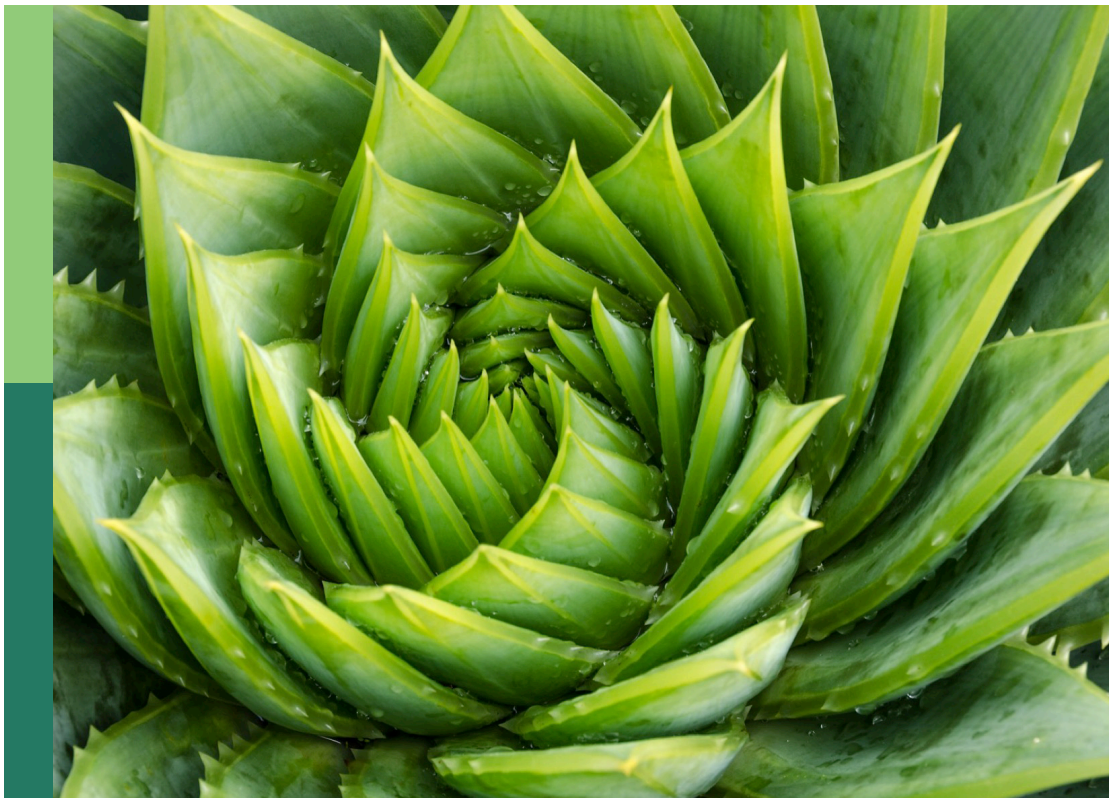
frontiers | Research Topics

Peanut genomics and biotechnology in breeding applications

Edited by
Weijian Zhuang and Rajeev K. Varshney

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-2781-8
DOI 10.3389/978-2-8325-2781-8

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

Peanut genomics and biotechnology in breeding applications

Topic editors

Weijian Zhuang — Fujian Agriculture and Forestry University, China

Rajeev K. Varshney — Murdoch University, Australia

Citation

Zhuang, W., Varshney, R. K., eds. (2023). *Peanut genomics and biotechnology in breeding applications.* Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-2781-8

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 07 June 2023
DOI 10.3389/fpls.2023.1226637

OPEN ACCESS

EDITED AND REVIEWED BY
Kyung Do Kim,
Myongji University, Republic of Korea

*CORRESPONDENCE
Weijian Zhuang
weijianz@fafu.edu.cn
Rajeev K. Varshney
rajeev.varshney@murdoch.edu.au

RECEIVED 22 May 2023
ACCEPTED 29 May 2023
PUBLISHED 07 June 2023

CITATION
Zhuang W and Varshney RK (2023) Editorial: Peanut genomics and biotechnology in breeding applications. *Front. Plant Sci.* 14:1226637. doi: 10.3389/fpls.2023.1226637

Editorial: Peanut genomics and biotechnology in breeding applications

Weijian Zhuang[1]* and Rajeev K. Varshney[1,2]*

[1]Center of Legume Plant Genetics and Systems Biology, College of Agriculture, Oil Crops Research Institute, Fujian Agriculture and Forestry University (FAFU), Fuzhou, China, [2]WA State Agricultural Biotechnology Centre, Crop Research Innovation Centre, Food Futures Institute, Murdoch University, Murdoch, WA, Australia

KEYWORDS

QTL mapping, genomic, food safely, molecular marker, biotechnology

Editorial on the Research Topic
Peanut genomics and biotechnology in breeding applications

Cultivated peanut (*Arachis hypogaea* L.) is a significant food and oil legume crop cultivated in tropical and subtropical regions globally. Climate change-mediated biotic and abiotic stresses significantly impact peanut productivity and agronomic traits. However, bioinformatics and next-generation sequencing technologies have facilitated the feasible application of peanut genomics and genetics resources in sustainable breeding programs. Genomics- and biotechnology-assisted breeding holds great potential to fast-track the rate of genetic improvement, and design improved peanut cultivars with high yield and quality to ensure food safety. Therefore, this Research Topic on "*Peanut genomics and biotechnology in breeding applications*" presents ten articles from leading experts in this field. This editorial highlights key advances reported in our Research Topic.

Genome-wide identification and characterization of new gene families in cultivated peanut helps to uncover their gene structures, evolutionary relationships, protein interactions, *cis*-elements, expression levels in different tissues and against various stresses, putative miRNAs, and their putative functions using numerous computational tools. In this context, Wang et al. identified 196 *R2R3-MYB* genes in peanut genome, arranged into 48 sub-groups. Of these, 90 genes displayed higher expression patterns against waterlogging stress. Association analysis identified a single nucleotide polymorphism (SNP) located in the third exon region of *AdMYB03-18* (*AhMYB033*), and the three haplotypes of the SNP were highly connected with total branch number, pod length, and root-shoot ratio, respectively, indicating that *AhMYB033* could help to improve the peanut yield. Likewise, Zhang et al. identified 22, 22 and 46 phospholipase Ds (*PLDs*) in *A. duranensis*, *A. ipaensis* and *A. hypogaea*, respectively, and classified them into α, β, γ, δ, ϵ, ζ and φ isoforms. AhPLDs interact with key proteins in lipid metabolic pathways, while ahy-miR3510, ahy-miR3513-3p, and ahy-miR3516 may act as hub regulators. qRT-PCR-based expression analysis indicated that mainly, *AhPLDα3A*, *AhPLDα5A*, *AhPLDβ1A*, *AhPLDβ2A* and *AhPLDδ4A* were highly up-regulated under abiotic stresses.

Wu et al. identified 178 plant protein phosphatase 2C (*PP2C*) genes in cultivated peanut, distributed across 20 chromosomes. Twenty two miRNAs from 14 diverse families targeted 57 *AhPP2Cs*. Some *AhPPC2Cs* were highly expressed in different tissues, while

AhPP2C45 and *AhPP2C134* were highly up-regulated under salinity stress, revealing their role in salinity tolerance in peanut. Another study by Cai et al. identified 116 o-methyltransferases (*OMTs*) genes, divided into three main classes. Twelve miRNAs from various families targeted 35 *AhOMTs* genes. Mostly AhOMTs were up-regulated in various tissues and under phytohormones, drought, and temperature stress. In the future, these novel genes could be genetically engineered to improve peanut yield, quality and stress tolerance.

Genome-wide association study by Oteng-Frimpong et al. discovered 97 SNPs and 17 candidate genes associated with early leaf spot (ELS) and late leaf spot (LLS) diseases in peanut. Of these, 29 unique SNPs were revealed for one or more traits across 16 chromosomes, explaining 0.01-62.76% phenotypic variation. These outcomes offer insights into the genetic structure of ELS and LLS diseases in African peanut germplasm. The identified SNPs and anticipated candidate genes could aid in breeding diseases resistant peanut varieties. Using whole genome resequencing method, Zhang et al. discovered a major quantitative trait loci (QTL, *qRGRB09*) associated with cold tolerance on chromosome B09 (between 46.74 cM-61.75 cM) that is confirmed by KASP markers. A regional QTL mapping assessment confirmed that *qRGRB09* was between the KASP markers, G22096 and G220967 (chrB09:155637831–155854093), and this region contained a total of 15 annotated genes, suggesting their vital role in cold tolerance in peanut. By performing transcriptome and bisulfite sequencing, Liu et al. exhibited that variations in DNA methylation between wild and cultivated peanuts can alter oil content by affecting the expression of peroxisomal acyl transporter protein (*Araip.H6S1B*). In short, they concluded that DNA methylation might act as a negative regulator of lipid metabolic genes and transcription factors, leading to elusive differences in oil accumulation between wild and cultivated peanuts.

After different transgenic screening assays, Huai et al. reported that the red fluorescence protein (DsRed2) could be potentially used as a visual reporter to attain the highest screening productivity and precision in peanut genetic transformation. In another study, Zhuang et al. functionally characterized a novel *Aspergillus flavus* inducible gene (*AhOMT1*) promoter from peanut. In transgenic *Arabidopsis*, *AhOMT1*-promoter showed highly inducible activities under *A. flavus* infection, presenting a new hope for future controlling of aflatoxins contamination through the induction of peanut resistance genes; thus, satisfying the safety concerns of the transgenes. These studies suggest that genetic engineering (gene editing or transgenic breeding) approaches must be fully exploited to improve important agronomic traits and stress tolerance in peanut.

In a review article, Huang et al. presented the scope of different omics approaches, including genomics, transcriptomics, proteomics, metabolomics, miRNAomics, epigenomics and phenomics, for identifying various biotic stress-related genes, proteins, metabolites, metabolic pathways and their networks. Omics understanding is important for developing biotic stress-resistant peanut cultivars to meet the peanut-food demands of the growing world population. The data obtained from these omics approaches can be further exploited *via* genetic or metabolic engineering to design improved peanut cultivars.

In short, this Research Topic contains a plethora of valuable insights into different areas of peanut genetics, genomics, biotechnology, and molecular biology. To further fast-track the development of modern genotypes that can meet future peanut-food necessities and withstand climate change, it is vital to constantly advance genomics and biotechnological techniques, as well as associated statistical methodologies, to help peanut breeders. With the recent advances in genome sequencing technologies, we are anticipated that the peanut community will join hands and develop peanut pan-genome and super pan-genome to further facilitate peanut research and breeding. The integration of different genomics, genetics, and biotechnology tools will be able to guide the peanut community to enhance the peanut production, quality and stress tolerance to meet the food and market demands.

Author contributions

WZ and RV prepared, reviewed and edited the draft of the editorial. All authors approved the submitted version.

Funding

WZ thanks Fujian Agriculture and Forestry University (FAFU), the National Natural Science Foundation of China, and the Science and Technology Foundation of Fujian Province of China, for supporting our peanut genomic and biotechnology research. RKV is thankful to Food Futures Institute, Murdoch University (Australia), for financial support.

Acknowledgments

We thank Ali Raza from FAFU, China, for help in preparing the first draft. We thank all the authors who have contributed to this Research Topic and the reviewers who have provided valuable feedback to enhance the quality of the articles. We also thank the Frontiers team for their support.

Conflict of interest

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

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 December 2022
DOI 10.3389/fpls.2022.1065267

OPEN ACCESS

EDITED BY
Daoquan Xiang,
National Research Council Canada (NRC), Canada

REVIEWED BY
Peng Gao,
Agriculture and Agri-Food Canada (AAFC), Canada
Zhaorong Hu,
China Agricultural University, China

*CORRESPONDENCE
Huifang Jiang
peanutlab@oilcrops.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

RECEIVED 19 October 2022
ACCEPTED 24 November 2022
PUBLISHED 16 December 2022

CITATION
Liu N, Wu B, Pandey MK, Huang L, Luo H, Chen Y, Zhou X, Chen W, Huai D, Yu B, Chen H, Guo J, Lei Y, Liao B, Varshney RK and Jiang H (2022) Gene expression and DNA methylation altering lead to the high oil content in wild allotetraploid peanut (*A. monticola*). *Front. Plant Sci.* 13:1065267. doi: 10.3389/fpls.2022.1065267

Gene expression and DNA methylation altering lead to the high oil content in wild allotetraploid peanut (*A. monticola*)

Nian Liu[1†], Bei Wu[1†], Manish K. Pandey[2], Li Huang[1], Huaiyong Luo[1], Yuning Chen[1], Xiaojing Zhou[1], Weigang Chen[1], Dongxin Huai[1], Bolun Yu[1], Hao Chen[3], Jianbin Guo[1], Yong Lei[1], Boshou Liao[1], Rajeev K. Varshney[2] and Huifang Jiang[1*]

[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China, [2]Center of Excellence in Genomics and Systems Biology (CEGSB), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [3]Institute of Crop Sciences, Fujian Academy of Agricultural Sciences, Fuzhou, China

Introduction: The wild allotetraploid peanut *Arachis monticola* contains a higher oil content than the cultivated allotetraploid *Arachis hypogaea*. Besides the fact that increasing oil content is the most important peanut breeding objective, a proper understanding of its molecular mechanism controlling oil accumulation is still lacking.

Methods: We investigated this aspect by performing comparative transcriptomics from developing seeds between three wild and five cultivated peanut varieties.

Results: The analyses not only showed species-specific grouping transcriptional profiles but also detected two gene clusters with divergent expression patterns between two species enriched in lipid metabolism. Further analysis revealed that expression alteration of lipid metabolic genes with co-expressed transcription factors in wild peanut led to enhanced activity of oil biogenesis and retarded the rate of lipid degradation. In addition, bisulfite sequencing was conducted to characterize the variation of DNA methylation between wild allotetraploid (245, WH 10025) and cultivated allotetraploid (Z16, Zhh 7720) genotypes. CG and CHG context methylation was found to antagonistically correlate with gene expression during seed development. Differentially methylated region analysis and transgenic assay further illustrated that variations of DNA methylation between wild and cultivated peanuts could affect the oil content *via* altering the expression of peroxisomal acyl transporter protein (*Araip.H6S1B*).

Discussion: From the results, we deduced that DNA methylation may negatively regulate lipid metabolic genes and transcription factors to subtly affect oil accumulation divergence between wild and cultivated peanuts. Our work provided the first glimpse on the regulatory mechanism of gene expression altering for oil accumulation in wild peanut and gene resources for future breeding applications.

KEYWORDS

Arachis monticola, *Arachis hypogaea*, oil biosynthesis, expression profiling, DNA methylation, functional genomics

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is one of the important oil crops, which is widely grown in more than 100 countries. The annual production is 48.76 Mt, with 29.60 Mha of global planting area for cultivated peanut (FAOSTAT, 2019, https://www.fao.org/faostat/en/#data/). It was domesticated from a wild relative, *Arachis monticola*, which harbors a high oil content and resistance to several biotic stresses (Bertioli et al., 2011; Huang et al., 2012; Moretzsohn et al., 2013; Yin et al., 2020). Since cultivated peanut is one of the major sources of edible oil in the world, enhancing oil content is the second most vital objective after yield in peanut breeding. There are examples in several crops that novel genes from wild counterparts were successfully introgressed into cultivated accessions for crop improvement (Hufford et al., 2012; Sang and Ge, 2013; Qi et al., 2014; Tian et al., 2019). The wild counterpart (*A. monticola*) not having any reproductive barrier with domesticated peanut (*A. hypogaea*) could provide favorable alleles for high oil accumulation. Developing a better understanding of the molecular mechanism and genomic control of high lipid accumulation in wild peanut would facilitate a significant increase in oil content in newly developed peanut varieties.

Lipids can be classified into fatty acids (FAs), galactolipids, phospholipids, sphingolipid, and acylglycerol (Manan et al., 2017). Tri-acylglycerol (TAG) is the major form of lipid in seed oil that provides calories and essential nutrients to the human body. In the model plant *Arabidopsis*, significant progress has been made in understanding lipid biosynthesis, transport, and degradation. Many genes encoding enzymes involved in lipid metabolism have been characterized, and several transcription families such as B3, NFY-B, AP2/EREBP, and bZIP have been reported to regulate these structural genes to control seed oil accumulation (Beisson, 2003; Li-Beisson, 2013; Manan et al., 2017). Only a couple of studies on expression variation underlying oil accumulation of peanut seed are available for domesticated peanut, *A. hypogaea* (Wang et al., 2018; Zhang et al., 2021). In the case of the wild tetraploid, *A. monticola*, a study very recently has been reported on the alteration of gene expression by structural variations affecting pod size (Yin et al., 2020) and not oil accumulation. Ample literature generated in the past provided convincing evidence related to the alteration of gene transcription contributing to phenotypic variations between domesticated crops and their wild counterparts (Koenig et al., 2013; Ichihashi et al., 2014; Yoo and Wendel, 2014; Lu et al., 2016). The last decade has witnessed monumental progress in terms of genomic resources including reference genomes, gene expression atlas, and genotyping assays in peanut to accelerate the genomics and breeding applications in peanut (Pandey et al., 2020). The availability of reference genomes for wild and cultivated tetraploid provides a unique opportunity to explore comparative structural and functional genomics to generate more information to further expand our understanding of the genomic and regulatory mechanisms of enhanced oil accumulation in *A. monticola*.

In addition to gene transcription, recent studies have also revealed the association between the alteration of DNA methylation and phenotypic variation in crops (Shen et al., 2018; Xu et al., 2019). In plants, DNA methylation occurs in the three sequence contexts of cytosine, namely, CG, CHG, and CHH (H represents A, T, or C) (Zhang et al., 2006; Zhang et al., 2018). In heterochromatin or euchromatic chromosome arms, DNA methylation plays a role in the control of transposon silencing and chromosome interaction (Zhang et al., 2006; Cokus et al., 2008). DNA methylation in gene promoters or bodies usually represses transcription but might increase transcription in some cases (Zhang et al., 2006; Wang et al., 2015; Bewick and Schmitz, 2017; Lang et al., 2017; Huang et al., 2019). The DNA methylation has been widely reported to play important roles in vegetable growth, fruit ripening, seed development, and response to biotic and abiotic stress (Zhong

et al., 2013; Baubec et al., 2014; Yong-Villalobos et al., 2015; Hewezi et al., 2017; Narsai et al., 2017; Huang et al., 2019; Rajkumar et al., 2020). However, the role of DNA methylation on the regulation of protein-coding genes involved in oil accumulation is largely unknown including in peanut. Integrated analysis of methylome and transcriptome data would investigate the alteration of DNA methylation and its association with genome-wide gene expression. Therefore, bisulfite sequencing (bisulfite-seq) and RNA sequencing (RNA-seq) technologies have been performed to reveal methylation regulation of genes involved in abiotic resistance, fruit development, and reproductive growth (Yong-Villalobos et al., 2015; Huang et al., 2019; Gutschker et al., 2022; You et al., 2022). These studies would expand our view of DNA methylation on the regulation of gene transcription.

In the present study, we conducted comparative transcriptome profiling of developing seeds from multiple accessions of wild and cultivated peanuts using RNA-seq. Two developmental stages in seed were studied, R5 and R8, representing stages of lipid initial accumulation and lipid rapid increase, respectively. In addition, bisulfite-seq was performed to investigate the variation of DNA methylation between wild accession (245, WH 10025) and cultivated accession (Z16, Zhh 7720). The objective of this study is to reveal a regulatory mechanism of gene expression altering for enhanced oil accumulation in peanut.

Materials and methods

Plant material

Three accessions, 171 (WH 4335), 172 (WH 4334), and 245 (WH 10025), were collected to represent wild peanuts (*A. monticola*). Five accessions, 003 (Zhh 0225), 145 (Zhh 0888), 492 (Zhh 0003), 502 (Zhh 0602), and Z16 (Zhh 7720), were used to represent cultivated peanuts (*A. hypogaea*). Among them, 003, 145, 492, and 502 belong to four mainly agronomic types (*vulgaris*, *hypogaea*, *fastigiata*, and *hirsuta*, respectively) in cultivated peanut. Z16 is a modern cultivar with mixed parentage. Wild and cultivated peanuts were planted in the experimental nursery of Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan. Nursery management followed standard agricultural practices. Three biological replications were grown for each accession. Developing seeds from eight accessions were collected at previously characterized stages (Pattee et al., 1974): lipid initial accumulation stage (R5) in which developing seeds are smaller and lipid rapid increase stage (R8) in which developing seeds are bigger. The collected samples were immediately frozen in liquid nitrogen and stored at -70°C for RNA and DNA isolation. The oil content (%) of mature seeds from each accession was measured as described previously (Liu et al., 2020).

Processing of RNA sequencing data

Total RNA was extracted using an RNAprep pure plant kit (Tiangen, Beijing, China) according to the manufacturer's instructions. Forty-eight libraries from the developing seeds of wild and cultivated peanuts were constructed and sequenced using a HiSeq XTen platform in Beijing Genomics Institute (BGI; https://www.genomics.cn/). After obtaining raw data, we used software SOAPnuke to perform quality filtering and read trimming (https://github.com/BGI-flexlab/SOAPnuke). The clean RNA-seq reads were mapped to the reference genome using HISAT2 software (Gao et al., 2016). The parameters were set as follows: –phred64 –sensitive –no-discordant –no-mixed –mp 6,2 -X 1000. The reference genome consisted of two diploid ancestors' genomes of *Arachis ipaensis* (V14167) and *Arachis duranensis* (V14167) (Bertioli et al., 2016). The RSEM package was used to calculate and normalize the gene expression level as fragments per kilobase of transcript per million mapped reads (FPKM) (Koenig et al., 2013). Differentially expressed genes (DEGs) with FPKM ≥1 in at least one sample (fold change ≥2 and adjusted *P*-value ≤0.001) were identified using DEGseq package (Lin et al., 2012).

Gene clustering, functional annotation, and weighted gene co-expression network analysis

K-means clustering was used to visualize genes exhibiting a similar expression pattern, and it was performed on normalized FPKM values using MeV software (Saeed et al., 2003). The distance metric for K-means clustering was set as Euclidean distance. Annotation analyses were performed by BLASTing public protein databases, including Nr (http://www.ncbi.nlm.nih.gov), Gene Ontology (GO) (http://www.geneontology.org), and Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg). GO annotation of all of the genes in the two diploid ancestors' genomes was downloaded from the website AgriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) and was set as background reference to identify and show overrepresented GO terms using software TBtools (Chen et al., 2020). The R package [weighted gene co-expression network analysis (WGCNA)] was used to build weighted gene co-expression networks (Langfelder and Horvath, 2008). Network construction was performed using the blockwiseModules function with default parameters. The topological overlap matrix (TOM) was calculated to measure the strength of a co-

expression relationship, i.e., connectivity between any two genes with respect to all other genes in the network. The expressed genes with FPKM ≥1 in at least one sample were selected to perform K-means clustering and WGCNA.

Whole-genome bisulfite sequencing

Genome DNA from each sample was extracted using a DNeasy Plant Maxi Kit (Qiagen, Germany). Then, DNA was fragmented to a mean size of 250 bp through Bioruptor (Diagenode, Belgium). Adapters were ligated to the fragment DNA and treated with sodium bisulfite using EZ DNA Methylation-Gold Kit (Zymo, USA). Sequencing was performed using Illumina HiSeq platform in paired-end mode at BGI (https://www.genomics.cn/). Three biological replications were sequenced for each stage of developing seeds in both wild and cultivated peanuts.

Read alignment and methylcytosine identification

The raw reads from each library were processed to remove low-quality reads, adaptor sequences, and contamination using software Trimmomatic (Bolger et al., 2014). Bisulfite sequence mapping program (BSMAP) was conducted to map clean reads to the reference genome with parameters (-u -v 8 -z 33 -p 4 -n 0 -w 20 -s 16 -f 10 -L 150), and only the reads mapped at unique positions were retained (Xi and Li, 2009). The binomial test was performed for each cytosine base to identify true methylcytosine (mC). Only the cytosines covered by at least four reads in all compared tissues were considered for further analysis. Cytosine sites with *P*-value <0.0001 were defined as methylated cytosine sites. The methylation level at each mC site was determined by the percentage of reads giving a methylation call to all of the reads aligned at the same site.

Identification of differentially methylated regions

Putative differentially methylated regions (DMRs) were identified using windows that contained at least five CG (CHG or CHH) sites with different methylation levels (cutoff value >0.1 between two compared samples) and Fisher test *P*-value ≤0.05. In addition, two nearby DMRs would be considered interdependent and joined into one continuous DMR if the genomic region from the start of an upstream DMR to the end of a downstream DMR also had different methylation levels (cutoff value >0.1 between two compared samples) with a *P*-value ≤0.05. Otherwise, the two DMRs were viewed as independent. When the genic region (2 kb upstream or body) was overlapped with DMR, the gene was defined as a DMR-associated gene. The fold enrichment of DEGs in DMR-associated genes was calculated as (DMR-associated DEGs/Total DEGs)/(DMR-associated genes/Total genes), and *P*-value significance was generated using the hypergeometric test.

Plant transformation

The full-length open reading frame (ORF) of *CTS* (*Araip.H6S1B*) was cloned and subsequently recombined into pBWA(V)HS plasmid (from Wuhan BioRun Biosciences Co., Ltd., China) to generate *35S:CTS* construct. *Arabidopsis* plant transformation was performed by the floral dip method (Zhang et al., 2006). Peanut cotyledons from germinating seeds were used as explant for hairy root transformation following a previous report (Yuan et al., 2019).

RNA isolation and qRT-PCR

Total RNA was isolated using plant RNA Extract Kit (Tiangen Biotech, China). The first-strand cDNA was synthesized following the instruction of Revert Aid First Strand cDNA Synthesis Kit (Thermo Fisher, USA). qRT-PCR was performed using a CFX Connect Real-time system (Bio-Rad, USA) with ChamQ Universal SYBR qPCR Mix (Vazyme, China). The relative expression levels were calculated using the $2^{-\Delta CT}$ method. *GAPDH* (Morgante et al., 2011) and *AtACTIN7* (AT5G09810) were chosen as reference genes to normalize the relative expression level of *CTS* (*Araip.H6S1B*) in peanut and *Arabidopsis*, respectively.

Fat red 7B staining

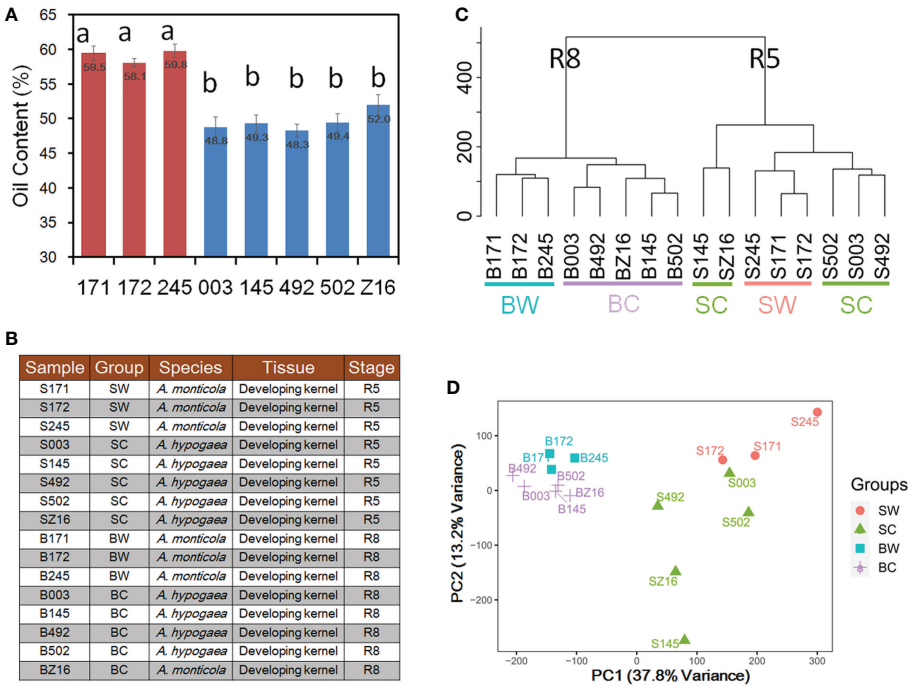
Fat red staining was performed by incubating *Arabidopsis* seedlings and peanut hairy roots in 0.1% (w/v) Fat red 7B solution for 3 h at room temperature. Samples were then rinsed with 70% ethanol to remove chlorophyll. The samples were quickly photographed using a stereomicroscope (Olympus SZX16, Japan).

Results

Evaluation of RNA sequencing data

To analyze phenotypic variations in seed oil accumulation between wild tetraploid and cultivated peanuts, we measured oil content in three accessions of *A. monticola* and five accessions of *A. hypogaea*. The oil content of mature seed in three wild accessions, namely, 171 (WH 4335), 172 (WH 4334), and 245

10.3389/fpls.2022.1065267



B

Sample	Group	Species	Tissue	Stage
S171	SW	*A. monticola*	Developing kernel	R5
S172	SW	*A. monticola*	Developing kernel	R5
S245	SW	*A. monticola*	Developing kernel	R5
S003	SC	*A. hypogaea*	Developing kernel	R5
S145	SC	*A. hypogaea*	Developing kernel	R5
S492	SC	*A. hypogaea*	Developing kernel	R5
S502	SC	*A. hypogaea*	Developing kernel	R5
SZ16	SC	*A. hypogaea*	Developing kernel	R5
B171	BW	*A. monticola*	Developing kernel	R8
B172	BW	*A. monticola*	Developing kernel	R8
B245	BW	*A. monticola*	Developing kernel	R8
B003	BC	*A. hypogaea*	Developing kernel	R8
B145	BC	*A. hypogaea*	Developing kernel	R8
B492	BC	*A. hypogaea*	Developing kernel	R8
B502	BC	*A. hypogaea*	Developing kernel	R8
BZ16	BC	*A. monticola*	Developing kernel	R8

FIGURE 1
Analysis of wild and cultivated peanuts at two seed developmental stages. **(A)** Seed oil content of three wild and five cultivated accessions. **(B)** Illustration of 16 samples for RNA sequencing (RNA-seq). Different letters indicated statistically significant differences in oil content between accessions according to Tukey's range test at the 0.05 level. **(C)** Hierarchical clustering between 16 samples. **(D)** Principal component analysis (PCA) of gene expression profiles in 16 samples. SW and SC indicated seeds of wild and cultivated peanuts at the R5 stage, respectively. BW and BC denoted seeds of wild and cultivated peanuts at the R8 stage, respectively.

(WH 10025), ranged from 58.1% to 59.8% and that of five cultivated peanuts, namely, 003 (Zhh 0225), 145 (Zhh 0888), 492 (Zhh 0003), 502 (Zhh 0602), and Z16 (Zhh 7720), ranged from 48.3% to 52.0% (Figure 1A). Multiple comparison analyses indicated that wild tetraploid species have significantly higher oil content than cultivated peanuts. To explore the reason for oil accumulation difference during peanut domestication, developing seeds at two stages (R5 and R8) were collected to generate RNA-seq datasets for both species (Figure 1B). Sixteen samples with three biological replications were used to construct RNA-seq libraries followed by generation of 44.1–45.4 million clean reads per library (Table S1). The union set of genomes of diploid ancestors *A. ipaensis* and *A. duranensis* (Bertioli et al., 2016) was used as a reference in this study. Averages of 83.72% and 83.59% of clean reads were mapped on the reference genome for wild and cultivated peanuts, respectively (Table S1). This similar mapping rate indicated that the reference can be used to quantify the gene expression level in both wild and cultivated species. The number of genes among different libraries ranged from 32,148 to 45,226 genes with the expression of a total of 56,236 genes among 16 samples (Tables S1, S2). Hierarchical clustering analysis was performed to cluster the samples according to developmental stages (Figure 1C). In the stage R8 group, the samples could be further divided into wild (BW) or cultivated (BC) subgroups. However, the samples in the R5 group could not be clearly distinguished between wild and cultivated subgroups. The first two principal component analysis (PCA) components (Figure 1D) of transcriptional profiles explained 37.8% (PC1) and 13.2% (PC2) of sample-to-sample variance. Similar to hierarchical clustering analysis, the PCA also grouped samples according to developmental stages (PC1) prior to species (PC2).

Differentiation in gene expression between wild and cultivated peanuts

To explore the divergence of gene expression patterns between two species, transcriptional profiles of 16 samples were used to perform K-mean clustering analysis. A total of 36,850 genes with FPKM ≥1 in at least one sample were selected to be grouped into nine clusters designated as C1–C9 (Figure 2A). Five clusters (C1–C5) showed a decreasing tendency from the R5 to R8 stage in both peanut species. Conversely, the C6 and C8 clusters showed an increasing trend across accessions from the R5 to R8 stage. For the C7

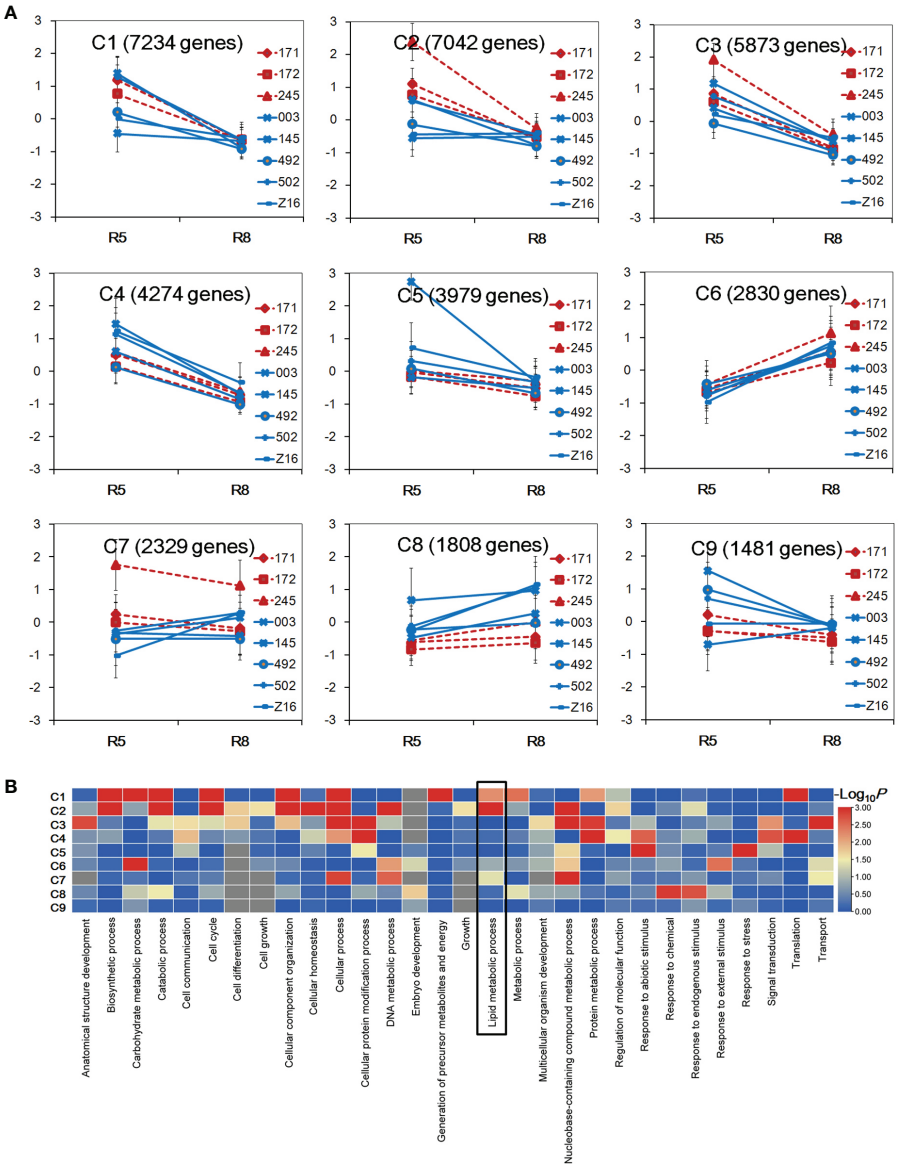


FIGURE 2
Gene expression pattern between wild and cultivated peanuts. **(A)** Identification of gene clusters in 16 samples. Nine gene clusters (C1–C9) were identified using K-means clustering. The x-axis represented the stage of developing seed. The y-axis represented log2 FPKM derived from RNA sequencing (RNA-seq) data for each sample FPKM fragments per kilobase of transcript per million mapped reads. **(B)** Heatmap of enrichment of biological process category among nine clusters. The color was indicated −log10 (*P*-value). The lipid metabolism process was boxed in the heatmap.

cluster, wild accessions (171, 172, and 245) showed a decreasing trend, while cultivated accessions showed an increasing pattern or no obvious change during seed developmental stages. In the C2 cluster, the gene expression level was higher in wild accessions at the R5 stage. In the C8 cluster, the gene expression level was in general higher in cultivated accessions at both stages. GO category analysis was performed to identify overrepresented GO terms of the nine clusters (Figure 2B, Table S3). A total of 30, 14, and 22 GO terms were enriched in biological process, cellular component, and molecular function, respectively. Each cluster had differently enriched GO terms, and the number of overrepresented GO terms ranged from 3 (C9) to 35 (C2). The GO term of lipid metabolism process was found overrepresented in clusters C1, C2, and C7 ($P < 0.05$). Interestingly, the gene expression level was different between wild and cultivated peanuts in C2 and C7 clusters, indicating that the transcriptional profile of lipid metabolism may be divergent between the two species.

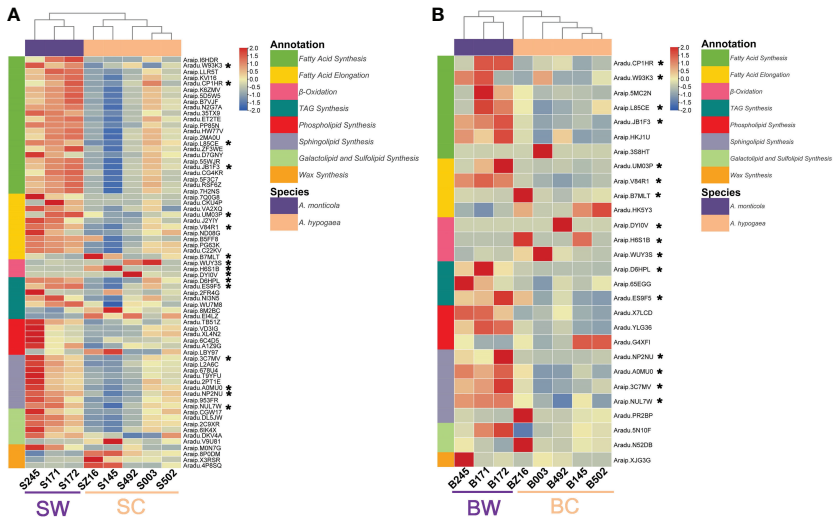


FIGURE 3
Heatmaps of lipid metabolism-related DEGs between wild and cultivated peanuts. Panels **(A)** and **(B)** represented the R5 and R8 stages, respectively. * denoted the differentially expressed genes identified at both stages. SW and SC indicated seeds of wild and cultivated peanuts at the R5 stage, respectively. BW and BC denoted seeds of wild and cultivated peanuts at the R8 stage, respectively.

To further dissect the expression difference of lipid metabolism, a comparative transcriptomic analysis was performed to identify DEGs between wild and cultivated peanuts. There were 5,647 and 3,184 DEGs at the R5 (SW vs. SC) and R8 (BW vs. BC) stages, respectively (Figure S1; Tables S4, S5). A total of 1,578 DEGs were detected at both R5 and R8 stages. The number of upregulated and downregulated DEGs was overall equivalent at both R5 and R8 stages (Figure S1). Heatmaps of DEGs involved in FA synthesis, FA elongation, TAG synthesis, phospholipid synthesis, sphingolipid synthesis, galactolipid and sulfolipid synthesis, wax synthesis, and β-oxidation were profiled to represent a transcriptional change in lipid metabolism (Figure 3). According to the expression pattern of DEGs involved in lipid metabolism, samples in both heatmaps (R5 and R8 stages) could be divided into two groups (wild and cultivated). At the R5 stage, most lipid metabolism-related pathways, such as FA synthesis, FA elongation, TAG synthesis, phospholipid synthesis, sphingolipid synthesis, and galactolipid and sulfolipid synthesis were dramatically upregulated in wild species (Figure 3A). Most downregulated DEGs at the R5 stage were mainly distributed on β-oxidation and wax synthesis pathways. Compared with the R5 stage, the portion of upregulated DEGs in wild species was generally lower at the R8 stage (Figure 3B). However, DEGs in FA synthesis, TAG synthesis, phospholipid synthesis, and sphingolipid synthesis were mainly upregulated in wild peanut accessions at the R8 stage. Conversely, all of the DEGs in β-oxidation were downregulated at the R8 stage. There were 16 lipid metabolic DEGs simultaneously identified at both R5 and R8 stages belonging to FA synthesis (5), FA elongation (3), TAG synthesis (2), sphingolipid synthesis (4), and β-oxidation (3) pathways. According to the results of the clustering analysis, 81 lipid metabolism-related DEGs were grouped into eight clusters. About 44% of lipid metabolism-related DEGs (36) belonged to C2 clusters in which the expression pattern was divergent between wild and cultivated peanuts. Another two "divergent" clusters (C7 and C8) harbored six and five lipid metabolism-related DEGs (Tables S4, S5).

Construction of the co-expression network

To further explore genes with high connectivity to lipid metabolism-related DEGs, WGCNA (Langfelder and Horvath, 2008) was performed to construct a TOM of expression similarity between genes (Table S6). Genes co-expressed with lipid metabolism-related DEGs (expression similarity ≥0.2) were selected to perform GO enrichment analysis. They were enriched in 14 GO terms for molecular function category, including transcription regulator activity, DNA-binding transcription factor (TF) activity, and DNA binding (Figure 4A). It was suggested that TFs may play a role in the co-expression network. Many TFs were differentially expressed between wild and cultivated peanuts (Figure S3). In specific TF families, such as AP2, B3, and bZIP, the family members were predominantly upregulated in wild peanuts at both R5 and R8 stages. A total of 391 TFs belong to 47 families that were co-expressed with DEGs involved in lipid synthesis pathways, such as FA synthesis, FA elongation, TAG synthesis, phospholipid

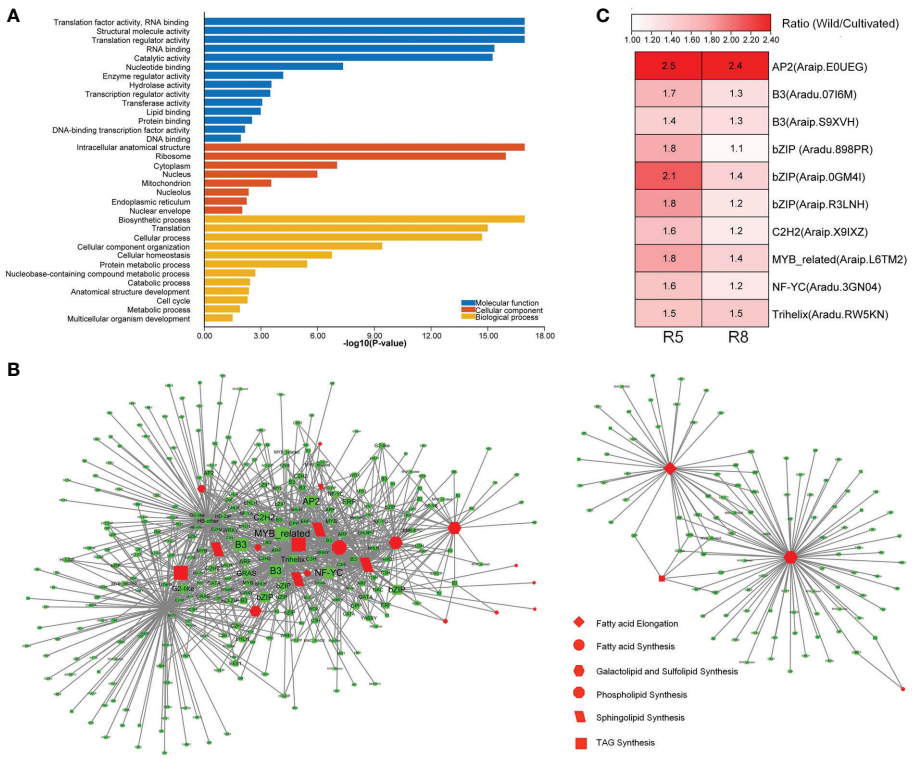


FIGURE 4
Identification of co-expression genes with DEGs involved in lipid metabolism. **(A)** GO enrichment analysis of co-expression genes. GO terms with *P*-value <0.05 were listed in the y-axis. **(B)** Network of co-expression TFs with lipid metabolic DEGs. **(C)** Differential expression of top 10 co-expression TFs between wild and cultivated peanuts. Ratio (Wild/Cultivated) represented fold change of FPKM values between wild and cultivated peanut. FPKM fragments per kilobase of transcript per million mapped reads.

synthesis, sphingolipid synthesis, and galactolipid and sulfolipid synthesis (Figure 4B). According to the interaction number, the top 10 notes in the co-expressed TFs (Table S6) were AP2 (*Araip.E0UEG*), B3 (*Aradu.07I6M*), B3 (*Araip.S9XVH*), bZIP (*Aradu.898PR*), bZIP (*Araip.0GM4I*), bZIP (*Araip.R3LNH*), C2H2 (*Araip.X9IXZ*), MYB-related (*Araip.L6TM2*), NF-YC (*Aradu.3GN04*), and Trihelix (*Aradu.RW5KN*). Fold change (wild/cultivated) of the TFs ranged from 1.4 to 2.5 at the R5 stage with a mean value of 1.8, while the values ranged from 1.1 to 2.4 at the R8 stage with a mean value of 1.4 (Figure 4C). Interestingly, six of the top 10 co-expressed TFs, i.e., AP2 (*Araip.E0UEG*), B3 (*Aradu.07I6M*), bZIP (*Aradu.898PR, Araip.0GM4I, Araip.R3LNH*), and Trihelix (*Aradu.RW5KN*), were divided into the C2 clusters (Table S6). There was a quarter of co-expressed TFs (101) belonging to the C2 clusters in which the gene expression pattern was divergent between wild and cultivated peanuts. Among the co-expressed TFs in the C2 "divergent" clusters, 43 members including AP2, bZIP, ARF, C3H, C2H2, and HD-ZIP were significantly upregulated in wild peanuts, especially at the R5 stage. Together with co-expressed TFs, ~44% of lipid metabolism-related DEGs were also categorized into the C2 clusters, suggesting that there was a divergent TF module to regulate the expression of lipid metabolism-related genes between wild and cultivated peanuts.

Influence of DNA methylation on gene expression

The available literature in multiple crops proved that DNA methylation could help regulate gene expression (Wang et al., 2015; Huang et al., 2019; Rajkumar et al., 2020; Wang et al., 2016; Xing et al., 2015). To reveal the relationship between DNA methylation and gene expression in peanut, wild peanut (245) with the highest oil content among the eight accessions and elite cultivar (Z16) with a relatively low oil content were selected to perform bisulfite-seq for developing seeds at the R5 and R8 stages. Four samples (S245, SZ16, B245, BZ16) representing seeds at the R5 (S) and R8 (B) stages for 245 and Z16 were sequenced with three biological replicates. About 480 M clean reads were generated for each sample (Table S7), and approximately 76% of the clean reads were uniquely mapped to the reference genomes covering >87% of the genomic cytosine positions. Each methylome had >16-fold average depth per

10.3389/fpls.2022.1065267

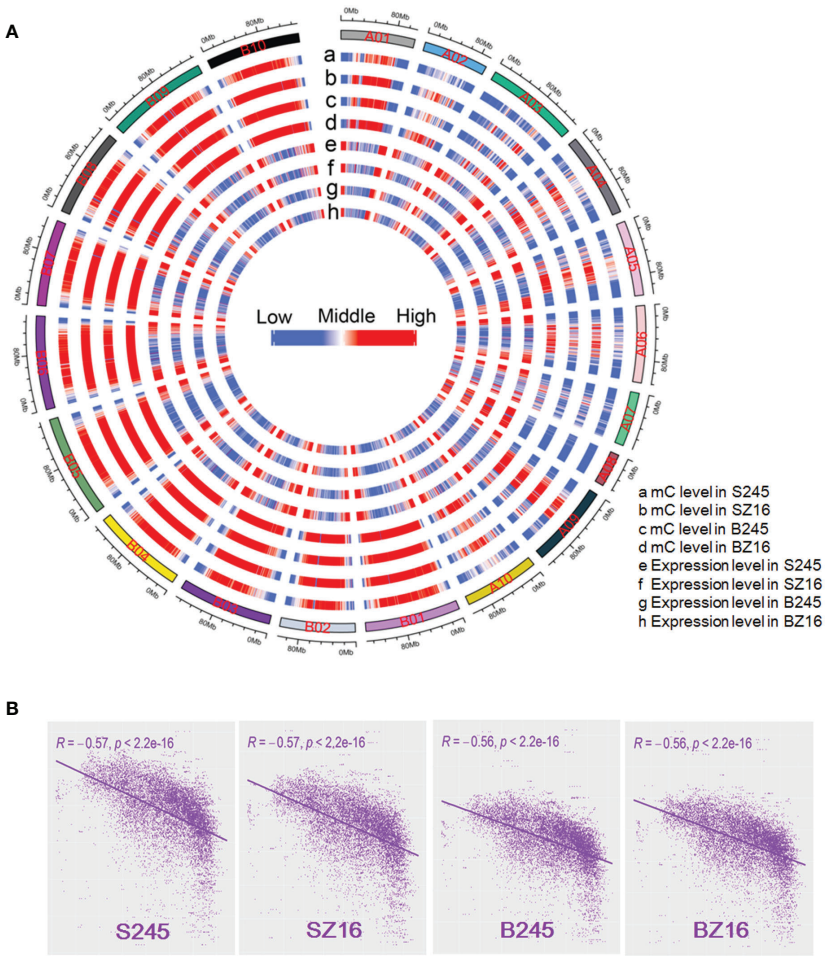


FIGURE 5
Genome-wide correlation between DNA methylation and gene expression. **(A)** A circle plot showed the overall DNA methylation level and gene expression level (log10 FPKM) in the reference chromosomes. FPKM fragments per kilobase of transcript per million mapped reads. mC indicated methylated cytosine. The data for each chromosome were analyzed in 1 Mb windows sliding 200 kb. **(B)** Analysis of correlation coefficient between DNA methylation and gene expression. S245 and SZ16 indicated seeds of 245 and Z16 at the R5 stage, respectively. B245 and BZ16 indicated seeds of 245 and Z16 at the R8 stage, respectively.

strand. Methylcytosines were identified in CG, CHG, and CHH contexts across samples. Compared with the average methylation of the CHH context (11.2%–22.3%), the level was much higher in CG (81.7%–86.0%) and CHG (73.2%–78.2%) contexts (Figure S3A). The fraction of mC was 20.1%–25.2% in CG, 25.6%–31.8% in CHG, and 43.0%–54.4% in CHH. The overall methylation levels in the four samples were similar (29.3%–34.7%) (Figure 5A, Figure S3A). The distribution of mC showed much lower methylation in the terminal chromosomes in contrast to a much higher gene expression in the terminal chromosomes (Figure 5A). The genome-wide correlation coefficient between overall DNA methylation and gene expression in the four samples was ~-0.56 ($P < 2.2$ e-16), indicating a significantly antagonistic correlation (Figure 5B).

To further characterize epigenetic variations between 245 (wild) and Z16 (cultivated), DMRs of CG, CHG, and CHH contexts were detected at the R5 and R8 stages (Table S8). The chromosome-wide view of DMR distribution indicated CG-DMR being the most likely to be enriched in gene-rich regions (Figures 6A, B). Approximately 40% of CG-DMRs were located in the genic region (gene body+2k upstream); the portions were reduced to ~15% and ~10% for CHG-DMRs and CHH-DMRs, respectively (Figure 6C). Conjoint analysis of the expression profile and DMRs indicated that CG-DMRs and CHG-DMRs were significantly enriched in the genic region (gene body+2k upstream) of DEGs between 245 and Z16 (Figure 6D). It is hinted that CG-DMR and CHG-DMR, not CHH-DMR, may play a role in the change of gene expression. Among DEG-overlapped CG-DMRs, the portion of hyper-DMRs in S245-upregulated DEGs was 50.3% and 51.9% at the R5 and R8 stages, respectively, and that in S245-downregulated DEGs could increase to 70.5% and 71.1% at the R5 and R8 stages,

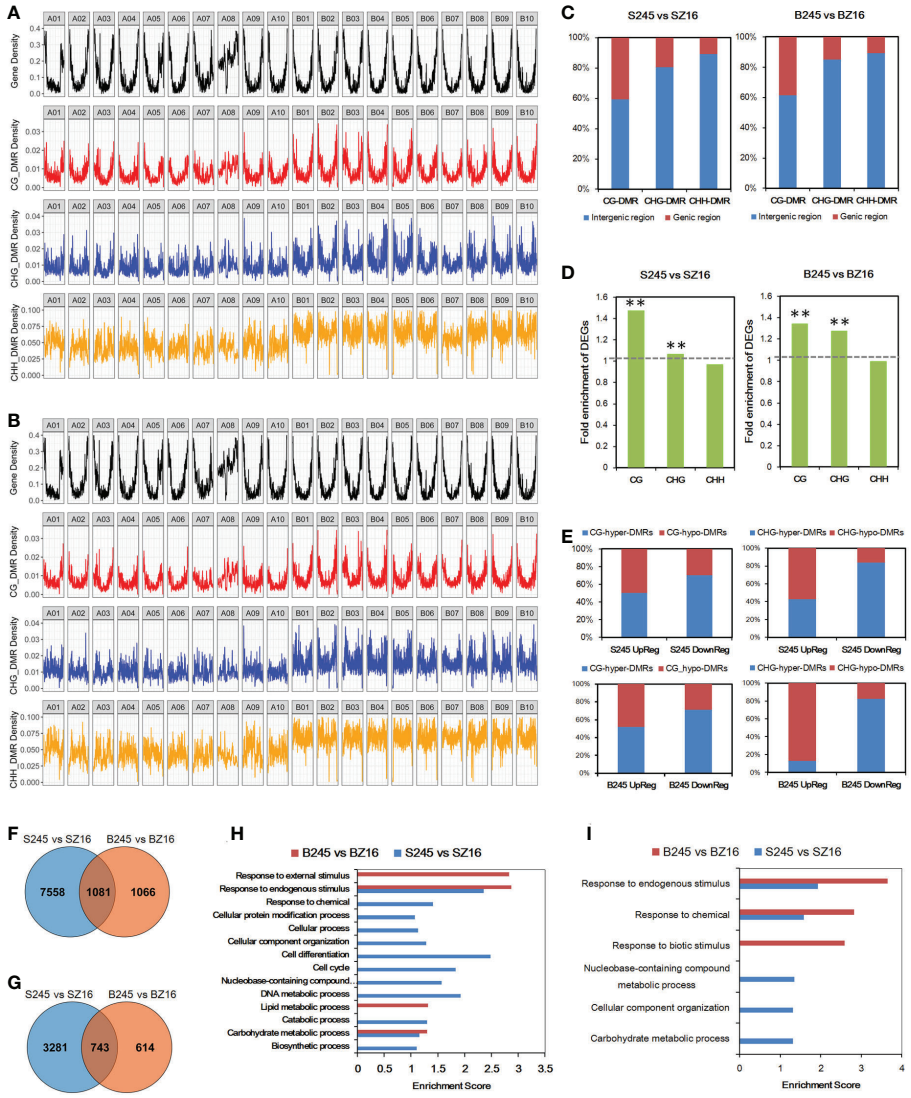


FIGURE 6
Influence of DMR on differential gene expression. **(A, B)** Genome-wide distribution of protein-coding genes and DMRs (CG, CHG, and CHH) between wild and cultivated seeds at the R5 stage **(A)** and R8 stage **(B)**. **(C)** Percentage of DMRs in the genic region and intergenic region. **(D)** Fold enrichment of DEGs in those overlapping with DMRs in the gene body with 2 kb upstream. The hypergeometric test was used to infer statistical significance (** $P < 0.01$). **(E)** The ratio of (hyper/hypo) DMRs in CG or CHG contexts overlapped with DEGs (up/down) between seeds of 245 and Z16 at the R5 and R8 stages. **(F)** Venn diagram of CG-DMR-associated DEGs at the R5 and R8 stages. **(G)** Venn diagram of CHG-DMR-associated DEGs between 245 and Z16 at the R5 and R8 stages. **(H)** GO enrichment analysis of CG-DMR-associated DEGs between 245 and Z16 at the R5 and R8 stages. GO terms with *P*-value <0.05 were listed in the y-axis. **(I)** GO enrichment analysis of CHG-DMR-associated DEGs between 245 and Z16 at the R5 and R8 stages. GO terms with *P*-value <0.05 were listed in the y-axis. S245 and SZ16 denoted seeds of 245 and Z16 at the R5 stage, respectively. B245 and BZ16 denoted seeds of 245 and Z16 at the R8 stage, respectively. DMR indicated differentially methylated regions.

respectively (Figure 6E). Similarly, the portion of hyper-DMRs was much higher in S245-downregulated DEGs (83.7% and 82.4% at R5 and R8, respectively) than in S245-upregulated DEGs (42.7% and 12.5% at R5 and R8, respectively) for DEG-overlapped CHG-DMRs. The results suggested that the dynamic change of methylated CG (mCG) and methylated CHG (mCHG) on a genic scale correlated negatively with the difference in gene abundance between 245 and Z16.

