2017). QTL Mapping of Adult-Plant Resistance to Leaf Rust in the Wheat Cross Zhou 8425B/Chinese Spring Using High-Density SNP Markers. *Front. Plant Sci.* 8, 793. doi:10.3389/fpls.2017.00793

Zhao, X. L., Zheng, T. C., Xia, X. C., He, Z. H., Liu, D. Q., Yang, W. X., et al. (2013). Molecular Mapping of Leaf Rust Resistance Gene *LrNJ97* in Chinese Wheat Line Neijiang 977671. *Theor. Appl. Genet.* 126, 2141–2147. doi:10.1007/s00122-013-2124-7

Zhao, X. L., Zheng, T. C., Xia, X. C., He, Z. H., Liu, D. Q., Yang, W. X., et al. (2008). Molecular Mapping of Leaf Rust Resistance Gene *LrZH84* in Chinese Wheat Line Zhou 8425B. *Theor. Appl. Genet.* 117, 1069–1075. doi:10.1007/s00122-008-0845-9

Zhou, Y., Ren, Y., Lillemo, M., Yao, Z., Zhang, P., Xia, X., et al. (2014). QTL Mapping of Adult-Plant Resistance to Leaf Rust in a RIL Population Derived from a Cross of Wheat Cultivars Shanghai 3/Catbird and Naxos. *Theor. Appl. Genet.* 127 (9), 1873–1883. doi:10.1007/s00122-014-2346-3

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 Kumar, Jan, Saripalli, Sharma, Mir, Balyan and Gupta. 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: 06 April 2022
doi: 10.3389/fgene.2022.772517

Genotyping-by-Sequencing Based Molecular Genetic Diversity of Pakistani Bread Wheat (*Triticum aestivum* L.) Accessions

Shabbir Hussain[1], Madiha Habib[1], Zaheer Ahmed[2], Bushra Sadia[1], Amy Bernardo[3], Paul St. Amand[3], Guihua Bai[3], Nida Ghori[3], Azeem I. Khan[2], Faisal S. Awan[1] and Rizwana Maqbool[2]*

[1]Center of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan, [2]Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan, [3]USDA, Hard Winter Wheat Genetics Research Unit, Manhattan, KS, United States

OPEN ACCESS

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

Reviewed by:
Ali Raza,
Fujian Agriculture and Forestry University, China
Mohd Anwar Khan,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

***Correspondence:**
Faisal S. Awan
faisal.saeed@uaf.edu.pk

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

Received: 08 September 2021
Accepted: 07 January 2022
Published: 06 April 2022

Citation:
Hussain S, Habib M, Ahmed Z, Sadia B, Bernardo A, Amand PS, Bai G, Ghori N, Khan AI, Awan FS and Maqbool R (2022) Genotyping-by-Sequencing Based Molecular Genetic Diversity of Pakistani Bread Wheat (Triticum aestivum L.) Accessions. Front. Genet. 13:772517. doi: 10.3389/fgene.2022.772517

Spring wheat (*Triticum aestivum* L.) is one of the most imperative staple food crops, with an annual production of 765 million tons globally to feed ~40% world population. Genetic diversity in available germplasm is crucial for sustainable wheat improvement to ensure global food security. A diversity panel of 184 Pakistani wheat accessions was genotyped using 123,596 high-quality single nucleotide polymorphism (SNP) markers generated by genotyping-by-sequencing with 42% of the SNPs mapped on B, 36% on A, and 22% on D sub-genomes of wheat. Chromosome 2B contains the most SNPs (9,126), whereas 4D has the least (2,660) markers. The mean polymorphic information content, genetic diversity, and major allele frequency of the population were 0.157, 0.1844, and 0.87, respectively. Analysis of molecular variance revealed a higher genetic diversity (80%) within the sub-population than among the sub-populations (20%). The genome-wide linkage disequilibrium was 0.34 Mbp for the whole wheat genome. Among the three subgenomes, A has the highest LD decay value (0.29 Mbp), followed by B (0.2 Mbp) and D (0.07 Mbp) genomes, respectively. The results of population structure, principal coordinate analysis, phylogenetic tree, and kinship analysis also divided the whole population into three clusters comprising 31, 33, and 120 accessions in group 1, group 2, and group 3, respectively. All groups were dominated by the local wheat accessions. Estimation of genetic diversity will be a baseline for the selection of breeding parents for mutations and the genome-wide association and marker-assisted selection studies.

Keywords: genotyping-by-sequencing, genetic diversity, population structure, *Triticum aestivum* L., single nucleotide polymorphisms, polymorphic information content, Pakistan

INTRODUCTION

Wheat (*Triticum aestivum* L.) is amongst the most prominent cereal crops in the world. Wheat is cultivated on more than 224 million hectares (Mha) and provides 20% (Bhatta et al., 2017) of dietary nutrients for 40% world population (Ray et al., 2013; Arzani and Ashraf, 2017; Food and Agriculture Organization of the United Nations, 2019). Moreover, 68% of global wheat is directly utilized for human food while the remaining 32% for livestock consumption and other miscellaneous uses such

as biofuel (Mohanty and Swain, 2019). The world population is increasing at the rate of 1.5% per annum and is expected to reach 9.7 billion by 2050 (Lobell et al., 2011). However, a 0.9% annual increase in bread wheat yield currently is not sufficient to ensure global food security (Ray et al., 2013). Pakistan is the fifth densely populated country in the world. Wheat was cultivated on an area of 8,825 thousand ha of land, which produced 24,946 MT of grain in 2020 (Pakistan Economic Survey 2019-20).

Wheat genetic gains can be improved by optimizing crop husbandry (Sener et al., 2009) and harnessing genetic diversity in native germplasm resources (Nielsen et al., 2014; Govindaraj et al., 2015). Dissection of genetic diversity is a prerequisite for plant breeding experiments, such as domestication, inheritance, conservation, and evaluation of wheat germplasm (Peterson et al., 2014). Narrow genetic bases, limited genetic diversity, and continuous reduction in arable farmland, as well as various climate-associated anomalies in the form of biotic and abiotic stresses, pose a continuous threat to world food security in developing countries (Fischer et al., 2014). Moreover, wheat has a long evolutionary history from the days of its early domestication from Einkorn (one of the primitive wheat ancestors) to modern bread wheat and assimilates plenty of genetic variation during this long period of evolution. Farmers' selection, uniform varietal seed production, continuous selfing, use of modern breeding techniques, domestication, and stringent selection pressure lead to prompt genetic erosion and squat gene pool, which causes continuous losses of favorable alleles in currently used wheat germplasm (Haudry et al., 2007; Sofalian et al., 2008).

Strategic trait-based wheat breeding is the most viable and sustainable solution for crop improvement. The success of a wheat breeding program can be directly associated with the availability of valuable genetic diversity in the program (Rufo et al., 2019; Voss-Fels et al., 2019). Genetic resources such as wild relatives, gene bank accessions, landraces, advanced breeding lines, and induced and natural mutants are considered indispensable genetic resources for maintaining genetic diversity and crop improvement (Ogbonnaya et al., 2013; Das et al., 2016). Natural variation selection, wide crossing, new gene introduction, genetic hybridization, induced mutagenesis, and horizontal and vertical gene transfer can contribute to enriching the genetic diversity of modern wheat varieties to meet the challenges of climate change and global food security (Arya et al., 2013).

DNA markers are considered indispensable tools for the genetic characterization of plants. Various types of DNA-based markers, including randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), restriction fragment length polymorphisms (RFLP), and simple sequence repeats (SSRs) (Röder et al., 1995), were successfully used for genetic characterization in plants. Nowadays, single nucleotide polymorphism (SNP) markers generated through next-generation sequencing (NGS,) such as targeted amplicon sequencing (TAS), Illumina bead chip array, DArT, Genotyping-by-Sequencing (GBS), and kompetitive allele-specific PCR (KASP), are gradually replacing other old marker systems due to high throughput and low cost per data point. GBS is a reduced representation sequencing method that identifies SNP for genotyping and discovers new SNP compared to other array-based genotyping technologies, namely, DArT/Illumina bead chip. An array-based genotyping chip targets and identifies a pre-labeled specific number of SNPs markers. GBS-based sequencing could be effectively used in the breeding of complex genome crops without any prior sequencing information or even in the absence of reference genome in many orphan crops (Getachew et al., 2019). GBS is a technology that can reduce genome complexity by using two restriction enzymes (*Pst*I/*Msp*I) and simultaneously discover and genotype genome-wide variations in complex genome crops, namely, bread wheat (Elshire et al., 2011; Poland et al., 2012a; Poland et al., 2012b). Currently, GBS has been successfully used in many crops to unblock the GD including wheat, barley, rice, maize, cassava, potato, and soybean (Elshire et al., 2011).

SNPs are the most abundant polymorphic markers in both plant and animal genomes (Rimbert et al., 2018). The bi-allelic nature, high level of polymorphisms, ubiquitous presence, uniform distribution across genomes, automated data acquisition, and analysis make SNPs the most suitable marker for genome-wide marker analysis (Verma et al., 2015). Quick advancement in NGS techniques significantly improved sequencing throughput and reduced sequencing cost, making the genome-wide SNP analysis a time- and cost-effective tool for genomic studies (He et al., 2014). Therefore, the NGS-based markers have been widely used for retrieving genetic diversity (GD), harnessing population structure (PS), studying linkage disequilibrium (LD), mapping quantitative trait loci (QTL), conducting genome-wide association studies (GWAS), and genomic selection (GS) in various crops (You et al., 2018).

Although the cost of whole-genome sequencing significantly reduced during the last decade, still it is not feasible to completely sequence all genotypes for routine screening of breeding materials especially for those species with a huge genome such as wheat in most breeding programs. Mourad et al. (2020) genotyped a panel of 103 spring wheat accessions that were collected from five continents and deposited in the USDA gene bank to evaluate the GD, PS, and LD patterns. The panel was genotyped with 36,720 high-quality SNPs, and the whole population was divided into three subpopulations on the basis of analysis of molecular variance (AMOVA), structure analysis, kinship, and principal component analysis. No high LD was observed on a whole bread wheat genome, but at the sub-genomic level, the D genome showed the highest LD decay value compared to the A and B sub-genomes.

Alipour et al. (2017) utilized 16,506 polymorphic GBS-SNPs to dissect the GD of an Iranian wheat diversity panel of 369 wheat genotypes. The B subgenome has the highest number of mapped SNPs compared to the other A and D subgenomes, respectively. The whole population was divided into three subgroups: one for cultivars and two for landraces. Heslot et al. (2013) analyzed 38,412 GBS-SNPs in 365 soft winter wheat to harness GD in advanced breeding lines. Yang et al. (2020) characterized a population of 180 bread wheat

genotypes from Asia and Europe to determine the indigenous PS and GD using 24,767 high-quality polymorphic SNPs using the GBS approach and to determine GD of the subjected population. The polymorphic information content (PIC) value of markers ranges from 0.1 to 0.4. Based on cluster and structure analysis, the whole diversity panel was divided into two groups: group 1 comprises European and partial Asian and group 2 consists of the Middle East and partial Asian accessions.

Most Pakistani wheat germplasm has not been characterized using DNA (SNP) markers. Local germplasm is always a key source of resistance against biotic and abiotic anomalies. The primary objective of this study is to investigate the extent and the pattern of GD, PS, LD, PIC, phylogeny, and kinship of the subjected population that will be useful for the selection of breeding parents for various stress breeding strategies through GWAS and association mapping (AM).

MATERIALS AND METHODS

Deoxy Nucleic Acid Isolation and Genotyping-by-Sequencing Library Preparation

A set of 184 Pakistani spring wheat varieties and germplasm accessions were collected from the Wheat Research Institute, Ayub Agricultural Research Institute (AARI), Pakistan, Faisalabad (**Supplementary Table S1**). The panel was sown into two 96-well plastic trays in the greenhouse at the Kansas State University, Manhattan, KS, United States. Three pieces of 15-day-old seedling leaf tissues (2 cm) were collected in 1.1 ml 96-deep-well plates with 3 mm stainless beads in each well and immediately freeze-dried for 2 days. Leaf tissues were ground by shaking the plates at 30 cycles per second for 3 min in a Mixer Mill (Retsch GmbH, Haan, Germany). Genomic DNA was isolated using the standard cetyltrimethylammonium bromide (CTAB) method with a slight modification (Zhao et al., 2020). DNA quality was checked with 1% agarose gel and quantified in a FLUOstar Omega microplate reader (BMG LABTECH, Germany) using Quant-iT™ PicoGreen dsDNA assay kits (Thermo Fisher Scientific, Waltham, MA, United States). The genomic DNA of all samples was normalized at a concentration of 20 ng/μL for preparing the GBS library.

GBS libraries of 184 spring wheat samples were prepared using the standard protocol (Poland et al., 2012a). In brief, normalized DNA (200 ng) was digested using restriction enzyme *Pst*I-HF and *Msp*I from New England BioLabs, Inc. (Ipswich, MA, United States) and ligated with barcoded adapters using a T4 DNA ligase (New England BioLabs, Inc., Ipswich, MA, United States). Ligated DNA fragments were pooled and purified using QIAquick PCR Purification Kit (QIAGEN GmbH, Hilden, Germany). Primers complementary on both adapters were used for PCR amplification. PCR product was cleaned using QIAquick PCR Purification Kit and size selected using an E-gel (Thermo Fisher Scientific, Waltham, MA, United States) to select 200–300 bp fragments and then quantified using Qubit 2.0 Fluorometer and Qubit dsDNA HS Assay Kits (Life Technologies Inc., Carlsbad, CA, United States).

Single Nucleotide Polymorphism Calling and Data Imputation

The final library was sequenced in an Ion Proton next-generation sequencer (Thermo Fisher Scientific, Waltham, MA, United States) in the USDA Central Small Grain Genotyping Lab, Kansas State University, Manhattan, KS, United States. SNPs were called using the GBS discovery pipeline v2.0 in Trait Analysis by Association, Evolution, and Linkage (TASSEL) v5.2.63 (Bradbury et al., 2007) by aligning the sequence reads with the International Wheat Genome Sequencing Consortium (IWGSC) reference genome RefSeq V2.0 (IWGSC, 2018). Initially called SNPs were filtered to remove these SNPs with >20% missing data and <0.01% minor allele frequency (MAF) and then imputed for missing data using BEAGLE v5.1, (Browning and Browning, 2009), in TASSEL v5.2.63.

Genetic Diversity, Polymorphic Information Content, and Analysis of Molecular Variance

TASSEL v5.2.67 was used to calculate the evolutionary relationship among the 184 Pakistani accessions. The dendrogram was constructed using the Neighbor-Joining (NJ) distance method (Saitou and Nei, 1987) in TASSEL. Principal coordinate analysis (PCoA) was implemented using the Euclidean distance method in GenAlEx v6.5 (Peakall and Smouse, 2006). Population GD, PIC, major allele frequency (MF), and percentage heterozygosity (HZ) were determined using POWER MARKER v3.25 (Liu and Muse, 2005). Analysis of molecular variance (AMOVA) (Excoffier et al., 1992) and Shannon's information index (*I*) were calculated using GenAlEx v6.5 (Peakall and Smouse, 2006).

Analyses of Population Structure, Linkage Disequilibrium, and Kinship Analysis

Population structure was inferred using an admixture model and Bayesian model-based clustering algorithm in STRUCTURE v2.3.4 (Pritchard et al., 2000). The best fit delta K value for the number of subpopulations was determined using K values from 1 to 10 and burned in 10,000 generations and 100,000 Markov chain Monte Carlo of (MCMC) iterations (Zorić et al., 2012; Chen et al., 2012). The output of STRUCTURE analysis was visualized using a STRUCTURE HARVESTER software (Earl and vonHoldt, 2012). TASSEL v5.2.67 was used to calculate LD as squared allele frequency (r^2) and physical distance (D) between each pair of SNPs using a sliding window size of 50 with 1,000 permutations. Pairwise LD (r^2) values were plotted against the relative physical distances (D), and a locally weighted polynomial curve regression (LOESS) model was fitted to determine genome-wide LD decay using the R package (https://www.r-project.org/). The LDs were calculated for the whole wheat genome and three sub-genomes (A, B, D) (Hill and Weir 1988; Remington et al., 2001). The critical value of r^2 beyond which LD likely starts to decay was set at r^2

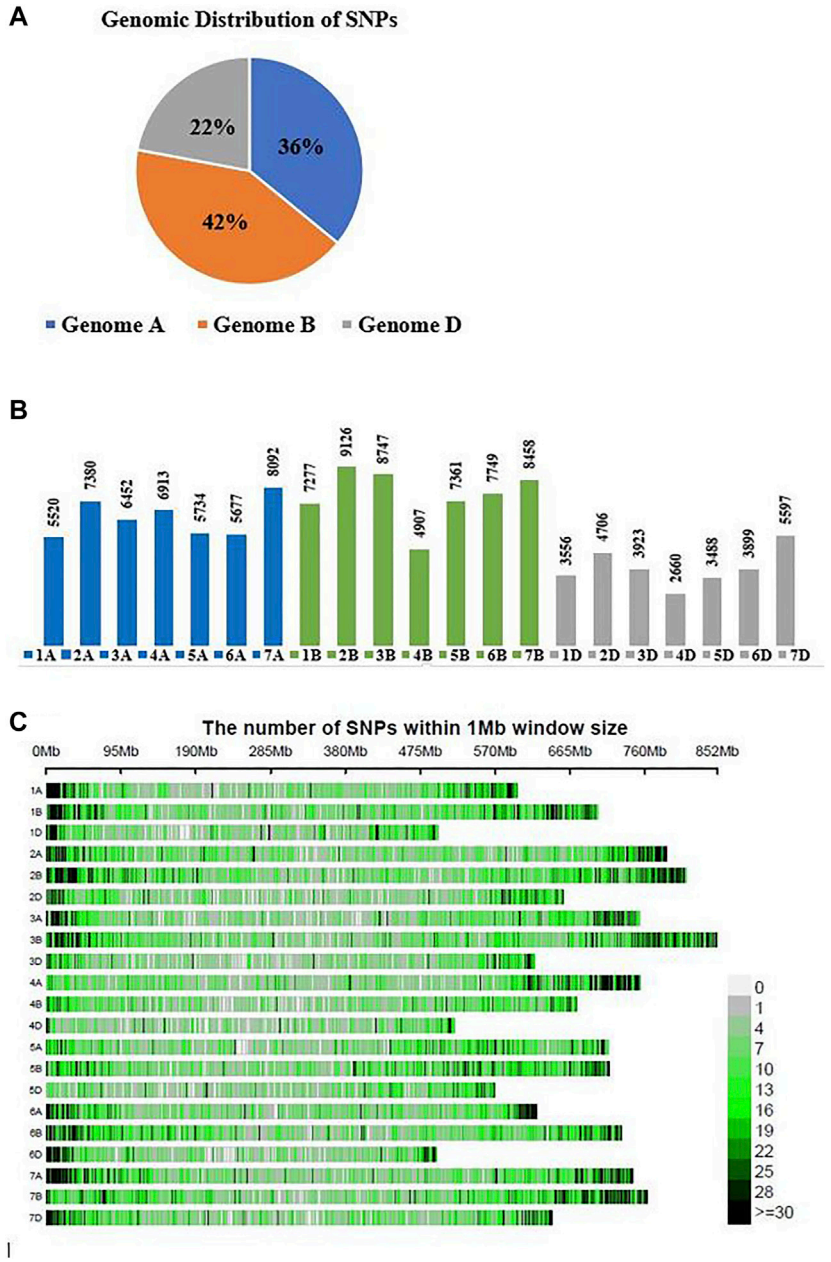


FIGURE 1 | (A) Sub-genomic distribution of 123,596 SNPs on three sub-genomes of spring wheat. **(B)** Chromosomal SNPs distribution on each chromosome of three sub-genomes (A, B, D), **(C)** Number of SNPs within 1 Mb window size.

= 0.1. This critical threshold value of r^2 was estimated using the 95th percentile in the distribution of r^2 below which the relationship among pairs of marker loci is not caused by physical linkage. The intersection of the fitted curve among correlation r^2 and physical distance (D) on a chromosome with this critical threshold was considered the estimated LD range. Population kinship heat matrix among all genotypes was calculated using GAPIT in the R package (Wang et al., 2014).

RESULTS

Genomic Distribution of Single Nucleotide Polymorphisms

The panel of 184 wheat accessions generated 202,147,814 sequence reads and 129,180 SNPs with 80% missing data points. A total of 123,596 polymorphic SNPs were retained after removing low-quality SNPs with MAF <0.01, heterozygote rate >0.2, and imputation using the Chinese

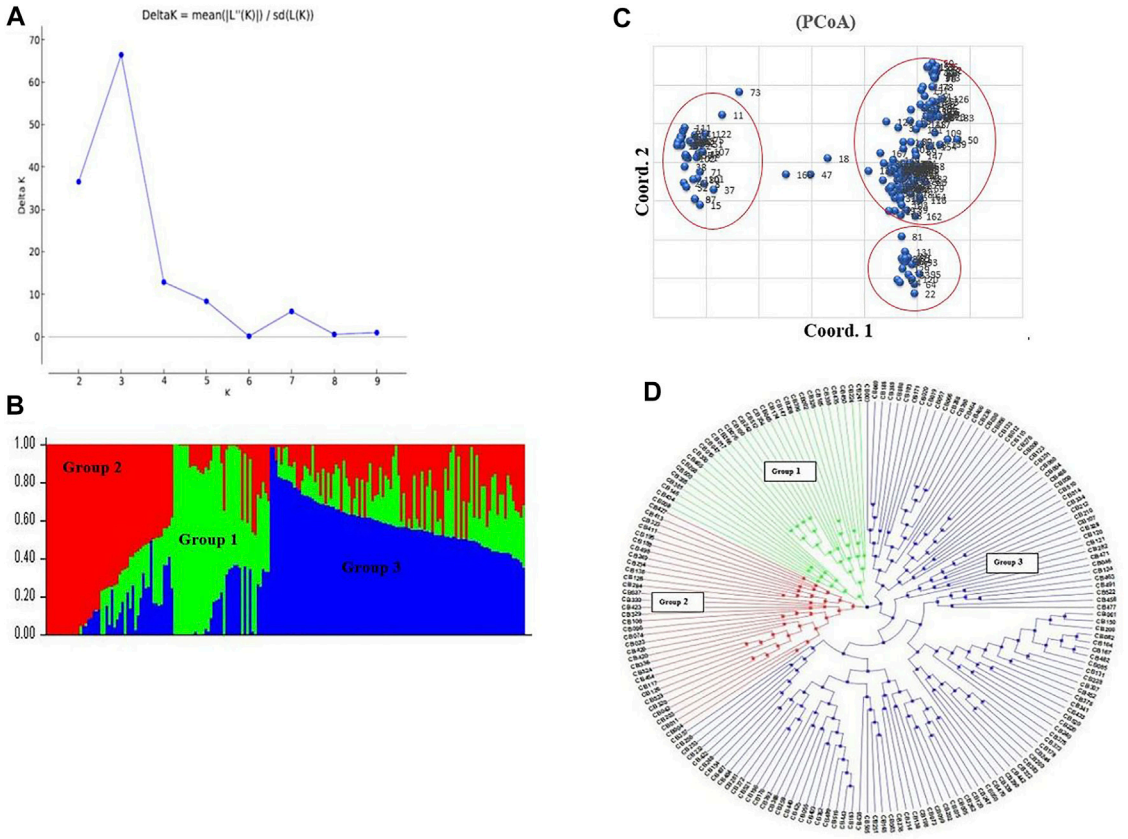


FIGURE 2 | Population structure analysis, **(A)** Delta K line graph, **(B)** structure analysis combine bar chart, **(C)** PCoA analysis of wheat accessions, **(D)** Neighbor Joining phylogenetic tree analysis.

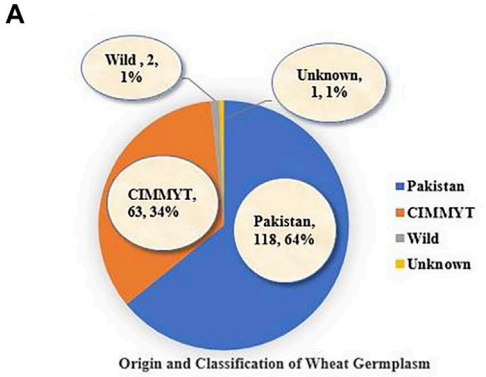
spring wheat reference genome RefSeq V2.0 (IWGSC, 2018) for downstream analysis.

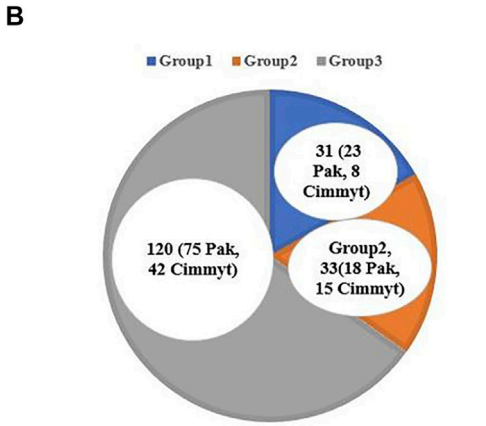
A total of 51,975 SNPs were mapped on the B genome, 44,400 on the A genome, and 27,221 on the D genome (**Figure 1A**). These results indicated that the B sub-genome of bread wheat holds the highest number of mapped markers (SNPs) followed by the A and D sub-genomes. However, the genome sequence sizes of all sub-genome are nearly equal to 5.5 Gb each. Chromosome 2B has the highest number of mapped SNPs (9,126), while chromosome 4D holds the lowest number of mapped SNPs (2,660) (**Figure 1B**). The chromosomal distribution of SNPs within 1 Mb window size is presented in **Figure 1C**, which depicts a higher density of SNPs on chromosomal arms rather than its centromeric region.

Population Structure

A total of 15,779 highly polymorphic SNPs markers with PIC values ranging from 0.30 to 0.37 were used for downstream data analysis. Population structure analysis was performed on 184 Pakistani wheat accessions, distributed into three sub-groups (group 1, group 2, and group 3) based on optimum K = 3. Delta K value gives the optimal number of subpopulations by plotting a graph with the number of clusters (K) (**Figures 2A,B**). Results of the STRUCTURE analysis were further confirmed by the PCoA analysis (**Figure 2C**), which was also supported by neighbor-joining phylogenetic tree analysis (**Figure 2D**).

The panel of 184 accessions was used in this study that includes 118 Pakistani local and 63 accessions from CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo; English: International Maize and Wheat Improvement Center, Mexico), two wild wheat relatives, and one accession with unknown origin (**Figure 3A**). Phylogenetic analysis divided the 184 accessions into three different clusters. Group 1 contained 31 accessions, those dominated by local lines including 23 Pakistani accessions, and only eight CIMMYT accessions. Group 2 consisted of 33 accessions, including 18 Pakistani and 15 from CIMMYT. Group 3 was the largest group that had 120 accessions, including 75 local accessions from Pakistan, 42 from CIMMYT, two wild wheat relatives, and one accession with unknown origin (**Figure 3B**). A total of 120 local wheat accessions, 63 from CIMMYT, two wild relatives of wheat, and one accession of unknown origin were presented with red, black, blue, and purple colors, respectively (**Figure 3C**).





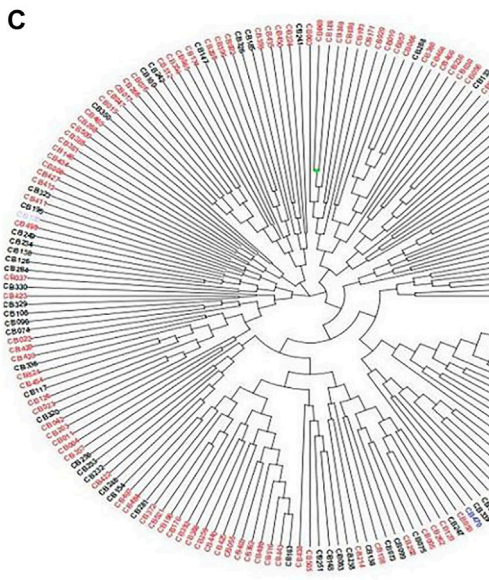


FIGURE 3 | (A) Origin of wheat accessions, **(B)** Classification of accessions in three different groups on the basis of structure analysis and phylogenetic tree analysis **(C)** Distribution of different accession in different sub-groups, the red color showed local genotypes, black color indicated CIMMYT accessions, blue color for wild type and purple color for unknown origin.

Major Allele Frequency, Genetic Diversity, Heterozygosity, Polymorphic Information Content, and Kinship Heat Map Analysis

Major allele frequency (MF), genetic diversity (GD), heterozygosity (HZ), and the polymorphic information content (PIC) of the wheat panel are listed in **Table 1**. The whole-genome mean MF was 0.87, with 0.86, 0.87, and 0.89 for A, B, and D sub-genomes, respectively. Mean GD was observed as 0.20, 0.19, and 0.16 for sub-genomes A, B, and D, respectively. The HZ was 0.07, 0.06, and 0.08 for A, B, and D genomes, respectively, and the PIC value was observed as 0.16, 0.16, and 0.14 for A, B, and D sub-genomes, respectively, with 0.15 cumulative PIC for whole wheat genome. Results of the current experiment showed that subgenomes A and B do not have any significant differences among MF, GD, HZ, and PIC values. However, the D sub-genome showed a higher and significant difference among the MF, HZ, relatively lower GD, and PIC values compared to the A and B sub-genomes (**Table 1**).

AMOVA analysis showed 20 and 80% genetic variation among and within the population, respectively. The fixation index F_{st} was used to genetically differentiate total genetic variability among the sub-populations. A low haploid Nm value (0.497) indicates limited gene flow between the sub-populations (**Table 2**). Shannon's information index (*I*) also reported very low variation among groups than within different groups (**Table 3**). The percentage of Shannon's information index (*I*) and scaled diversity overlapped among and within groups, as presented in **Table 3**. Kinship analysis also divided the panel into three distinct clusters, suggesting a considerable genetic difference among accessions in this panel. The phylogenetic tree is shown on the top, and the left side of the heat map also confirms the results of structure analysis. However, the intensity of color in the heat map also indicated the high LD regions in the heat map (**Figure 4**).

Evolution of Linkage Disequilibrium

The scatter plot of (r^2) and pairwise physical distance (D) revealed that LD decay increases with an increase in marker physical distances on chromosomes or a genome. The average cutoff value $r^2 = 0.1$ was used to determine the LD decay level. The LD decay value (0.34 Mbp) was observed for the whole genome (**Figure 5A**). However, among sub-genomes, the LD decay was the highest on the A subgenome (0.29 Mbp), followed by the B (0.2 Mbp) and D (0.07 Mbp) subgenomes, respectively (**Figures 5B–D**). Chromosome-wide LD of each sub-genome is shown in **Table 4**. The LD decay of the A sub-genome ranges from 0.08 to 0.12, and chromosome 3A showed the highest value of LD (0.12). The LD decay of the B sub-genome ranges from 0.05 to 0.13, and chromosome 3B showed a maximum value of LD for the 3B chromosome. However, in the case of the D sub-genome, the LD value ranged from 0.01 to 0.16, and chromosome 2D showed the highest value (0.16) of LD. A low LD value in this population may be due to higher heterozygosity in the sub-genome reflecting the actual hemizygous state of SNPs markers on the D genome.

TABLE 1 | Mean major allele frequency, genetic diversity, heterozygosity, and polymorphic information content calculated using 123,596 genome-wide SNPs derived from the panel of 184 wheat genotypes.