Role of DNA methylation in the divergence of lipid metabolism

The CG-DMRs between 245 and Z16 were overlapped with 8,639 and 2,147 DEGs at the R5 and R8 stages, respectively (Figure 6F). For CHG-DMRs, there were 4,024 and 1,357 CHG-DMR-associated DEGs at the R5 and R8 stages, respectively (Figure 6G). In general, more than 50% and 25% of DEGs

between 245 and Z16 may be associated with CG-DMRs and CHG-DMRs, respectively. GO enrichment analysis showed that CG-DMR-associated DEGs were enriched in the carbohydrate metabolic process and response to endogenous stimulus at both stages (Figure 6H). The CHG-DMR-associated DEGs were enriched in response to chemical and response to endogenous stimulus at both stages (Figure 6I). Interestingly, lipid metabolism was one of the enriched terms at the R8 stage for CG-DMR-associated DEGs.

There were 41 and 19 lipid metabolic DEGs that showed a negative correlation with DMRs in the CG and CHG contexts between 245 and Z16 at the R5 stage, respectively (Figure 7A). Most genes involved in FA synthesis, FA elongation, phospholipid synthesis, sphingolipid synthesis, galactolipid and sulfolipid synthesis, and TAG synthesis pathways exhibited hypomethylation with a higher expression in 245 at the R5 stage. In addition, 33 and 20 TFs known to be involved in the regulation of lipid metabolism were found to be negatively correlated with CG-DMR and CHG-DMR. Among the CG-DMR- or CHG-DMR-associated TFs, 15 members had been identified to co-express with lipid metabolic DEGs (Figure 7A, Table S6). Most TFs involved in the regulation of lipid metabolism, especially the co-expressed TFs, showed hypomethylation with higher abundance in 245 at the R5 stage. At the R8 stage, six and six lipid metabolic DEGs were found to be negatively correlated with CG-DMR and CHG-DMR between 245 and Z16, respectively (Figure 7B). Genes involved in β-oxidation showed hypermethylation with lower expression in 245 at the R8 stage. Reversely, genes involved in FA elongation, phospholipid synthesis, and TAG synthesis pathways showed hypermethylation with lower abundance in 245. Meanwhile, 17 TFs known to regulate lipid metabolism exhibited a negative correlation with DMR in the CG or CHG context at the R8 stage. Among them, three members were identified to co-express with lipid metabolic DEGs (Figure 7B, Table S6). Two DEGs showed hypermethylation with lower abundance, while another one showed hypomethylation with higher expression in 245 at the R8 stage. *Araip.H6S1B* encoding peroxisomal acyl transporter protein (*CTS*) was an example to exhibit a relationship between DMRs and lipid metabolic DEGs. The abundance of *CTS* was lower in 245 than that in Z16 at the two stages. Correspondingly, hyper-DMRs were observed at 5' Untranslated Region (UTR) and upstream of *CTS* in 245 compared with Z16 (Figure 7C).

To further investigate whether altering the expression level of *CTS* would influence oil accumulation, transgenic analysis in

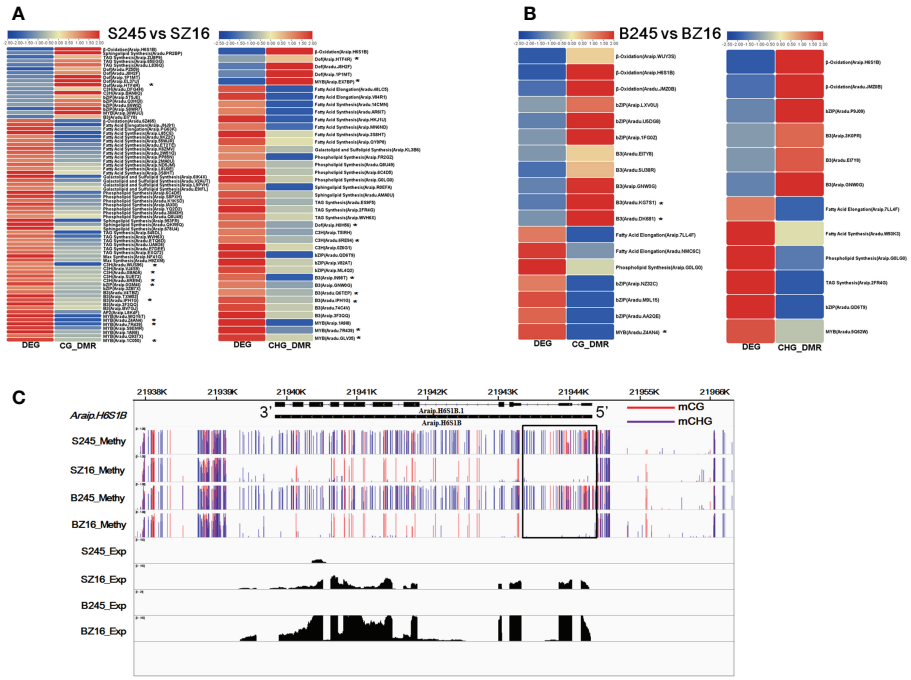


FIGURE 7
CG-DMR or CHG-DMR and their association with the expression of lipid metabolism-related genes. **(A, B)** Heatmaps showing CG-DMR and CHG-DMR between 245 and Z16 associated with differential expression at the R5 **(A)** and R8 **(B)** stages for sets of enzymes and TFs involved in lipid metabolism, respectively. * denote the CG-DMR- or CHG-DMR-associated TFs which were identified to co-express with lipid metabolic DEGs. Scales represent differential expression and differential methylation in log2 fold change. CG-DMR and CHG-DMR indicated differentially methylated regions in CG and CHG context, respectively. **(C)** A visualization showing the methylation level and expression level of acyl transporter (*Araip.H6S1B*) in seeds of 245 and Z16 at the R5 and R8 stages. S245 and SZ16 represented seeds of 245 and Z16 at the R5 stage, respectively. B245 and BZ16 denoted seeds of 245 and Z16 at the R8 stage, respectively.

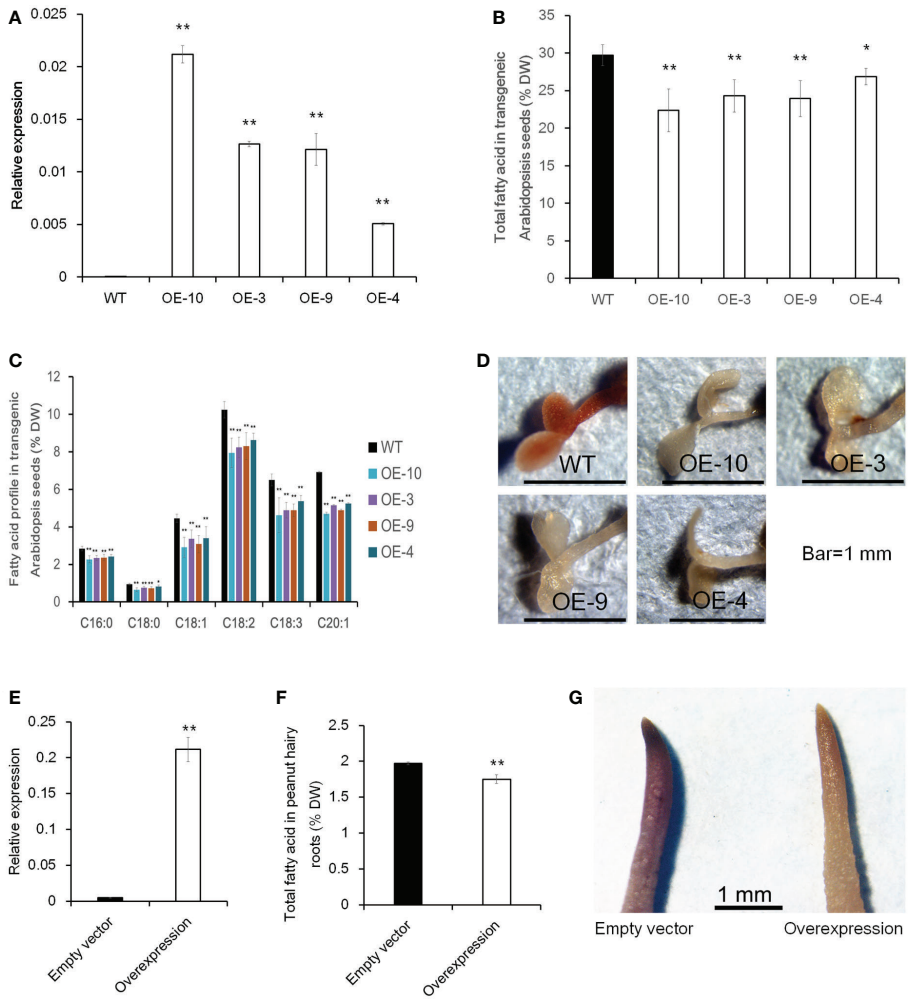


FIGURE 8
The effect of *CTS* overexpression on oil content in transgenic *Arabidopsis* and peanut hairy roots. **(A)** Expression level of *CTS* in Col-0 and transgenic *Arabidopsis* lines. Error bars indicated SD (n = 3). *AtACTIN7* (AT5G09810) was used as the internal reference gene. **(B)** Total fatty acid content of mature seeds in Col-0 and transgenic *Arabidopsis* plants. Error bars indicated SD (n = 4). **(C)** Fatty acid profile in Col-0 and transgenic *Arabidopsis* seeds. Error bars indicated SD (n = 4). The x-axis represented the species of fatty acid. **(D)** Fat red 7B staining of Col-0 and transgenic *Arabidopsis* plants. **(E)** Expression level of *CTS* in peanut transgenic hairy roots. Error bars indicated SD (n = 4). *GAPDH* (Morgante et al., 2011) was used as the internal reference gene. **(F)** Total fatty acid content in peanut transgenic hairy roots. Error bars indicated SD (n = 4). **(G)** Fat red 7B staining of peanut hairy roots. Student's *t*-test was used for significance analysis. Asterisks indicated significant differences between transgenic plants with WT plants (empty vector) under the same treatment (* $P < 0.05$ and ** $P < 0.01$). *CTS* indicated acyl transporter (*Araip.H6S1B*).

Arabidopsis and peanut hairy root was performed. The ORF of *CTS* was overexpressed using the CaMV 35S promoter in *Arabidopsis*, and 10 homozygous transgenic lines were obtained (Figure S4, Figure 8A). Then, four lines with a relatively higher expression level were selected to investigate total FA content in mature seed. In comparison with wild-type plants (Col-0), the overexpressing lines had a significantly lower total FA content and showed a decrease of 24.7% in OE-10, 18.3% in OE-3, 19.4% in OE-9, and 9.6% in OE-4 (Figure 8B). FA composition was also analyzed, and the examined FA species exhibited a significantly lower level in transgenic lines than those in Col-0 (Figure 8C). Five-day-old *Arabidopsis* seedlings were further stained with Sudan red 7B; less red color in overexpressing plants than Col-0 was observed (Figure 8D). In peanut, we used a transgenic hairy root system to test whether increasing *CTS* transcript abundance would decrease FA content in hairy roots. The expression level of *CTS* was much higher in overexpressing transgenic hairy roots than that in the control hairy roots (empty vector). Meanwhile, the total FA content of transgenic hairy roots was significantly lower than that of control roots (Figures 8E, F). The stain assay also revealed a much lighter red color in overexpressing transgenic roots than in the roots containing empty vector (Figure 8G). These results indicated that overexpression of *CTS* would decrease oil content.

Discussion

Cultivated peanut (*A. hypogaea*) was domesticated from the wild tetraploid *A. monticola*, originated by hybridization between two diploid *A. duranensis* and *A. ipaensis* (Bertioli et al., 2011; Moretzsohn et al., 2013; Yin et al., 2020). Compared to cultivated species, wild peanut possesses high genetic diversity and several superior features including a high oil content in seed (Huang et al., 2012). Understanding lipid accumulation in *A. monticola* at the level of regulation of gene expression would contribute to high-oil improvement in peanut breeding. In this study, we used the RNA-seq approach to analyze transcriptome divergence between wild and cultivated peanut at two seed developing stages (R5 and R8) and investigated the molecular mechanism underlying the difference in seed oil content. More than 70% of the reference genes (56,236) were identified and profiled in at least one sample. According to the result of PCA for expression pattern, the developmental stage prior to species was the first principal component to distinguish the samples. However, the expression profiles at the R8 stage could be obviously divided into two species in the hierarchical clustering diagram, indicating that gene expression profiles in the developing seed (R8 stage) have been altered during peanut domestication. Available literature suggests that gene expression divergence is essential to drive phenotypic variation during domestication (Lin et al., 2012; Koenig et al., 2013; Lu et al., 2016). Interestingly, we detected two "divergent" gene clusters (C2 and C7) in peanuts enriched in lipid metabolism (Figure 2). Our results suggest that there may be the existence of divergence at the gene expression level between wild and cultivated peanut contributing to the difference in lipid accumulation.

The lipid metabolism-related DEGs were presumed to play key roles in oil accumulation. Based on the expression pattern of lipid metabolic DEGs, samples could be clearly clustered into two groups (wild and cultivated) (Figure 3). Most DEGs in the lipid biosynthesis pathways were upregulated in wild species, especially at the lipid initial accumulation stage (R5) (Figure 3). Meanwhile, β-oxidation-related DEGs, which are involved in the degradation of lipid, were downregulated in wild peanut from lipid initiation to lipid rapid increase stages (R5 and R8). The DEG analysis revealed a robust activity of oil biosynthesis with a constraint of lipid degradation in developing seed of wild peanuts. It is worth noting that 16 lipid metabolic DEGs between wild and cultivated peanut were repeatedly identified at both R5 and R8 stages. These consistently divergent genes during seed development may contribute to a part of the difference in oil accumulation between wild and cultivated species. For example, *Aradu.CP1HR* encoded a biotin carboxyl carrier protein, which is a subunit of acetyl-CoA carboxylase (ACC). ACC catalyzes malonyl-CoA and bicarbonate to yield malonyl-CoA, which is the first committed step in FA synthesis (Li-Beisson, 2013). The expression level of *Aradu.CP1HR* was upregulated in wild species. In contrast, the expression abundance of another gene (*Araip.H6S1B*) was lower in wild species. It was an acyl transporter protein involved in the import of β-oxidation substrate (FAs) to peroxisome to break down FAs (Li-Beisson, 2013). In addition, many TFs known in the regulatory circuitry during seed development (Agarwal et al., 2011; Li and Li, 2016) were identified to co-express with lipid metabolic DEGs (Figure 4). The TF families, such as AP2, B3, b-ZIP, and C3H, whose members were grouped into "divergent" gene cluster C2 with different abundance between wild and cultivated peanuts (Figure 4, Figure S2, Table S6), have been well known in the regulation of oil production (Pouvreau et al., 2011; Song et al., 2013; Kim et al., 2015; Manan et al., 2017; Lu et al., 2021). Meanwhile, approximately 44% of lipid metabolic-related DEGs were also categorized into gene cluster C2. Since a set of genes with a similar expression pattern was likely to take part in the same biological process, we deduced that the co-expressed network of TFs and lipid metabolic DEGs may construct transcriptional modules affecting differential oil accumulation between wild and cultivated peanuts. Altogether, TFs may coordinate the expression abundance of lipid metabolic genes to promote oil biosynthesis in wild species. It might explain the higher oil content in wild peanuts than in cultivated peanuts.

DNA methylation is a conserved epigenetic marker that regulates gene expression. There are examples of natural epialleles in several crops showing varied DNA methylation affecting multiple biological processes (Manning et al., 2006; Zhang et al., 2015; Song et al., 2017). Here, we sought to investigate the role of DNA methylation on genes involved in lipid accumulation. The genome-wide DNA methylome and trancriptome for developing seeds of 245 (wild species) and Z16 (cultivated species) were conjointly profiled (Figure 5). A significantly negative correlation on chromosome-scale between DNA methylation and gene expression was observed, indicating that DNA methylation generally inhibits gene transcription in developing peanut seeds (R5 and R8). Meanwhile, DMRs between 245 and Z16 were displayed on 20 reference chromosomes (Figures 6A, B), providing a first glimpse of the epigenetic changes in seed development during peanut domestication. DMRs in CG and CHG, but not in the CHH context, were positively correlated with gene density on chromosomes and tended to enrich in genic regions (gene body +2k upstream) of DEGs (Figure 6D). Previous studies have shown that DNA methylation occurring in the promoter or within the transcribed gene body would regulate gene transcription (Zhang et al., 2006; Wang et al., 2015; Bewick and Schmitz, 2017; Lang et al., 2017; Huang et al., 2019). Therefore, the DMRs, especially in CG and CHG contexts, would play a role in the differential expression of genes in the present study. There were 50% and 25% of DEGs between 245 and Z16 that were associated with CG-DMRs and CHG-DMRs, respectively. GO terms of CG-DMR- and CHG-DMR-associated

DEGs were enriched in several biological processes, including lipid metabolism. There was an obvious trend that most differentially expressed enzymes and co-expressed TFs involved in lipid production were hypomethylated in wild peanut (245) at the R5 stage. In contrast, DEGs involved in β-oxidation were found hypermethylated in 245 at the R8 stage. Acyl transporter protein (*CTS*, *Araip.H6S1B*) that takes part in the β-oxidation process to degrade lipid was an example to demonstrate that the influence of DNA methylation on the expression of lipid-related genes varied between wild (245) and cultivated peanut (Z16, Figure 7C). Hypomethylated promoter and 5' UTR of *CTS* (*Araip.H6S1B*) with increased expression abundance were observed in low-oil content peanut (Z16), and the transgenic assay further confirmed that enhancing the expression of *CTS* (*Araip.H6S1B*) would reduce oil content in *Arabidopsis* seeds and peanut hairy root (Figure 8). It was suggested that DNA methylation may play an important role in oil accumulation through regulating the expression of specific genes including *CTS*. In addition, 20 TFs including AP2 (*Aradu.C87QH*) were predicted to bind the promoter (2 kb upstream) of the *CTS* gene (Figure S5). The abundance of most putative binding TFs was relatively higher in wild peanut (245) than that in cultivated peanut (Z16). In contrast, their target gene (*CTS*) was expressed more in cultivated peanut (Z16), suggesting that AP2 and other TFs may bind promote *CTS* to suppress its transcription. Altogether, the methylome and transcriptome data depicted a possible regulatory network in which DNA methylation and TFs coregulate the expression of lipid metabolism genes in peanut seed (Figure 9). In wild peanut, comprehensive alteration of gene transcription by TFs and DNA methylation would promote oil production and constrain oil degradation simultaneously, which finally contribute to higher oil accumulation.

In summary, transcriptomic and methylomic comparisons revealed gene expression and DNA methylation variations in seed development between wild and cultivated peanuts. In wild

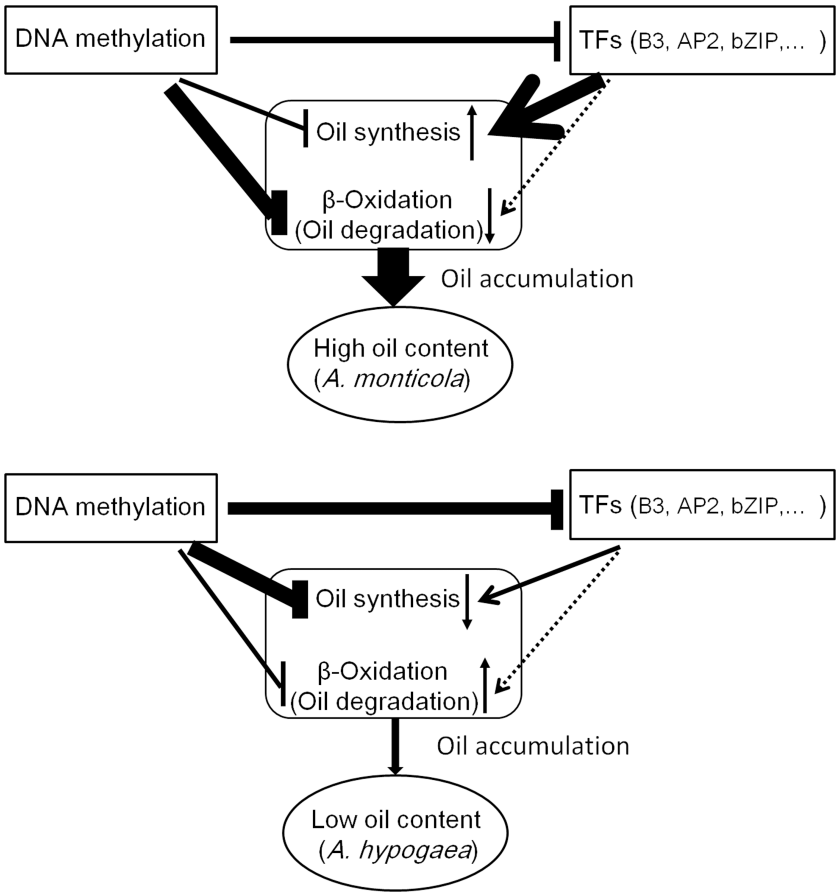


FIGURE 9
A regulatory model of lipid accumulation in peanut seed. Compared to *A. hypogaea*, *A. monticola* activated the transcription of genes involved in oil synthesis and depressed the transcription of genes involved in β-oxidation *via* DNA methylation and TFs in seed. Thus, DNA methylation and TFs coordinately tuning the expression of genes contribute to high oil accumulation in *A. monticola*.

peanut seed, TFs and DNA methylation may coordinately regulate specific lipid metabolic genes to activate oil biosynthesis and simultaneously constrain lipid degradation. Thus, gene expression change would contribute to increasing oil content in wild peanut. Our study reveals a regulation mechanism of oil accumulation in seed and provides gene resources for oil improvement in cultivated peanut.

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/, PRJNA781013, https://www.ncbi.nlm.nih.gov/, PRJNA782686.

Author contributions

NL, MP, RV, and HJ conceived, designed, and supervised the experiments. LH, YC, XZ, and JG managed materials planted in the experimental nursery. NL, BW, HL, WC, and DH conducted sampling and phenotyping. NL, BW, and BY performed DNA and RNA extraction. NL and BW performed bioinformatic analysis and interpreted the results. NL and BW prepared the first draft, and NL, MP, HC, YL, BL, RV, and HJ contributed to the final editing of manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31801403, No. 31871666, and No. 31761143005), the National Peanut Industry Technology System Construction (CARS-13), National Crop Germplasm Resources Center (NCGRC-2021-036), National Program for Crop Germplasm Protection of China (19210163), Central Public-interest Scientific Institution Basal Research Fund (No.1610172019003), and the Science and technology innovation Project of Chinese Academy of Agricultural Sciences.

Acknowledgments

The authors thank Dr. Xin Wang from OCRI of CAAS for kindly providing technical guidance in genetic transform of *Arabidopsis* (Col-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.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1065267/full#supplementary-material

References

Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *BioEssays* 33, 189–202. doi: 10.1002/bies.201000107

Baubec, T., Finke, A., Mittelsten Scheid, O., and Pecinka, A. (2014). Meristem-specific expression of epigenetic regulators safeguards transposon silencing in arabidopsis. *EMBO Rep.* 15, 446–452. doi: 10.1002/embr.201337915

Beisson, F. (2003). Arabidopsis genes involved in acyl lipid metabolism. a 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.* 132, 681–697. doi: 10.1104/pp.103.022988

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of arachis duranensis and arachis ipaensis, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517

Bertioli, D. J., Seijo, G., Freitas, F. O., Valls, J. F. M., Leal-Bertioli, S. C. M., and Moretzsohn, M. C. (2011). An overview of peanut and its wild relatives. *Plant Genet. Resour.* 9, 134–149. doi: 10.1017/S1479262110000444

Bewick, A. J., and Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* 36, 103–110. doi: 10.1016/j.pbi.2016.12.007

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

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

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745

FAOSTAT (2019). Available at: https://www.fao.org/faostat/en/#data/QCL.

Gao, X., Chen, J., Dai, X., Zhang, D., and Zhao, Y. (2016). An effective strategy for reliably isolating heritable and Cas9-free arabidopsis mutants generated by

CRISPR/Cas9-mediated genome editing. *Plant Physiol.* 171, 1794. doi: 10.1104/pp.16.00663

Gutschker, S., Corral, J. M., Schmiedl, A., Ludewig, F., Koch, W., Fiedler-Wiechers, K., et al. (2022). Multi-omics data integration reveals link between epigenetic modifications and gene expression in sugar beet (Beta vulgaris subsp. vulgaris) in response to cold. *BMC Genomics* 23, 144. doi: 10.1186/s12864-022-08312-2

Hewezi, T., Lane, T., Piya, S., Rambani, A., Rice, J. H., and Staton, M. (2017). Cyst nematode parasitism induces dynamic changes in the root epigenome. *Plant Physiol.* 174, 405–420. doi: 10.1104/pp.16.01948

Huang, L., Jiang, H. F., Ren, X. P., Chen, Y. N., Xiao, Y. J., Zhao, X. Y., et al. (2012). Abundant microsatellite diversity and oil content in wild arachis species. *PloS One* 7, e50002. doi: 10.1371/journal.pone.0050002

Huang, H., Liu, R. E., Niu, Q. F., Tang, K., Zhang, B., Zhang, H., et al. (2019). Global increase in DNA methylation during orange fruit development and ripening. *Proc. Natl. Acad. Sci. United States America* 116, 1430–1436. doi: 10.1073/pnas.1815441116

Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhajarvi, T., Chia, J. M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309

Ichihashi, Y., Aguilarmartínez, J. A., Farhi, M., Chitwood, D. H., Kumar, R., Millon, L. V., et al. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. United States America* 111, 2616–2621. doi: 10.1073/pnas.1402835111

Kim, H. U., Lee, K. R., Jung, S. J., Shin, H. A., Go, Y. S., Suh, M. C., et al. (2015). Senescence-inducible LEC2 enhances triacylglycerol accumulation in leaves without negatively affecting plant growth. *Plant Biotechnol. J.* 13, 1346–1359. doi: 10.1111/pbi.12354

Koenig, D., Jimenez-Gomez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. United States America* 110, E2655–E2662. doi: 10.1073/pnas.1309606110

Langfelder, P., and Horvath, S. (2008). WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., et al. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc. Natl. Acad. Sci. United States America* 114, E4511–E4519. doi: 10.1073/pnas.1705233114

Li-Beisson, Y., Shorrosh, B. , Beisson, F., Andersson, M., Arondel, V., Bates, P., et al. (2013). *Acyl-lipid metabolism. arabidopsis book*, Vol. 11. Rockville: American Society of Plant Biologists.

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008

Lin, Z. W., Li, X. R., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G. H., et al. (2012). Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* 44, 720–U154. doi: 10.1038/ng.2281

Liu, N., Guo, J., Zhou, X., Wu, B., Huang, L., Luo, H., et al. (2020). High-resolution mapping of a major and consensus quantitative trait locus for oil content to a ~ 0.8-Mb region on chromosome A08 in peanut (Arachis hypogaea l.). *Theor. Appl. Genet.* 133, 37–49. doi: 10.1007/s00122-019-03438-6

Lu, X., Li, Q.-T., Xiong, Q., Li, W., Bi, Y.-D., Lai, Y.-C., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* 86, 530–544. doi: 10.1111/tpj.13181

Lu, L., Wei, W., Li, Q. T., Bian, X. H., Lu, X., Hu, Y., et al. (2021). A transcriptional regulatory module controls lipid accumulation in soybean. *New Phytol.* 231, 661–678. doi: 10.1111/nph.17401

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

Manning, K., Tor, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38, 948–952. doi: 10.1038/ng1841

Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C., Valls, J. F., and Bertioli, D. J. (2013). A study of the relationships of cultivated peanut (Arachis hypogaea) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* 111, 113–126. doi: 10.1093/aob/mcs237

Morgante, C. V., Guimarães, P. M., Martins, A. C., Araújo, A. C., Leal-Bertioli, S. C., Bertioli, D. J., et al. (2011). Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. *BMC Res. Notes* 4, 1756–0500. doi: 10.1186/1756-0500-4-339

Narsai, R., Secco, D., Schultz, M. D., Ecker, J. R., Lister, R., and Whelan, J. (2017). Dynamic and rapid changes in the transcriptome and epigenome during germination and in developing rice (Oryza sativa) coleoptiles under anoxia and re-oxygenation. *Plant J.* 89, 805–824. doi: 10.1111/tpj.13418

Pandey, M. K., Pandey, A. K., Kumar, R., Nwosu, C. V., Guo, B., Wright, G. C., et al. (2020). Translational genomics for achieving higher genetic gains in groundnut. *Theor. Appl. Genet.* 133, 1679–1702. doi: 10.1007/s00122-020-03592-2

Pattee, H. E., Johns, E. B., Singleton, J. A., and Sanders, T. H. (1974). Composition changes of peanut fruit parts during maturation. *Peanut Sci.* 1, 57–62. doi: 10.3146/i0095-3679-1-2-6

Pouvreau, B., Baud, S., Vernoud, V., Morin, V., Py, C., Gendrot, G., et al. (2011). Duplicate maize Wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis. *Plant Physiol.* 156, 674–686. doi: 10.1104/pp.111.173641

Qi, X., Li, M. W., Xie, M., Liu, X., Ni, M., Shao, G., et al. (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat. Commun.* 5, 4340. doi: 10.1038/ncomms5340

Rajkumar, M. S., Gupta, K., Khemka, N. K., Garg, R., and Jain, M. (2020). DNA Methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. *Commun. Biol.* 3, 340. doi: 10.1038/s42003-020-1059-1

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374–378. doi: 10.2144/03342mt01

Sang, T., and Ge, S. (2013). Understanding rice domestication and implications for cultivar improvement. *Curr. Opin. Plant Biol.* 16, 139–146. doi: 10.1016/j.pbi.2013.03.003

Shen, Y., Zhang, J., Liu, Y., Liu, S., Liu, Z., Duan, Z., et al. (2018). DNA Methylation footprints during soybean domestication and improvement. *Genome Biol.* 19, 128. doi: 10.1186/s13059-018-1516-z

Song, Q.-X., Li, Q.-T., Liu, Y.-F., Zhang, F.-X., Ma, B., Zhang, W.-K., et al. (2013). Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic arabidopsis plants. *J. Exp. Bot.* 64, 4329–4341. doi: 10.1093/jxb/ert238

Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18, 99. doi: 10.1186/s13059-017-1229-8

Tian, J., Wang, C., Xia, J., Wu, L., Xu, G., Wu, W., et al. (2019). Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* 365, 658–664. doi: 10.1126/science.aax5482

Wang, M., Wang, P., Tu, L., Zhu, S., Zhang, L., Li, Z., et al. (2016). Multi-omics maps of cotton fibre reveal epigenetic basis for staged single-cell differentiation. *Nucleic Acids Res.* 44, 4067–4079. doi: 10.1093/nar/gkw238

Wang, H., Beyene, G., Zhai, J., Feng, S., Fahlgren, N., Taylor, N. J., et al. (2015). CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proc. Natl. Acad. Sci. United States America* 112, 13729–13734. doi: 10.1073/pnas.1519067112

Wang, X., Xu, P., Yin, L., Ren, Y., Li, S., Shi, Y., et al. (2018). Genomic and transcriptomic analysis identified gene clusters and candidate genes for oil content in peanut (Arachis hypogaea l.). *Plant Mol. Biol. Rep* 36, 518–529. doi: 10.1007/s11105-018-1088-9

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinf.* 10, 232. doi: 10.1186/1471-2105-10-232

Xing, M., Zhang, Y., Zhou, S., Hu, W., Wu, X., Ye, Y., et al. (2015). Global analysis reveals the crucial roles of DNA methylation during rice seed development. *Plant Physiol.* 168, 1417–1554. doi: 10.1104/pp.15.00414

Xu, J., Chen, G., Hermanson, P. J., Xu, Q., Sun, C., Chen, W., et al. (2019). Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. *Genome Biol.* 20, 243. doi: 10.1186/s13059-019-1859-0

Yin, D., Ji, C., Song, Q., Zhang, W., Zhang, X., Zhao, K., et al. (2020). Comparison of arachis monticola with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of peanut. *Adv. Sci. (Weinh)* 7, 1901672. doi: 10.1002/advs.201901672

Yong-Villalobos, L., Gonzalez-Morales, S. I., Wrobel, K., Gutierrez-Alanis, D., Cervantes-Perez, S. A., Hayano-Kanashiro, C., et al. (2015). Methylome analysis reveals an important role for epigenetic changes in the regulation of the arabidopsis response to phosphate starvation. *Proc. Natl. Acad. Sci. United States America* 112, E7293–E7302. doi: 10.1073/pnas.1522301112

Yoo, M.-J., and Wendel, J. F. (2014). Comparative evolutionary and developmental dynamics of the cotton (Gossypium hirsutum) fiber transcriptome. *PloS Genet.* 10, e1004073. doi: 10.1371/journal.pgen.1004073

You, J., Li, M., Li, H., Bai, Y., Zhu, X., Kong, X., et al. (2022). Integrated methylome and transcriptome analysis widen the knowledge of cytoplasmic Male sterility in cotton (Gossypium barbadense l.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.770098

Yuan, M., Zhu, J., Gong, L., He, L., Lee, C., Han, S., et al. (2019). Mutagenesis of FAD2 genes in peanut with CRISPR/Cas9 based gene editing. *BMC Biotechnol.* 19, 24. doi: 10.1186/s12896-019-0516-8

Zhang, X. R., Henriques, R., Lin, S. S., Niu, Q. W., and Chua, N. H. (2006). Agrobacterium-mediated transformation of arabidopsis thaliana using the floral dip method. *Nat. Protoc.* 1, 641–646. doi: 10.1038/nprot.2006.97

Zhang, H. M., Lang, Z. B., and Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z

Zhang, X., Sun, J., Cao, X., and Song, X. (2015). Epigenetic mutation of RAV6 affects leaf angle and seed size in rice. *Plant Physiol.* 169, 2118–2128. doi: 10.1104/pp.15.00836

Zhang, H., Wang, M. L., Dang, P., Jiang, T., Zhao, S. Z., Lamb, M., et al. (2021). Identification of potential QTLs and genes associated with seed composition traits in peanut (Arachis hypogaea l.) using GWAS and RNA-seq analysis. *Gene* 769, 145215. doi: 10.1016/j.gene.2020.145215.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189–1201. doi: 10.1016/j.cell.2006.08.003

Zhong, S., Fei, Z., Chen, Y. R., and Zheng, Y. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat. Biotechnol.* 31, 154–159. doi: 10.1038/nbt.2462

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 January 2023
DOI 10.3389/fpls.2022.1076744

OPEN ACCESS

EDITED BY
Weijian Zhuang,
Fujian Agriculture and Forestry University, China

REVIEWED BY
Diaa Abd El Moneim,
Arish University, Egypt
David Sewordor Gaikpa,
University of Hohenheim, Germany
Abdulwahab Saliu Shaibu,
Bayero University Kano, Nigeria
Muhammad Amjad Nawaz,
Far Eastern Federal University, Russia

*CORRESPONDENCE
Richard Oteng-Frimpong
kotengfrimpong@gmail.com
Benjamin Karikari
benkarikari1@gmail.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 27 October 2022
ACCEPTED 24 November 2022
PUBLISHED 05 January 2023

CITATION
Oteng-Frimpong R, Karikari B, Sie EK, Kassim YB, Puozaa DK, Rasheed MA, Fonceka D, Okello DK, Balota M, Burow M and Ozias-Akins P (2023) Multi-locus genome-wide association studies reveal genomic regions and putative candidate genes associated with leaf spot diseases in African groundnut (*Arachis hypogaea* L.) germplasm. *Front. Plant Sci.* 13:1076744. doi: 10.3389/fpls.2022.1076744

Multi-locus genome-wide association studies reveal genomic regions and putative candidate genes associated with leaf spot diseases in African groundnut (*Arachis hypogaea* L.) germplasm

Richard Oteng-Frimpong[1]*, Benjamin Karikari[2]*, Emmanuel Kofi Sie[1], Yussif Baba Kassim[1], Doris Kanvenaa Puozaa[1], Masawudu Abdul Rasheed[1], Daniel Fonceka[3], David Kallule Okello[4], Maria Balota[5], Mark Burow[6] and Peggy Ozias-Akins[7]

[1]Groundnut Improvement Program, Council for Scientific and Industrial Research (CSIR)-Savanna Agricultural Research Institute, Tamale, Ghana, [2]Department of Agricultural Biotechnology, Faculty of Agriculture, Food and Consumer Sciences, University for Development Studies, Tamale, Ghana, [3]Centre d'Etude Régional pour l'Amélioration de l'Adaptation àla Sécheresse (CERAAS), Institut Sénégalais de Recherches Agricoles (ISRA), Thiès, Senegal, [4]Oil Crops Research Program, National Semi-Arid Resources Research Institute (NaSARRI), Soroti, Uganda, [5]School of Plant and Environmental Sciences, Tidewater Agricultural Research and Extension Center (AREC), Virginia Tech, Suffolk, VA, United States, [6]Texas A&M AgriLife Research and Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, United States, [7]Institute of Plant Breeding Genetics and Genomics, University of Georgia, Tifton, GA, United States

Early leaf spot (ELS) and late leaf spot (LLS) diseases are the two most destructive groundnut diseases in Ghana resulting in ≤ 70% yield losses which is controlled largely by chemical method. To develop leaf spot resistant varieties, the present study was undertaken to identify single nucleotide polymorphism (SNP) markers and putative candidate genes underlying both ELS and LLS. In this study, six multi-locus models of genome-wide association study were conducted with the best linear unbiased predictor obtained from 294 African groundnut germplasm screened for ELS and LLS as well as image-based indices of leaf spot diseases severity in 2020 and 2021 and 8,772 high-quality SNPs from a 48 K SNP array Axiom platform. Ninety-seven SNPs associated with ELS, LLS and five image-based indices across the chromosomes in the 2 two sub-genomes. From these, twenty-nine unique SNPs were detected by at least two models for one or more traits across 16 chromosomes with explained phenotypic variation ranging from 0.01 - 62.76%, with exception of chromosome (Chr) 08 (Chr08), Chr10, Chr11, and Chr19. Seventeen potential candidate genes were predicted at ± 300 kbp of the stable/prominent SNP positions (12 and 5, down- and

upstream, respectively). The results from this study provide a basis for understanding the genetic architecture of ELS and LLS diseases in African groundnut germplasm, and the associated SNPs and predicted candidate genes would be valuable for breeding leaf spot diseases resistant varieties upon further validation.

KEYWORDS

candidate genes, environmentally friendly, genomics, markerassisted selection, oilseed, early leaf spot, late leaf spot

Introduction

Groundnut [*Arachis hypogaea* L. (2n = 4x = 40)] is an allotetraploid originating from South America and is a grain legume that is largely grown in the tropical and subtropical regions of the world (Bertioli et al., 2016). It comprises of two genomes with their origins coming from different diploid wild ancestors (Bertioli et al., 2016). Groundnut plays a pivotal role in the life of small-holder farmers in Ghana, and it is a suitable vehicle for making improvements in the areas such as poverty alleviation, food and nutritional security (Tyroler, 2018). It is a good source of plant protein for resource poor-households who are unable to buy animal protein. Groundnut also provides vitamins, minerals and unsaturated oil for most Ghanaians (Asibuo et al., 2008). Increase production and consumption of groundnut will reduce the number of the over 800 million people in many developing countries who are chronically hungry as well as the about 2 billion people who suffer micronutrients deficiencies (FAO et al., 2019). Globally, it provided 8, 3 and 2% of oilseed produced, vegetable oil and protein meal consumed, respectively (http://soystats.com/, accessed 03.07.2022).

Despite the importance of groundnut, its cultivation is hindered by numerous biotic and abiotic factors. Early leaf spot (ELS) caused by the fungus *Passalora arachidicola* (S. Hori) and late leaf spot (LLS) also caused by the fungus *Nothopassalora personata* (Berk. & Curt.) are the two most destructive groundnut diseases in Ghana with a potential yield loss of ≤ 70% (Naab et al., 2005; Denwar et al., 2021). Chemical control of these diseases is not feasible in farmers field in Ghana due to their inability to afford these chemicals and therefore ends up not controlling the diseases (Nutsugah et al., 2007; Denwar et al., 2021). Farmers often confuse leaf spots severities with maturity indicator further affecting their mitigation measures. Development and cultivation of leaf spot resistant varieties is cheaper and environmentally friendly (Gaikpa et al., 2017). Resistance is largely controlled by polygenes and influenced by genotype (G), environment (E) and their interactions (G×E) (Johnson, 1984; Wiesner-Hanks and Nelson, 2016). This makes the traditionally commonly used screening methods for identifying leaf spot resistant varieties in Ghana difficult due to the partial and polygenic nature of these diseases (Dwivedi et al., 2002; Pasupuleti et al., 2013). In addition, conventional phenotyping procedures are laborious, time-consuming, destructive, subjective, costly, inefficient and lack inter or intra-rater repeatability (Araus et al., 2018; Awada et al., 2018).

To overcome errors and expenses by manual phenotyping, the red-green-blue (RGB)-image method (which is the science of making measurements through the use of an RGB camera), together with conventional and marker-assisted selection (MAS) may overcome the flaws of current breeding methods (Li et al., 2014; Wang et al., 2017; Yang et al., 2020; Sarkar et al., 2021a). The application of the RGB image method for screening has generated much interest in agricultural research because of its importance in crop production (Sie et al., 2022; Chapu et al., 2022). RGB image method is more efficient, offers inter or intra-rater repeatability, is easy to apply, is less expensive, non-destructive, and offers the chance to take multiple measurements on a specific plant due to the non-destructive nature of the technology (Araus et al., 2018; Awada et al., 2018; Gill et al., 2022). Moreover, the application of the RGB image method in phenotyping will allow the screening of a large set of genotypes using a small fraction of the time that would have been used in conventional phenotyping. For instance, previous studies have shown the efficacy of the RGB imaging for assessment of a number of diseases in several crops verticillium wilt (caused by *Verticillium dahliae* Kleb) in olive (*Olea europaea L.*) (Sancho-Adamson et al., 2019), yellow rust (*Puccinia striiformis* f. sp tritici) in wheat (Zaman-Allah et al., 2015; Zhou et al., 2015), and lethal necrosis (caused by a combination of maize chlorotic mottle virus (MCMV) and sugar cane mosaic virus (SCMV) in maize (*Zea mays* L.)) (Kefauver et al., 2015).

In the last decade, there has been rapid development of next-generation sequencing, high-throughput genotype data together with phenotypic data for utilization to identify marker-trait associations *via* genome-wide association study (GWAS) (Varshney et al., 2019; Varshney et al., 2020). Compared to linkage mapping, GWAS has emerged as a powerful tool to detect markers (single nucleotide polymorphisms (SNPs)) closely linked to quantitative trait loci (QTL), based on the

principle of linkage disequilibrium (LD) between genetic markers and QTL that affect the trait (Geng et al., 2015). By this strategy, Zhang et al. (2020a) detected a total of 46 and 28 QTL for ELS and LLS, respectively, with Efficient Mixed-Model Association eXpedited (EMMAX), while Pandey et al. (2014) detected 6 QTL for ELS and 1 QTL for LLS. In contrast, seven and five major QTL for ELS and LLS have been detected and reported on chromosomes 2 and 3, respectively (Chu et al., 2019). In addition, Shaibu et al. (2021) detected 25 SNPs for ELS in mini-core collection of 168 accessions in Nigeria by efficient mixed-model association (EMMA) and compression mixed linear model (CMLM).

The statistical model adopted is one of the setbacks to the power of detection in GWAS (Gupta et al., 2014; Ibrahim et al., 2020; Yoosefzadeh-Najafabadi et al., 2022). Traditional popular statistical models (single-marker genome-wide scan models), mixed linear model (MLM), and general linear model (GLM), among others, have a number of limitations such as the stringent threshold of significance and mapping power (Wen et al., 2018). To overcome these limitations, several multi-locus models have been developed and utilized for GWAS in several crops (Zhang et al., 2019; Karikari et al., 2020; Berhe et al., 2021; Vikas et al., 2022). Among them include a multi-locus random-SNP-effect mixed linear model (mrMLM) (Wang et al., 2016), a fast mrMLM (FASTmrMLM) (Zhang et al., 2018), a fast mrMLM efficient mixed-model association (FASTmrEMMA) (Wen et al., 2018), polygene-background-control-based least-angle regression plus empirical Bayes (pLARmEB) (Zhang et al., 2017), Kruskal-Wallis test with empirical Bayes under polygenic background control (pKWmEB) (Ren et al., 2018) and integrative sure independence screening expectation maximization Bayesian least absolute shrinkage and selection operator model (ISIS EM-BLASSO) (Tamba et al., 2017).

The multi-locus models have become the state-of-the-art procedure to identify genetic bases for complex traits due to their power of detection and robustness (Zhang et al., 2019). Therefore, the present study applied the six multi-locus models to identify genomic regions and potential candidates associated with ELS and LLS diseases. A total of 294 groundnut accessions were collected from different African countries and screened in two years with manual scoring of ELS and LLS together with 5 imaged-based indices and area under disease progression curve (AUDPC) for both diseases. The results from this study lay the foundation for MAS to speed up breeding for leaf spot-resistant cultivars.

Materials and methods

Planting materials and experimental condition

Two hundred and ninety-four African groundnut collections (Supplementary Table S1) were planted during the main planting season, from June 2020 to September 2020 and June 2021 to September 2021 at the experimental site (09° 25′ 41″ N, 00° 58′ 42″ W) of Council for Scientific and Industrial Research-Savannah Agricultural Research Institute (CSIR-SARI) located in Nyankpala, Northern region, Ghana. This population comprised 54, 49, 44, 32, 31, 27, 22, 18, and 17 accessions from Uganda, Ghana, Niger, Malawi, Senegal, Mali, Mozambique, Togo, and Zambia, respectively (Supplementary Table S1), mainly from African Groundnut Germplasm Collection leaf spots resistant and yield phenotyping programs. This panel was selected for the current study to lay foundation for future molecular breeding. The experimental area is characterized by a relatively dry climate with unimodal rainfall ranging between 500 and 1200 mm annually (Atiah et al., 2019; Atiah et al., 2020). The inception of the rains is in May and ends in October with small scattered precipitations in November. The soils of the research area belong to Ferric Luvisols of the Tingoli series with a brown color, moderately drained, and free from concretions (Atakora and Kwakye, 2016). The experiment was carried out in a location that is a hotspot for the disease and therefore can sufficiently discriminate between susceptible and resistant lines (Danful et al., 2019; Oteng-Frimpong et al., 2021)

The accessions were arranged in lattice design with three replicates. A plot was made up of one row of 2 m long with a spacing of 0.5 m between rows and 0.2 m between plants. One seed was planted per hill. Weeding was carried out whenever necessary to ensure a weed-free trial.

Agronomic practices

Pre-emergence weed control was done by spraying (Alligator® 400EC, Pendimethaline 400g/L, EC) and glyphosate (480g/L SL) at 200 ml/15 liters of water immediately after planting. Weeds were manually controlled regularly by hoeing between the rows and pulling weeds within rows as well as on top of plots using hands to ensure a weed-free experiment. Earthen-up was done 40 days after planting to enhance aeration. A compound fertilizer made of nitrogen(N), phosphorus (P), potassium (K) together with sulphur (S), zinc (Zn) and boron (B), i.e., (N:P:K: 11:22:21+5S+0.7Zn+0.5B) was applied on the sides of the plants two weeks after seedling emergence at a rate of 150 kg/ha. At the same time, the experiment was sprayed against aphis infestation using *K*-Optimal (Lambda-cyhalothrin 15 g/l + acetamiprid 20 g/l; EC) at 40 ml in a 15L Knapsack sprayer.

Collection of phenotypic data

Visual scoring for leaf spot disease

Visual scoring for the severity of ELS and LLS infection was evaluated using the scale described by Subrahmanyam et al.

(1995) at 70 and 90 days after planting (Sie et al., 2022). Values of 1 to 4 indicate increasing leaf spot incidence on leaflets within the lower or upper canopy, but no defoliation. Ratings from 4 to 10 are associated with increasing levels of severity with defoliation (Chiteka et al., 1988). The average score of the two-sampling time was computed. AUDPC values were computed for each plot from these disease ratings using the formula: AUDPC = $\sum_{i=1}^{a}[\{\frac{y_i + y_{i+1}}{2}\}x(t_{i+1} - t_i)]$, where y_i is the level of disease severity score at a point in time, $_{t(i+1)}$-t_i is the number of days between two successive scores (Shaner and Finney, 1977).

Measurement of normalized difference vegetation index

GreenSeeker® handheld sensor: (Model HCS-100 manufactured by Trimble Navigation Limited, Sunnyvale, USA) was used to measure the canopy normalized difference vegetation index (NDVI) of the vegetation from each plot. The instrument was aligned horizontally and maintained at a constant height of 50 cm over the plants' canopy with a walking speed to cover the row within 60 seconds. The GreenSeeker optical sensor uses radiation of 650 ± 10 of red and 770 ± 15 of near-infrared band independently. The sensor uses built-in software to directly calculate the NDVI value using the formula: (NIR-RED)/(NIR+RED) (Rouse et al., 1974). The NDVI value which ranges from 0.00 to 0.99 was recorded from the screen of the device. Readings were taken at 70 and 90 days after planting when the sun was at its zenith.

RGB images

The RGB digital camera (Samsung Galaxy NX300) was used to take close-up images of one plot at a time. The camera was set to "auto" to allow the camera to adjust the required sharpness, brightness, and hue depending on the light available, with the zoom of the lens being at 0. A part representing the plot was selected for the image. The camera was maintained at the same height of 80 cm over the row for all pictures and facing the sun to avoid any shadows on the pictures. Pictures were taken at 70, and 90 days after planting. Digital image analysis was carried out in Image J software by converting hue (H), saturation (S), and brightness (B) values into the dark green color index (DGCI).

Green area, greener area, and crop senescence index

Green area (GA=H 60-120°), greener area (GGA= H 80-120°), Hue angle, and crop senescence index (CSI=(100*(GA-GGA)/GA) (Gracia-Romero et al., 2018) were extracted using Breedpix 2.0 option from the CIMMYT maize scanner 1.16 plugin (http://github.com/george-haddad/CIMMYT open software; Copyright 2015 Shawn Carlisle Kefauver, University of Barcelona); produced as part of Image J/Fiji (open source software; http://fiji.sc/Fiji) (Schindelin et al., 2012; Schindelin et al., 2015). Both GA and GGA measure the number of green pixels on an image. However, the GGA removes green tones that are yellowish from the image and, accordingly, differentiates leaf senescence and active photosynthetic biomass more accurately.

Statistical analysis of phenotypic data

Data collected manually and imaged based across the two years (2020 and 2021) were subjected to analysis of variance (ANOVA) in SAS (SAS Institute, 2010. SAS/STAT software version 9.2. SAS Institute Inc, Cary, NC) with a general linear model procedure (PROC GLM), following statistical model $y_{pqr}=\mu+G_p+E_q+GE_{pq}+R_{r(q)}+\epsilon_{pqr}$, where y_{pqr} stands for the individual observation of pqrth experiment unit, μ is the total average phenotypic value, G_p is the effect of the pth genotype, $E_1$is the effect of the qth year, GE_{pq} is the interaction effect between the pth genotype and the qth year, $R_{r(q)}$ is the effect of the rth block within the qth year, and ε_{pqr} is the residual error. All factors were considered random.

Descriptive statistics: mean, standard error of the mean, kurtosis, and skewness were calculated in SAS (SAS Institute, 2010. SAS/STAT software version 9.2. SAS Institute Inc, Cary, NC) from the two years data. Pearson correlation coefficients were computed and visualized in R with the corrplot package (Wei et al., 2017).

In addition to the above, broad-sense heritability (H^2) for each trait was computed following the formula proposed by Nyquist and Baker (1991), thus $H^2 = \sigma_g^2 /(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$ where σ_g^2 is the genotypic variance, σ_{ge}^2 is the genotype by environment interaction variance, σ_e^2 is the error variance, n is the number of environments, and r is the number of replications.

Genotyping and population structure analysis

Prior to genotyping, fresh and healthy leaf samples were collected from the panel evaluated in this study and stored at -80 °C for DNA isolation. The genomic DNA was extracted using the modified CTAB method (Porebski et al., 1997). Purified DNA was dissolved in TE buffer for further analysis. The quantity and quality of the DNA were assessed with NanoDrop™ 2000 Spectrophotometer (Thermo Scientific, Wilmington, DC, USA). The genotyping was performed using SNP array (Affymetrix 2). The SNP array used in this study was the 48 K SNP array that was developed for Arachis Axiom Arachis. Quality control was conducted following the procedure outlined by Clevenger et al. (2018) on 8,911 SNPs.

Population structure was analyzed *via* Structure software 2.3.3 (Pritchard et al., 2000) with the number of presumed population (K) set from 1-7 and replicated 5 times with a burn-in period of 50,000 steps and Monte Carlo Marko Chain

of 100,000. An admixture model with correlated allele frequency was adopted in this study (Falush et al., 2003). After analysis, the Structure Harvester online program (https://taylor0.biology.ucla.edu/structureHarvester/) was used to retrieve the optimum K (ΔK) (Earl and Vonholdt, 2012). Only the accessions with a membership coefficient (Q) ≥ 0.60 were assigned to a genetic group and those with Q<0.60 were classified as admixture (Delfini et al., 2021). We further constructed a neighbor-joining (NJ) phylogenetic tree *via* TASSEL 5.2.31 software (Bradbury et al., 2007). A kinship plot was produced with the kinship2 package in R (Sinnwell et al., 2014). LD between pairwise SNPs was computed with RTM-GWAS V1.1 software with squared allele frequency correlation model (He et al., 2017). The panel's LD decay rate was estimated as the chromosomal distance when the LD decay (r^2) fell to half of its highest value. The graph of the LD decay was produced with the help of GraphPad Prism version 5.01 (GraphPad Software, San Diego California USA) within the pairwise distance of 5 Mb in the genome.

Marker-trait association analysis

To prevent environmental (year) variation in phenotypic data, the best linear unbiased predictor (BLUPs) for each accession for all traits were calculated using R package lme4 (Bates et al., 2015) with the effect of environment (year), replicate within E, G, GE and error as random effects. Six multi-locus models, i.e., mrMLM (Wang et al., 2016), FASTmrMLM (Zhang et al., 2018), FASTmrEMMA (Wen et al., 2018), pLARmEB (Zhang et al., 2017), pKWmEB (Ren et al., 2018) and ISIS EM-BLASSO (Tamba et al., 2017) were conducted in R with mrMLM package (V4.0.2) (Zhang et al., 2020b) with both Q matrix and principal component (PC) together with kinship matrix. The threshold with a critical logarithm of odd value was set at 3.