Chromosome	Sample size	Marker	MF[a]	GD[b]	HZ[c]	PIC[d]
1A	184	5,344	0.8557	0.2118	0.0763	0.1776
2A	184	7,166	0.8628	0.2055	0.0750	0.1733
3A	184	6,253	0.8720	0.1956	0.0713	0.1660
4A	184	6,668	0.8713	0.1949	0.0731	0.1652
5A	184	5,556	0.8730	0.1938	0.0642	0.1648
6A	184	5,496	0.8696	0.1966	0.0704	0.1667
7A	184	7,889	0.8641	0.2025	0.0679	0.1705
Means			**0.8669**	**0.2001**	**0.0712**	**0.1692**
1B	184	7,062	0.8730	0.1954	0.0651	0.1664
2B	184	8,146	0.8592	0.2114	0.0645	0.1783
3B	184	8,513	0.8710	0.1957	0.0634	0.1660
4B	184	4,700	0.8927	0.1721	0.0630	0.1491
5B	184	7,086	0.8621	0.2067	0.0655	0.1744
6B	184	7,549	0.8628	0.2046	0.0658	0.1723
7B	184	8,218	0.8751	0.1900	0.0607	0.1614
Means			**0.8708**	**0.1965**	**0.0640**	**0.1668**
1D	184	3,466	0.8844	0.1779	0.0902	0.1523
2D	184	4,588	0.8873	0.1779	0.0862	0.1534
3D	184	3,847	0.8995	0.1592	0.0839	0.1377
4D	184	2,591	0.9018	0.1571	0.0868	0.1366
5D	184	3,389	0.8983	0.1623	0.0836	0.1407
6D	184	3,850	0.8869	0.1742	0.0902	0.1492
7D	184	5,490	0.9005	0.1567	0.0798	0.1355
Means			**0.8941**	**0.1665**	**0.0858**	**0.1436**
Genome-wide mean			**0.8772**	**0.1843**	**0.0736**	**0.1598**

MF[a]: *Major allele frequency,* ***GD[b]:*** *Genetic Diversity,* ***HZ[c]:*** *Heterozygosity,* ***PIC[d]:*** *Polymorphic Information Content.*

TABLE 2 | Analysis of molecular variance in the panel of 184 genotypes.

Source	df	SS	MS	Est. Var.	%	*p*-value
Among pops	1	16.860	16.860	0.126	20	0.001
Within pops	182	91.502	0.497	0.497	80	0.001
Total	183	108.858	17.357	0.623	100	0.001
Haploid (Nm)	0.49					
Fst	0.67					

Df: degree of freedom, SS: sum of squares, MS: mean sum of squares, %: percentage variation.

Genetic differentiation among and within two subpopulations has been estimated, along with F_{st} gene flow Nm with 9,999 permutations. AP: estimated variance among pops, WP: estimated variance within pops.

***p *value <0.001 (based on 9,999 permutations).*

DISCUSSION

Genetic diversity among the wheat germplasm is exceptionally important for the genetic improvement of wheat cultivars. Wheat widely grows in South Asia (Pakistan, India, and Bangladesh), producing more than 15% of the world's wheat production. Pakistan ranks seventh in world wheat production. However, the genetic diversity of most Pakistani wheat germplasm has not been well characterized using genome-wide SNPs markers. Information on the genetic diversity of the germplasm would be crucial for breeding programs that can play a critical role for wheat breeders to make efficient use of available germplasm resources in crop improvement

TABLE 3 | Population Shannon information index (*I*).

Source of information	Degrees freedom	Log-like Chi-Sq	Shannon inform	Percent of total	Diversity estimate	[0,1] scaled diversity	[0,1] scaled overlap	Estimated probability
	DF	G-test	*I*	Inform	Exp(*I*)	D′	O′ = 1-D′	*p* (rand ≥ data)
Among pops	1	155.425	0.425	8.604	1.529	0.816	0.184	0.923
Within pops	182	1,651.067	4.511	91.396	91.023	0.998	0.002	0.082
Total	183	1806.492	4.936	100.000	139.180	0.998	0.002	

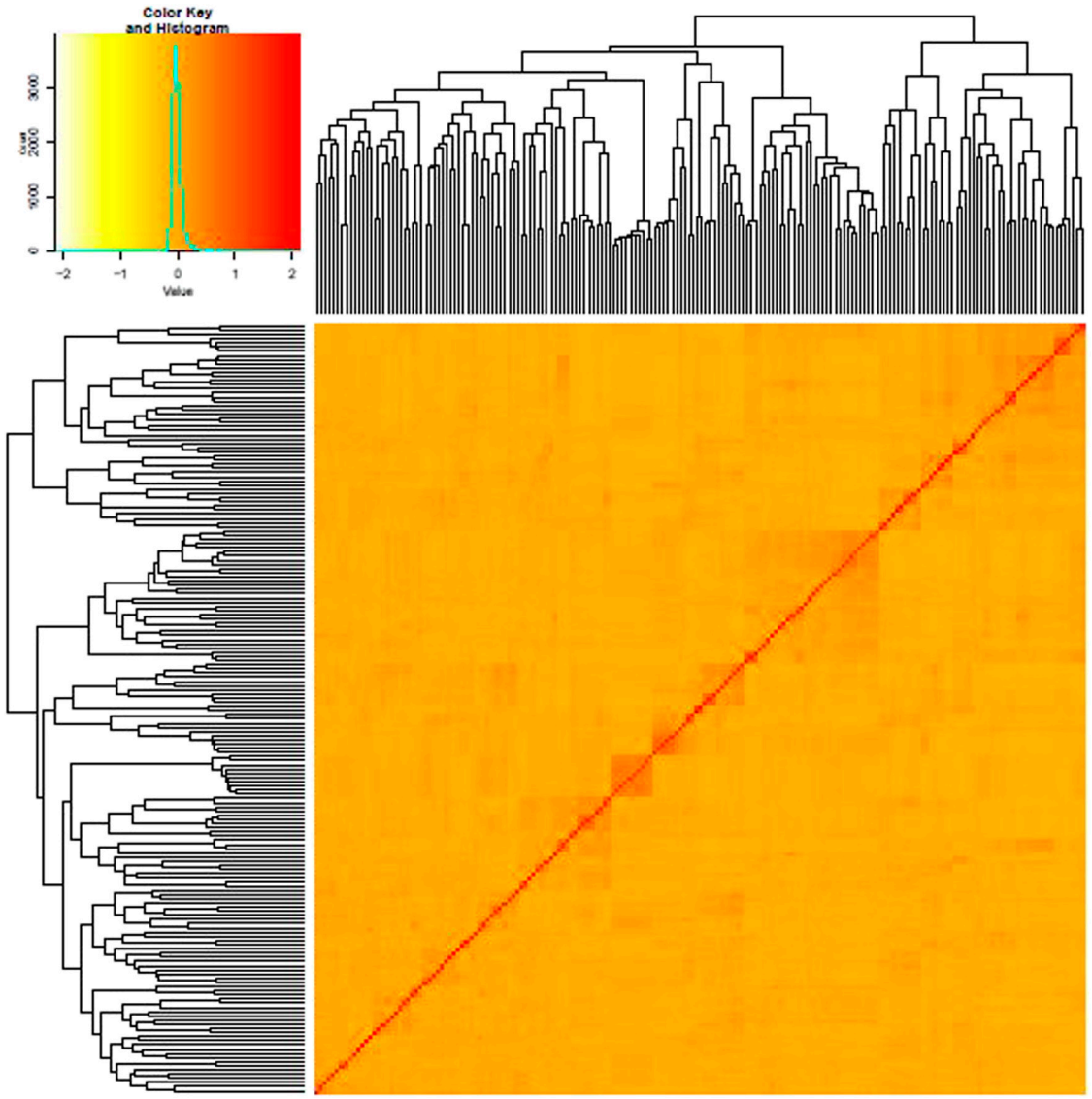


FIGURE 4 | Heat map of kinship matrix with the dendrogram shown on the top-left side based on 123,596 markers.

(Hawkes, 1981). In this study, 184 Pakistani spring wheat accessions were genotyped using genome-wide SNPs generated by GBS technology. Most SNPs were mapped on the B genome (42%), followed by the A (36%) and D (22%) sub-genomes (**Figure 1A**), with the maximum number of SNPs on chromosome 2B (9,126) and the minimum number (2,660) on chromosome 4D (**Figure 1B**). These findings are consistent with previous studies (Alipour et al., 2017; Bhatta et al., 2017; Eltaher et al., 2018; Rufo et al., 2019; Kumar et al., 2020; Yang et al., 2020).

In the current study, the number of SNPs from the D sub-genome (22%) was about half of these present on the B (42%) sub-genome, which agrees with some of the previous reports that the B genome harbors twice as many of the mapped markers as these present on the D sub-genome (Wang et al., 2013; Iehisa et al., 2014; Edae et al., 2015). However, a much lower ratio of the polymorphic markers was found on the D sub-genome, as low as five times compared to the A or B sub-genomes (Allen et al., 2013; Cavanagh et al., 2013). Low polymorphism in the D genome in most of the germplasm was observed not only for SNPs markers but also for RFLP, AFLP, SSR, and DArT markers (Liu and Tsunewaki, 1991; Röder et al., 1998; Peng et al., 2000; Chao et al., 2007; Akhunov et al., 2010; Würschum et al., 2011; Poland et al., 2012a; Berkman et al., 2013; Marcussen et al., 2014; Nielsen et al., 2014; Shavrukov et al., 2014; Edae et al., 2015; Alipour et al., 2017; Eltaher et al., 2018; Rufo et al., 2019). In wheat, A and B are predecessor sub-genomes that accumulated more genetic recombination, duplication,

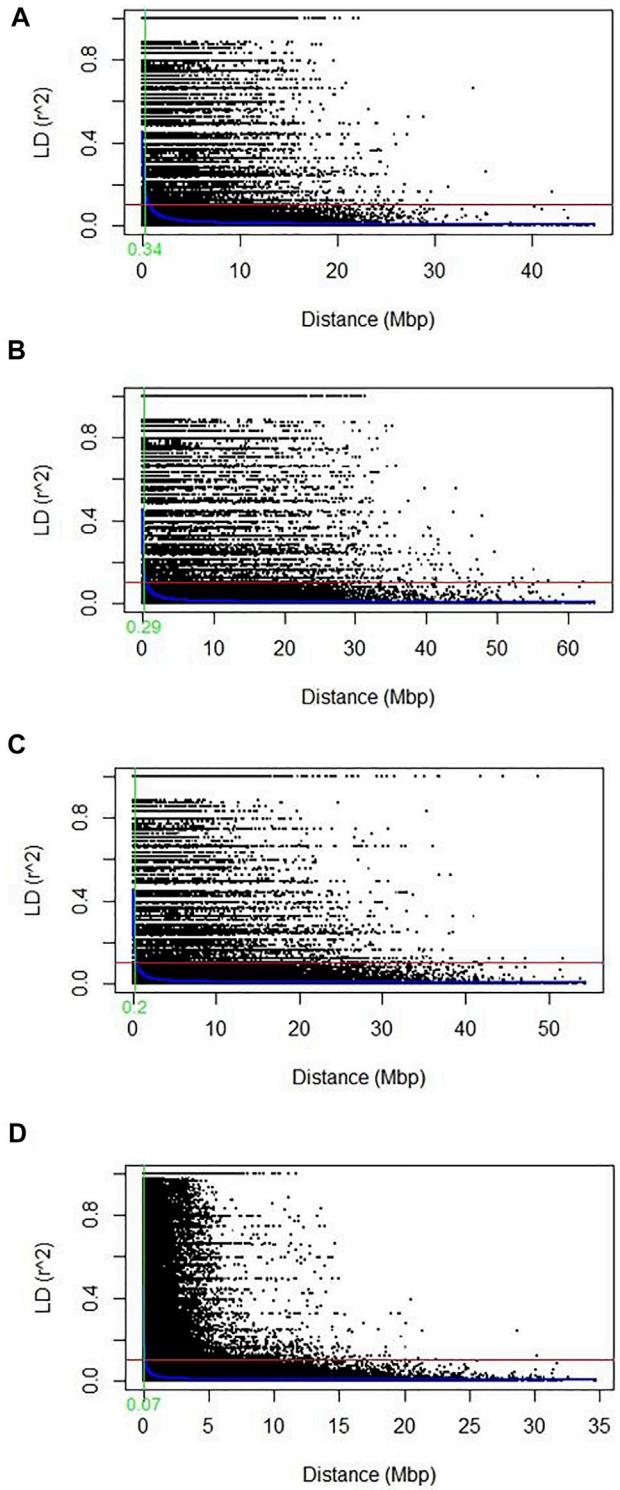


FIGURE 5 | Scatter plot showing the linkage disequilibrium (LD) decay across the genomes for 184 diverse Pakistani bread wheat accessions. The genetic distance in megabase pair (Mbp) plotted against the LD estimate (r^2) for pairs of SNPs. The solid red line showed the threshold LD value at r^2 = 0.1 and the solid green line showed the average LD decay point at which the LD curve intercepts the critical r^2. The LOESS LD decay curve was presented (*Continued*)

FIGURE 5 | with the solid blue curve **(A)** Genome-wide average LD decay plot using all genomes **(B)** LD decay plot of the A genome **(C)** LD decay plot of the B genome; **(D)** LD decay plot of the D genome.

mutation, and gene flow events than those in the relatively newer D subgenome (Berkman et al., 2013). These results suggest that Pakistani wheat genotypes, like the Iranian germplasm (Alipour et al., 2017), hold more genetic diversity on the D genome than other sources of germplasm. Therefore it is a valuable source for crop improvement against various climate anomalies (Jia et al., 2013). Pakistani wheat accessions as *T. aestivum* L. ssp. Sphaerococcum indigenous from northern Pakistan and northwestern India an early flowering, yellow rustresistant, semi-dwarf plant with a semispherical grain shape (Percival, 1922). This notion points out the ancestral relationship between global populations of *Puccinia striiformis* f. sp. tritici (*Pst*) and the putative origin of *Pst* in the Himalayan, as well as their neighboring plains and foothill regions. The presence of putative alleles, high level of diversity, ability to have the sex-related structure in (*Pst*) races, and clear signature of recombination further support this hypothesis (Ali et al., 2014). Archeological remains and evolutionary studies indicated that hexaploid spring wheat was already grown in the region of India and Pakistan somewhat between 4000 and 2000 BC, which indicates that the region (India and Pakistan) has the longest cultivation and interactions history with *Pst*. This statement also indicates that Pakistani wheat accessions were evolved with comparatively more recombination and ultimately enriched source of genetic diversity (Tengberg, 1999; Habib et al., 2020).

PIC is another important parameter for the selection of markers for breeding programs. Botstein et al. (1980) coined a scale range from 0 to 1.0 to categorize multilocus markers such as SSR according to their PIC value. In the current study, PIC values ranged from 0.004 to 0.37, which is low to moderately high with the mean of 0.158 for the whole genome that was smaller than 0.25 reported by previous literature (Chao et al., 2009; Novoselović et al., 2016; Alipour et al., 2017; El-Esawi et al., 2018; Eltaher et al., 2018; Alemu et al., 2020; Mourad et al., 2020). Among the three sub-genomes, the mean PIC for the D sub-genome (0.14) was slightly lower than that of the A or B sub-genomes (0.16) (**Table 1**) due to limited hybridization event and gene flow in *Aegilops tauschii* during evolution (Chao et al., 2009; Lopes et al., 2015; Liu Y. et al., 2017; Eltaher et al., 2018; Tyrka et al., 2021). However, much higher PIC (0.26–0.33) were reported in other studies (Allen et al., 2013; Baloch et al., 2017; Chen et al., 2018; Kumar et al., 2020; Yang et al., 2020).

Gene diversity (expected heterozygosity, He) and PIC are considered as primary measures for the dissection of genetic diversity and shedding light on the mutation rate, as well as the evolutionary pressure on a specific allele in a population over the period of time (Botstein et al., 1980; Shete et al., 2000). Overall GD of a population is mainly explained by the distribution of informative markers on a genome (Nielsen et al., 2014). Genetic diversity provides gene diversity of haploid bi-allelic SNPs and ranges of genetic distance, as well as average heterozygosity among individuals of a population (Nei, 1978). As expected, in

TABLE 4 | Chromosome-wise linkage disequilibrium (LD) of individual chromosome.

	A genome	LD (Mbp)	B genome	LD (Mbp)	D genome	LD (Mbp)
1	**A1**	0.09	**B1**	0.05	**D1**	0.12
2	**A2**	0.09	**B2**	0.01	**D2**	0.16
3	**A3**	0.12	**B3**	0.13	**D3**	0.08
4	**A4**	0.08	**B4**	0.08	**D4**	0.01
5	**A5**	0.11	**B5**	0.11	**D5**	0.06
6	**A6**	0.11	**B6**	0.1	**D6**	0.03
7	**A7**	0.1	**B7**	0.11	**D7**	0.06

the current study, the overall mean GD value was greater than PIC (**Table 1**). In the absence of more polymorphic alleles and ubiquitously even distribution of allele frequency of markers, PIC is always lower than its GD. The PIC values of SNP markers affect the classification of informative markers: highly informative, moderately informative, and slightly informative. Highly informative markers could be used in studying population genetics and GD in various plant species (Salem et al., 2021).

Understanding GD and population structure is a prerequisite to comprehending the genetic variability of germplasm before it can be used in a biotic or abiotic stress breeding program. In the current study, genome-wide mean GD was 0.18, ranging from 0.005 to 0.97 in the panel evaluated, which is slightly lower than 0.26 by Yang et al. (2020) and 0.29 by Mourad et al. (2020) but slightly higher than 0.14 observed by Alipour et al. (2017). The genome-wide heterozygosity (HZ) was 0.0736, ranging from 0 to 0.89. Mean HZ varied greatly among sub-genomes, with the lowest value (0.064) for B and the highest (0.085) for the D sub-genome. A similar trend of HZ values was also reported in other studies (Eltaher et al., 2018; Rimbert et al., 2018; Liu et al., 2019; Kumar et al., 2020).

AMOVA data suggested significant GD within sub-groups (80%), which might result from continuous selection for specific traits under certain environments by wheat breeders. A low level of genetic variability between populations (20%) (**Table 2**) may be due to schematic gene flow in the form of germplasm exchange among different countries and regions (Arora et al., 2014). The high genetic diversity (80%) within sub-groups means that diversity was attributed to variation within population sub-groups. The results of clustering and AMOVA suggested that inter-cluster cultivars crossing may be useful for developing promising agronomic and disease resistance associated traits in local germplasm. Higher diversity within sub-population than among sub-populations was in line with the findings from other studies (Alipour et al., 2017; Bhatta et al., 2018; Eltaher et al., 2018; Kumar et al., 2020). The Nm value is another important parameter to estimate the gene flow among subpopulations. Nm haploid value of 1 or >1 indicates the low rate of gene flow between subpopulations. In our experiment, the Nm haploid value of 0.497 indicates minor gene flow among three sub-populations because the current germplasm collection does not represent a diverse geographical area. The current population was mainly collected from the Punjab province of Pakistan.

LD is a non-random co-segregation of single/multiple loci among the same or different chromosomes of a genome. The magnitude estimation of LD decay helps determine the resolution of association mapping and the number of SNPs needed for effective association studies and marker-assisted selection (Davey et al., 2011). LD extent varies with sub-genomes or even with chromosomes of the same sub-genome and depends on various factors such as recombination rate, allele frequency, genetic drift, population structure, natural selection, and chromosomal rearrangements (Sukumaran et al., 2015; Habib et al., 2020). Higher values of LD among pairs of markers indicate slower decay rates on a genome or a chromosome, suggesting that fewer markers are needed for GWAS. The current study was conducted on a diverse set of 184 Pakistani spring wheat accessions, and the LD decay at 0.34 Mbp for the whole genome, and 0.29 Mbp, 0.2 Mbp, and 0.07 Mbp for all subgenomes A, B, and D, respectively, observed (**Figure 5**). In the current study, the A sub-genome showed the highest value of LD (0.29) which shows relatively fewer markers needed for genome coverage and comparatively slower LD decay due to its evolutionary history of origination. However, at the chromosomal level, chromosome 3A showed the highest value (0.12) (**Table 4**) of LD and chromosome 1A showed the minimum value of LD decay (0.09).

In the case of the B sub-genome chromosomes, 3B and 2B showed maximum and minimum values of LD decay of 0.13 and 0.01, respectively. In the case of D, the sub-genome chromosomes 2D and 4D showed LD values of 0.16 and 0.01, respectively (**Table 4**). Chromosomal level LD decay helps determine the level of decay, LD hotspots, and number of markers needed for genome coverage on various regions of chromosomes. Results of this study were consistent with the findings of Chao et al. (2010) and Ladejobi et al. (2018) reported the LD decay of 6.4, 4.5, 4, and 4.98 Mbp for A, B, D, and whole-genome, respectively. Aleksandrov et al. (2021) performed a study on Bulgarian bread wheat germplasm to dissect the population structure and linkage disequilibrium on two populations, including one population of old germplasm and one set of modern semi-dwarf cultivars. Old germplasm showed an LD decay of 3.6, 3.3, 3.1, and 3.3 Mb in subgenomes A, B, and D and the whole-genome, respectively. These results support the results of the current study. Li et al. (2019) reported the lowest LD decay value at the D genome, such as 0.35, 0.75, and 0.25 Mb at the A, B, and D subgenomes, respectively. However, the highest LD value was observed in the B sub-genome of the population. Zhang et al. (2013) reported that the highest LD decay rate for the A sub-genome ranges from 25 to 30 cM and on the B and D sub-genomes, LD decay ranges from 15 to 20 cM). Sehgal et al. (2017) reported the LD decay of

values of 10, 4, 8, and 5 cM for A, B, D, and whole-genome of the wheat population, respectively. These results also confirm our results of the highest LD decay at the A sub-genome.

Usually, the highest LD value was detected in the D genome in some previous reports (Edae et al., 2015; Sukumaran et al., 2015; Liu W. et al., 2017; Ayana et al., 2018; Bhatta et al., 2018; Jamil et al., 2019; Sandeep et al., 2020). The higher LD value for the D sub-genome might be due to limited and shallow infusion of *Aegilops tauschii* with tetraploid wheat in evolutionary history. The slow LD decay rate in the D sub-genome suggests that more markers may be needed for GWAS, association mapping, and marker-assisted selection (MAS) than those that are needed for the A and B sub-genomes. High LD value in the A sub-genome indicates a few markers needed to cover the genome.

CONCLUSION

Analysis of population structure, PIC, LD, PCoA, and NJ phylogenetic tree was used to dissect the genetic diversity of 184 Pakistani genotypes. The whole wheat population was divided into three distinct subpopulations, with group 1 and group 3 dominated by the local lines and group 2 somewhat intermediate type mixture of local and CIMMYT lines. Most of the local genotypes were distributed in all three sub-groups with CIMMYT lines because local accessions were developed by crossing one or both CIMMYT lines as breeding parents. This shows that a huge amount of exotic blood mixing among local germplasm leads toward the ample amount of genetic diversity in local germplasm. Pakistan continuously faces problems such as sudden heatwaves, unusual rainfall, water shortage, and unexpected weather changes due to prompt climate change. Most Pakistani spring wheat genotypes are prone to yellow stripe rust because pathogens get resistant to hot, humid conditions due to unexpected rainfall at the vegetative stage. The results of population structure, phylogenetic tree based genetic distribution, moderately informative nature, and the number of SNP markers, PIC value, and LD regions of this population could be used for the detection of genetic diversity in the local germplasm, and this high genetic diversity would be used for association mapping studies and the selection of crossing parents in national wheat breeding programs. The results of the current genetic diversity study provided the information necessary to broaden the genetic bases and conservation studies. This information will urge national wheat breeders to speed up wheat breeding by using the genetic diversity of local and exotic germplasm. The results of the current study will help dissect dissecting genetic diversity in local germplasm, as well as for GWAS-related studies, to improve the local wheat germplasm of the country.

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 SRA; PRJNA783303.

AUTHOR CONTRIBUTIONS

SH and MH were involved in manuscript witting and data analysis. SH, FA, BS, AK, RM, and ZA were involved in the planning and execution of the experiment. SH, AB, PA, and NG were involved in sample preparation, DNA extraction, SNP library preparation, and GBS sequencing. SH, GB, FA, and RM were involved in reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was funded by the Higher Education Commission (HEC) of Pakistan through the International Research Support Initiative Program (IRSIP) (Grant No: 1-8/HEC/HRD/2019/8749).

ACKNOWLEDGMENTS

The authors thank the Wheat Research Institute, Ayub Agricultural Research Institute (AARI), and Faisalabad for providing germplasm for this experiment. They are especially thankful to the USDA Central Small Grain Hard Winter Wheat Genotyping Lab, Kansas State University, Manhattan, KS, United States, for GBS and molecular data analysis facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.772517/full#supplementary-material

REFERENCES

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., et al. (2010). Nucleotide Diversity Maps Reveal Variation in Diversity Among Wheat Genomes and Chromosomes. *BMC Genomics* 11, 1–22. doi:10.1186/1471-2164-11-702

Aleksandrov, V., Kartseva, T., Alqudah, A. M., Kocheva, K., Tasheva, K., Börner, A., et al. (2021). Genetic Diversity, Linkage Disequilibrium and Population Structure of Bulgarian Bread Wheat Assessed by Genome-wide Distributed SNP Markers: From Old Germplasm to Semi-Dwarf Cultivars. *Plants* 10, 1116. doi:10.3390/plants10061116

Alemu, A., Feyissa, T., and Letta, T. (2020). Abeyo, B. Genetic Diversity and Population Structure Analysis Based on the High Density SNP Markers in Ethiopian Durum Wheat (*Triticum Turgidum Ssp. Durum*). *BMC Genet.* 21, 1–2. doi:10.1186/s12863-020-0825-x

Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., et al. (2014). Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia Striiformis F. Sp. Tritici. *Plos. Pathog.* 10, e1003903. doi:10.1371/journal.ppat.1003903

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., and Zhang, G. (2017). Genotyping-by-sequencing (GBS) Revealed Molecular Genetic Diversity of Iranian Wheat Landraces and Cultivars. *Front. Plant Sci.* 8, 1293. doi:10.3389/fpls.2017.01293

Allen, A. M., Barker, G. L., Wilkinson, P., Burridge, A., Winfield, M., Coghill, J., et al. (2013). Discovery and Development of Exome-based, Co-dominant Single Nucleotide Polymorphism Markers in Hexaploid Wheat (*Triticum aestivum L.*). *Plant Biotechnol. J.* 11, 279–295. doi:10.1111/pbi.12009

Arora, A., Kundu, S., Dilbaghi, N., Sharma, I., and Tiwari, R. (2014). Population Structure and Genetic Diversity Among Indian Wheat Varieties Using Microsatellite (SSR) Markers. *Aust. J. Crop Sci.* 8, 9. doi:10.1080/15427528.2020.1818342

Arya, S., Mishra, D. K., and Bornare, S. S. (2013). Screening Genetic Variability in advance Lines for Drought Tolerance of Bread Wheat (*Triticum aestivum*). *Biosc* 8, 1193–1196.

Arzani, A., and Ashraf, M. (2017). Cultivated Ancient Wheats (*Triticum spp.*): A Potential Source of Health-beneficial Food Products. *Compr. Revi. Food Sci. Food Saf.* 16, 477–488. doi:10.1111/1541-4337.12262

Ayana, G. T., Ali, S., Sidhu, J. S., Gonzalez Hernandez, J. L., Turnipseed, B., and Sehgal, S. K. (2018). Genome-wide Association Study for Spot Blotch Resistance in Hard winter Wheat. *Front. Plant Sci.* 9, 926. doi:10.3389/fpls.2018.00926

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., Sáenz de Miera, E., Aasim, M., et al. (2017). A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from central fertile-crescent. *Plos one* 12, e0167821. doi:10.1371/journal.pone.0167821

Berkman, P. J., Visendi, P., Lee, H. C., Stiller, J., Manoli, S., Lorenc, M. T., et al. (2013). Dispersion and Domestication Shaped the Genome of Bread Wheat. *Plant Biotechnol. J.* 11, 564–571. doi:10.1111/pbi.12044

Bhatta, M., Regassa, T., Rose, D. J., Baenziger, P. S., Eskridge, K. M., Santra, D. K., et al. (2017). Genotype, Environment, Seeding Rate, and Top-dressed Nitrogen Effects on End-use Quality of Modern Nebraska winter Wheat. *J. Sci. Food Agr.* 97, 5311–5318. doi:10.1002/jsfa.8417

Bhatta, M., Morgounov, A., Belamkar, V., Poland, J., and Baenziger, P. S. (2018). Unlocking the Novel Genetic Diversity and Population Structure of Synthetic Hexaploid Wheat. *BMC Genomics* 19, 1–12. doi:10.1186/s12864-018-4969-2

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* 32, 314.

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, 26332635. doi:10.1093/bioinformatics/btm308

Browning, B. L., and Browning, S. R. (2009). A Unified Approach to Genotype Imputation and Haplotype-phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Ameri. J. Hum. Genet.* 84, 210–223. doi:10.1016/j.ajhg.2009.01.005

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., and Kiani, S. (2013). Genome-wide Comparative Diversity Uncovers Multiple Targets of Selection for Improvement in Hexaploid Wheat Landraces and Cultivars. *P. Nat. A. Sci.* 110, 8057–8062. doi:10.1073/pnas.1217133110

Chao, S., Zhang, W., Dubcovsky, J., and Sorrells, M. (2007). Evaluation of Genetic Diversity and Genome-wide Linkage Disequilibrium Among U.S. Wheat (*Triticum aestivum L.*) Germplasm Representing Different Market Classes. *Crop Sci.* 47, 1018–1030. doi:10.2135/cropsci2006.06.0434

Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., and Luo, M. C. (2009). Analysis of Gene-Derived SNP Marker Polymorphism in US Wheat (*Triticum aestivum L.*) Cultivars. *Mol. Breed.* 23, 23–33. doi:10.1007/s11032-008-9210-6

Chao, S., Dubcovsky, J., Dvorak, J., Luo, M. C., Baenziger, S. P., and Matnyazov, R. (2010). Population-and Genome-specific Patterns of Linkage Disequilibrium and SNP Variation in spring and winter Wheat (*Triticum aestivum L.*). *BMC Genomics* 11, 1–7. doi:10.1186/1471-2164-11-727

Chen, X., Min, D., Yasir, T. A., and Hu, Y. G. (2012). Genetic Diversity, Population Structure and Linkage Disequilibrium in Elite Chinese winter Wheat Investigated with SSR Markers. *PLoS One* 7, e44510. doi:10.1371/journal.pone.0044510

Chen, T., Tantasawat, P. A., Wang, W., Gao, X., and Zhang, L. (2018). Population Structure of Chinese Southwest Wheat Germplasms Resistant to Stripe Rust and Powdery Mildew Using the DArT-Seq Technique. *Ciência Rural* 16, 48. doi:10.1590/0103-8478cr20160066

Das, M. K., Bai, G., Mujeeb-Kazi, A., and Rajaram, S. (2016). Genetic Diversity Among Synthetic Hexaploid Wheat Accessions (*Triticum aestivum*) with Resistance to Several Fungal Diseases. *Genet. Resour. Crop Ev.* 63, 1285–1296. doi:10.1007/s10722-015-0312-9

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide Genetic Marker Discovery and Genotyping Using Next-Generation Sequencing. *Nat. Rev. Genet.* 12, 499–510. doi:10.1038/nrg3012

Earl, D. A., and VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Methods. *Conser. Genet. Resour.* 4 (2), 359–361. doi:10.1007/s12686-011-9548-7

Edae, E. A., Bowden, R. L., and Poland, J. (2015). Application of Population Sequencing (POPSEQ) for Ordering and Imputing Genotyping-By-Sequencing Markers in Hexaploid Wheat. *G3: Gen. Genom Genet.* 5, 2547–2553. doi:10.1534/g3.115.020362

El-Esawi, M. A., Witczak, J., Abomohra, A. E., Ali, H. M., and Elshikh, M. S. (2018). Analysis of the Genetic Diversity and Population Structure of Austrian and Belgian Wheat Germplasm within a Regional Context Based on DArT Markers. *Genes* 9, 47. doi:10.3390/genes9010047

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

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2018). Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Front. Genet.* 9, 1–9. doi:10.3389/fgene.2018.00076

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of Molecular Variance Inferred from Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 131, 479–491. doi:10.1093/genetics/131.2.479

Fischer, R. A., Byerlee, D., and Edmeades, G. (2014). *Crop Yields and Global Food Security*. Canberra, ACT: ACIAR, 8–11.

Food and Agriculture Organization of the United Nations (2019). World Food Situation. Available at: http://www.fao.org/worldfoodsituation/csdb/en/ (Accessed August 19, 2020).

Getachew, S. E., Ngalle, H. B., Joseph, M. B., and Wosene, G. (2019). Genotyping by Sequencing for Plant Breeding- A Review. *Adv. Biotechnol. Microbiol.* 14, 555891. doi:10.19080/AIBM.2019.14.555891

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, 431487. doi:10.1155/2015/431487

Habib, M., Awan, F. S., Sadia, B., and Zia, M. A. (2020). Genome-Wide Association Mapping for Stripe Rust Resistance in Pakistani Spring Wheat Genotypes. *Plants* 9, 1056. doi:10.3390/plants9091056

Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., et al. (2007). Grinding up Wheat: a Massive Loss of Nucleotide Diversity since Domestication. *Mol. Bio. Evol.* 24, 1506–1517. doi:10.1093/molbev/msm077

Hawkes, J. G. (1981). "Germplasm Collection, Preservation, and Use," in *2 Plant Breeding Symposium* (Iowa: Iowa State University Press).

He, J., Zhao, X., Andre, L., Lu, Z. H., HongKui, L., and Ziqin, L. (2014). Genotyping-by-Sequencing (GBS), an Ultimate Marker-Assisted Selection (MAS) Tool to Accelerate Plant Breeding. *Front. Plant Sci.* 5, 484. doi:10.3389/fpls.2014.00484

Heslot, N., Rutkoski, J., Poland, J., Jannink, J. L., and Sorrells, M. E. (2013). Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. *PloS one* 8, e74612. doi:10.1371/journal.pone.0074612

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

Iehisa, J. C., Shimizu, A., Sato, K., Nishijima, R., Sakaguchi, K., Matsuda, R., et al. (2014). Genome-wide Marker Development for the Wheat D Genome Based on Single Nucleotide Polymorphisms Identified from Transcripts in the Wild Wheat Progenitor Aegilops Tauschii. *Theor. Appl. Genet.* 127, 261–271. doi:10.1007/s00122-013-2215-5

IWGSC (2018). Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome. *Science* 361, 7191. doi:10.1126/science.aar7191

Jamil, M., Ali, A., Gul, A., Ghafoor, A., Napar, A. A., Ibrahim, A. M., et al. (2019). Genome-wide Association Studies of Seven Agronomic Traits under Two Sowing Conditions in Bread Wheat. *BMC Plant Biol.* 19, 1–8. doi:10.1186/s12870-019-1754-6

Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., et al. (2013). International Wheat Genome Sequencing, Aegilops Tauschii Draft Genome Sequence Reveals a Gene Repertoire for Wheat Adaptation. *Nature* 496, 91–95. doi:10.1038/nature12028

Kumar, S., Kumari, J., Bhusal, N., Pradhan, A. K., Budhlakoti, N., Mishra, D. C., et al. (2020). Genome-wide Association Study Reveals Genomic Regions Associated with Ten Agronomical Traits in Wheat under Late-Sown Conditions. *Front. Plant Sci.* 11, 1420. doi:10.3389/fpls.2020.549743

Ladejobi, O., Mackay, I. J., Poland, J., Praud, S., Hibberd, J/M., and Bentley, A. R. (2018). Reference Genome Anchoring of High-Density Markers for Association Mapping and Genomic Prediction in European winter Wheat. *Front. Plant Sci.* 10, 1278. doi:10.3389/fpls.2019.01278

Li, L., Peng, Z., Mao, X., Wang, J., Chang, X., Reynolds, M., et al. (2019). Genome-wide Association Study Reveals Genomic Regions Controlling Root and Shoot Traits at Late Growth Stages in Wheat. *Ann. Bot.* 124, 993–1006. doi:10.1093/aob/mcz041

Liu, K., and Muse, S. V. (2005). PowerMaker: an Integrated Analysis Environment for Genetic Maker Analysis. *Bioinformatics* 21, 2128–2129. doi:10.1093/bioinformatics/bti282

Liu, Y. G., and Tsunewaki, K. (1991). Restriction Fragment Length Polymorphism (RFLP) Analysis in Wheat. II. Linkage Maps of the RFLP Sites in Common Wheat. *Jpn. J. Genet.* 66, 617–633. doi:10.1266/jjg.66.617

Liu, J., Rasheed, A., He, Z., Imtiaz, M., Arif, A., Mahmood, T., et al. (2019). Genome-wide Variation Patterns between Landraces and Cultivars Uncover Divergent Selection during Modern Wheat Breeding. *Theor. Appl. Genet.* 132, 2509–2523. doi:10.1093/aob/mcz04110.1007/s00122-019-03367-4

Liu, W., Maccaferri, M., Chen, X., Laghetti, G., Pignone, D., Pumphrey, M., et al. (2017). Genome-wide Association Mapping Reveals a Rich Genetic Architecture of Stripe Rust Resistance Loci in Emmer Wheat (*Triticum Turgidum Ssp. Dicoccum*). *Theor. Appl. Genet.* 130, 2249–2270. doi:10.1007/s00122-017-2957-6

Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., et al. (2017). A Genome-wide Association Study of 23 Agronomic Traits in Chinese Wheat Landraces. *Plant J.* 91, 861–873. doi:10.1111/tpj.13614

Lobell, D. B., Bänziger, M., Magorokosho, C., and Vivek, B. (2011). Nonlinear Heat Effects on African maize as Evidenced by Historical Yield Trials. *Nat. Clim. Change.* 1, 42–45. doi:10.1038/nclimate1043

Lopes, M. S., Dreisigacker, S., Peña, R. J., Sukumaran, S., and Reynolds, M. P. (2015). Genetic Characterization of the Wheat Association Mapping Initiative (WAMI) Panel for Dissection of Complex Traits in spring Wheat. *Theor. Appl. Geneti.* 128, 453–464. doi:10.1007/s00122-014-2444-2

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., et al. (2014). Ancient Hybridizations Among the Ancestral Genomes of Bread Wheat. *Science* 345, 6194. doi:10.1126/science.1250092

Mohanty, S. K., and Swain, M. R. (2019). "Bioethanol Production from Corn and Wheat: Food, Fuel, and Future," in *Bioethanol Production from Food Crops* (Academic Press), 45–59. doi:10.1016/b978-0-12-813766-6.00003-5

Mourad, A. M., Belamkar, V., and Baenziger, P. S. (2020). Molecular Genetic Analysis of spring Wheat Core Collection Using Genetic Diversity, Population Structure, and Linkage Disequilibrium. *BMC Genomics* 21, 1–2. doi:10.1186/s12864-020-06835-0

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics* 89, 583–590. doi:10.1093/genetics/89.3.583

Nielsen, N. H., Backes, G., Stougaard, J., Andersen, S. U., and Jahoor, A. (2014). Genetic Diversity and Population Structure Analysis of European Hexaploid Bread Wheat (*Triticum aestivum L.*) Varieties. *PloS ONE* 9, e94000. doi:10.1371/journal.pone.0094000

Novoselović, D., Bentley, A. R., Šimek, R., Dvojković, K., Sorrells, M/E., Gosman, N., et al. (2016). Characterizing Croatian Wheat Germplasm Diversity and Structure in a European Context by DArT Markers. *Front. Plant Sci.* 7, 184. doi:10.3389/fpls.2016.00184

Ogbonnaya, F. C., Abdalla, O., Mujeeb-Kazi, A., Kazi, A. G., Xu, S. S., Gosman, N., et al. (2013). Synthetic Hexaploids: Harnessing Species of the Primary Gene Pool for Wheat Improvement. *Plant Breed. Rev.* 37, 35–122. doi:10.1002/9781118497869.ch2

Peakall, R. O., and Smouse, P. E. (2006). GENALEX 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes.* 6, 288–295. doi:10.1111/j.1471-8286.2005.01155.x

Peng, J., Korol, A. B., Fahima, T., Röder, M. S., Ronin, Y. I., Li, Y. C., et al. (2000). Molecular Genetic Maps in Wild Emmer Wheat, Triticum Dicoccoides: Genome-wide Coverage, Massive Negative Interference, and Putative Quasi-Linkage. *Genome Res.* 10, 1509–1531. doi:10.1101/gr.150300

Percival, J. (1922). *The Wheat Plant; A Monograph*. Duckworth and Company. doi:10.1038/109366a0

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, 665–680. doi:10.3390/d6040665

Poland, J. A., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012a). Genomic Selection in Wheat Breeding Using Genotyping-By-Sequencing. *Plant Genome* 5, 103–113. doi:10.3835/plantgenome2012.06.0006

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012b). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-By-Sequencing Approach. *PloS One* 7, e32253. doi:10.1371/journal.pone.0032253

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

Ray, D. K., Mueller, N. D., West, P. C., and Jonathan, A. F. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One* 8, 1–8. doi:10.1371/journal.pone.0066428

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. *P. Nat. A. Sci.* 98, 11479–11484. doi:10.1073/pnas.201394398

Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., et al. (2018). High Throughput Snp Discovery and Genotyping in Hexaploid Wheat. *PLoS One* 13, 1–9. doi:10.1371/journal.pone.0186329

Röder, M. S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M. H., Leroy, P., et al. (1998). A Microsatellite Map of Wheat. *Genetics* 149, 2007–2023. doi:10.1093/genetics/149.4.2007

Röder, M. S., Plaschke, J., König, S. U., Börner, A., Sorrells, M. E., Tanksley, S. D., et al. (1995). Abundance, Variability and Chromosomal Location of Microsatellites in Wheat. *Mol. Gen. Genet.* 246 (3), 327–333. doi:10.1007/BF00288605

Rufo, R., Alvaro, F., Royo, C., and Soriano, J. M. (2019). From Landraces to Improved Cultivars: Assessment of Genetic Diversity and Population Structure of Mediterranean Wheat Using SNP Markers. *PloS One* 14, e0219867. doi:10.1371/journal.pone.0219867

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method: a New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Salem, K. F., Elabsawy, E. A., Khidr, Y. A., and Elshamy, H. (2021). Assessment Genetic Diversity Among Some Egyptian Bread Wheat (*Triticum aestivum L.*) Cultivars Using RAPD Markers. *J. Genet. Environ. Reso. Conser.* 26, 168–175.

Sandeep, C., Kumar, A., Rodrigues, V., Viswanath, S., Shukla, A. K., and Sundaresan, V. (2020). Morpho-genetic Divergence and Population Structure in Indian *Santalum Album L. Trees* 34, 1113–1129. doi:10.1007/s00468-020-01963-2

Sehgal, D., Autrique, E., Singh, R., Ellis, M., Singh, S., and Dreisigacker, S. (2017). Identification of Genomic Regions for Grain Yield and Yield Stability and Their Epistatic Interactions. *Sci. Rep.* 7, 1–2. doi:10.1038/srep41578

Sener, O., Arslan, M., Soysal, Y., and Erayman, M. (2009). Estimates of Relative Yield Potential and Genetic Improvement of Wheat Cultivars in the Mediterranean Region. *J. Agri. Sci.* 147, 323–332. doi:10.1017/S0021859609008454

Shavrukov, Y., Suchecki, R., Eliby, S., Abugalieva, A., Kenebayev, S., and Langridge, P. (2014). Application of Next-Generation Sequencing Technology to Study Genetic Diversity and Identify Unique SNP Markers in Bread Wheat from Kazakhstan. *BMC Plant Biol.* 14, 1–3. doi:10.1186/s12870-014-0258-7

Shete, S., Tiwari, H., and Elston, R. C. (2000). On Estimating the Heterozygosity and Polymorphism Information Content Value. *Theor. Popul. Biol.* 57, 265–271. doi:10.1006/tpbi.2000.1452

Sofalian, O., Chaparzadeh, N., Javanmard, A., and Hejazi, M. S. (2008). Study the Genetic Diversity of Wheat Landraces from Northwest of Iran Based on ISSR Molecular Markers. *Int. J. Agri. Bio.* 10, 466–468.

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide Association Study for Grain Yield and Related Traits in an Elite spring Wheat Population Grown in Temperate Irrigated Environments. *Theor. Appl. Genet.* 128, 353–363. doi:10.1007/s00122-014-2435-3

Tengberg, M. (1999). Crop Husbandry at Miri Qalat Makran, SW Pakistan (4000–2000 BC). *Veg. Hist. Archaeobot.* 8, 3–12. doi:10.1007/bf02042836

Tyrka, M., Mokrzycka, M., Bakera, B., Tyrka, D., Szeliga, M., Stojałowski, S., et al. (2021). Evaluation of Genetic Structure in European Wheat Cultivars and Advanced Breeding Lines Using High-Density Genotyping-By-Sequencing Approach. *BMC Genomics* 22, 1–7. doi:10.1186/s12864-020-07351-x

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 (GBS). *Sci. Rep.* 5, 1–4. doi:10.1038/srep17512

Voss-Fels, K. P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., et al. (2019). Breeding Improves Wheat Productivity under Contrasting Agrochemical Input Levels. *Nat. Plants* 5, 706–714. doi:10.1038/s41477-019-0445-5

Wang, N., Thomson, M., Bodles, W. J., Crawford, R. M., Hunt, H. V., Featherstone, A. W., et al. (2013). Genome Sequence of dwarf Birch (Betula Nana) and Cross-species RAD Markers. *Mol. Ecol.* 22, 3098–3111. doi:10.1111/mec.12131

Wang, Q., Tian, F., Pan, Y., Buckler, E., and Zhang, Z. (2014). A SUPER Powerful Method for Genome Wide Association Study. *PloS One* 9, e107684. doi:10.1371/journal.pone.0107684

Würschum, T., Maurer, H. P., Kraft, T., Janssen, G., Nilsson, C., and Reif, J. C. (2011). Genome-wide Association Mapping of Agronomic Traits in Sugar Beet. *Theor. Appl. Genet.* 123, 1121–1131. doi:10.1007/s00122-011-1653-1

Yang, X., Tan, B., Liu, H., Zhu, W., Xu, L., Wang, Y., et al. (2020). Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing. *Front. Genet.* 11, 1157. doi:10.3389/fgene.2020.580782

You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. (2018). Development and Applications of a High Throughput Genotyping Tool for Polyploid Crops: Single Nucleotide Polymorphism (SNP) Array. *Front. Plant Sci.* 9, 104. doi:10.3389/fpls.2018.00104

Zhang, K., Wang, J., Zhang, L., Rong, C., Zhao, F., Peng, T., et al. (2013). Association Analysis of Genomic Loci Important for Grain Weight Control in Elite Common Wheat Varieties Cultivated with Variable Water and Fertiliser Supply. *PloS one* 8, e57853. doi:10.1371/journal.pone.0057853

Zhao, L., Abdelsalam, N. R., Xu, Y., Chen, M. S., Feng, Y., Kong, L., et al. (2020). Identification of Two Novel Hessian Fly Resistance Genes H35 and H36 in a Hard winter Wheat Line SD06165. *Theor. Appl. Genet.* 133, 2343–2353. doi:10.1007/s00122-020-03602-3

Zorić, M., Dodig, D., Kobiljski, B., Quarrie, S., and Barnes, J. (2012). Population Structure in a Wheat Core Collection and Genomic Loci Associated with Yield under Contrasting Environments. *Genetica* 140, 259–275. doi:10.1007/s10709-012-9677-2

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 Hussain, Habib, Ahmed, Sadia, Bernardo, Amand, Bai, Ghori, Khan, Awan and Maqbool. 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: 26 April 2022
doi: 10.3389/fgene.2022.848627

Novel GBS-Based SNP Markers for Finger Millet and Their Use in Genetic Diversity Analyses

Haftom Brhane[1,2], Teklehaimanot Haileselassie[1], Kassahun Tesfaye[1,3], Rodomiro Ortiz[2], Cecilia Hammenhag[2], Kibrom B. Abreha[2] and Mulatu Geleta[2]*

[1]*Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia,* [2]*Department of Plant Breeding, Swedish University of Agricultural Sciences, Lomma, Sweden,* [3]*Ethiopian Biotechnology Institute, Ministry of Science and Technology, Addis Ababa, Ethiopia*

Eleusine coracana (L.) Gaertn., commonly known as finger millet, is a multipurpose crop used for food and feed. Genomic tools are required for the characterization of crop gene pools and their genomics-led breeding. High-throughput sequencing-based characterization of finger millet germplasm representing diverse agro-ecologies was considered an effective method for determining its genetic diversity, thereby suggesting potential candidates for breeding. In this study, the genotyping-by-sequencing (GBS) method was used to simultaneously identify novel single nucleotide polymorphism (SNP) markers and genotype 288 finger millet accessions collected from Ethiopia and Zimbabwe. The accessions were characterized at individual and group levels using 5,226 bi-allelic SNPs, with a minimum allele frequency (MAF) of above 0.05, distributed across 2,500 scaffolds of the finger millet reference genome. The polymorphism information content (PIC) of the SNPs was 0.23 on average, and a quarter of them have PIC values over 0.32, making them highly informative. The grouping of the 288 accessions into seven populations based on geographic proximity and the potential for germplasm exchange revealed a narrow range of observed heterozygosity (Ho; 0.09–0.11) and expected heterozygosity (He) that ranged over twofold, from 0.11 to 0.26. Alleles unique to the different groups were also identified, which merit further investigation for their potential association with desirable traits. The analysis of molecular variance (AMOVA) revealed a highly significant genetic differentiation among groups of accessions classified based on the geographic region, country of origin, days to flowering, panicle type, and Al tolerance ($p < 0.01$). The high genetic differentiation between Ethiopian and Zimbabwean accessions was evident in the AMOVA, cluster, principal coordinate, and population structure analyses. The level of genetic diversity of finger millet accessions varies moderately among locations within Ethiopia, with accessions from the northern region having the lowest level. In the neighbor-joining cluster analysis, most of the improved cultivars included in this study were closely clustered, probably because they were developed using genetically less diverse germplasm and/or selected for similar traits, such as grain yield. The recombination of alleles *via* crossbreeding genetically distinct accessions from different regions of the two countries can potentially lead to the development of superior cultivars.

Keywords: finger millet, gene diversity, genotyping-by-sequencing, single nucleotide polymorphism, tetraploid

OPEN ACCESS

Edited by:
Ramsey Steven Lewis, North Carolina State University, United States

Reviewed by:
Jaime Andrés Osorio-Guarín, Colombian Corporation for Agricultural Research (AGROSAVIA), Colombia
Hea-Young Lee, Seoul National University, South Korea

****Correspondence:***
Haftom Brhane
haftom1@gmail.com

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

Received: *04 January 2022*
Accepted: *21 March 2022*
Published: *26 April 2022*

Citation:
Brhane H, Haileselassie T, Tesfaye K, Ortiz R, Hammenhag C, Abreha KB and Geleta M (2022) Novel GBS-Based SNP Markers for Finger Millet and Their Use in Genetic Diversity Analyses. Front. Genet. 13:848627. doi: 10.3389/fgene.2022.848627

INTRODUCTION

Eleusine coracana (L.) Gaertn. (finger millet) is an annual crop that belongs to the Chloridoideae subfamily of the Poaceae grass family. There are nine species under the genus *Eleusine* Gaertn., among which *E. coracana* is the only domesticated species (Bisht and Mukai, 2001, 2002). According to archeological records, finger millet was first cultivated in Ethiopia and Uganda (Hilu and De Wet, 1976). Thereafter, the crop was distributed to other African and Asian countries (Hilu and De Wet, 1976; Fakrudin et al., 2004; Goron and Raizada, 2015). Finger millet is an important cereal crop cultivated for its nutritious grains. In Africa, it is mainly grown in Ethiopia, Zimbabwe, Tanzania, Malawi, Zaire, Zambia, and Kenya (National Research Council, 1996). In Ethiopia, it is commonly grown across the whole Tigray, in parts of Amhara (Gonder and Gojjam), Oromia (Wellega, IIluababora, and Hararghe), and Southern Nations and Nationalities and People's region (Gamo-Gofa) (Admassu et al., 2009). More than 2000 diverse finger millet accessions collected from its major cultivation areas in Ethiopia and abroad are conserved *ex situ* at the Ethiopian Biodiversity Institute (EBI; https://www.ebi.gov.et/) and used for research and breeding purposes. In Zimbabwe, it is a major crop in regions IV and V, which are the driest agro-ecological regions in the country (Mukarumbwa and Mushunje, 2010). The climatic conditions of these regions are similar to those of areas in Ethiopia where finger millet is widely cultivated.

Finger millet is a disomic tetraploid ($2n = 4x = 36$; AABB genome) crop. The diploid species *E. indica* (L.) Gaertn. ($2n = 2x = 18$; AA genome) and *E. floccifolia (Forssk.) Spreng* ($2n = 2x = 18$; BB genome) are considered the genome donors of *E. coracana* (Bisht and Mukai, 2001). It exhibits a great deal of phenotypic diversity for different traits. These include diversity in grain color (dark brown, light brown, radish brown, and white), growth habit (erect, recumbent, and prostrate), panicle shape (open, semi-curved, and curved), and flowering time (Upadhyaya et al., 2007). About 33 million hectares are estimated to be devoted to the production of different millet crops globally, with an estimated production quantity of 32.7 million tons of millet grains (FAOSTAT, 2020). About 4.5 million hectares of finger millet is produced globally every year (Antony Ceasar et al., 2018), and hence finger millet accounts for about 13.6% of the production of all millets. Ethiopia's total finger millet harvested area is 480,511 ha (FAOSTAT, 2020), which produced an estimated 1.2 million tons of grains (FAOSTAT, 2020). Finger millet is consumed as porridge, pancake, injera, and traditional alcoholic beverages in Ethiopia (Bezaweletaw et al., 2006). It is favored because it is rich in nutrients (e.g., calcium, iron, thiamine, riboflavin, and nicotinic acid) and health-promoting components (tackle bone-related diseases, anemia, and cholesterol) and has better performance under stressful conditions (such as drought, disease, pest, salt, and soil acidity) (Barbeau and Hilu, 1993; Chethan and Malleshi, 2007; Sri et al., 2007; Nakarani et al., 2021a). It is a multipurpose crop because the grains are used for food, and its straw serves as a preferred animal feed (Chethan and Malleshi, 2007; Nakarani et al., 2021a).

Even though finger millet is an important food security, health-promoting, and climate-resilient crop, its productivity in Ethiopia is low due to various factors. The attention given to finger millet research and breeding has been insufficient and modern technologies were not well adopted to increase the efficiency of its production. Due to the highly limited availability of improved cultivars or their potential limitations, farmers are widely using local landraces. Consequently, major factors such as threshing difficulty, weeds, lodging, soil acidity, and blast disease caused by *Magnaporthe grisea* are limiting finger millet production and productivity, especially in Ethiopia (Erenso et al., 2007; Molla, 2010; Kebede et al., 2019). Aluminum is phytotoxic to plants through disrupting their root system, increasing the rigidity of cell wall and cell membrane, interrupting cell division, inducing oxidative stress, and blocking the influx of essential nutrients (Delhaize and Ryan, 1995; Zhang et al., 2014; Eekhout et al., 2017). Hence developing widely adaptable stable cultivars tolerant to the aforementioned biotic and abiotic factors to promote its production and productivity.

Plant breeding relies on the selection of diverse germplasm with desirable characteristics to develop new cultivars, achieved through the understanding of the genetic differences between germplasm used for breeding. Hence, the development and use of genome-wide markers are highly desirable to reveal the genetic diversity within a crop gene pool, followed by efficient use in conservation and plant breeding. Quantifying the genetic diversity of finger millet using such genome-wide DNA markers provides valuable information to breeders who assist in developing cultivars with various desirable traits, such as high yield and stress tolerance. Genotyping-by-sequencing (GBS) is a widely used next-generation sequencing (NGS) method developed for the simultaneous discovery of new markers and genotyping of target germplasm (Elshire et al., 2011). It is a high-throughput and cost-effective method used in various crops for different applications (Elshire et al., 2011; Poland and Rife, 2012; Huang et al., 2014; Alipour et al., 2017; Geleta et al., 2020; Hammenhag et al., 2020). The most popular DNA markers generated through the GBS method are single nucleotide polymorphism (SNP) markers. SNP markers are the most abundant sequence variation across crop genomes, which are suitable for analysis of genetic variation, population structure, marker-trait association, genomic selection, mapping quantitative trait loci (QTL), map-based cloning, and other plant breeding applications that need genome-wide scanning (Batley and Edwards, 2007; Kumar et al., 2012; Tsehay et al., 2020). The genetic diversity analysis of finger millet using GBS-derived SNP markers has received insufficient attention, despite its potential for providing markers useful for genetic improvement of the crop. The present study aimed at genotyping genetically diverse finger millet germplasm sampled from diverse agro-ecologies in Ethiopia and Zimbabwe using the GBS method as a means

to develop new genomic resources for finger millet and molecular characterization of the germplasm.

MATERIALS AND METHODS

Plant Material

This study used two hundred eighty-eight finger millet accessions composed of 274 landrace accessions (228 from Ethiopia and 46 from Zimbabwe) and 14 improved cultivars released in Ethiopia. The 228 Ethiopian landrace accessions were originally collected from Amhara (130), Benishangul-Gumuz (2), Oromia (51), Southern Nations, Nationalities and Peoples' Region (SNNPR) (4), Tigray (32), and unknown locations in Ethiopia (9) (**Supplementary Table S1**). Landraces were obtained from the Ethiopian Biodiversity Institute (EBI), whereas the improved cultivars were obtained from Bako Agricultural Research Center (BARC).

Testing Accessions for Aluminum Stress Tolerance

Aluminum (Al^{3+}) stress tolerance of the accessions was initially screened using the hydroponic nutrient solution. The solution containing 500 μM KNO_3, 500 μM $CaCl_2$, 500 μM NH_4NO_3, 150 μM $MgSO_4.7H_2O$, 10 μM KH_2PO_4, 2 μM $FeCl_3$ (III), and 75 μM Al_2 $(SO_4)_3.18H_2O$ was prepared according to Zhou et al. (2013). The pH of the nutrient solution was adjusted to 4.3 (using 1 M HCl or NaOH) and renewed every day to maintain the pH and Al^{3+} concentration relatively constant. For this experiment, similar-sized seeds from each accession were selected and surface-sterilized by soaking in 3% sodium hypochlorite solution for 5 min and rinsed thoroughly with water. The sterilized seeds of each accession were then wrapped in tissue paper, moistened with distilled water, and placed in a Petri dish for 36 h under dark conditions to initiate germination. Then, the seeds of each accession were grouped into two, and one group was grown in the hydroponic nutrient solution containing 100 μM Al. In contrast, the other group was grown as a control in the nutrient solution without aluminum for 10 days. This was followed by measuring the root length of the seedlings in cm using a ruler and calculating the relative root length (RRL) in percent as follows:

$$\text{RRL } (\%) = \frac{\text{Root length under Al treatment}}{\text{Root length under control}} \times 100\%.$$

Planting, Leaf Tissue Sampling, and Phenotyping

In order to ensure at least one germinating seeds of each of the 288 accessions, five seeds were planted in a 5 L plastic pot for each accession in a greenhouse at the Swedish University of Agricultural Sciences (SLU, Alnarp, Sweden). After seed germination, extra seedlings were removed, and only a single seedling was maintained for each accession. Three weeks after planting, the leaf tissue of each accession was separately sampled using the BioArk Leaf collection kit from LGC, Biosearch Technologies (https://biosearchassets.blob.core.windows.net/assetsv6/guide_bioark-leaf-collection-kit.pdf), as described in Osterman et al. (2021). Afterward, the samples were sent to LGC, Biosearch Technologies (Berlin, Germany) for genomic DNA extraction, followed by GBS-based genotyping. The plants were hand-weeded and treated with pesticides and fertilizers during their growth in the greenhouse. Afterward, data on days to flowering and panicle shape were recorded.

DNA Extraction, Genotyping-By-Sequencing Optimization, Sequencing, and Read Pre-Processing

A high molecular weight genomic DNA of the 288 accessions was extracted using the Sbeadex plant kit (https://biosearch-cdn.azureedge.net/assetsv6/sbeadex-plant-data-sheet.pdf) for the GBS analysis. For the construction of a GBS library, *Pst*I (CTGCA*G, a six-base cutter) and *Ape*KI (G*CWGC, a four-base cutter) restriction enzymes were used following the recommendations of the experts at LGC, Biosearch Technologies, who optimized restriction enzymes for various crops. The combined use of *Pst*I-*Ape*KI produced a fragment size distribution suitable for sequencing on Illumina platforms with a mean insert size of ~220 bp. The GBS was conducted in 2 × 150 bp (paired-end) sequencing mode using NexSeq 500/550 v2 and NovaSeq SP FC NGS platforms to generate the reads. This produced about 1.5 million read pairs per sample. The reads were then adapter clipped, and those containing Ns were discarded along with reads whose 5′ ends did not match the restriction enzyme site. This was followed by trimming the 3′-end of the remaining reads using Trimmomatic v. 0.3 (Bolger et al., 2014) to obtain an average Phred quality score of ≥20 over a window of ten bases and discarding reads with a final length <20 bp.

Read Alignment, Single Nucleotide Polymorphism Discovery and Filtering, and Genotype Calling

For this GBS analysis, *Eleusine coracana* subsp. *coracana* cultivar ML-365 scaffold-level genome assembly (Hittalmani et al., 2017; https://www.ncbi.nlm.nih.gov/assembly/GCA_002180455.1/) was used as a reference genome, as there is no chromosome-level assembly available for finger millet yet. BWA-MEM v. 0.7.12 software package (Li and Durbin, 2009) was used to align quality trimmed reads against the reference genome. The mapping rate of the GBS reads to the reference genome (containing 525,627 scaffolds with a total size of 1.3 Gb) was 99.16%. Then, Freebayes v. 1.2.0 (Garrison and Marth, 2012) was used for variant discovery and genotype calling as diploids. The alignment of the reads from the 288 finger millet genotypes resulted in the discovery of 101,889 SNPs. The SNPs were then filtered using GBS-specific criteria (minimum read count > 8, minimum allele frequency (MAF) ≥ 0.05, and percentage of

samples with assigned genotype ≥66%). Out of 71,140 SNPs with average read counts of above eight, 11,120 SNPs were fully covered (2 × 150 bp reads), had MAF of ≥5%, and scored in ≥66% of the samples. The 11,120 SNPs spanned 4,729 scaffolds, with 51.4% of the scaffolds containing a single SNP, although up to 32 SNPs were recorded per scaffold (**Supplementary Figure S1**). The 11,120 SNPs were further filtered, and 5,226 biallelic SNPs with genotypic data across all samples (no missing data) were used for genetic diversity analyses described below.