Candidate gene prediction and *in silico* analyses

SNPs detected for at least two traits were considered stable, hence were selected for downstream analysis including candidate gene prediction from reference genome (*A. hypogaea* V1) (Bertioli et al., 2019) available on phytozome (https://phytozome-next.jgi.doe.gov/). The gff3 file was retrieved from the phytozome website (https://phytozome-next.jgi.doe.gov/) and information gene ontology (GO) (Ashburner et al., 2000), protein families (Pfam) (Bateman et al., 2004), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), and transcription factors (Jin et al., 2014)) were further explored in selecting candidate genes with the SNP ± LD.

Results

Phenotypic variation, broad-sense heritability, and correlation among the 294 accessions of groundnut

Descriptive statistics and H^2 of the four image-based traits (Hue, GA, GGA, and CSI), vegetative index trait (NDVI), manually scored traits (ELS and LLS), and quantitatively computed traits (AUDPC-ELS and -LLS) among the 294 groundnut accessions based on the two years (2020 and 2021) evaluation in this study are shown in Table. The hue, GA, GGA, CSI (from the image-based phenotyping) and NDVI values ranged (mean ± Standard error of mean) from 38.37-90.20 (68.03 ± 0.56), 0.30-0.92 (0.63 ± 0.01), 0.23-0.81 (0.48 ± 0.01), 9.27-45.90 (26.19 ± 0.45) and 0.36-7.09 (0.68 ± 0.02), respectively (Table 1). The qualitatively scoring of ELS and LLS incidence followed a normal distribution (Figures 1A, B), with the mean scores of 4.28 ± 0.03 and 4.85 ± 0.04 (Table 1), respectively. This indicates that the population used for this study exhibited a wide range of variation in response to ELS and LLS diseases. Also, the quantitatively computed AUDPC for ELS and LLS ranged from 30.00-140.00 (86.00 ± 0.60) and 40.00-140.00 (97.00 ± 0.47) (Table 1), respectively, and this followed normal distribution among the population used in this study (Figures 1C, D).

Combined ANOVA of the two years data revealed that eight traits varied significantly due to genotypes (G), environment (E), and their interaction (GE), while NDVI differed due to E (Supplementary Tables 2A–I). These coupled with high H^2 (Table 1) suggest that these ELS and LLS are largely controlled by polygenes with both major and minor effects. This highlights that selection based only on phenotypic variation may be misleading.

We further conducted a Pearson correlation analysis among the five image-based indices, ELS, LLS, AUDPC-ELS, and -LLS with Corrplot package in R (Wei et al., 2017) and the results are shown in Supplementary Figure 1. Among the five image-based indices, only CSI positively correlated with LLS (correlation coefficient, r = 0.71), AUDPC-LLS (r=0.66), ELS (r=0.56) and AUDPC-ELS (r=0.44) (Supplementary Figure 1). This indicates that CSI could be a relatively good indicator for the assessment of ELS and LLS diseases. However, the qualitatively and quantitatively scoring of ELS and LLS negatively correlated with the other four image-based indices (GA, GGA, Hue, and NDVI) (r=-0.02 to -0.82) (Supplementary Figure 1) confirming that leaf spot diseases affect the photosynthetic capacity of the leaf. These together with leaf spot diseases ratings and ANOVA highlight that the phenotypic data from the population qualify for GWAS mapping.

TABLE 1 Descriptive statistics and broad-sense heritability of nine traits.

Parameter [a]	Mean ± SEM [b]	Range	CV (%) [c]	Skewness	Kurtosis	H^2 (%) [d]
Hue*	68.03 ± 0.56	38.37-90.20	13.85	0.16	-0.26	90.09
GA*	0.63 ± 0.01	0.30-0.92	17.96	0.34	-0.48	90.68
GGA*	0.48 ± 0.01	0.23-0.81	25.39	0.54	-0.85	92.38
CSI*	26.19 ± 0.45	9.27-45.90	29.68	-0.12	-0.99	82.12
NDVI*	0.68 ± 0.02	0.36-7.09	55.69	16.42	276.80	66.23
ELS+	4.28 ± 0.03	1.58-6.17	12.66	-0.88	4.01	73.08
LLS+	4.85 ± 0.04	1.58-6.67	14.33	-0.85	1.56	85.55
AUDPC-ELS	86.00 ± 0.60	30.00-140.00	28.53	0.05	-0.93	75.97
AUDPC-LLS	97.00 ± 0.47	40.00-140.00	19.87	-0.13	-0.54	86.00

[a]Hue, hue angle; GA, green area; GGA, greener area; CSI, crop senescence index; NDVI, normalized difference vegetation index; ELS, early leaf spot; LLS, late leaf spot; AUDPC-ELS and -LLS, area under disease progress curve for ELS and LLS, respectively. * RGB-image-based phenotyping.+ Manual disease scoring with a scale of 1-9 by Subrahmanyam et al. (1995). [b]mean± standard error of mean. [c]Coefficient of variation. [d]Broad-sense heritability.

Genetic differentiation and LD estimation among the mapping population

After quality control on the obtained 8,911 SNPs, a total of 8,772 SNP markers were distributed between the two sub-genomes (A and B) with quality criteria of minor allele frequency< 0.05 and call rate<0.95. Out of these, 8,152 SNPs were located on one of the linkage groups in the two sub-genomes (A and B), while 619 SNPs were located on scaffolds. The SNPs located on the scaffolds were excluded from the downstream analysis. The longest and shortest chromosome (Chr) spanned 149.79 Mb (Chr15) and 49.33 Mb (Chr08), respectively (Table 2; Supplementary Figure 2). However, Chr08 and Chr19 contained the highest and least number of SNPs (877 and 191, respectively).

The Bayesian model based population structure analysis was carried out with K=1-7 with 5 independent runs for each K. The estimated ΔK plot shown in Figure 2A reveals that 294 accessions are optimally grouped into two sub-populations (I

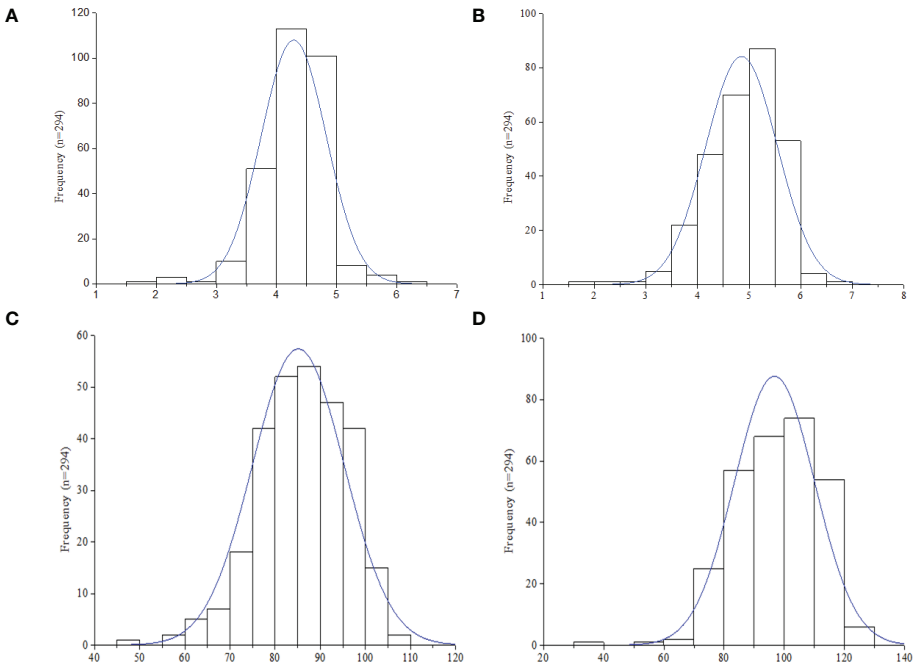


FIGURE 1
Frequency distribution of leaf spot diseases rating in the 294 accessions (n) in this study. **(A)** Early leaf spot (ELS). **(B)** Late leaf spot (LLS). **(C)** Area under disease progression curve (AUDPC) -ELS. **(D)** AUDPC - LLS.

and II) (Figure 2B). The sub-population I consisted of 99 accessions (≈33.67%) with nearly 50% of them from Uganda and Senegal, while those in sub-population II comprised 173 accessions (≈58.84%) with 43, 26, 20, and 20 accessions from Niger, Ghana, Malawi, and Mali, respectively (Table 3; Supplementary Table 1). Based on the Q≥ 0.60 as pure lines, twenty-two accessions representing 7.48% of the population are considered admixtures (Figure 2C). Allele frequency divergence between two sub-populations was estimated as 0.38, while the expected heterozygosity between individual accessions with sub-population I and II are 0.15 and 0.18, indicating that sub-population II is relatively more diverse than sub-population I. The population structure stratification was consistent with kinship matrix, phylogenetic trees, and principal component analysis (Figures 2D, E). The LD decay across the two sub-genomes of the studied panel was estimated to be about 300 kbp (Figure 2F).

Marker-trait association

In order to remove environmental effect during the marker-trait association mapping, BLUP values were used where the effect of environment (year), replicates within E, G, GE, and error were considered as random effects. With the six multi-locus models, a total of 97 SNPs distributed across the 20 chromosomes with an average of ≈5 SNPs per Chr and range of 1 SNP on Chr15 to 11 SNPs on Chr16 were detected (Figure 3A; Supplementary Table 3). Of these, the power of detection among the six models followed pKWmEB (46 SNPs) > pLARmEB (40 SNPs) > mrMLM (25 SNPs) > FASTmrMLM (18 SNPs) > ISIS EM-BLASSO (16 SNPs) > FASTmrEMMA (12 SNPs) (Figure 3B).

Comparative analysis among the six multi-locus models revealed that twenty-nine unique SNPs were detected by at least two models for one or more traits across 16 chromosomes with exception of Chr08, Chr10, Chr11 and Chr19 (LOD ≥ 3.00) (Table 4). Among these, seven SNPs (AX-176823205 (Ch01), AX-176823123 (Chr01), AX176799357 (Chr02), AX-176796174 (Chr05), AX-147224865 (Chr06), AX-147239793 (Chr06) and AX-177643984 (Chr20)) associated with two traits (Table 4). For example, SNP, AX-176823205 associated with both Hue and LLS with allele C at 91414269, LOD of 3.40-5.04 and phenotypic variation explained (PVE) of ≤ 3.12%. The C allele had negative and positive effect on Hue and LLS, respectively. This and the other six SNPs (AX-176823123, AX176799357, AX-176796174, AX-147224865, AX-147239793 and AX-177643984) may be the basis for the r values observed on Supplementary Figure 1. Therefore, these SNPs could be valuable genetic resources to understand the relationship among the evaluated indices associated with leaf spot diseases.

TABLE 2 Distribution of single nucleotide polymorphism (SNP) markers between A and B sub-genomes of groundnut (chromosome) of the studied population.

Genome/LG	Chr.	Length (bp)	Length (kb)	Length (Mb)	No. of SNPs	Kbs/SNP	SNPs/Mb
A01	1	106806005	106806.01	106.81	444	240.6	4.16
A02	2	93528862	93528.86	93.53	327	286.0	3.50
A03	3	134894015	134894.02	134.89	406	332.3	3.01
A04	4	121312866	121312.87	121.31	429	282.8	3.54
A05	5	109393004	109393.00	109.39	393	278.4	3.59
A06	6	112315382	112315.38	112.32	428	262.4	3.81
A07	7	78545074	78545.07	78.55	312	251.7	3.97
A08	8	49330572	49330.57	49.33	191	258.3	3.87
A09	9	120497462	120497.46	120.50	267	451.3	2.22
A10	10	109302486	109302.49	109.30	230	475.2	2.10
B01	11	137285820	137285.82	137.29	325	422.4	2.37
B02	12	108946667	108946.67	108.95	247	441.1	2.27
B03	13	135317267	135317.27	135.32	325	416.4	2.40
B04	14	132798623	132798.62	132.80	365	363.8	2.75
B05	15	149794401	149794.40	149.79	334	448.5	2.23
B06	16	136159078	136159.08	136.16	433	314.5	3.18
B07	17	126126609	126126.61	126.13	507	248.8	4.02
B08	18	129540280	129540.28	129.54	610	212.4	4.71
B09	19	147063990	147063.99	147.06	877	167.7	5.96
B10	20	135921470	135921.47	135.92	702	193.6	5.16
Total		**2374879933**	**2374879.93**	**2374.88**	**8152**	**6348.00**	**68.82**

LG, Linkage group; Chr, Chromosome.

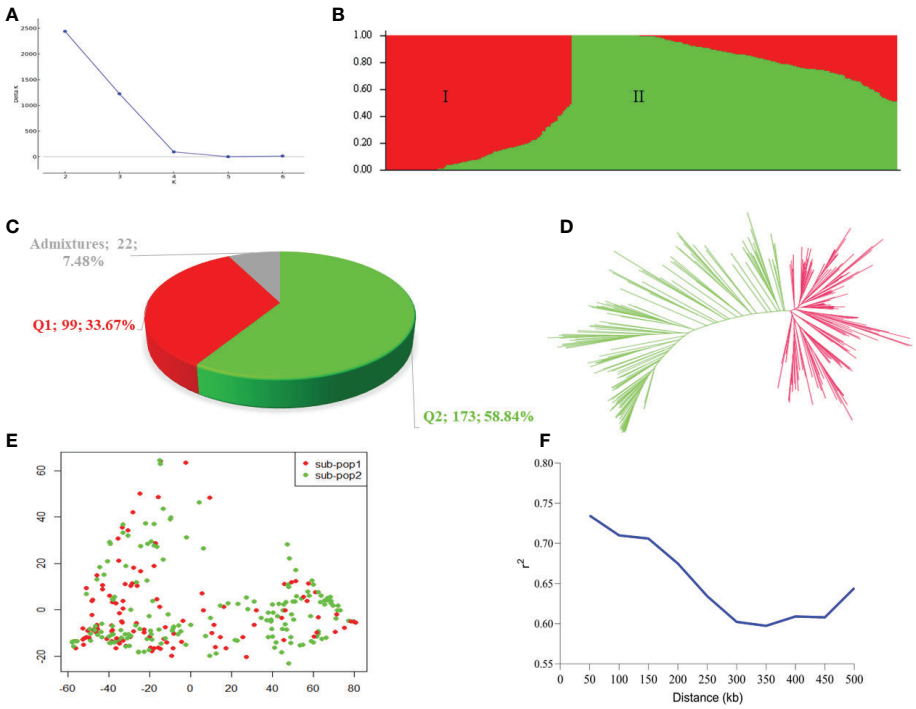


FIGURE 2
Population stratification of the 294 accessions in this study. **(A)** Estimated ΔK in the population structure analysis retrieved from STRUCTURE HARVESTER online programme. **(B)** Population structure by STRUCTURE software. The two colours (red and green) represent two sub-populations. Each colour represent one inferred ancestral population with the vertical column representing one individual accession and coloured segment in each column denotes percentage of the individual inferred ancestral population in the studied panel. **(C)** Proportion of pure and admixture lines based on membership coefficients (Q) ≥0.60 as pure lines, while those with Q< 0.60 considered as admixture lines. **(D)** A neigbour-joining phylogenetic tress. **(E)** Principal component analysis plot. **(F)** Linkage disequilibrium decay plot within the 500 kb across the two sub-genomes **(A, B)**.

In addition, nine SNPs (AX-176799357, AX-176806210, AX-176796174, AX-176793720, AX-147224496, AX-176818776, AX-176820950, AX-177643984 and AX-133120520) were associated with at least one model with PVE ≥ 10%, hence these were considered as major SNPs for downstream analysis for candidate genes prediction (Table 4). Interestingly, AX-176799357 was linked to both LLS and AUDPC-LLS, AX-176796174 associated with both Hue and AUDPC-ELS, and AX-133120520 associated with both GA and GGA.

As typical of quantitative traits, seventy SNPs were trait and model specific with LOD and PVE ranging 3.04 (AX-177637712 on Chr17 with LLS) to 35.90 (AX-147227883 on Chr07 with GGA), and<0.01% (AX-177637712 on Chr17 with LLS) to 21.79% (AX-176797562 on Chr02 with CSI) (Supplementary Table 3). These markers may need further verification for their possible use in practical plant breeding.

Candidate gene prediction around stable SNPs

To predict candidate genes, significant SNPs with at least 1 of the following criteria: linked to more than 1 trait, detected by at least 4 GWAS model and with PVE ≥ 10% were used for gene mining. Two hundred and fifty-three genes located within 300 kbp of 14 SNPs were mined and *in silico* analysis performed. In all, 17 potential candidate genes were predicted, of which 12 and 5 are located down- and up-stream of the linked SNP positions, respectively (Table 5). BAEJ4E gene is located 172.3 kbp up-stream of AX-176823205 (chr01) was predicted for Hue and LLS traits, and this gene encodes for glutathione S-transferase which have been reported to regulate plant response to fungal infection (Dean et al., 2005; Gullner et al., 2018). In addition, four genes (7U0T6N, CA7A2G, YE4BG5 and AZU29N predicted for CSI & ELS, LLS & AUDPC-LLS, AUDPC-LLS, and GA & GGA, respectively) are involved in photosynthesis pathway, hence could be involved in regulating groundnut response to leaf spot pathogens. Phytohormones are reported to regulate plant's response to disease attack or resistance (Denancé et al., 2013). 5L52L8 gene is involved in ethylene biosynthesis, and L76SLB gene for auxin efflux carrier. Both of genes (5L52L8 and L76SLB) were located ≤ 289 kbp from the AX-176799357 on Chr02 linked to LLS and AUDPC-LLS, and AX-176793720 on Chr05 linked AUDPC-LLS, respectively (Table 5).

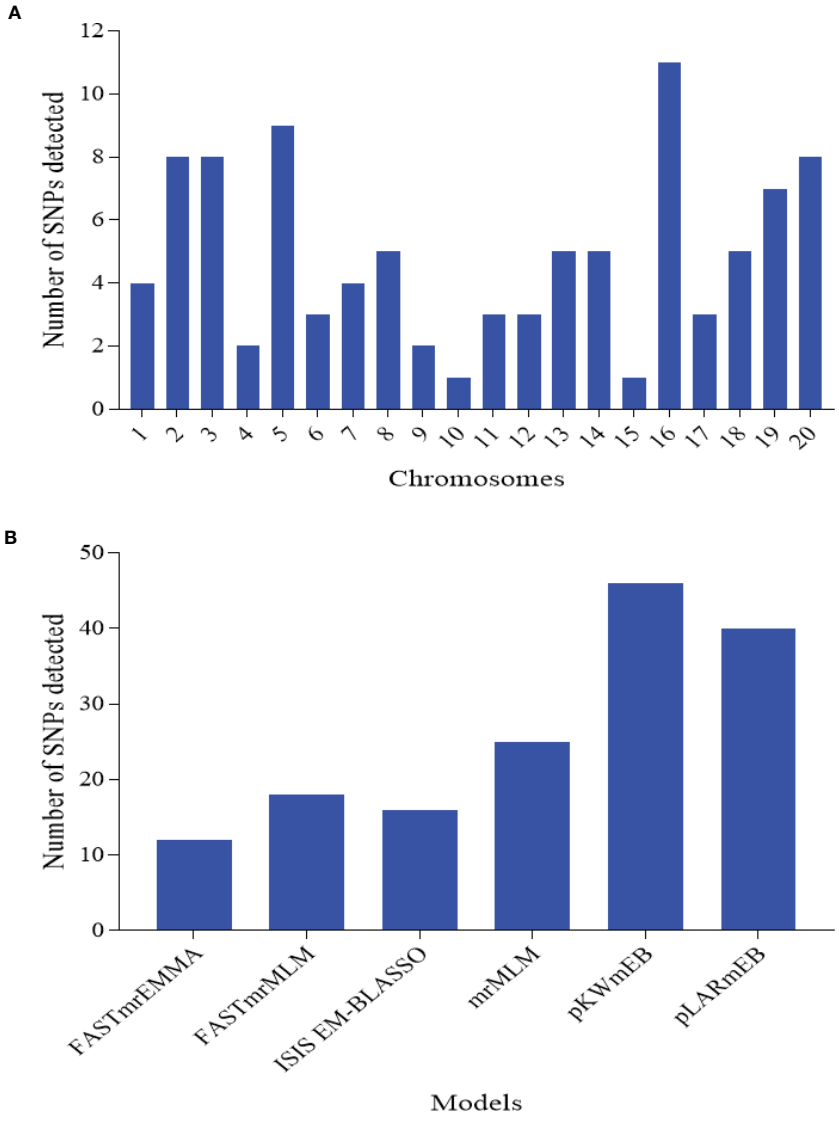


FIGURE 3
Number of significantly associated single nucleotide polymorphism (SNP) markers detected for the nine traits associated with leaf spot diseases rating. **(A)** SNPs detected on each chromosome of the two sub-genomes. A sub-genome comprised chromosome (Chr) 1 to 10, while B sub-genome consisted of Chr11-20. **(B)** SNPs detected by each of the six multi-locus models implemented in this study.

We further explored the promoter region (2 kbp) of the 17 candidate genes predicted for Cis-acting regulatory elements (CAREs) with PlantCare database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html (Lescot et al., 2002) that may be involved in the modulating groundnut response to leaf spot diseases. Aside essential CAREs (CAAT- and TATA-box) as well as 20 light responsiveness CAREs (such as AE-box, AT1-motif, Box 4, G-Box, GATA-motif, etc.), twenty CAREs with potential in modulating gene expression were identified (Supplementary Table 5; Figure 4). Seven genes (29A50M, 7U0T6N, AZU29N, CA7A2G, G540IJ, KVF40G, and YE4BG5) were found to contain at least one ATTCTCTAAC (TC-rich repeats) demonstrated to involve in defence and stress responsiveness in Nicotiana tabacum (Figure 4) (Diaz-De-Leon et al., 1993; Xu et al., 2011; Wang et al., 2020). Moreover, the role of salicylic acid (SA) in plant defence is well documented, thus SA is required for basal resistance against pathogens as well as for the inducible defence mechanism, systemic acquired resistance which confers resistance against a broad-spectrum of pathogens including *P. arachidicola* and *N. personata* (Chaturvedi and Shah, 2007). Five predicted candidate genes (29A50M, BAEJ4E, G540IJ, P6RS4K and YE4BG5) (Table 4) contain at least one TCA-element (TCAGAAGAGG) involved in SA responsiveness (Figure 4).

TABLE 3 Population stratification based on membership coefficient (Q) from the STRUCTURE software.

Country	Number of accessions	Sub-pop I	Sub-pop II	Admixtures
Ghana	49	19	26	4
Malawi	32	10	20	2
Mali	27	4	20	3
Mozambique	22	4	15	3
Niger	44	1	43	0
Senegal	31	19	8	4
Togo	18	6	11	1
Uganda	54	28	22	4
Zambia	17	8	8	1
Total (%)	**294**	**99 (37.67%)**	**173 (58.84%)**	**22 (7.48%)**

Accessions assigned to sub-population had Q≥0.60, while those assigned as admixture had Q<0.60.

Other CAREs that may be implicated in groundnut response to leaf spot pathogens identified include (abscisic acid (ABA), gibberellin (GA), auxin, methyl jasmonate (MeJA), MYB binding site involved in flavonoid biosynthetic genes regulation (MBSI) and so) were detected in at least 1 of the predicted candidate genes highlighting the possibility of their involvement in modulating groundnut response to leaf spot pathogen. The actual roles of the 17 predicted candidate genes warrant further screening and functional validation to unravel the bases of correlation among the studied traits (Supplementary Table S1).

Discussion

Leaf spots (ELS and LLS) diseases are the two most destructive groundnut diseases in Ghana resulting in ≤ 70% yield losses which is controlled largely by chemical method (Naab et al., 2005; Denwar et al., 2021). To speed up breeding efforts, the present study was undertaken to identify significantly associated molecular markers and putative candidate genes linked to ELS and LLS diseases' indicators. Irrespective of the breeding strategy, germplasm serves as lifeblood for breeding effort (Acquaah, 2009; Allier et al., 2020). With this, the present study utilized 294 groundnut germplasm assembled from nine African countries (Ghana, Malawi, Mali, Mozambique, Niger, Senegal, Togo, Uganda and Zambia) (Supplementary Table 1) to assess their response to leaf spot diseases in 2020 and 2021. These germplasm exhibited wide range of responses to leaf spot diseases (resistance to susceptibility) (Table 1; Figure 1) of which a portion was recently published by Sie et al. (2022). This suggest that the germplasm hold a promise for breeding groundnut cultivars resistant to leaf spot diseases as well as other demand driven traits. In addition, the high H^2 (Table 1) suggest that these ELS and LLS as well as other indicators used are largely controlled by polygenes with both major minor effects. This highlights that selection based only on phenotypic variation may be misleading.

One of setbacks in phenotyping large germplasm is the laborious nature coupled with human error among others (Sandhu et al., 2022), which have necessitated a number of phenotyping strategies/platforms which complement the conventional phenotyping. Among these include multispectral imaging (Kobayashi et al., 2001; Chang et al., 2021), RGB imaging (Duan et al., 2018) and others (see Sandhu et al. (2022) for more). The present study manually scored leaf spot incidences and these scores were converted to quantitative scores by AUDPC which were correlated by r=0.47-0.76, giving credence to our scorings (Supplementary Figure 1). In addition, we used four RGB imaging indices (Hue, GA, GGA and CSI) and one vegetation index (NDVI) to confirm both the manual scoring of ELS and LLS as well as ELS- and LLS-AUDPC. It was observed that CSI positively correlated with ELS, LLS, ELS-AUDPC and LLS-AUDPC with r=0.44-0.71 (Supplementary Figure 1), thus the higher the CSI, the more developed the leaf spot diseases, giving an indication that CSI could be used as a screening criterion for leaf spot diseases in groundnut. The predominance and distribution of leaf spots vary according to regions. In most cases there is dual occurrence but the predominance towards physiological maturities varies. For instance, in Eastern Africa, LLS predominates, whereas in Western and central Africa, ELS predominates. However, in the current study, CSI could not distinguish between ELS and LLS, hence further study is needed to develop a model or phenotyping platform to create the distinction.

Recent advances in plant phenotyping involve the use of unmanned aerial vehicles (UAVs) to collect several images generating large amounts of data. Several studies have reported that UAVs are faster and more effective for phenotyping large populations for traits such as height and drought tolerance in groundnut breeding (Sarkar et al., 2021b; Chapu et al., 2022), hence providing the desired high-throughput. This study therefore lays the foundation for investment in such more advanced equipment in groundnut breeding for selection for resistance to late leaf spot and

TABLE 4 SNPs detected by at least two of the six models for one or more traits.

Trait name [a]	SNP markers	Chr [b]	Position (bp)	Models [c]	LOD [d]	QTN effect	PVE (%) [e]	MAF [f]	Alleles [g]
Hue	AX-176823205	1	91414269	1,2,3,5	3.40-4.94	-1.09- -1.90	1.49-3.12	0.21	C
LLS				2	5.04	<0.01	0.01		
CSI	AX-176823123	1	96303630	1,2	3.79	0.26-0.47	0.24-1.35	0.22	T
ELS				6	4.01	0.0029	3.02		
AUDPC-ELS	AX-147212206	2	453115	1,5,6	3.68-4.14	0.10-0.20	3.19-7.49	0.14	A
ELS	AX-176801892	2	38320573	1,2,3	3.18-3.99	0.01-0.02	0.05-3.49	0.10	G
LLS	AX-176799357	2	76275147	1,3,6	7.73-16.45	0.09-0.12	12.78-20.69	0.49	T
AUDPC-LLS				4	4.18	1.83	5.16		
ELS	AX-176801892	2	38320573	1,2,3	3.18-3.99	0.01-0.02	0.05-3.49	0.10	G
AUDPC-ELS	AX-147217628	3	118446096	1,2,3,4,6	3.13-7.02	0.14-0.37	5.10-8.51	0.28	G
AUDPC-LLS	AX-176806210	4	27792120	1,2,3,4,5,6	3.20-9.84	3.14-9.91	7.21-38.50	0.13	G
Hue	AX-176796174	5	15450246	1,2,6	4.72-7.31	-3.21- -2.48	10.90-24.54	0.23	T
AUDPC-ELS				5	6.86	0.26	12.22		
Hue	AX-147222698	5	81969778	3,4,5	3.25-7.87	1.15-3.22	3.24-6.16	0.46	G
AUDPC-LLS	AX-176793720	5	102154664	1,2,4,5,6	3.22-5.65	-3.03- -1.83	4.64-11.00	0.35	G
LLS	AX-176808070	6	2451915	1,6	3.09-3.80	0.06	2.34-2.91	0.14	T
NDVI	AX-147224496	6	5557950	3,4,5	4.36-5.46	<0.01	0.01-10.15	0.49	C
GA & GGA	AX-147224865	6	11297985	6	5.47-6.37	<0.01	3.54-3.92	0.12	G
CSI	AX-147228765	7	70634323	1,2,5,6	3.46-5.03	0.92-1.13	1.28-3.74	0.10	T
AUDPC-ELS	AX-147232168	9	1308812	1,2	3.78-3.88	-0.13- -0.09	2.22-4.89	0.32	T
GA & GGA	AX-147239793	12	295929	6	3.88-5.67	<0.01	0.06-2.77	0.50	C
LLS	AX-176804113	13	14456209	1,5,6	4.61-7.46	0.05-0.08	4.03-8.34	0.38	C
CSI	AX-176818776	13	115500602	1,2,6	3.01-3.03	1.21-3.69	12.12-62.76	0.19	C
AUDPC-ELS	AX-147246588	14	2233380	1,2	3.25-3.58	-0.13- -0.08	2.19-5.15	0.45	T
LLS	AX-147247867	14	101149286	1,2,6	3.16-5.88	0.06-0.08	2.56-5.97	0.17	G
AUDPC-ELS	AX-176820950	16	18466678	1,6	5.32-5.41	0.14-0.15	6.84-10.05	0.42	C
CSI	AX-147253729	16	128084428	5,6	3.41-4.21	-0.64- -0.61	3.14-6.33	0.30	C
ELS	AX-177643647	18	7462734	5,6	3.91-6.83	<0.01	0.01-0.58	0.10	G
Hue	AX-147259422	18	127104626	1,2,5	4.21-6.36	-1.95- -1.69	2.03-3.51	0.10	C
Hue	AX-176824170	19	145041727	1,2,5,6	3.08-6.29	-1.84- -1.23	3.03-6.54	0.28	C
AUDPC-ELS	AX-177644589	20	53038882	1,2,6	3.19-4.51	-0.16 - -0.14	5.38-8.79	0.43	C
LLS	AX-177639015	20	121402090	1,5,6	3.96-4.88	0.04-0.06	2.34-5.28	0.43	G
GA & GGA	AX-177643984	20	133120520	6	3.01-3.84	<0.01	2.97-10.44	0.48	G

[a]Hue, hue angle; GA, green area; GGA, greener area; CSI, crop senescence index; NDVI, normalized difference vegetation index; ELS, early leaf spot; LLS, late leaf spot; AUDPC-ELS, area under disease progress curve for early leaf spot; AUDPC-LLS, area under disease progress curve for late leaf spot. [b]Chromosomes. [c]mrMLM (1), FASTmrMLM (2), FASTmrEMMA (3), ISIS EM-BLASSO (4), pLARmEB (5) and pKWmEB (6). [d]Logarithm of odd. [e]Phenotypic variation explained (%). [f]Minor allele frequency. [g]Associated allele. SNP positions, GWAS models and PVE underlined were considered as stable for candidate genes prediction.

groundnut rosette disease which are the most important foliar diseases in SSA (Chapu et al., 2022).

Leaf spot diseases are well known to affect photosynthetic capacity of leaves (Singh et al., 2011). Three image-based indices, i.e., GA, GGA, Hue and one vegetation index, i.e., NDVI negatively correlated weakly to strongly with ELS, LLS, ELS-AUDPC and LLS-AUDPC (Supplementary Figure 1). However, GA, GGA and Hue seem more better reflective of leaf spot diseases based on our correlation analysis (Supplementary Figure 1) According to Wang et al. (2017), NDVI is a proxy of green biomass, which is linked to canopy photosynthesis. Pronounced changes take place in the visible portion of the electromagnetic spectrum due to the effect caused by a disease or physiological stress to the reflectance properties of the vegetation. Healthy plants absorb both the red and blue light of the electromagnetic spectrum, whiles reflecting near-infrared and some part of the green light (Knipling, 1970). Less of the red light is absorbed by a stressed plant as well as reflecting less of the near-infrared light. Wu et al. (2014) have provided evidence that healthy and unhealthy plants differ in their absorption and reflection of visible and near-infrared lights.

The advancement in high-throughput genotyping, next generation sequencing, bioinformatics tools, statistical models, etc., have served as catalyst to access valuable information from

TABLE 5 Predicted candidate genes around fourteen stable/major SNPs for the studied traits.

	SNP markers			Predicted candidate genes		
Trait name [a]	Identity [b]	Chr[c]	Position (bp)	Gene [d]	Position from SNP (kb)[e]	Domain/function/pathway [f]
Hue & LLS	AX-176823205	1	91414269	*BAEJ4E*	172.3 (u)	Glutathione S-transferase
CSI & ELS	AX-176823123	1	96303630	*7U0T6N*	299.6 (d)	Oxygenic photosynthesis (pathway), RuBP carboxylase (enzyme)
				46C7PY	216.1 (d)	Serine/threonine specific protein phosphatase (Enzyme)
LLS & AUDPC-LLS	AX-176799357	2	76275147	*5L52L8*	247.4 (d)	Ethylene biosynthesis/Superoxide dismutase
				CA7A2G	21.4 (d)	Photosystem antenna protein-like
AUDPC-ELS	AX-147217628	3	118446096	*UB9N1X*	111.4 (u)	Ammonium/urea transporter
AUDPC-LLS	AX-176806210	4	27792120	*YE4BG5*	116.7 (d)	Oxygenic photosynthesis (Pathway)
Hue & AUDPC-ELS	AX-176796174	5	15450246	*29A50M*	16.3 (d)	Myc-type, basic helix-loop-helix (bHLH) domain
AUDPC-LLS	AX-176793720	5	102154664	*L76SLB*	288.6 (d)	Auxin efflux carrier
GA & GGA	AX-147224865	6	11297985	*AZU29N*	269.4 (d)	Geranylgeranyl reductase involved in chlorophyll a biosynthesis
CSI	AX-147228765	7	70634323	*UJVA88*	75.9 (u)	Peroxidase
				P6RS4K	284.5 (u)	Pentatricopeptide repeat
GA & GGA	AX-147239793	12	295929	*D1YXCG*	216.1 (d)	Ribosomal protein S5 domain 2-type fold
CSI	AX-176818776	13	115500602	*1EP8UK*	23.6 (d)	Thaumatin
AUDPC-ELS	AX-176820950	16	18466678	*8W91IE*	91.5 (d)	Superoxide dismutase
Hue	AX-176824170	19	145041727	*KVF40G*	101.2 (d)	Homeobox-leucine zipper protein
GA & GGA	AX-177643984	20	133120520	*G540IJ*	168.0 (u)	Glutaredoxin family protein

[a]Hue, hue angle; GA, green area; GGA, greener area; CSI, crop senescence index; NDVI, normalized difference vegetation index; ELS, early leaf spot; LLS, late leaf spot; AUDPC-ELS, area under disease progress curve for early leaf spot; AUDPC-LLS, area under disease progress curve for late leaf spot. [b]Marker name. c Chromosomes. [c]Marker position on the chromosome. [d]Gene symbol obtained from phytozome (https://phytozome-next.jgi.doe.gov/). [e]Position from the marker (u=upstream and d=downstream). [f]Gene annotation obtained from phytozome.

genomic databases and a large number of germplasm, allowing effective harnessing of genetic diversity of a crop (Vikas et al., 2022). Such diversity is vital for broadening the genetic base, as it increases the probability of identifying more unique genes for which two parents have different alleles (Mascher et al., 2019). The 294 germplasm optimally divided into 2 sub-populations/clusters/clade (Figures 2A-E). This information is not only useful for the utilization of the 294 germplasm, but also provide valuable information in their conservation as the cost of maintenance and uncertainty about their genetic similarity and dissimilarity (Wambugu et al., 2018; Mascher et al., 2019). The results from this study could contribute to tracking the identity of accessions, avoiding unnecessary duplications within and between genebanks and breeding programmes, while maintaining the genetic integrity of accessions (Mascher et al., 2019). The germplasm in sub-populations/clusters/clade did not follow strictly to a country/sub-region (Table 3; Supplementary Table 1), suggesting that some germplasm may have a duplicated version(s) in other country(ies) breeding programmes.

A number of studies by linkage and association mappings have been conducted on leaf spot diseases in groundnut (Pasupuleti et al., 2013; Pandey et al., 2014; Zhang et al., 2020a; Shaibu et al., 2021). However, most of the reported genomic regions/markers are not consistent due to population specific or environment sensitivity (Patil et al., 2018). This necessitated the present study in our attempt to identify potential markers for marker-assisted selection (MAS) (Barmukh et al., 2022). To increase the chances of detecting more possible SNPs that could be associated with the various indices used to assess leaf spot diseases, we applied 6 multi-locus models (mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB and ISIS EM-BLASSO) of GWAS. These models differed in their power of detection of associated SNPs, which are in consonance with several earlier studies (Zhang et al., 2019; Karikari et al., 2020). In comparison to study of Zhang et al.

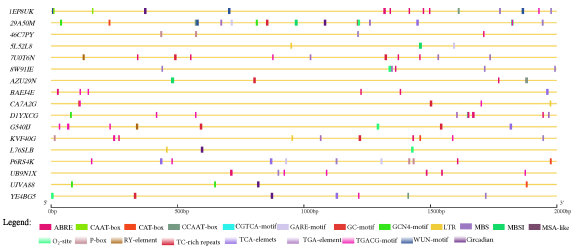


FIGURE 4
Key cis-acting regulatory elements (CAREs) identified in the promoter regions (2 kbp upstream) of the 17 predicted candidate genes in this study retrieved from PlantCare database (Lescot et al., 2002). ABRE= abscisic acid responsiveness; CAAT-box= essential CARE, CAT-box= CARE related to meristem expression, CCAAT-box= MYBHv1 binding site, GARE-motif= gibberellin-responsive element, GC-motif= anoxic specific inducibility, GCN4-motif= involved in endosperm expression, LTR= low temperature responsiveness, MBS= MYB binding site involved in drought-inducibility, MBSI= MYB binding site involved in flavonoid biosynthetic genes regulation, MSA-like= cell cycle regulation, O_2-site= zein metabolism regulation, P-box= gibberellin-responsive element, RY-element= seed-specific regulation, TC-rich repeats= defense and stress responsiveness, TCA-elements= salicylic acid responsiveness, TGA-element= auxin-responsive element, TGACG-motif= MeJA-responsiveness, and, WUN-motif= and Circadian= circadian control.

(2020a) who reported only SNPs with major effects (PVE ≥ 10%), the present study detected numerous SNPs with both minor (PVE< 10%) and major effects. According to Zhou et al. (2021) SNPs identified by multiple models are usually reliable when several multi-locus models of GWAS are applied on the same dataset. Hence, AX-176801892 (Chr02, LOD= 3.18-3.99; PVE= 0.05-3.49%) linked to ELS, AX-176799357 (Chr04, LOD= 7.73-16.45; PVE= 12.78-20.69%) linked to LLS, AX-176806210 (Chr04, LOD=3.20-9.84; PVE=7.21-38.50%) linked to AUDPC-LLS, AX-147224496 (Chr04, LOD= 4.36-5.46; PVE= 0.01-10.15%) linked to NDVI and several others (Table 4; Supplementary Table 3) could be targeted for verification and use for practical plant breeding. The numerous genomic regions/SNPs suggest that leaf spot diseases are regulated by multiple loci with both minor and major effects, hence selection based on only phenotypic data may be misleading (Sood et al., 2020). In addition to the above, a number of SNPs that colocalized with multiple indices assessed in this study could be useful in developing models that could distinguish between ELS and LLS.

One advantage of GWAS is the power to detect SNPs in a narrow genomic regions of which causal variants could result in variation in traits of economic importance such as leaf spot diseases (Jiang et al., 2009; Zhao et al., 2021). However, one of the factors that determine resolution of GWAS is LD (Korte and Farlow, 2013; Ibrahim et al., 2020) which is population specific and influenced by recombination, genetic drift and mating system (Yao et al., 2009). In this study, the LD decay across the two sub-genomes of groundnut was estimated to be about 300 kbp which is nearly 50% higher than the estimation by Zhang et al. (2020a) within 120 kbp in US accessions with 13,382 SNPs. Upon application of LD for the 294 African germplasm, seventeen candidate genes were predicted based on in-silico analyses (Table 5; Figure 4). A number of these genes encode for phytohormone/contain CAREs/photosynthesis pathway which could be implicated in groundnut response to leaf spot diseases. For plants to defend themselves from pathogen attack, plants often rely on elaborate signaling networks regulated by phytohormones (Kazan and Lyons, 2014). These genes would be valuable for future functional validation by gene overexpression, CRISPR/Cas9 technology, among others (Zaidi et al., 2020).

Conclusions

In all, ninety-seven SNPs were detected by the six multi-locus GWAS models. Out of which twenty-nine SNPs were detected by at least two models for one or more traits across 16 chromosomes with explained phenotypic variation ranging from 0.01 - 62.76%, with exception of Chr08, Chr10, Chr11, and Chr19. Two hundred and fifty-three genes were located within 300 kbp of 14 SNPs, from seventeen potential candidate genes were predicted. Most of the predicted candidate genes were found to SA, ABA, GA, auxin and MeJA responsive and MBSI CAREs implicated in plant response to biotic stresses. The results from this study would be useful for breeding leaf spots resistance cultivars.

Data availability statement

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

Author contributions

RO-F, ES, MR, DF, DO, and PO-A conducted the experiments. BK, RO-F, DP, and MBa organized and supervised the overall project. BK, YK, and ES performed the data analyses and wrote the manuscript. RO-F, DO and MBu edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was made possible by the generous support of the BMGF funded Accelerated Varietal Improvement and Seed Systems for Legumes and Cereals in Africa (AVISA) Project Grant Number OPP1198373 and the American people through the United States Agency for International Development (USAID) Cooperative Agreement No. 7200AA 18CA00003 to the University of Georgia as management entity for U.S. Feed the

Future Innovation Lab for Peanut (2018–2023). The funders had no role in experiments, data analyses and publication.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1076744/full#supplementary-material

SUPPLEMENTARY FIGURE 3
Heatmap of the genomic kinship matrix obtained by the VanRaden (2008) Method among the single nucleotide polymorphism markers and 294 groundnut germplasm.

References

Acquaah, G. (2009). *Principles of plant genetics and breeding* (The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK:John Wiley & Sons).

Allier, A., Teyssèdre, S., Lehermeier, C., Moreau, L., and Charcosset, A. (2020). Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics* 21 (1), 349. doi: 10.1186/s12864-020-6756-0

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23 (5), 451–466. doi: 10.1016/j.tplants.2018.02.001

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556

Asibuo, J. Y., Akromah, R., Safo-Kantanka, O., Adu-Dapaah, H. K., Ohemeng-Dapaah, S., and Agyeman, A. (2008). Chemical composition of groundnut, *Arachis hypogaea* (L) landraces. *Afr. J. Biotechnol.* 7 (13), 2203–2208.

Atakora, W., and Kwakye, P. (2016). Measurement and modeling nitrous oxide emissions from ferric luvisols in the guinea savanna agro-ecological zone of Ghana. *Int. J. Plant Soil Sci.* 10 (3), 1–8. doi: 10.9734/IJPSS/2016/24165

Atiah, W. A., Amekudzi, L. K., Quansah, E., and Preko, K. (2019). The spatio-temporal variability of rainfall over the agro-ecological zones of Ghana. *Atmospheric Climate Sci.* 9 (03), 527. doi: 10.4236/acs.2019.93034

Atiah, W. A., Mengistu Tsidu, G., Amekudzi, L. K., and Yorke, C. (2020). Trends and interannual variability of extreme rainfall indices over Ghana, West Africa. *Theor. Appl. Climatol.* 140 (3), 1393–1407. doi: 10.1007/s00704-020-03114-6

Awada, L., Phillips, P. W. B., and Smyth, S. J. (2018). The adoption of automated phenotyping by plant breeders. *Euphytica* 214 (8), 148. doi: 10.1007/s10681-018-2226-z

Barmukh, R., Roorkiwal, M., Dixit, G. P., Bajaj, P., Kholova, J., Smith, M. R., et al. (2022). Characterization of 'QTL-hotspot' introgression lines reveals physiological mechanisms and candidate genes associated with drought adaptation in chickpea. *J. Exp. Bot.*, erac348. doi: 10.1093/jxb/erac348

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al. (2004). The pfam protein families database. *Nucleic Acids Res.* 32 (Database issue), D138–D141. doi: 10.1093/nar/gkh121

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *J. Stat. Software* 67 (1), 1–48. doi: 10.18637/jss.v067.i01

Berhe, M., Dossa, K., You, J., Mboup, P. A., Diallo, I. N., Diouf, D., et al. (2021). Genome-wide association study and its applications in the non-model crop *Sesamum indicum*. *BMC Plant Biol.* 21 (1), 283. doi: 10.1186/s12870-021-03046-x

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48 (4), 438–446. doi: 10.1038/ng.3517

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

Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T., Buckler, E., and Ramdoss, Y. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23 (19), 2633–2635. doi: 10.1093/bioinformatics/btm308

Chang, F., Lv, W., Lv, P., Xiao, Y., Yan, W., Chen, S., et al. (2021). Exploring genetic architecture for pod-related traits in soybean using image-based phenotyping. *Mol. Breed.* 41 (4), 28. doi: 10.1007/s11032-021-01223-2

Chapu, I., Okello, D. K., Okello, R. C. O., Odong, T. L., Sarkar, S., and Balota, M. (2022). Exploration of alternative approaches to phenotyping of late leaf spot and groundnut rosette virus disease for groundnut breeding. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.912332

Chaturvedi, R., and Shah, J. (2007). "Salicylic acid in plant disease resistance," in *Salicylic acid: A plant hormone*. Eds. S. Hayat and A. Ahmad (Netherlands, Dordrecht: Springer), 335–370.

Chiteka, Z. A., Gorbet, D. W., Knauft, D. A., Shokes, F. M., and Kucharek, T. A. (1988). Components of resistance to late leafspot in peanut. II. correlations among components and their significance in breeding for resistance. *Peanut Sci.* 15 (2), 76–81. doi: 10.3146/i0095-3679-15-2-9

Chu, Y., Chee, P., Culbreath, A., Isleib, T. G., Holbrook, C. C., and Ozias-Akins, P. (2019). Major QTLs for resistance to early and late leaf spot diseases are identified on chromosomes 3 and 5 in peanut (*Arachis hypogaea*). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00883

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

Danful, R., Kassim, Y. B., Puozaa, D. K., Oteng-Frimpong, R., Rasheed, M. A., Wireko-Kena, A., et al. (2019). Genetics of stay-green trait and its association with leaf spot tolerance and pod yield in groundnut. *Int. J. Agron.* 30640262019. doi: 10.1155/2019/3064026

Dean, J. D., Goodwin, P. H., and Hsiang, T. (2005). Induction of glutathione s-transferase genes of *Nicotiana benthamiana* following infection by *Colletotrichum destructivum* and *C. orbiculare* and involvement of one in resistance. *J. Exp. Bot.* 56 (416), 1525–1533. doi: 10.1093/jxb/eri145

Delfini, J., Moda-Cirino, V., Dos Santos Neto, J., Ruas, P. M., Sant'Ana, G. C., Gepts, P., et al. (2021). Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. *Sci. Rep.* 11 (1), 2964–2964. doi: 10.1038/s41598-021-82437-4

Denancé, N., Sánchez-Vallet, A., Goffner, D., and Molina, A. (2013). Disease resistance or growth: the role of plant hormones in balancing immune responses and fitness costs. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00155

Denwar, N. N., Simpson, C. E., Starr, J. L., Wheeler, T. A., and Burow, M. D. (2021). Evaluation and selection of interspecific lines of groundnut (*Arachis hypogaea* l.) for resistance to leaf spot disease and for yield improvement. *Plants* 10 (5), 873. doi: 10.3390/plants10050873

Diaz-De-Leon, F., Klotz, K. L., and Lagrimini, L. M. (1993). Nucleotide sequence of the tobacco (*Nicotiana tabacum*) *Anionic peroxidase* gene. *Plant Physiol.* 101 (3), 1117–1118. doi: 10.1104/pp.101.3.1117

Duan, L., Han, J., Guo, Z., Tu, H., Yang, P., Zhang, D., et al. (2018). Novel digital features discriminate between drought resistant and drought sensitive rice under controlled and field conditions. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00492

Dwivedi, S. L., Pande, S., Rao, J. N., and Nigam, S. N. (2002). Components of resistance to late leaf spot and rust among interspecific derivatives and their significance in a foliar disease resistance breeding in groundnut (*Arachis hypogaea* l.). *Euphytica* 125 (1), 81–88. doi: 10.1023/A:1015707301659

Earl, D. A., and Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4 (2), 359–361. doi: 10.1007/s12686-011-9548-7

Falush, D., Matthew, S., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164 (4), 1567–1587. doi: 10.3410/f.1015548.197423

FAO, IFAD, UNICEF, WFP and WHO. (2019). *The state of food security and nutrition in the world. in ieee journal of selected topics in applied earth observations and remote sensing* Vol. 7 (Rome, Italy:FAO).

Gaikpa, D. S., Akromah, R., Asibuo, J. Y., and Nyadanu, D. (2017). Molecular and phenotypic resistance of groundnut varieties to leaf spots disease in ghana. *J. Microbiol. Biotechnol. Food Sci.* 6, 1043–1048. doi: 10.15414/jmbfs.2017.6.4.1043-1048

Geng, X., Sha, J., Liu, S., Bao, L., Zhang, J., Wang, R., et al. (2015). A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* 16 (1), 196. doi: 10.1186/s12864-015-1409-4

Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., and Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics* 2 (3), 156–183. doi: 10.1007/s43657-022-00048-z

Gracia-Romero, A., Vergara-Díaz, O., Thierfelder, C., Cairns, J. E., Kefauver, S. C., and Araus, J. L. (2018). Phenotyping conservation agriculture management effects on ground and aerial remote sensing assessments of maize hybrids performance in Zimbabwe. *Remote Sens.* 10 (2), 349. doi: 10.3390/rs10020349

Gullner, G., Komives, T., Király, L., and Schröder, P. (2018). Glutathione s-transferase enzymes in plant-pathogen interactions. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01836

Gupta, P. K., Kulwal, P. L., and Jaiswal, V. (2014). "Chapter two - association mapping in crop plants: Opportunities and challenges," in *Advances in genetics*, vol. 85 . Eds. T. Friedmann, J. C. Dunlap and S. F. Goodwin (USA: Academic Press), 109–147.

He, J., Meng, S., Zhao, T., Xing, G., Yang, S., Li, Y., et al. (2017). An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theor. Appl. Genet.* 130 (11), 2327–2343. doi: 10.1007/s00122-017-2962-9

Ibrahim, A. K., Zhang, L., Niyitanga, S., Afzal, M. Z., Xu, Y., Zhang, L., et al. (2020). Principles and approaches of association mapping in plant breeding. *Trop. Plant Biol.* 13 (3), 212–224. doi: 10.1007/s12042-020-09261-4

Jiang, Z., Michal, J. J., Chen, J., Daniels, T. F., Kunej, T., Garcia, M. D., et al. (2009). Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. *Int. J. Biol. Sci.* 5 (6), 528–542. doi: 10.7150/ijbs.5.528

Jin, J., Zhang, H., Kong, L., Gao, G., and Luo, J. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 42 (D1), D1182–D1187. doi: 10.1093/nar/gkt1016

Johnson, R. (1984). A critical analysis of durable resistance. *Annu. Rev. Phytopathol.* 22 (1), 309–330. doi: 10.1146/annurev.py.22.090184.001521

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27

Karikari, B., Wang, Z., Zhou, Y., Yan, W., Feng, J., and Zhao, T. (2020). Identification of quantitative trait nucleotides and candidate genes for soybean seed weight by multiple models of genome-wide association study. *BMC Plant Biol.* 20 (1), 404. doi: 10.1186/s12870-020-02604-z

Kazan, K., and Lyons, R. (2014). Intervention of phytohormone pathways by pathogen effectors. *Plant Cell* 26 (6), 2285–2309. doi: 10.1105/tpc.114.125419

Kefauver, S. C., El-Haddad, G., Vergara-Diaz, O., and Araus, J. L. (2015). "RGB Picture vegetation indexes for high-throughput phenotyping platforms (HTPPs)," in *In remote sensing for agriculture, ecosystems, and hydrology XVII*, (Toulouse, France:SPIE Remote Sensing) 9637, 82–90.

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens. Environ.* 1 (3), 155–159. doi: 10.1016/S0034-4257(70)80021-9

Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K., and Torigoe, Y. (2001). Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. *Phytopathology* 91 (3), 316–323. doi: 10.1094/phyto.2001.91.3.316

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9 (1), 29. doi: 10.1186/1746-4811-9-29

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

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Basel)* 14 (11), 20078–20111. doi: 10.3390/s141120078

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., and Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51 (7), 1076–1081. doi: 10.1038/s41588-019-0443-6

Naab, J. B., Tsigbey, F. K., Prasad, P. V. V., Boote, K. J., Bailey, J. E., and Brandenburg, R. L. (2005). Effects of sowing date and fungicide application on yield of early and late maturing peanut cultivars grown under rainfed conditions in Ghana. *Crop Prot.* 24 (4), 325–332. doi: 10.1016/j.cropro.2004.09.002

Nutsugah, S. K., Abudulai, M., Oti-Boateng, C., Brandenburg, R. L., and Jordan, D. L. (2007). Management of leaf spot diseases of peanut with fungicides and local detergents in Ghana. *Plant Pathol. J.* 6 (3), 248–253.

Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10 (3), 235–322. doi: 10.1080/07352689109382313

Oteng-Frimpong, R., Kassim, Y. B., Puozaa, D. K., Nboyine, J. A., Issah, A.-R., Rasheed, M. A., et al. (2021). Characterization of groundnut (*Arachis hypogaea* l.) test locations using representative testing environments with farmer-preferred traits. *Front. Plant Sci.* 12.

Pandey, M. K., Upadhyaya, H. D., Rathore, A., Vadez, V., Sheshshayee, M. S., Sriswathi, M., et al. (2014). Genomewide association studies for 50 agronomic traits in peanut using the 'Reference set' comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. *PloS One* 9 (8), e105228. doi: 10.1371/journal.pone.0105228

Pasupuleti, J., Ramaiah, V., Rathore, A., Rupakula, A., Reddy, R. K., Waliyar, F., et al. (2013). Genetic analysis of resistance to late leaf spot in interspecific groundnuts. *Euphytica* 193 (1), 13–25. doi: 10.1007/s10681-013-0881-7

Patil, G., Vuong, T. D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., et al. (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnol. J.* 16 (11), 1939–1953. doi: 10.1111/pbi.12929

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15 (1), 8–15. doi: 10.1007/BF02772108

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959. doi: 10.1093/genetics/155.2.945

Ren, W.-L., Wen, Y.-J., Dunwell, J. M., and Zhang, Y.-M. (2018). pKWmEB: integration of kruskal–Wallis test with empirical bayes under polygenic background control for multi-locus genome-wide association study. *Heredity* 120 (3), 208–218. doi: 10.1038/s41437-017-0007-4

Rouse, J. J.W., Haas, R. H., Deering, D. W., Schell, J. A., and Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (Texas 77843 No. E75-10354:Texas A&M University, Remote Sensing Center, College).