Data Analysis

The vast majority of the finger millet accessions used in this study are landraces, originally collected from diverse agro-ecologies. Additionally, each accession was represented by a single genotype as described above. Hence, to facilitate data analysis and better understand the pattern of genetic diversity, the accessions were grouped based on different criteria prior to data analyses. First, the 288 accessions were grouped into seven groups according to their geographic origin, which will be referred to as "populations" (Pop-1 to Pop-7) from here on for the sake of simplicity. Pop-1 to Pop-5 represent Ethiopian landrace accessions, and their descriptions are as follows: Pop-1 represents accessions collected from Agew-Awi, Gojam, Bahrdar, and Metekel (northwestern accessions); Pop-2 represents accession from western Tigray and Gonder (northern accessions); Pop-3 represents accessions collected from Wellega and IIluababora (western accessions); Pop-4 represents accessions from central, eastern, and southern Tigray and northern Wello (northeastern accessions); Pop-5 represents accessions with unknown sampling location; and Pop-6 represents the 14 improved cultivars whereas the 46 Zimbabwean landrace accessions form Pop-7.

Second, the accessions were grouped based on days to flowering, as early flowering and late flowering types (**Supplementary Figure S2**). Third, the accessions were grouped based on their panicle shape as open (fingers straight), semi-compact (top of fingers curved inward), and compact (fingers fully curved inward) (**Supplementary Figure S3**). Fourth, the accessions were broadly grouped into two groups based on the performance of accessions against aluminum toxicity (Al), where accessions having relative root length (RRL) above and below 50% were grouped as Al-tolerant and Al-susceptible, respectively (**Supplementary Figure S4**). Finger millet production is hampered by abiotic factors such as Al toxicity, so it is necessary to use this information to group the accessions according to their level of Al tolerance.

The genetic diversity of the 288 accessions grouped based on different criteria was assessed using different statistical programs. For this data analysis, 5,226 SNPs with minor allele frequency (MAF) above 5% and no missing values were used. Polymorphism information content (PIC) of the SNPs was calculated as described by Hildebrand et al. (1994). The frequency spectrum of segregating sites of each population was analyzed using the command analysis through segregating sites option in population size changes added to a toolbar of DNASP v 6 (Rozas et al., 2003). The number of alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe), Shannon's information index (I), percentage of polymorphic loci (PPL), and number of private alleles were calculated using GeneAlEx v. 6.51b2 (Peakall and Smouse, 2006). The analysis of molecular variance (AMOVA) and Hardy–Weinberg equilibrium test were performed using Arlequine v. 3.5.2.2 (Excoffier and Lischer, 2010). The average number of pairwise differences within and among the populations and Nei's genetic distance (Nei, 1972) were also calculated using the same software, and their graphic representation was generated using scripts of a console version of the R statistical package (Rcmd) triggered through a command button added to Arlequin's toolbar.

Neighbor-joining (NJ) cluster analysis was done based on Nei's genetic distance using the MEGA7 program (Kumar S. et al., 2016). GeneAlEx v. 6.51b2 was used for principal coordinate analysis (PCoA). Furthermore, the genetic structure of the accessions was evaluated using STRUCTURE 2.3.4 software (Pritchard et al., 2000). STRUCTURE analysis was run for K = 1 to K = 10 with 10 independent runs at each K using the admixture model with burn-in period length of 1,000,000 and a Markov chain Monte Carlo of 1,000,000 replications. STRUCTURESELECTOR (Li and Liu, 2018) was used to determine the optimum K based on the different approaches described by Puechmaille (2016). After the optimum K was determined, the CLUMPACK beta version (Kopelman et al., 2015) was used to display the graphical representation of the population structure.

RESULTS

Characteristics and Distribution of the Single Nucleotide Polymorphism Markers

A total of 5,226 SNP markers with MAF above 5% and without missing data were used for genetic diversity analyses of the finger millet accessions (**Supplementary Table S2**). The 5,226 SNPs were spread across 2,500 scaffolds, with 52.7% of these scaffolds containing only one of the SNPs although the number of SNPs per scaffold ranged from 1 to 18 (**Figure 1**). The mean and median of the MAF of the SNPs were 0.19 and 0.15, respectively, with 25% of the SNPs having a MAF above 0.27 (**Figure 2**). The polymorphism information content (PIC) of the SNPs varied from 0.09 to 0.38 with a mean and third quartile of 0.23 and 0.32, respectively. Similarly, the mean values of the effective number of alleles (Ne), Shannon information index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe), and fixation index (F) of the SNPs were 1.43, 0.44, 0.10, 0.28, 0.28, and 0.73, respectively (**Figure 2**). A significant number of SNPs were outliers in terms of Ho and F. The Ho of 75% of the SNPs was below 0.03, whereas the F of the SNPs varied from −0.99 to 1.0 (**Figure 2**). Based on the site-frequency spectrum, it was revealed that the MAF distribution of the SNP loci varied substantially among the seven finger millet populations (**Figure 3**). Only 4.5% of the individuals across all populations had minor alleles at more than 10% of the SNP loci, with all individuals having minor alleles at less than 35% of the loci. At a population level, only

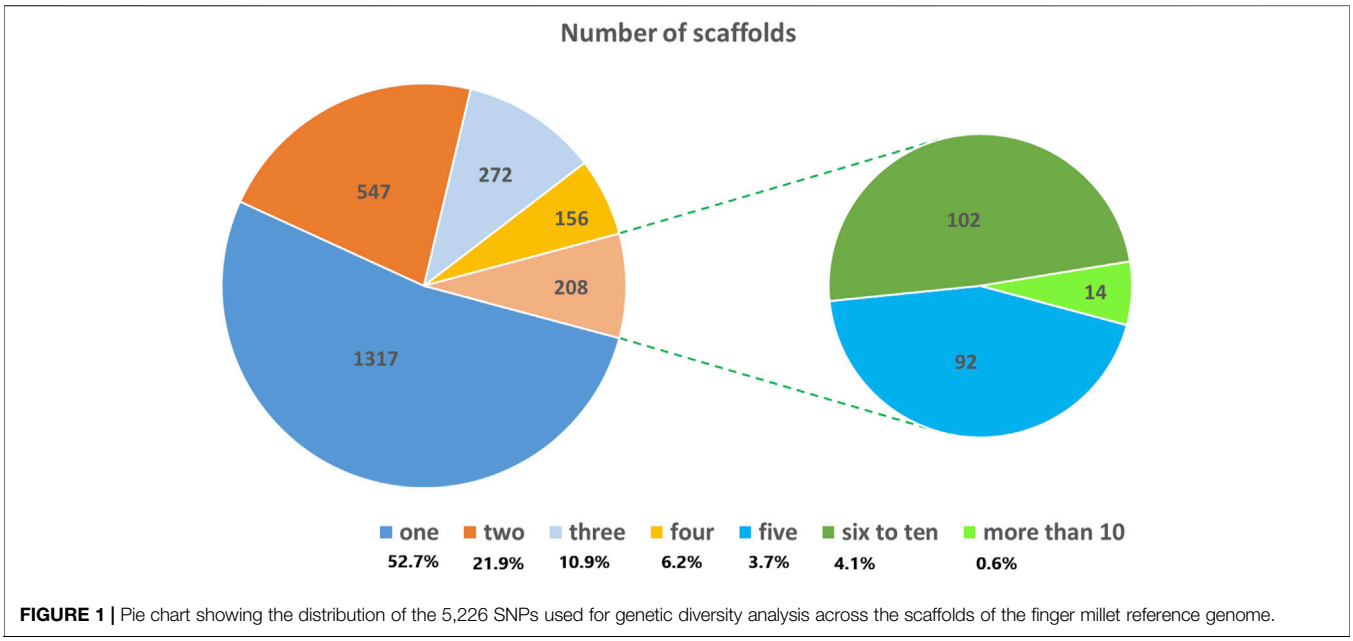


FIGURE 1 | Pie chart showing the distribution of the 5,226 SNPs used for genetic diversity analysis across the scaffolds of the finger millet reference genome.

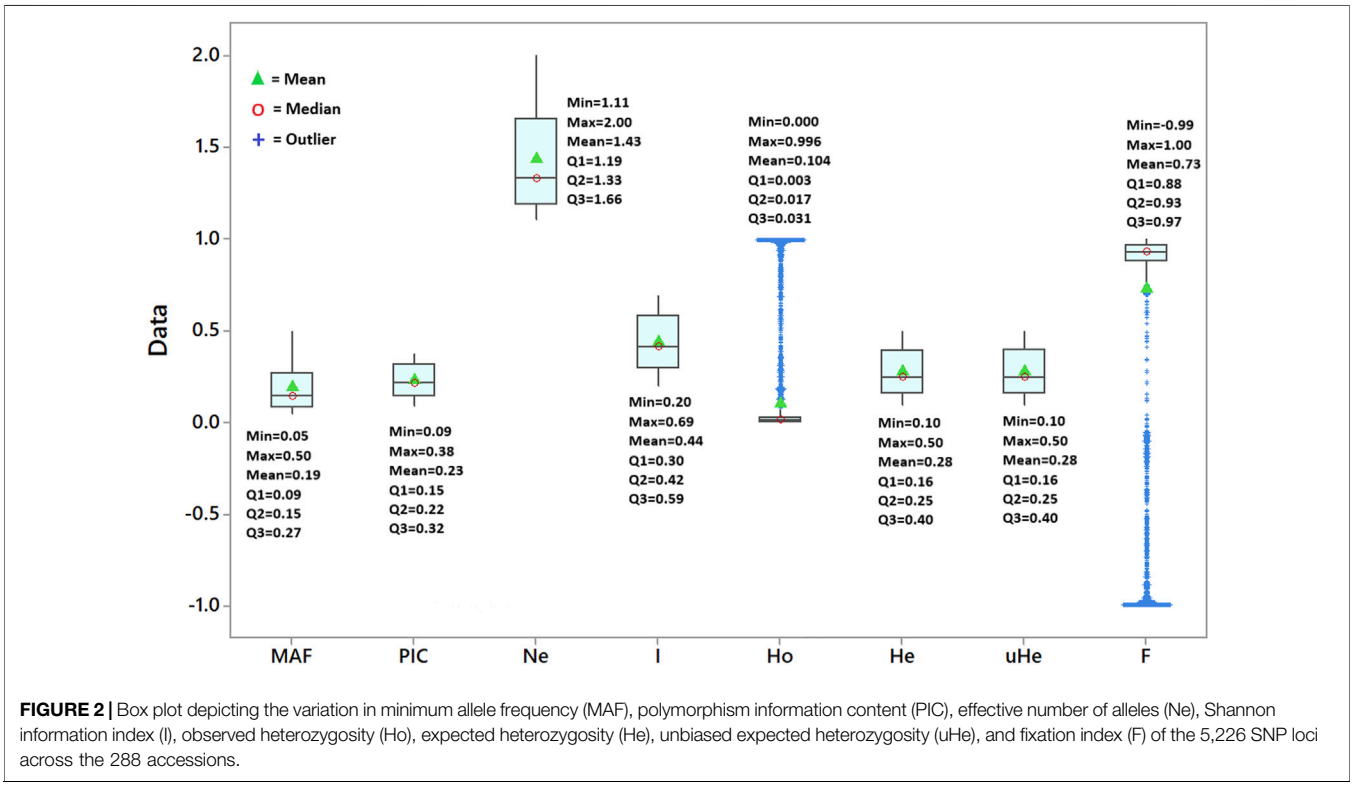


FIGURE 2 | Box plot depicting the variation in minimum allele frequency (MAF), polymorphism information content (PIC), effective number of alleles (Ne), Shannon information index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe), and fixation index (F) of the 5,226 SNP loci across the 288 accessions.

1.0%, 1.8%, 0%, 7.1%, 55.6%, 21.4%, and 2.2% of the individuals had minor alleles in ≥10% of the SNP loci in Pop-1 to Pop-7, respectively. A coalescent analysis based on the expected site-frequency spectrum did not match the observed frequency distributions well, with a vast majority of individuals in all populations exhibiting lower expected values than observed (**Figure 3**).

The Hardy–Weinberg equilibrium (HWE) test was conducted using 1000,000 steps in the Markov chain and 100,000 steps in dememorization by assuming that the 288 accessions are random samples from a single population. This analysis revealed that 2.5% of the loci were at HWE, 11% were with excess heterozygosity, and 86.5% were with heterozygosity deficit. Interestingly, the vast majority of the loci deviated from HWE at a highly significant

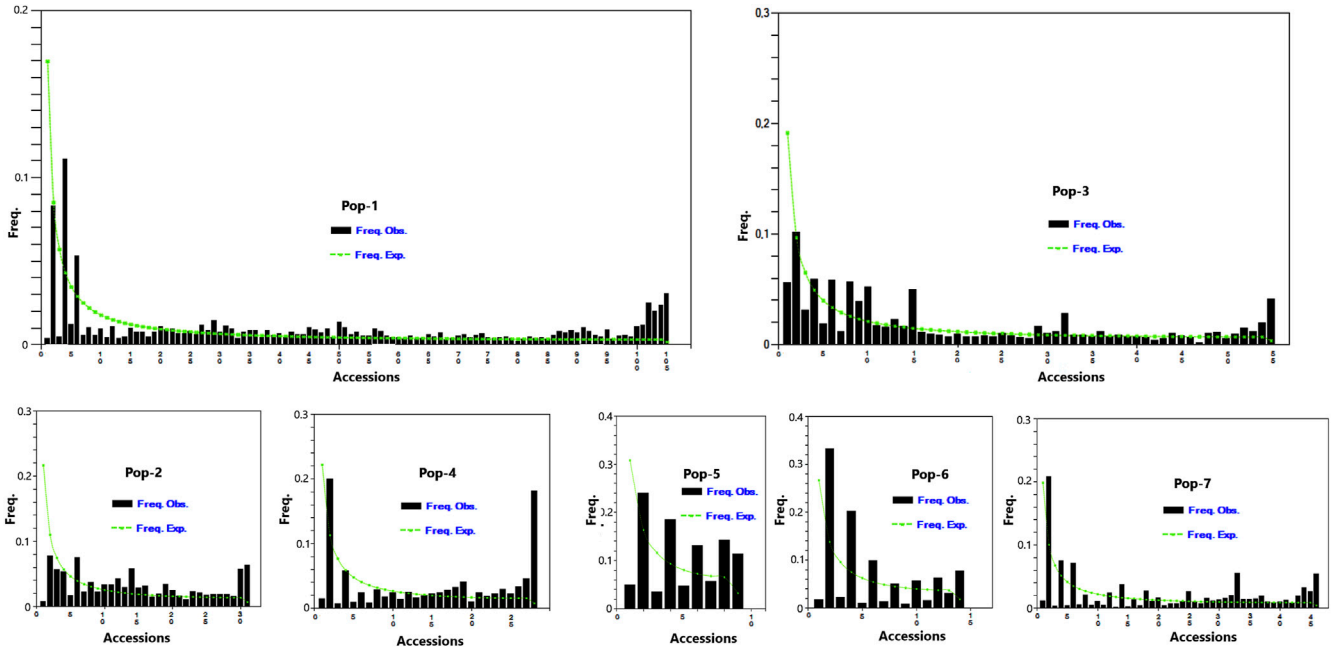


FIGURE 3 | Graph depicting minor allele frequency (MAF) based site-frequency spectrum across 5,226 SNP loci in the seven finger millet populations.

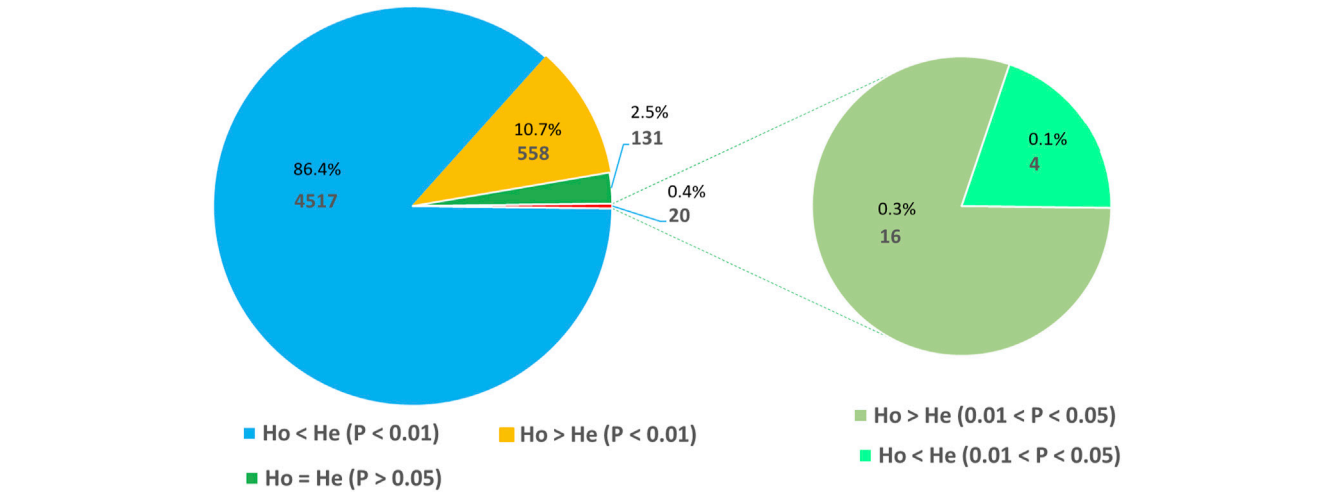


FIGURE 4 | Pie chart illustrating the number and proportion of SNP loci at Hardy–Weinberg equilibrium (HWE) and those with a significant deviation from HWE in terms of heterozygote excess and deficiency.

level (p < 0.01) with 10.7% and 86.4% of them showing heterozygosity excess and deficit, respectively (**Figure 4**).

Genetic Diversity of the Finger Millet Accessions

As a means to facilitate data analysis and determine patterns of genetic diversity, the 288 accessions were divided into seven populations based on geographic proximity and the potential for germplasm exchange. Pop-1 to Pop-7 comprised 105, 31, 55, 28, 9, 14, and 46 accessions, respectively. Based on their flowering time under greenhouse conditions, we broadly grouped the accessions into "early" and "late" flowering types, with 115 days to flowering as the dividing point (**Supplementary Figure S2**). The early and late flowering types accounted for 69.4 and 30.6% of the accessions, respectively. In terms of panicle shape (**Supplementary Figure S3**), the majority of the accessions had open panicles (71.2%), whereas those with curved and semi-curved panicles accounted for 19.1% and 9.7% of the accessions, respectively.

TABLE 1 | Estimates of different genetic diversity parameters for different groups of finger millet accessions: mean values for the observed number of alleles (N_A), effective number of alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased genetic diversity (uHe), percentage of polymorphic loci (PPL), and number of private alleles (NPA).

Groups	N	Na	Ne	I	Ho	He	uHe	PPL	NPA
Pop-1[a]	105	1.96	1.44	0.40	0.11	0.26	0.26	0.96	13
Pop-2[a]	31	1.69	1.36	0.33	0.10	0.22	0.22	0.69	0
Pop-3[a]	55	1.87	1.33	0.32	0.11	0.20	0.21	0.87	0
Pop-4[a]	28	1.32	1.19	0.16	0.09	0.11	0.11	0.32	0
Pop-5[a]	9	1.53	1.33	0.28	0.10	0.19	0.20	0.53	0
Pop-6[a]	14	1.69	1.31	0.31	0.09	0.19	0.20	0.69	0
Pop-7[a]	46	1.76	1.36	0.32	0.09	0.21	0.21	0.76	11
Mean	41.1	1.69	1.33	0.30	0.10	0.20	0.20	0.69	3.4
EF[b]	126	2.00	1.45	0.43	0.10	0.28	0.28	1.00	25
LF[b]	162	2.00	1.42	0.40	0.11	0.26	0.26	1.00	1
Mean	144	2.00	1.43	0.42	0.10	0.27	0.27	1.00	13
Compact[c]	55	1.70	1.38	0.33	0.09	0.22	0.22	0.70	64
Semi-compact[c]	31	1.87	1.34	0.33	0.11	0.21	0.21	0.87	0
Open[c]	202	1.99	1.40	0.38	0.11	0.25	0.25	0.99	111
Mean	96	1.85	1.37	0.35	0.10	0.23	0.23	0.85	58.3
Tolerant[d]	60	1.93	1.45	0.41	0.10	0.27	0.27	0.93	11
Susceptible[d]	228	2.00	1.39	0.39	0.11	0.24	0.24	1.00	387
Mean	144	1.96	1.42	0.40	0.10	0.26	0.26	0.96	199

EF, early flowering; LF, late flowering; N, number of accessions.

[a]*Accessions were grouped according to geographic proximity and the potential for germplasm exchange.*

[b]*Accessions were grouped according to days to flowering.*

[c]*Accessions were grouped according to panicle shape.*

[d]*Accessions were grouped according to tolerance to aluminum toxicity.*

The Al tolerance experiment revealed that 20.8% of the accessions had RRL values above 50% and were hence considered tolerant of Al toxicity. However, the RRL values of the vast majority of accessions (77.2%) were below 50% and so were considered susceptible to Al toxicity. The results of genetic diversity analyses for the different groups of accessions are presented below.

Among the seven populations, the effective number of alleles (Ne) varied from 1.19 (Pop-4) to 1.44 (Pop-1), with an overall mean of 1.33 (**Table 1**). Shannon information index (I) ranged from 0.16 (Pop-4) to 0.40 (Pop-1) with an average value of 0.30. In terms of observed heterozygosity (Ho), there were slight differences between the populations, which only varied from 0.09 to 0.11. On the contrary, expected heterozygosity (He) or gene diversity ranged from 0.11 (Pop-4) to 0.26 (Pop-1) with an average value of 0.20, which were similar to the corresponding values of unbiased expected heterozygosity (uHe). The percentage of polymorphic loci (PPL) showed threefold variation among the populations, with Pop-1 having the highest (96%) and Pop-4 the lowest (32%) values, and on average 69% of the loci were polymorphic. Interestingly, the level of genetic diversity of the improved cultivars (Pop-6) was similar to the overall average values regardless of the relatively small sample size (**Table 1**). Pop-1, Pop-2, Pop-3, and Pop-7 were more diverse and Pop-4 and Pop-5 less diverse when compared with Pop-6 (improved cultivars).

The genetic analyses of the accessions grouped based on panicle type revealed that accession with open panicle had higher mean values for most of the diversity parameters (Na = 1.99, Ne = 1.40, I = 0.38, Ho = 0.11, He = 0.25, and PPL = 100%) than semi-compact accessions (Na = 1.87, Ne = 1.34, I = 0.33, Ho = 0.11, He = 0.21, and PPL = 87%) and compact accessions (Na = 1.70, Ne = 1.38, I = 0.33, Ho = 0.09, He = 0.22, and PPL = 70%) (**Table 1**). The early and late flowering groups had quite similar values for the different genetic diversity estimates, with the early flowering group having Ne = 1.45, I = 0.43, Ho = 0.10, He = 0.28, and PPL = 100% and the late flowering group having Ne = 1.42, I = 0.40, Ho = 0.11, He = 0.26, and PPL = 100% (**Table 1**). The Al-tolerant group had slightly higher values in effective number of alleles, Shannon information index, and expected heterozygosity (Ne = 1.45, I = 0.41, and He = 0.27) than the Al-susceptible group (Ne = 1.39, I = 0.39, and He = 0.24). All loci were polymorphic in the Al-susceptible group as opposed to 93% polymorphic loci in the Al-tolerant group. Among the seven populations, private alleles were recorded only in Pop-1 (NPA = 13) and Pop-7 (NPA = 11) (**Table 1**). The frequency of the unique alleles in Pop-1 varied from 0.138 to 0.148, while in Pop-7, it varied from 0.315 to 0.576 (**Table 2**). In the case of panicle-type group, the number of unique alleles to open, semi-compact, and compact panicle groups were 111, 0, and 64, respectively. Moreover, 25 unique alleles were specific to early flowering group as opposed to only one in late flowering group. Likewise,

TABLE 2 | List of SNP loci with a private allele in Pop-1 and Pop-7 with their alleles and allele frequencies.

No.	Locus name	Population	Private allele	Private allele frequency
1	LXGH01033309.1_4802	Pop-1	T	0.148
2	LXGH01036535.1_25361	Pop-1	G	0.138
3	LXGH01036535.1_25481	Pop-1	G	0.138
4	LXGH01224870.1_62858	Pop-1	C	0.138
5	LXGH01296805.1_22457	Pop-1	T	0.138
6	LXGH01296805.1_22532	Pop-1	A	0.138
7	LXGH01296805.1_22641	Pop-1	G	0.138
8	LXGH01296805.1_22649	Pop-1	G	0.138
9	LXGH01296805.1_4809	Pop-1	A	0.138
10	LXGH01296805.1_4820	Pop-1	G	0.138
11	LXGH01405315.1_45826	Pop-1	G	0.138
12	LXGH01405315.1_45892	Pop-1	C	0.138
13	LXGH01405315.1_45929	Pop-1	G	0.138
14	LXGH01094451.1_9437	Pop-7	T	0.435
15	LXGH01151768.1_642	Pop-7	T	0.315
16	LXGH01174315.1_91057	Pop-7	A	0.576
17	LXGH01211987.1_24770	Pop-7	G	0.348
18	LXGH01246444.1_13014	Pop-7	A	0.391
19	LXGH01321131.1_13891	Pop-7	A	0.348
20	LXGH01321131.1_14094	Pop-7	C	0.348
21	LXGH01321131.1_14706	Pop-7	C	0.348
22	LXGH01321131.1_14787	Pop-7	C	0.359
23	LXGH01321131.1_16828	Pop-7	A	0.348
24	LXGH01482256.1_2202	Pop-7	T	0.380

TABLE 3 | Analysis of molecular variance (AMOVA) of different groups formed from the 288 finger millet accessions based on their geographic origin, country of origin, panicle type, days to flowering, and aluminum tolerance using 5,226 SNP markers.

Source of variation	df	SS	MS	Est. Var.	PV	FI	*p*-value
Seven populations							
Among pops	6	183,071.9	30,511.98	767.44	30	0.305	$p < 0.001$
Within pops	281	492,412.4	1,752.357	1752.36	70		
Total	287	675,484.3		2519.80			
Ethiopia vs. Zimbabwe							
Among countries	1	115,499.5	115,499.5	1,468.74	42.9	0.429	$p < 0.001$
Within countries	286	559,984.8	1,957.989	1,957.99	57.1		
Total	287	675,484.3		3,426.73			
Open vs. semi-compact vs. compact panicles							
Among ear shape groups	2	62,139.14	31,069.57	453.09	30	0.300	$p < 0.001$
Within ear shape groups	285	301,174.6	1,056.753	1,056.75	70		
Total	287	363,313.7		1,509.85			
Early vs. Late flowering							
Among flowering groups	1	34,807.07	34,807.07	229.73	9.3	0.09	$p < 0.001$
Within flowering groups	286	641,277.1	2,242.228	2,242.23	90.7		
Total	287	676,084.2		2,471.96			
Al-tolerant vs. Al-susceptible							
Among tolerance groups	1	42,568.6	42,568.6	436.29	28	0.280	$p < 0.001$
Within tolerance groups	286	320,745.1	1,121.486	1,121.49	72		
Total	287	363,313.7		1,557.77			

DF, degrees of freedom; SS, sum of squares; MS, mean square; Est. Var, estimated variance; PV, percentage of variation; FI, fixation index (F_{ST}).

TABLE 4 | Pairwise Nei's standard genetic distance between the seven populations of finger millet evaluated using 5,226 SNP markers. A diagonal value is the mean genetic distance of the corresponding populations from the other populations.

	Pop-1	Pop-2	Pop-3	Pop-4	Pop-5	Pop-6	Pop-7
Pop-1	**0.18**						
Pop-2	0.04	**0.17**					
Pop-3	0.10	0.09	**0.16**				
Pop-4	0.20	0.19	0.14	**0.25**			
Pop-5	0.06	0.01	0.09	0.19	**0.17**		
Pop-6	0.24	0.24	0.16	0.30	0.24	**0.24**	
Pop-7	0.41	0.44	0.38	0.48	0.43	0.27	**0.40**

11 and 387 private alleles were recorded in the Al-tolerant and Al-susceptible groups, respectively (**Table 1**).

Genetic Variation Within and Among Different Groups of Accessions

Analysis of molecular variance (AMOVA) was conducted to estimate the genetic variation within and among different groups of finger millet accessions grouped based on different criteria. AMOVA for the seven populations revealed that the variation within and among the populations accounted for 70% and 30% of the total variation, respectively (**Table 3**). When accessions were grouped based on country of origin (Ethiopia vs. Zimbabwe), AMOVA revealed that the variation within and among countries accounted for 57.1% and 42.9% of the total variation, respectively. Similarly, the variation among days to flowering groups (early vs. late), panicle-type groups (open vs. semi-compact vs. compact), and aluminum tolerance groups (Al-tolerant vs. Al-susceptible) accounted for 30%, 9.3%, and 28% of the corresponding total variations, respectively (**Table 3**). Interestingly, the variations among the groups formed based on all criteria were highly significant ($p < 0.01$). The estimates for the differentiation among the groups (F_{ST}) were 0.31 (for the seven populations), 0.43 (for country of origin), 0.30 (for panicle type), 0.09 (for days to flowering), and 0.28 (for aluminum tolerance) (**Table 3**).

Pairwise Genetic Distance, Cluster, and Principal Coordinate Analyses

Pairwise Nei's standard genetic distance among the seven finger millet populations ranged from 0.01 to 0.48. The lowest genetic distance was found between Pop-2 and Pop-5 (0.01), followed by Pop-1 vs. Pop-2 (0.04) and Pop-1 vs. Pop-5

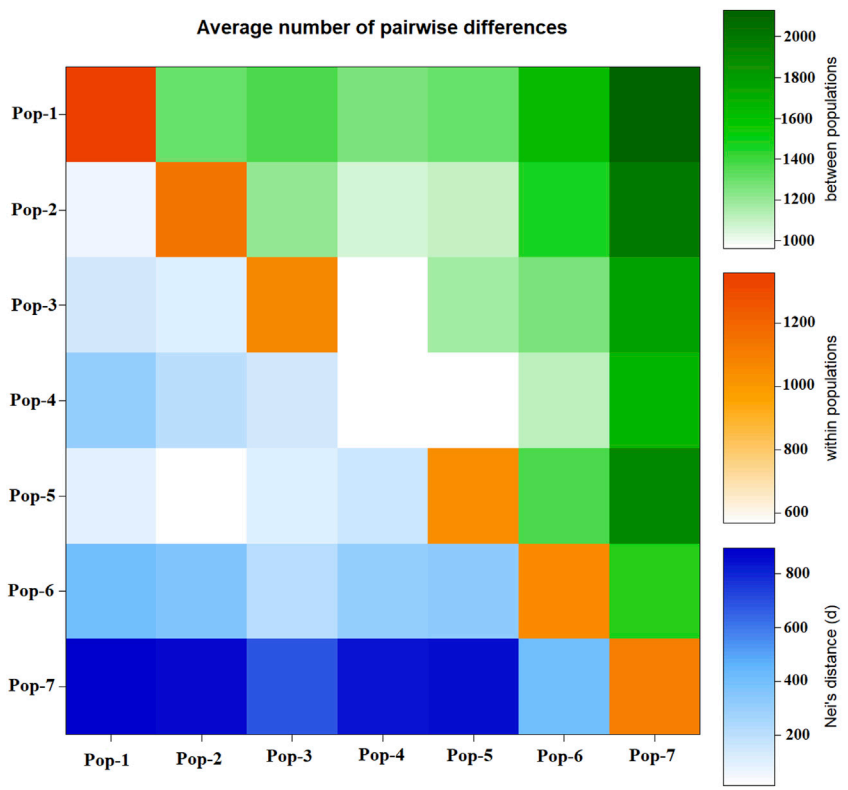


FIGURE 5 | Graphical representation of the average number of pairwise differences between the seven finger millet populations generated using 5,226 SNP markers. Below diagonal represents Nei's genetic distance, diagonal represents pairwise differences within populations, and above diagonal represents pairwise differences between populations.

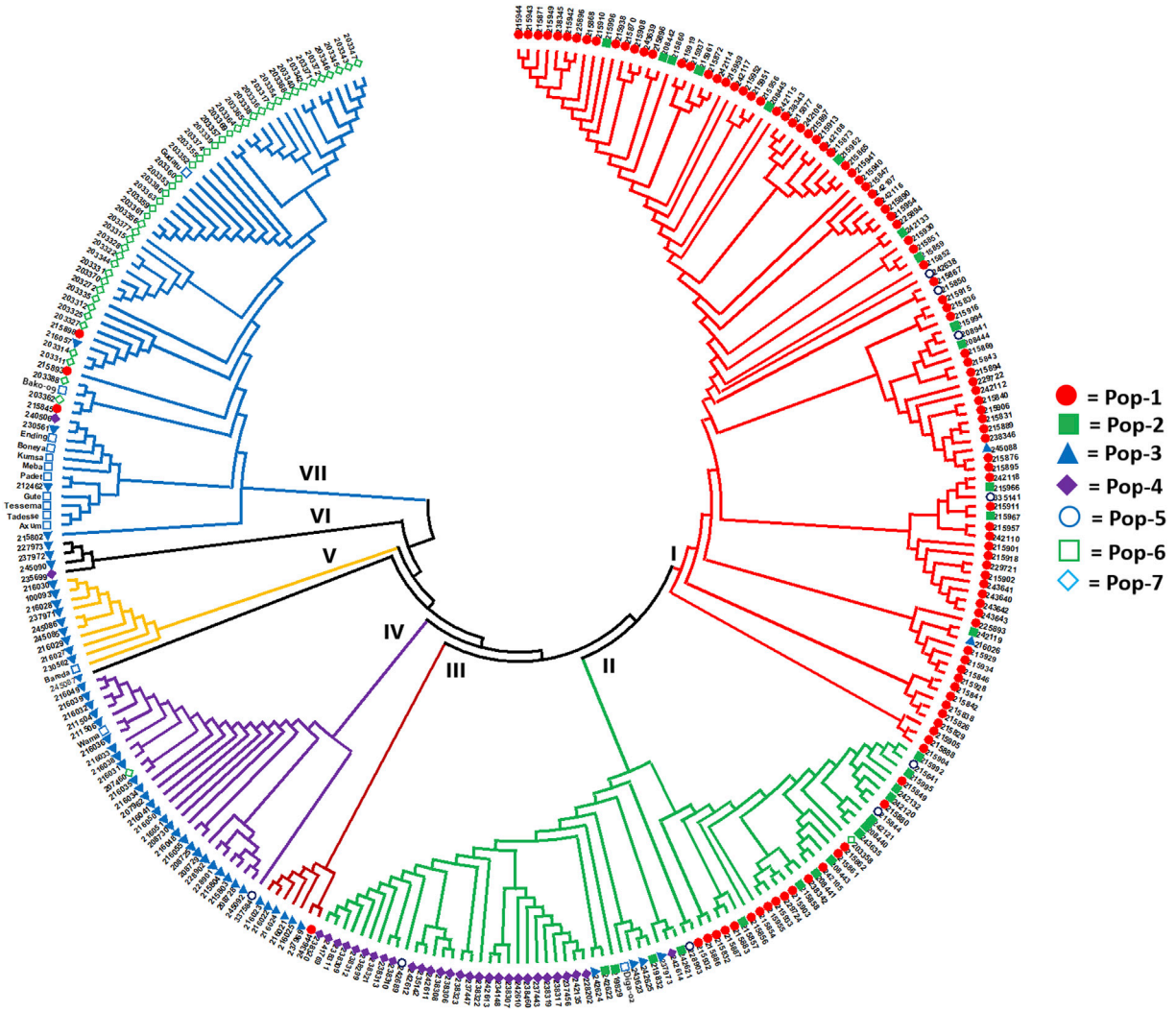


FIGURE 6 | Nei's standard genetic distance-based neighbor-joining tree showing the clustering pattern of the 288 finger millet accessions. Accessions sharing a symbol with the same shape and color belong to the same population.

(0.06) (**Table 4**), which is in agreement with the average number of pairwise differences (**Figure 5**). The highest genetic distance was obtained between Pop-4 and Pop-7 (0.48), followed by Pop-2 vs. Pop-7 (0.44) (**Table 4**). Pop-7 (representing Zimbabwean accessions) had a higher genetic distance to most of the finger millet populations from Ethiopia (0.27–0.48) when compared with the genetic distance between pairs of the Ethiopian populations (0.01–0.30). Pop-7 was more similar to Pop-6 (released cultivars) among the six populations representing the Ethiopian finger millet gene pool with a genetic distance of 0.27 (**Table 4**; **Figure 5**). The mean Nei's genetic distance of each population from the other populations varied from 0.16 (Pop-3) to 0.40 (Pop-7). The analysis of the average number of pairwise differences within populations revealed that Pop-1 and Pop-4 had the highest and lowest differences, respectively (**Figure 5**). In agreement with the genetic distance data, most Zimbabwean accessions were phenotypically distinct from most Ethiopian accessions by being early flowering type and having compact panicles and shorter and stronger stems.

The Nei's standard genetic distance-based neighbor-joining (NJ) cluster analysis of the 288 accessions revealed seven clusters with a good clustering pattern according to their geographic origin, except that few accessions collected from different regions were clustered together (**Figure 6**). Cluster-I was the largest and more diverse group comprising 36 early-flowering and 67 late-flowering accessions from Pop-1 (84), Pop-2 (13), Pop-3 (two), and Pop-5 (four). Cluster-II was also a diverse group in terms of populations containing 38 early- and 33 late-flowering accessions from Pop-1 (17), Pop-2 (15), Pop-3 (four), Pop-4 (29), Pop-5 (four), Pop-6 (one), and Pop-7 (one). Cluster-III comprised one accession of Pop-1 and six accessions of Pop-3, all of which were a late flowering type.

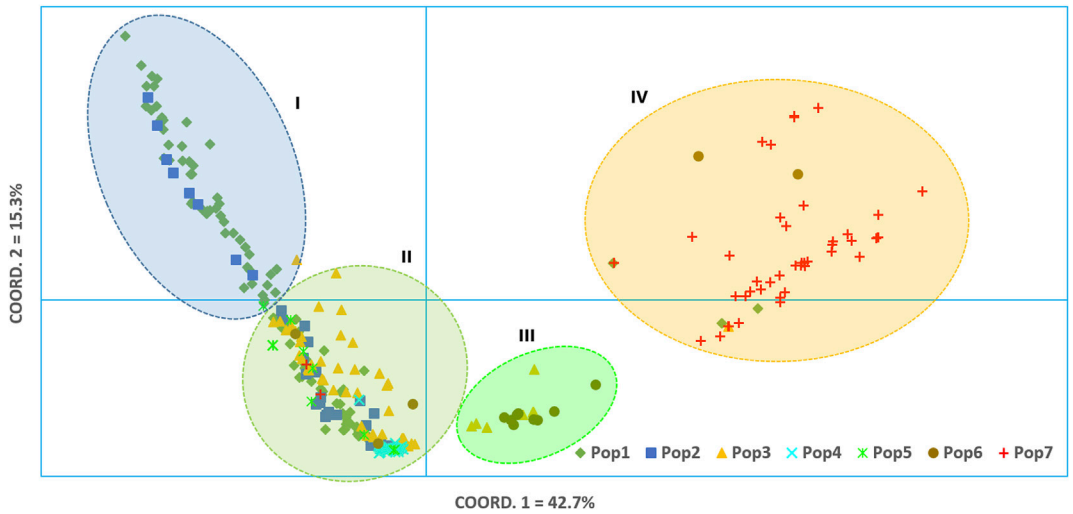


FIGURE 7 | Bi-plot of principal coordinate analysis (PCoA) generated using 5,226 SNP markers depicting the genetic relationship among the 288 finger millet accessions. A symbol of the same shape and color represents accessions of the same population.

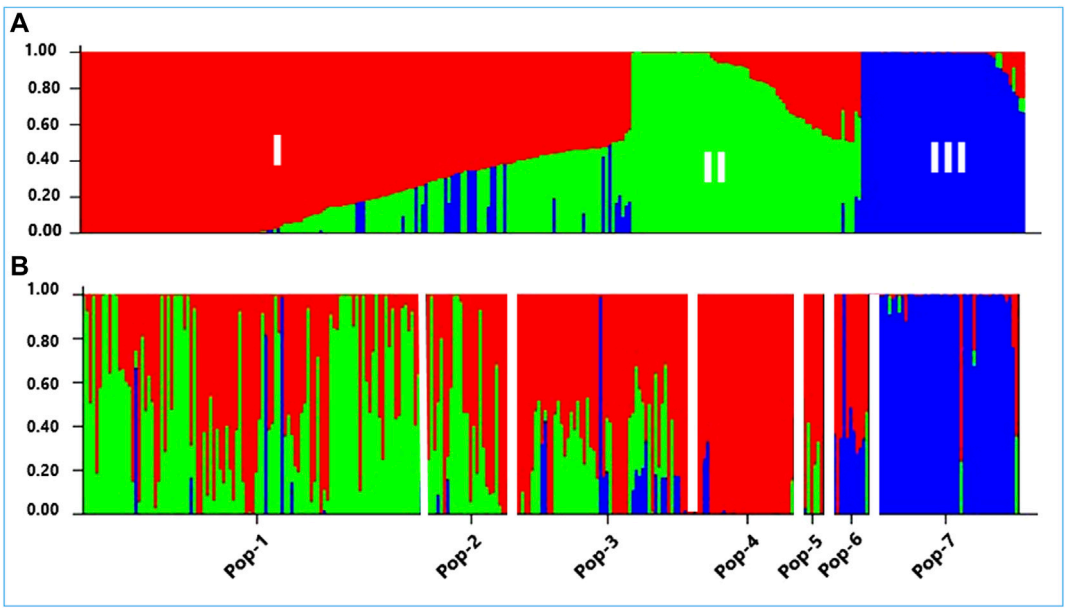


FIGURE 8 | A graphical display of the genetic structure of 288 accessions of finger millet at K = 3, forming three clusters (shown by different colors) and exhibiting different levels of admixture. **(A)** Individual accessions were arranged according to the level of their membership in different clusters and **(B)** individual accessions were arranged according to their assigned populations (Pop-1 to Pop-7) comprising 105, 31, 55, 28, 9, 14, and 46 accessions, respectively.

Cluster-IV is composed of 14 early-flowering and 15 late-flowering type, of which 26 accessions were from Pop-3, and the remaining three accessions were from Pop-5, Pop-6, and Pop-7. Cluster-V comprised 10 accessions (two early- and eight late-flowering type) from Pop-3 (nine) and Pop-6 (one). Cluster-VI was the smallest cluster with four late-flowering accessions from Pop-3 (three) and Pop-4 (one). Another large and heterogeneous group was Cluster-VII comprising 46 early- and 17 late-flowering accessions from Pop-1 (three), Pop-3 (four), Pop-6 (12), and Pop-7 (44) (**Figure 6**).

Principal coordinate analysis (PCoA) was used to visualize the differentiation among the 288 accessions (**Figure 7**). The analysis revealed that the first and second coordinates together explained 58% of the total variation. In spite of the lack of distinct clustering, the 288 accessions could be broadly categorized into four clusters, mainly based on the first coordinate that

explained 42.7% of the total variation. Cluster-I comprised almost exclusively accessions from Pop-1 and Pop-2, only with additional two accessions from Pop-3 and Pop-5. Cluster-II was the most heterogenous group comprising accessions from all the seven populations. Interestingly, all accessions of Pop-4 were tightly grouped in this cluster. Cluster-III was the smallest, comprising six accessions of Pop-3 and nine accessions of Pop-6. The vast majority of accessions from Zimbabwe (Pop-7) were placed in Cluster-IV, containing three accessions from Pop-1, one accession from Pop-3, and two accessions from Pop-6 (**Figure 7**).

Population Structure Analysis

According to the admixture model-based population genetic structure analysis conducted using the 5,226 SNP markers, three genetic populations represent the 288 accessions the best (**Figure 8**). This analysis revealed that Pop-1, Pop-2, Pop-3, and Pop-5 were dominated by alleles from the first two genetic populations (red and green). Population-4 was the least diverse among the Ethiopian finger millet populations, with its alleles almost entirely originating from the first genetic population (red). The improved cultivars appeared to be the results of strong admixture between the second and third genetic populations (green and blue). In agreement with the results of cluster analysis and PCoA, accessions from Zimbabwe were, in general, distinctly separated from the Ethiopian accessions by having the vast majority of their alleles from the third genetic population (blue) (**Figure 8**).

DISCUSSION

The Single Nucleotide Polymorphism Markers

The GBS method that uses two restriction enzymes is a highly effective method for reducing genome complexity and generating high-quality SNP markers in several crops, including barley, wheat (Poland et al., 2012), and cacao (Osorio-Guarin et al., 2020), which was also demonstrated in the present study. The GBS method has been used previously for simultaneous SNP discovery and genotyping in finger millet (Kumar A. et al., 2016; Gimode et al., 2016; Qi et al., 2018; Tiwari et al., 2020). However, none of them used the *Pst*I/*ApeK*I methylation-sensitive restriction enzyme combination. Hence, the SNPs identified in this study represent a novel genomic resource for finger millet as the target genomic regions were different from those in the previous study. The final set of 5,226 SNP markers used in this study were distributed across 2,500 scaffolds with only one SNP per scaffold in most cases, indicating a wide distribution of these SNPs across the finger millet genome. Hence, the *Pst*I/*ApeK*I enzyme combination is suitable for GBS-based SNP discovery in finger millet and closely related species. The allele frequencies of most of these SNPs exceed 10%, with a quarter exceeding 27%, which makes them ideal for population genetics and genome-wide association studies (GWAS). The polymorphism information content (PIC) is a measure of the informativeness of molecular markers, and the higher the PIC value of a locus is, the more informative it is (Hildebrand et al., 1994; Shete et al., 2000). A bi-allelic SNP locus can have a maximum PIC value of 0.38, which is attained when both alleles are equally frequent. The SNP loci in the present study have an average PIC of 0.23, and a quarter have PIC values over 0.32, making them highly informative. The average PIC (0.23) obtained in this study was lower than the average PIC (0.29) value estimated for finger millet accessions in Gimode et al. (2016). The slightly higher mean PIC value in Gimode et al. (2016) could most likely be due to the use of more diverse accessions (wild and cultivated) and much fewer SNP loci (80) with more frequent alleles when compared to the present study (MAF > 0.05). Similarly, it is also lower than the average PIC of earlier studies using expressed sequence tags and genomic microsatellites (SSR) markers (Gimode et al., 2016; Krishna et al., 2018; Lule et al., 2018; Pandian et al., 2018; Brhane et al., 2021). The PIC values of SNP markers are generally lower than those of SSR markers, as the average number of alleles per locus is higher in SSRs than in bi-allelic SNPs (Chao et al., 2009).

Finger millet is a highly self-pollinating crop. Hence, the overall expected heterozygosity (He) is expected to be higher than the observed heterozygosity (Ho) if all other HWE assumptions are met. In line with the expectation, the mean He value (0.28) was higher than the mean Ho value (0.10) in the present study. A recent study in sorghum (a self-pollinating cereal crop) using 3001 SNP markers reported mean He and Ho values of 0.10 and 0.06 (Enyew et al., 2021). The lower values in the latter case are most likely because most of the SNP markers used were located within genes (Enyew et al., 2021) and hence are expected to be less diverse than GBS-based SNP markers. Similar bi-allelic SNP-based studies in noug (*Guizotia abyssinica*; Tsehay et al., 2020) and red clover (*Trifolium pratense*; Osterman et al., 2021) reported Ho values of 0.24 and 0.22, respectively. It is expected that noug and red clover have higher Ho values than finger millet and sorghum because both of these crops are strictly outcrossing, enabling a higher level of heterozygosity. According to the HWE test, 97.5% of the loci showed significant deviation from HWE, with 86.5% of loci showing heterozygote deficiency. Due to the self-pollinating nature of finger millet, there is a high likelihood of heterozygote deficiency at the vast majority of loci, although other evolutionary forces may have played a role as well. In contrast, approximately one-tenth of the SNP loci displayed excess heterozygosity, which is interesting. The result of the present study suggests that selection pressures that favor heterozygosity are at work at these loci, and further studies targeting these loci will shed more light on their significance. As indicated by Tsehay et al. (2020), loci with heterozygote excess or deficit contrary to the reproductive mechanism of a plant indicate that such loci are under various types and intensity of selection and other evolutionary forces.

Genetic Diversity Within Groups of Accessions

The genetic characterization of germplasm using codominant DNA markers is a crucial step toward developing stable and productive cultivars with various desirable traits. In the present study, 5,226 GBS-based SNP markers were used to examine the pattern and level of genetic diversity across 288 finger millet accessions representing the finger millet gene pool in Ethiopia and Zimbabwe. According to the result of site-frequency spectrum analysis, there is substantial genetic diversity among the seven finger millet populations, as evidenced by variation in minor allele frequency distributions between them. However, individual accessions in different populations generally bear minor alleles at fewer loci compared to results obtained in a similar study in sorghum (Enyew et al., 2021). More than 50% of individuals in Pop-5 and over 20% of individuals in Pop-6 had minor alleles in at least 10% of the SNP loci, as opposed to the other populations with minor alleles in less than 8% of the loci. Pop-6 represents released cultivars, and hence it is likely that minor alleles play a significant role in their improved traits. Although Pop-5 lacks passport data, it would be interesting to investigate them further to uncover their phenotypic diversity of desirable traits.

The level of genetic diversity observed across the populations in this study was comparable to that of earlier studies in finger millet based on SSR markers (Babu et al., 2014; Nirgude et al., 2014; Ramakrishnan et al., 2016; Pandian et al., 2018). Among the seven finger millet populations, Pop-1 was the most diverse as estimated based on different parameters (Ne = 1.44; I = 0.40; He = 0.26; PPL = 0.96). Because the sample size of this population is larger than that of the other populations and the accessions originated from broader areas and diverse agro-ecological environments, their genetic diversity is expected to be higher. However, sample size does not fully explain the lowest genetic diversity recorded in Pop-4 as Pop-5 and Pop-6 had a smaller sample size than Pop-4. For example, the sample size of Pop-2 (northern accessions) and Pop-4 (northeastern accessions) is quite similar, but the gene diversity of the former was twice that of the latter, clearly indicating low genetic diversity in finger millet cultivated in the northeastern part of Ethiopia. In addition to its relatively high genetic diversity, Pop-1 (northwestern accessions) came on top in terms of the private alleles it harbors (13). Hence, the region represented by Pop-1 may be considered a hotspot for finger millet genetic diversity in Ethiopia, which is appropriate for the crop's *in situ* conservation because finger millet is cultivated in large quantities in this region. The average values of different estimates of genetic variation of the four Ethiopian landrace populations (Pop-1 to Pop-4) were similar to those of the Zimbabwean landrace population (Pop-7). Hence, the finger millet gene pools in the two countries can be considered equally diverse. The level of genetic diversity of the population representing the cultivars (Pop-6) was within the diversity range of the landrace populations, which suggests that at least some cultivars were developed based on different finger millet genetic resources. It is noteworthy that the observed heterozygosity (Ho) of the cultivars was quite similar to that of the landraces, indicating that the cultivars are not pure lines. This indicates that their further improvement is possible through self-pollination.

Among the three groups of accessions classified based on the panicle types, the genetic diversity of those with open panicles was slightly higher than that of the other two groups. This could be mainly because of the sample size effects as accessions with open panicles accounted for two-thirds of the 288 accessions, representing wider geographic areas. Finger millet with open panicles is preferred for cultivation in Ethiopia due to its better grain yield, drought tolerance, resistance to bird attack and shattering, and ease to thresh (Tsehaye and Kebebew, 2002; Bezaweletaw et al., 2007). Hence, their higher genetic diversity is desirable for further improvement through breeding. On the contrary, slightly higher genetic diversity was recorded in the early flowering group than in the late flowering group, regardless of the smaller sample size used. Similarly, a slightly greater genetic diversity was found in the Al-tolerant group than in the Al-susceptible group, despite a roughly fourfold larger sample size in the latter. The accessions in the late-maturing group and Al-susceptible group were exclusively Ethiopian, whereas both Ethiopia and Zimbabwe were widely represented in the early maturing group and Al-tolerant group, which may explain in part the higher diversity level in the latter groups. Private alleles were found in both the early maturing and the Al-tolerant groups, and their potential associations with desirable characteristics require further investigation. Private alleles that are strongly associated with a desirable trait can be used in marker-assisted selection (MAS) and selection of parental lines for crossbreeding based development of superior cultivars with multiple desirable traits (Park et al., 2008). Early flowering and tolerance to aluminum toxicity are desirable traits in finger millet. Hence, based on landraces that combine early flowering and Al tolerance, improved cultivars with characteristics of interest can be developed, taking advantage of their genetic diversity.

Genetic Relationships Among Accessions and Their Groups and Population Genetic Structure

Analysis of molecular variance (AMOVA) revealed highly significant genetic differentiations among groups of accessions classified based on the geographic region of origin, country of origin, flowering time, panicle shape, and Al tolerance ($p < 0.01$). The significant differentiation among groups of finger millet accessions was in agreement with the previously published research using EST and genomic SSR markers (Babu et al., 2014; Pandian et al., 2018; Brhane et al., 2021). In the present study, the highest level of differentiation was observed between Ethiopian and Zimbabwean accessions accounting for 43% of the total genetic variation (**Table 2**). The generally distinct differentiation between the two groups was also evident from the genetic distance data, average pairwise differences, cluster analysis, PCoA, and population structure (**Table 3**; **Figures 5–8**). Phenotypically, Zimbabwean accessions are dominated by early flowering plants with shorter and stronger stems, compact

panicles, and better Al tolerance compared to the Ethiopian accessions. Accessions with desirable traits such as early flowering and Al tolerance exist in both groups. Hence, crossbreeding between selected finger millet genotypes from the two countries increases the likelihood of recombination of alleles that will eventually lead to superior cultivars. The Ethiopian accession 208440 and Zimbabwean accession 203347 are excellent examples in this regard as both are early flowering and Al-tolerant (**Supplementary Table S1**). However, the high genetic distance between them placed them separately in Cluster-II and Cluster-VII, respectively (**Figure 6**).

Eleven of the 14 released cultivars were clustered together in Cluster-VII, with nine of them showing tight clustering. A recent study based on SSR markers reported similar close clustering of finger millet cultivars (Brhane et al., 2021). The clustering of cultivars close together indicates that they were developed based on genetically similar germplasm and/or selected for similar traits during breeding. Because most Ethiopian cultivars clustered with the Zimbabwean landrace accessions, germplasm genetically similar to the latter may have been used in the development of the cultivars. Among the Ethiopian landrace accessions, those in Pop-3 (western accessions) appeared to be genetically closer to the cultivars, except for cultivar Diga-2 (**Figure 6**). Thus, finger millet cultivar development might have heavily relied on germplasm from western Ethiopia. Nevertheless, further research into the pedigrees of the cultivars and potential gene flow after their release are required to shed more light on this. The clustering pattern of accessions along the lines of days to flowering, aluminum tolerance, and panicle types did not resolve well in the cluster analysis and PCoA. However, accessions carrying the desirable characteristics of these traits are genetically diverse and can be improved through breeding due to potential transgressive segregations (Ortiz et al., 2020).

The cluster analysis and PCoA clearly showed that Pop-1 (northwestern accessions) and Pop-2 (northern accessions) were poorly differentiated, suggesting stronger gene flow between them. However, some of the accessions in these populations are different from those in other populations, as shown in Cluster-I of the PCoA (**Figure 7**). Pop-4 is quite unique in that most of its accessions are genetically highly similar (low genetic diversity) and show close genetic relationships with some members of all other populations, indicating its low genetic differentiation. The fact that accessions representing all populations were found within Cluster-II of PCoA strongly suggests a countrywide gene flow, albeit to a different degree. Overall, the clustering pattern of accessions generally agrees with a recent study on Ethiopian finger millet accessions based on genomic and EST-derived SSR markers (Brhane et al., 2021).

An important feature of the Bayesian statistical approach to population structure analysis is that it enables the identification of genotypes that originate purely from one genetic population or are the results of genetic admixture (Pritchard et al., 2000; Kumar A. et al., 2016). This study revealed that the 288 genotypes from diverse sources originate from three genetic populations (K = 3). Similar to the present study, population structure analyses in previous research on finger millet using SSR and SNP markers revealed three genetic groups with different levels of admixture (Dida et al., 2008; Kumar A. et al., 2016; Ramakrishnan et al., 2016; Lule et al., 2018; Pandian et al., 2018; Brhane et al., 2021). In the present study, the Zimbabwean accessions and Ethiopian accessions appeared to have originated from different genetic populations with few exceptions, which is expected due to the clear geographic separation between the two countries. In addition, it is interesting to note that the Zimbabwean accessions were much less admixed than the Ethiopian landrace accessions, except for those in Pop-4 (**Figure 8**). The genetic structure of Pop-1 and Pop-2 is highly similar, which indicates that gene flow is stronger between the geographic areas the accessions represent.

CONCLUSION

In the present study, over ten thousand SNP markers were detected using GBS, providing a new genomic resource for finger millet. The characterization of 5,226 of these SNPs in diverse finger millet populations from Ethiopia and Zimbabwe shows that they are highly informative and, therefore, suitable for different applications, such as genome-wide association studies and population genetics. The observed heterozygosity among landraces and cultivars is low, agreeing with the crop's reproductive mechanism. However, there was excess heterozygosity at about one-tenth of the SNP loci, although the crop is mainly self-pollinating, suggesting evolutionary forces favoring heterozygosity may be at play at these loci. Future research focusing on these loci will provide more insight into their role. In order to determine whether private alleles found in different accession groups have potential associations with desired traits, further investigation is needed. Within Ethiopia, finger millet landrace accessions from different geographic regions differ moderately in terms of genetic diversity, with those from northeastern Ethiopia being the least diverse. The fact that some accessions from different regions clustered closely together suggests a countrywide gene flow, though to a different extent. The genetic differentiation among accessions classified by geographic region, country of origin, days to flowering, panicle type, and Al tolerance is significantly high, with differentiation among countries being the highest. In the case of improved cultivars, most of them clustered tightly together, suggesting that they were developed from similar germplasm and/or selected for the same traits, mainly grain yield. In addition, the level of their heterozygosity suggests that they can be further improved through self-pollination-based breeding. By using genetically distinct accessions from different geographic regions or countries, crossbreeding can potentially lead to the development of superior cultivars due to the recombination of alleles.

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 the GBS reads: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA791522.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology: HB, MG, TH, KT, CH, and RO. Data collection: HB under the guidance of MG and CH. Data analysis: HB under the guidance of MG. Original draft preparation: HB. Review and editing: HB, TH, KT, CH, RO, KBA, and MG. Project funding acquisition and administration: TH, KT, RO, MG, and CH. All authors have read and approved the final version of the manuscript.

FUNDING

This study was funded by the Swedish International Development Cooperation Agency (Sida) Research and Training grant awarded to Addis Ababa University and the Swedish University of Agricultural Sciences (AAU-SLU Biotech; https://sida.aau.edu.et/index.php/biotechnology-phd-program/; accessed on 27 December 2021).

ACKNOWLEDGMENTS

The authors wish to thank the Swedish International Development Cooperation Agency (Sida) for financing this work. They would also like to thank the Ethiopian Biodiversity Institute and Bako Agricultural Research Center for providing the finger millet germplasm used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.848627/full#supplementary-material

REFERENCES

Admassu, S., Teamir, M., and Alemu, D. (2009). Chemical Composition of Local and Improved finger Millet [*Eleusine Corocana* (L.) Gaetrtin] Varieties Grown in Ethiopia. *Ethiopian J. Health Sci.* 19.

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., and Zhang, G. (2017). Genotyping-by-sequencing (GBS) Revealed Molecular Genetic Diversity of Iranian Wheat Landraces and Cultivars. *Front. Plant Sci.* 8, 1293. doi:10.3389/fpls.2017.01293

Antony Ceasar, S., Maharajan, T., Ajeesh Krishna, T. P., Ramakrishnan, M., Victor Roch, G., Satish, L., et al. (2018). Finger Millet [*Eleusine Coracana* (L.) Gaertn.] Improvement: Current Status and Future Interventions of Whole Genome Sequence. *Front. Plant Sci.* 9, 1054. doi:10.3389/fpls.2018.01054

Arora, A., Kundu, S., Dilbaghi, N., Sharma, I., and Tiwari, R. (2014). Population Structure and Genetic Diversity Among Indian Wheat Varieties Using Microsatellite (SSR) Markers. *Aust. J. Crop Sci.* 8.

Babu, B. K., Agrawal, P. K., Pandey, D., and Kumar, A. (2014). Comparative Genomics and Association Mapping Approaches for Opaque2 Modifier Genes in finger Millet Accessions Using Genic, Genomic and Candidate Gene-Based Simple Sequence Repeat Markers. *Mol. Breed.* 34, 1261–1279. doi:10.1007/s11032-014-0115-2

Barbeau, W. E., and Hilu, K. W. (1993). Protein, Calcium, Iron, and Amino Acid Content of Selected Wild and Domesticated Cultivars of finger Millet. *Plant Food Hum. Nutr.* 43, 97–104. doi:10.1007/BF01087914

Batley, J., and Edwards, D. (2007). "SNP Applications in Plants," in *Association Mapping in Plants* (Springer), 95–102. doi:10.1007/978-0-387-36011-9_6

Bezaweletaw, K., Sripichitt, P., Wongyai, W., and Hongtrakul, V. (2006). Genetic Variation, Heritability and Path-Analysis in Ethiopian finger Millet [*Eleusine Corocana* (L.) Gaetrtin] Landraces. *Agric. Nat. Resour.* 40, 322–334.

Bezaweletaw, K., Sripichitt, P., Wongyai, W., and Hongtrakul, V. (2007). Phenotypic Diversity of Ethiopian finger Millet [*Eleusine Coracana* (L.) Gaertn] in Relation to Geographical Regions as an Aid to Germplasm Collection and Conservation Strategy. *Agric. Nat. Resour.* 41, 7–16.

Bisht, M. S., and Mukai, Y. (2002). Genome Organization and Polyploid Evolution in the Genus *Eleusine* (Poaceae). *Plant Syst. Evol.* 233, 243–258. doi:10.1007/s00606-002-0201-5

Bisht, M. S., and Mukai, Y. (2001). Genomic *In Situ* Hybridization Identifies Genome Donor of finger Millet (*Eleusine Coracana*). *Theor. Appl. Genet.* 102, 825–832. doi:10.1007/s001220000497

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120.