Sancho-Adamson, M., Trillas, M. I., Bort, J., Fernandez-Gallego, J. A., and Romanyà, J. (2019). Use of RGB vegetation indexes in assessing early effects of verticillium wilt of olive in asymptomatic plants in high and low fertility scenarios remote sensing. *Remote Sens.* 11 (6), 607. doi: 10.3390/rs11060607

Sandhu, K. S., Merrick, L. F., Sankaran, S., Zhang, Z., and Carter, A. H. (2022). Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. *Front. Genet.* 12. doi: 10.3389/fgene.2021.829131

Sarkar, S., Cazenave, A.-B., Oakes, J., McCall, D., Thomason, W., Abbott, L., et al. (2021a). Aerial high-throughput phenotyping of peanut leaf area index and lateral growth. *Sci. Rep.* 11 (1), 21661. doi: 10.1038/s41598-021-00936-w

Sarkar, S., Ramsey, A. F., Cazenave, A.-B., and Balota, M. (2021b). Peanut leaf wilting estimation from RGB color indices and logistic models. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.658621

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9 (7), 676–682. doi: 10.1038/nmeth.2019

Schindelin, J., Rueden, C. T., Hiner, M. C., and Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* 82 (7-8), 518–529. doi: 10.1002/mrd.22489

Shaibu, A. S., Miko, Z. L., Ajeigbe, H. A., Mohammed, S. G., Usman, A., Muhammad, M. S., et al. (2021). Genome-wide detection of markers associated

with early leaf spot and pod weight in groundnut using SNP and DArT markers. *J. Crop Improvement* 35 (4), 522–535. doi: 10.1080/15427528.2020.1846102

Shaner, G., and Finney, R. E. (1977). The effect of nitrogen fertilization on the expression of slow-mildewing resistance in Knox wheat. *Phytopathology* 67 (8), 1051–1056. doi: 10.1094/Phyto-67-1051

Sie, E. K., Oteng-Frimpong, R., Kassim, Y. B., Puozaa, D. K., Adjebeng-Danquah, J., Masawudu, A. R., et al. (2022). RGB-Image method enables indirect selection for leaf spot resistance and yield estimation in a groundnut breeding program in Western Africa. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.957061

Singh, M. P., Erickson, J. E., Boote, K. J., Tillman, B. L., van Bruggen, A. H. C., and Jones, J. W. (2011). Photosynthetic consequences of late leaf spot differ between two peanut cultivars with variable levels of resistance. *Crop Sci.* 51 (6), 2741–2748. doi: 10.2135/cropsci2011.03.0144

Sinnwell, J. P., Therneau, T. M., and Schaid, D. J. (2014). The *kinship2* r package for pedigree data. *Hum. Hered* 78 (2), 91–93. doi: 10.1159/000363105

Sood, S., Bhardwaj, V., Kaushik, S. K., and Sharma, S. (2020). Prediction based on estimated breeding values using genealogy for tuber yield and late blight resistance in auto-tetraploid potato (*Solanum tuberosum* l.). *Heliyon* 6 (11), e05624. doi: 10.1016/j.heliyon.2020.e05624

Subrahmanyam, P., McDonald, D., Waliyar, F., Reddy, L. J., Nigam, S. N., Gibbons, R. W., et al. (1995). Screening methods and sources of resistance to rust and late leaf spot of groundnut. Information Bulletin no. 47. *Technical Report*. (Andhra Pradesh, India:International Crops Research Institute for the Semi-Arid Tropics, Patancheru).

Tamba, C. L., Ni, Y.-L., and Zhang, Y.-M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PloS Comput. Biol.* 13 (1), e1005357. doi: 10.1371/journal.pcbi.1005357

Tyroler, C. (2018) *Gender considerations for researchers working in groundnuts. USAID feed future*. Available at: https://ftfpeanutlab.caes.uga.edu/content/dam/caes-subsite/ftf-peanut-lab/documents/peanut-lab/Gender%20Considerations.pdf (Accessed 03.07.2022).

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91 (11), 4414–4423. doi: 10.3168/jds.2007-0980

Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* 132 (3), 797–816. doi: 10.1007/s00122-018-3252-x

Varshney, R. K., Sinha, P., Singh, V. K., Kumar, A., Zhang, Q., and Bennetzen, J. L. (2020). 5Gs for crop genetic improvement. *Curr. Opin. Plant Biol.* 56, 190–196. doi: 10.1016/j.pbi.2019.12.004

Vikas, V. K., Pradhan, A. K., Budhlakoti, N., Mishra, D. C., Chandra, T., Bhardwaj, S. C., et al. (2022). Multi-locus genome-wide association studies (ML-GWAS) reveal novel genomic regions associated with seedling and adult plant stage leaf rust resistance in bread wheat (*Triticum aestivum* l.). *Heredity* 128 (6), 434–449. doi: 10.1038/s41437-022-00525-1

Wambugu, P. W., Ndjiondjop, M. N., and Henry, R. J. (2018). Role of genomics in promoting the utilization of plant genetic resources in genebanks. *Briefings Funct. Genomics* 17 (3), 198–206. doi: 10.1093/bfgp/ely014

Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., et al. (2016). Improving power and accuracy of genome-wide association studies *via* a multi-locus mixed linear model methodology. *Sci. Rep.* 6 (1), 19444. doi: 10.1038/srep19444

Wang, Y., Wang, C., Rajaofera, M. J. N., Zhu, L., Liu, W., Zheng, F., et al. (2020). WY7 is a newly identified promoter from the rubber powdery mildew pathogen that regulates exogenous gene expression in both monocots and dicots. *PloS One* 15 (6), e0233911–e0233911. doi: 10.1371/journal.pone.0233911

Wang, S., Zhang, L., Huang, C., and Qiao, N. (2017). *An NDVI-based vegetation phenology is improved to be more consistent with photosynthesis dynamics through applying a light use efficiency model over Boreal high-latitude forests remote sensing*, (Remote Sens) 9, 695. doi: 10.3390/rs9070695

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., and Zemla, J. (2017). Package '*corrplot*'. *Statistician* 56 (316), e24.

Wen, Y.-J., Zhang, H., Ni, Y.-L., Huang, B., Zhang, J., Feng, J.-Y., et al. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief Bioinform.* 19 (4), 700–712. doi: 10.1093/bib/bbw145

Wiesner-Hanks, T., and Nelson, R. (2016). Multiple disease resistance in plants. *Annu. Rev. Phytopathol.* 54 (1), 229–252. doi: 10.1146/annurev-phyto-080615-100037

Wu, C.-D., McNeely, E., Cedeño-Laurent, J. G., Pan, W.-C., Adamkiewicz, G., Dominici, F., et al. (2014). Linking student performance in Massachusetts elementary schools with the "Greenness" of school surroundings using remote sensing. *PloS One* 9 (10), e108548. doi: 10.1371/journal.pone.0108548

Xu, W., Yu, Y., Zhou, Q., Ding, J., Dai, L., Xie, X., et al. (2011). Expression pattern, genomic structure, and promoter analysis of the gene encoding stilbene synthase from Chinese wild *Vitis pseudoreticulata*. *J. Exp. Bot.* 62 (8), 2745–2761. doi: 10.1093/jxb/erq447

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* 13 (2), 187–214. doi: 10.1016/j.molp.2020.01.008

Yao, J., Wang, L., Liu, L., Zhao, C., and Zheng, Y. (2009). Association mapping of agronomic traits on chromosome 2A of wheat. *Genetica* 137 (1), 67–75. doi: 10.1007/s10709-009-9351-5

Yoosefzadeh-Najafabadi, M., Eskandari, M., Belzile, F., and Torkamaneh, D. (2022). "Genome-wide association study statistical models: A review," in *Genome-wide association studies. methods in molecular biology*, vol. 2481. Eds. D. Torkamaneh and F. Belzile (New York: Humana), 43–62.

Zaidi, S.S.-e., Mahas, A., Vanderschuren, H., and Mahfouz, M. M. (2020). Engineering crops of the future: CRISPR approaches to develop climate-resilient and disease-resistant plants. *Genome Biol.* 21 (1), 289. doi: 10.1186/s13059-020-02204-y

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., et al. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* 1 (1), 35. doi: 10.1186/s13007-015-0078-2

Zhang, H., Chu, Y., Dang, P., Tang, Y., Jiang, T., Clevenger, J. P., et al. (2020a). Identification of QTLs for resistance to leaf spots in cultivated peanut (*Arachis hypogaea* l.) through GWAS analysis. *Theor. Appl. Genet.* 133 (7), 2051–2061. doi: 10.1007/s00122-020-03576-2

Zhang, J., Feng, J. Y., Ni, Y. L., Wen, Y. J., Niu, Y., Tamba, C. L., et al. (2017). pLARmEB: integration of least angle regression with empirical bayes for multilocus genome-wide association studies. *Heredity* 118 (6), 517–524. doi: 10.1038/hdy.2017.8

Zhang, Y.-M., Jia, Z., and Dunwell, J. M. (2019). Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00100

Zhang, Y., Liu, P., Zhang, X., Zheng, Q., Chen, M., Ge, F., et al. (2018). Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00611

Zhang, Y., Tamba, C. L., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., et al. (2020b). mrMLM v4.0.2: An r platform for multi-locus genome-wide association studies. *Genomics Proteomics Bioinf.* 18 (4), 481–487. doi: 10.1016/j.gpb.2020.06.006

Zhao, H., Guo, T., Lu, Z., Liu, J., Zhu, S., Qiao, G., et al. (2021). Genome-wide association studies detects candidate genes for wool traits by re-sequencing in Chinese fine-wool sheep. *BMC Genomics* 22 (1), 127–127. doi: 10.1186/s12864-021-07399-3

Zhou, B., Elazab, A., Bort, J., Vergara, O., Serret, M. D., and Araus, J. L. (2015). Low-cost assessment of wheat resistance to yellow rust through conventional RGB images. *Comput. Electron. Agric.* 116, 20–29. doi: 10.1016/j.compag.2015.05.017

Zhou, G., Zhu, Q., Mao, Y., Chen, G., Xue, L., Lu, H., et al. (2021). Multi-locus genome-wide association study and genomic selection of kernel moisture content at the harvest stage in maize. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.697688

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 January 2023
DOI 10.3389/fpls.2023.1102200

OPEN ACCESS

EDITED BY
Weijian Zhuang,
Fujian Agriculture and Forestry University, China

REVIEWED BY
Bin Zhang,
Beijing Vegetable Research Center, China
Jiedan Chen,
Tea Research Institute (CAAS), China
Hafiz Muhammad Rizwan,
Shenzhen University, China

*CORRESPONDENCE
Haiqiu Yu
yuhaiqiu@syau.edu.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 18 November 2022
ACCEPTED 09 January 2023
PUBLISHED 20 January 2023

CITATION
Zhang H, Yu Y, Wang S, Yang J, Ai X, Zhang N, Zhao X, Liu X, Zhong C and Yu H (2023) Genome-wide characterization of phospholipase D family genes in allotetraploid peanut and its diploid progenitors revealed their crucial roles in growth and abiotic stress responses. *Front. Plant Sci.* 14:1102200. doi: 10.3389/fpls.2023.1102200

Genome-wide characterization of phospholipase D family genes in allotetraploid peanut and its diploid progenitors revealed their crucial roles in growth and abiotic stress responses

He Zhang, Yang Yu, Shiyu Wang, Jiaxin Yang, Xin Ai, Nan Zhang, Xinhua Zhao, Xibo Liu, Chao Zhong and Haiqiu Yu*

Peanut Research Institute, College of Agronomy, Shenyang Agricultural University, Shenyang, China

Abiotic stresses such as cold, drought and salinity are the key environmental factors that limit the yield and quality of oil crop peanut. Phospholipase Ds (PLDs) are crucial hydrolyzing enzymes involved in lipid mediated signaling and have valuable functions in plant growth, development and stress tolerance. Here, 22, 22 and 46 *PLD* genes were identified in *Arachis duranensis*, *Arachis ipaensis* and *Arachis hypogaea*, respectively, and divided into α, β, γ, δ, ε, ζ and φ isoforms. Phylogenetic relationships, structural domains and molecular evolution proved the conservation of *PLD*s between allotetraploid peanut and its diploid progenitors. Almost each *A. hypogaea PLD* except for *AhPLDα6B* had a corresponding homolog in *A. duranensis* and *A. ipaensis* genomes. The expansion of *Arachis PLD* gene families were mainly attributed to segmental and tandem duplications under strong purifying selection. Functionally, the most proteins interacting with AhPLDs were crucial components of lipid metabolic pathways, in which ahy-miR3510, ahy-miR3513-3p and ahy-miR3516 might be hub regulators. Furthermore, plenty of *cis*-regulatory elements involved in plant growth and development, hormones and stress responses were identified. The tissue-specific transcription profiling revealed the broad and unique expression patterns of *AhPLD*s in various developmental stages. The qRT-PCR analysis indicated that most *AhPLD*s could be induced by specific or multiple abiotic stresses. Especially, *AhPLDα3A*, *AhPLDα5A*, *AhPLDβ1A*, *AhPLDβ2A* and *AhPLDδ4A* were highly up-regulated under all three abiotic stresses, whereas *AhPLDα9A* was neither expressed in 22 peanut tissues nor induced by any abiotic stresses. This genome-wide study provides a systematic analysis of the *Arachis PLD* gene families and valuable information for further functional study of candidate *AhPLD*s in peanut growth and abiotic stress responses.

KEYWORDS

PLDs, comparative genomics, molecular evolution, lipid metabolic network, growth and development, stress tolerance, *Arachis*

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is one of the most important grain legumes worldwide, ranking second in production among all grain legumes and fifth among oilseeds. China contributes the highest share and India ranks second by 36.48% and 13.97% in world production, respectively (United States Department of Agriculture (USDA) Foreign Agricultural Service, 2020). Peanut seeds, containing 40%-56% oil, 20%-30% protein and 10%-20% carbohydrates, have been primarily used to provide vegetable oil and proteins for human nutrition (Huang et al., 2015). In some Third World countries, peanut shows greater potential to reduce hunger and malnutrition as it is also a good source of quality fodder, calories, vitamins, minerals, and other antioxidant molecules (Krishna et al., 2015). However, majority of the world's peanut is often grown on marginal soils with lesser inputs or intercropped with cereals in many developing countries. A huge gap is developed between its demand and supply due to the constrained quality and productivity resulting from various abiotic factors such as drought, salinity and temperature aberrations (Zhang et al., 2020; Shi et al., 2021). Thus, identification of key genes that can confer abiotic stress tolerance and can be utilized in biotechnological programs to generate improved varieties is an urgent requirement in peanut production (Raza et al., 2022a; Raza et al., 2022b).

One of the most crucial signaling networks for plants in response to multiple stimuli is mediated by lipid molecules (Rizwan et al., 2022a). Environmental cues can increase the activities of phospholipases and trigger the hydrolysis of membrane phospholipids, thus leading to the generation of different classes of lipids and lipid-derived signal messengers (Hou et al., 2016). Phospholipase D (PLD) represents a major family of membrane phospholipases in plants. It acts upon and cleaves the terminal phosphodiester bond of glycerophospholipids to produce phosphatidic acid (PA) and water-soluble free head group (Wang et al., 2012). *PLD* was first identified in plants as early as in 1940s (Hanahan and Chaikoff, 1947), but did not receive detailed attention until the 1980s (Bocckino et al., 1987). In the past two decades, the disclosure of numerous genomic resources facilitates the characterization of plant *PLD*s at a genome-wide level. The *PLD* gene families have been identified in *Arabidopsis* (*Arabidopsis thaliana*), rice (*Oryza sativa*), soybean (*Glycine max*), cotton (*Gossypium* spp.), rape (*Brassica napus* L.), chickpea (*Cicer arietinum*) and other plants successively (Qin and Wang, 2002; Li et al., 2007; Liu et al., 2010; Zhao et al., 2012; Tang et al., 2016a; Lu et al., 2019; Sagar et al., 2021). Commonly, all plant PLDs contain two conserved HKD domains (HxKxxxxD) responsible for hydrolysis activity and can be divided into three sub-classes based on the presence of different domains near the N-terminus: the PLDs with a calcium/phospholipid-binding C2 domain belong to C2-PLD subclass; the PLDs with phox homology (PX) or pleckstrin homology (PH) domains belong to PX/PH-PLD subclass; the PLDs possessing a signal peptide (SP) in place of the usual C2 or PX/PH domains belong to SP-PLD subclass (Tang et al., 2016b). According to sequence characteristics and biochemical properties, the PLD members can be further subdivided into different isoforms, including PLDαs, PLDβs, PLDγs, PLDδs, PLDεs, PLDζs, and/or PLDφs. The isoforms α, β, γ, δ and ε are part of C2-PLDs; ζ and φ isoforms are attached to PX/PH-PLDs and SP-PLDs, respectively (Yao et al., 2021).

The different PLD isoforms are found to have specific subcellular localizations, lipid selectivity and reaction requirements, which lead to their unique cytological and biological functions in particular signaling pathways (Wang, 2005). In *Arabidopsis*, the phenotypic changes caused by the absence of one *PLD* member cannot be offset by the other 11 *PLD*s (Hong et al., 2016). PLDα1 is the predominant PLD and has been proved to regulate drought and salt tolerance by stimulating the accumulation of abscisic acid (ABA) and jasmonic acid (JA) (Li et al., 2009). The repressed expression of *PLDα1* may result in a reduced sensitivity to ABA and drought-induced stomatal closure (Sang et al., 2001). The *PLDα1* knock-out mutants display increased sensitivities to salinity and water deficiency and tend to induce ABA-responsive genes more readily, whereas the overexpression of *PLDα1* have decreased sensitivities (Bargmann et al., 2009). PLDδ is a signal enzyme that can connect microtubules with plasma membrane. The expression levels of *PLDδ* significantly increase under dehydration, salinity, and ABA treatments (Angelini et al., 2018). The knockout of *PLDδ* makes plants sensitive to freezing, heat and oxidative stresses (Li et al., 2004; Liu et al., 2021; Song et al., 2021). PLDζs are similar to animal PLDs in structure and have unique functions in root growth. PLDζ2-derived PA can promote root hair development under phosphorus deficiency by suppressing the vacuolar degradation of auxin efflux carrier PIN-FORMED2 (Lin et al., 2020). PLDζ1 shows crucial roles in both ionic and osmotic stress-induced auxin carrier dynamics during salt stress (Korver et al., 2020). Overall, most of the knowledge about *PLDs* and PLD-mediated lipid signaling has been revealed from studies on model plant *Arabidopsis* and some other crops. But the information about peanut *PLDs* and their roles in regulating developmental features and abiotic stresses yet need to be studied.

Cultivated peanut is a classic natural allotetraploid (AABB, 2n=4x=40). It arose from the interspecific hybridization and subsequent chromosome doubling of two diploid species *Arachis duranensis* (AA, 2n=2x=20) and *Arachis ipaensis* (BB, 2n=2x=20). Recently, the genomes of *A. duranensis*, *A. ipaensis* and *A. hypogaea* have been completely sequenced, which open a new chapter of *Arachis* genomic studies (Bertioli et al., 2016; Bertioli et al., 2019). In present study, we identified 90 *PLD* genes in these three *Arachis* species and evaluated their sequence characteristics, gene structures, conserved domains, phylogenetic relationships, expansion patterns, physiochemical properties of proteins, *cis*-regulatory elements, protein-protein interactions, miRNA-genes regulatory networks and expression profiles in different tissues and multiple abiotic stresses. These fundings may provide a comprehensive characterization of *Arachis PLDs* and lay a theoretical basis for further functional analysis of *PLDs* in regulating abiotic stress tolerance in peanut.

Materials and methods

Identification of *PLD* gene families in peanut and its two progenitors

The genome data (version 1.0) of *A. duranensis*, *A. ipaensis* and *A. hypogaea* were downloaded from the PeanutBase database (https://www.peanutbase.org/peanut_genome). The protein and nucleotide

sequences of *PLDs* in *Arabidopsis* (TAIR10), soybean (version 2.1) and cotton (version 2.0) were retrieved from the Ensemble database (http://plants.ensembl.org/index.html) and used as queries to perform BLASTP and BLASTN searches against the *Arachis* genome data. The Hidden Markov Model (HMM) profiles for *PLDs* (PF00614) were obtained from the Pfam (https://pfam.xfam.org/) and used to perform HMMER searches in the *Arachis* proteome database. The protein sequences identified by both above methods were integrated and parsed by manual editing to remove the redundant. The remaining were considered as candidate *Arachis* PLD proteins and finally submitted to the SMART (http://smart.embl-heidelberg.de/) and InterProScan (http://www.ebi.ac.uk/interpro/) to analyze the presence of characteristic and functional domains. The protein size (aa), molecular weight (Mw), theoretical isoelectric point (pI), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) of *Arachis* PLD proteins were calculated by the ExPASy (https://web.expasy.org/protparam/). The subcellular localization was predicted using the Plant-mPLoc server (http://www.csbio.sjtu.edu.cn/).

Phylogenetic analysis, chromosomal localization and gene nomenclature

The multiple sequence alignment of non-redundant PLD protein sequences in *A. duranensis*, *A. ipaensis*, *A. hypogaea*, *Arabidopsis*, soybean and cotton was performed using the Clustal W with the default settings (Thompson et al., 1994). A phylogenetic tree was constructed using the neighbor joining (NJ) method in MEGA7 software with following parameters: P-distance, pairwise gap deletion and bootstrap with 1000 replicates (Kumar et al., 2016).

The information regarding the detailed positions of *Arachis PLDs* on chromosomes were obtained from the *Arachis* genome database and visualized using the MapChart 2.32 software (Voorrips, 2002). The nomenclature of *Arachis PLD* genes was based on the results of phylogenetic analysis and chromosomal localization. Identified *PLD* genes in *A. duranensis*, *A. ipaensis* and *A. hypogaea* were named as *AdPLD*, *AiPLD* and *AhPLD* followed by roman letters, numbers and capital letters corresponding to their respective orthologs and chromosomal coordinates.

Gene structure, conserved domain and protein motifs

The exon-intron organization of *Arachis PLD* genes was displayed using the Gene Structure Display Server (GSDS2.0) (http://gsds.cbi.pku.edu.cn/) by comparing their coding sequences (CDS) and corresponding genomic sequences. The conserved domains of *Arachis* PLD proteins were identify using the Pfam and SMART tools with the default cut off parameters. The conserved motifs of *Arachis* PLD proteins were investigated by the online tool Multiple Expectation maximization for Motif Elicitation (MEME) (https://meme-suite.org/meme/tools/meme) with following parameters: the maximum motif number was 30; the minimum motif width was 6; and the maximum motif width was 50. The identified protein motifs were further annotated with InterProScan.

Gene duplication events and adaptive evolution analysis

The homologous *PLD* gene pairs were identified based on the results of chromosomal localization, multiple sequence alignments and phylogenetic analysis. The adopted criteria for gene duplication events were that the shared aligned sequence covered over 70% of the longer sequence and the minimum similarity of aligned regions was 70% (Tang et al., 2016b). The tandem duplications have been characterized as multiple members of one gene family occurring within the same or neighboring intergenic regions. The segmental duplications have been defined as homologous genes that result from large-scale events, such as whole genome duplications or large chromosomal region duplications. The duplication events and collinear relationships of *Arachis PLDs* were analyzed using the MCScanX and MCScanX-transposed toolkits (Wang et al., 2013) and visualized by the Circos software (Krzywinski et al., 2009).

The non-synonymous substitution rate (*Ka*) and synonymous substitution rate (*Ks*) of the duplicated gene pairs were calculated to assess the molecular selection effect using the *KaKs*_Calculator 2.0 software (Wang et al., 2010). The gene pairs with the *Ka* and *Ks* value of 0 as well as the *Ks* value more than 2 were discarded, as they might result from the sequence saturation or misalignment. The *Ka*/*Ks* ratio was then calculated to show the selection pressure for duplicated *PLD* genes. The *Ka*/*Ks* ratio >1, <1 or =1 represented positive, negative (purifying selection) and neutral evolution, respectively. The divergence time of *AhPLD* gene pairs was estimated by the formula T= Ks/2r, where r indicates the neutral substitution rate ($r = 8.12 \times 10^{-9}$ Ks yr^{-1}) (Bertioli et al., 2016).

Cis-regulatory elements prediction and interaction networks analysis

The 2000 base pair (bp) DNA sequences in upstream regions of *AhPLDs* were extracted from the peanut genome database and submitted to the PlantCare (http://bioinformatice.psb.ugent.be/webtools/plantcare/) for the identification of putative *cis*-regulatory elements. The results were finally visualized by the TBtools software (Chen et al., 2020).

The STRING database (https://string-db.org/) was used to analyze the interaction relationships between AhPLDs and other proteins with the confidence parameter set at the threshold of 0.4. The well-characterized plant *Arabidopsis* was as the query organism. The predicted protein-protein interaction (PPI) network was displayed by the Cytoscape 3.7.2 software (Shannon et al., 2003).

The targeting relationship between *AhPLDs* and microRNA (miRNA) were predicated using the psRNATarget Server (http://plantgrn.noble.org/psRNATarget/) with the expectation score of 3.5 (Rizwan et al., 2022b). The peanut miRNA sequences were obtained from the miRbase database (http://www.mirbase.org/). The cDNA

sequences of *AhPLDs* were extracted as the candidate targets. The linkage of the predicted miRNAs and corresponding target genes were displayed by the Cytoscape 3.7.2 software (Shannon et al., 2003).

Expression profiles of *AhPLDs* in different tissues based on RNA-sequencing analysis

The RNA-sequencing (RNA-seq) data of 22 peanut tissues at vegetative, reproductive and seed development stages were retrieved from the National Center for Biotechnology Information BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) with the accession number PRJNA291488. The fragments per kilobase of exon model per million mapped reads (FPKM) method in Cufflinks software (http://cufflnks.cbcb.umd.edu/) were used to calculate the transcript abundance. The $\log_2$FPKM values were displayed in the form of heatmaps by the HemI software (Deng et al., 2014).

Plant materials, growth conditions and stress treatments

The allotetraploid peanut (*A. hypogaea* cultivar 'Nonghua 5') planted in large areas of northeast China was used as experimental materials in this study. The seeds were surface sterilized with 3% sodium hypochlorite, and washed five times with distilled water, and kept in dark to germinate. The germinated seeds were sown in round plastic pots filled with clean sandy soil and grown in a climate chamber with a 16 h light (28 °C)/8 h dark (23 °C) cycle, a photosynthetic photon flux density of 400 μmol m^{-2} s^{-1}, and a relative humidity of 70%. After 14 days, the three-leaf seedlings were transferred from sandy soil into hydroponic cultures and grown for 3 days to recover before initiating any stress treatments. For drought and salt stresses, seedlings were incubated in 20% (w/v) polyethylene glycol (PEG-6000) and 250mM NaCl solution, respectively. For cold stress, the temperature of the climate chamber was reduced to 6°C without changing other growth conditions. The second leaves were collected at 0, 6, 12, 24 and 48 h of each treatment with three biological replicates, and frozen in liquid nitrogen immediately and stored at -80°C.

RNA extraction and expression analysis by quantitative real-time RT-PCR

The total RNA for each sample was extracted by TRIzol reagent (Carlsbad, CA, USA) according to the manufacturer's protocol. The RNA concentration was tested using a micro-spectrophotometer (OD260/280). The RNA integrity was tested using the Agilent Bioanalyzer 2100 system. 1 μg of total RNA was used to synthesize first strand cDNA by Takara Reverse Transcription System (TaKaRa, Shuzo, Otsu, Japan). The expression profiles of *AhPLDs* under various abiotic stresses was detected by quantitative real-time PCR (qRT-PCR) using the specific primers as listed in Table S1. The real-time quantification was performed with SYBR Premix Ex Taq™ (TaKaRa, Shuzo, Otsu, Japan) according to the manufacturer's instructions. PCR mixtures (10 μL) contained 1.0 μL cDNA, 0.3 μL each primer, 3.4 μL ddH_2O and 5.0 μL SYBRfi Green Master Mix. The amplification conditions were as follows: 60 s denaturation at 95°C, followed by 40 cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 60 s. Three biological replicates per sample and three technical replicates per biological replicate were used for the analysis. The *AhACT11* was served as the internal reference gene, and the relative expression values were calculated using the $2^{-\Delta\Delta ct}$ method (Chi et al., 2012).

Results

Genome-wide identification of *PLD* genes in peanut and its two progenitors

Using the method as described above, a total of 22 *AdPLD*s, 22 *AiPLD*s and 46 *AhPLD*s were identified in *A. duranensis*, *A. ipaensis* and *A. hypogaea* genomes, respectively (Table S2). Chromosome localization found that the identified *AdPLD*s and *AiPLD*s were unevenly distributed in nine out of ten chromosomes of *A. duranensis* (A genome) and *A. ipaensis* (B genome), respectively (Figure 1A). Among them, A01 and A08 chromosomes possessed the most abundant *PLDs* with each containing five *AdPLD*s, followed by B01, B08 and B09 chromosomes with each containing four *AiPLD*s, but no *PLDs* was found on chromosomes A06 and B06. In *A. hypogaea*, 46 *AhPLD*s were located on 18 out of 20 chromosomes (Figure 1B), of which chromosomes 01, 08, 11 and 18 contained the most *AhPLD*s but chromosomes 06 and 16 had no *AhPLD*.

Furthermore, the sequence characteristics of PLD proteins in three *Arachis* species were analyzed (Table 1). The opening reading frame (ORF) lengths of all the 90 *PLDs* ranged from 1305 bp to 3417 bp, which encoded polypeptides of 507 aa to 1138 aa with predicted MWs ranging from 57.9 kD to 126.94 kD and theoretical p*I*s ranging from 5.16 to 8.44. According to the II values, 49 *Arachis* PLDs were considered as the stable proteins (II<40), and 41 *Arachis* PLDs belonged to the unstable proteins (II>40). Besides, all *Arachis* PLD proteins had high AI values from 69.47 to 86.41 but minus GRAVY values from -0.179 to -0.756, indicating they were stable over a wide temperature range and were hydrophilic and highly soluble in water. Subcellular localization prediction revealed that most *Arachis* PLD proteins were in the cytoplasm, endoplasmic reticulum and vacuole, only a few were in the chloroplast, nucleus and plasma membrane.

Phylogenetic analysis, gene structure and conserved motifs of *Arachis PLD* genes

To explore the phylogenetic relationships and evolutionary patterns of the *PLDs* in peanut and its two progenitors, a neighbor-joining tree was constructed using the full-length protein sequence alignments of 90 *Arachis* PLDs, 12 AtPLDs, 23 GmPLDs and 20 GrPLDs. All the PLDs from different plant species were clearly divided into seven well-supported sub-clades based on the similarity with cotton PLDs, comprising α, β, γ, δ, ε, ζ and φ isoforms (Figure 2). Among them, the α constituted the largest clade containing 38 *Arachis* PLDs (20 AhPLDαs, nine AdPLDαs and AiPLDαs), the δ formed the second largest clade containing 16 *Arachis* PLDs (eight AhPLDδs, four AdPLDδs and AiPLDδs), and the

10.3389/fpls.2023.1102200

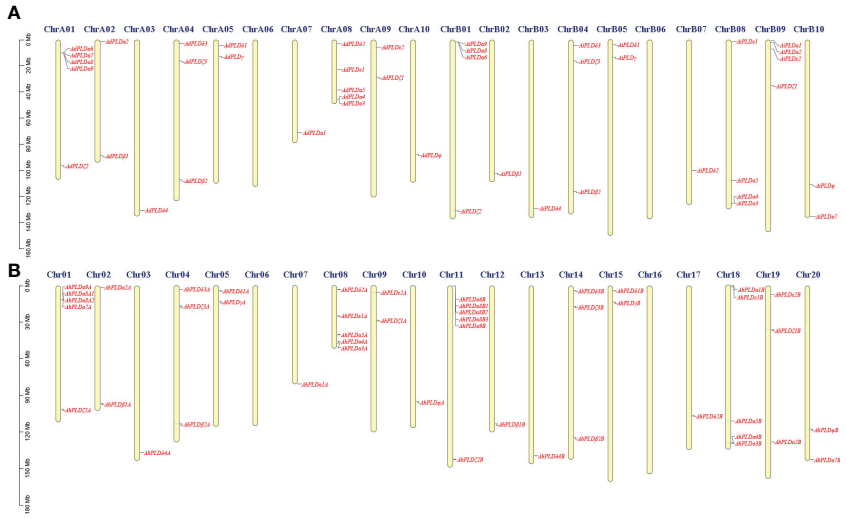


FIGURE 1
Chromosomal distribution of *Arachis PLD* genes. **(A)** Chromosomal distribution of *PLD* genes in *A. duranensis* and *A. ipaensis*. **(B)** Chromosomal distribution of *PLD* genes in *A. hypogaea*. Yellow color bars represent the chromosomes, and the location of *PLD* genes has been marked alongside.

TABLE 1 Detailed information of PLD genes identified in *A. duranensis*, *A. ipaensis* and *A. hypogaea*.

Gene name	PeanutBase ID	Chr	ORF length (bp)	Deduced protein						Subcellular location
				Size (aa)	MW (kDa)	pI	II	AI	GRAVY	
AdPLDα1	Aradu.4Q29Q.1	A07	2292	763	86.76	5.74	40.87	82.66	-0.380	Endoplasmic reticulum; Vacuole
AdPLDα2	Aradu.FS7LG.1	A02	2325	774	87.90	6.20	43.62	80.74	-0.402	Endoplasmic reticulum; Vacuole
AdPLDα3	Aradu.H7I4I.1	A08	2448	815	93.29	6.19	37.99	81.47	-0.485	Endoplasmic reticulum; Vacuole
AdPLDα4	Aradu.H7I4I.2	A08	2469	822	94.43	6.37	35.86	80.47	-0.432	Endoplasmic reticulum; Vacuole
AdPLDα5	Aradu.B9ITB.1	A08	2454	817	93.36	6.22	41.63	84.35	-0.373	Endoplasmic reticulum; Vacuole
AdPLDα6	Aradu.ZT8PK.1	A01	2103	700	81.03	5.21	42.81	69.47	-0.756	Cytoplasm; Nucleus
AdPLDα7	Aradu.PE7A6.1	A01	2388	795	91.26	8.44	38.26	79.55	-0.479	Cytoplasm
AdPLDα8	Aradu.SYE4G.1	A01	2400	799	91.30	6.25	37.42	83.43	-0.448	Cytoplasm
AdPLDα9	Aradu.D9A5M.1	A01	1305	693	78.89	6.67	35.49	82.87	-0.392	Cytoplasm
AdPLDβ1	Aradu.7N75D.1	A02	3417	1138	126.58	8.30	52.61	73.76	-0.447	Chloroplast; Nucleus
AdPLDβ2	Aradu.NT5AR.1	A04	3351	1116	124.57	7.03	46.39	72.31	-0.476	Chloroplast; Nucleus
AdPLDγ	Aradu.X0Z7E.1	A05	2550	849	95.66	6.68	32.27	80.45	-0.409	Cytoplasm
AdPLDδ1	Aradu.UM7P3.1	A05	2592	863	98.18	6.62	34.59	76.95	-0.424	Cytoplasm
AdPLDδ2	Aradu.FN19Y.1	A08	1524	507	57.90	6.58	36.63	78.24	-0.407	Cytoplasm
AdPLDδ3	Aradu.401JR.1	A04	2544	847	95.84	6.59	32.70	83.21	-0.369	Cytoplasm
AdPLDδ4	Aradu.J7XF6.1	A03	2526	841	95.32	6.66	34.17	80.19	-0.394	Cytoplasm
AdPLDε1	Aradu.76V5M.1	A08	2292	763	87.69	6.25	35.22	78.09	-0.423	Cytoplasm
AdPLDε2	Aradu.AA5P6.1	A09	2325	774	88.47	6.15	39.94	78.89	-0.478	Cytoplasm
AdPLDζ1	Aradu.KSI0H.1	A09	3049	1015	115.78	6.38	41.52	86.41	-0.438	Cytoplasm

(Continued)

10.3389/fpls.2023.1102200

TABLE 1 Continued

Gene name	PeanutBase ID	Chr	ORF length (bp)	Deduced protein						Subcellular location
				Size (aa)	MW (kDa)	pI	II	AI	GRAVY	
AdPLDζ2	Aradu.AF7PL.1	A01	3330	1109	126.46	6.53	43.17	82.38	-0.398	Cytoplasm
AdPLDζ3	Aradu.NR2Z3.1	A04	3354	1117	126.94	6.06	49.88	80.84	-0.407	Cytoplasm
AdPLDφ	Aradu.W0B0I.1	A10	1593	530	59.78	6.28	36.67	80.02	-0.179	Plasma membrane
AiPLDα1	Araip.YJB9F.1	B09	2355	784	89.36	5.65	41.53	83.05	-0.400	Endoplasmic reticulum; Vacuole
AiPLDα2	Araip.MYU90.1	B09	2424	807	91.60	6.14	44.57	80.22	-0.408	Endoplasmic reticulum; Vacuole
AiPLDα3	Araip.0C2UG.1	B08	2448	815	93.19	6.27	40.76	80.87	-0.490	Endoplasmic reticulum; Vacuole
AiPLDα4	Araip.0C2UG.2	B08	2136	711	81.35	5.72	38.08	81.66	-0.421	Endoplasmic reticulum; Vacuole
AiPLDα5	Araip.6R3VE.1	B08	2454	817	93.57	6.27	40.50	85.19	-0.363	Endoplasmic reticulum; Vacuole
AiPLDα6	Araip.03APC.2	B01	2394	797	91.57	6.28	46.77	76.21	-0.532	Cytoplasm; Nucleus
AiPLDα7	Araip.I4SWQ.1	B10	2439	812	93.02	6.44	43.50	80.60	-0.527	Cytoplasm
AiPLDα8	Araip.03APC.1	B01	2385	794	90.68	6.38	37.40	81.75	-0.456	Cytoplasm
AiPLDα9	Araip.72Y3Y.1	B01	2121	706	80.63	6.19	31.25	78.73	-0.541	Cytoplasm
AiPLDβ1	Araip.10YD3.1	B02	3312	1103	122.24	8.14	53.50	71.70	-0.499	Chloroplast; Nucleus
AiPLDβ2	Araip.CEE4L.1	B04	3351	1116	124.53	7.02	45.51	72.75	-0.467	Chloroplast; Nucleus
AiPLDγ	Araip.67DP4.1	B05	2550	894	95.54	6.83	31.68	81.59	-0.385	Cytoplasm
AiPLDδ1	Araip.5S0ML.1	B05	2592	863	98.11	6.54	34.67	77.18	-0.410	Cytoplasm
AiPLDδ2	Araip.Z1PQN.1	B07	2547	848	96.41	6.58	37.50	79.65	-0.370	Cytoplasm
AiPLDδ3	Araip.SC81T.1	B04	2544	847	95.93	6.61	32.03	83.32	-0.373	Cytoplasm
AiPLDδ4	Araip.N118V.1	B03	2526	841	95.12	6.73	34.88	80.19	-0.395	Cytoplasm
AiPLDε1	Araip.ZU18G.1	B08	2292	763	87.65	6.36	35.66	76.42	-0.431	Cytoplasm
AiPLDε2	Araip.F3NWG.1	B09	2816	938	106.91	5.91	43.06	78.82	-0.484	Cytoplasm
AiPLDζ1	Araip.QK1BG.1	B09	3237	1078	122.20	6.34	43.03	85.53	-0.435	Cytoplasm
AiPLDζ2	Araip.C34P5.1	B01	2676	891	101.55	7.05	41.19	81.40	-0.408	Cytoplasm
AiPLDζ3	Araip.3PC35.1	B04	2949	982	111.65	6.02	49.11	83.30	-0.388	Cytoplasm
AiPLDφ	Araip.37594.1	B10	1593	530	59.92	6.40	35.75	80.19	-0.198	Plasma membrane
AhPLDα1A	Arahy.PYKR9B.1	07	2424	807	91.88	5.54	42.32	83.82	-0.391	Endoplasmic reticulum; Vacuole
AhPLDα1B	Arahy.JUK473.1	18	2424	807	91.88	5.54	42.32	83.82	-0.391	Endoplasmic reticulum; Vacuole
AhPLDα2A	Arahy.019LBF.1	02	2424	807	91.62	6.11	44.47	79.62	-0.408	Endoplasmic reticulum; Vacuole
AhPLDα2B	Arahy.7PV95K.1	19	2424	807	91.60	6.14	44.57	80.22	-0.408	Endoplasmic reticulum; Vacuole
AhPLDα3A	Arahy.GVK7JC.1	08	2448	815	93.26	6.19	38.21	80.87	-0.491	Endoplasmic reticulum; Vacuole
AhPLDα3B	Arahy.ES1PUL.1	18	2448	815	93.19	6.27	40.76	80.87	-0.490	Endoplasmic reticulum; Vacuole

(Continued)

TABLE 1 Continued

Gene name	PeanutBase ID	Chr	ORF length (bp)	Deduced protein						Subcellular location
				Size (aa)	MW (kDa)	pI	II	AI	GRAVY	
AhPLDα4A	Arahy.A46I61.1	08	2469	822	93.74	6.21	37.28	82.62	-0.413	Endoplasmic reticulum; Vacuole
AhPLDα4B	Arahy.ES1PUL.2	18	2469	822	93.69	6.18	36.38	82.49	-0.405	Endoplasmic reticulum; Vacuole
AhPLDα5A	Arahy.ITK9EF.1	08	2454	817	93.39	6.22	41.54	84.35	-0.374	Endoplasmic reticulum; Vacuole
AhPLDα5B	Arahy.F6RD2F.1	18	2454	817	93.57	6.27	40.50	85.19	-0.363	Endoplasmic reticulum; Vacuole
AhPLDα6B	Arahy.JL8VMK.1	11	2532	843	97.16	5.16	52.00	72.98	-0.652	Cytoplasm; Nucleus
AhPLDα7A	Arahy.80NXVT.1	01	2388	795	91.26	8.44	38.26	79.55	-0.479	Cytoplasm
AhPLDα7B	Arahy.79BS07.1	20	2439	812	93.02	6.44	43.50	80.60	-0.527	Cytoplasm
AhPLDα8A1	Arahy.C52HL3.1	01	2400	799	91.50	6.23	36.04	82.47	-0.447	Cytoplasm
AhPLDα8A2	Arahy.F67ZD9.1	01	2400	799	91.19	6.25	37.41	83.30	-0.445	Cytoplasm
AhPLDα8B1	Arahy.LN5JTA.1	11	2385	794	90.64	6.25	35.28	82.12	-0.443	Cytoplasm
AhPLDα8B2	Arahy.19WSGG.1	11	2385	794	90.76	6.38	37.87	82.62	-0.443	Cytoplasm
AhPLDα8B3	Arahy.6051AB.1	11	2382	793	90.91	6.17	36.78	80.87	-0.468	Cytoplasm
AhPLDα9A	Arahy.KC14LW.1	01	2061	686	77.51	7.31	34.09	82.87	-0.378	Cytoplasm
AhPLDα9B	Arahy.0K14UN.1	11	2121	706	80.63	6.19	31.25	78.73	-0.541	Cytoplasm
AhPLDβ1A	Arahy.3G00JX.1	02	3312	1103	122.33	8.13	53.27	72.31	-0.492	Chloroplast; Nucleus
AhPLDβ1B	Arahy.ZGV9T5.1	12	3312	1103	122.24	8.14	53.50	71.70	-0.499	Chloroplast; Nucleus
AhPLDβ2A	Arahy.Z6D9E8.1	04	3351	1116	124.60	7.03	46.14	72.40	-0.471	Chloroplast; Nucleus
AhPLDβ2B	Arahy.IIM5IQ.1	14	3351	1116	124.53	7.02	45.51	72.75	-0.467	Chloroplast; Nucleus
AhPLDγA	Arahy.NAJ0WW.1	05	2550	849	95.63	6.60	33.72	81.59	-0.393	Cytoplasm
AhPLDγB	Arahy.GAK0W0.1	15	2550	849	95.54	6.83	31.68	81.59	-0.385	Cytoplasm
AhPLDδ1A	Arahy.UU4P1L.1	05	2592	863	98.19	6.62	34.59	76.95	-0.424	Cytoplasm
AhPLDδ1B	Arahy.PYQ6LW.1	15	2592	863	98.18	6.62	34.59	76.95	-0.424	Cytoplasm
AhPLDδ2A	Arahy.J2YZ53.1	08	2547	848	96.38	6.73	37.27	80.09	-0.358	Cytoplasm
AhPLDδ2B	Arahy.I9WKA5.1	17	2547	848	96.41	6.58	37.50	79.65	-0.370	Cytoplasm
AhPLDδ3A	Arahy.NJG8XE.1	04	2544	847	95.84	6.59	32.70	83.21	-0.369	Cytoplasm
AhPLDδ3B	Arahy.XJ1GJM.1	14	2580	859	97.41	6.61	32.95	82.72	-0.369	Cytoplasm
AhPLDδ4A	Arahy.87HDGW.1	03	2529	842	95.33	6.66	34.34	80.10	-0.397	Cytoplasm
AhPLDδ4B	Arahy.UD7LPN.1	13	2526	841	95.12	6.73	34.88	80.19	-0.395	Cytoplasm
AhPLDε1A	Arahy.PTFL6P.1	08	2292	763	87.73	6.27	34.83	77.44	-0.434	Cytoplasm
AhPLDε1B	Arahy.SE7ST2.1	18	2292	763	87.65	6.36	35.66	76.42	-0.431	Cytoplasm
AhPLDε2A	Arahy.1L37A5.1	09	2427	808	92.19	6.25	39.46	76.88	-0.509	Cytoplasm
AhPLDε2B	Arahy.63QAK2.1	19	2433	810	92.57	6.14	41.40	76.33	-0.530	Cytoplasm
AhPLDζ1A	Arahy.T1UA0C.1	09	3231	1076	122.63	6.21	43.65	85.50	-0.443	Cytoplasm
AhPLDζ1B	Arahy.Y1617F.1	19	3231	1076	122.48	6.14	43.42	84.52	-0.446	Cytoplasm
AhPLDζ2A	Arahy.HZZY15.1	01	2676	891	101.72	7.19	41.48	81.40	-0.409	Cytoplasm

(Continued)

10.3389/fpls.2023.1102200

TABLE 1 Continued

Gene name	PeanutBase ID	Chr	ORF length (bp)	Deduced protein						Subcellular location
				Size (aa)	MW (kDa)	pI	II	AI	GRAVY	
AhPLDζ2B	Arahy.M1WUGI.1	11	2676	891	101.55	7.05	41.19	81.40	-0.408	Cytoplasm
AhPLDζ3A	Arahy.I1DXP3.1	04	3354	1117	126.94	6.06	49.88	80.84	-0.407	Cytoplasm
AhPLDζ3B	Arahy.4IL771.1	14	3099	1032	117.75	6.11	48.71	83.04	-0.396	Cytoplasm
AhPLDφA	Arahy.XZ9AUI.1	10	1593	530	59.84	6.40	37.50	80.02	-0.192	Plasma membrane
AhPLDφB	Arahy.G1SIKM.1	20	1593	530	59.92	6.40	35.75	80.19	-0.198	Plasma membrane

φ was the smallest subgroup containing only four *Arachis* PLDs (two AhPLDφs, one AdPLDφs and AiPLDφs). Besides, the β and γ showed closely together and were not explicitly separated from each other. Within the separate clades, isoforms ε and φ were absent in *Arabidopsis*, isoforms γ and φ were absent in soybean, but all seven isoforms were present in *Arachis* and cotton. Obviously, *Arachis* PLDs showed closer to GmPLDs but more distant from AtPLDs. According to the phylogenetic relationships with orthologs in soybean, cotton and *Arabidopsis*, as well as the physical location on chromosomes, 22 AdPLDs and 22 AiPLDs were renamed as Ad/AiPLDα1-9, Ad/AiPLDβ1, Ad/AiPLDβ2, Ad/AiPLDγ, Ad/AiPLDδ1-4, Ad/AiPLDε1, Ad/AiPLDε2, Ad/AiPLDζ1-3 and Ad/AiPLDφ. The 46 AhPLDs were renamed as AhPLDα1A/1B-5A/5B, AhPLDα6B, AhPLDα7A/7B, AhPLDα8A1/8A2/8B1/8B2/8B3, AhPLDα9A/9B, AhPLDβ1A/1B, AhPLDβ2A/2B, AhPLDγA/B, AhPLDδ1A/1B-4A/4B, AhPLDε1A/1B, AhPLDε2A/2B, AhPLDζ1A/1B- 3A/3B and AhPLDφA/B.

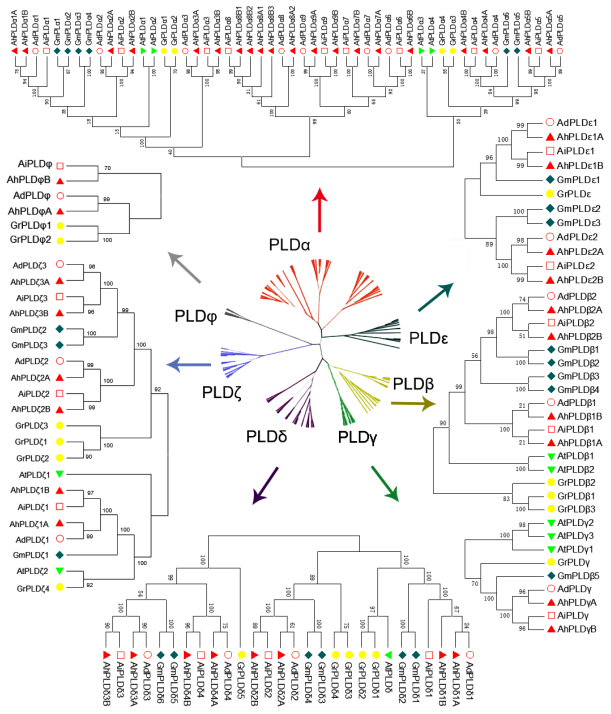


FIGURE 2
Phylogenetic tree of *PLD* gene family with bootstrap values among three *Arachis* species, *Arabidopsis*, soybean and cotton. The red, yellow, green, purple, dark green, blue and gray branches represent α, β, γ, δ, ε, ζ and ϕ isoforms, respectively. The red circles, red squares, red triangles, green triangles, dark green diamonds and yellow circles represent the proteins of *A. duranensis* (Ad), *A. ipaensis* (Ai), *A. hypogaea* (Ah), *Arabidopsis* (At), soybean (Gm) and cotton (Gr), respectively.

To confirm the authenticity and integrity of identified *Arachis PLD* genes, their functional domains were analyzed by searching the Pfam and SMART databases. As expected, all *Arachis* PLD proteins possessed two structurally conserved HKD domains at C-terminal and were divided into three subgroups (Figure 3A). Isoforms α, β, γ, δ and ε contained C2 domain at the N-terminal and belonged to the C2-PLDs. PLDζs contained PX and/or PH domain at the N-terminal and belonged to the PX/PH-PLDs. PLDφs contained signal peptide at the N-terminal instead of C2 or PX/PH domain and belonged to the SP-PLDs. Furthermore, the distribution of exons and introns within each PLD protein was also analyzed. As shown in Figure 3B, all the 90 *Arachis* PLDs were comprised of multiple exons and introns. The number of introns determined for *Arachis* PLDs ranged from one in AdPLDα6, AdPLDα9, AiPLDα9 and AhPLDα9B to 20 in AiPLDζ1. The members of all *Arachis* PLD isoforms showed similar exon-intron structure withing their respective subgroups. It was consistent with the phylogenetic classification depicted in the left panel of Figure 3A. For instance, both α and ε subgroups included members with one to five introns, the members in β/γ and δ subgroups possessed eight or nine introns (except AdPLDδ2 that had six introns), the members in ζ subgroup contained 15 to 20 introns, and all six members in φ subgroup had six introns.

To reveal the typical domain characteristics of *Arachis* PLD subgroups, the conservation of amino acid residues in functional domains were analyzed based on the alignments of these PLDs. A total of 30 distinct motifs designated as Motif 1 to Motif 30 were identified (Figures 4, S1; Table S3). The members within the same subfamilies were usually shared similar motif composition, for example, C2-PLDs possessed 17 to 25 motifs, of which Motif 1-11, Motif 13, Motif 15 and Motif 25 were highly conserved; PX/PH-PLDs possessed 18 motifs, including Motif 1-9, Motif 13, Motif 15, Motif 19, Motif 24 and Motif 26-30; whereas SP-PLDs only contained five motifs, including Motif 1, Motif 4, Motif 15, Motif 28 and Motif 29. By comparison, the Motif 26 and Motif 27 were only present in the C-terminus of PX/PH-PLDs but absent in C2-PLDs and SP-PLDs, suggesting these two motifs were specific to the PX and PH domains. Moreover, almost all these motifs could be matched to the annotated motifs in InterProScan and were functionally associated with PLD activity. The Motif 1, Motif 6, Motif 14 and

10.3389/fpls.2023.1102200

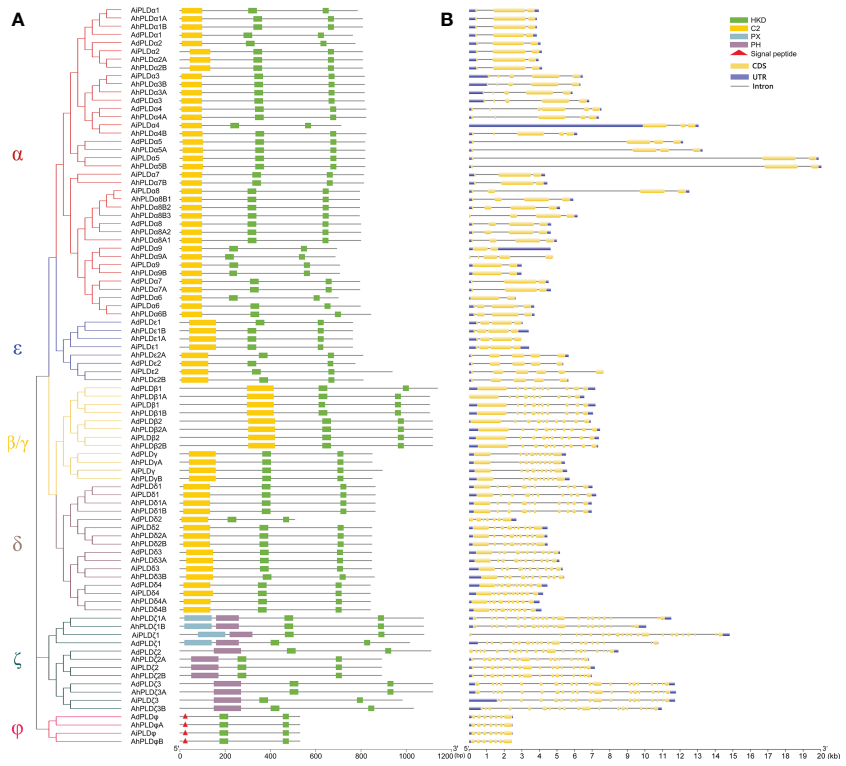


FIGURE 3
Phylogenetic relationship, conserved domains and gene structure of *PLD* genes in peanut and its two progenitors. **(A)** Schematic representation of the phylogenetic tree and conserved domain of *PLD* genes in three *Arachis* species. The red, blue, yellow, purple, dark green and pink branches on phylogenetic tree represent the α, ε, β/γ, δ, ζ and φ isoforms, respectively. The yellow, green, blue and purple rectangles represent the C2, HKD, PX and PH domains, respectively. The red triangles represent the signal peptides. **(B)** Exon-intron structure of *PLD* genes in three *Arachis* species. The blue boxes represent exons. The yellow boxes represent coding sequences. The black lines represent introns.

Motif 29 were annotated as the conserved HKD domains (HxKxxxxD), especially the Motif 1 was uniformly observed in all *Arachis* PLDs. The Motif 4, also observed in all PLD proteins, contained a core triplet of amino acids "ERF" followed by a highly conserved hydrophobic region "VYVVV". The Motif 2 contained a regular-expression sequence "IYIENQ[F/Y]F", of which the seventh amino acid, Phenylalanine (F), appeared in all PX/PH-PLDs but was often substituted by the Tyrosine (Y) in C2-PLDs. The Motif 3 contained the conserved amino acid"xxGPRxPWHDxHx xxxGPAxxDVLTNFExRWRKxGx", which was considered as the binding sites of PIP2.

Gene duplication events and molecular evolution of *Arachis PLD* genes

To clarify the roles of gene duplications in the expansion of *Arachis PLD* gene families, the synteny analysis including tandem duplication and whole genome duplication (WGD)/segmental duplication was performed based on the multiple and pairwise alignments of *Arachis PLDs*. As a result, a total of 131 orthologous/paralogous gene pairs were identified, of which 54 pairs were predicted to form paralogous gene pairs within the *A. duranensis*, *A. ipaensis* and *A. hypogaea* genomes, and nine pairs underwent the tandem duplications (*AdPLDα3-AdPLDα4*, *AdPLDα8-AdPLDα9*, *AiPLDα1-AiPLDα2*, *AiPLDα3-AiPLDα4*, *AiPLDα6-AiPLDα8-AiPLDα9*, *AhPLDα3A-AhPLDα4A*, *AhPLDα3B-AhPLDα4B*, *AhPLDα8A1-AhPLDα8A2-AhPLDα9A*, *AhPLDα6B-AhPLDα8B1-AhPLDα8B2-AhPLDα8B3-AhPLDα9B*) (Figure 5A; Table S4). Besides, 25, 28 and 24 segmental duplications were found in groups *AhPLDs-AdPLDs*, *AhPLDs-AiPLDs* and *AdPLDs-AiPLDs* respectively, but one group only contained *AiPLDα6* homoeologous gene (*AhPLDα6B*), suggesting the corresponding *AhPLDα6A* had lost during peanut evolution (Figure 5B; Table S4). Consequently, we presumed that the putative gene duplication events were main causes of *Arachis PLD* gene family expansion, and homoeologous gene pairs were generally raised from tandem or WGD/segmental duplication before polyploidization involved in evolution process.

To investigate which type of selection pressure had been involved in *PLD* gene divergence after duplication, the *Ka*/*Ks* ratios of duplicated *Arachis PLD* gene pairs were calculated on the basis of coding sequences (Table S5). The resulting pairwise comparison data showed that most homoeologous gene pairs had *Ka*/*Ks* ratios of < 1, indicating these gene pairs might have undergone strong purifying selection pressure with limited functional divergence that occurred after tandem or WGD/segmental duplication. Only three gene pairs (*AhPLDα8A2-AdPLDα8*, *AhPLDδ3B-AiPLDδ3* and *AhPLDδ3A-AhPLDδ3B*) had *Ka*/*Ks* ratios of > 1, suggesting these gene pairs might have experienced relatively rapid evolution following duplication. Meanwhile, according to Ks values, the divergence time of duplicated *Arachis PLD* gene pairs were estimated. As a result, these tandem and segmental duplications may occur in 3.02-70.62 and 0.10-119.83 Mya, respectively (Table S5).

10.3389/fpls.2023.1102200

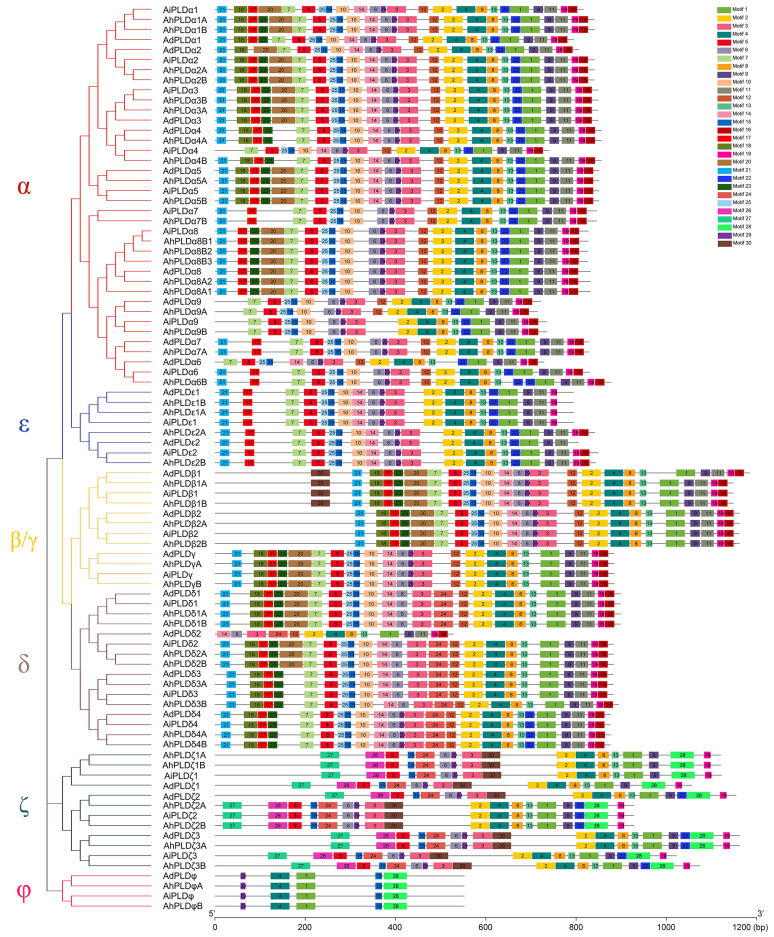


FIGURE 4
Motif composition of PLD proteins in peanut and its two progenitors. The red, blue, yellow, purple, dark green and pink branches on phylogenetic tree represent the α, ε, β/γ, δ, ζ and ϕ isoforms, respectively. The motifs, numbers 1-30, are displayed in different colored boxes.

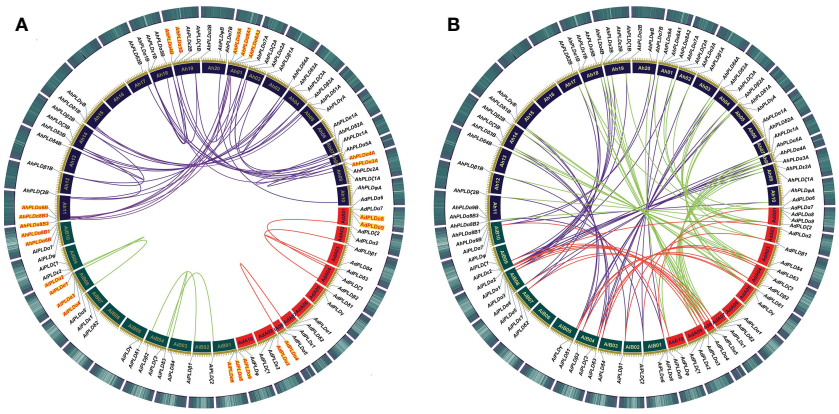


FIGURE 5
Syntenic relationships among *PLD* genes in peanut and its two progenitors. **(A)** Syntenic relationships of *PLD* genes within *A. duranensis*, *A. ipaensis* and *A. hypogaea*, respectively. **(B)** Syntenic relationships of *PLD* genes between *A. duranensis*, *A. ipaensis* and *A. hypogaea*. The chromosomes of *A. duranensis*, *A. ipaensis* and *A. hypogaea* were shown with red, dark green and dark purple colors, respectively. The putative homologous *PLD* genes are linked by different colored lines. The tandem duplicated *PLD* genes are marked with red font next to the chromosomes.

Cis-regulatory elements prediction in *AhPLDs* promotors

To better understand the transcriptional regulation and potential functions of peanut *PLD* genes, the *cis*-regulatory elements present in promoters (2000 bp of 5' upstream regions) of *AhPLD* genes were identified. Totally, 39 functional *cis*-elements related to growth and development, plant hormones and stress responses were obtained from 46 *AhPLD*s (Table S6). All *AhPLD* promoters had variable number of *cis*-regulatory elements and most of them were present in multiples (Figure 6). Among the 25 growth and development-related elements, light-responsive elements (87.59%), such as Box 4, G-box, GT1-motif, TCT-motif and GATA-motif, accounted for the most and widely distributed in all *AhPLD* promoters; others mainly included the circadian element (3.28%) related to circadian control, the O_2-site (2.57%) related to zein metabolism regulation, the CAT-box (2.00%) related to meristem expression and the GCN4_motif (1.85%) related to endosperm expression.

The nine hormone-responsive *cis*-elements were related to five plant hormones. Among them, abscisic acid responsiveness element (35.84%, ABRE) was identified in 39 *AhPLD*s; MeJA-responsive elements (37.99%, CGTCA-motif/TGACG-motif) were identified in 27 *AhPLD*s; gibberellin-responsive elements (12.19%, GARE-motif/P-box/TATC-box) were identified in 17 *AhPLD*s; salicylic acid-responsive element (7.17%, TCA-element) was identified in 15 *AhPLD*s; auxin-responsive elements (6.81%, TGA-element/AuxRR-core) were identified in 14 *AhPLD*s (Figure 6; Table S6). There were three *AhPLD*s (*AhPLDα5A*, *AhPLDζ3A* and *AhPLDζ3B*) contained the five hormone-responsive elements above-mentioned, but *AhPLDα9A* did not contain any hormone-responsive elements. Besides, the promoter regions of eight *AhPLD*s (*AhPLDα1A*, *AhPLDα1B*, *AhPLDα3A*, *AhPLDα3B*, *AhPLDα8B1*, *AhPLDα8B2*, *AhPLDα9B* and *AhPLDφB*) only contained ABA-responsive elements, whereas *AhPLDε2B* and *AhPLDφA* only contained MeJA- and auxin- responsive elements, respectively. However, the ethylene-responsive *cis*-elements were not found in any *AhPLD* promoters.

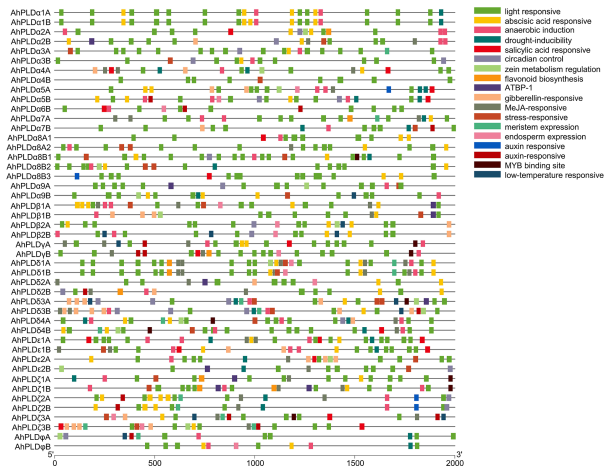


FIGURE 6
Cis-regulatory elements in the promoters of peanut *PLD* genes. Various *cis*-regulatory elements are displayed in different colored boxes.

Moreover, there were five *cis*-elements found to be responsive to various stresses, including anaerobic element (42.76%, ARE), MYB binding site involved in drought-inducibility (27.59%, MBS), low temperature-responsive (8.28%, LTR) element and stress-responsive element (21.38%, TC-rich repeats) (Figure 6; Table S6). Except for *AhPLDα8B3* and *AhPLDδ2A*, almost all *AhPLD*s contained at least one stress-responsive element, indicating *AhPLD*s played an important role not only in peanut growth and development, but also in stress responses. The simultaneous appearance of stress and hormone related *cis*-elements in some *AhPLD*s suggesting that *AhPLD*s could mediate the cross-talk of stress and hormone signaling pathways and may have potential role in hormone mediated abiotic stress signaling in peanut.

Protein-protein interactions and miRNA-genes regulatory networks of peanut PLDs

To elucidate the metabolic regulation network mediated by PLDs in peanut and further understand the biological function of *AhPLDs*, the protein-protein interactions (PPIs) of AhPLDs were analyzed using ortholog-based method. Totally, 24 AhPLDs had orthologous relationships with 12 *Arabidopsis* PLDs and interacted with 55 functional proteins (Figure 7A; Table S7). As expected, most proteins were the functionally validated components of lipid biosynthetic and lipid metabolic processes, such as amino alcohol-phosphotransferase (AAPT), diacylglycerol O-acyltransferase 1 (DGAT1), diacylglycerol kinases (DGKs), lysophosphatidyl acyltransferase 2 (LPAT2) and triglyceride lipase (TGL4). Some others like phospholipase C1 (PLC1), phosphatidylinositol 3-kinase (PI3K), phosphatidylinositol 4-OH kinase 1 (PI4K1) and phosphatidylinositol-4-phosphate 5-kinase 1 (PIP5K1) mainly participated in the phosphatidylinositol-mediated signaling. Besides, several AhPLDs could interact with phospholipid/glycerol acyltransferase family protein ATS2, G protein alpha subunit 1 (GPA1), lipoxygenase 1 (LOX1) and plasma-membrane choline transporter-like protein CTL1. These proteins may regulate plant root development, pollen tube growth, post-embryonic development, seed development and other system development processes.

Moreover, the pathways in response to stress or abiotic stimulus were also significantly enriched by abundant proteins that interacted with AhPLDs, for example, there were five proteins (DGAT1, DGK2, PLC1, respiratory burst oxidase homologue D (RBOHD) and phospholipid-transporting ATPase (PTAT-1)) involved in the response to temperature stimulus, four proteins (DGAT1, PI3K, PLC1 and aldehyde dehydrogenase 3H1 (ALDH3H1)) involved in the response to salt stress, four proteins (DGAT1, PI3K, PLC1 and RBOHD) involved in the response to osmotic stress, three proteins (phosphatidate phosphatase PAH1, monogalactosyldiacylglycerol synthase type C (MGDC) and sulfoquinovosyldiacylglycerol 2 (SQD2)) involved in the response to phosphate starvation, and six proteins (four DGKs, phosphatidic acid phosphatase 1 (LPP1) and RBOHD) involved in the defense response. Notably, there were 14 predicted proteins only interacted with AhPLDζ2A/2B and mainly responsible for lipid transport and phospholipid translocation.

Furthermore, the putative miRNA-targeted *AhPLDs* were predicted to obtain the potential association between lipid

10.3389/fpls.2023.1102200

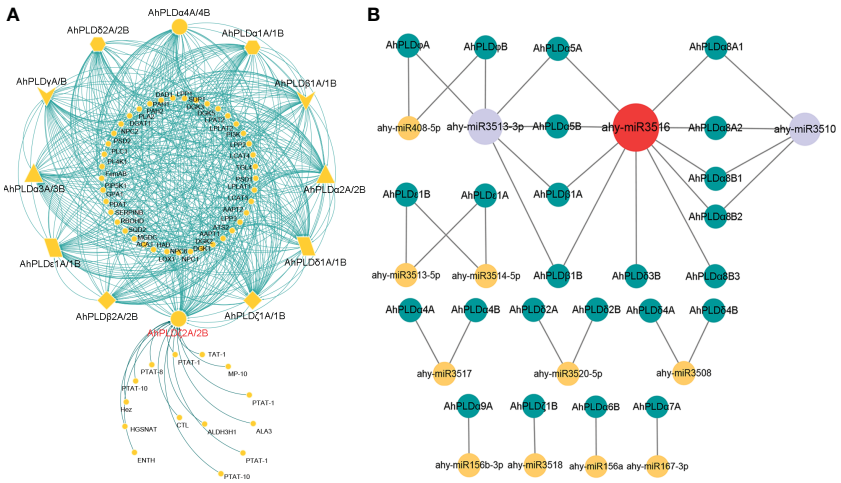


FIGURE 7
Interactions and regulations of peanut *PLD* genes with other proteins and miRNAs. **(A)** Protein-protein interaction network of peanut PLD proteins with other proteins. The yellow geometric figures represent peanut PLD proteins. The little yellow dots represent other proteins that interact with AhPLDs. The interaction relationships are shown by green lines. **(B)** Regulatory network of putative miRNAs and their targeted *AhPLDs*. The yellow, purple and red circles represented miRNAs. The green circles represent targeted *AhPLDs*. The putative regulatory relationships between miRNAs and their targeted *AhPLDs* are shown in grey lines.

metabolic pathways and miRNA regulations. As shown in Figure 7B, a total of 24 *AhPLDs* were targeted by 13 miRNAs, of which 12 *AhPLDs* were targeted by only one miRNA, and 12 *AhPLDs* were targeted by two miRNAs. These miRNAs belonged to 12 different families. The comprehensive data of all miRNAs targeted sites/genes was given in Table S8. Most miRNAs only targeted one or two *AhPLDs*, but ahy-miR3516, ahy-miR3510 and ahy-miR3513-3p could target ten (*AhPLDα5A/5B*, *AhPLDα8A1/8A2/8B1/8B2/8B3*, *AhPLDβ1A/1B* and *AhPLDδ3B*), six (*AhPLDα5A/5B*, *AhPLDβ1A/1B* and *AhPLDφA/B*) and four *AhPLDs* (*AhPLDα8A1/8A2/8B1/8B2*), respectively, indicating these three miRNAs may be crucial in regulating the lipid metabolic processes.

Spatial expression profiles of *PLD* genes in peanut

The tissue-specific expression profiles of *AhPLDs* were investigated by using the RNA-Seq data from Clevenger et al. (2016). A total of 22 different tissues encompassing almost all peanut tissues and developmental stages were analyzed (Figure 8; Table S9). The results showed that the expression patterns of 46 *AhPLDs* in various tissues were quite different, though the expression levels of most homologous copies from the A genome and the B genome (such as *AhPLDα1A* and *AhPLDα1B*) were similar because of their extremely high similarity of mRNAs and transcript sizes. The *AhPLDα1A/1B*, *AhPLDα7B*, *AhPLDβ2A/2B*, *AhPLDγA/B*, *AhPLDδ1A/1B*, *AhPLDδ3A/3B*, *AhPLDζ3A/3B* and *AhPLDφA/B* were expressed in all 22 tissues, especially the *AhPLDδ3A/3B* exhibited higher expressions in all tissues. Conversely, *AhPLDα9A/9B*, the members of PLDα subgroup, had weak or even undetected expression in any tissues. Besides, almost all *AhPLDs* except for *AhPLDα9A/9B* were highly expressed in all above-ground tissues during the vegetative growth stage, indicating *AhPLDs* might play essential roles in the growth and development of peanut seedlings. The expression levels of *AhPLDεs*, *AhPLDζs* and *AhPLDφs* in pods and seeds were significantly lower than in leaves, shoots, roots and flowers, suggesting these *AhPLDs* might mainly involve in peanut seedling growth, root elongation and flowering.

Moreover, some *AhPLDs* displayed tissue-specific or preferential expression patterns, for example, the *AhPLDα1A/1B* showed preferential expressions in flowers and pod development; the *AhPLDα2A/2B* had higher expressions in shoot at the reproductive growth stage but hardly expressed in pod or seed developmental processes; the *AhPLDα6B* preferred to express in root and nodule; the

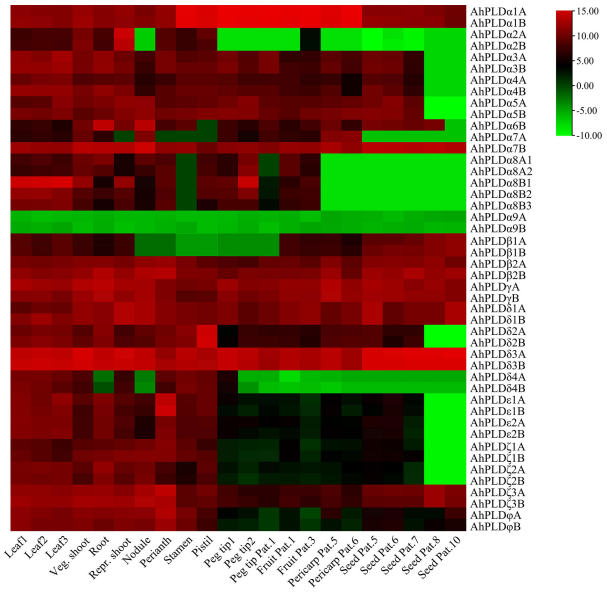


FIGURE 8
Expression profiles of 46 *AhPLDs* in 22 different tissues. The X axis represents various tissues. The Y axis represents 46 *AhPLDs*. The color scale represents Log2 expression values (FPKM).

10.3389/fpls.2023.1102200

AhPLDα7B maintained obvious expression levels in all 22 tissues, but its homologue *AhPLDα7A* had almost undetected expression in flowers and nearly mature seeds; the *AhPLDδ2A/2B* and *AhPLDε1A/1B* showed highest expression levels in pistil and perianth, respectively, indicating their significant roles in the floral organ development in peanut; but only a few *AhPLDs* were highly expressed in seed Pat. 8 and seed Pat. 10.

Expression profiles of peanut *PLD* genes in response to abiotic stresses

To explore the potential involvement of peanut *PLD* genes in abiotic stresses, their expression profiles under cold, drought and salt were investigated by qRT-PCR (Figure 9; Table S10). Since the paralogous pairs of *AhPLDs* displayed highly similar nucleotide sequence and expression pattern in 22 different tissues, only *AhPLDs* from the A subgenome were selected for analysis. Similarly, most *AhPLDs* within the same phylogenetic subgroup were also similar in nucleotide sequence and expression pattern, so one representative *AhPLD* in each subclade was chosen for analysis, such as *AhPLDα8A1*. Besides, given that *AhPLDα7B* showed distinct expression pattern in 22 different tissues compared with its homologue *AhPLDα7A*, it was also included. Finally, the expression profiles of 23 *AhPLDs* under three major abiotic stresses were analyzed. The results in Figure 9 showed that all *AhPLDs* except for *AhPLDα9A* could be induced by at least one abiotic stress, of which 13 *AhPLDs* (*AhPLDα3A*, *AhPLDα5A-7A*, *AhPLDβ1A*, *AhPLDβ2A*, *AhPLDδ1A*, *AhPLDδ3A*, *AhPLDδ4A*, *AhPLDζ1A-3A* and *AhPLDφA*) were found to be induced by all three abiotic stresses commonly, and seven *AhPLDs* (*AhPLDα1A*, *AhPLDα2A*, *AhPLDα7B*, *AhPLDα8A1*, *AhPLDδ2A*, *AhPLDε1A* and *AhPLDε2A*) could be induced by both cold and drought stresses, and one *AhPLD* gene (*AhPLDα4A*) could be induced by both cold and salt stresses, and *AhPLDγA* was only

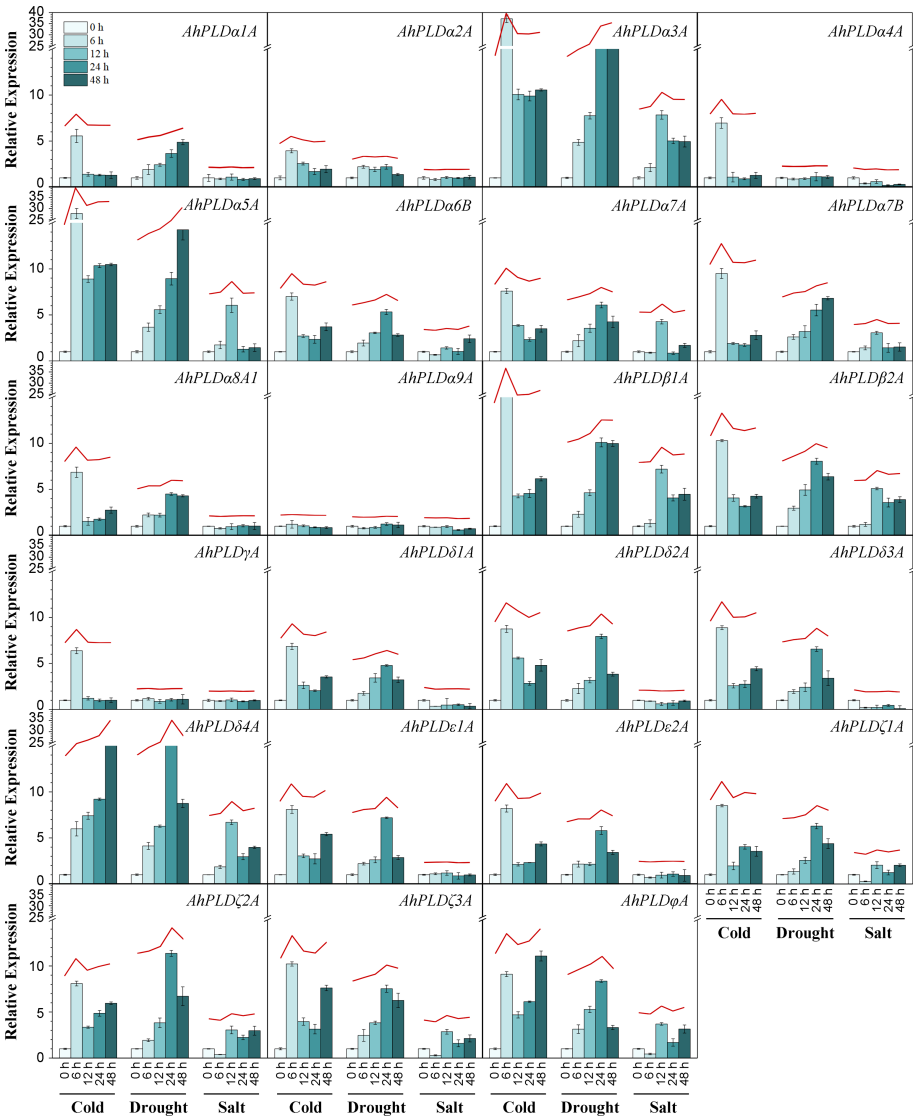


FIGURE 9
Expression profiles of *AhPLDs* under abiotic stresses. The expression levels of *AhPLDs* under cold, drought and salt were generated using qRT-PCR. The X axis represents different treatments. The Y axis represents relative expression levels of *AhPLDs*. Each bar represents average of three replicates. Standard error is indicated by error bars. The red lines above bars represent the overall trends of *AhPLDs* expressions under different abiotic stresses.

induced by cold stress. Surprisingly, almost all *AhPLDs* induced by cold and drought were up-regulated compared with controls.

Under cold stress, the expressions of *AhPLDα3A*, *AhPLDα5A*, *AhPLDβ1A*, *AhPLDδ4A*, *AhPLDζ3A* and *AhPLDφA* were most significantly up-regulated. Among them, *AhPLDδ4A* showed a continuous increase with the prolongation of cold time, the fold change (FC) at 48 h of cold stress reached 23.48; *AhPLDα3A*, *AhPLDα5A* and *AhPLDβ1A* showed a trend of increasing (0-6 h) first and then decreasing (6-48 h), peaking at 6 h of cold stress with the FC values of 37.19, 27.86 and 16.11, respectively; similarly, *AhPLDζ3A* and *AhPLDφA* also showed a trend of increasing first and then decreasing, while their expression level was increased again after 48 h of cold stress. Under drought stress, the expressions of most *AhPLDs* were significantly up-regulated in the early stage (24 h) of stress condition, but their expressions did not continue to increase with the duration of stress, typically like *AhPLDδ4A*, *AhPLDζ2A* and *AhPLDφA*; while such *AhPLDs* as *AhPLDα3A* (19.36 FC), *AhPLDα5A* (14.25 FC) and *AhPLDβ1A* (9.99 FC) kept up-regulated expression to 48 h of drought stress. Compared with cold and drought, the transcript levels of *AhPLDs* under salt tress were lower, some of which were even not differentially expressed or down-regulated. For example, *AhPLDα3A* (7.84 FC), *AhPLDα5A* (6.05 FC), *AhPLDβ1A* (7.20 FC), *AhPLDβ2A* (5.09 FC) and *AhPLDδ4A* (6.70 FC) were found to be induced to highly express exclusively at 12 h of salt treatment, the expressions of *AhPLDζ1A*, *AhPLDζ2A*, *AhPLDζ3A* and *AhPLDφA* were down-regulated at 6 h but up-regulated with the duration of salt stress, whereas *AhPLDα4A*, *AhPLDδ1A* and *AhPLDδ3A* were continuously down-regulated with the $\log_2$FC values of -1.74, -1.47 and -3.32 at 48 h, respectively. Above fundings suggested that *AhPLDs* were potentially involved in signaling triggered by multiple abiotic stresses, and many of them may act as positive regulators of cold and drought stresses in peanut, especially such as *AhPLDα3A*, *AhPLDα5A*, *AhPLDβ1A* and *AhPLDδ4A*.

Discussion

As one of the most representative phospholipases in plants, PLD can catalyze the hydrolysis of membrane lipids for lipid remodeling and mediate many physiological processes in plant growth, development and responses to abiotic stresses (Wang, 2005). The identification and functional validation of the *PLDs* may hold the promise to breed the improved crops with excellent agronomic traits and stress tolerances to combat the challenge of global climate change. In this study, we obtained 46 *AhPLDs* in allotetraploid peanut as well as 22 *AdPLDs* and 22 *AiPLDs* in its two diploid progenitors *A. duranensis* and *A. ipaensis*, respectively (Table 1). Obviously, the number of *PLDs* in *A. hypogaea* was greater than already reported plant species (Qin and Wang, 2002; Li et al., 2007; Liu et al., 2010; Zhao et al., 2012; Lu et al., 2019; Sagar et al., 2021), which might be due to the hybridization of *A. duranensis* and *A. ipaensis* with succeeding polyploidization. The similar results are also reported in cotton (Tang et al., 2016a; Tang et al., 2016b). Based on the conserved structural domains and sequence properties, these 90 *Arachis PLD* genes could be divided into seven subgroups with distinct biochemical, regulatory and catalytic properties, including α, β, γ, δ, ε, ζ and φ isoforms but excluding isoform κ (Figure 2). At present, PLDκ has only been identified in rice and encodes a C2-PLD (Li et al., 2007).

The specific patterns of retention or dispersion of family genes are vital clues to understand the homoeologous chromosome interaction and genetic evolution during plant allo-polyploidization. In this study, the integrated results of phylogeny, gene structure and gene duplication showed that most allotetraploid peanut *PLDs* had at least two homoeologous copies in A and B subgenomes as well as orthologous genes with its progenitors (Figures 3–5). It proved that a specific large-scale genome duplication event has occurred during the peanut origin, with the segmental duplication and tandem duplication jointly taking place at some locations (Figure 5; Table S4). But the orthologous gene of *AhPLDα6B* was not found in peanut genome, suggesting a *AhPLDα6A* loss event occurred during peanut evolution. Like *Arabidopsis* and other plant species, the small clade of PLDγs was closer to the clades of PLDβs and PLDδs on the phylogenetic tree (Figure 2), and they contained similar numbers of exons and introns (Figure 3). This suggested that the isoforms β, γ and δ might have originated as one group in plants but be separated into different functional isoforms during the evolution due to the possible gene duplication events. Similarly, α and ε isoforms might also originate from the common ancestor. But PLDζs and PLDφs had dissimilar intron numbers and belonged to the PX/PH-PLD and SP-PLD subfamilies, respectively, suggesting their convergent evolution *via* two independent evolutionary paths. Besides, the conservation of isoform φ could be also proved by the rate of molecular evolution, in which the *Ka*/*Ks* ratios of gene pairs in subgroup PLDφ were smaller than those in other subgroups (Table S5).

Based on the sequence similarity, structural conservation and close evolutionary relationships of ortholog genes among different species, a functional conservation of peanut *PLDs* might also be predicted. The PPI network of AhPLDs showed that most proteins that interacted with AhPLDs (like non-specific phospholipase C (NPC), DGAT1 and DGKs) were the major components of lipid biosynthesis, lipid metabolism and lipid signaling pathways (Figure 7A; Table S7), which proved the central and conservative functions of peanut PLDs in lipid-related biological processes. Currently, it has become increasingly difficult to find an area of cell biology in which lipids do not have important roles as signaling and regulatory molecules (Hou et al., 2016). For example, *DGAT1*, *DGK2*, *DGK3* and *DGK5* could enhance plant cold tolerance by balancing triacylglycerol (TAG) and PA production (Tan et al., 2018); *NPC4* knockout plants displayed increased sensitivity to salt stress in root elongation, seedling biomass, and seed germination (Kocourkova et al., 2011); PLDζs could hydrolyze phosphatidylcholine to supply phosphorus for cell metabolism and DAG for galactolipid synthesis during phosphorus starvation (Lin et al., 2020). Besides, AhPLDs also interacted with other proteins such as protease inhibitors, GPA1 and RBOHD (Figure 7A, Table S7). These interactions can regulate PLD activity and intracellular locations, thus affecting cellular functions. For instance, PLDα1 could interact with the GPA1 through its DRY motif to mediate ABA signaling in *Arabidopsis* (Zhao and Wang, 2004). Furthermore, there were 24 *AhPLDs* targeted by 13 miRNAs, suggesting the complex regulation network of *AhPLDs* involved in and providing the clues to genetically engineer *AhPLDs* precisely through miRNA mediation.

Hormones are important regulators of plant growth and development. Many basic biological processes in peanut, such as seed germination, root hair growth, pollen tube elongation, blossom and leaf senescence, are known to be regulated by auxin, ABA, gibberellic acid (GA), JA and ethylene (Wang et al., 2018; Guo et al., 2021). Here, the *AhPLDs* expressed in various peanut tissues were found to contain at least one class of hormone-responsive *cis*-elements in their respective promoters (Figure 6; Table S6). *AhPLDα1A/1B* that contained ABA-responsive elements could be expressed in all peanut tissues and showed preferential expression in flowers and pod development. In *Arabidopsis*, PLDα1 can interact with a low-affinity nitrate transporter NRT1.2 to positively regulate ABA-mediated seed germination and seedling development (Li et al., 2020). It proved that the regulation of *AhPLDα1A/1B* on peanut growth and development may be mediated by ABA. Isoform PLDδ is the second abundant subfamily next only to isoform PLDα. It represents the majority of PLD isoforms expressed in male gametophyte throughout angiosperms evolution and has been found to be expressed higher in old leaves, stem, roots and flowers than in young leaves and siliques (Qin and Wang, 2002). Here, all *AhPLDδ*s had high expression levels in perianths, stamens and pistil, suggesting their central roles in peanut floral organ development. *AhPLDδ3A/B* even showed high expressions in all 22 tissues. This might be supported by the abundant hormone-responsive elements in their promoters. In tobacco (*Nicotiana tabacum* L.), PLDδ3 is the most important member active in pollen tubes. Tightly controlled production of PA generated by PLDδ3 is crucial for maintaining the balance between various membrane trafficking processes that are vital for plant cell tip growth (Pejchar et al., 2020). However, *AhPLDα9A* did not contain any *cis*-regulatory elements that respond to hormones, which may be the major factor that caused its non-expression in all peanut tissues. There is also evidence that *PLD* genes can be induced by ethylene to regulate the programmed death of plant cells (Lanteri et al., 2008), but no ethylene-responsive element could be found in any peanut PLD promoters here.

Besides, many *cis*-elements in response to diverse environmental stimuli were also found in *AhPLDs*' promoters, including ARE, MBS, LTR and T-rich (Figure 6; Table S6). It has been proved that abiotic stresses, such as cold, drought and salinity, could trigger high expressions of most *PLDs* low or weakly expressed under normal growth conditions, as a result of that a number of specific elements are located in their promoters (Wei et al., 2022). As expected, almost all *AhPLDs* (except for *AhPLDα9A*) could be induced by specific or multiple adversities (Figure 9), proving their potential roles in abiotic stress tolerances. At present, the functions of 14 *AhPLDs*' *Arabidopsis* orthologs in stress resistances have been determined (Table 2). For example, *AtPLDα1* (the ortholog of *AhPLDα1A/B*) has been found to participate in ABA signaling and responses to cold, drought and salt stress (Bargmann et al., 2009); *AtPLDα3* (the ortholog of *AhPLDα3A/B*) knockout mutant plants exhibit high sensitivity towards salinity, dehydration and ABA, while gain-of-function of *AtPLDα3* leads to reduced sensitivity in transgenic plants (Hong et al., 2008); *AtPLDδ* (the ortholog of *AhPLDδ1A/1B*) also has been proved to regulate stress resistances, such as freezing, severe dehydration, high salt, oxidative assault and ultra-violet irradiation (Zhang et al., 2003; Li et al., 2004; Liu et al., 2021); *AtPLDδ* and/or *AtPLDα1* can form a regulatory feedback loop with MPK3 and MPK6 to regulate PLD stability and submergence-induced PA production (Zhou et al., 2022).

In this study, there were five *AhPLDs* (*AhPLDα3A*, *AhPLDα5A*, *AhPLDβ1A*, *AhPLDβ2A* and *AhPLDδ4A*) found to be highly up-regulated under all three abiotic stresses commonly, suggesting these five *AhPLDs* might be involved in multiple regulatory pathways at the same time and lead to a wider range of stress resistances in peanut. But the up-regulated expressions of *AhPLDα1A* and several other *AhPLDs* were only induced by cold and drought rather than salinity. *AhPLDα4A*, *AhPLDδ1A* and *AhPLDδ3A* were even continuously down-regulated under salt stress. These results suggested that most *AhPLDs* were mainly involved in cold and drought tolerances but had little or even negative regulation on salt tolerance of peanut. Besides,

TABLE 2 Orthologous *PLD* genes in peanut and *Arabidopsis* with known abiotic stress tolerant functions.

AhPLDs	Orthologs	Functions	References
AhPLDα1A/1B	*AtPLDα1* (*AT3G15730*)	Salt and ABA responses; drought tolerance; freezing tolerance; hypoxia signaling; high-Mg_2^+ stress response; wounding response; mediating superoxide production;	Sang et al., 2001; Bargmann et al., 2009; Hou et al., 2016; Kocourkova et al., 2020; Zhou et al., 2022
AhPLDα2A/2B	*AtPLDα2* (*AT1G52570*)	Heat stress memory	Urrea Castellanos et al., 2020
AhPLDα3A/3B	*AtPLDα3* (*AT5G25370*)	Hyperosmotic response	Hong et al., 2008
AhPLDα4A/4B	*AtPLDα4/ε* (*AT1G55180*)	Hyperosmotic response; nitrogen deficiency response	Hong et al., 2016; Yao et al., 2022
AhPLDγA/B	*AtPLDγ1* (*AT4G11850*)	Wounding response	Wang et al., 2000
AhPLDδ1A/1B	*AtPLDδ* (*AT4G35790*)	Guard cell signaling and drought tolerance; hypoxia signaling; osmotic stress-triggered stomatal closure; salt stress tolerance; heat stress defense; freezing tolerance	Li et al., 2004; Angelini et al., 2018; Liu et al., 2021; Song et al., 2021; Zhou et al., 2022
AhPLDζ1A/1B	*AtPLDζ1* (*AT3G16785*)	Phosphate deficiency response; salt stress response	Li et al., 2006; Korver et al., 2020
	AtPLDζ2 (*AT3G05630*)	Phosphate deficiency response; root hydrotropism under drought stress; salt stress response	Li et al., 2006; Taniguchi et al., 2010; Su et al., 2018

AhPLDγA was significantly up-regulated only at the early stage (6 h) of cold stress, which may be caused by the low-temperature responsive element (LTR) in its promoter. *AhPLDα9A* could neither be expressed in any peanut tissues, nor be induced by any abiotic stresses in peanut. This deviation proved that *AhPLDα9A* might have functional roles in some processes other than peanut growth or stress signaling and its specific biological mechanism need to be further studied.

Conclusion

In conclusion, a total of 22, 22 and 46 *PLD* genes were identified in *A. duranensis*, *A. ipaensis* and *A. hypogaea*, respectively. Our comparative analyses provided valuable insight into the understanding of sequence characteristics, conserved domains, phylogenetic relationships and molecular evolution of PLD genes in allotetraploid peanut and its diploid progenitors. The predictive analytics of *cis*-regulatory elements, protein-protein interactions, putative miRNA expanded the view of transcriptional regulation and potential functions of *AhPLD* genes. Importantly, the expression patterns of tissue-specific and abiotic stress-responsive *AhPLD*s obtained from RNA-seq and qRT-PCR results offered useful information for further functional investigations. Several candidate *AhPLD*s, such as *AhPLDα3A*, *AhPLDα5A*, *AhPLDβ1A*, *AhPLDβ2A* and *AhPLDδ4A*, can be utilized for genetic manipulation of peanut and other legume crops for improved abiotic stress tolerance and productivity.

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

HZ and HY designed the research study. YY, SW and CZ conducted the bioinformatics analysis. JY, XA, NZ and XL performed the experiments. HZ and XZ analyzed the data. HZ and HY wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the earmarked fund for CARS-13 and LiaoNing Revitalization Talents Program (XLYC1902002).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1102200/full#supplementary-material

References

Angelini, J., Vosolsobě, S., Skůpa, P., Ho, A. Y. Y., Bellinvia, E., Valentová, O., et al. (2018). Phospholipase dδ assists to cortical microtubule recovery after salt stress. *Protoplasma* 255 (4), 1195–1204. doi: 10.1007/s00709-018-1204-6

Bargmann, B. O., Laxalt, A. M., ter Riet, B., van Schooten, B., Merquiol, E., Testerink, C., et al. (2009). Multiple PLDs required for high salinity and water deficit tolerance in plants. *Plant Cell Physiol.* 50 (1), 78–89. doi: 10.1093/pcp/pcn173

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48 (4), 438–446. doi: 10.1038/ng.3517

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

Bocckino, S. B., Blackmore, P. F., Wilson, P. B., and Exton, J. H. (1987). Phosphatidate accumulation in hormone-treated hepatocytes *via* a phospholipase d mechanism. *J. Biol. Chem.* 262 (31), 15309–15315. doi: 10.1016/S0021-9258(18)48176-8

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 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chi, X., Hu, R., Yang, Q., Zhang, X., Pan, L., Chen, N., et al. (2012). Validation of reference genes for gene expression studies in peanut by quantitative real-time RT-PCR. *Mol. Genet. Genomics* 287 (2), 167–176. doi: 10.1007/s00438-011-0665-5

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

Deng, W., Wang, Y., Liu, Z., Cheng, H., and Xue, Y. (2014). HemI: a toolkit for illustrating heatmaps. *PloS One* 9 (11), e111988. doi: 10.1371/journal.pone.0111988

Guo, F., Hou, L., Ma, C., Li, G., Lin, R., Zhao, Y., et al. (2021). Comparative transcriptome analysis of the peanut semi-dwarf mutant 1 reveals regulatory mechanism involved in plant height. *Gene* 791, 145722. doi: 10.1016/j.gene.2021.145722

Hanahan, D. J., and Chaikoff, I. L. (1947). A new phospholipide-splitting enzyme specific for the ester linkage between the nitrogenous base and the phosphoric acid grouping. *J. Biol. Chem.* 169 (3), 699–705. doi: 10.1016/S0021-9258(17)30887-6

Hong, Y., Pan, X., Welti, R., and Wang, X. (2008). Phospholipase Dalpha3 is involved in the hyperosmotic response in arabidopsis. *Plant Cell* 20 (3), 803–816. doi: 10.1105/tpc.107.056390

Hong, Y., Zhao, J., Guo, L., Kim, S. C., Deng, X., Wang, G., et al. (2016). Plant phospholipases d and c and their diverse functions in stress responses. *Prog. Lipid Res.* 62, 55–74. doi: 10.1016/j.plipres.2016.01.002

Hou, Q., Ufer, G., and Bartels, D. (2016). Lipid signalling in plant responses to abiotic stress. *Plant Cell Environ.* 39 (5), 1029–1048. doi: 10.1111/pce.12666

Huang, L., He, H., Chen, W., Ren, X., Chen, Y., Zhou, X., et al. (2015). Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea* l.). *Theor. Appl. Genet.* 128 (6), 1103–1115. doi: 10.1007/s00122-015-2493-1

Kocourkova, D., Krckova, Z., Pejchar, P., Kroumanova, K., Podmanicka, T., Danek, M., et al. (2020). Phospholipase Dα1 mediates the high-Mg^{2+} stress response partially through regulation of k^+ homeostasis. *Plant Cell Environ.* 43 (10), 2460–2475. doi: 10.1111/pce.13831

Kocourkova, D., Krckova, Z., Pejchar, P., Veselkova, S., Valentova, O., Wimalasekera, R., et al. (2011). The phosphatidylcholine-hydrolysing phospholipase c NPC4 plays a role in response of *Arabidopsis* roots to salt stress. *J. Exp. Bot.* 62 (11), 3753–3763. doi: 10.1093/jxb/err039

Korver, R. A., van den Berg, T., Meyer, A. J., Galvan-Ampudia, C. S., Ten Tusscher, K. H. W. J., and Testerink, C. (2020). Halotropism requires phospholipase Dζ1-mediated modulation of cellular polarity of auxin transport carriers. *Plant Cell Environ.* 43 (1), 143–158. doi: 10.1111/pce.13646

Krishna, G., Singh, B. K., Kim, E. K., Morya, V. K., and Ramteke, P. W. (2015). Progress in genetic engineering of peanut (*Arachis hypogaea* l.) –a review. *Plant Biotechnol. J.* 13 (2), 147–162. doi: 10.1111/pbi.12339

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

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

Lanteri, M. L., Laxalt, A. M., and Lamattina, L. (2008). Nitric oxide triggers phosphatidic acid accumulation *via* phospholipase d during auxin-induced adventitious root formation in cucumber. *Plant Physiol.* 147 (1), 188–198. doi: 10.1104/pp.107.111815

Li, M., Hong, Y., and Wang, X. (2009). Phospholipase d- and phosphatidic acid-mediated signaling in plants. *Biochim. Biophys. Acta* 1791 (9), 927–935. doi: 10.1016/j.bbalip.2009.02.017

Li, G., Lin, F., and Xue, H. W. (2007). Genome-wide analysis of the phospholipase d family in *Oryza sativa* and functional characterization of PLD beta 1 in seed germination. *Cell Res.* 17 (10), 881–894. doi: 10.1038/cr.2007.77

Li, W., Li, M., Zhang, W., Welti, R., and Wang, X. (2004). The plasma membrane-bound phospholipase ddelta enhances freezing tolerance in arabidopsis thaliana. *Nat. Biotechnol.* 22 (4), 427–433. doi: 10.1038/nbt949

Li, M., Welti, R., and Wang, X. (2006). Quantitative profiling of *Arabidopsis* polar glycerolipids in response to phosphorus starvation. roles of phospholipases d zeta1 and d zeta2 in phosphatidylcholine hydrolysis and digalactosyldiacylglycerol accumulation in phosphorus-starved plants. *Plant Physiol.* 142 (2), 750–761. doi: 10.1104/pp.106.085647

Li, J., Zhao, C., Hu, S., Song, X., Lv, M., Yao, D., et al. (2020). *Arabidopsis* NRT1.2 interacts with the PHOSPHOLIPASE Dα1 (PLDα1) to positively regulate seed germination and seedling development in response to ABA treatment. biochem. biophys. *Res. Commun.* 533 (1), 104–109. doi: 10.1016/j.bbrc.2020.08.025

Lin, D. L., Yao, H. Y., Jia, L. H., Tan, J. F., Xu, Z. H., Zheng, W. M., et al. (2020). Phospholipase d-derived phosphatidic acid promotes root hair development under phosphorus deficiency by suppressing vacuolar degradation of PIN-FORMED2. *New Phytol.* 226 (1), 142–155. doi: 10.1111/nph.16330

Liu, Q., Zhang, C., Yang, Y., and Hu, X. (2010). Genome-wide and molecular evolution analyses of the phospholipase d gene family in poplar and grape. *BMC Plant Biol.* 10, 117. doi: 10.1186/1471-2229-10-117

Liu, Q., Zhou, Y., Li, H., Liu, R., Wang, W., Wu, W., et al. (2021). Osmotic stress-triggered stomatal closure requires phospholipase dδ and hydrogen sulfide in arabidopsis thaliana. *Biochem. Biophys. Res. Commun.* 534, 914–920. doi: 10.1016/j.bbrc.2020.10.074

Lu, S., Fadlalla, T., Tang, S., Li, L., Ali, U., Li, Q., et al. (2019). Genome-wide analysis of phospholipase d gene family and profiling of phospholipids under abiotic stresses in brassica napus. *Plant Cell Physiol.* 60 (7), 1556–1566. doi: 10.1093/pcp/pcz071

Pejchar, P., Sekereš, J., Novotný, O., Žárský, V., and Potocký, M. (2020). Functional analysis of phospholipase dδ family in tobacco pollen tubes. *Plant J.* 103 (1), 212–226. doi: 10.1111/tpj.14720

Qin, C., and Wang, X. (2002). The arabidopsis phospholipase d family. Characterization of a calcium-independent and phosphatidylcholine-selective PLD zeta 1 with distinct regulatory domains. *Plant Physiol.* 128 (3), 1057–1068. doi: 10.1104/pp.010928

Raza, A., Mubarik, M. S., Sharif, R., Habib, M., Jabeen, W., Zhang, C., et al. (2022a). Developing drought-smart, ready-to-grow future crops. *Plant Genome* 10, e20279. doi: 10.1002/tpg2.20279

Raza, A., Tabassum, J., Fakhar, A. Z., Sharif, R., Chen, H., Zhang, C., et al. (2022b). Smart reprograming of plants against salinity stress using modern biotechnological tools. *Crit. Rev. Biotechnol.* 15, 1–28. doi: 10.1080/07388551.2022.2093695

Rizwan, H. M., Shaozhong, F., Li, X., Bilal Arshad, M., Yousef, A. F., Chenglong, Y., et al. (2022a). Genome-wide identification and expression profiling of KCS gene family in passion fruit (*Passiflora edulis*) under fusarium kyushuense and drought stress conditions. *Front. Plant Sci.* 13, 872263 doi: 10.3389/fpls.2022.872263.

Rizwan, H. M., Waheed, A., Ma, S., Li, J., Arshad, M. B., Irshad, M., et al. (2022b). Comprehensive genome-wide identification and expression profiling of eceriferum (CER) gene family in passion fruit (*Passiflora edulis*) under fusarium kyushuense and drought stress conditions. *Front. Plant Sci.* 13, 898307. doi: 10.3389/fpls.2022.898307

Sagar, S., Deepika,, Biswas, D. K., Chandrasekar, R., and Singh, A. (2021). Genome-wide identification, structure analysis and expression profiling of phospholipases d under hormone and abiotic stress treatment in chickpea (*Cicer arietinum*). *Int. J. Biol. Macromol.* 169, 264–273. doi: 10.1016/j.ijbiomac.2020.12.102

Sang, Y., Zheng, S., Li, W., Huang, B., and Wang, X. (2001). Regulation of plant water loss by manipulating the expression of phospholipase dalpha. *Plant J.* 28 (2), 135–144. doi: 10.1046/j.1365-313x.2001.01138.x

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Shi, X., Zhao, X., Ren, J., Dong, J., Zhang, H., Dong, Q., et al. (2021). Influence of peanut, sorghum, and soil salinity on microbial community composition in interspecific interaction zone. *Front. Microbiol.* 12, 678250. doi: 10.3389/fmicb.2021.678250

Song, P., Jia, Q., Xiao, X., Tang, Y., Liu, C., Li, W., et al. (2021). HSP70-3 interacts with phospholipase dδ and participates in heat stress defense. *Plant Physiol.* 185 (3), 1148–1165. doi: 10.1093/plphys/kiaa083

Su, Y., Li, M., Guo, L., and Wang, X. (2018). Different effects of phospholipase Dζ2 and non-specific phospholipase C4 on lipid remodeling and root hair growth in arabidopsis response to phosphate deficiency. *Plant J.* 94 (2), 315–326. doi: 10.1111/tpj.13858

Tang, K., Dong, C. J., and Liu, J. Y. (2016a). Genome-wide comparative analysis of the phospholipase d gene families among allotetraploid cotton and its diploid progenitors. *PloS One* 11 (5), e0156281. doi: 10.1371/journal.pone.0156281

Tang, K., Dong, C. J., and Liu, J. Y. (2016b). Genome-wide analysis and expression profiling of the phospholipase d gene family in gossypium arboreum. *Sci. China Life Sci.* 59 (2), 130–141. doi: 10.1007/s11427-015-4916-2

Taniguchi, Y. Y., Taniguchi, M., Tsuge, T., Oka, A., and Aoyama, T. (2010). Involvement of arabidopsis thaliana phospholipase Dzeta2 in root hydrotropism through the suppression of root gravitropism. *Planta* 231 (2), 491–497. doi: 10.1007/s00425-009-1052-x

Tan, W. J., Yang, Y. C., Zhou, Y., Huang, L. P., Xu, L., Chen, Q. F., et al. (2018). DIACYLGLYCEROL ACYLTRANSFERASE and DIACYLGLYCEROL KINASE modulate triacylglycerol and phosphatidic acid production in the plant response to freezing stress. *Plant Physiol.* 177 (3), 1303–1318. doi: 10.1104/pp.18.00402

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22), 4673–4680. doi: 10.1093/nar/22.22.4673

United States Department of Agriculture (USDA) Foreign Agricultural Service (2020). Available at: www.fas.usda.gov.

Urrea Castellanos, R., Friedrich, T., Petrovic, N., Altmann, S., Brzezinka, K., Gorka, M., et al. (2020). FORGETTER2 protein phosphatase and phospholipase d modulate heat stress memory in arabidopsis. *Plant J.* 104 (1), 7–17. doi: 10.1111/tpj.14927

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, X. (2005). Regulatory functions of phospholipase d and phosphatidic acid in plant growth, development, and stress responses. *Plant Physiol.* 139 (2), 566–573. doi: 10.1104/pp.105.068809

Wang, Y., Li, J., and Paterson, A. H. (2013). MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. *Bioinformatics* 29 (11), 1458–1460. doi: 10.1093/bioinformatics/btt150

Wang, G., Ryu, S., and Wang, X. (2012). Plant phospholipases: an overview. *Methods Mol. Biol.* 861, 123–137. doi: 10.1007/978-1-61779-600-5_8

Wang, P., Shi, S., Ma, J., Song, H., Zhang, Y., Gao, C., et al. (2018). Global methylome and gene expression analysis during early peanut pod development. *BMC Plant Biol.* 18 (1), 352. doi: 10.1186/s12870-018-1546-4

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinf.* 8 (1), 77–80. doi: 10.1016/S1672-0229(10)60008-3

Wang, C., Zien, C. A., Afitlhile, M., Welti, R., Hildebrand, D. F., and Wang, X. (2000). Involvement of phospholipase d in wound-induced accumulation of jasmonic acid in arabidopsis. *Plant Cell* 12 (11), 2237–2246. doi: 10.1105/tpc.12.11.2237

Wei, J., Shao, W., Liu, X., He, L., Zhao, C., Yu, G., et al. (2022). Genome-wide identification and expression analysis of phospholipase d gene in leaves of sorghum in response to abiotic stresses. *Physiol. Mol. Biol. Plants* 28 (6), 1261–1276. doi: 10.1007/s12298-022-01200-9

Yao, Y., Li, J., Lin, Y., Zhou, J., Zhang, P., and Xu, Y. (2021). Structural insights into phospholipase d function. *Prog. Lipid Res.* 81, 101070. doi: 10.1016/j.plipres.2020.101070

Yao, S., Wang, G., and Wang, X. (2022). Effects of phospholipase dε overexpression on soybean response to nitrogen and nodulation. *Front. Plant Sci.* 13, 852923. doi: 10.3389/fpls.2022.852923

Zhang, H., Jiang, C., Ren, J., Dong, J., Shi, X., Zhao, X., et al. (2020). An advanced lipid metabolism system revealed by transcriptomic and lipidomic analyses plays a central role in peanut cold tolerance. *Front. Plant Sci.* 11, 1110. doi: 10.3389/fpls.2020.01110

Zhang, W., Wang, C., Qin, C., Wood, T., Olafsdottir, G., Welti, R., et al. (2003). The oleate-stimulated phospholipase d, PLDdelta, and phosphatidic acid decrease H_2O_2-induced cell death in *Arabidopsis*. *Plant Cell* 15 (10), 2285–2295. doi: 10.1105/tpc.013961

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

Zhao, J., Zhou, D., Zhang, Q., and Zhang, W. (2012). Genomic analysis of phospholipase d family and characterization of *GmPLDα*s in soybean (*Glycine max*). *J. Plant Res.* 125 (4), 569–578. doi: 10.1007/s10265-011-0468-0

Zhou, Y., Zhou, D. M., Yu, W. W., Shi, L. L., Zhang, Y., Lai, Y. X., et al. (2022). Phosphatidic acid modulates MPK3- and MPK6-mediated hypoxia signaling in *Arabidopsis*. *Plant Cell* 34 (2), 889–909. doi: 10.1093/plcell/koab289

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 January 2023
DOI 10.3389/fpls.2023.1093913

OPEN ACCESS

EDITED BY
Weijian Zhuang,
Fujian Agriculture and Forestry University, China

REVIEWED BY
Benjamin Karikari,
University for Development Studies, Ghana
Chuanzhi Zhao,
Shandong Academy of Agricultural Sciences, China
Yong Lei,
Oil Crops Research Institute (CAAS), China

*CORRESPONDENCE
Fengzhen Liu
liufz@sdau.edu.cn
Yongshan Wan
yswan@sdau.edu.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 09 November 2022
ACCEPTED 09 January 2023
PUBLISHED 27 January 2023

CITATION
Wu Z, Luo L, Wan Y and Liu F (2023) Genome-wide characterization of the *PP2C* gene family in peanut (*Arachis hypogaea* L.) and the identification of candidate genes involved in salinity-stress response. *Front. Plant Sci.* 14:1093913. doi: 10.3389/fpls.2023.1093913

Genome-wide characterization of the *PP2C* gene family in peanut (*Arachis hypogaea* L.) and the identification of candidate genes involved in salinity-stress response

Zhanwei Wu[1,2], Lu Luo[1,2], Yongshan Wan[1,2]* and Fengzhen Liu[1,2]*

[1]State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, China, [2]College of Agronomy, Shandong Agricultural University, Tai'an, China

Plant protein phosphatase 2C (PP2C) play important roles in response to salt stress by influencing metabolic processes, hormone levels, growth factors, etc. Members of the PP2C family have been identified in many plant species. However, they are rarely reported in peanut. In this study, 178 *PP2C* genes were identified in peanut, which were unevenly distributed across the 20 chromosomes, with segmental duplication in 78 gene pairs. AhPP2Cs could be divided into 10 clades (A-J) by phylogenetic analysis. *AhPP2C*s had experienced segmental duplications and strong purifying selection pressure. 22 miRNAs from 14 different families were identified, targeting 57 *AhPP2C* genes. Gene structures and motifs analysis exhibited *PP2C*s in subclades AI and AII had high structural and functional similarities. Phosphorylation sites of *AhPP2C45/59/134/150/35/121* were predicted in motifs 2 and 4, which located within the catalytic site at the C-terminus. We discovered multiple MYB binding factors and ABA response elements in the promoter regions of the six genes (*AhPP2C45/59/134/150/35/121*) by *cis*-elements analysis. GO and KEGG enrichment analysis confirmed *AhPP2C-A* genes in protein binding, signal transduction, protein modification process response to abiotic stimulus through environmental information processing. Based on RNA-Seq data of 22 peanut tissues, clade A *AhPP2C*s showed a varying degree of tissue specificity, of which, *AhPP2C35* and *AhPP2C121* specifically expressed in seeds, while *AhPP2C45/59/134/150* expressed in leaves and roots. qRT-PCR indicated that *AhPP2C45* and *AhPP2C134* displayed significantly up-regulated expression in response to salt stress. These results indicated that *AhPP2C45* and *AhPP2C134* could be candidate *PP2Cs* conferring salt tolerance. These results provide further insights into the peanut *PP2C* gene family and indicate *PP2Cs* potentially involved in the response to salt stress, which can now be further investigated in peanut breeding efforts to obtain cultivars with improved salt tolerance.