Brhane, H., Haileselassie, T., Tesfaye, K., Hammenhag, C., Ortiz, R., Abreha, K. B., et al. (2021). Novel Expressed Sequence Tag-Derived and Other Genomic Simple Sequence Repeat Markers Revealed Genetic Diversity in Ethiopian Finger Millet Landrace Populations and Cultivars. *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.735610

Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M.-C., et al. (2009). Analysis of Gene-Derived SNP Marker Polymorphism in US Wheat (*Triticum aestivum* L.) Cultivars. *Mol. Breed.* 23, 23–33. doi:10.1007/s11032-008-9210-6

Chethan, S., and Malleshi, N. G. (2007). Finger Millet Polyphenols: Characterization and Their Nutraceutical Potential. *Am. J. Food Tech.* 2, 582–592. doi:10.3923/ajft.2007.582.592

Cooke, T. F., Yee, M.-C., Muzzio, M., Sockell, A., Bell, R., Cornejo, O. E., et al. (2016). GBStools: a Statistical Method for Estimating Allelic Dropout in Reduced Representation Sequencing Data. *PLoS Genet.* 12, e1005631. doi:10.1371/journal.pgen.1005631

Delhaize, E., and Ryan, P. R. (1995). Aluminum Toxicity and Tolerance in Plants. *Plant Physiol.* 107, 315. doi:10.1104/pp.107.2.315

Dida, M. M., Wanyera, N., Dunn, M. L. H., Bennetzen, J. L., and Devos, K. M. (2008). Population Structure and Diversity in finger Millet (*Eleusine Coracana*) Germplasm. *Trop. Plant Biol.* 1, 131–141. doi:10.1007/s12042-008-9012-3

Eekhout, T., Larsen, P., and De Veylder, L. (2017). Modification of DNA Checkpoints to Confer Aluminum Tolerance. *Trends Plant Science* 22, 102–105. doi:10.1016/j.tplants.2016.12.003

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

Enyew, M., Feyissa, T., Carlsson, A. S., Tesfaye, K., Hammenhag, C., and Geleta, M. (2021). Genetic Diversity and Population Structure of Sorghum [*Sorghum Bicolor* (L.) Moench] Accessions as Revealed by Single Nucleotide Polymorphism Markers. *Fronters Plant Sci.* 12, 799482. doi:10.3389/fpls.2021.799482

Erenso, D., Asfaw, A., Taye, T., and Tesso, T. (2007). "Genetic Resources, Breeding and Production of Millets in Ethiopia," in Organizing Committee of the International Conference on New Approaches to, Bern, Switzerland, 19-21 September 2007, 43–56. Proceedings of an International Conference.*(New Approaches to Plant Breeding of Orphan Crops in Africa*

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin Suite Ver 3.5: a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Fakrudin, B., Shashidhar, H., Kulkarni, R., and Hittalmani, S. (2004). Genetic Diversity Assessment of finger Millet, *Eleusine Coracana* (Gaertn), Germplasm through RAPD Analysis. *PGR Newslett* 138, 50–54.

Faostat (2020). *Statistical Databases and Data-Sets of the Food and Agriculture Organization of the United Nations*.

Garrison, E., and Marth, G. (2012). Haplotype-based Variant Detection from Short-Read Sequencing. *arXiv preprint arXiv:1207.3907*.

Geleta, M., Gustafsson, C., Glaubitz, J., and Ortiz, R. (2020). High-density Genetic Linkage Mapping of Lepidium Based on Genotype-By-Sequencing and Segregating Contig Tag Haplotypes. *Front. Plant Sci.* 11, 448. doi:10.3389/fpls.2020.00448

Gimode, D., Odeny, D. A., De Villiers, E. P., Wanyonyi, S., Dida, M. M., Mneney, E. E., et al. (2016). Identification of SNP and SSR Markers in finger Millet Using Next Generation Sequencing Technologies. *PLoS one* 11, e0159437. doi:10.1371/journal.pone.0159437

Goron, T. L., and Raizada, M. N. (2015). Genetic Diversity and Genomic Resources Available for the Small Millet Crops to Accelerate a New Green Revolution. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.00157

Hammenhag, C., Saripella, G. V., Ortiz, R., and Geleta, M. (2020). QTL Mapping for Domestication-Related Characteristics in Field Cress (*Lepidium Campestre*)—a Novel Oil Crop for the Subarctic Region. *Genes* 11, 1223. doi:10.3390/genes11101223

Hildebrand, C. E., Torney, D. V., and Wagner, R. P. (1994). "Informativeness of Polymorphic DNA Markers," in, *The Human Genome Project: Decipering the Blueprint of Heredity*. Editor N. G. Cooper (Mill Valley, CA, USA: University Science Books), 100–102.

Hilu, K. W., and De Wet, J. (1976). Domestication of *Eleusine Coracana*. *Econ. Bot.* 30, 199–208. doi:10.1007/bf02909728

Hittalmani, S., Mahesh, H., Shirke, M. D., Biradar, H., Uday, G., Aruna, Y., et al. (2017). Genome and Transcriptome Sequence of finger Millet (*Eleusine Coracana* (L.) Gaertn.) Provides Insights into Drought Tolerance and Nutraceutical Properties. *BMC Genomics* 18, 465. doi:10.1186/s12864-017-3850-z

Huang, Y.-F., Poland, J. A., Wight, C. P., Jackson, E. W., and Tinker, N. A. (2014). Using Genotyping-By-Sequencing (GBS) for Genomic Discovery in Cultivated Oat. *PloS one* 9, e102448. doi:10.1371/journal.pone.0102448

Kebede, D., Dagnachew, L., Megersa, D., Chemeda, B., Girma, M., Geleta, G., et al. (2019). Genotype by Environment Interaction and Grain Yield Stability of Ethiopian Black Seeded finger Millet Genotypes. *Afr. Crop Sci. J.* 27. doi:10.4314/acsj.v27i2.12

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a Program for Identifying Clustering Modes and Packaging Population Structure Inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi:10.1111/1755-0998.12387

Kopittke, P. M., and Blamey, F. P. C. (2016). Theoretical and Experimental Assessment of Nutrient Solution Composition in Short-Term Studies of Aluminium Rhizotoxicity. *Plant and Soil* 406, 311–326. doi:10.1007/s11104-016-2890-5

Krishna, T. P. A., Maharajan, T., Antony David, R. H., Ramakrishnan, M., Ceasar, S. A., Duraipandiyan, V., et al. (2018). Microsatellite Markers of finger Millet (Eleusine Coracana (L.) Gaertn) and Foxtail Millet (*Setaria Italica* (L.) Beauv) Provide Resources for Cross-Genome Transferability and Genetic Diversity Analyses in Other Millets. *Biocatal. Agric. Biotechnol.* 16, 493–501. doi:10.1016/j.bcab.2018.09.009

Kumar, A., Sharma, D., Tiwari, A., Jaiswal, J., Singh, N., and Sood, S. (2016a). Genotyping-by-sequencing Analysis for Determining Population Structure of finger Millet Germplasm of Diverse Origins. *The plant genome* 9, 1–15. doi:10.3835/plantgenome2015.07.0058

Kumar, S., Banks, T. W., and Cloutier, S. (20122012). SNP Discovery through Next-Generation Sequencing and its Applications. *Int. J. Plant genomics*. doi:10.1155/2012/831460

Kumar, S., Stecher, G., and Tamura, K. (2016b). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, Y. L., and Liu, J. X. (2018). StructureSelector: A Web-based Software to Select and Visualize the Optimal Number of Clusters Using Multiple Methods. *Mol. Ecol. Resour.* 18, 176–177. doi:10.1111/1755-0998.12719

Lule, D., De Villiers, S., Fetene, M., Odeny, D. A., Rathore, A., Das, R. R., et al. (2018). Genetic Diversity and Association Mapping of Ethiopian and Exotic finger Millet Accessions. *Crop Pasture Sci.* 69, 879–891. doi:10.1071/CP18175

Molla, F. (2010). *Genotype X Environment Interaction and Stability Analyses of Yield and Yield Related Traits of finger Millet (Eleusine Coracana (L) Gaertn) Varieties in North Western Ethiopia*. Haramaya, Ethiopia: M. Sc. thesis presented to the School of Graduate Studies of Haramaya University.

Mukarumbwa, P., and Mushunje, A. (2010). Potential of Sorghum and finger Millet to Enhance Household Food Security in Zimbabwe's Semi-arid Regions: A Review. doi.doi:10.22004/ag.econ.96430

Nakarani, U. M., Singh, D., Suthar, K. P., Karmakar, N., Faldu, P., and Patil, H. E. (2021a). Nutritional and Phytochemical Profiling of Nutracereal finger Millet (*Eleusine Coracana* L.) Genotypes. *Food Chem.* 341, 128271. doi:10.1016/j.foodchem.2020.128271

National Research Council (1996). *Lost Crops of Africa: Volume I: Grains*. Washington, D.C.: National Academies Press. doi:10.17226/2305

Nei, M. (1972). Genetic Distance between Populations. *The Am. Naturalist* 106, 283–292. doi:10.1086/282771

Nirgude, M., Babu, B. K., Shambhavi, Y., Singh, U., Upadhyaya, H., and Kumar, A. (2014). Development and Molecular Characterization of Genic Molecular Markers for Grain Protein and Calcium Content in finger Millet (*Eleusine Coracana* (L.) Gaertn.). *Mol. Biol. Rep.* 41, 1189–1200. doi:10.1007/s11033-013-2825-7

Osorio-Guarín, J. A., Berdugo-Cely, J. A., Coronado-Silva, R. A., Baez, E., Jaimes, Y., and Yockteng, R. (2020). Genome-wide Association Study Reveals Novel Candidate Genes Associated with Productivity and Disease Resistance to Moniliophthora Spp. In Cacao (*Theobroma Cacao* L.). *G3: Genes, Genomes, Genet.* 10, 1713–1725. doi:10.1534/g3.120.401153

Osterman, J., Hammenhag, C., Ortiz, R., and Geleta, M. (2021). Insights into the Genetic Diversity of Nordic Red clover (*Trifolium Pratense*) Revealed by SeqSNP-Based Genetic Markers. *Front. Plant Sci.* 12, 748750. doi:10.3389/fpls.2021.748750

Pandian, S., Satish, L., Rameshkumar, R., Muthuramalingam, P., Rency, A. S., Rathinapriya, P., et al. (2018). Analysis of Population Structure and Genetic Diversity in an Exotic Germplasm Collection of Eleusine Coracana (L.) Gaertn. Using Genic-SSR Markers. *Gene* 653, 80–90. doi:10.1016/j.gene.2018.02.018

Park, Y.-J., Dixit, A., Ma, K.-H., Lee, J.-K., Lee, M.-H., Chung, C.-S., et al. (2008). Evaluation of Genetic Diversity and Relationships within an On-Farm Collection of *Perilla Frutescens* (L.) Britt. Using Microsatellite Markers. *Genet. Resour. Crop Evol.* 55, 523–535. doi:10.1007/s10722-007-9258-x

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes* 6, 288–295. doi:10.1111/j.1471-8286.2005.01155.x

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-By-Sequencing Approach. *PloS one* 7, e32253. doi:10.1371/journal.pone.0032253

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for Plant Breeding and Genetics. *The Plant Genome* 5. doi:10.3835/plantgenome2012.05.0005

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

Puechmaille, S. J. (2016). The Program Structure Does Not Reliably Recover the Correct Population Structure when Sampling Is Uneven: Subsampling and New Estimators Alleviate the Problem. *Mol. Ecol. Resour.* 16, 608–627. doi:10.1111/1755-0998.12512

Qi, P., Gimode, D., Saha, D., Schröder, S., Chakraborty, D., Wang, X., et al. (2018). UGbS-Flex, a Novel Bioinformatics Pipeline for Imputation-free SNP Discovery in Polyploids without a Reference Genome: finger Millet as a Case Study. *BMC Plant Biol.* 18, 1–19. doi:10.1186/s12870-018-1316-3

Ramakrishnan, M., Ceasar, S. A., Duraipandiyan, V., Al-Dhabi, N. A., and Ignacimuthu, S. (2016). Assessment of Genetic Diversity, Population Structure and Relationships in Indian and Non-Indian Genotypes of finger Millet (*Eleusine Coracana* (L.) Gaertn) Using Genomic SSR Markers. *Springerplus* 5. doi:10.1186/s40064-015-1626-y

Rozas, J., Sánchez-Delbarrio, J. C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods. *Bioinformatics* 19, 2496–2497. doi:10.1093/bioinformatics/btg359

Shete, S., Tiwari, H., and Elston, R. C. (2000). On Estimating the Heterozygosity and Polymorphism Information Content Value. Theoretical Population Biology 57, 265–271. doi:10.1006/tpbi.2000.1452

Sri, S., Dida, M. M., Gale, M. D., and Devos, K. M. (2007). Comparative Analyses Reveal High Levels of Conserved Colinearity between the finger Millet and rice Genomes. *Theor. Appl. Genet.* 115, 489–499. doi:10.1007/s00122-007-0582-5

Tiwari, A., Sharma, D., Sood, S., Jaiswal, J. P., Pachauri, S. P., Ramteke, P. W., et al. (2020). Genome-wide Association Mapping for Seed Protein Content in finger Millet (*Eleusine Coracana*) Global Collection through Genotyping by Sequencing. *J. Cereal Sci.* 91, 102888. doi:10.1016/j.jcs.2019.102888

Tsehay, S., Ortiz, R., Johansson, E., Bekele, E., Tesfaye, K., Hammenhag, C., et al. (2020). New Transcriptome-Based SNP Markers for Noug (*Guizotia Abyssinica*) and Their Conversion to KASP Markers for Population Genetics Analyses. *Genes* 11, 1373. doi:10.3390/genes11111373

Tsehaye, T., and Kebebew, F. (2002). Morphological Diversity and Geographic Distribution of Adaptive Traits in finger Millet (*Eleusine Coracana* (L.) Gaertn.[Poaceae]) Populations from Ethiopia. *Ethiopian J. Biol. Sci.* 1, 37–62.

Upadhyaya, H., Gowda, C., and Reddy, V. G. (2007). Morphological Diversity in finger Millet Germplasm Introduced from Southern and Eastern Africa. *J. SAT Agric. Res.* 3, 1–3.

Zhang, H., Jiang, Z., Qin, R., Zhang, H., Zou, J., Jiang, W., et al. (2014). Accumulation and Cellular Toxicity of Aluminum in Seedling of Pinus Massoniana. *BMC Plant Biol.* 14, 1–16. doi:10.1186/s12870-014-0264-9

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 Brhane, Haileselassie, Tesfaye, Ortiz, Hammenhag, Abreha and Geleta. 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 May 2022
doi: 10.3389/fgene.2022.849357

Heterologous Expression of Arabidopsis *AtARA6* in Soybean Enhances Salt Tolerance

Zhipeng Hong[1†], Yang Li[1†], Yang Zhao[1†], Mingyu Yang[1], Xiaoming Zhang[1], Yuhan Teng[1], Linjie Jing[1], Danxun Kong[1], Tongxin Liu[1], Shuanglin Li[1], Fanli Meng[1], Qi Wang[2]* and Ling Zhang[3]**

[1]Key Laboratory of Soybean Biology in Chinese Ministry of Education, Northeast Agricultural University, Harbin, China, [2]Institute of Crop Cultivation and Tillage, Heilongjiang Academy of Agricultural Sciences, Harbin, China, [3]Agro-Biotechnology Research Institute, Jilin Academy of Agricultural Sciences, Changchun, China

OPEN ACCESS

Edited by:
Sreepriya Pramod,
Altria, United States

Reviewed by:
Chathurani Ranathunge,
Eastern Virginia Medical School,
United States
Riza Putranto,
Indonesian Research Institute for Biotechnology and Bioindustry (IRIBB),
Indonesia

***Correspondence:**
Fanli Meng
mengfanli@neau.edu.cn
Qi Wang
neauwq@163.com
Ling Zhang
zly_jaas@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: 06 January 2022
Accepted: 23 March 2022
Published: 12 May 2022

Citation:
Hong Z, Li Y, Zhao Y, Yang M, Zhang X, Teng Y, Jing L, Kong D, Liu T, Li S, Meng F, Wang Q and Zhang L (2022) Heterologous Expression of Arabidopsis AtARA6 in Soybean Enhances Salt Tolerance. Front. Genet. 13:849357. doi: 10.3389/fgene.2022.849357

Salt damage is an important abiotic stress affecting the agronomic traits of soybean. Soybeans rapidly sense and transmit adverse signals when salt-damaged, inducing a set of response mechanisms to resist salt stress. *AtARA6* encodes a small GTPase, which plays an important role in Arabidopsis vesicle transport and salt tolerance. In this study, we transformed the Arabidopsis gene *AtARA6* into the cultivated soybean Shen Nong 9 (SN9). To investigate the salt tolerance pathways affected by *AtARA6* in soybean, we performed transcriptome sequencing using transgenic soybean and wild-type (SN9) under salt treatment and water treatment. Our results suggest that *AtARA6* is involved in the regulation of soybean SNARE complexes in the vesicle transport pathway, which may directly strengthen salt tolerance. In addition, we comprehensively analyzed the RNA-seq data of transgenic soybean and SN9 under different treatments and obtained 935 DEGs. GO analysis showed that these DEGs were significantly enriched in transcription factor activity, sequence-specific DNA binding, and the inositol catabolic process. Three salt-responsive negative regulator transcription factors, namely *MYC2*, *WRKY6*, and *WRKY86*, were found to be significantly downregulated after salt treatment in transgenic soybeans. Moreover, four genes encoding inositol oxygenase were significantly enriched in the inositol catabolic process pathway, which could improve the salt tolerance of transgenic soybeans by reducing their reactive oxygen species content. These are unique salt tolerance effects produced by transgenic soybeans. Our results provide basic insights into the function of *AtARA6* in soybeans and its role in abiotic stress processes in plants.

Keywords: soybean, *AtARA6*, salt tolerance, RAB GTPase, RAB5, SNARE pathway

INTRODUCTION

Soybean (*Glycine max* [L.] Merr.) is an important oilseed crop that provides high-quality nutrients and is rich in proteins, unsaturated fatty acids, and many bioactive substances, such as isoflavones. Soybean oil is not only used as food but also as a fuel (Qi and Lee, 2014). Due to the diversity of uses, the demand for soybeans is continuously increasing. However, soybean yields are chronically compromised by abiotic stresses, such as drought and salt stress (Phang et al., 2008; Manavalan et al., 2009). Soil salinity affects 20% of the world's arable land and severely affects soybean yield and

quality (Hamwieh and Xu, 2008; Wang et al., 2010). Currently, more than 100 countries worldwide are facing serious soil salinization (Tester and Davenport, 2003; Munns and Tester, 2008). To address this problem, the search for salt-tolerant genes and the breeding of salt-tolerant soybean varieties have become a research priority for breeders.

A variety of techniques have been used to breed salt-tolerant soybean. While traditional breeding methods have limitations due to long cycle times, transgenic techniques can quickly yield soybean germplasm with good agronomic traits and genetic stability, through overexpression or downregulation of target genes. With the advent of advanced molecular biology techniques, salt tolerance genes and salt tolerance mechanisms have been widely studied. Many genes can improve salt tolerance in soybeans by coordinating ion transport. For example, overexpression of the Arabidopsis *AtNHX* (Na^+, K^+/H^+ exchangers) antiporter gene, *AtNHX5*, enhances transgenic soybean salt resistance by increasing the concentration of Na^+ in leaves and roots (Wu et al., 2016).

Drought and salt damage are the most common stressors that have a significant impact on plant growth and development, causing significant losses in crop yield (Beeson et al., 2007; Tester and Langridge, 2010; Agarwal et al., 2013). Plants have multiple ways to deal with abiotic stresses, such as vesicular transport. Substantial evidence suggests that vesicular trafficking ensures the efficient transport of stress-related ions to vacuoles. For example, the involvement of *SchRabGDI1* in endocytic transport and tolerance to salt stress in *Solanum chilense* has been reported (Martin-Davison et al., 2017).

In eukaryotic organisms, membrane transport is a conserved cellular process that moves materials between organelles. This process occurs in three steps: vesicle budding, shuttling between organelles, and fusion. Small extracellular molecules are first screened by clathrin, then the donor cell membrane sags to form vesicles through endocytosis. The vesicles first bud from the donor membrane and then move along the cytoskeleton to the target membrane, where they undergo docking and tethering processes, among others. Finally, the fusion of vesicles and target membranes mediated by SNARE proteins complete material transport. As a house-keeping process, it is tightly regulated. Membrane transport mediated by vesicles/tubules is an important component of substance transport in plants. RAB-GTPase, a small GTPase with a molecular size of 20–25 kDa, is involved in vesicular transport and functions as a molecular switch in the cycling between GDP and GTP (Bucci and Chiariello, 2006; Schwartz et al., 2007). RAB5 belongs to the RAB family. Arabidopsis has three RAB5 homolog genes, *AtARA7* (RABF2b), *AtRHA1* (RABF2a), and *AtARA6* (RABF1) (Ueda et al., 2001). Both *AtARA7* and *AtRHA1* are localized in the perivacuolar compartment (PVC), while *AtARA6* is localized at the plasma membrane, in the Golgi network (TGN), and in multivesicular endosomes (MVEs), suggesting that it may be involved in intracellular endosomal transport (Hoepflinger et al., 2013). In the TGN, *AtARA6* affects SNARE complexes (soluble N-ethylmaleimide-sensitive factor attachment protein receptor), which are crucial for the fusion step, by regulating the R-SNARE/VAMP727 complex on the endosome and the Q-SNARE/SYP121 complex on the plasma membrane (Ebine et al., 2011).

The importance of the SNARE pathway has been investigated under salt stress conditions (Singh et al., 2016). In Arabidopsis, *the AtARA6* knockout line has a moderately salt-sensitive phenotype, while *the AtARA6*-overexpression line has enhanced salt tolerance (Ebine et al., 2011). *SYP121* functions in drought and ABA signaling pathways in tomato (Leyman et al., 1999). Disruption of *SYP121* decreases the water permeability of the cell membrane (Besserer et al., 2012). In Arabidopsis, *the SYP121* mutant, *AtSYP121*, shows suppressed stomatal opening, indicating that the gene may be involved in the drought response (Eisenach et al., 2012). Recently, *VAMP727* was found to physically interact with *BRI1*, a brassinosteroid receptor (Zhang et al., 2019b), thus responding to broad stress stimuli. Another RAB family member, Rab7, has been shown to negatively affect salt tolerance in Arabidopsis (Mazel et al., 2004).

AtARA6 is different from other RAB5 members because of its unique N-terminal myristoylation and palmitoylation, but it has a common activator, *VPS9a* (Grosshans et al., 2006). To date, *AtARA6* homologues have been found in all sequenced plants, except for single-celled green algae, and have not been found in animals (Ebine et al., 2012). *AtARA6* exhibits a variety of biological functions, including endosomal transport, signal transduction, stress response, and growth regulation. In this study, we heterologously expressed the *AtARA6* gene in soybean and found that salt tolerance was enhanced (**Supplementary Figure S1**). We performed RNA-seq to uncover potential pathways affected by the *AtARA6* transgene. Our study provides insights into the function of *AtARA6* and lays a foundation for breeding new germplasms of transgenic salt-tolerant soybean.

MATERIALS AND METHODS

Plant Material and Growing Environment

Col-0 wild-type Arabidopsis was incubated in a normal growth chamber with a light/dark cycle of 16 h/8 h at 22°C/16°C. Transgenic soybean and SN9 were cultured in a greenhouse environment with a light/dark cycle of 16 h/8 h at 28°C/25°C.

Construction of Expression Vector and Soybean Genetic Transformation

Total RNA was extracted from 3 week-old wild-type Arabidopsis leaves and was reverse transcribed into cDNA using the PrimeScript RT kit (Takara, Shiga, Japan). The 609 bp CDS of Arabidopsis *AtARA6* (AT3G54840) was amplified with primers (5′-GAAGAGAAGAAGCACATCCCAT-3′ 5′- ATGGGATGT GCTTCTTCTCTTC-3′) designed using Primer 5.0. The amplification products were recovered using a DNA Gel Extraction Kit (Tiangen Biotechnology Co. Ltd., Beijing, China) and sent for sequencing. The *AtARA6* CDS was cloned into the pTF101 vector *via* Spe1 and Sac1 sites. The plasmid was transformed into *Agrobacterium tumefaciens* strain EHA101.

Cultivated soybean SN9 was infected with the strain by the cotyledon nodulation method. Regenerated plants were obtained after exosome infiltration, clump shoot induction, clump shoot screening, and rooting culture (Paz et al., 2006; Zhang et al., 2019a). The *bar* gene driven by the CaMV35S promoter in the PTF101 vector was used for the positive progeny screen.

Identification of Transgenic Positive Progeny

The transgenic progeny was first tested using LibertyLink test strips, according to the manufacturer's instructions (EnviroLogix Inc., Portland, ME, United States). Genomic DNA from test strip-positive soybean leaves was extracted using a DNA extraction kit (Kangwei Century Biotechnology Co., Ltd., Beijing, China), and primers were designed according to the CDS sequences of *AtARA6* and *Bar* genes. The soybean actin gene was used as a control (**Supplementary Table S1**).

Salt Stress Treatment

The transgenic soybeans and recipient soybean seeds were sowed in vermiculite and watered with 0, 100 mM, or 200 mM NaCl solutions (Wang et al., 2018). Root length was measured after 1 week. For the salt stress treatment at emergence stages, the NaCl concentrations chosen followed the method of Li et al. (2017), with a slight modification: we first sowed the seeds in vermiculite and treated them with water for 10 days until the main leaves fully unfolded, then they were treated with 200 mM NaCl. The salt solution was changed every 3 days. Soybean phenotypes were recorded at weeks 1, 2, and 3 after treatment.

Determination of Salt Tolerance Parameters

We measured representative physiological indicators of salt tolerance in transgenic lines and SN9, including superoxide dismutase (SOD), peroxidase (POD), and catalase (CAT) activities after 3 weeks of salt treatment. SOD activity was determined based on the photochemical reduction capacity of p-Nitro-Blue tetrazolium chloride, and the absorbance value was read at 560 nm (Beauchamp and Fridovich, 1971). POD activity was determined based on the amount of reaction per unit time using guaiacol as a substrate (Reuveni, 1992). During the disappearance of H_2O_2, the activity of CAT is detected by monitoring the change in absorbance at 240 nm (Yang et al., 2008). Malondialdehyde (MDA) reflects the level of peroxidation of plant cell membrane lipids, and we measured MDA contents to assess the degree of damage to transgenic lines and WT plants (Landi, 2017). Chlorophyll was extracted in a mixture of acetone and ethanol, and the absorbance was measured at 645, 652, and 663 nm (Arnon, 1949). The determination of proline was based on the method described by Irigoyen et al. (1992).

RNA Extraction, Library Construction, and RNA Sequencing

The transgenic positive line2 and WT, after salt stress treatment for 5 days, were named S1 and S2, respectively. The transgenic positive line2 and WT, after water treatment for 5 days, were used as controls, and named H1 and H2, respectively. Total leaf RNA was extracted using the OminiPlant RNA Kit (Dnase I) (Kangwei Century Biotechnology Co., Ltd., Beijing, China), according to the manufacturer's instructions. Transcriptome sequencing was performed at Biomarker Biotechnology (Beijing, China). The mRNA was enriched with magnetic beads with Oligo (dT), and cDNA was synthesized using random primers. The purified double-stranded cDNA was end-repaired, A-tail was added, sequencing connectors were ligated, and the cDNA library was obtained by PCR. The RNA-seq library was constructed using Illumina novaseq 6000.

qRT-PCR Validation of Gene Expression

Leaves collected 5 days after salt treatment were used for qPCR validation. Total plant RNA was extracted using a Plant RNA Extraction Kit (Kangwei Century Biotechnology Co., Ltd., Beijing, China), and gDNA Remover Mix was used to remove genomic DNA contamination. An ABScript II RT Mix (ABClonal technology, Wuhan, China) was used for the reverse transcription reaction, according to the manufacturer's instructions. To investigate the expression of *AtARA6* at different treatment times and the expression of genes with which *AtARA6* may have a reciprocal relationship, gene primers and GmActin (NM_001289231) primers were designed using the NCBI online website (**Supplementary Table S2**). The cDNA was amplified using SYBR Green Fast qRT-PCR Mix (ABClonal Technology Co. Ltd., Wuhan, China) using the chimeric fluorescence method. Each treatment had three replicates. Two internal reference primers of GmActin were used to ensure the accuracy of qPCR. qPCR experiments were performed using the LightCycler 480 II machine. Relative expression of GmActin was calculated using the $2^{-\Delta\Delta CT}$ method.

Data Quality Control and Read Mapping

Raw data generated by sequencing were processed using FastQC and Trim_galore software (the quality control threshold was 25, reads with a length lower than 50 were discarded, the allowable error rate was 0.1, and the minimum number of bases overlapping with the adapter sequence was 3) to obtain high-quality data for mapping (reads with a ratio of N greater than 10% and reads (Q ≤ 10%) whose bases accounted for more than 50% of the entire reads were removed). The Williams 82 reference genome was constructed using HISAT2 software (version 2.1.0) (Parameter; hisat2-build -p 4 W82_ref. fa -snp W82_ref.snp -ss W82_ref. ss -exon W82_ref. exon W82_hisat2_index) (Kim et al., 2015), and unique reads were compared. Reads on pairs were assembled and quantified by DESeq2 (the test method was "Wald significance tests" and the fit type was "parametric") using the StringTie comparison (the TPM value for the minimum allowed input transcript was 0.1 and the minimum isoform fraction was 0.1) (Love et al., 2014; Pertea

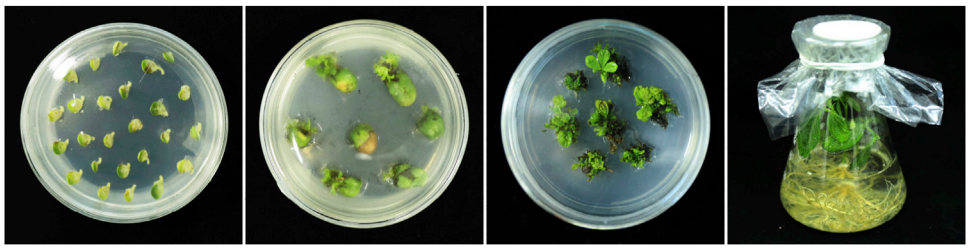


FIGURE 1 | Genetic transformation of *AtARA6*. Agrobacterium tumefaciens EHA101-mediated genetic transformation of soybean, including exosome infiltration, clump shoot induction, clump shoot screening, and rooting culture.

et al., 2015). StringTie used FPKM (transcript per million fragments mapped) as a measure of transcript or gene expression level. Upregulated and downregulated genes between transgenic soybean and SN9 under salt treatments were determined using the edgeR package in R (Robinson et al., 2010), and all analyzed False Discovery Rate (FDR) thresholds were less than 0.05 (Fold Change ≥ 2). All raw data have been uploaded to the SRA database in NCBI (ID: PRJNA789350).

Gene Ontology and KEGG Pathway Enrichment

Differentially expressed genes (DEGs) generated between each treatment were analyzed for gene ontology (GO) gene function clustering and enrichment via online websites (http://www.geneontology.org and http://bioinfo.cau.edu.cn/agriGO) with an FDR threshold set at 0.05 (Huang da et al., 2009; Du et al., 2010). Gene ontology was based on the National Center for Biotechnology Information (NCBI) non-redundant protein sequences (Nr), Swiss-Prot, protein family, and clusters of orthologous groups of proteins (KOG/COG) database. KEGG [Kyoto Encyclopedia of Genes and Genomes (KEGG) and Ortholog database (KO)] pathway enrichment analysis of DEGs was performed using the (http://www.genome.jp/kegg) website (Kanehisa et al., 2021), and we screened for pathways with p values less than 0.05.

All data were analyzed using two-way Analysis of Variance to evaluate significant differences at $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. All statistical analyses were carried out IBM SPSS Statistics 22 (IBM Corp., Armonk, NY, United States). All DEGs and pathways used for GO analysis and KEGG analysis can be found in **Supplementary Table S3** and **Supplementary File S1**.

RESULTS

Soybean Transformation and Positive Line Identification

Cultivated soybean, strain SN9, was transformed by agrobacterium harboring *AtARA6,* and a total of six transformation lines, named Line1, Line2, Line3, Line4, Line5, and Line6, were obtained in the T0 generation (**Figure 1**). PCR was used to detect the presence of *AtARA6* and *Bar* genes separately. Five positive transgenic lines were identified, consistent with the results of the LibertyLink test strips (**Figures 2A–C**). The positive lines were propagated to T3 to ensure stable transgene inheritance. qRT-PCR results demonstrated that *AtARA6* showed the highest expression level at 5 days after salt treatment (**Figures 2D,E**).