KEYWORDS

peanut, *PP2C* gene family, salt stress, genome-wide identification, gene expression analysis

1 Introduction

Protein phosphorylation/dephosphorylation by protein kinases (PKs) and protein phosphatases (PPs) regulates targeted protein function in signal transduction cascades, leading to alterations in metabolism and/or gene expression in many biological processes, such as the development and stimulation of responses to the external environment (Lessard et al., 1997). PKs primarily phosphorylate serine (Ser), threonine (Thr) and tyrosine (Tyr), whereas PPs reverse this process (Li et al., 2014). PPs can be divided into three categories based on their substrate specificity: serine/threonine phosphatases, protein tyrosine phosphatases and dual specificity phosphatases (Kerk et al., 2008; Cao et al., 2016). Among them, protein tyrosine phosphatases can be further divided into phosphoprotein metallophosphatases and phosphoprotein phosphatases. Mn^{2+} or Mg^{2+}-dependent protein phosphatases (PP2Cs) are important members of the phosphoprotein metallophosphatase family (Kerk et al., 2002).

PP2Cs are highly conserved in evolution and regulate stress signaling pathways in a targeted manner (Cao et al., 2016). Most PP2Cs in plants have a conserved catalytic region at the C-terminus, while the N-terminus is a poorly conserved region of varying lengths. The N-terminus is important for the definition of the PP2C member's functionality, since this region can contain transmembrane regions and/or sequence motifs related with intracellular signaling, including those for interaction with kinases (Stone et al., 1994; Carrasco et al., 2003). It has been reported that some plant PP2Cs have significant roles in the plant response to environmental pressures, including salt, through their effects on metabolism, hormone levels and growth factors (Lim et al., 2015; Chu et al., 2021).

Relatively large PP2C gene families have been reported in several plant species, including soybean (103) (Chen et al., 2018), cotton (181) (Shazadee et al., 2019), *Arabidopsis* (76) (Kerk et al., 2002) and rice (132) (Singh et al., 2010). In *Arabidopsis*, PP2Cs (AtPP2Cs) are divided into clades A-J (Schweighofer et al., 2004). In clade A, ABA-INSENSITIVE 1 (ABI1) and ABA-INSENSITIVE 2 (ABI2) are important components of ABA signaling (Rodriguez et al., 1998; Kaliff et al., 2007). ABI1 and ABI2 act as negative regulators of ABA signaling, which play a pivotal role in the ABA-dependent salt stress response process. Under normal conditions, the ABA receptor (Pyrabatin Resistance 1-Like proteins; PYLs) is conformationally inactive and certain PP2Cs bind and dephosphorylate the sucrose non-fermenting-1-related protein kinases 2 (SnRK2) to inactivate ABA signaling. However, under osmotic stress ABA levels increase and their binding to the PYL induces conformational changes which enhance PYL-PP2C binding and inactivation of the PP2C. Consequently, the SnRK2 remains active and ABA-mediated stress signaling is initiated (reviewed in Gong et al., 2020; Zhang et al., 2022). Four members of the AtPP2Cs in clade B (APP2C1-4) are mitogen-activated protein kinase (MAPK) phosphatases that are negative regulators of phytohormones and defense responses (Ayatollahi et al., 2022). POLTERGEIST (AtPOL) and POLTERGEIST LIKE 1-5 (AtPLL1-5) of clade C are essential for cell maintenance and differentiation (Song et al., 2020). The functions of some AtPP2Cs in clades A, D, E, F and G have been reported, but the functions of most *Arabidopsis* PP2Cs remain unknown. The *PP2C* gene family in rice is divided into 11 clades, and most genes in clade A are triggered by various external abiotic stimuli, indicating that they play an important role in stress tolerance, particularly in salt stress (Xue et al., 2008).

Many *PP2C* gene family members have been reported to participate in plant salt stress regulation directly. (Leung and Giraudat et al., 1998; Kuhn et al., 2006; Liu et al., 2008). *AtPP2CG1* (*At2G33700*), a *PP2C* gene from *Arabidopsis*, is fully expressed under salt stress and positively regulates salt tolerance in an ABA-dependent manner (Liu et al., 2012). *At3G63320* and *At3G63340* actively respond to external stress stimuli and play a key role in the regulation of stomatal opening and closing, revealing that protein phosphorylation and dephosphorylation of signal regulators are significant for controlling stomatal pore size (Pang et al., 2020). Among *OsPP2Cs*, *OsPP2C53* (*Os05g51510*) and *OsPP2C51* (*Os05g49730*) are significantly expressed under 150mM NaCl stress (Xue et al., 2008). *OsPP2C08* (*Os01g46760*) is reported to be distinctly expressed in response to saline stress at the seedling stage compared to other *PP2Cs* (Sun et al., 2019). In rice, ubiquitination and degradation of OsPP2C09 (Os01g62760), a clade A PP2C, by abscisic acid-responsive RING Finger E3 Ligase (OsRF1) induced the salt tolerance of rice. (Kim et al., 2022).

Peanut (*Arachis hypogaea* L.) is an important oilseed and food crop worldwide. Cultivated peanut evolved from a cross between *Arachis duranensis* (A) and *Arachis ipaensis* (B) which subsequently underwent chromosome doubling to form the modern heterotetraploid (AABB genome, 2n=4x=40) with a total genome size of approximately 2.7 Gb (Bertioli et al., 2019). The main obstacles to peanut growth in semi-arid regions, where more than half of the global peanut production takes place, are drought and soil salinity (Banavath et al., 2018). There has been considerable advancement in the study of salt tolerance in peanut. Important examples include the stress-induced expression of *Arabidopsis* homeodomain-leucine zipper transcription factor (*AtHDG11*) to enhance salt tolerance, the identification of crucial intrinsic proteins that regulate salt stress in peanuts, the overexpression of the sodium/proton antiporter gene (*AtNHX1*) to improve salt tolerance, the vacuolar H^+-pyrophosphatase gene (*AVP1*) (Asif et al., 2011; Banavath et al., 2018; Han et al., 2021). Compared to other classes of proteins, the role of AhPP2Cs in the peanut response to saline and other stresses are substantially less well understood. In order to advance the development of peanut salt-tolerant mechanisms, we identified the peanut *PP*2C gene family and selected genes associated with salt tolerance. In this study, we gained insights into the *AhPP2C* genes by using some computer analysis, such as characterization, genomic evolution, gene structure, motifs, *cis*-acting elements, catalytic sites and gene functional annotations, etc., and expression analysis in different tissues and under salt stress for research purposes. 178 *AhPP2Cs* were identified in the *A. hypogea* genome. 18 members that clustered together with clade A of *Arabidopsis* were studied further. Of these, six *AhPP2Cs* were clustered together with salt tolerant *PP2C* genes from rice and *Arabidopsis*. There were total 20 ABA response elements and 21 MYB binding sites in promoter region of the six genes (*AhPP2C45/59/134/150/35/121*). Moreover, their expression patterns under salt stress were analyzed from publicly available RNA-Seq data and qRT-PCR. *AhPP2C45* and *AhPP2C134*

displayed significantly up-regulated expression under salt stress. They were considered as strong candidates for PP2Cs mediating salt stress signaling in peanut. These results provide theoretical support for biological functions of the peanut *PP2C* gene family under salt stress. Also, identification of *AhPP2C45/134* furnish new research threads for mechanism study of salt response differences between peanut germplasms, which play important roles in selection of salt tolerance germplasms and breeding.

2 Materials and methods

2.1 Identification of *PP2C* genes in cultivated peanut

Protein and gene sequences of *A. hypogaea* cv. Tifrunner and their annotated gene models were downloaded from PeanutBase[1] (Version 1). Protein sequences were scanned using the HMMER v3 (Finn et al., 2011) using the hidden Markov model (HMM) for PP2Cs (PF00481) from the Pfam database[1] (Finn et al., 2014). Proteins carrying the raw PP2C HMM with an E-value ($\leq 1\times10^{-20}$) were used to construct a peanut-specific hidden Markov model using hmmbuild from HMMER v3. The resultant peanut-specific HMM was used to search the protein sequences once again, and proteins that conformed to the criteria were submitted to NCBI-CDS[2] (Conserved Domain Search) in order to select those containing the tPP2C domain, PF00481.

2.2 Phylogenetic analysis of *Arabidopsis* and peanut PP2Cs

Phylogenetic trees were constructed with MEGA11 software (Tamura et al., 2021) using the ClustalW algorithm (Thompson et al., 1994) for sequence alignment with the Maximum Likelihood (ML) method with 1000 bootstrap replicates and Pearson correction.

2.3 Chromosomal location and gene duplication analysis

The chromosomal locations of *AhPP2C*s were mapped with Map Gene 2 Chrom web v2[3] (Chao et al., 2021). *AhPP2C* gene duplications were identified based on a sequence similarity of their CDS ≥75% and a sequence overlap of ≥75% of the longest sequence of the pair. Gene duplications were depicted using a circos diagram (Krzywinski et al., 2009). The Ka/Ks ratios of all *AhPP2C*s were predicted *via* simple Ka/Ks calculator in TBtools (Chen et al., 2020).

2.4 Phylogenetic analysis and physicochemical parameters of *AhPP2C-A*, salt tolerant *PP2C* genes from *Arabidopsis* and rice

AhPP2Cs in clade A were further clustered with *PP2C* genes associated with salt tolerance in *Arabidopsis* and rice. The physicochemical properties of *AhPP2C-A*, salt tolerant *PP2C* genes from *Arabidopsis* and rice, including amino acid length, molecular weight (MW), isoelectric point, were calculated using Expasy ProtPrarm tools[4] (Wilkins et al., 1999).

2.5 Protein and gene structures, the prediction of the catalytic site and *Cis*-acting elements in *AhPP2Cs*

The gene structures of clade A *AhPP2Cs* were predicted and visualized using Gene Structure Display Server 2.0[5] (Hu et al., 2015). Their conserved protein motifs were analyzed by MEME Suite Version 4.12.0 (Bailey and Elkan, 1994) using a maximum of 10 motifs with an a.a. length of 6-50. The other parameters were set as default. For *AhPP2C*s in subclades AI and AII, *cis*-acting elements were searched for in the 1500 bp region upstream of their start codons using PlantCARE[6] (Lescot et al., 2002). Their protein secondary and tertiary structures were predicted using SOPMA[7] (Geourjon and Deléage, 1994) and SwissModel[8] (Waterhouse et al., 2018), respectively. The protein sequences were submitted to NCBI-Blast[9], NetPhos3.1[10] (Blom et al., 1999; Blom et al., 2004) and Phyre2[11] (Kelley et al., 2015), for prediction of their catalytic sites.

2.6 Identification of miRNAs targeting *AhPP2C*s and gene function evaluation analysis

Through the psRNATarget website, the CDS of all *AhPP2C*s was used to predict miRNA target sites (Dai et al., 2018) with default. Figure of the interaction network among miRNAs and *AhPP2C*s were mapped by Cytoscape software (Shannon et al., 2003). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)

1 https://www.peanutbase.org/peanut_genome

2 https://www.ncbi.nlm.nih.gov/cdd/

3 http://mg2c.iask.in/mg2c_v2.1/

4 https://web.expasy.org/protparam/

5 http://gsds.gao-lab.org/

6 http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

7 https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html

8 https://swissmodel.expasy.org/in•teractive

9 https://blast.ncbi.nlm.nih.gov/Blast.cgi

10 https://services.healthtech.dtu.dk/service.php?NetPhos-3.1

11 http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

annotation evaluation was performed by submitting AhPP2C-A protein sequences to eggNOG-mapper[12] (Cantalapiedra et al., 2021). GO and KEGG enrichment evaluations were undertaken using TBtools.

2.7 Plant material and treatment

In this study, the peanut cultivar Shanhua 11, which was bred and is maintained by our research group, was used as the plant material. Mature seeds were germinated on cotton wool soaked in distilled water within seed cultivation discs placed in darkness for three days at 26°C, after which, intact seedlings were transplanted to a hydroponic box and cultured with 1/5 Hoagland's nutrient solution under a light/dark cycle of 16/8 h. Salt stress was applied to two-week-old seedlings by the adjustment of the nutrient solution to 200 mM NaCl. Leaves or roots from peanut seedlings were harvested after 72 h of salt or control treatments and were flash-frozen in liquid nitrogen before storage at -80°C until further use. Three biological replicates were used for analysis, each formed from materials obtained from five randomly chosen seedlings.

2.8 RNA extraction and qRT-PCR analysis

Total RNA was isolated with the Quick RNA Isolation Kit (Waryong, Beijing, China) and reverse transcribed into cDNA using Advantage RT-for-PCR Kit (TaKaRa, Dalian, China) following the manufacturer's instructions. The specific primers were designed using Beacon Designer 7.9. SYBR Green real-time PCR was performed according to the guidelines of the PCR machine. The relative expression levels were calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Statistical differences were determined by T-test (**P<0.01, *P<0.05).

2.9 The analysis of *AhPP2C* gene expression in peanut tissues and seedlings under salt stress

The tissue specificity of *AhPP2Cs* was determined from RNA-Seq data obtained for 22 different tissues at different developmental stages during peanut growth (Clevenger et al., 2016) obtained from the peanut database. The RNA-Seq data from salt-stressed peanut seedlings (Luo et al., 2021) was obtained from the public repository of NCBI[13] under the biological project accession number, PRJNA560660. The FPKM values of *AhPP2C-A*, including subclades AI and AII and other *AhPP2Cs* were extracted from the data set. The heatmap of was generated by TBtools after log2-transformation of FPKM values.

12 http://eggnog-mapper.embl.de/

13 https://www.NCBI.nlm.nih.gov

3 Results

3.1 Identification and phylogenetic analysis of the peanut *PP2C* gene family

A total of 183 PP2C-coding candidate genes were identified *via* HMM searching in peanut (*A. hypogaea* L.). After screening for the presence of the typical PP2C domain (PF00481), 178 genes were retained. A phylogenetic tree was created after multiple sequence alignment of the 178 *AhPP2Cs* 76 *Arabidopsis PP2C* genes. The 178 *AhPP2Cs* were divided into the 10 clades, A-J, as reported in other plants before, while 6 *AhPP2Cs* did not belong to any reported *PP2Cs* clades (Figure 1 and Supplementary Table 1). Of these, clade E was the largest, containing 34 *AhPP2Cs*, while clade B was the smallest, including 4 *AhPP2Cs*. The cladistics analysis indicated that the *PP2Cs* of peanut and *Arabidopsis* were similarly distributed between the clades, indicating the maintenance of the same types of functional diversity in *PP2Cs* following their speciation.

3.2 Chromosomal distribution and gene duplication analysis of *AhPP2Cs*

As shown in Figure 2, all 178 *AhPP2Cs* are unevenly distributed on 20 chromosomes, with no more than five genes on chromosomes 2, 7, 10 and 12, while the others are greater than five, with the greatest number of genes on chromosome 20. Most *AhPP2Cs* were concentrated near the ends of each chromosome, which was similar to the chromosomal distribution of PP2Cs reported for other plant species (Shazadee et al., 2019; Yu et al., 2019). A gene duplication analysis based on sequence similarity of *AhPP2Cs* indicated 78 genes pairs and found that there was segmental duplication in 78 genes pairs (Figure 3 and Supplementary Table 2). As expected, the paired *AhPP2Cs* clustered together in the phylogenetic tree. However, their distribution positions on 20 chromosomes were similar. Thus, to understand the mode of selection of the *AhPP2Cs*, the Ka, Ks, and Ka/Ks ratio was revealed for all gene pairs (Supplementary Table 2). The dataset unveiled that all duplicated *AhPP2C* gene pairs had a Ka/Ks ratio of <1, except for *AhPP2C54/154*, suggesting that *AhPP2Cs* in these groups underwent purifying selective pressure. The Ka/Ks value of two gene pairs (*AhPP2C42/131*, *AhPP2C81/137*) was zero, indicating strong purifying selection.

3.3 Investigation of putative miRNAs targeting *AhPP2C* genes

To better investigate how the *AhPP2C* genes were regulated by miRNAs during translation, we identified 22 miRNAs targeting 57 genes (Figure 4 and Supplementary Table 3). These miRNAs belonged to 14 different families. The result showed that ahy-miR3513-5p targeted the most number (8) of genes, followed by ahy-miR159 that targeted seven genes. Two miRNAs, including ahy-miR156a and ahy-miR3511-5p, targeted four genes, followed by ahy-miR156c, ahy-miR394, ahy-miR3516, ahy-miR3514-3p and ahy-miR3520-5p that targeted three genes, while ahy-miR408-3p, ahy-miR156b-3p, ahy-miR3519, ahy-miR3514-5p, ahy-miR156b-5p and ahy-miR3509-5p targeted two

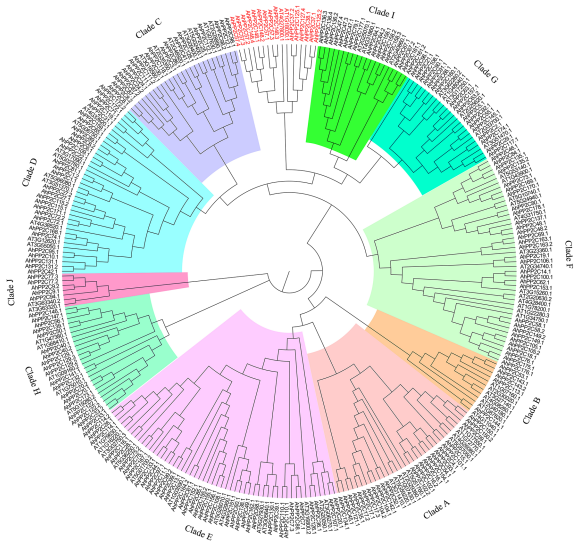


FIGURE 1
A Phylogenetic tree of PP2Cs from peanut and *Arabidopsis*. PP2Cs in different clades are labeled with different colors. 6 *AhPP2C* genes in red were not classified.

different genes. Only four miRNAs, including ahy-miR3520-3p, ahy-miR167-3p, ahy-miR160-3p and ahy-miR398 targeted one gene, *AhPP2C163*, *AhPP2C34*, *AhPP2C47* and *AhPP2C15* respectively. In *AhPP2C-A*, only *AhPP2C15*, *AhPP2C35* and *AhPP2C121* were targeted by miRNAs. Some genes like *AhPP2C30*, *AhPP2C117 AhPP2C19*, *AhPP2C41*, *AhPP2C128* and *AhPP2C155* were found to be targeted by more than one miRNA.

3.4 Phylogenetic analysis and physicochemical parameters of *AhPP2C-A*

Previous studies have demonstrated that clade A *PP2C* genes has been identified to be associated with salt tolerance in rice and *Arabidopsis* (Schweighofer et al., 2004; Xue et al., 2008). In peanut, eighteen *PP2C* genes belong to clade A. To further investigate their biological functions, *AhPP2C-A* were further compared with the reported salt response *PP2C* genes in rice and *Arabidopsis*. As shown in Figure 5, the latter were clustered together with closely related *AhPP2C*s in subclades AI and AII. The bootstrap values of these two branches were 74.3% and 74.6%, respectively, indicating a high degree of confidence, suggesting that *AhPP2Cs* in these subclades may be functionally associated with salt tolerance. Their identification regresses to finding peanut sequences with the high homology with known *PP2C* genes associated with salt tolerance. The physicochemical properties of subclades AI and AII are shown in Table 1, which indicates that all members of each subclade shared similar physiochemical properties.

3.5 Gene structures and motifs analysis of *PP2C*s in subclades AI and AII

To further analyze the potential roles of *AhPP2C*s in subclades AI and AII, their phylogeny was analyzed together with their gene structures and motifs (Figure 6). In subclade AI, all *PP2Cs*, including the salt tolerant genes from *Arabidopsis* and rice in the same branch of the phylogenetic tree were similar in structures and contained 4 exons. In subclade AII, *AhPP2C134* and *45*, and two rice *PP2Cs* involved in the response to salt stress contained 3 exons, while the remaining two *AhPP2Cs* contained only 2 exons.

A total of 10 motifs was identified using MEME Suite. Motifs 5, 9 and 10 were exclusive to subclade AI, motif 8 was exclusive to subclade AII, while motifs 1, 2, 3, 4, 6 and 7 were present in both

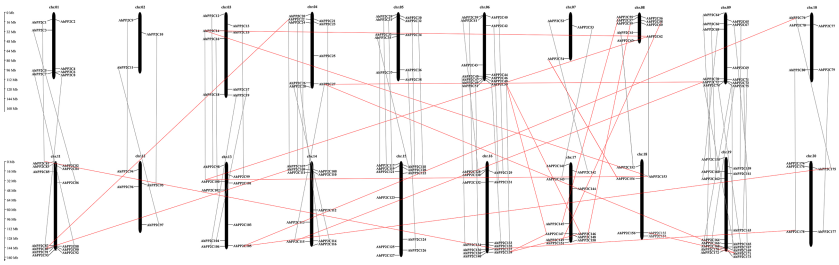


FIGURE 2
Chromosomal distribution of *AhPP2C*s. Orthologous genes were linked by gray lines. Paralogous genes were linked by red lines.

10.3389/fpls.2023.1093913

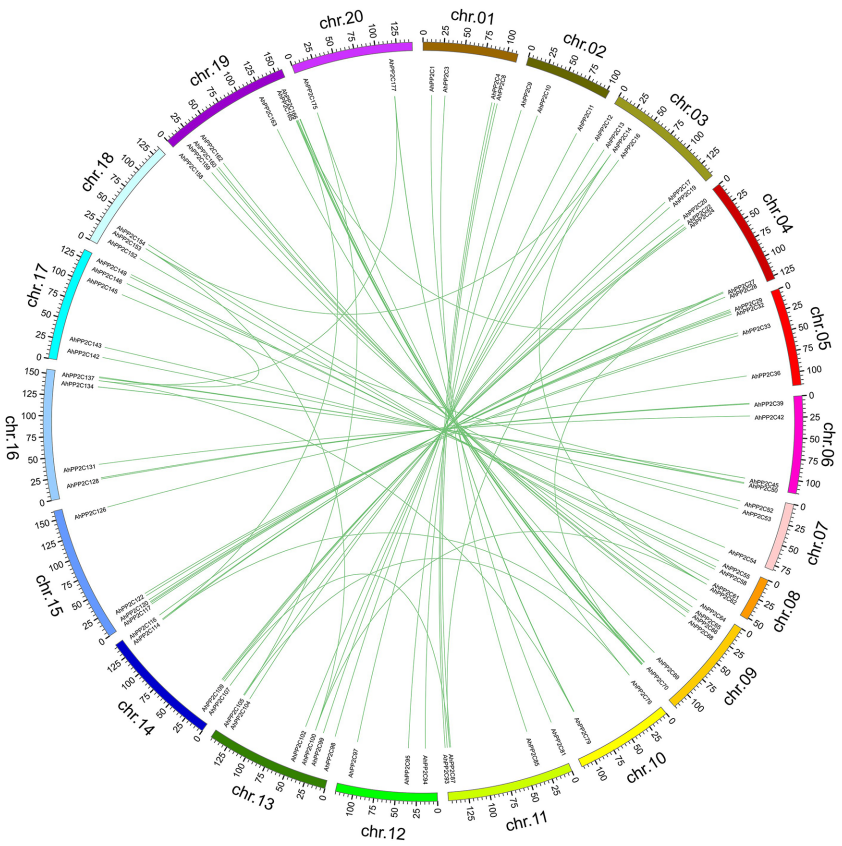


FIGURE 3
Gene duplication analysis of *AhPP2Cs*. Chromosomes are drawn in different colors. The scale provided represents the chromosome size (Mbp). Duplicated *AhPP2Cs* are linked by green lines.

subclades. After alignment of protein sequences, conserved domain sequences and motifs, it was found motifs 2 and 4 are located in the catalytic site at the C-terminus (Bork et al., 1996). These sites were highly conserved across the PP2Cs of subclades AI and AII, except Os1g40094 (Figure 6 and Supplementary Figure 1). Motif 10 was located at the N-terminal, but it appeared to be highly conserved, because it was found to exist on fewer number of sequences (Supplementary Table 4). Motif 5 and motif 9 were only present in subclade AI. This may account for the functional differences between the two subclades. AhPP2C45/134 in subclade AII had a greater number of motif 6, where it may confer unique functions to these PP2Cs, which will need be investigated further.

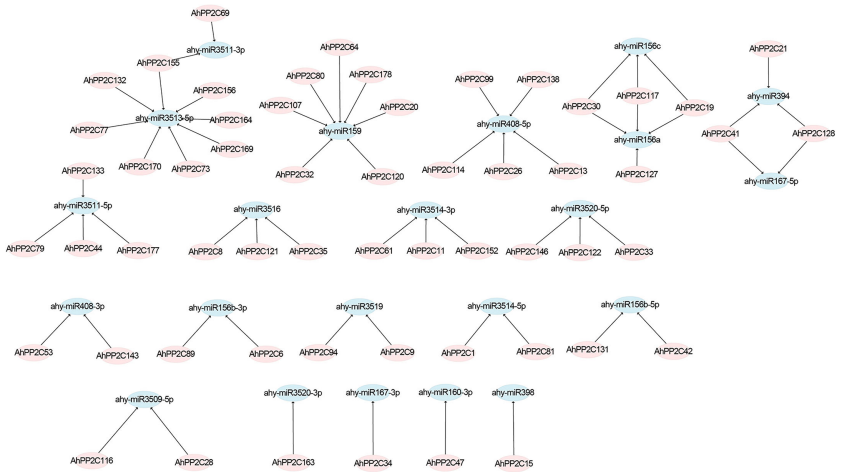


FIGURE 4
Prediction of putative miRNAs targeting *AhPP2Cs*. The pink shapes correspond to *AhPP2Cs*, and blue shapes indicate predicted miRNAs.

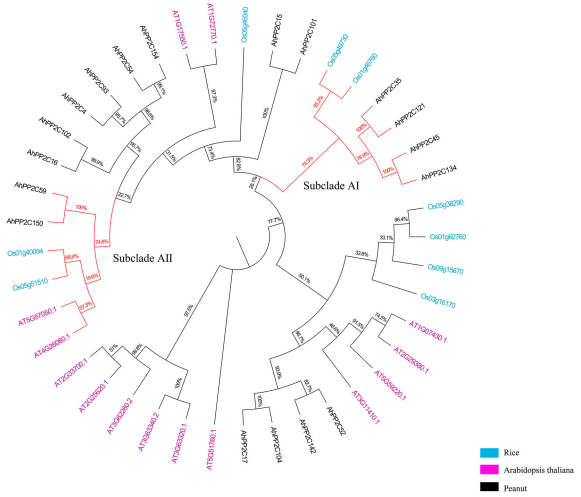


FIGURE 5
Phylogenetic tree of clade A *AhPP2Cs*, with salt tolerant genes from *Arabidopsis* and rice. The branches of subclades AI and AII are indicated in red.

3.6 Protein structure and catalytic site prediction of *AhPP2Cs* in subclades AI and AII

The secondary structures of *AhPP2Cs* in Subclades AI and AII consisted of 31.16%-43.64% α-helices, 30.26%-46.11% irregular curls, 15.91%-23.68% β-sheets and 5.25%-9.21% β-turns. Analysis of the secondary structures of orthologous gene pairs *AhPP2C150*/*AhPP2C59*, *AhPP2C134*/*AhPP2C45* and *AhPP2C121*/*AhPP2C35* showed they were very similar. The tertiary structures were also highly similar (Figure 7). Moreover, their catalytic structural domains all consist of a central β-sandwich that binds Mn^{2+}/Mg^{2+} and is surrounded by α-helices (Das et al., 1996; Shi, 2009). The phosphorylation sites, which can be found in motifs 4 and 2 within their predicted catalytic sites were also found to be similar: S362 and S309 for AhPP2C134/AhPP2C45, S322 and S276 for AhPP2C35/AhPP2C121, S532 and S473 for AhPP2C59 and S531 and S472 for AhPP2C150.

3.7 Gene ontology and Kyoto Encyclopedia of genes and genomics enrichment analysis of *AhPP2C-A* genes

To further understand the role of *AhPP2C-A* genes at the molecular level, we performed GO and KEGG enrichment analysis (Supplementary Figure 2). The GO enrichment analysis can be divided into three main categories: biological processes (BP), molecular functions (MF) and cellular components (CC). For instance, in MF class, the most enriched term was protein binding (GO:0005515), followed by binding (GO:0005488) and hydrolase activity (GO:0016787). In CC class, the highly enriched terms were cytosol (GO:0005829) and nucleus

TABLE 1 Physicochemical parameter of PP2Cs in Subclades AI and AII.

Subclade	Name	Numbers of Amino Acid/aa	MW/kDa	Theoretical pI
AI	AhPP2C35	346	36.9	5.22
	AhPP2C121	346	37.2	5.24
	AhPP2C45	380	41.5	6.1
	AhPP2C134	380	41.5	5.78
	Os05g49730	381	40.3	7.64
	Os01g46760	403	43.0	5.57
AII	Os01g40094	356	36.4	4.68
	Os05g51510	445	46.7	4.90
	AT5G57050.1	423	46.3	5.93
	AT4G26080.1	434	47.5	5.81
	AhPP2C150	552	59.4	4.73
	AhPP2C59	553	59.5	4.74

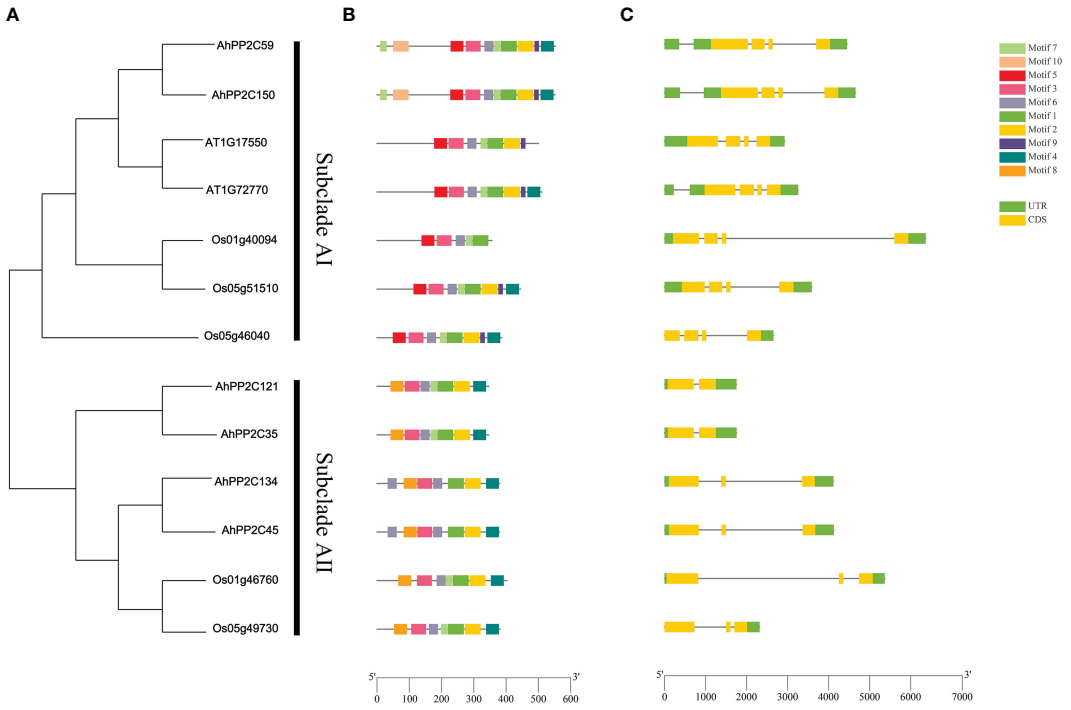


FIGURE 6
Phylogenetic relationship, gene structures and motifs analysis of PP2Cs in subclades AI and AII. **(A)** Phylogenetic tree of PP2Cs **(B)** Motifs analysis of PP2Cs. Colored boxes indicate the different conserved motifs as indicated in the scheme to the right of the figure. **(C)** Exon-Intron Structure of *PP2Cs*, where exons are shown as yellow boxes. Where present, 5′ and 3′ UTRs are depicted as green boxes.

(GO:0005634). Whereas in BP class, the highly enriched terms were cell communication (GO:0007154), signal transduction (GO:0007165), regulation of molecular function (GO:0065009), response to endogenous stimulus (GO:0009719), response to abiotic stimulus (GO:0009628), etc. In addition, KEGG pathway enrichment study identified four pathways involved in the different functions of the *AhPP2C-A* genes. The highly enriched pathways include signal transduction (B09132), plant hormone signal transduction (04075), environmental information processing (A09130), MAPK signaling pathway – plant (04016) (Supplementary Table 5). In brief, it could be concluded that *AhPP2C-A* genes related to abiotic stresses response, protein modification process and plant hormone signaling.

3.8 *Cis*-acting elements analysis of *AhPP2Cs* in subclades AI and AII

In higher plants, PP2C is involved in many stress resistance responses and is induced by a variety of abiotic stresses. *Cis*-acting elements were therefore searched in the 1500 bp sequences upstream of the start codons of peanut *PP2Cs* of subclades AI and AII (Table 2). There were obvious differences in the types and numbers of *cis*-acting elements among different *AhPP2Cs*. The results showed that many phytohormones and stress-related response elements were identified, including ABRE/ABRE4/ABRE3a, Aux RR-core/TGA-element, TCA-element, MYB and CGTCA-motif/TGACG-motifs. *AhPP2Cs* in subclades AI and AII all contained MYB binding factors, which are essential for regulating how plants respond to abiotic stress, particularly in response to salt stress. *AhPP2C150*, *35*, *134 and 45* contained *cis*-acting elements involved in the regulation of the ABA response (ABRE/ABRE4/ABRE3a), while *AhPP2C121* and *59* did not. *AhPP2C45* contained the maximum number of *cis*-elements (21), while *AhPP2C35* contained the minimum number of *cis*-elements (4). The differences in number and kind of elements may lead to functional difference of *AhPP2Cs*.

3.9 Expression profiling of clade A *AhPP2Cs* in different tissues

Based on the transcriptome data of 22 peanut tissues (Clevenger et al., 2016), the expression patterns of clade A *AhPP2Cs* were extracted and shown in Figure 8 and Supplementary Table 6. Clade A *AhPP2Cs* diverse with differing levels of tissue specificity. *AhPP2C17*, *59*, *150*, *134*, *45*, *15 and 101* expressed at relatively high levels in several of the 22 different tissues. In contrast, the overall expression levels of *AhPP2C54*, *35* and *121* were lower, and mainly detected in seeds tissues. The remaining genes (7) were moderately expressed in all tissues. Considering *AhPP2Cs* of subclades AI and AII, *AhPP2C121* and *35* were only expressed in seeds, while the other four genes were also expressed in leaves and roots. The above results indicated that clade A *AhPP2Cs* exhibited different expression patterns.

3.10 The salt stress effects on the expression of *AhPP2Cs* in subclades AI and AII

An initial assay of the expression of subclade AI and AII *AhPP2Cs* in seedlings under salt stress used publicly available RNA-Seq data

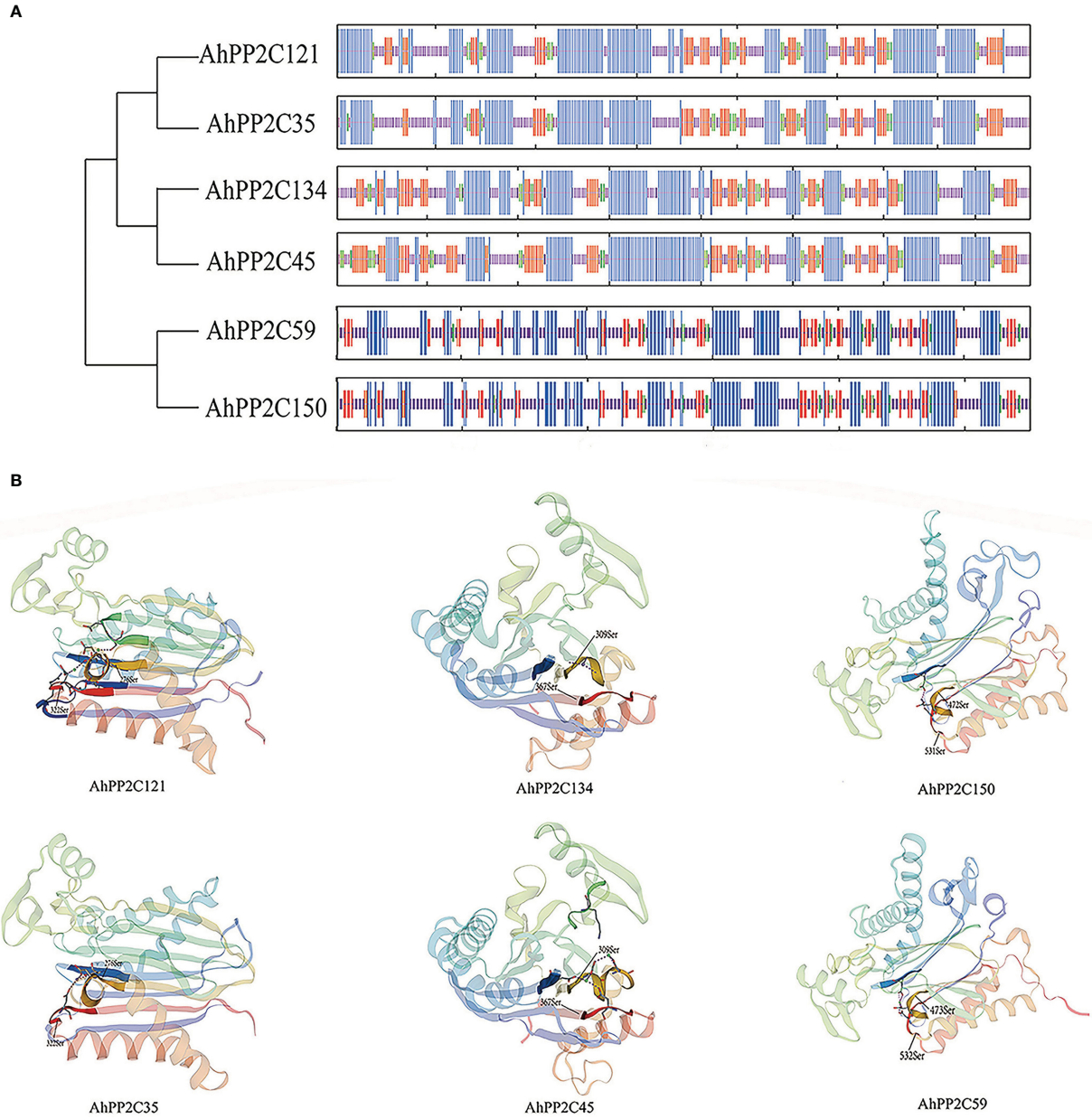


FIGURE 7
Protein structures and catalytic site prediction of *AhPP2C*s in subclades AI and AII. **(A)** protein secondary structures of *AhPP2C*s. α-helices were in blue; irregular curls were in purple; β-sheets were in red; β-turns were in green. **(B)** protein tertiary structures and catalytic sites of *AhPP2C*s in subclades AI and AII.

from leaves and roots. The heatmap of relative expression is shown in Figure 9A. Under salt stress, in leaves, *AhPP2C45*, *134*, *59* and *150* were up-regulated compared with the control, but this was only significant in *AhPP2C45* and *134*. In roots, *AhPP2C45* and *AhPP2C134* were also up-regulated, while *AhPP2C59* and *AhPP2C150* were down-regulated relative to the control. The expression of *AhPP2C121* was not detected in roots or leaves with or without salt stress.

To validate the results for the four up-regulated genes under salt stress one step further, the relative expression levels of these *AhPP2Cs* (*45, 134, 59 and 150*) in leaves and roots were analyzed by qRT-PCR (Figures 9B, C). After 72 h of salt treatment (200 mM NaCl), the relative expression levels of *AhPP2C45* and *134* were up-regulated significantly in both leaves and roots. These results indicated that AhPP2C45 and 134 may play important roles in peanut salt stress response.

4 Discussion

Salt stress is the second worst abiotic factor affecting global agricultural productivity, disrupting many physiological, biochemical and molecular processes in plants. Salinity affects plant growth, development and productivity (Raza et al., 2022). Salt stress also has a serious impact on peanut production (Aravind et al., 2022). PP2Cs are widespread and highly conserved in prokaryotes and

TABLE 2 The Distribution of *Cis*-Acting Elements in *AhPP2Cs* of Subclades AI and AII.

Gene name	Numbers of *cis*-acting element								
	A	B	C	D	E	F	G	H	I
AhPP2C150	2	3	—	3	—	—	5	2	2
AhPP2C121	—	—	1	—	—	1	3	—	1
AhPP2C35	1	1	1	—	—	—	1	—	—
AhPP2C134	7	—	2	—	2	—	6	2	—
AhPP2C59	—	—	—	1	—	—	2	—	2
AhPP2C45	10	—	2	—	2	—	4	2	1

Key: A, ABRE/ABRE4/ABRE3a; B, Aux RR-core/TGA-element; C, ERE; D, TCA-element; E, MBS; F, LTR; G, MYB binding factor; H, CGTCA-motif/TGACG-motif; I, WUN-motif; —, no response element.

eukaryotes (Yang et al., 2018). Several studies of plant PP2Cs have indicated that of a subfamily of PP2Cs (clade A) have regulatory roles in stress responses through dephosphorylating the substrate proteins, especially in ABA-dependent responses, such as the drought and salinity response (Xue et al., 2008; Cui et al., 2013; Singh et al., 2015; Krzywińska et al., 2016). However, less information is available for *PP2Cs* in the economically important crop plant peanut, whose productivity is considerably affected by salinity stresses. This research therefore aimed to provide a systematic analysis of *AhPP2Cs* and to identify candidate *AhPP2Cs* participate in salt stress response.

In this study, we identified 178 *AhPP2Cs* that showed a phylogenetic clustering similar to that of *Arabidopsis*, including a distinct clade A, which is considered to contain a subfamily of *PP2Cs* with roles in ABA-mediated salt stress responses (Zhang et al., 2013; Singh et al., 2015; He et al., 2019). Numbers of *PP2C* genes differ a lot among species (Kerk et al., 2002; Chen et al., 2018; Shazadee et al., 2019). Deviations in the number of *PP2C* members from different plant species may be attributed to gene duplication events, including tandem duplication and segmental duplication, which play an important role in expanding *PP2Cs* members. Duplications of *PP2C* genes have also been found in some plant species (Yang et al., 2018; Zhang et al., 2022). Our results confirmed that *AhPP2Cs* had underwent segmental duplication strong purifying selective pressure (Figure 3 and Supplementary Table 2). Previous phylogenetic tree of *Arabidopsis* showed that there were 6 *AtPP2Cs* could not clustered to any clades (Schweighofer et al., 2004; Bhaskara et al., 2019). In this work, we found 6 *AhPP2Cs* clustered with 2 of the 6 *AtPP2Cs*, and we speculated that this may be universal phenomena in plant *PP2C* gene family. Some AhPP2Cs of clade A clustered together with PP2Cs known to confer salt tolerance in *Arabidopsis* and rice (Figure 5). The AhPP2C35, 121, 45 and 134 of subclade AI were found to be closely clustered with the OsPP2Cs Os05g49730 (OsPP2C51) and Os01g46760 (OsPP2C08). *OsPP2C08* (*Os01g46760*) was identified as a potential candidate gene conferring improved salt tolerance through the integration of genome-wide polymorphism information with QTL mapping and expression profiling data (Subudhi et al., 2020). The heterogeneous expression of *OsPP2C51* (*Os05g49730*) in *Arabidopsis* revealed that it could positively regulate ABA-induced rice seed germination under conditions of salt stress (Bhatnagar et al., 2017).

Similarly, the AhPP2C59 and 150 of subclade AII were closely clustered with two rice Os01g40094 (OsPP2C06) and Os05g51510 (OsPP2C53) and two *Arabidopsis* PP2Cs (AT5G57050.1 and AT4G26080.1). ABI2 (AT5G57050.1) as a SOS2-interacting protein can affect salt stress response in *Arabidopsis* (Ohta et al., 2003).

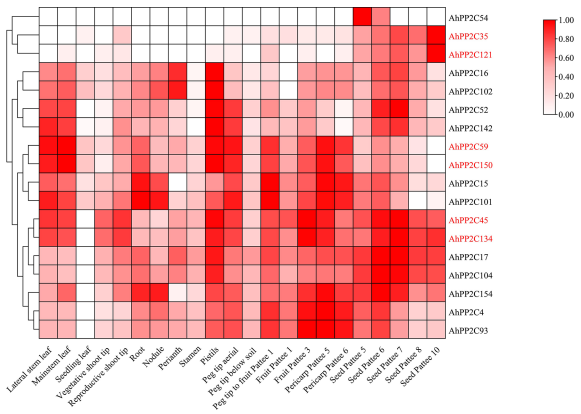


FIGURE 8
Expression pattern analysis of clade A *AhPP2Cs*. The heat map of log2-transformed averaged FPKM data was mapped using the intensity scale shown on the right. The visualization and clustering utilized TBtools. The *AhPP2Cs* of subclades AI and AII are labelled in red.

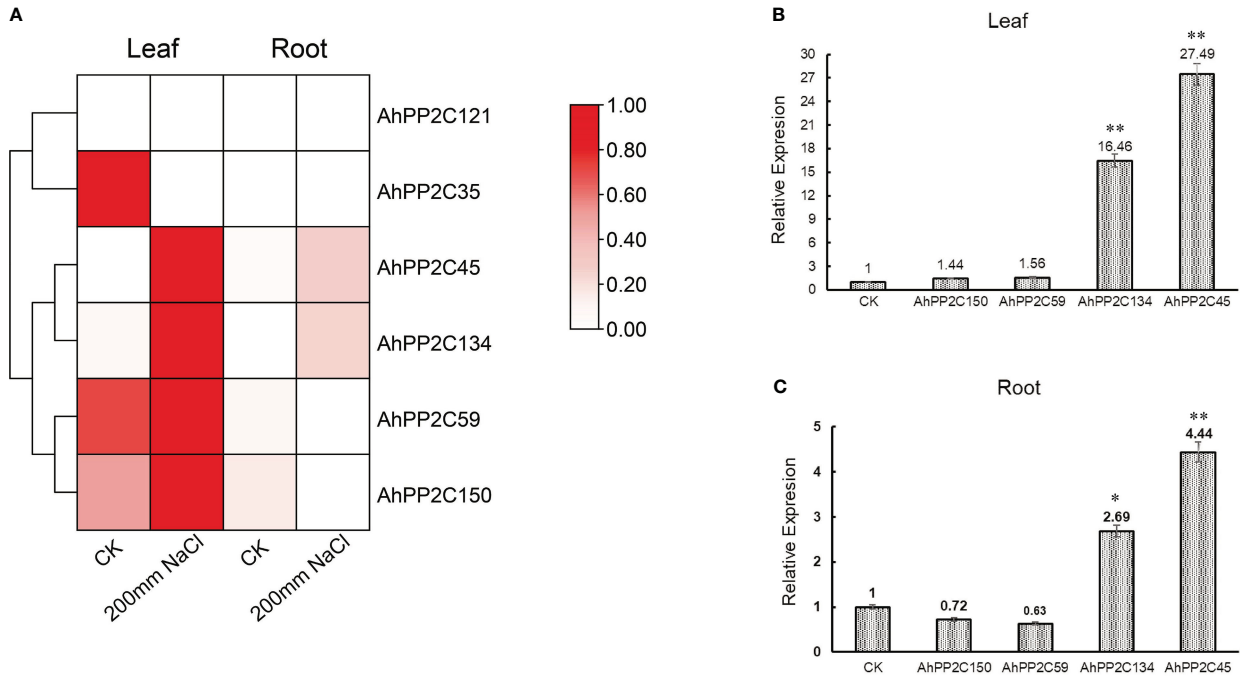


FIGURE 9
Expression profiles under salt stress. **(A)** Relative expression pattern of *AhPP2Cs* in subclades AI and AII in response to salt stress. **(B, C)** qRT-PCR analysis of *AhPP2Cs 45*, *59*, *134* and *150* under salt stress in leaves **(B)** and roots **(C)**. "*" indicated that the genes were significant differences compared with the untreated control (**P<0.01, *P<0.05).

Mutations ABI1 (AT4G26080.1), affecting ABA-perception in *Arabidopsis*, can reduce the accumulation of both AtP5CS mRNAs during salt-stress and result in higher proline accumulation. (Strizhov et al., 1997). OsPP2C53 (Os05g51510) negatively regulates OsSLAC1 in stomatal closure and transgenic rice overexpressing *OsPP2C53* showed significantly higher water loss than control (Min et al., 2019). Microarray analysis of transgenic rices overexpressing *OsNAP* in high salinity, drought and low temperature revealed that a stress-related gene *OsPP2C06* (*Os01g40094*) was up-regulated (Chen et al., 2014). Salt tolerant *PP2C* genes from *Arabidopsis* and rice are located in both subclades AI and AII, then this functionality occurs in two phylogenetically distinct subclades. This could, for example, be because function of salt tolerant *PP2C* genes can be achieved in mechanistically different ways and these evolved independently, or these genes differ in other aspects of their sequence (domains/motifs) that have more weight in their phylogenetic placing. In addition, the peanut *AhPP2Cs* that closely clustered with *PP2Cs* conferring salt tolerance also shared similar motifs and gene structures with their *Arabidopsis* and rice counterparts (Figure 6), indicating that they may also share similar functionalities.

Additionally, we found nine *cis*-elements, two of them, ABRE/ABRE4/ABRE3a and MYB binding factor are related to salt tolerance. ABRE/ABRE4/ABRE3a has been discovered in previous studies of *PP2Cs* in stress (Xue et al., 2008; Chen et al., 2018), however there is less available information regarding MYB binding factors in this process. Furthermore, *AhPP2Cs* gene functions were further predicted by GO enrichment analysis, which supported the role of these genes in dephosphorylating the substrate proteins and salt stress response. It is therefore of interest that after our analysis of the expression profiles of *AhPP2Cs* under salt stress, we observed the *PP2C* genes with ABRE/ABRE4/ABRE3a and MYB binding factors showed greater relative changes in expression. The genes (*AhPP2C45/134*) contained both *cis*-elements and having the maximum number of both *cis*-elements. The greater number of salt tolerance-related cis-elements *AhPP2Cs* have, the greater expression levels under salt stress. The number of salt tolerance-related *cis*-elements may therefore be correlated with either high or low gene expression in response to salt stress.

The potential phosphorylation sites and protein structures of the AhPP2Cs were also examined. The conserved catalytic core domain of PP2Cs contains a central β-sandwich with a pair of α-helices on either side of each β-sheet (Das et al., 1996). Within the conserved catalytic structural domain, the PP2Cs contain three characteristic sequences: DG××G, DG, and G××DN (where x is any amino acid) (Shi, 2009). These characteristic sequences located in motifs 3, 2 and 4, respectively (Supplementary Figure 3). Moreover, phosphorylation sites were predicted to be located in motifs 2 and 4, but not in motif 3. More importantly, some of phosphorylation sites, S362 for AhPP2C134/AhPP2C45, S322 for AhPP2C35/AhPP2C121, S532 for AhPP2C59 and S531 for AhPP2C150, were within the G××DN characteristic sequence. These phosphorylation sites may provide an important direction for future studies of AhPP2Cs in ABA-related signal transduction under salt stress.

Based on phylogenetic, gene structure, motif and *cis*-acting elements analysis of *AhPP2C-A*, *AhPP2C45* and *AhPP2C134* had similar structures to the reported *PP2C* genes associated with salt tolerance in rice and *Arabidopsis* and contained more *cis*-elements that related to salt tolerance. According to the salt stress effects on the expression, *AhPP2C45* and *AhPP2C134* had significant differences compared with the control. These results indicated that *AhPP2C45*

and *AhPP2C134* could be candidate *PP2Cs* conferring salt tolerance. The study of *AhPP2C*s in response to salt stress may be used in conjunction with phenotypic data to screen for more salt tolerant genotypes and salt tolerant varieties, providing a crucial foundation for the next step breeding.

5 Conclusion

In this study, we identified 178 *AhPP2C* genes which distributed across the 20 chromosomes, with segmental duplication in 78 gene pairs. Phylogenetic tree, gene and protein structures, *cis*-elements and expression analysis indicated that *AhPP2C45* and *AhPP2C134* could be candidate *PP2Cs* conferring salt tolerance. These results provide important information for the research of salt-stress mechanism, selection of salt-resistant germplasm and breeding of salt-tolerance cultivar in peanut.

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: The public repository of NCBI under the biological project accession number, PRJNA560660.