AtARA6 Gene Improves Soybean Salt Tolerance

We assessed the salt tolerance of transgenic soybean and WT plants in the germination and emergence stages separately. Three positive T3 generation transgenic lines and WT were sown in vermiculite treated with different concentrations of salt solution. The length of the roots was compared after 7 days of treatment. As shown in **Figures 3A,B**, there was no significant difference in root length between transgenic soybean and WT plants in the water treatment group. However, transgenic soybean roots were longer than WT in both 100 mM and 200 mM NaCl solutions (**Supplementary Table S4**), suggesting that transgenic soybean has stronger germination ability and adaptability under salt stress.

We then treated transgenic soybean with unfolded opposite true leaves in 200 mM NaCl solution WT leaves started to turn yellow and dry at 1 week, and the true leaves wilted and fell off. There were no significant changes in the transgenic lines Line1 and Line3, although the true leaves of Line2 turned slightly yellow (**Figure 4A**). After 2 weeks of salt treatment, most of the lower leaves of WT plants lost their green color, appeared dehydrated, the petioles withered and fell off, the upper leaves turned yellow and appeared dehydrated, and only a few of the apical ternate leaves were unaffected. In contrast, only 1-2 true leaves of transgenic soybean Line1 and Line2 turned yellow, and the upper part of the plants stayed upright and grew normally (**Figure 4B**). After 3 weeks, the entire WT plant was withered, the growth point was necrotic, and the leaves were severely shed. The true leaves of Line1 and Line2 transgenic soybean did not fall off, and Line3 only had true leaves that turned yellow. The degree of salt damage in transgenic soybean was thus significantly lower than that in WT (**Figure 4C**). These results showed that the transformation using the *AtARA6* gene significantly improved the salt tolerance pathway of SN9 and enhanced soybean resistance to salt stress.

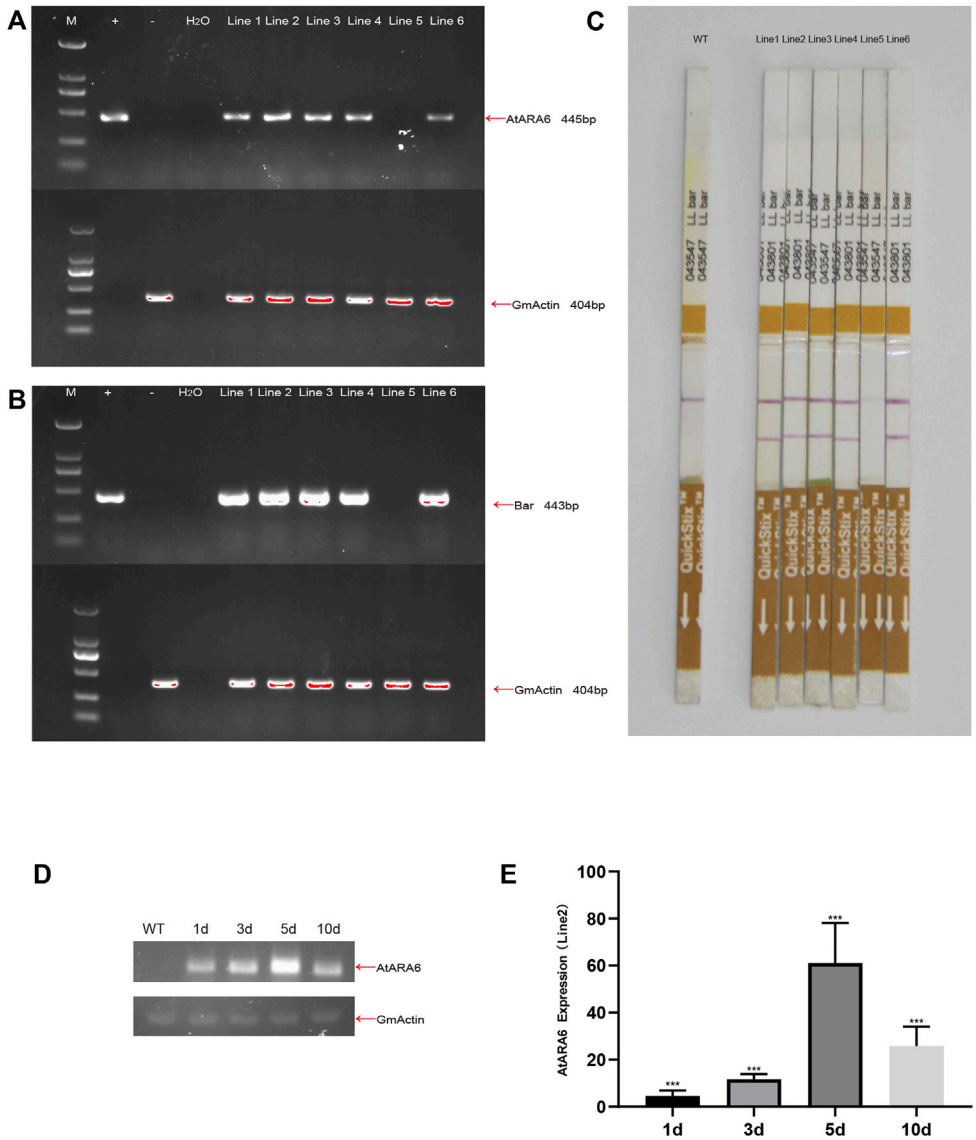


FIGURE 2 | Confirmation of *AtARA6* transgenesis. **(A,B)** Positive transgenic soybean progeny Lines1-6: 404 bp of GmActin was used to detect plant genomic DNA, *AtARA6* 445 bp, *Bar* 443=bp (M, Trans2K Plus II DNA marker; +, *AtARA6* expression vector plasmid DNA; –, WT genomic DNA). **(C)** LibertyLink test strips for genetically transformed soybeans. **(D)** RT-PCR analysis of *AtARA6* gene expression at different times under 200 mM NaCl treatment. **(E)** qRT-PCR analysis of the relative expression of *AtARA6* in different durations under salt treatment.

Physiological Evaluation of Salt Tolerance

To assess the salt tolerance level of transgenic soybean, we measured the SOD, POD, and CAT activities of transgenic soybean and WT plants under salt stress. All these activities in both transgenic soybean and WT increased in the 200 mM NaCl treatment group. However, the increase in these activities in transgenic soybean was significantly greater than that in WT, suggesting that transgenic soybean has a stronger tolerance capacity under salt stress (**Figures 5A–C**). Similarly, the increased proline content in the WT was much lower than that in the transgenic line (**Figure 5D**). MDA content reflects the degree of damage to plant membrane lipids. Both transgenic soybean and WT MDA contents increased under salt stress, but the transgenic line increased less than SN9 (**Figure 5E**). The chlorophyll content of the soybeans decreased under salt treatment, but that of the transgenic line was over 2-fold higher than that of the WT plants (**Figure 5F**). Under stress, the plant's defense system responds by generating protective enzymes such as SOD, POD, and CAT, which can reduce ROS levels in cells. Proline, an osmoregulatory substance, can regulate the osmotic potential of plant cells. Protective enzyme activity and proline content increased in both transgenic soybean and WT under salt stress. However, the elevation was much higher in transgenic soybean than in WT. In addition, the transgenic soybean plasma membrane was less damaged than the WT, and chlorophyll levels were higher than in WT. Such

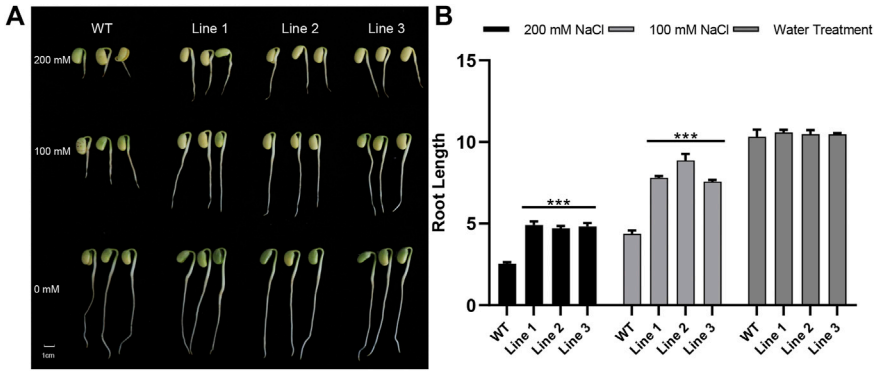


FIGURE 3 | Salt treatment phenotypes of transgenic soybean and WT during germination. **(A)** Phenotypes of transgenic soybean Lines1-3 and WT treated with 200 mM NaCl, 100 mM NaCl, or water for 7 days at the germination stage. **(B)** Root length phenotypes of WT and transgenic soybean under salt stress during germination ($p < 0.05$).



FIGURE 4 | Salt treatment phenotypes of transgenic soybean and WT at emergence stages. **(A–C)** Phenotypes of transgenic soybean Lines1-3 and WT (left) treated with 200 mM NaCl at emergence stages for 1, 2, and 3 weeks.

differences suggest that *AtARA6* can aid soybean resistance to salt stress and reduce the degree of damage to the organism under such conditions.

RNA-Seq Analysis of Transgenic Lines and WT

To investigate the potential pathways by which *AtARA6* affects salt tolerance, we performed RNA-seq analysis using Line2 and WT in the true leaf stage under salt stress for 5 days, named S1 and S2, respectively. The same plant materials were mock-treated with water, named H1 and H2. The clean data obtained from each of these samples was over 5.84 Gb, and the percentage of Q30 bases was above 92.13%. The clean data of each sample were compared with the Williams 82 reference genome separately, and the comparison efficiency ranged from 89.56% to 94.08% (**Table 1**). Between H1 and S1, we detected 3,916 DEGs, including 2,399 downregulated DEGs and 1,517 upregulated DEGs in transgenic soybean under salt treatment. In the H2 versus S2 group, we detected 5,027 DEGs, including 1,729 downregulated DEGs and 3,298 upregulated DEGs in SN9 under salt treatment (**Figure 6**). Furthermore, we compared the DEGs in the H2-S2 and H1-S1 groups, which shared only 246 upregulated DEGs (~5.4%) and 310 downregulated DEGs (~8.1%) (**Figure 6**). These results indicate that overexpression of the *AtARA6* gene in soybean resulted in significant alterations in gene transcription in SN9, which may be the reason for improved salt tolerance.

AtARA6 and the Soybean SNARE Regulatory Pathway

In the vesicular transport pathway, SNARE complexes are responsible for the transport of substances from endosomes to the plasma membrane. *AtARA6* regulates the formation of SNARE complexes containing *VAMP727* and *SYP121* (Ebine et al., 2011). In S1 and S2 DEGs, we found two homologs of *SYP121* in soybean (Glyma.02G069700 and Glyma.16G151200). The KEGG results showed that they were significantly enriched in the SNARE pathway (**Supplementary Table S5**). *VAMP727* encodes a vesicle-associated membrane protein 727 on the target membrane and plays a role in the fusion of vesicles with the target membrane. We found two homologous genes of *VAMP727* in soybean (Glyma.01G179300 and

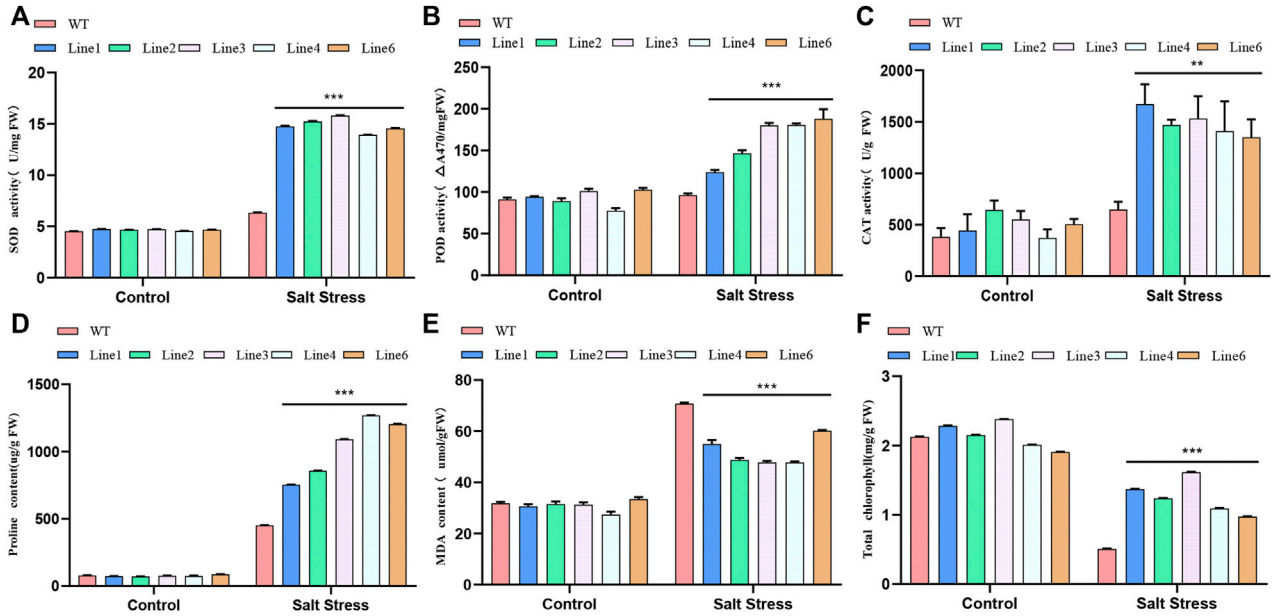


FIGURE 5 | Determination of physiological parameters of transgenic soybean and WT treated with salt stress. **(A)** Superoxide dismutase (SOD) activity. **(B)** Peroxidase (POD) activity. **(C)** Catalase (CAT) activity. **(D)** Proline content in leaves (μg/g). **(E)** Malondialdehyde content in leaves (μmol/g). **(F)** Chlorophyll content of leaves after salt treatment (mg/g). All values were measured after 21 days of salt treatment. Statistical analysis was performed using one-way ANOVA, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $p > 0.05$ indicates the difference between transgenic line and WT was not significant.

TABLE 1 | Summary statistics of the RNA quality and sequencing results.

Clean data	59.68GB		
Q30	>92.13%		
Rades mapped rate	89.56%–94.08%		
All database annotation	57,265		
Clean data	59.68GB		
	Total reads on average	Mapped reads on average	Unmapped reads on average
H1	49,314,341	44,299,190 (89.81%)	5,015,151 (10.19%)
H2	48,795,122	45,559,462 (93.37%)	3,235,660 (6.64%)
S1	47,200,475	44,389,190 (94.05%)	2,811,286 (5.95%)
S2	54,559,953	51,243,100 (93.93%)	3,316,853 (6.08%)

Glyma.11G062900), which were significantly upregulated in transgenic soybean after salt treatment. KEGG enrichment results showed that these two genes are in the endocytosis pathway (**Supplementary Table S5**). We verified 4 upregulated genes using qRT-PCR (**Figure 7**). Our results show that *AtARA6* upregulates the SNARE genes *SYP121* and *VAMP727*.

Gene Ontology Enrichment Analysis of Differentially Expressed Genes

To investigate the functions of DEGs, we performed GO analysis of DEGs in transgenic soybean and WT plants under salt and water treatments. We focused on 1,271 upregulated DEGs and 2,089 downregulated genes in the transgenic line under salt treatment (**Figure 6**). The expression of some of these genes may be perturbed by the *AtARA6* transgene. For 1,271 upregulated DEGs, significant GO terms in RNA-related processes were enriched, including RNA metabolic processes (GO:0016070, p = 1.10E-05), RNA biosynthetic processes (GO: 0032774, p = 4.00E-05), regulation of transcription (GO:0006355, p = 1.10E-05), and regulation of gene expression (GO:0010468, p = 1.60E-05) (**Table 2**). We hypothesize that such upregulated RNA-related pathways may help to explain the stronger salt resistance of the transgenic line. It is well-known that plants use a wide array of mechanisms, including transcriptional regulation and posttranscriptional mechanisms, to survive under diverse stress conditions. At the post-transcriptional level, many studies have demonstrated that Ca^{2+}-dependent transcriptional reprogramming is important for the stress response (Galon et al., 2008; Yuan et al., 2018a; Yuan et al., 2018b). In addition, RNA modification has been shown to be related to stress responses, such as splicing, capping, polyadenylation, and translocation. Overexpression of the



FIGURE 6 | Comparison of DEGs between transgenic positive lines and SN9 under salt treatment and water treatment.

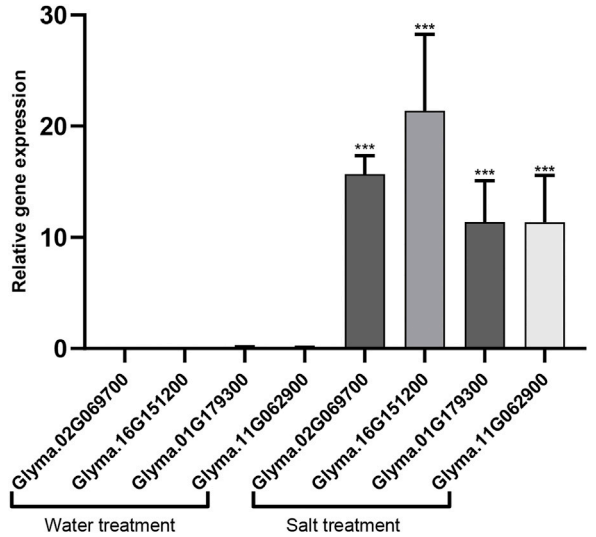


FIGURE 7 | qRT_PCR analysis of homologous genes of *SYP121* and *VAMP727*. The homologous genes of *SYP121*, Glyma.02G069700 and Glyma.16G151200. The homologous genes of *VAMP727*, 01G179300 and Glyma.11G062900. They were significantly up-regulated under in transgenic soybean after salt treatment. Three biological replicates were used. Data was calculated using the $2^{-\Delta\Delta CT}$ method, Statistical analysis was performed using ANOVA ($p < 0.05$), $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Arabidopsis RNA binding protein *AtRGGA* enhances salt tolerance (Ambrosone et al., 2015). Similarly, overexpression of *Beta vulgaris* RNA binding protein, salt tolerant 3 (*BvSATO3*), in Arabidopsis increased plant salt resistance (Rosa Téllez et al., 2020).

For 2,089 downregulated DEGs in the transgenic line under salt treatment, significant GO terms in metabolism-related processes, (GO:0008152, p = 1.50E-27), primary metabolic processes (GO:0044238, p = 0.0037), cellular metabolism (GO: 0044237, p = 2.50E-08), macromolecule metabolism (GO: 0043170, p = 0.019), protein metabolism (GO:0019538, p = 0.0019), carbohydrate metabolism (GO:0044262, p = 0.00077), and glucose metabolism (GO:0006006, p = 1.60E-05) (**Table 3**). We believe that the reduction of the expression of certain metabolism-related genes may indirectly reduce the pressure to maintain cellular homeostasis, which is perturbed by salt stress. A similar result was found for gene expression changes in the sugar beet under salt stress (Rosa Téllez et al., 2020).

In addition, we analyzed the direct mechanism via which *AtARA6* improves salt tolerance in soybean. Between the S1 and S2 groups, we identified 6,572 DEGs that represent unique processes carried out by transgenic soybean under salt treatment. Among the GO terms, Rho guanyl-nucleotide exchange factor activity (GO:0005089, p = 0.0045), Ras guanyl-nucleotide exchange factor activity (GO:0005088, p = 0.0045), GTPase regulator activity (GO:0030695, p = 0.0093), and GTPase activity (GO:0003924, p = 0.044) pathways were significantly enriched in the molecular function module in upregulated DEGs in transgenic soybean under salt treatment (**Supplementary Figure S2**). Rho-like GTPases are plant-specific molecular switches that play a critical role in plant survival under abiotic stresses (Miao et al., 2018). The expression levels of DEGs enriched in these pathways under different treatments were found to be higher in transgenic soybean than in WT (**Figure 8**). Therefore, we suggest that the upregulation of these four pathways by the *AtARA6* gene acted as a direct factor to improve the salt tolerance of transgenic soybean.

After the GO analysis of DEGs between different treatments, we concluded that the transgenic soybean showed significant enrichment in terms of transcription level and substance metabolism. In order to explore the changes caused by the transformation of *AtARA6* in soybean, we combined S1, S2 and H1, H2 for analysis. We first found 3,815 DEGs between H1 vs. S1 and 4,933 DEGs between H2 vs. S2. Among these DEGs, 1,344 DEGs were shared and could be considered as changes in soybean caused by salt stress. At the same time, we identified 6,456 DEGs between S1 vs. S2 as a result of transfection of the *AtARA6* gene. We found 935 DEGs between 1,344 and 6,456 DEGs, which represented the unique effects of *AtARA6* gene transformation under salt stress (**Supplementary Figure S3**). GO analysis of 935 DEGs showed that they had the highest levels of transcription factor activity, sequence-specific DNA binding (GO:0003700, p = 5.2e-11, MF), and inositol catabolic process (GO:0019310, p = 4.0e-10, BP), and the difference was significant (**Figure 9**). Furthermore, we identified multiple salt stress-

TABLE 2 | GO enrichment of 1,271 upregulated DEGs.

GO term	Ontology	Description	Genes	Total genes	*p*-value	FDR
GO:0016070	BP	RNA metabolic process	84	2,485	0.00021	0.0054
GO:0032774	BP	RNA biosynthetic process	80	2015	1.30E-06	4.00E-05
GO:0010468	BP	Regulation of gene expression	77	1850	3.80E-07	1.60E-05
GO:0006355	BP	Regulation of transcription, DNA-dependent	77	1796	1.20E-07	1.10E-05

TABLE 3 | GO enrichment of 2089 downregulated DEGs.

GO term	Ontology	Description	Genes	Total genes	*p*-value	FDR
GO:0006066	BP	Alcohol metabolic process	17	24	2.10E-12	3.40E-10
GO:0043170	BP	Macromolecule metabolic process	427	7,301	0.00075	0.019
GO:0044238	BP	Primary metabolic process	529	9,070	0.00013	0.0037
GO:0019538	BP	Protein metabolic process	281	4,388	4.30E-05	0.0019
GO:0044262	BP	Cellular carbohydrate metabolic process	30	235	1.60E-05	0.00077
GO:0044237	BP	Cellular metabolic process	521	8,043	2.20E-10	2.50E-08
GO:0008152	BP	Metabolic process	842	12,346	2.70E-30	1.50E-27

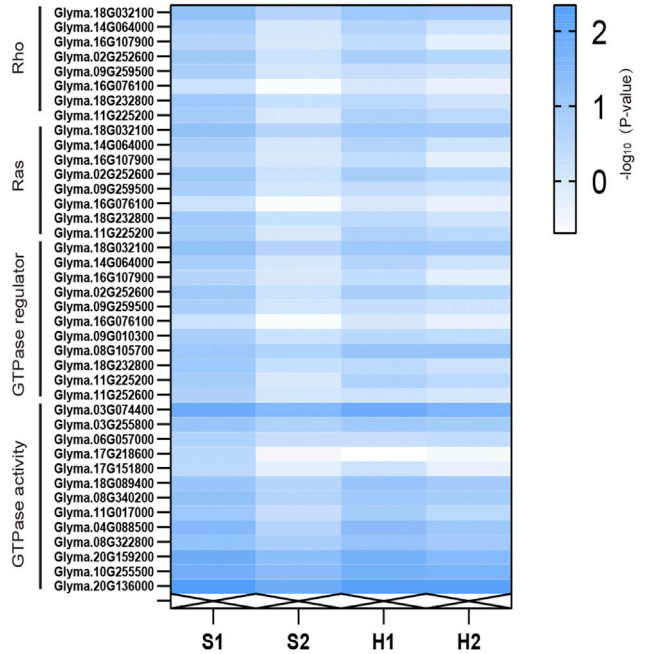


FIGURE 8 | RNA-seq reveals the expression levels of DEGs enriched in the Rho/Ras/GTPase regulator/GTPase activity pathways under different treatments.

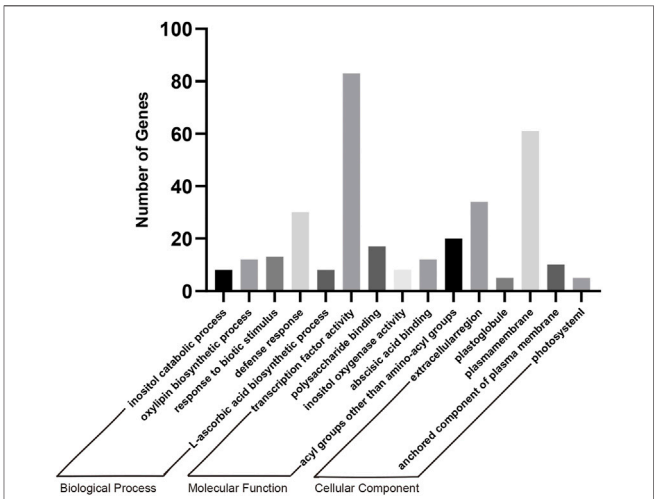


FIGURE 9 | GO enrichment analysis of 935 DEGs including BP, CC and MF. *p*-values < 0.01 for all GOs. Transcription factor activity, sequence-specific DNA binding (GO:0003700, p = 5.2e-11, MF) and inositol catabolic process (GO:0019310, p = 4.0e-10, BP) are the most significant.

responsive transcription factors. *MYC2* is a negative regulator of proline synthesis and improves salt tolerance by regulating the proline content in Arabidopsis (Verma et al., 2020). In addition, *WRKY6* and *WRKY86* have been reported to negatively regulate plant salt tolerance (Li et al., 2019; Fang et al., 2021). In our RNA-seq results, the expression levels of *MYC2*, *WRKY6*, and *WRKY86* were all downregulated (**Supplementary Table S5**), indicating that they are involved in the salt tolerance pathway of transgenic soybean. Moreover, the phospholipase signaling pathway is an important pathway for plant salt tolerance. The N-terminus of phospholipase D contains a domain that binds to phosphoinositide. At the same time, *PLDα* can bind to the heterotrimeric protein subunit to regulate GTPase activity and control stomatal movement and loss of water (Zhao and Wang, 2004). GO results showed that transgenic soybeans were significantly enriched in the inositol catabolic process and inositol oxygenase activity (GO:0050113, p = 2.7e-9) pathways. Four DEGs were found to be enriched in this pathway (Glyma.08G199300, Glyma.07G126600, Glyma.05G224500, and Glyma.07G013900), all of which functioned to encode inositol oxygenase. In rice, inositol oxygenase has been

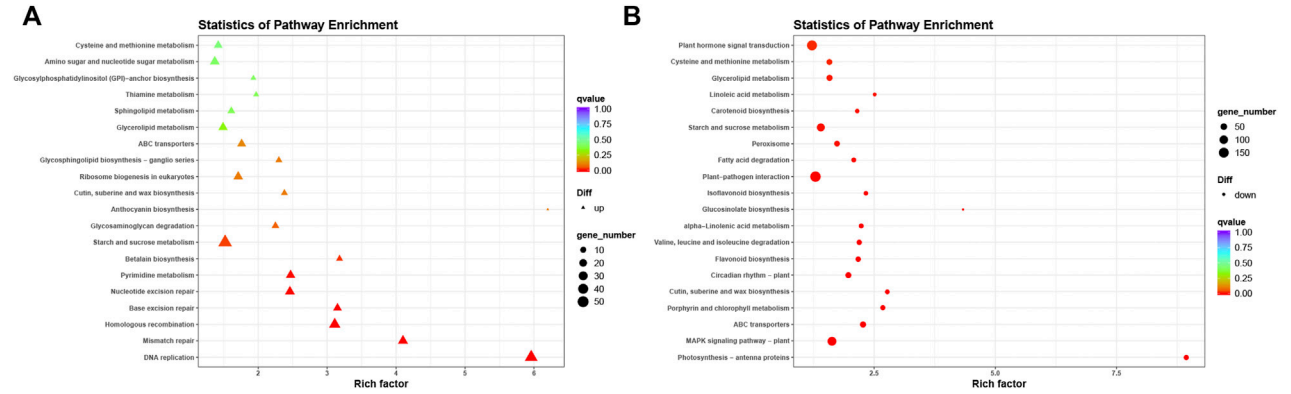


FIGURE 10 | Kyoto Encyclopedia of Genes and Genomes analysis of up-regulation and down-regulation of DEGs. **(A,B)** KEGG analysis on up-regulated (left) and down-regulated (right) DEGs under salt stress in transgenic soybean compared to WT.

reported to scavenge reactive oxygen species (Duan et al., 2012). In our study, these four genes were strongly upregulated after transgenic soybean salt treatment, indicating that transgenic soybeans could regulate the inositol catabolic process to improve salt tolerance.

Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analysis of Differentially Expressed Genes

To further investigate the mechanisms for the increased salt tolerance of transgenic soybean, we performed KEGG analysis on up and downregulated DEGs under salt stress in transgenic soybean compared to WT (**Figure 10A**). Among the upregulated DEGs, there was enrichment in DNA replication (ko03030, p = 6.04e-80), mismatch repair (ko03430, p = 7.22e-25), pyrimidine metabolism (ko00240, p = 5.26e-43), homologous recombination (ko03440, p = 8.41e-32), base excision repair (ko03410, p = 1.17e-26), nucleotide excision repair (ko03420, p = 7.24e-32), and starch and sucrose metabolism (ko00500, p = 1.68e-98) (**Supplementary Table S7**). Previous studies have shown that salt stress disrupts genome stability in plants (Roy et al., 2013). Exposure to high salinity increases the risk of DNA double-strand breaks (Roy et al., 2013). Our transgenic soybean showed upregulated genes related to DNA replication and mismatch repair pathways, possibly partially underlying its stronger ability to cope with salt stress.

In the downregulated DEGs, significantly enriched pathways included photosynthesis-antenna proteins (ko00196, p = 4.53e-51), porphyrin and chlorophyll metabolism (ko00860, p = 2.64e-41), cutin suberine and wax biosynthesis (ko00073, p = 1.10e-48), flavonoid biosynthesis (ko00941, p = 2.80e-43), cysteine and methionine metabolism (ko00270, p = 8.57e-88), and linoleic acid metabolism (ko00591, p = 1.84e-37; **Figure 10B**) (**Supplementary Table S8**). Many downregulated pathways were concentrated in biosynthesis and metabolism processes, which were in line with our hypothesis based on GO analysis that reduction of certain metabolism-related gene expression may indirectly reduce the pressure to maintain cellular homeostasis, which is perturbed by salt stress. The downregulated pathways in our study were consistent with those of previous studies. For example, cysteine and methionine levels decreased significantly in broccoli after salt treatment (Lopez-Berenguer et al., 2008). In peanuts, linoleic acid synthesis genes were also downregulated after salt treatment (Sui et al., 2018).

DISCUSSION

Salt stress is an important factor that affects soybean yield and growth. Transgene breeding has proven to be a promising method for obtaining salt-tolerant cultivars. For example, soybean *GmNHX1* can accelerate the transport of Na^+ from the cytoplasm to the vesicles. Salt tolerance was improved in both soybean and Arabidopsis by overexpression of *GmNHX1* (Sun et al., 2019). In rice, overexpression of *OsRab7* increases endocytosis and leads to increased salt tolerance (Peng et al., 2014).

In this study, an Arabidopsis key vesicle trafficking gene, *AtARA6*, was transformed into soybean and stable expression lines were confirmed by PCR and qRT-PCR. Some stress-related indicators, such as protective enzymes, proline, and MDA, were found to be upregulated, perhaps underlying the ability of transgenic soybean to withstand adversity. Our results showed that the transgenic soybean grew normally at the germination and emergence stages under salt treatment, compared to WT, which showed growth stagnation and eventual death.

AtARA6 belongs to the RAB5 family. Its N-terminal myristoylation and palmitoylation sites distinguish it from other RAB5 members. The relationship between these sites and salt tolerance has been demonstrated in both *AtARA6* and other genes. For example, overexpression *of AtARA6* in Arabidopsis under salt treatment yields longer roots compared to WT, indicating stronger salt resistance. However, overexpression of *AtARA6* with a 29 amino acid truncation at the N-terminus shows a similar root length to WT, indicating that the N-terminus is important for salt resistance (Yin et al., 2017). Similarly, the Arabidopsis *AtSOS3* gene plays a role in the salt

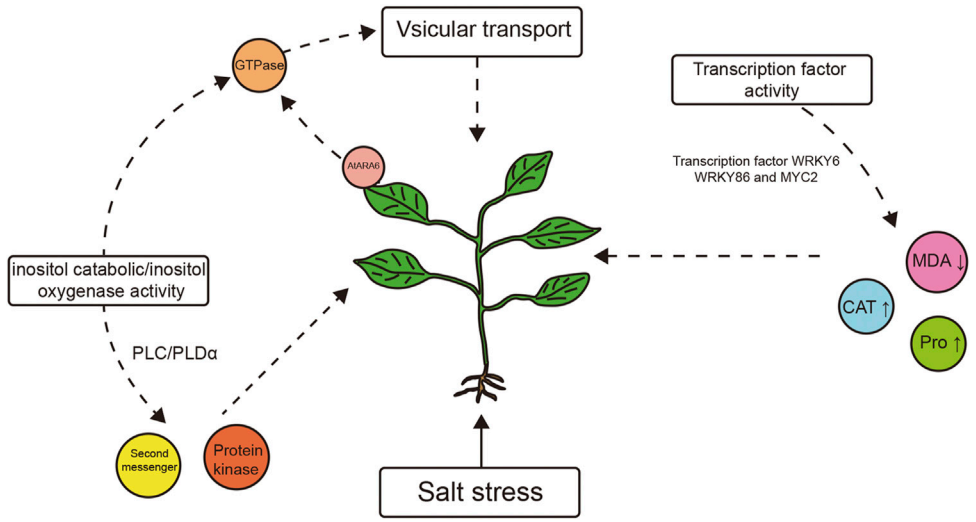


FIGURE 11 | Salt tolerance pathway of *AtARA6* transgenic soybean.

response. The rescue of the salt-hypersensitive phenotype of *AtSOS3-1* knockout plants required the full-length *AtSOS3* gene, but the *AtSOS3* gene with an N-terminal mutant did not rescue the phenotype (Ishitani et al., 2000).

In Arabidopsis, SNARE is a pathway in which *AtARA6* is directly involved, and *AtARA6* has been shown to enhance salt tolerance in Arabidopsis through vesicular transport. In this study, we focused on *SYP121* and *VAMP727* genes. qRT-PCR results showed that their expression was upregulated under salt treatment, more in transgenic plants than in WT plants, suggesting that the transformation of the *AtARA6* gene accelerated vesicle movement in soybean and that the improvement in salt tolerance in soybean is due to this, as it is in Arabidopsis. Many functions of *AtARA6* in Arabidopsis have been reported: for example, the *AtARA6* overexpression line is more tolerant to dark-induced senescence (DIS) (Yin et al., 2017). Recent studies have shown that *AtARA6* can affect the expression of the starch gene *QQS* in Arabidopsis, thus participating in the dynamic balance between starch and soluble sugars in Arabidopsis. A comparison of wild-type and *AtARA6* mutants revealed that *AtARA6* inhibited the proliferation of pathogenic bacteria (Tsutsui et al., 2015). We explored the function of *AtARA6* in transgenic soybeans. Here, we performed RNA-seq on transgenic soybean and SN9 under different treatments using leaves but not roots. Although leaves are not the first organ to respond to salt stress, considering the function of *AtARA6*, it has been reported that Qc-SNAREs transfer excess Na^+ to vacuoles, thus increasing salt tolerance in Arabidopsis (Liu et al., 2007). As the vacuoles in the primary roots were immature, leaves became the first choice. In addition, plants can also regulate body water through the opening and closing of stomata. *SYP121* and *VAMP71* have been reported to be related to stomata closure (Grefen et al., 2015; Salinas-Cornejo et al., 2019). Leaves contain a larger number of stomata compared to roots; hence, they may be an important factor affecting the salt tolerance of soybeans. We identified significant enrichment of RNA-related processes and metabolic processes such as protein metabolism (GO:0019538), suggesting that transgenic plants can adapt to high salt stress through transcriptional regulation and metabolic regulation. At the same time, we comprehensively analyzed the DEGs of transgenic soybean and SN9 under different treatments and found that inositol catabolic process and transcription factor activity and sequence-specific DNA binding were significantly enriched in transgenic soybean under stress conditions, indicating that the two pathways are used for salt tolerance improvement. We have summarized the salt tolerance pathway of the *AtARA6* transgenic soybean in a map (**Figure 11**). Reactive oxygen species scavenging mechanisms are one of the important pathways for plants to cope with abiotic stress. In our study, four genes encoding inositol oxygenases were found to be significantly upregulated in transgenic soybeans after salt treatment, which was consistent with our physiological indicator measurements. In addition, we identified multiple transcription factors that respond to salt stress, which can serve as candidate genes for studying salt tolerance in transgenic soybeans. Under drought and salt stress, *WRKY6* negatively regulates the ability of Arabidopsis and cotton to tolerate abiotic stress; *WRKY86* can bind to maize W-box, thereby increasing maize catalase activity, reducing malondialdehyde content, and improving salt tolerance. Moreover, the transcription factor *MYC2* of the bHLH family enhances salt tolerance in Arabidopsis by regulating proline content. They provided the basis for the changes in physiological indexes of transgenic soybean under salt stress. These findings suggest that transgenic soybeans have salt tolerance pathways other than the SNARE pathway, in contrast to Arabidopsis.

Meanwhile, terms like Rho guanyl-nucleotide exchange factor activity, Ras guanyl-nucleotide exchange factor activity, GTPase regulator activity, and GTPase activity were found to be significantly enriched. To demonstrate that the enrichment of

these pathways is not caused by *AtARA6* homologous genes in soybean, we investigated the expression of five *AtARA6* homologous genes. It was found that the expression levels of these five genes were not significantly different between SN9 and transgenic soybean under salt treatments, indicating that the upregulation of these pathways is caused by the *AtARA6* gene itself.

Sometimes, plants need to cope with diverse stress threats simultaneously. On one hand, combined stress conditions require a trade-off between stress adaptation and growth. Starch and sucrose metabolism was one of the most significant upregulated terms in our KEGG analysis. Starch has been proved to be a key player in response to water deficit, high salinity, and extreme temperatures. Under these adverse environmental conditions, chloroplasts and photosynthesis may be compromised, as shown in our study. Starch metabolism and remobilization provide extra energy and carbon for the plant to survive. The correlation between starch metabolism and diverse stress resistance has been widely proved. For example, in the model plant, Arabidopsis, salt stress induces rapid carbohydrate metabolism and increases soluble sugars (Kempa et al., 2008). Similarly, in sugar beet roots, upregulated starch and sucrose metabolism was found by RNA-seq of the plants with higher salt resistance levels (Liu et al., 2020).

Previous studies have shown that salt stress affects genome stability and induces DNA double-strand breaks in plants (Roy et al., 2013). The notable enriched upregulated KEGG terms in our transgenic soybean, mismatch repair and DNA replication, may be associated with salt tolerance by maintaining genome stability. Although the connection is not fully understood, some evidence has linked genome stability and stress resistance. For example, the eukaryotic pre-replicative complex (PreRC), including heterohexameric minichromosome maintenance proteins (MCM), is responsible for making sure the DNA is replicated only once per cell division cycle. Overexpression of a MCM family member, *MCM6*, in tobacco confers higher salinity tolerance (Tuteja et al., 2011). Moreover, topoisomerase is the specific enzyme that can remove or add DNA supercoils and untangle snarled DNA. Enhanced salt tolerance was found in transgenic tobacco overexpressing *NtTopoIIα-1*, a topoisomerase gene (John et al., 2016).

On the other hand, prioritization of different stress adaptations is also critical. In Arabidopsis, salt treatment induces decreased resistance to pathogens such as *Pseudomonas syringae*, *Alternaria brassicicola*, and *Botrytis cinerea* (Haller et al., 2020). Similarly, Matthias et al. found that ABA, a stress plant hormone, can repress immunity against a bacterial strain, *Pto hrcC*–, in old leaves (Berens et al., 2019). In cotton, overexpression of the *GhMKK1* gene enhances salt tolerance, as well as the susceptibility to the pathogen Ralstonia solanacearum (Lu et al., 2013). In our study, we found overexpression of an Arabidopsis gene, AtARA6, enhances soybean salt resistance. Through KEGG analysis, we found that downregulated genes were significantly enriched in plant-pathogen pathways. Our data suggest that heterologous expression of the AtARA6 gene in soybean may prioritize the salt tolerance response over the pathogen response.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA789350.

AUTHOR CONTRIBUTIONS

FM and ZH designed the experiment and wrote the manuscript; ZH, YL and YZ carried out experiments; MY, YT, LJ and DK carried out the salt treatment experiment; TL and SL carried out the germination experiment; FM and ZH carried out the statistical analysis.

FUNDING

This study was financially supported by the Basic Research Funds of JAAS (KYJF2021JQ001), the Chinese National Natural Science Foundation (32172032), China National Novel Transgenic Organisms Breeding Project (2016ZX08004-004-006) and the national project (CARS-04-PS04), the Major Project of New Varieties of Genetically Modified Organism of China (2016ZX08004-002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.849357/full#supplementary-material

REFERENCES

Agarwal, P. K., Shukla, P. S., Gupta, K., and Jha, B. (2013). Bioengineering for Salinity Tolerance in Plants: State of the Art. *Mol. Biotechnol.* 54 (1), 102–123. doi:10.1007/s12033-012-9538-3

Ambrosone, A., Batelli, G., Nurcato, R., Aurilia, V., Punzo, P., Bangarusamy, D. K., et al. (2015). The Arabidopsis RNA-Binding Protein AtRGGA Regulates Tolerance to Salt and Drought Stress. *Plant Physiol.* 168 (1), 292–306. doi:10.1104/pp.114.255802

Arnon, D. I. (1949). Copper Enzymes in Isolated Chloroplasts. Polyphenoloxidase in Beta Vulgaris. *Plant Physiol.* 24 (1), 1–15. doi:10.1104/pp.24.1.1

Beauchamp, C., and Fridovich, I. (1971). Superoxide Dismutase: Improved Assays and an Assay Applicable to Acrylamide Gels. *Anal. Biochem.* 44 (1), 276–287. doi:10.1016/0003-2697(71)90370-8

Beeson, T. D., Mastracchio, A., Hong, J.-B., Ashton, K., and MacMillan, D. W. C. (2007). Enantioselective Organocatalysis Using SOMO Activation. *Science* 316 (5824), 582–585. doi:10.1126/science.1142696

Berens, M. L., Wolinska, K. W., Spaepen, S., Ziegler, J., Nobori, T., Nair, A., et al. (2019). Balancing Trade-Offs between Biotic and Abiotic Stress Responses

through Leaf Age-dependent Variation in Stress Hormone Cross-Talk. *Proc. Natl. Acad. Sci. U S A.* 116 (6), 2364–2373. doi:10.1073/pnas.1817233116

Besserer, A., Burnotte, E., Bienert, G. P., Chevalier, A. S., Errachid, A., Grefen, C., et al. (2012). Selective Regulation of maize Plasma Membrane Aquaporin Trafficking and Activity by the SNARE SYP121. *Plant Cell* 24 (8), 3463–3481. doi:10.1105/tpc.112.101758

Bucci, C., and Chiariello, M. (2006). Signal Transduction gRABs Attention. *Cell Signal* 18 (1), 1–8. doi:10.1016/j.cellsig.2005.07.001

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO Analysis Toolkit for the Agricultural Community. *Nucleic Acids Res.* 38 (Web Server issue), W64–W70. doi:10.1093/nar/gkq310

Duan, J., Zhang, M., Zhang, H., Xiong, H., Liu, P., Ali, J., et al. (2012). OsMIOX, a Myo-Inositol Oxygenase Gene, Improves Drought Tolerance through Scavenging of Reactive Oxygen Species in rice (Oryza Sativa L.). *Plant Sci.* 196, 143–151. doi:10.1016/j.plantsci.2012.08.003

Ebine, K., Fujimoto, M., Okatani, Y., Nishiyama, T., Goh, T., Ito, E., et al. (2011). A Membrane Trafficking Pathway Regulated by the Plant-specific RAB GTPase ARA6. *Nat. Cel Biol.* 13 (7), 853–859. doi:10.1038/ncb2270

Ebine, K., Miyakawa, N., Fujimoto, M., Uemura, T., Nakano, A., and Ueda, T. (2012). Endosomal Trafficking Pathway Regulated by ARA6, a RAB5 GTPase Unique to Plants. *Small GTPases* 3 (1), 23–27. doi:10.4161/sgtp.18299

Eisenach, C., Chen, Z. H., Grefen, C., and Blatt, M. R. (2012). The Trafficking Protein SYP121 of Arabidopsis Connects Programmed Stomatal Closure and K(+) Channel Activity with Vegetative Growth. *Plant J.* 69 (2), 241–251. doi:10.1111/j.1365-313X.2011.04786.x

Fang, X., Li, W., Yuan, H., Chen, H., Bo, C., Ma, Q., et al. (2021). Mutation of ZmWRKY86 Confers Enhanced Salt Stress Tolerance in maize. *Plant Physiol. Biochem.* 167, 840–850. doi:10.1016/j.plaphy.2021.09.010

Galon, Y., Nave, R., Boyce, J. M., Nachmias, D., Knight, M. R., and Fromm, H. (2008). Calmodulin-binding Transcription Activator (CAMTA) 3 Mediates Biotic Defense Responses in Arabidopsis. *FEBS Lett.* 582 (6), 943–948. doi:10.1016/j.febslet.2008.02.037

Grefen, C., Karnik, R., Larson, E., Lefoulon, C., Wang, Y., Waghmare, S., et al. (2015). A Vesicle-Trafficking Protein Commandeers Kv Channel Voltage Sensors for Voltage-dependent Secretion. *Nat. Plants* 1, 15108. doi:10.1038/nplants.2015.108

Grosshans, B. L., Ortiz, D., and Novick, P. (2006). Rabs and Their Effectors: Achieving Specificity in Membrane Traffic. *Proc. Natl. Acad. Sci. U. S. A.* 103 (32), 11821–11827. doi:10.1073/pnas.0601617103

Haller, E., Iven, T., Feussner, I., Stahl, M., Frohlich, K., Loffelhardt, B., et al. (2020). ABA-Dependent Salt Stress Tolerance Attenuates Botrytis Immunity in Arabidopsis. *Front. Plant Sci.* 11, 594827. doi:10.3389/fpls.2020.594827

Hamwieh, A., and Xu, D. (2008). Conserved Salt Tolerance Quantitative Trait Locus (QTL) in Wild and Cultivated Soybeans. *Breed. Sci.* 58 (4), 355–359. doi:10.1270/jsbbs.58.355

Hoepflinger, M. C., Geretschlaeger, A., Sommer, A., Hoeftberger, M., Nishiyama, T., Sakayama, H., et al. (2013). Molecular and Biochemical Analysis of the First ARA6 Homologue, a RAB5 GTPase, from green Algae. *J. Exp. Bot.* 64 (18), 5553–5568. doi:10.1093/jxb/ert322

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211

Irigoyen, J. J., Einerich, D. W., and Sánchez-Díaz, M. (1992). Water Stress Induced Changes in Concentrations of Proline and Total Soluble Sugars in Nodulated Alfalfa (Medicago Sativd) Plants. *Physiol. Plantarum* 84, 55–60. doi:10.1111/j.1399-3054.1992.tb08764.x

Ishitani, M., Liu, J., Halfter, U., Kim, C. S., Shi, W., and Zhu, J. K. (2000). SOS3 Function in Plant Salt Tolerance Requires N-Myristoylation and Calcium Binding. *Plant cell* 12 (9), 1667–1678. doi:10.1105/tpc.12.9.166710.2307/3871181

John, R., Ganeshan, U., Singh, B. N., Kaul, T., Reddy, M. K., Sopory, S. K., et al. (2016). Over-expression of Topoisomerase II Enhances Salt Stress Tolerance in Tobacco. *Front. Plant Sci.* 7, 1280. doi:10.3389/fpls.2016.01280

Kanehisa, M., Sato, Y., and Kawashima, M. (2021). KEGG Mapping Tools for Uncovering Hidden Features in Biological Data. *Protein Sci.* 31 (1), 47–53. doi:10.1002/pro.4172

Kempa, S., Krasensky, J., Dal Santo, S., Kopka, J., and Jonak, C. (2008). A central Role of Abscisic Acid in Stress-Regulated Carbohydrate Metabolism. *PLoS One* 3 (12), e3935. doi:10.1371/journal.pone.0003935

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* 12 (4), 357–360. doi:10.1038/nmeth.3317

Landi, M. (2017). Commentary to: "Improving the thiobarbituric acid-reactive-substances assay for estimating lipid peroxidation in plant tissues containing anthocyanin and other interfering compounds" by Hodges et al., Planta (1999) 207:604-611. *Planta* 245 (6), 1067. doi:10.1007/s00425-017-2699-3

Leyman, B., Geelen, D., Quintero, F. J., and Blatt, M. R. (1999). A Tobacco Syntaxin with a Role in Hormonal Control of Guard Cell Ion Channels. *Science (New York, N.Y.)* 283 (5401), 537–540. doi:10.1126/science.283.5401.537

Li, Y., Chen, Q., Nan, H., Li, X., Lu, S., Zhao, X., et al. (2017). Overexpression of GmFDL19 Enhances Tolerance to Drought and Salt Stresses in Soybean. *PLoS One* 12 (6), e0179554. doi:10.1371/journal.pone.0179554

Li, Z., Li, L., Zhou, K., Zhang, Y., Han, X., Din, Y., et al. (2019). GhWRKY6 Acts as a Negative Regulator in Both Transgenic Arabidopsis and Cotton During Drought and Salt Stress. *Front. Genet.* 10, 392. doi:10.3389/fgene.2019.00392

Liu, J. X., Srivastava, R., Che, P., and Howell, S. H. (2007). Salt Stress Responses in Arabidopsis Utilize a Signal Transduction Pathway Related to Endoplasmic Reticulum Stress Signaling. *Plant J.* 51 (5), 897–909. doi:10.1111/j.1365-313X.2007.03195.x

Liu, L., Wang, B., Liu, D., Zou, C., Wu, P., Wang, Z., et al. (2020). Transcriptomic and Metabolomic Analyses Reveal Mechanisms of Adaptation to Salinity in Which Carbon and Nitrogen Metabolism Is Altered in Sugar Beet Roots. *BMC Plant Biol.* 20 (1), 138. doi:10.1186/s12870-020-02349-9

Lopez-Berenguer, C., Martinez-Ballesta, M. C., Garcia-Viguera, C., and Carvajal, M. (2008). Leaf Water Balance Mediated by Aquaporins under Salt Stress and Associated Glucosinolate Synthesis in Broccoli. *Plant Sci.* 174 (3), 321–328. doi:10.1016/j.plantsci.2007.11.012

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Lu, W., Chu, X., Li, Y., Wang, C., and Guo, X. (2013). Cotton GhMKK1 Induces the Tolerance of Salt and Drought Stress, and Mediates Defence Responses to Pathogen Infection in Transgenic Nicotiana Benthamiana. *PLoS One* 8 (7), e68503. doi:10.1371/journal.pone.0068503

Manavalan, L. P., Guttikonda, S. K., Tran, L. S., and Nguyen, H. T. (2009). Physiological and Molecular Approaches to Improve Drought Resistance in Soybean. *Plant Cel. Physiol.* 50 (7), 1260–1276. doi:10.1093/pcp/pcp082

Martin-Davison, A. S., Perez-Diaz, R., Soto, F., Madrid-Espinoza, J., Gonzalez-Villanueva, E., Pizarro, L., et al. (2017). Involvement of SchRabGDI1 from Solanum Chilense in Endocytic Trafficking and Tolerance to Salt Stress. *Plant Sci.* 263, 1–11. doi:10.1016/j.plantsci.2017.06.007

Mazel, A., Leshem, Y., Tiwari, B. S., and Levine, A. (2004). Induction of salt and osmotic stress tolerance by overexpression of an intracellular vesicle trafficking protein AtRab7 (AtRabG3e). *Plant Physiol.* 134 (1), 118–128. doi:10.1104/pp.103.025379

Miao, H., Sun, P., Liu, J., Wang, J., Xu, B., and Jin, Z. (2018). Overexpression of a Novel ROP Gene from the Banana (MaROP5g) Confers Increased Salt Stress Tolerance. *Int. J. Mol. Sci.* 19 (10), 3108. doi:10.3390/ijms19103108

Munns, R., and Tester, M. (2008). Mechanisms of Salinity Tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi:10.1146/annurev.arplant.59.032607.092911

Paz, M. M., Martinez, J. C., Kalvig, A. B., Fonger, T. M., and Wang, K. (2006). Improved Cotyledonary Node Method Using an Alternative Explant Derived from Mature Seed for Efficient Agrobacterium-Mediated Soybean Transformation. *Plant Cel. Rep.* 25 (3), 206–213. doi:10.1007/s00299-005-0048-7

Peng, X., Ding, X., Chang, T., Wang, Z., Liu, R., Zeng, X., et al. (2014). Overexpression of a Vesicle Trafficking Gene, OsRab7, Enhances Salt Tolerance in rice. *ScientificWorldJournal* 2014, 483526. doi:10.1155/2014/483526

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33 (3), 290–295. doi:10.1038/nbt.3122

Phang, T. H., Shao, G., and Lam, H. M. (2008). Salt Tolerance in Soybean. *J. Integr. Plant Biol.* 50 (10), 1196–1212. doi:10.1111/j.1744-7909.2008.00760.x

Qi, D. H., and Lee, C. F. (2014). Influence of Soybean Biodiesel Content on Basic Properties of Biodiesel-Diesel Blends. *J. Taiwan Inst. Chem. Eng.* 45 (2), 504–507. doi:10.1016/j.jtice.2013.06.021

Reuveni, R. (1992). Peroxidase Activity as a Biochemical Marker for Resistance of Muskmelon (Cucumis Melo) to Pseudoperonospora Cubensis. *Phytopathology* 82, 749–753. doi:10.1094/phyto-82-749

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 (1), 139–140. doi:10.1093/bioinformatics/btp616

Rosa Téllez, S., Kanhonou, R., Castellote Bellés, C., Serrano, R., Alepuz, P., and Ros, R. (2020). RNA-Binding Proteins as Targets to Improve Salt Stress Tolerance in Crops. *Agronomy* 10 (2). doi:10.3390/agronomy10020250

Roy, S., Choudhury, S. R., Sengupta, D. N., and Das, K. P. (2013). Involvement of AtPol Lambda in the Repair of High Salt- and DNA Cross-Linking Agent-Induced Double Strand Breaks in Arabidopsis. *Plant Physiol.* 162 (2), 1195–1210. doi:10.1104/pp.113.219022

Salinas-Cornejo, J., Madrid-Espinoza, J., and Ruiz-Lara, S. (2019). Identification and Transcriptional Analysis of SNARE Vesicle Fusion Regulators in Tomato (Solanum lycopersicum) during Plant Development and a Comparative Analysis of the Response to Salt Stress with Wild Relatives. *J. Plant Physiol.* 242, 153018. doi:10.1016/j.jplph.2019.153018

Schwartz, S. L., Cao, C., Pylypenko, O., Rak, A., and Wandinger-Ness, A. (2007). Rab GTPases at a Glance. *J. Cel. Sci.* 120 (Pt 22), 3905–3910. doi:10.1242/jcs.015909

Singh, D., Yadav, N. S., Tiwari, V., Agarwal, P. K., and Jha, B. (2016). A SNARE-Like Superfamily Protein SbSLSP from the Halophyte Salicornia Brachiata Confers Salt and Drought Tolerance by Maintaining Membrane Stability, K^+/Na^+ Ratio, and Antioxidant Machinery. *Front. Plant Sci.* 7, 737. doi:10.3389/fpls.2016.00737

Sui, N., Wang, Y., Liu, S., Yang, Z., Wang, F., and Wan, S. (2018). Transcriptomic and Physiological Evidence for the Relationship between Unsaturated Fatty Acid and Salt Stress in Peanut. *Front. Plant Sci.* 9, 7. doi:10.3389/fpls.2018.00007

Sun, T. J., Fan, L., Yang, J., Cao, R. Z., Yang, C. Y., Zhang, J., et al. (2019). A Glycine max Sodium/hydrogen Exchanger Enhances Salt Tolerance through Maintaining Higher Na^+ Efflux Rate and K^+/Na^+ Ratio in Arabidopsis. *BMC Plant Biol.* 19 (1), 469. doi:10.1186/s12870-019-2084-4

Tester, M., and Davenport, R. (2003). Na+ Tolerance and Na+ Transport in Higher Plants. *Ann. Bot.* 91 (5), 503–527. doi:10.1093/aob/mcg058

Tester, M., and Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science* 327 (5967), 818–822. doi:10.1126/science.1183700

Tsutsui, T., Nakano, A., and Ueda, T. (2015). The Plant-Specific RAB5 GTPase ARA6 Is Required for Starch and Sugar Homeostasis in *Arabidopsis thaliana*. *Plant Cel Physiol.* 56 (6), 1073–1083. doi:10.1093/pcp/pcv029

Tuteja, N., Tran, N. Q., Dang, H. Q., and Tuteja, R. (2011). Plant MCM Proteins: Role in DNA Replication and beyond. *Plant Mol. Biol.* 77 (6), 537–545. doi:10.1007/s11103-011-9836-3

Ueda, T., Yamaguchi, M., Uchimiya, H., and Nakano, A. (2001). Ara6, a Plant-Unique Novel Type Rab GTPase, Functions in the Endocytic Pathway of *Arabidopsis thaliana*. *EMBO J.* 20 (17), 4730–4741. doi:10.1093/emboj/20.17.4730

Verma, D., Jalmi, S. K., Bhagat, P. K., Verma, N., and Sinha, A. K. (2020). A bHLH Transcription Factor, MYC2, Imparts Salt Intolerance by Regulating Proline Biosynthesis in Arabidopsis. *FEBS J.* 287 (12), 2560–2576. doi:10.1111/febs.15157

Wang, Z., Wang, J., Bao, Y., Wu, Y., and Zhang, H. (2010). Quantitative Trait Loci Controlling rice Seed Germination under Salt Stress. *Euphytica* 178 (3), 297–307. doi:10.1007/s10681-010-0287-8

Wang, Y., Jiang, L., Chen, J., Tao, L., An, Y., Cai, H., et al. (2018). Overexpression of the Alfalfa WRKY11 Gene Enhances Salt Tolerance in Soybean. *PLoS One* 13 (2), e0192382. doi:10.1371/journal.pone.0192382

Wu, X. X., Li, J., Wu, X. D., Liu, Q., Wang, Z. K., Liu, S. S., et al. (2016). Ectopic Expression of *Arabidopsis thaliana* Na+(K+)/H+ Antiporter Gene, AtNHX5, Enhances Soybean Salt Tolerance. *Genet. Mol. Res.* 15 (2). doi:10.4238/gmr.15027483

Yang, L., Tang, R., Zhu, J., Liu, H., Mueller-Roeber, B., Xia, H., et al. (2008). Enhancement of Stress Tolerance in Transgenic Tobacco Plants Constitutively Expressing AtIpk2beta, an Inositol Polyphosphate 6-/3-kinase from *Arabidopsis thaliana*. *Plant Mol. Biol.* 66 (4), 329–343. doi:10.1007/s11103-007-9267-3

Yin, C., Karim, S., Zhang, H., and Aronsson, H. (2017). Arabidopsis RabF1 (ARA6) Is Involved in Salt Stress and Dark-Induced Senescence (DIS). *Int. J. Mol. Sci.* 18 (2), 309. doi:10.3390/ijms18020309

Yuan, P., Du, L., and Poovaiah, B. W. (2018a). Ca^{2+}/Calmodulin-Dependent AtSR1/CAMTA3 Plays Critical Roles in Balancing Plant Growth and Immunity. *Int. J. Mol. Sci.* 19 (6), 1764. doi:10.3390/ijms19061764

Yuan, P., Tanaka, K., Du, L., and Poovaiah, B. W. (2018b). Calcium Signaling in Plant Autoimmunity: A Guard Model for AtSR1/CAMTA3-Mediated Immune Response. *Mol. Plant* 11 (5), 637–639. doi:10.1016/j.molp.2018.02.014

Zhang, L., Li, T., Wang, Y., Zhang, Y., and Dong, Y. S. (2019a). FvC5SD Overexpression Enhances Drought Tolerance in Soybean by Reactive Oxygen Species Scavenging and Modulating Stress-Responsive Gene Expression. *Plant Cel Rep.* 38 (9), 1039–1051. doi:10.1007/s00299-019-02424-y

Zhang, L., Liu, Y., Zhu, X. F., Jung, J. H., Sun, Q., Li, T. Y., et al. (2019b). SYP22 and VAMP727 Regulate BRI1 Plasma Membrane Targeting to Control Plant Growth in Arabidopsis. *New Phytol.* 223 (3), 1059–1065. doi:10.1111/nph.15759

Zhao, J., and Wang, X. (2004). Arabidopsis Phospholipase Dalpha1 Interacts with the Heterotrimeric G-Protein Alpha-Subunit through a Motif Analogous to the DRY Motif in G-Protein-Coupled Receptors. *J. Biol. Chem.* 279 (3), 1794–1800. doi:10.1074/jbc.M309529200

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 Hong, Li, Zhao, Yang, Zhang, Teng, Jing, Kong, Liu, Li, Meng, Wang and Zhang. 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.

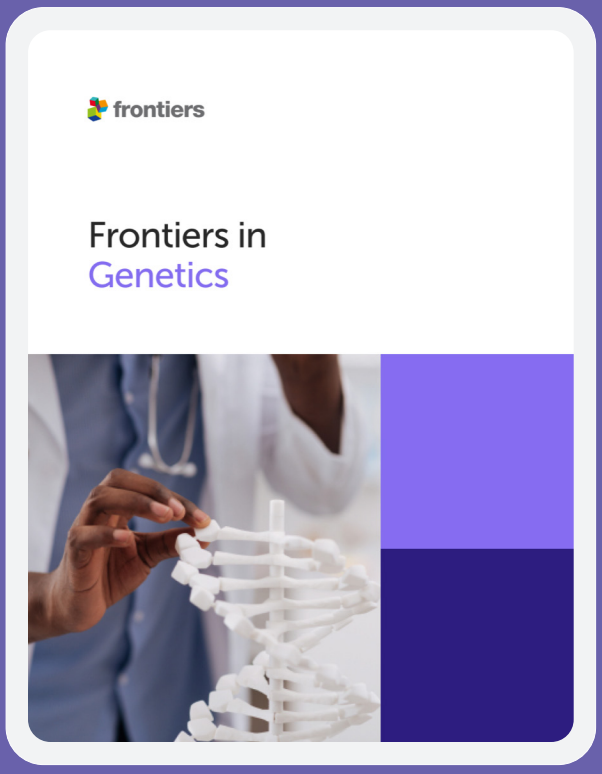
Frontiers in Genetics

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