Author contributions

FL conceived the study, designed the experiments, supervised and complemented the writing. YW provided suggestions and supervision. ZW and LL produced the figures and wrote the original article. All authors discussed the results and commented on the article. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Peanut Seed Industry Project in Shandong Province, China (No. 2022LZGC007 to FL), and funding from the Agriculture Research System in Shandong Province, China (No. SDAIT-04-03 to FL) and the Chinese Agriculture Research System (CARS-13 to YW).

Acknowledgments

We gratefully acknowledge students including Hui Yang, Yuanyuan Sang, Yijun Li, Haiyang Yu, and Junjie Zhang for assisting us with this study.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1093913/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Motifs analysis of PP2Cs in Subclades AI and AII.

SUPPLEMENTARY FIGURE 2
GO and KEGG enrichment analysis of clade A *AhPP2C*s. **(A)** The enriched GO terms from MF, CC, BP classifications in *AhPP2C*s. **(B)** The enriched KEGG pathways in *AhPP2C*s.

SUPPLEMENTARY FIGURE 3
Sequence alignment of PP2Cs in subclades AI and AII.

References

Aravind, B., Nayak, S. N., Choudhary, R. S., Gandhadmath, S. S., Prasad, P. V. V., Pandey, M. K., et al. (2022). "Integration of genomics approaches in abiotic stress tolerance in groundnut (L.): An overview," in *Genomic Designing Abiotic Stress Resistant Oilseed Crops*, ed. C. Kole (Cham: Springer), 149197. doi: 10.1007/978-3-030-90044-1_4

Asif, M. A., Zafar, Y., Iqbal, J., Iqbal, M. M., Rashid, U., Ali, G. M., et al. (2011). Enhanced expression of *AtNHX1*, in transgenic groundnut (*Arachis hypogaea* l.) improves salt and drought tolerence. *Mol. Biotechnol.* 49 (3), 250–256. doi: 10.1007/s12033-011-9399-1

Ayatollahi, Z., Kazanaviciute, V., Shubchynskyy, V., Kvederaviciute, K., Schwanninger, M., Rozhon, W., et al. (2022). Dual control of MAPK activities by AP2C1 and MKP1 MAPK phosphatases regulates defence responses in *Arabidopsis*. *J. Exp. Bot.* 73 (8), 2369–2384. doi: 10.1093/jxb/erac018

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intelligent Syst. Mol. Biol.* 2, 28–36.

Banavath, J. N., Chakradhar, T., Pandit, V., Konduru, S., Guduru, K. K., Akila, C. S., et al. (2018). Stress inducible overexpression of *AtHDG11* leads to improved drought and salt stress tolerance in peanut (*Arachis hypogaea* l.). *Front. Chem.* 6. doi: 10.3389/fchem.2018.00034

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* l.). *Nat. Genet.* 51 (5), 877–884. doi: 10.1038/s41588-019-0405-z

Bhaskara, G. B., Wong, M. M., and Verslues, P. E. (2019). The flip side of phospho-signalling: regulation of protein dephosphorylation and the protein phosphatase 2Cs. *Plant Cell Environ.* 42 (10), 2913–2930. doi: 10.1111/pce.13616

Bhatnagar, N., Min, M. K., Choi, E. H., Kim, N., Moon, S. J., Yoon, I., et al. (2017). The protein phosphatase 2C clade a protein OsPP2C51 positively regulates seed germination by directly inactivating OsbZIP10. *Plant Mol. Biol.* 93 (4-5), 389–401. doi: 10.1007/s11103-016-0568-2

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294 (5), 1351–1362. doi: 10.1006/jmbi.1999.3310

Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4 (6), 1633–1649. doi: 10.1002/pmic.200300771

Bork, P., Brown, N. P., Hegyi, H., and Schultz, J. (1996). The protein phosphatase 2C (PP2C) superfamily: Detection of bacterial homologues. *Protein Sci. A Publ. Protein Soc.* 5 (7), 1421–1425. doi: 10.1002/pro.5560050720

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38 (12), 5825–5829. doi: 10.1093/molbev/msab293

Cao, J., Jiang, M., Li, P., and Chu, Z. (2016). Genome-wide identification and evolutionary analyses of the *PP2C* gene family with their expression profiling in response to multiple stresses in brachypodium distachyon. *BMC Genomics* 17, 175. doi: 10.1186/s12864-016-2526-4

Carrasco, J. L., Ancillo, G., Mayda, E., and Vera, P. (2003). A novel transcription factor involved in plant defense endowed with protein phosphatase activity. *EMBO J.* 22 (13), 3376–3384. doi: 10.1093/emboj/cdg323

Chao, J., Li, Z., Sun, Y., Aluko, O. O., Wu, X., Wang, Q., et al. (2021). MG2C: A user-friendly online tool for drawing genetic maps. *Mol. Horticul.* 1 (1), 1–4. doi: 10.1186/s43897-021-00020-x

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

Chen, X., Wang, Y., Lv, B., Li, J., Luo, L., Lu, S., et al. (2014). The NAC family transcription factor OsNAP confers abiotic stress response through the ABA pathway. *Plant Cell Physiol.* 55 (3), 604–619. doi: 10.1093/pcp/pct204

Chen, C., Yu, Y., Ding, X., Liu, B., Duanmu, H., Zhu, D., et al. (2018). Genome-wide analysis and expression profiling of PP2C clade d under saline and alkali stresses in wild soybean and. *Arabidopsis. Protoplasma* 255 (2), 643–654. doi: 10.1007/s00709-017-1172-2

Chu, M., Chen, P., Meng, S., Xu, P., and Lan, W. (2021). The *Arabidopsis* phosphatase PP2C49 negatively regulates salt tolerance through inhibition of AtHKT1;1. *J. Integr. Plant Biol.* 63 (3), 528–542. doi: 10.1111/jipb.13008

Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A developmental transcripttome map for allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* 7 1446. doi: 10.3389/fpls.2016.01446

Cui, M. H., Yoo, K. S., Hyoung, S., Nguyen, H. T., Kim, Y. Y., Kim, H. J., et al. (2013). An *Arabidopsis* R2R3-MYB transcription factor, AtMYB20, negatively regulates type 2C serine/threonine protein phosphatases to enhance salt tolerance. *FEBS Lett.* 587 (12), 1773–1778. doi: 10.1016/j.febslet.2013.04.028

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server, (2017 release). *Nucleic Acids Res.* 46 (W1), W49–W54. doi: 10.1093/nar/gky316

Das, A. K., Helps, N. R., Cohen, P. T., and Barford, D. (1996). Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 a resolution. *EMBO J.* 15 (24), 6798–6809. doi: 10.1002/j.1460-2075.1996.tb01071.x

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: The protein families database. *Nucleic 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

Geourjon, C., and Deléage, G. (1994). SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* 7 (2), 157–164. doi: 10.1093/protein/7.2.157

Gong, Z., Xiong, L., Shi, H., Yang, S., Herrera-Estrella, L. R., Xu, G., et al. (2020). Plant abiotic stress response and nutrient use efficiency. *Sci. China Life Sci.* 63 (5), 635674. doi: 10.1007/s11427-020-1683-x

Han, Y., Li, R., Liu, Y., Fan, S., Wan, S., Zhang, X., et al. (2021). The major intrinsic protein family and their function under salt-stress in peanut. *Front. Genet.* 12. doi: 10.3389/fgene.2021.639585

He, Z., Wu, J., Sun, X., and Dai, M. (2019). The maize clade a PP2C phosphatases play critical roles in multiple abiotic stress responses. *Int. J. Mol. Sci.* 20 (14), 3573. doi: 10.3390/ijms20143573

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinf. (Oxford England)* 31 (8), 1296–1297. doi: 10.1093/bioinformatics/btu817

Kaliff, M., Staal, J., Myrenås, M., and Dixelius, C. (2007). ABA is required for leptosphaeria maculans resistance *via* ABI1- and ABI4-dependent signaling. *Mol. Plant-Microbe interactions: MPMI* 20 (4), 335–345. doi: 10.1094/MPMI-20-4-0335

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10 (6), 845–858. doi: 10.1038/nprot.2015.053

Kerk, D., Bulgrien, J., Smith, D. W., Barsam, B., Veretnik, S., and Gribskov, M. (2002). The complement of protein phosphatase catalytic subunits encoded in the genome of *Arabidopsis*. *Plant Physiol.* 129 (2), 908–925. doi: 10.1104/pp.004002

Kerk, D., Templeton, G., and Moorhead, G. B. (2008). Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. *Plant Physiol.* 146 (2), 351–367. doi: 10.1104/pp.107.111393

Kim, S., Park, S. I., Kwon, H., Cho, M. H., Kim, B. G., Chung, J. H., et al. (2022). The rice abscisic acid-responsive RING finger E3 ligase OsRF1 targets OsPP2C09 for degradation and confers drought and salinity tolerance in rice. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.797940

Krzywińska, E., Kulik, A., Bucholc, M., Fernandez, M. A., Rodriguez, P. L., and Dobrowolska, G. (2016). Protein phosphatase type 2C PP2CA together with ABI1 inhibits SnRK2.4 activity and regulates plant responses to salinity. *Plant Signaling Behav.* 11 (12), e1253647. doi: 10.1080/15592324.2016.1253647

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

Kuhn, J. M., Boisson-Dernier, A., Dizon, M. B., Maktabi, M. H., and Schroeder, J. I. (2006). The protein phosphatase AtPP2CA negatively regulates abscisic acid signal transduction in *Arabidopsis*, and effects of abh1 on AtPP2CA mRNA. *Plant Physiol.* 140 (1), 127–139. doi: 10.1104/pp.105.070318

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

Lessard, P., Kreis, M., and Thomas, M. (1997). Protein phosphatases and protein kinases in higher plants. *Comptes rendus l'Academie Des. Sci. Serie III Sci. la vie* 320 (9), 675–688. doi: 10.1016/s0764-4469(97)84815-9

Leung, J., and Giraudat, J. (1998). Abscisic acid signal transduction. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 199–222. doi: 10.1146/annurev.arplant.49.1.199

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., et al. (2014). Genome sequence of the cultivated cotton *Gossypium arboretum*. *Nat. Genet.* 46 (6), 567–572. doi: 10.1038/ng.2987

Lim, C. W., Baek, W., Jung, J., Kim, J. H., and Lee, S. C. (2015). Function of ABA in stomatal defense against biotic and drought stresses. *Int. J. Mol. Sci.* 16 (7), 15251–15270. doi: 10.3390/ijms160715251

Liu, P. F., Chang, W. C., Wang, Y. K., Chang, H. Y., and Pan, R. L. (2008). Signaling pathways mediating the suppression of *Arabidopsis thaliana* Ku gene expression by abscisic acid. *Biochimica. Biophy. Acta* 1779 (3), 164–174. doi: 10.1016/j.bbagrm.2007.12.005

Liu, X., Zhu, Y., Zhai, H., Cai, H., Ji, W., Luo, X., et al. (2012). AtPP2CG1, a protein phosphatase 2C, positively regulates salt tolerance of *Arabidopsis* in abscisic acid-dependent manner. *Biochem. Biophys. Res. Commun.* 422 (4), 710–715. doi: 10.1016/j.bbrc.2012.05.064

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods (San Diego Calif.)* 25 (4), 402–408. doi: 10.1006/meth.2001.1262

Luo, L., Wan, Q., Zhang, K., Zhang, X., Guo, R., Wang, C., et al. (2021). AhABI4s negatively regulate salt-stress response in peanut. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.741641

Min, M. K., Choi, E. H., Kim, J. A., Yoon, I. S., Han, S., Lee, Y., et al. (2019). Two clade a phosphatase 2Cs expressed in guard cells physically interact with abscisic acid signaling components to induce stomatal closure in rice. *Rice (New York N.Y.)* 12 (1), 37. doi: 10.1186/s12284-019-0297-7

Ohta, M., Guo, Y., Halfter, U., and Zhu, J. K. (2003). A novel domain in the protein kinase SOS2 mediates interaction with the protein phosphatase 2C ABI2. *Proc. Natl. Acad. Sci. United States America* 100 (20), 11771–11776. doi: 10.1073/pnas.2034853100

Pang, Q., Zhang, T., Zhang, A., Lin, C., Kong, W., and Chen, S. (2020). Proteomics and phosphoproteomics revealed molecular networks of stomatal immune responses. *Planta* 252 (4), 66. doi: 10.1007/s00425-020-03474-3

Raza, A., Tabassum, J., Fakhar, A. Z., Sharif, R., Chen, H., Zhang, C., et al. (2022). Smart reprograming of plants against salinity stress using modern biotechnological tools. *Crit. Rev. Biotechnol.*, 1–28. doi: 10.1080/07388551.2022.2093695. Advance online publication.

Rodriguez, P. L., Benning, G., and Grill, E. (1998). ABI2, a second protein phosphatase 2C involved in abscisic acid signal transduction in *Arabidopsis*. *FEBS Lett.* 421 (3), 185–190. doi: 10.1016/s0014-5793(97)01558-5

Schweighofer, A., Hirt, H., and Meskiene, I. (2004). Plant PP2C phosphatases: emerging functions in stress signaling. *Trends Plant Sci.* 9 (5), 236–243. doi: 10.1016/j.tplants.2004.03.007

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Shazadee, H., Khan, N., Wang, J., Wang, C., Zeng, J., Huang, Z., et al. (2019). Identification and expression profiling of protein phosphatases (*PP2C)* gene family in *Gossypium hirsutum* l. *Int. J. Mol. Sci.* 20 (6), 1395. doi: 10.3390/ijms20061395

Shi, Y. (2009). Serine/threonine phosphatases: mechanism through structure. *Cell* 139 (3), 468–484. doi: 10.1016/j.cell.2009.10.006

Singh, A., Giri, J., Kapoor, S., Tyagi, A. K., and Pandey, G. K. (2010). Protein phosphatase complement in rice: Genome-wide identification and transcriptional analysis under abiotic stress conditions and reproductive development. *BMC Genomics* 11, 435. doi: 10.1186/1471-2164-11-435

Singh, A., Jha, S. K., Bagri, J., and Pandey, G. K. (2015). ABA inducible rice protein phosphatase 2C confers ABA insensitivity and abiotic stress tolerance in *Arabidopsis*. *PloS One* 10 (4), e0125168. doi: 10.1371/journal.pone.0125168

Song, S. K., Yun, Y. B., and Lee, M. M. (2020). POLTERGEIST and POLTERGEIST-LIKE1 are essential for the maintenance of post-embryonic shoot and root apical meristems as revealed by a partial loss-of-function mutant allele of pll1 in *Arabidopsis*. *Genes Genomics* 42 (1), 107–116. doi: 10.1007/s13258-019-00894-8

Stone, J. M., Collinge, M. A., Smith, R. D., Horn, M. A., and Walker, J. C. (1994). Interaction of a protein phosphatase with an arabidopsis serine-threonine receptor kinase. *Sci. (New York N.Y.)* 266 (5186), 793–795. doi: 10.1126/science.7973632

Strizhov, N., Abrahám, E., Okrész, L., Blickling, S., Zilberstein, A., Schell, J., et al. (1997). Differential expression of two *P5CS* genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 and AXR2 in *Arabidopsis*. *Plant J. Cell Mol. Biol.* 12 (3), 557–569. doi: 10.1046/j.1365-313x.1997.00557.x

Subudhi, P. K., Shankar, R., and Jain, M. (2020). Whole genome sequence analysis of rice genotypes with contrasting response to salinity stress. *Sci. Rep.* 10 (1), 21259. doi: 10.1038/s41598-020-78256-8

Sun, B. R., Fu, C. Y., Fan, Z. L., Chen, Y., Chen, W. F., Zhang, J., et al. (2019). Genomic and transcriptomic analysis reveal molecular basis of salinity tolerance in a novel strong salt-tolerant rice landrace changmaogu. *Rice (New York N.Y.)* 12 (1), 99. doi: 10.1186/s12284-019-0360-4

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi: 10.1093/molbev/msab120

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22 (22), 4673–4680. doi: 10.1093/nar/22.22.4673

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 (W1), W296–W303. doi: 10.1093/nar/gky427

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., et al. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol. (Clifton N.J.)* 112, 531–552. doi: 10.1385/1-59259-584-7:531

Xue, T., Wang, D., Zhang, S., Ehlting, J., Ni, F., Jakab, S., et al. (2008). Genome-wide and expression analysis of protein phosphatase 2C in rice and *Arabidopsis*. *BMC Genomics* 9, 550. doi: 10.1186/1471-2164-9-550

Yang, Q., Liu, K., Niu, X., Wang, Q., Wan, Y., Yang, F., et al. (2018). Genome-wide identification of *PP2C* genes and their expression profiling in response to drought and cold stresses in *Medicago truncatula*. *Sci. Rep.* 8 (1), 12841. doi: 10.1038/s41598-018-29627-9

Yu, X., Han, J., Wang, E., Xiao, J., Hu, R., Yang, G., et al. (2019). Genome-wide identification and homoeologous expression analysis of *PP2C* genes in wheat (*Triticum aestivum* l.). *Front. Genet.* 10. doi: 10.3389/fgene.2019.00561

Zhang, J., Li, X., He, Z., Zhao, X., Wang, Q., Zhou, B., et al. (2013). Molecular character of a phosphatase 2C (*PP2C*) gene relation to stress tolerance in *Arabidopsis thaliana*. *Mol. Biol. Rep.* 40 (3), 2633–2644. doi: 10.1007/s11033-012-2350-0

Zhang, Z., Luo, S., Liu, Z., Wan, Z., Gao, X., Qiao, Y., et al. (2022). Genome-wide identification and expression analysis of the cucumber *PYL* gene family. *PeerJ* 10, e12786. doi: 10.7717/peerj.12786

Zhang, G., Zhang, Z., Luo, S., Li, X., Lyu, J., Liu, Z., et al. (2022). Genome-wide identification and expression analysis of the cucumber PP2C gene family. *BMC Genomics* 23 (1), 563. doi: 10.1186/s12864-022-08734-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fpls.2023.1102181

OPEN ACCESS

EDITED BY
Fei Shen,
Beijing Academy of Agricultural and Forestry Sciences, China

REVIEWED BY
Sunil S. Gangurde,
University of Georgia, United States
Zhichao Wu,
National Institutes of Health (NIH), United States

*CORRESPONDENCE
Zhuang Weijian
weijianz@fafu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 18 November 2022
ACCEPTED 06 January 2023
PUBLISHED 09 February 2023

CITATION
Zhuang Y, Sharif Y, Zeng X, Chen S, Chen H, Zhuang C, Deng Y, Ruan M, Chen S and Weijian Z (2023) Molecular cloning and functional characterization of the promoter of a novel *Aspergillus flavus* inducible gene (*AhOMT1*) from peanut. *Front. Plant Sci.* 14:1102181. doi: 10.3389/fpls.2023.1102181

Molecular cloning and functional characterization of the promoter of a novel *Aspergillus flavus* inducible gene (*AhOMT1*) from peanut

Yuhui Zhuang[1,2†], Yasir Sharif[1†], Xiaohong Zeng[1], Suzheng Chen[1], Hua Chen[1], Chunhong Zhuang[1], Ye Deng[1], Miaohong Ruan[3], Shuanglong Chen[3] and Zhuang Weijian[1*]

[1]Center of Legume Plant Genetics and Systems Biology, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China, [2]College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, China, [3]Fujian Seed General Station, Fuzhou, China

Peanut is an important oil and food legume crop grown in more than one hundred countries, but the yield and quality are often impaired by different pathogens and diseases, especially aflatoxins jeopardizing human health and causing global concerns. For better management of aflatoxin contamination, we report the cloning and characterization of a novel *A. flavus* inducible promoter of the O-methyltransferase gene (*AhOMT1*) from peanut. The *AhOMT1* gene was identified as the highest inducible gene by *A. flavus* infection through genome-wide microarray analysis and verified by qRT-PCR analysis. *AhOMT1* gene was studied in detail, and its promoter, fussed with the *GUS* gene, was introduced into Arabidopsis to generate homozygous transgenic lines. Expression of GUS gene was studied in transgenic plants under the infection of *A. flavus*. The analysis of *AhOMT1* gene characterized by in silico assay, RNAseq, and qRT-PCR revealed minute expression in different organs and tissues with trace or no response to low temperature, drought, hormones, Ca2+, and bacterial stresses, but highly induced by *A. flavus* infection. It contains four exons encoding 297 aa predicted to transfer the methyl group of S-adenosyl-L-methionine (SAM). The promoter contains different cis-elements responsible for its expression characteristics. Functional characterization of *AhOMT1*P in transgenic Arabidopsis plants demonstrated highly inducible behavior only under *A. flavus* infection. The transgenic plants did not show GUS expression in any tissue(s) without inoculation of *A. flavus* spores. However, *GUS* activity increased significantly after inoculation of *A. flavus* and maintained a high level of expression after 48 hours of infection. These results provided a novel way for future management of peanut aflatoxins contamination through driving resistance genes in *A. flavus* inducible manner.

KEYWORDS

aflatoxins, *Arachis hypogaea*, hepatocellular carcinoma, inducible promoter, pathogens, functional characterization

Introduction

Aflatoxin (AF) contamination produced by *Aspergillus flavus* is the key limitation for peanut production worldwide, which not only causes enormous economic losses but also jeopardizes animal and human health. Since the first discovery of aflatoxin-contaminated peanut meals leading to the death of more than 100,000 turkey birds in England in the 1960s (Richard, 2008), aflatoxin contamination has caused heavy loss to human and live stocks now and again (Pandey et al., 2019). Aflatoxins are a class of mycotoxins produced by species of the *Aspergillus* genus, mainly including *A. flavus*, *A. parasiticus*, *A. nomius*, and *A. tamarii* (Goto et al., 1996; Payne et al., 2006; Pandey et al., 2019). Among the aflatoxins, B1 and B2 synthesized by *A. flavus* and *A. parasiticus* in peanut and corn are the most toxic secondary metabolites, carcinogenic, hepatotoxic, immunosuppressive, and teratogenic to mankind and animals (Amaike and Keller, 2011; Kensler et al., 2011). As aflatoxin is stable in nature and decomposes at 237–306 °C, it may contaminate ordinary foods like meat, eggs, and milk in the food chain leading to wide detriments to humans (Awasthi et al., 2012; Pandey et al., 2019). *A. flavus* can contaminate peanut during the growth stage, called pre-harvest infection, and after harvest, in drying, storage, transport, and processing, called post-harvest infection (Khan et al., 2020; Soni et al., 2020). The pre-harvest infection in crops is the key to post-harvest contamination (Khan and Zhuang, 2019). How to reduce or even resolve the aflatoxin contamination problem in peanut and peanut products has been attracting global attention.

As an important protein, oil-providing crop, and valuable nutrition source for humankind and animals, peanut is affected by several bacterial and fungal diseases. Peanut seeds develop underground and are in continuous contact with soil fungal microbiota. Aflatoxin-producing fungal species live in soil as conidia or sclerotia, while in plants, they live in mycelial form. Under harsh environmental conditions, sclerotia in the soil develop conidia and probably ascospores (Horn et al., 2009), resulting in an increased population under warm and dry conditions. Under suitable growth conditions, *Aspergillus* species invade peanut seeds and deteriorate peanut yield and quality. Breeding Aflatoxin-resistant varieties are the most effective and safe method for managing aflatoxin contamination. But, so far, no bred or released varieties have shown stable resistance to *A. flavus* in the world (Pandey et al., 2020), which makes aflatoxin control a challenging task.

In recent years, the successes of genetic engineering and breeding methods have made it possible to breed aflatoxin-resistant peanut varieties. Success stories are available in some crops, such as the success of Bt-cotton (Fleming et al., 2018; Gutierrez, 2018), which has highly increased yields and reduced the use of insecticides, and the golden rice producing functional nutritions (Tang et al., 2009). At present, nearly all promoters being used in genetic engineering are constitutive, like CaMV 35S promoter, Ubi, and Act-1 (Kay et al., 1987; McElroy et al., 1990; Cornejo et al., 1993), all of them can direct expression virtually in all tissues without any influence by internal and external factors. These promoters not only waste the energy of plants but also bring some unsafety defects to consumers, which has caused broad transgenic safety concerns. To breed the aflatoxin-resistant peanut varieties, *A. flavus*-induced promoters, which only express when plants undergo *A. flavus* infection, would be clearly a good choice. Inducible promoters are of great importance as they don't pose an additional load on plant metabolism compared to constitutive promoters (Divya et al., 2019). Thus, identification and investigation of Aflatoxins-induced genes and their promoters should be important and have potential applications in creating resistance peanut to aflatoxin contamination.

Plant O-methyltransferases (OMTs) transfer the methyl group of *S-adenosyl-L-methionine* (SAM) to the hydroxyl group of numerous organic chemical compounds, ultimately resulting in the synthesis of the methyl ether variants of these substances (Struck et al., 2012). Plant OMTs are primarily classified into three groups: C-methyltransferases, N-methyltransferases, and O-methyltransferases (OMTs) (Roje, 2006). O-methyltransferases are involved in concern for cresistance mechanism in crop plants against several fungal and bacterial pathogens. OMT members are reported to have fusarium blight resistance in wheat (Yang et al., 2021). OMT members are involved in lodging resistance in wheat by playing a role in regulating lignin biosynthesis and resistance against various environmental stresses (Nguyen et al., 2016). Yang et al. (2019) reported that a wheat OMT improved drought resistance in transgenic *Arabidopsis* (Yang et al., 2019). Another member of wheat OMT (*TaCOMT-3D*) enhances tolerance against eyespot disease and improves the mechanical strength of the stem (Wang et al., 2018). In maize, OMT members enhance PEG and drought tolerance (Qian et al., 2014). Plant OMTs participate in the structural diversification of different chemical compounds; the methylation alters the solubility and reactivity of natural compounds, resulting in a change in their biological processes (Gang et al., 2002). Due to their important roles in secondary metabolism, OMTs are widely investigated in plants. However, no reports have touched the *OMTs* associated with *A. flavus* infection, especially its highly inducible response to the *A. flavus* inoculation, implying an important use for aflatoxin management.

In searching for *A. flavus*-induced promoters, a genome-wide large-scale microarray study was performed (Chen et al., 2016; Zhang et al., 2017), an O-methyltransferase gene (*AhOMT1*) was found to have a gigantic response to the *A. flavus* infection. Then, we studied the aflatoxin inducible promoter of that O-methyltransferase gene (*AhOMT1*). Our study aimed to identify and clone the *OMT1* and the aflatoxins inducible promoter that can be used to drive aflatoxins-resistant genes for future use. We also performed the functional investigation of this promoter in *Arabidopsis* plants and provided its practical significance. If genes with high resistance under the control of aflatoxin inducible promoters are transformed in peanut, it would be an outstanding achievement in obtaining aflatoxin resistance.

Materials and methods

Plant materials and growth conditions

In this study, we used peanut cultivars Minhua-6 (M6), Xinhuixiaoli (XHXL), and *Arabidopsis* Columbia-0 (Col-0). XHXL is a highly *A. flavus*-resistant variety. The peanut and *Arabidopsis* seeds were maintained by the Oil Crops Research Institute, Fujian Agriculture and Forestry University. Peanut plants were grown in

research fields and the greenhouse of Fujian Agriculture University. *Arabidopsis* plants were grown in small plastic pots (8 cm diameter). Peanut and *Arabidopsis* plants were maintained at 25°C and 16/8 h day/night photoperiod.

Fungal strain growth and inoculation

A highly toxic strain of *A. flavus* (AF2) was grown on Potato Dextrose Agar medium PDA (Potato=200g, Dextrose=20g, Agar=15g, water up to 1L). The fungal isolates were grown on PDA in a 100 mL flask at 28 °C for one week. Green spores of AF2 were collected in a solution of 0.01% tween 20. The spore concentration was diluted to 1.9 $\times 10^6$ conidia mm^{-1} (Li et al., 2006). The spore solution was used to spray the peanut leaves, and samples were taken at 3h, 6h, 9h, 12h, 24h, 36h, and 48h after inoculation. Non-treated sampled (0h) were used as a control to compare the expression of the *AhOMT1* gene.

Selection of aflatoxins inducible gene

The screening of aflatoxin inducible genes was based on our previous microarray studies (Chen et al., 2016; Zhang et al., 2017). Differential expression profiles were compared among different peanut tissues, including the maturing pods which were infected by *A. flavus* plus drought (PAFDR, peanut *A. flavus* inoculation plus drought), dealing with drought (PDR-af, peanut in drought stress, no *A. flavus* inoculation) and normal watering (PAFCK, peanut without *A. flavus* inoculation as for control), using microarray with oligonucleotides of 60bp containing 12x135K arrays. Trizol reagent was used to extract the total RNA, and array hybridization, washing, scanning, and data analysis were performed according to NimbleGen's Expression user's guide for expression comparisons. From the microarray expression, *A. flavus* inducible specific expression gene fragments were isolated. A member of the peanut OMT family, "Isoliquiritigenin 2'-O-methyltransferase", was found as a highly upregulated gene.

Cloning and analysis of the *AhOMT1* gene

The full-length *AhOMT1* gene was cloned from the peanut pericarp cDNA infected by *A. flavus*. Based on the target gene sequence, the full-length gene was amplified by RACE (Rapid Amplification of cDNA Ends) with the special primers, RACE-F and *AhOMT1*-R for the 5'-upstream fragment cloning; RACE-R and *AhOMT1*-F for the 3' downstream fragment cloning. Primer sequences are given in Table S2. The "*Isoliquiritigenin 2'-O-methyltransferase*" gene with an mRNA Id "AH13G54850.1" is present on the chromosome 13th of cultivated peanut (http://peanutgr.fafu.edu.cn/) (Zhuang et al., 2019). As this gene is being studied for the first time in peanut, we renamed it *AhOMT1* for convenience.

We used BioXM 2.6 software to predict the open reading frame of the splicing sequence. Sequence homology analysis was performed using BLASTP at NCBI, TAIR (https://www.arabidopsis.org/) (Lamesch et al., 2012), and Legume Information System LIS (https://legacy.legumeinfo.org/) (Gonzales et al., 2005). The protein sequence was also submitted to ScanProsite and NCBI databases to analyze the protein binding sites and functional domains. Protein physical and chemical parameters were analyzed by ExPASy (Gasteiger et al., 2003). The evolutionary reconstructions of OMT1 proteins from different legume species and peanut were studied by constructing a rooted phylogenetic tree. Protein sequences were aligned by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/), and a rooted tree was constructed by Dendroscope 3 (Huson and Scornavacca, 2012).

Cloning and analysis of *AhOMT1* promoter

Cloning of the *AhOMT1* promoter is based on the previous Nested PCR analysis. Genomic DNA was extracted from the young roots of peanut M6 using the CTAB method. Peanut Genome Walker libraries were constructed using our own library's approach. The genomic DNA was incompletely digested by three restriction enzymes (AseI, EcoRI, and HindIII) and then ligated the fragments to adapters to construct three genomic libraries using the adaptor primers, ongAd, ShortHindIII, ShortEcoRI, and ShortAseI (Table S2).

Genomic libraries were used as the PCR temple to obtain the upstream sequence by Flanking PCR. Two forward primers (AP1, AP2-C) and two reverse primers (AhOMT1-SP1, AhOMT1-SP2) were designed to clone the promoter of the *AhOMT1* gene (Table S2). Finally, a primers pair was designed to clone the promoter from the genomic DNA of peanut; the forward primer (AhAF7-F) was designed upstream of the coding sequence, and the reverse primer (AhAF7-R) was designed at the *AhOMT1* gene sequence (Table S2). The PCR amplified fragment was visualized through 2% agarose gel, purified by TIANGEN Universal DNA Purification Kit (TIANGEN Biotech Beijing, China), and sequence was verified by Beijing Genomics Institute (BGI, Shenzhen, China).

A 1698 bp region upstream of the start codon of the *AhOMT1* gene was selected for promoter analysis. *Cis*-regulatory elements of the promoter region were predicted by the PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace) (Higo et al., 1998) and the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts et al., 1999).

Verification of *A. flavus* inducible expression of *AhOMT1* gene

Following the *A. flavus* inoculation of peanut leaves, samples were taken and preserved at -80°C. RNA was extracted by the Cetyl Trimethyl Ammonium Bromide (CTAB) method with few modifications (Chen et al., 2019). Following the manufacturer's protocol, 1μg RNA was reverse transcribed to synthesize the cDNA with the PrimeScript 1st strand cDNA Synthesis Kit (Takara, Dalian, China). To check the real-time expression of the *AhAOMT1* gene in response to *A. flavus* infection, a qRT-PCR reaction was performed using the MonAmp^{TM} ChemoHS qPCR Mix (Monad Biotech, Wuhan, China). The reaction was prepared according to the product guidelines. The peanut *Actin* gene was used as a control

(Chi et al., 2012), and the reaction was carried out in Applied Biosystems 7500 real-time PCR system (USA) at 94°C (1 min), 60° C (1 min), and 72°C (1 min) for 40 cycles. Primers used for RT-qPCR are given in Table S2.

Vector construction and transformation into *Arabidopsis*

A two-step Gateway cloning system was used to construct the plant expression vector using the pMDC164 vector. The pMDC164 vector is a plant expression vector that contains the *GUS* reporter gene and hygromycin antibiotic resistance gene. After sequence verification, the *AhOMT1* promoter fragment was ligated between *attP* sites of entry vector pDONR-207 by BP reaction. The promoter sequence was verified again and then cloned into the pMDC164 vector between *attR* sites by Gateway LR cloning with the primers containing universal Gateway adapter sequences.

Constructed vector *AhOMT1P::GUS* was transformed into *A. tumefaciens* (GV3101) and KM+ (50 μg/mL) resistance colonies were selected on YEB and verified by PCR. The Transformed *A. tumefaciens* cells were grown to the OD_{600} of 1.0-1.5. Bacterial cells were harvested by centrifugation and resuspended sucrose solution (5%), also containing Silwet L-77 (0.02%) and Acetosyringone (AS) (100μg/mL). The floral dip method (Clough and Bent, 1998) was employed for the genetic transformation of mature *Arabidopsis* plants. Opened flowers were clipped off before floral dipping, and the transformation was repeated after five days. Plants were grown normally following the transformation procedure, and mature seeds were collected. T0 seeds were sterilized with ethanol (75%) and 10% H_2O_2 and grown on MS medium containing 50μg/mL Hygromycin for screening of positively transformed plants. From Hygromycin-resistant plants, eight were randomly selected for genetic verification for the presence of promoter by PCR with the promoter and *GUS* gene-specific forward and reverse primers (Table S2). PCR-verified plants were grown to get the homozygous T3 generation for further functional studies. In each generation, greater care was carried out to avoid outcrossing, while Hygromycin and PCR-based verification were repeated in each generation.

Functional study of *AhOMT1* gene in response to *A. flavus*

The *AhOMT1* promoter expression was analyzed by studying the behavior of the *GUS* gene in transgenic plants. Histochemical staining and quantitative expression of the *GUS* gene were checked in response to *A. flavus* infection. Transgenic plants were divided into three groups, each infected with *A. flavus*. For *A. flavus* infection, the spore solution was prepared as mentioned above and used to spray the young plants. Leaf samples were collected before treatment (0h) and after infection, 1h, 3h, 6h, 12h, 24h, and 36h. GUS staining solution 2mM 5-Bromo-4-chloro-3-indolyl β-D-glucuronide (X-Gluc) was prepared in 10 mM EDTA, 0.1% Triton X-100, 2 mM Potassium ferricyanide, 50 mM sodium phosphate buffer, and 2 mM Potassium ferrocyanide (Jefferson et al., 1987). Samples were incubated in GUS solution at 37°C for 12 hours. After that, samples were washed and decolorized with 75% ethanol. Photographs of stained samples were taken with an OLYMPUS microscope (BX3-CBH) and digital camera. For quantitative expression analysis of *GUS* gene, qRT-PCR was performed with samples taken from *A. flavus*-stressed plants. RNA was extracted with Trizol reagent, as mentioned above. Transgenic plants were also treated with different plant hormones, including salicylic acid (SA, 3mmol/L), ethephon (ETH, 1mg/ml), Brassinolide (BR, 0.1 mg/L), abscisic acid (ABA, 10μg/ml) and paclobutrazol (PAC, 150mg/L) to check that either *AhOMT1* promoter is affected by these hormones or not. Plants were treated with distilled water as a control group.

Results

Selection and verification of *A. flavus* inducible *AhOMT1* gene

Aflatoxin inducible genes were selected from a previous microarray expression study. Microarray analysis was performed on peanut cultivar Minhua-6 plants treated with drought, a combination of drought and *A. flavus* infection, and a control group. Plants with normal watering were taken as control group. Gene with high microarray expression under drought and *A. flavus* infection were selected as *A. flavus* inducible genes. Among the upregulated genes, a member of peanut O-methyltransferase family "Isoliquiritigenin 2'-O-methyltransferase" (*AhOMT1*) with the probe Id "GFBZZHF01DILSM" was found to be highly inducible. Expression levels of ten highly *A. flavus* inducible genes are shown in Figures S1A, B (Log2 normalized microarray expression values are used), while the relative expression values are given in Table 1. High-density oligo-nucleotide microarray with 100,000 unigenes, including the *AhOMT1* gene and relative expression levels of the *AhOMT1* gene across different tissues, is given in Figures S1A–C. Expression level of the *AhOMT1* gene under control and drought conditions was almost similar, but it recorded a high increase in microarray expression in response to a combination of drought and *A. flavus* infection. An almost 200-fold increase was observed upon *A. flavus* inoculation. *AhOMT1* recorded the highest increase in expression among inducible genes, so it was selected for the promoter analysis.

For evaluation of *AhOMT1*, RT–PCR was performed using total RNA obtained from root, stem, leaf, flower, pericarp, testa embryo, and aflatoxins-inoculated pericarp tissues. The result showed that the gene was nearly unexpressed in these peanut organs without any stress treatments (Figures 1A, B). We also performed the semi-quantitative RT-PCR analysis and got consistent results. However, *AhOMT1* detected high expression in the peanut pericarp infected with *A. flavus* (Figure 1B). These findings are in accordance with the analysis of the cDNA microarray in two sets of *in silic* expression for *AhOMT1* gene evaluation (Figure S1C; Table S1-1, 2). The result confirmed that *AhOMT1* is an *A. flavus*-induced gene.

Further evaluation of the *AhOMT1* inducible behaviors was performed by qRT-PCR analysis and another set of microarray analyses with various stresses (Figures 2A, B, S1-B). For the qRT assay, the peanut plants were challenged with *A. flavus*. The expression level of *AhOMT1* in leaves was observed at 1h, 3h, 6h, 9h, 12h, 24h, 36h, and 48h after inoculation. Data were analyzed by the $2^{-\Delta\Delta Ct}$ method while using Actin as control gene. *AhOMT1* gene

10.3389/fpls.2023.1102181

TABLE 1 Microarray expression of some *A. flavus* inducible genes.

Probe Id	Gene name	Microarray expression			
		Normal watering	Drought	Drought + *A. flavus*	Folds of upregulation
GFBZZHF02GJYBV	40S ribosomal protein S5 (Fragment)	12	14	9194	656.71
GALJ6PP01B5DBY	Probable F-box protein At4g22030	15	13	3562	274.00
GAYHEI101EGW7N	Protein GLUTAMINE DUMPER 1	888	31	7672	247.48
GALJ6PP01DP2GR	EG45-like domain containing protein	13	11	2751	250.09
GD3IWDR02CZXC6	Uncharacterized protein	19	22	4752	216.00
GFBZZHF01DILSM	Isoliquiritigenin 2'-O-methyltransferase	49	86	17355	201.80
GD3IWDR02DXEVJ	Laccase-14	31	29	4462	153.86
GFBZZHF02F852U	Probable inactive purple acid phosphatase 1	11	26	3856	148.31
GALJ6PP01A9Y8O	Unknown function	17	26	3793	145.88
GAYHEI101CFH5I	Cysteine-rich receptor-like protein kinase 3	40	28	2358	84.21

* Microarray expression under different treatments.

was expressed gradually upon *A. flavus* treatment at 1h, 3h, and 6h after inoculation (Figure 2A). After 6h, *AhOMT1* showed decreased expression, but the expression level was still much higher compared to the control (without treatment). Microarray studies showed that compared with actin-7, *AhOMT1* expressed weakly in different organs and tissues and did not respond to environmental challenges such as low temperature, drought, deficient Ca^{2+}, bacterium *Ralstonia solanacearum* inoculation, and also to hormones treatments but upregulated by Ethephon (Figure S2B; Tables S1, 2). These results indicate that the *AhOMT1* gene is highly expressed under *A. flavus* infection.

Cloning and characterization of *AhOMT1* gene

The cDNA of the *AhOMT1* gene was first obtained from *Arachis hypogaea* cv. Minhua-6 pericarp cDNA infected with *A. flavus* by RACE approach using the designed specific primers (Table S2). After obtaining the 5' upstream and 3' downstream cDNA sequence of *AhOMT1*, two specific oligo primers were designed to clone the entire cDNA and genomic DNA. The gene *AhOMT1* contained a 1,293 bp full-length cDNA (not counting 31bp of polyA) comprised of a 241 bp 5'UTR, an 894 bp long ORF and a 158bp of 3'UTR (Figure S2).

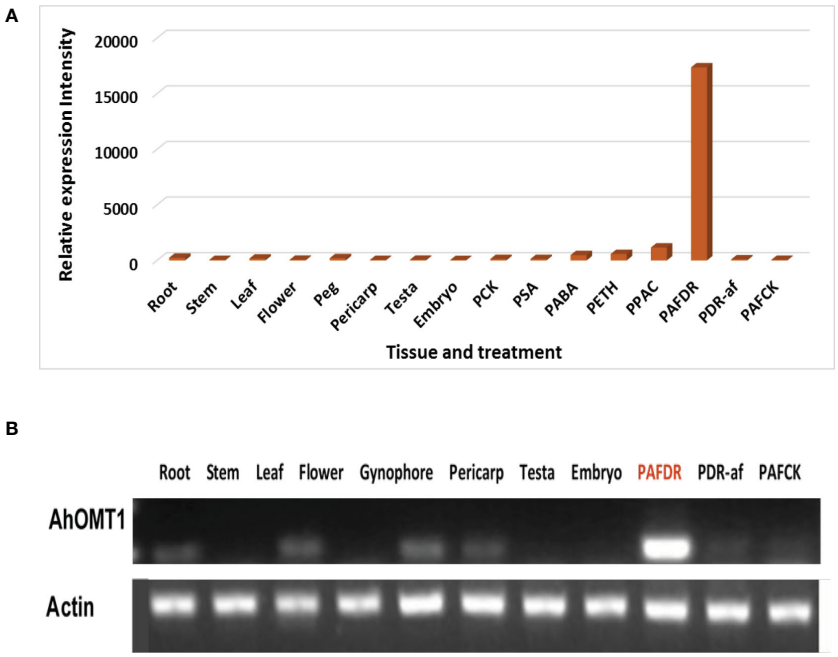


FIGURE 1
A. flavus inducible gene screening and *AhOMT1* expression in peanut. **(A, B)** *AhOMT1* gene showed very week or no expression in different tissues (root, stem, leaf, flower, pericarp, testa, and embryo), but gigantic upregulation induced by *A. flavus* by both microarray and RT-PCR analysis. The DNA template from the pericarp treated with drought (PDR-af), drought plus *Aspergillus flavus* infection (PAFDR), and normal watering (PAFCK).

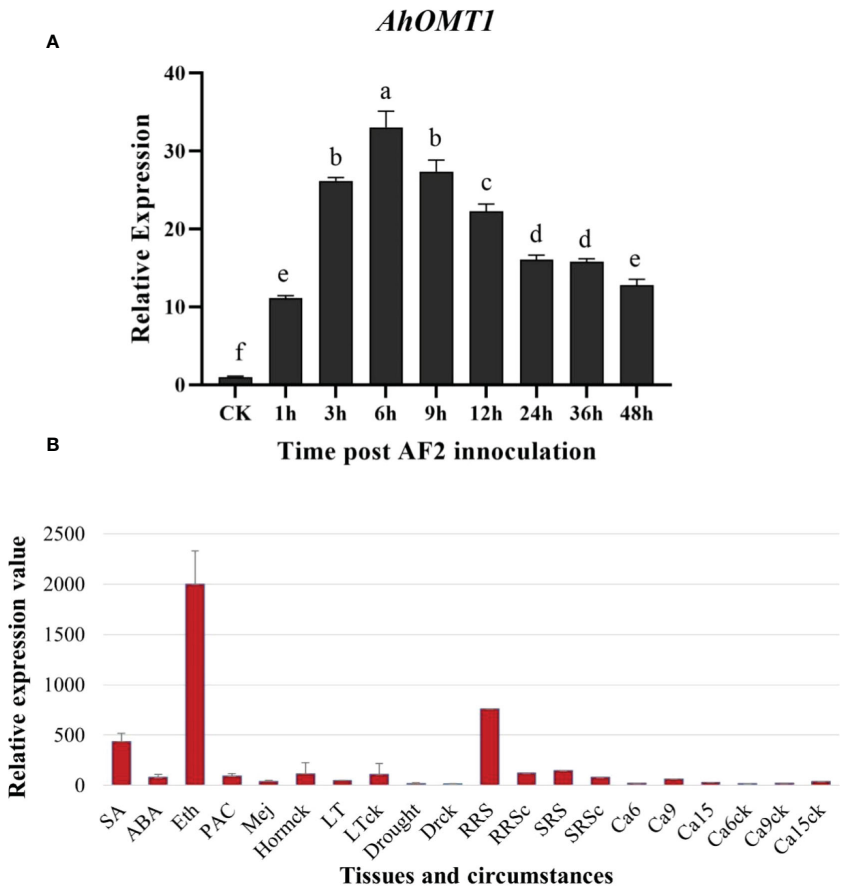


FIGURE 2
Characteristics of *AhOMT1* expression in peanut. **(A)** Expression analysis of *AhOMT1* gene in response to *A. flavus* infection determined by qRT-PCR. *AhOMT1* showed high level of expression upregulation upon *A. flavus* infection in peanut leaves. Expression was increased after one hour of inoculation that showed increasing trend upto 6 hours. After 6 hours, expression showed a decreased trend but still was very high as compared to non-treated leaves. **(B)** *AhOMT1* gene demonstrated little responses to hormones and environmental stresses besides Ethephon treatment (Eth) with upregulation. SA, salicylic acid; ABA, abscisic acid; PAC, paclobutrazol; Mej, Methyl jasmonate; LT, low temperature; RRS, *Ralstonia solanacearum* inoculation on resistance peanut; RRSc, resistance cultivar without inoculation; SRS, *R. solanacearum* inoculated on susceptible cultivar. SRSc, susceptible cultivar without inoculation; Ca6/9/15: embryo in deficient calcium at 6, 9 and 15 days after pegging; Ca6/9/15ck: embryos in sufficient calcium at 6, 9 and 15 days after pegging.

The translated coding sequence encodes a 297 amino acid protein. The comparison of genomic and cDNA sequences showed that ORF contains three introns (Figures 3A, S2) of 130 to 816 bp length, and all introns had a usual 5'GT and 3'AG ends (Ballance, 1986). This supported but revised the annotation of *AhOMT1* gene structure with 1086 bp ORF and 49 bp 5'UTR (*AH13G54850.1*) in Chromosome 13, a 192bp upstream fragment being predicted to be a coding region (Zhuang et al., 2019). Presence of Methyltransf_2 superfamily domain provided that *AhOMT1* belongs to O-Methyltansferase gene family (Figure 3B).

OMT1 homologous genes from other legume species were identified from LIS and TAIR databases. The percent identity was found by BLASTP in DNAMAN8.0 software. As *AhOMT1* is present on 13th chromosome (B sub-genome), derived from ancestral species *A. ipaensis*, and the results were consistent to expectation. It showed 97.41% identity with *A. ipaensis* (Araip.Z3XZX.1), while identity with *A. duranensis* (Aradu.PF1EJ.1) is 56.15%, indicating a rapid diversification of it. *AhOMT1* showed 70.99% identity with *G. max* OMT1 homolog (Glyma.09G281900.1) and 57.76% identity with *Arabidopsis* (At5G54160.1) (Table 2). Phylogenetic analysis revealed a well-supported division between methyltransferase family groups, and the *AhOMT1* lies within a distinct clade containing the *AhOMT1* (Figure 3C), providing further support that *AhOMT1* is a gene encoding Isoliquiritigenin 2'-O-methyltransferase.

The protein sequence analysis by ExPASy showed that the *AhOMT1* had a molecular weight (MW) of 33532.5 and a theoretical isoelectric point (pI) of 6.49. The instability index (II) is computed to be 44.34, indicating an unstable protein. The protein was predicted to be a hydrophilic protein with no signal peptide region (Figure S3A–4B). And the prediction of protein hydropathicity also showed that *AhOMT1* is a hydrophilic protein (Figure S3C). These indicated that *AhOMT1* might be a peripheral protein. The conserved domain analyzed by ScanProsite showed that the domain contains five S-adenosyl-L-methionine binding sites and a proton acceptor site (Figure S3D). The protein binding sites and functional domains of the *AhOMT1* gene analyzed by the NCBI database suggested that the gene is a member of the Methyltransf_2 superfamily (Figure S3E). This family contains a range of O-methyltransferases. These enzymes utilize S-adenosyl methionine (Kozbial and Mushegian, 2005). All these indicated this gene encoded a putative O-methyltransferase.

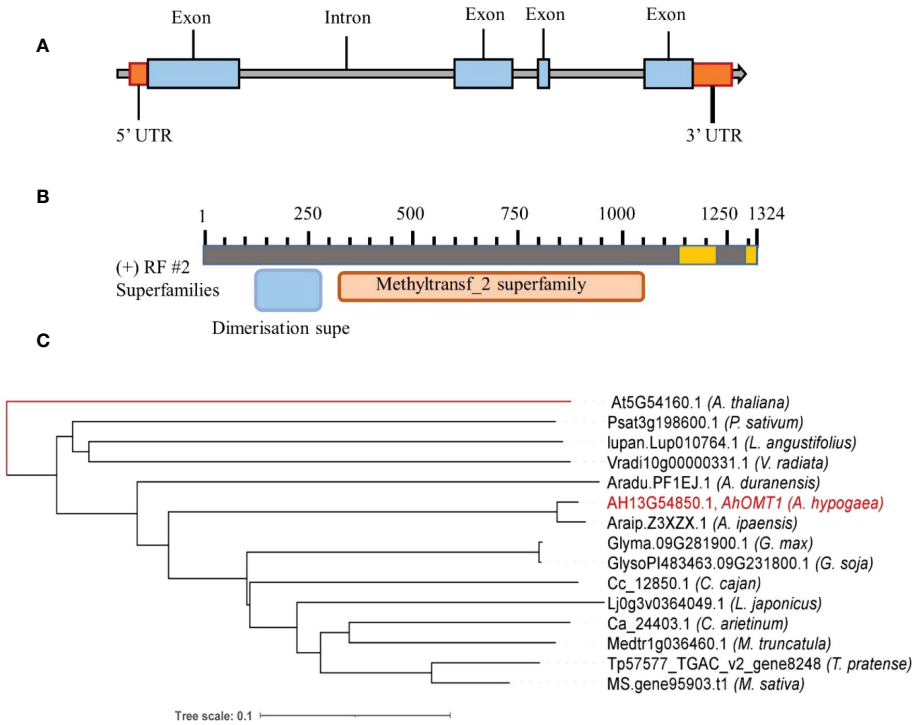


FIGURE 3
Structure and classification of *AhOMT1* gene. **(A)** Structure of *AhOMT1* gene. **(B)** Functional domains structure of *AhOMT1*. **(C)** Phylogenetic relationship of *AhOMT1* gene with *OMT1* genes of other legume species. Unrooted tree was constructed by taking *AtOMT1* as reference.

Isolation of *AhOMT1* promoter sequences

Genomic DNA was extracted from leaf samples of peanut (Figure S4A). 4μg genomic DNA was incompletely digested with the restriction enzyme *AseI, EcoRI*, and *HindIII* for 5min, 15min, and 25min, respectively (Figure S4B). Then we ligated the corresponding Adapter to the digestion products of the three restriction enzymes by T4 DNA Ligase overnight at 20 °C.

With the digested DNA connected with the complementary adaptors as the template, more than 1700bp fragment was isolated by three times nest PCR. The promoter was amplified using genomic DNA (50ng/μl) with primers AhAF7-F and AhAF7-R. After agarose gel electrophoresis, we got two bands that showed in Figure S4C. We linked them to the vector pMD18-T and sequenced them. The sequencing results showed both bands containing a 197bp downstream fragment that overlapped the upstream *AhOMT1* gene with the same sequence, indicating that the cloned promoters were correct starting from *AhOMT1*. Detach the overlapped sequence with the *AhOMT1* gene, the length of the two promoters region of the *AhOMT1* before the ATG were 1,690bp and 1103bp, respectively (Figure S5). We named them *AhOMT1* promoter L and *AhOMT1* promoter S. The two sequences, long and short, were then approved to be the promoters of *AhOMTs* from respective Ah13G54850.1 and Ah10G32250.1 genes (Zhuang et al., 2019). For future studies, we used the AhOMT1 L promoter from high aflatoxin resistance peanut cultivar XHXL.

TABLE 2 Comparison of *AhOMT1* to homologous genes from other legume species.

Species	Genes description	Accession number	Identity %
Glycine max	O-methyltransferase	Glyma.09G281900.1	70.99
Medicago truncatula	Caffeate O-methyltransferase	Medtr1g036460.1	61.58
Lotus japonicus	O-methyltransferase 1	Lj0g3v0364049.1	61.45
Arabidopsis thaliana	O-methyltransferase 1	At5G54160.1	57.76
Arachis ipaensis	O-methyltransferase 1	Araip.Z3XZX.1	97.41
Arachis duranensis	O-methyltransferase 1	Aradu.PF1EJ.1	56.15
Arachis hypogaea	Isoliquiritigenin 2'-O-methyltransferase	AH13G54850.1	100
Arachis hypogaea	Isoliquiritigenin 2'-O-methyltransferase	AH10G32250.1	98.6

The gene accession numbers are as given in LIS, TAIR, and PGR databases.

Analysis of *cis*-regulatory elements of *AhOMT1* promoter

The *cis-acting* elements of the L promoter region of the *AhOMT1* gene were predicted by online databases viz, PLACE (Higo et al., 1998) and PlantCARE (Rombauts et al., 1999). A 1698 bp sequence upstream of the *AhOMT1* coding region contained the core elements, including the TATA box, required for precise initiation of transcription (Grace et al., 2004), the CAAT box needed for tissue-specificity (Shirsat et al., 1989). Other key elements include light-responsive elements; Box 4, GATA-motif, GT1-motif, Gap-box, and I-box. Several others were also present, including defense-related elements, WUN-motif, MYB binding sites, ethylene-responsive element ERE, and wound-responsive elements. Zein metabolism-responsive element (O2-site), flavonoid biosynthesis elements (MBSI), and seed-specific elements (RY-element) were also present. More details of *cis*-regulatory elements in *AhOMT1P* are given in Figure 4. The presence of these elements suggests that *AhOMT1P* can be a suitable candidate for a gene's own promoter. Some novel elements were predicted in promoter region, except these previously identified *cis*-elements (Figure 4). Further details for the *cis*-regulatory elements predicted in *AhOMT1* promoter through PlantCARE database are given in Table S3. The *AhOMT1* promoter sequence are given in Supplementary File 1, while the elements predicted through the new PLACE database are given in Table S4.

Generation of transgenic plants with *AhOMT1P: GUS* fission unit

A 1698bp upstream region of the *AhOMT1* gene was PCR amplified (Figure 5A) using the DNA of a high *A. flavus*-resistant peanut variety XHXL by promoter-specific primers (Table S2). Following the Gateway BP and LR cloning steps, the expression vector *AhOMT1P::GUS* was successfully constructed (Figures 5B, C) and transformed to *A. tumefaciens*. Genetic transformation of *Arabidopsis* plants was carried out by the floral dip method, and hygromycin-resistant positive transgenic plants were confirmed by PCR amplification. Eight plants showing resistance to Hygromycin were selected for PCR confirmation (Figure 6A).

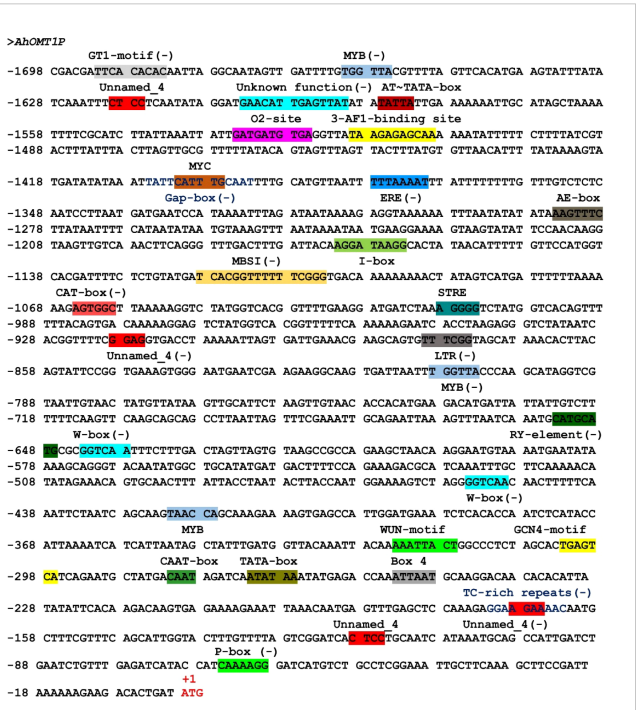


FIGURE 4
Sequence analysis of *AhOMT1* promoter. Presence of *cis*-elements in promoter sequences predicted by the PlantCARE database.

Characterization of *AhOMT1P* expression function

Histochemical GUS staining was checked under control (ddH_2O) and *A. flavus* infection. Transgenic *Arabidopsis* plants were sprayed with *A. flavus* spores. Leaf samples were taken before *A. flavus* inoculation and 1h, 3h, 6h, 12h, 24h, and 36h after spray and incubated in staining solution to check the histochemical expression of the *GUS* gene. Results showed that staining was not present in control plants (0h, before spray). At the same time, a strong blue color was present in treated leaves at all time points, indicating that the *GUS* gene was induced in response to *A. flavus* treatment after 1h, and it continuously showed high expression at 3h, 6h. 12h, 24h, and 36h samples with three replications (Figure 6B). More results for GUS staining of whole leaves of transgenic *Arabidopsis* plants can be seen in Figure S6. While the small sections of stained leaves are shown in Figures S7, 6C. Small seedlings of transgenic plants are shown in Figures S8, 6D. These results showed that the *AhOMT1* promoter was inactive under normal conditions, as staining was not present in non-treated leaves. *AhOMT1P* was induced only after *A. flavus* infection. These results supported that the *AhOMT1* gene is an *A. flavus* inducible gene and a suitable candidate for the native promoter of a gene. Based on these results, it can be predicted that the *AhOMT1* promoter can activate anti-aflatoxin genes upon *A. flavus* infection. Using this promoter to drive aflatoxin resistance genes in transgenic peanut will provide an excellent way to improve the aflatoxin resistance of transgenic peanuts.

Quantitative expression of *AhOMT1P* controlled *GUS* gene in response to *A. flavus* infection and plant hormones

Because the GUS staining just reflects continuous gene expression in *A. flavus* treated leaves after 1h, to quantify gene expression intensity at different time points for comparison more detailedly after *A. flavus* infection, the *GUS* gene expression was determined by qRT-PCR. The expression of the *GUS* gene was enhanced in response to *A. flavus* infection after 6 hours, and this increased expression was still maintained after 48 hours of inoculation (Figure 7). Although the *GUS* gene showed a reduced expression after six hours, the expression level was still too high compared to non-treated plants. Maximum expression was observed at 36 hours post-inoculation. The expression level of *AhOMT1* promoters under different phytohormones stress was determined to check that either this promoter is only induced by *A. flavus* or it can be induced by different hormones. So, the real-time expression of the *GUS* gene in transgenes treated with hormones was

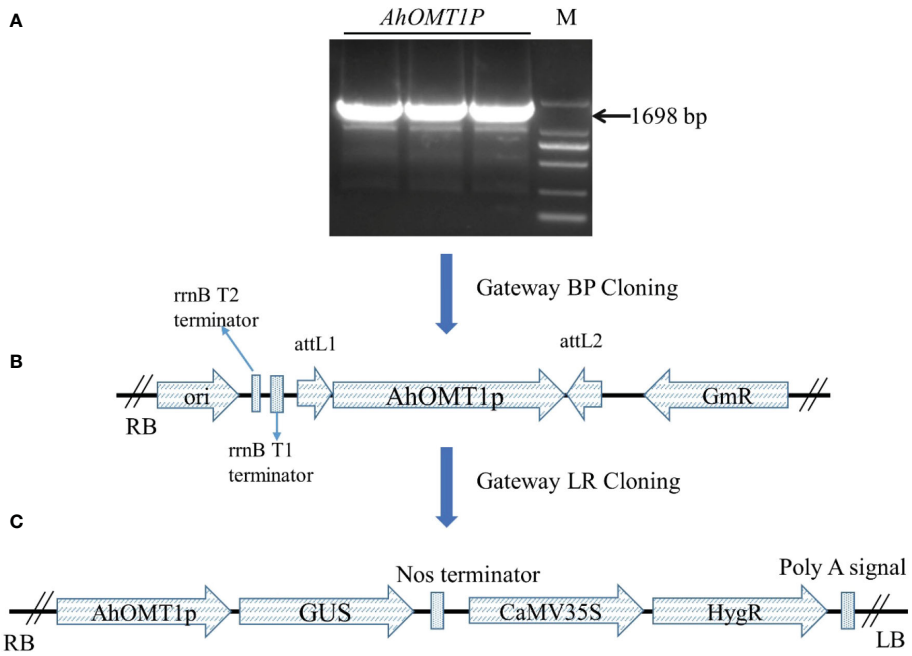


FIGURE 5
Construction of vectors using the backbone of *pMDC164* vector by Gateway cloning. **(A)** Amplification of *AhOMT1* promoter, M=2kb marker. **(B)** Construction of Gateway entry clone using *pDONR207* vector. **(C)** Construction of Gateway expression vector using the backbone of binary vector *pMDC164*.

determined by qRT-PCR. For that purpose, transgenic *Arabidopsis* plants were treated with abscisic acid, Brassinolide, ethephon, paclobutrazol, and salicylic acid solutions. Under all hormonal treatments and distilled water, the *GUS* gene did not show any remarkable increase in expression. The expression level of the *GUS* gene was a little reduced in some cases and a little increased in others. Overall, the expression pattern was comparable to the transcriptome expression of the *AhOMT1* gene in peanut compared to control or non-treated plants (Figure 8). The *GUS* gene's expression profiling in response to different hormonal treatments provided that the *AhOMT1* promoter did not induce the expression of the *GUS* gene under hormone treatments. Deep GUS staining and increased quantitative expression of the *GUS* gene upon *A. flavus* infection is indicative of the inducible behavior of the *AhOMT1* promoter. Our results are supported with our transcriptome analysis in another peanut cultivar (Zhuang et al., 2019) (Table S1). The same results were obtained that *AhOMT1P* was non-responsive to any hormone treatment, and very weak expression was found in all tissues or organs in peanut. Thereby, it is evident that plant hormones do not induce the *AhOMT1* promoter. These results collectively suggest that *AhOMT1P* is a suitable candidate for *A. flavus* resistance transgenic breeding.

Discussion

As an important geocarpic crop supplying humans with oil, food, and feed worldwide, peanut is vulnerable to many pathogens and diseases (Severns et al., 2003; Pandey et al., 2019; Soni et al., 2020). Aflatoxin is known as the strongest carcinogen caused by species of the genus *Aspergillus* (Bordin et al., 2014). Several researchers have suggested different strategies to cope with the attack by *Aspergillus* species and avoid aflatoxin contamination (Cobos et al., 2018; Sharma et al., 2018; Khan and Zhuang, 2019). Genetic engineering with *Aspergillus* species-resistant genes is one of the most promising ways to tackle the issue (Nigam et al., 2009; Sharma et al., 2018; Ncube and Maphosa, 2020), especially when no germplasm has found stable resistance to *A. flavus*, which makes breeding resistance variety challenging and very slowly. However, most transgenes are expressed under the control of constitutive promoters that often make plants experience an increased metabolic burden. Constitutive expression of transgenes wastes a lot of metabolic energy that may result in undesirable phenotypes and decreased output (Yuan et al., 2019). Under normal and stressful situations, plants need to transfer important nutrients to the target locations in order to survive and develop smoothly (Divya et al., 2019). As a result, inducible or specific promoters outperform the constitutive promoters to improve the performance of transgene by changing the genetic architecture in an inducible or time- and space-defined manner as plants need. In this study we first found out and characterized the *AhOMT1P* as an *A. flavus* inducible promoter. It shows trace expression under normal conditions but highly upregulated the *OMT1* gene under *A. flavus* infection. Thereby, this provides a new resource for managing aflatoxin contamination in the future peanut enhancement.

Promoters are key regulatory elements that switch on and off the gene functioning. Promoters of *A. flavus* inducible genes can hypothetically drive the expression of a gene upon *A. flavus* infection. Without functional evidence that an inducible promoter can drive a resistance-related gene, we cannot proceed further with the transgenic program. For screening a target promoter, we identified the *AhOMT1* through a genome-wide transcriptomic microarray analysis under drought and a combination of drought

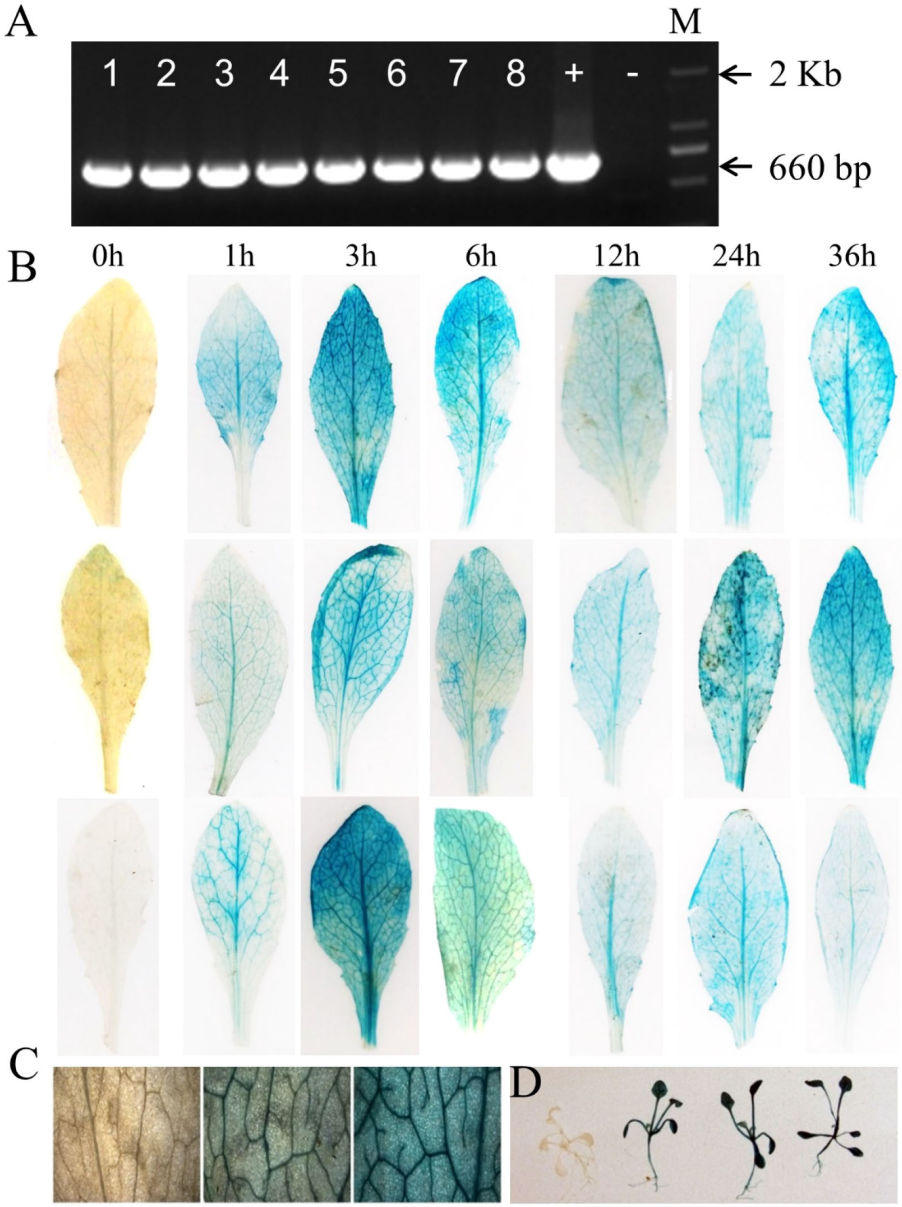


FIGURE 6
Evaluation of the function of *AhOMT1* promoter. **(A)** Confirmation of transgenic *Arabidopsis* plants with *AhOMT1P* (660bp fragment) PCR amplification with promoter-specific forward and *GUS* gene specific reverse primers pair. *Arabidopsis* Col-0 was used as negative control, and Gateway LR constructs used as positive control. M shows 2kb marker. **(B)** Strong GUS staining of *AhOMT1P* transgenic leaves is only induced by *A. flavus* infection with spores after 1 h of infection. *GUS gene* was not active in non-treated leaves. GUS staining was present in leaf samples at all time points after inoculation. **(C)** Magnified leaf surface staining indicated that *AhOMT1P* activated the *GUS* gene expression after infection with *A. flavus* spores. **(D)** Transgenic seedlings showing deep GUS staining only after *A. flavus* challenging.

and *A. flavus* infection (Chen et al., 2016; Zhang et al., 2017). The microarray expression analysis showed that the *AhOMT1* belonging to peanut methyltransferase family named "Isoliquiritigenin 2'-O-methyltransferase" was highly expressed under drought and *A. flavus* inoculation. It showed a more than 200-fold increase in expression compared to drought control(Table S1). The *AhOMT1* gene was cloned from the roots of peanut cultivar Minhua-6 by RACE PCR using a set of specifically designed primers (Table S2). The promoter region was also cloned from the roots of the same cultivar systematically. From the available datasets, we identified that it had a trace expression in different tissues or organs, low expression induced by wide environmental stresses like hormones, low temperature, drought, Ca^{2+}, and biotic stress like *R. solanacerum* in comparison with the house-keep gene the Actin-7 (Table S1-1~3), but a gigantic and specific increase in expression was induced by *A. flavus* infection. We checked the expression of the *AhOMT1* gene in a widely cultivated variety M-6 by qRT-PCR in response to *A. flavus* infection. All these showed the *AhOMT1P* is proper for the anti-aflatoxin purpose. Thus far, a lot of work has been done on the functional characterization of abiotic stress-inducible promoters in crop plants (Pino et al., 2007; Rai et al., 2009; Divya et al., 2019), but there is a lack of work on the functional studies of biotic stress-inducible promoters. Especially it is very hard to find any study reporting *A. flavus* inducible promoters in crop plants.

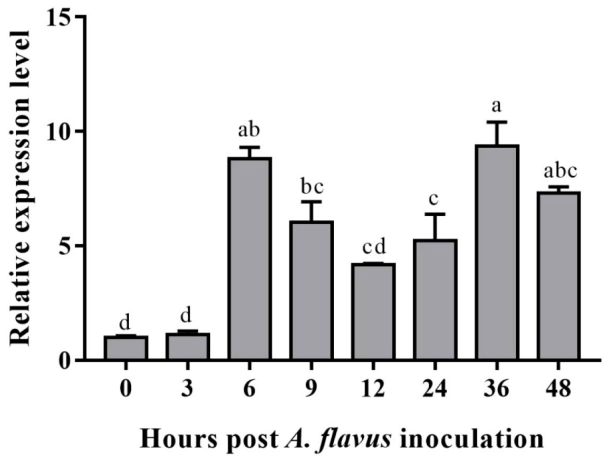


FIGURE 7
Quantitative expression of *GUS* gene in response to *A. flavus* infection. The *GUS* gene recorded an increased expression at 6h post inoculation and maintained higher expression even after 48 hours. The qRT-PCR data were analyzed by ΔΔCT method, statistical significance was assessed by LSD model with α=0.05.

As the available datasets and expression validation supported that the *AhOMT1* gene is an *A. flavus* inducible gene, its promoter might be a suitable alternative to drive the expression of any foreign gene (related to *A. flavus* resistance) in transgenic plants. To validate its functions, we cloned the promoter of the *AhOMT1* gene and used it to drive the GUS gene under *A. flavus* infection in transgenic plants. The online promoter analysis tools predicted several important *cis*-regulatory elements, including TATA Box, CAAT Box, GC Box (Muthusamy et al., 2017), and many others related to hormones, light, growth, regulation, and stress responsiveness. Different kinds of elements and their number critically determine the strength of a promoter (Stålberg et al., 1993; Tang et al., 2015). Seed-specific elements, including RY-repeats, were also present (Fujiwara and Beachy, 1994). Although these predicted elements are very important for promoter functioning, not a single element related to biotic stress was identified. Perhaps it is due to the reason that not extensive work is available for the binding sites of biotic stress-responsive elements. The promoter analysis also revealed some novel elements (CTCC). It is possible this is a binding sequence for some element involved in Aflatoxin-inducible activity.

Although peanut is much more suitable for functional studies of genes and promoters, but challenging and time-consuming work of genetic transformation makes it less favourite as a model species. In contrast, the short life cycle, easy handling, and well-established genetic transformation methods make *Arabidopsis* a good alternative for functional investigations (Sunkara et al., 2014). So, we choose *Arabidopsis* for the genetic transformation and functional characterization of the *AhOMT1* gene. The results indicated that *AhOMT1* is highly inducible and strongly drives the GUS gene in transgenic plants upon *A. flavus* infection. Although minor deviations in expression patterns were found in response to hormone treatments in transgenic plants, the reason for this deviation can be the different species and another because it was the *GUS* gene under the *AhOMT1* promoter, not the *AhOMT1* gene itself. We are highly convinced that the *AhOMT1P* promoter analyzed here could potentially drive aflatoxin-resistant genes in an inducible manner upon *A. flavus* infection to cope with this serious issue.

Conclusion

We identified a novel A. flavus inducible gene in this study and worked out its promoter. This gene belonged to peanut O-methyltransferase gene family (Isoliquiritigenin 2'-O-methyltransferase), transferring the methyl

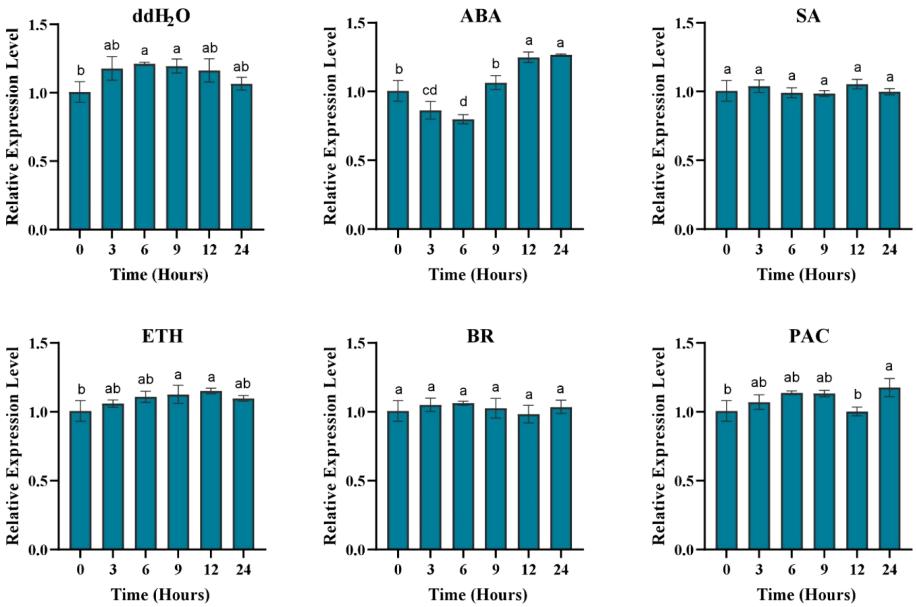


FIGURE 8
Quantitative expression of *GUS* gene in response to different hormone treatments. The *GUS* gene showed reduced expression in response to different hormones and ddH_2O spray at all time points. ABA, abscisic acid; BR, Brassinolide; ETH, ethephon; PAC, paclobutrazol; and SA, salicylic acid. Data were analyzed by analysis of variance (ANOVA, α=0.05).

group of S-adenosyl-L-methionine (SAM) to the hydroxyl group of numerous different organic chemical compounds, ultimately leading to the synthesis of the methyl ether variants of these substances. This gene was identified by the genome-wide microarray expression under drought and *A. flavus* inoculation. The *AhOMT1* promoter showed trace or no expression in any organs/tissues, nor was it inducible by abiotic or bacterial challenges. Upon infection of *A. flavus* spores, this gene was highly induced in peanut leaves and pericarp. A 1698 bp upstream region of the *AhOMT1* gene was fused with the GUS reporter gene and transformed into *Arabidopsis* plants. Histochemical GUS staining of transgenic *Arabidopsis* plants showed that the *AhOMT1* promoter activated the *GUS* gene only after *A. flavus* infection. While it did not drive the *GUS* expression in normal conditions, nor was it induced by hormone treatments. The real-time expression of the *GUS* gene in transgenic plants confirmed increased expression in response to *A. flavus* infection. These results provided the practical significance of *AhOMT1P* to drive a gene in response to *A. flavus* infection.

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

Idea was conceived by ZW. ZW, YZ designed the study. YZ, YS investigated the project and wrote the manuscript. XZ, SZC, HC, CZ, YD, MR, and SLC participated in research work, wrote or revised the manuscript. YZ and YS equally contributed to the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation (NSFC) of China (U1705233, 32072103, 32272155, 31701463 and 31601337), the Science and Technology Foundation of Fujian Province of China (2021N5007 and 2017N0006) and the Special Fund for Scientific and Technological Innovation of Fujian Agriculture and Forestry University (KFb22011XA and KFb22010XA).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1102181/full#supplementary-material

References

Amaike, S., and Keller, N. P. (2011). Aspergillus flavus. *Annu. Rev. Phytopathol.* 49, 107–133. doi: 10.1146/annurev-phyto-072910-095221

Awasthi, V., Bahman, S., Thakur, L. K., Singh, S. K., Dua, A., and Ganguly, S. (2012). Contaminants in milk and impact of heating: An assessment study. *Indian J. Public Health* 56, 95. doi: 10.4103/0019-557X.96985

Ballance, D. (1986). Sequences important for gene expression in filamentous fungi. *Yeast* 2, 229–236. doi: 10.1002/yea.320020404

Bordin, K., Sawada, M. M., Rodrigues, C. E., d., C., da Fonseca, C. R., and Oliveira, C. A. F. (2014). Incidence of aflatoxins in oil seeds and possible transfer to oil: A review. *Food Eng. Rev.* 6, 20–28. doi: 10.1007/s12393-014-9076-9

Chen, H., Yang, Q., Chen, K., Zhao, S., Zhang, C., Pan, R., et al. (2019). Integrated microRNA and transcriptome profiling reveals a miRNA-mediated regulatory network of embryo abortion under calcium deficiency in peanut (Arachis hypogaea l.). *BMC Genomics* 20, 1–17. doi: 10.1186/s12864-019-5770-6

Chen, H., Zhang, C., Cai, T. C., Deng, Y., Zhou, S., Zheng, Y., et al. (2016). Identification of low Ca2+ stress-induced embryo apoptosis response genes in arachis hypogaea by SSH-associated library lift (SSHaLL). *Plant Biotechnol. J.* 14, 682–698. doi: 10.1111/pbi.12415

Chi, X., Hu, R., Yang, Q., Zhang, X., Pan, L., Chen, N., et al. (2012). Validation of reference genes for gene expression studies in peanut by quantitative real-time RT-PCR. *Mol. Genet. Genomics* 287, 167–176. doi: 10.1007/s00438-011-0665-5

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

Cobos, C. J., Tengey, T. K., Balasubramanian, V. K., Williams, L. D., Sudini, H. K., Varshney, R. K., et al. (2018). Employing peanut seed coat cell wall mediated resistance against aspergillus flavus infection and aflatoxin contamination. doi: 10.20944/preprints201808.0292.v1

Cornejo, M.-J., Luth, D., Blankenship, K. M., Anderson, O. D., and Blechl, A. E. (1993). Activity of a maize ubiquitin promoter in transgenic rice. *Plant Mol Biol.* 23, 567–581. doi: 10.1007/BF00019304

Divya, K., Kishor, P. K., 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

Fleming, D., Musser, F., Reisig, D., Greene, J., Taylor, S., Parajulee, M., et al. (2018). Effects of transgenic bacillus thuringiensis cotton on insecticide use, heliothine counts, plant damage, and cotton yield: A meta-analysis 1996-2015. *PLoS ONE* 13. doi: 10.1371/journal.pone.0200131

Fujiwara, T., and Beachy, R. N. (1994). Tissue-specific and temporal regulation of a β-conglycinin gene: Roles of the RY repeat and other cis-acting elements. *Plant Mol Biol.* 24, 261–272. doi: 10.1007/BF00020166

Gang, D. R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., et al. (2002). Characterization of phenylpropene O-methyltransferases from sweet basil: Facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. *Plant Cell* 14, 505–519. doi: 10.1105/tpc.010327

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, 3784–3788. doi: 10.1093/nar/gkg563

Gonzales, M. D., Archuleta, E., Farmer, A., Gajendran, K., Grant, D., Shoemaker, R., et al. (2005). The legume information system (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Res.* 33, D660–D665. doi: 10.1093/nar/gki128

Goto, T., Wicklow, D. T., Ito, Y. J. A., and Microbiology, E. (1996). Aflatoxin and cyclopiazonic acid production by a sclerotium-producing aspergillus tamarii strain. *Appl Environ Microbiol.* 62, 4036–4038. doi: 10.1128/aem.62.11.4036-4038.1996

Grace, M. L., Chandrasekharan, M. B., Hall, T. C., and Crowe, A. J. (2004). Sequence and spacing of TATA box elements are critical for accurate initiation from the β-phaseolin promoter. *J. Biol. Chem.* 279, 8102–8110. doi: 10.1074/jbc.M309376200

Gutierrez, A. (2018). Hybrid bt cotton. *Curr. Sci.* 115, 2206–2210. doi: 10.18520/cs/v115/i12/2206-2210

Higo, K., Ugawa, Y., Iwamoto, M., and Higo, H. (1998). PLACE: A database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res.* 26, 358–359. doi: 10.1093/nar/26.1.358

Horn, B. W., Moore, G. G., and Carbone, I. J. M. (2009). Sexual reproduction in aspergillus flavus. *Mycologia* 101, 423–429. doi: 10.3852/09-011

Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO Journal* 6, 3901–3907. doi: 10.1002/j.1460-2075.1987.tb02730.x

Kay, R., Chan, A., Daly, M., and McPherson, J. J. S. (1987). Duplication of CaMV 35 s promoter sequences creates a strong enhancer for plant genes. *Science* 236, 1299–1302. doi: 10.1126/science.236.4806.1299

Kensler, T. W., Roebuck, B. D., Wogan, G. N., and Groopman, J. D. (2011). Aflatoxin: A 50-year odyssey of mechanistic and translational toxicology. *Toxicol Sci.* 120, S28–S48. doi: 10.1093/toxsci/kfq283

Khan, S. A., Chen, H., Deng, Y., Chen, Y., Zhang, C., Cai, T., et al. (2020). High-density SNP map facilitates fine mapping of QTLs and candidate genes discovery for aspergillus flavus resistance in peanut (Arachis hypogaea). *Theor. Appl. Genet.* 133, 2239–2257. doi: 10.1007/s00122-020-03594-0

Khan, S. A., and Zhuang, W. (2019). Obliging tactics to mitigate the intricate problem of aflatoxin contamination in peanut: A review. 4, 986–1005. doi: 10.25177/JFST.4.9.RA.597

Kozbial, P. Z., and Mushegian, A. R. (2005). Natural history of s-adenosylmethionine-binding proteins. *BMC Struct Biol.* 5, 1–26. doi: 10.1186/1472-6807-5-19

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, D1202–D1210. doi: 10.1093/nar/gkr1090

Li, A., Yue, S., and Ma, H. (2006). Correlativity of three counting methods of fungal spore. *J Microbiol.* 26, 1071110. doi: 10.1360/yc-006-1627

McElroy, D., Zhang, W., Cao, J., and Wu, R. (1990). Isolation of an efficient actin promoter for use in rice transformation. *Mol Gen Genet. MGG* 2, 163–171. doi: 10.1105/tpc.2.2.163

Muthusamy, S., Sivalingam, P., Sridhar, J., Singh, D., Haldhar, S., and Kaushal, P. (2017). Biotic stress inducible promoters in crop plants-a review. *J. Agric. Ecol.* 4, 14–24. doi: 10.53911/JAE.2017.4202

Ncube, J., and Maphosa, M. (2020). Current state of knowledge on groundnut aflatoxins and their management from a plant breeding perspective: Lessons for Africa. *Sci Afr.* 7. doi: 10.1016/j.sciaf.2020.e00264

Nguyen, T.-N., Son, S., Jordan, M. C., Levin, D. B., and Ayele, B. (2016). Lignin biosynthesis in wheat (Triticum aestivum l.): Its response to waterlogging and association with hormonal levels. *BMC Plant Biol.* 16, 1–16. doi: 10.1186/s12870-016-0717-4

Nigam, S., Waliyar, F., Aruna, R., Reddy, S., Kumar, P. L., Craufurd, P. Q., et al. (2009). Breeding peanut for resistance to aflatoxin contamination at ICRISAT. *Peanut Science* 36, 42–49. doi: 10.3146/AT07-008.1

Pandey, M. K., Gangurde, S. S., Sharma, V., Pattanashetti, S. K., Naidu, G. K., Faye, I., et al. (2020). Improved genetic map identified major QTLs for drought tolerance-and iron deficiency tolerance-related traits in groundnut. *Genes* 12, 37. doi: 10.3390/genes12010037

Pandey, M. K., Kumar, R., Pandey, A. K., Soni, P., Gangurde, S. S., Sudini, H. K., et al. (2019). Mitigating aflatoxin contamination in groundnut through a combination of genetic resistance and post-harvest management practices. *Toxins* 11, 315. doi: 10.3390/toxins11060315

Payne, G., Nierman, W., Wortman, J., Pritchard, B., Brown, D., Dean, R., et al. (2006). Whole genome comparison of aspergillus flavus and a. oryzae. *Med Mycol.* 44, S9–S11. doi: 10.1080/13693780600835716

Pino, M. T., Skinner, J. S., Park, E. J., Jeknić, Z., Hayes, P. M., Thomashow, M. F., et al. (2007). Use of a stress inducible promoter to drive ectopic AtCBF expression improves potato freezing tolerance while minimizing negative effects on tuber yield. *Plant Biotechnol J.* 5, 591–604. doi: 10.1111/j.1467-7652.2007.00269.x

Qian, Y., Xi, Y., Cheng, B., and Zhu, S. (2014). Genome-wide identification and expression profiling of DNA methyltransferase gene family in maize. *Plant Cell Rep* 33, 1661–1672. doi: 10.1007/s00299-014-1645-0

Rai, M., He, C., and Wu, R. (2009). Comparative functional analysis of three abiotic stress-inducible promoters in transgenic rice. *Transgenic Res.* 18, 787–799. doi: 10.1007/s11248-009-9263-2

Richard, J. L. (2008). Discovery of aflatoxins and significant historical features. *Toxin Rev.* 27, 171–201. doi: 10.1080/15569540802462040

Roje, S. (2006). S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* 67, 1686–1698. doi: 10.1016/j.phytochem.2006.04.019

Rombauts, S., Déhais, P., Van Montagu, M., and Rouzé, P. (1999). PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.* 27, 295–296. doi: 10.1093/nar/27.1.295

Severns, D. E., Clements, M. J., Lambert, R. J., and White, D. G. (2003). Comparison of aspergillus ear rot and aflatoxin contamination in grain of high-oil and normal-oil corn hybrids. *J Food Prot.* 66, 637–643. doi: 10.4315/0362-028X-66.4.637

Sharma, K. K., Pothana, A., Prasad, K., Shah, D., Kaur, J., Bhatnagar, D., et al. (2018). Peanuts that keep aflatoxin at bay: a threshold that matters. *Plant Biotechnol J* 16, 1024–1033. doi: 10.1111/pbi.12846

Shirsat, A., Wilford, N., Croy, R., and Boulter, D. (1989). Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol. Gen. Genet. MGG* 215, 326–331. doi: 10.1007/BF00339737

Soni, P., Gangurde, S. S., Ortega-Beltran, A., Kumar, R., Parmar, S., Sudini, H. K., et al. (2020). Functional biology and molecular mechanisms of host-pathogen interactions for aflatoxin contamination in groundnut (Arachis hypogaea l.) and maize (Zea mays l.). *Front. Microbiol.* 11, 227. doi: 10.3389/fmicb.2020.00227

Stålberg, K., Ellerström, M., Josefsson, L.-G., and Rask, L. (1993). Deletion analysis of a 2S seed storage protein promoter of brassica napus in transgenic tobacco. *Plant Mol. Biol.* 23, 671–683. doi: 10.1007/BF00021523

Struck, A. W., Thompson, M. L., Wong, L. S., and Micklefield, J. (2012). S-adenosyl-methionine-dependent methyltransferases: Highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* 13, 2642–2655. doi: 10.1002/cbic.201200556

Sunkara, S., Bhatnagar-Mathur, P., and Sharma, K. K. (2014). Isolation and functional characterization of a novel seed-specific promoter region from peanut. *Appl Biochem Biotechnol.* 172, 325–339. doi: 10.1007/s12010-013-0482-x

Tang, G., Qin, J., Dolnikowski, G. G., Russell, R. M., and Grusak, M. A. (2009). Golden rice is an effective source of vitamin a. *Am J Clin Nutr* 89, 1776–1783. doi: 10.3945/ajcn.2008.27119

Tang, G., Xu, P., Liu, W., Liu, Z., and Shan, L. (2015). Cloning and characterization of 5′ flanking regulatory sequences of AhLEC1B gene from arachis hypogaea l. *PLoS One* 10, e0139213. doi: 10.3945/ajcn.2008.27119

Wang, M., Zhu, X., Wang, K., Lu, C., Luo, M., Shan, T., et al. (2018). A wheat caffeic acid 3-o-methyltransferase TaCOMT-3D positively contributes to both resistance to sharp eyespot disease and stem mechanical strength. *Agronomy Journal* 8, 1–14. doi: 10.1038/s41598-018-24884-0

Yang, W.-J., Du, Y.-T., Zhou, Y.-B., Chen, J., Xu, Z.-S., Ma, Y.-Z., et al. (2019). Overexpression of TaCOMT improves melatonin production and enhances drought tolerance in transgenic arabidopsis. *BMC Genomics* 20, 652. doi: 10.3390/ijms20030652

Yang, G., Pan, W., Zhang, R., Pan, Y., Guo, Q., Song, W., et al. (2021). Genome-wide identification and characterization of caffeoyl-coenzyme a O-methyltransferase genes related to the fusarium head blight response in wheat. *BMC Genomics* 22, 1–16. doi: 10.1186/s12864-021-07849-y

Yuan, C., Sun, Q., and Kong, Y. (2019). Genome-wide mining seed-specific candidate genes from peanut for promoter cloning. *PLoS One* 14, e0214025. doi: 10.1371/journal.pone.0214025

Zhang, C., Chen, H., Cai, T., Deng, Y., Zhuang, R., Zhang, N., et al. (2017). Overexpression of a novel peanut NBS-LRR gene a h RRS 5 enhances disease resistance to r alstonia solanacearum in tobacco. *Plant Biotechnol. J.* 15, 39–55. doi: 10.1111/pbi.12589

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876. doi: 10.1038/s41588-019-0402-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 February 2023
DOI 10.3389/fpls.2023.1102174

OPEN ACCESS

EDITED BY
Weijian Zhuang,
Fujian Agriculture and Forestry University, China

REVIEWED BY
Xin Wei,
Shanghai Normal University, China
Qing Zhang,
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences (CAAS), China
Yunpeng Cao,
Wuhan Botanical Garden, Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Jiahai Fang
fangjiahai2019@126.com
Liyun Wan
wanliyun2019@163.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 18 November 2022
ACCEPTED 06 January 2023
PUBLISHED 14 February 2023

CITATION
Wang S, Xu Z, Yang Y, Ren W, Fang J and Wan L (2023) Genome-wide analysis of R2R3-MYB genes in cultivated peanut (*Arachis hypogaea* L.): Gene duplications, functional conservation, and diversification. *Front. Plant Sci.* 14:1102174. doi: 10.3389/fpls.2023.1102174

Genome-wide analysis of R2R3-MYB genes in cultivated peanut (*Arachis hypogaea* L.): Gene duplications, functional conservation, and diversification

Sijian Wang, Zhe Xu, Yiwen Yang, Weifang Ren, Jiahai Fang* and Liyun Wan*

Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang, China

The cultivated Peanut (*Arachis hypogaea* L.), an important oilseed and edible legume, are widely grown worldwide. The R2R3-MYB transcription factor, one of the largest gene families in plants, is involved in various plant developmental processes and responds to multiple stresses. In this study we identified 196 typical *R2R3-MYB* genes in the genome of cultivated peanut. Comparative phylogenetic analysis with *Arabidopsis* divided them into 48 subgroups. The motif composition and gene structure independently supported the subgroup delineation. Collinearity analysis indicated polyploidization, tandem, and segmental duplication were the main driver of the *R2R3-MYB* gene amplification in peanut. Homologous gene pairs between the two subgroups showed tissue specific biased expression. In addition, a total of 90 R2R3-MYB genes showed significant differential expression levels in response to waterlogging stress. Furthermore, we identified an SNP located in the third exon region of *AdMYB03-18 (AhMYB033)* by association analysis, and the three haplotypes of the SNP were significantly correlated with total branch number (TBN), pod length (PL) and root-shoot ratio (RS ratio), respectively, revealing the potential function of *AdMYB03-18 (AhMYB033)* in improving peanut yield. Together, these studies provide evidence for functional diversity in the *R2R3-MYB* genes and will contribute to understanding the function of *R2R3-MYB* genes in peanut.

KEYWORDS

cultivated peanut, R2R3-MYB transcription factors, allotetraploid, gene duplications, gene expression, functional diversity

Introduction

MYB transcription factor is one of the most numerous families of transcription factors in plants. They are distinguished by a conserved sequence of four incomplete amino acid repeats (R), with about 52 amino acids serving as the DNA-binding amino acids. Three a-helices are formed by each repeat, and the second and third helices of each repeat create a three-

dimensional (3D) HTH structure with three evenly spaced tryptophan (or hydrophobic) residues (Ogata et al., 1996). Each repeat's third helix serves as a "recognition helix," coming into touch with the DNA and becoming enmeshed in the main groove (Jia et al., 2004). According to the number of adjacent duplicates, MYB proteins can be classified into several types, with R2R3-MYB proteins accounting for the majority of them in plants (Dubos et al., 2010; Liu et al., 2015).

In many aspects of plant life, including primary and secondary metabolism, cell fate, developmental processes, and response to biotic and abiotic stimuli, R2R3-MYB transcription factors are crucial players (Martin and Paz-Ares, 1997; Jin and Martin, 1999; Dubos et al., 2010; Liu et al., 2015; Cao et al., 2020). The maize *C1* gene, related to the mammalian transcription factor C-MYB and is involved in the control of anthocyanin production, is the first *MYB* gene discovered in plants (Paz-Ares et al., 1987). ErMYB1 and ErMYB2 are regarded as inhibitors and activators, respectively, of the development of secondary cell walls in the eucalyptus (Goicoechea et al. 2005; Legay et al. 2007; Legay et al. 2010). ABA-mediated responses to environmental cues are mediated by the AtMYB13, AtMYB15, AtMYB33, and AtMYB101 (Reyes and Chua, 2007). AdMYB3 has been reported to be involved in anthocyanin biosynthesis and flower development in apples (Vimolmangkang et al., 2013). Members of the transcription factors that resemble MYBMIXTA are involved in starting the formation of cotton seed fiber (Bedon et al. 2014). *AhTc1*, encoding an R2R3-MYB transcription factor, play important role in regulating anthocyanin biosynthesis in peanut (Zhao et al., 2019). In rice, OsMYB30, an R2R3-MYB transcription factor, regulates the phenylalanine ammonia-lyase pathway to give brown planthopper resistance (He et al. 2020). GmMYB14 controls plant structure *via* the brassinosteroid pathway, contributing to high-density yield and drought resistance in soybean (Chen et al. 2021). MsMYB741 is involved in alfalfa resistance to aluminum stress by regulating flavonoid biosynthesis (Su et al. 2022). OsMYB60 positively regulates cuticular wax biosynthesis and this helps rice (*Oryza sativa*) plants tolerate drought stress (Jian et al., 2022).

Cultivated peanut (*Arachis hypogaea* L.), one of the most widely consumed legumes worldwide, has been used to meet the nutritional needs of developing countries globally (Toomer, 2018). It originated in South America from a heterozygous cross between two diploid ancestors and was domesticated and widely grown in the tropics and subtropics (Bertioli et al., 2016). Most of the Arachis genus is diploid (Sharma and Bhatnagar-Mathur, 2006). However, only tetraploid peanuts have been domesticated and widely grown to meet human nutritional requirements (Bertioli et al., 2016). Polyploid plants often exhibit greater environmental adaptability (Shimizu-Inatsugi et al., 2017). Recently, the contribution of polyploidization to important agronomic traits, including seed quality, fruit shape, and flowering time, has been reported for several crop species lineages such as soybean (*Glycine max*), wheat (*Triticum aestivum*), and cotton (*Gossypium hirsutum*) (Zhang et al., 2015). The large and close subgenomes of cultivated peanut genomes make genome assembly difficult (Bertioli et al., 2016). Due to advanced high-throughput sequencing technology and high-quality assembly and annotation, the sequencing of cultivated peanut was completed in 2019, the cultivated peanut (Arachis hypogaea L.) is of hybrid origin and has a polyploid genome that contains essentially complete sets of chromosomes from two ancestral species. (Bertioli et al., 2019; Chen et al., 2019; Zhuang et al., 2019). Based on high-quality whole-genome sequencing and assembly engineering, genome-wide characterization of the R2R3-MYB gene has been accomplished in various plants, such as *Arabidopsis thaliana*, rice, maize, soybean, eucalyptus, tomato, Chinese bayberry (Stracke et al., 2001; Jia et al., 2004; Chen et al., 2006; Du et al., 2012a; Soler et al., 2015; Li et al., 2016; Cao et al., 2021).

Most of the previous studies focused on the regulation mechanism of light-inducible anthocyanin (Dubos et al., 2010; Liu et al., 2015; Cao et al., 2020), but peanut is a very special and important crop. Given the fact that this crop possesses the unique characteristics of "aerial flowers and subterranean fruit," the genes responsible for flavonoid synthesis (in peanut testa) and stress response are likely distinct from those in model plants such as *Arabidopsis* and rice. In this study, we characterized 196 R2R3-MYB transcription factors genome-wide and analyzed their phylogenetic relationships, motif composition, gene structure, chromosome distribution, gene duplication, tissue and stress response expression pattern and. Furthermore, association analysis identified a candidate gene highly correlated with total branch number (TBN), pod length (PL) and root-shoot ratio (RS ratio).Our study will contribute to the understanding of the function of the *R2R3-MYB* genes in cultivated peanut and provide candidate genes for development and stress response.

Materials and methods

Identification and conserved DNA-binding domain analysis of R2R3-MYBs in peanut

To identify R2R3-MYBs in peanut, the *A. hypogaea* cv. Tifrunner protein sequences were retrieved from the PeanutBase (https://peanutbase.org). The MYB DNA-binding domain (PF00249) was exploited for the identification of *R2R3-MYB* genes in the peanut genome by using the HMMER 3.3.2 program at a standard E value $<1\times10^{-5}$ (http://pfam.xfam.org/search#tabview=tab1). In total, 204 predicted gene models were found with two consecutive repeats of the MYB domain. All but five (199) were also retrieved when performing a BLASTP analysis using the previously identified genes of *A. hypogaea* against the whole *A. thaliana R2R3-MYB* gene data set (Dubos et al., 2010) with a cut-off e-value of e^{-40}. Subsequently, protein sequences were evaluated for the presence of the MYB domain against the repository of the NCBI CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and SMART databases (http://smart.embl-heidelberg.de/). Finally, 196 *R2R3-MYB* genes were obtained after eliminating incomplete and uncertain sequences. The predicted molecular weights and the theoretical isoelectric point (pI) were obtained by the Expasy proteomics server (https://web.expasy.org/compute_pi/). The prediction of transmembrane helices in AhR2R3-MYB proteins was analyzed by TMHMM - 2.0 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0).

The 196 R2R3-MYB protein sequences in peanuts were performed by multiple sequence alignment using ClustalX (Larkin et al., 2007). The alignment of 196 peanut R2R3-MYB domains was performed using ClustalX and DNAMAN (Lynnon Biosoft). The

amino acid residue distributions of the conserved MYB domains of AhR2R3-MYBs were created using the WebLogo program with default parameters (http://weblogo.berkeley.edu/logo.cgi) (Crooks et al., 2004).

Construction of phylogenetic tree

The protein sequences of the 126 *A.thaliana* R2R3-MYBs were downloaded from the TAIR (http://www.arabidopsis.org/). The protein sequences of R2R3-MYB proteins from *A. hypogaea* and *A.thaliana* (protein sequence information is listed in Supplementary Table 1) were aligned by the MAFFT with the FFT-NS-i algorithm (Katoh et al., 2019), and the multiple sequence alignments were used for phylogenetic analysis. The phylogenetic tree was constructed by the neighbor-joining method of MEGA 7.0 with 1000 bootstrap replicates based on the p-distance model and pairwise deletion for gap treatment (Kumar et al., 2016). The phylogenetic tree was retouched by FigTree (http://tree.bio.ed.ac.uk/software/figtree/). For the construction of the phylogenetic trees of R2R3-MYB proteins from *Arachis. hypogaea* the same method described above was adopted.

Motif and gene structure analysis

Information on the intron-exon position and splicing sites for each gene model in the corresponding chromosome scaffold was downloaded from PeanutBase. Furthermore, the MEME program (https://meme-suite.org/meme/meme_5.3.3/tools/meme) was used for the identification of motifs of 196 R2R3-MYB protein sequences in peanut. The optimized parameters of MEME were employed as follows: the number of motifs that MEME find, 10; and the optimum width of each motif, 6–60 residues (Bailey et al., 2009). The gene exon-intron pattern and MEME results were also visualized by CFVisual_V2.1.5 (Chen et al., 2022).

Chromosome localization, duplications, and evolutionary analysis of *AhR2R3-MYBs*

The information on chromosome length and *R2R3-MYB* gene locations was acquired from the PeanutBase (https://peanutbase.org) (Supplementary Table 2), and the figure was created by TBtools (Chen et al., 2020). The whole-genome sequences and annotation documents of *A. hypogaea* were downloaded to PeanutBase (https://peanutbase.org). Then, the One Step MCScanx program of TBtools was executed to analyze the synteny relationships of genomes. We identified gene pairs with physical distance within 100 kb, with no more than 10 genes spaced in between, and in the same subgroup as tandem repeat gene pairs according to Hanada et al. (2008). The duplication pattern of the *AhR2R3-MYB* genes was visualized by the Amazing Super Circos package of TBtools. The Ka/Ks value was completed by the Simple Ka/Ks Calculator program of TBtools. Duplication time was calculated by the following formula as described by Bertioli et al. (2016): $T = Ks/2\lambda$ ($\lambda = 8.12 \times 10^{-9}$).

Plant material and stress treatment

The peanut (*Arachis hypogaea* L.) cultivar 'Changhua18', a germplasm resource preserved in our laboratory, was planted in a pot with a 1:1 mixture of nutrient soil and vermiculite, 450 mm long, 335 mm wide, and 170 mm high. When the plants were grown for about 8 weeks, the treatment group was subjected to waterlogging treatment according to Zeng et al. (2021), while the control group was kept under normal growth conditions. The samples (three seedlings per repeat) were collected at 0 h, 6 h, 24 h, 3 days, and 5 days after treatment, respectively. Subsequently, the samples were rapidly frozen using liquid nitrogen and stored at -80°C for RNAseq.

RNA-seq expression analysis

The raw read counts in *Arachis hypogaea* RNAseq samples were downloaded from PeanutBase (https://peanutbase.org). The data were obtained from 22 tissues at different developmental stages in peanut with three biological repeats (Clevenger et al., 2016) with all raw data deposited as BioSamples SAMN03944933–SAMN03944990. HTSeq was used to generate raw reads that were uniquely mapped on the *Arachis hypogaea* genome. StringTie and Ballgown were used for the FPKM calculation (Pertea et al., 2016). The transcript profiles for *AhR2R3-MYB* genes were displayed in TBtools (Chen et al., 2020).

These cDNA libraries generated from the samples were sequenced by Metware Biotechnology Ltd. on the Illumina sequencing platform. (Wuhan, China). Download the reference genome and its annotation files from NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/), use HISAT v2.1.0 to construct the index and compare clean reads to the reference genome. The featureCounts v1.6.2/StringTie v1.3.4d was used to calculate the gene alignment and FPKM. Gene expression patterns were also charted by TBtools. Quantitative RT-PCR (qRT-PCR) was performed to verify the transcriptome data. RNAprep Pure Plant Plus Kit (Tiangen Biotech, Co., Beijing, China) was utilized to extract RNA from control and treated peanut samples. The PrimeScrit^{TM} RT Kit with gDNA eraser user manual was used to prepare the cDNA (perfect real-time, Takara Biomedical Technology, Ltd., Beijing, China). The primers were designed by TBtools and were shown in Supplementary Table 3. Subsequently, qRT-PCR was performed using the ABI 7500 qRT-PCR detection system (ABI, United States) with SYBR Green Kit (Tiangen, Beijing, China). The ABI 7500 real-time PCR program was 95°C for 15 min, followed by 40 cycles of 95° C for 10 s, and 60°C for 30 s in a 20 μl volume. Three technical repeats of qRT-PCR were carried out, and the relative expression level was determined using $2^{-\triangle\triangle Ct}$ technique.The results of differential expression analysis between homeolog pairs of A and B subgenomes across tissues and pod developmental stages in *Arachis hypogaea* were downloaded from PeanutBase (https://peanutbase.org) (Bertioli et al., 2019). Differential gene expression analysis was performed using the

DESeq2 package (v1.14.1) with $\log_2$ fold change >= 1 and Benjamini-Hochberg adjusted P-value < 0.05 as the statistical cutoff for differentially expressed genes. Charts generated by GraphPad Prism 9.

Association analysis of the peanut *R2R3-MYBs* with TBN, PL and RS ratio

Total branch number, length of root and shoot, pod size and RS phenotypes were assessed using a randomized complete block design and replicated in five environments. SNPs of the *R2R3-MYBs* were obtained from transcriptome data set of a peanut germplasm population with 146 accessions (alleles in each polymorphism with minor allele frequency >0.05). Association analysis was performed with TASSEL 3.0 using an MLM Q + K model. Figures were generated by GraphPad Prism 9.

Results

Identification and conserved DBD analysis of R2R3-MYB genes in peanut

A total of 196 *R2R3-MYB* genes were obtained after genome-wide screening and exclusion in *A. hypogaea* cv. Tifrunner. The 196 genes were named *AhMYB001-AhMYB196* according to their physical location on chromosome. The mRNA length of the *AhR2R3-MYBs* ranged from 753 to 3712 bps (Supplementary Table 4); the proteins length ranged from 192 to 916 amino acids, with predicted molecular weights from 21.95 to 103.72 kDa; the theoretical isoelectric point (pI) ranged from 4.60 to 10.32. Furthermore, none of the proteins were predicted to contain transmembrane domains (Supplementary Table 4).

To investigate the MYB conserved domains of R2R3-MYB transcription factors in peanut, we performed WebLogo and multiple alignment (ClustalX) analysis using amino acid sequences of the R2R3 repeats (Supplementary Figure 1). The results showed that the R2 and R3 repeat consisted of two repeats of approximately 51 amino acid residues, and all of them contained highly conserved tryptophan residues (Trp, W) (Supplementary Figure 2). The R2 repeat contains three highly conserved W residues at positions 5, 26, and 47, while only two highly conserved W residues were uncovered at positions 24 and 43 in the R3 repeat, and the W residue at position 5 was generally replaced by phenylalanine or isoleucine (Phe, F or Ile, I). Tryptophan residues formed a hydrophobic core maintaining the stability of the helix-turn-helix (HTH) structure in the DNA binding domain, and the basic and polar amino acids adjacent to that tryptophan were thought to be directly involved in DNA binding (Saikumar et al., 1990). Several highly conserved polar amino acid residues were also found around the third W residue in each repeat, for example, asparagine (Asn, N), arginine (Arg, R), and lysine (Lys, K) around the third tryptophan in the R2 repeat, and K and N near the third W residue in the R3 repeat, which might be directly involved in binding to DNA. These polar amino acids, which are conserved around tryptophan, were also highly conserved in different species, such as *Arabidopsis* and tomato (Stracke et al., 2001; Li et al., 2016). Overall, most of the conserved amino acid residues were mainly distributed between the second and third conserved tryptophan residues, with the first tryptophan residue being relatively less conserved. Therefore, those conserved residues probably maintain the function of the DNA binding domain together with the conserved tryptophan.

Phylogenetic analysis of AhR2R3-MYBs in cultivated peanut and *Arabidopsis*

To investigate the evolutionary relationship between the *R2R3-MYB* genes in peanuts and *Arabidopsis*, the 196 predicted AhR2R3-MYB proteins were subjected to multiple sequence alignment along with 126 *Arabidopsis* R2R3-MYB proteins, and their evolutionary relationships were inferred by constructing a neighbor-joining phylogenetic tree. The 322 R2R3-MYB proteins were classified into 48 subgroups based on the topology and bootstrap value of the phylogenetic tree (Figure 1 and Supplementary Figure 3). Since the *R2R3-MYB* genes have been intensively studied in *A. thaliana* and most subgroups contained at least one *AtR2R3-MYB* gene, we named these subgroups according to the nomenclature of (Kranz et al., 1998) revised by Stracke et al. (2001) and Dubos et al. (2010). When a subgroup name was not yet determined in *A. thaliana*, we named the subgroup after the member of *A. thaliana* with the most distinct functional characteristics. In general, the phylogenetic characteristics of *A. thaliana* described in this paper were generally consistent with those described previously (Stracke et al., 2001; Dubos et al., 2010). The only exception was the *A. thaliana* genes of the subgroups 20 and 25, were split in two respectively (S20a and S20b; S25a and S25b), and the genes from 10 and 24 were merged (S10&24). As shown in Figure 1, 40 subgroups contained at least one gene from peanut, and the other eight subgroups contained genes only from peanut, and they were named as new subgroups 1-8 in this study. The distribution of *AhR2R3-MYBs* in the 40 subgroups was biased, varying from one (NS-8) to 17 members (S14). Notably, the number of *R2R3-MYB* genes in almost all subgroups was unbiased in both subgenomes, revealing a close association between two subgenomes of cultivated peanut as described by (Bertioli et al., 2016; Bertioli et al., 2019). The topology of the neighbor-joining tree for *AhR2R3-MYB* genes was in good agreement with the subgroup described above (Supplementary Figure 4).

Motif composition and gene structure of the *AhR2R3-MYB* genes

To investigate the relationship between subgroup classification and function of the peanut R2R3-MYBs, 10 conserved motifs were identified in the AhR2R3-MYBs through MEME program search (Supplementary Figure 5). The DNA binding domain of AhR2R3-MYB was represented by motifs 3, 6, 1, 2. Motifs 1 and 2 contained the amino acid sequence of the third helix forming the MYB domain, which is involved in the recognition and binding of cis-acting elements (Ogata et al., 1996; Jia et al., 2004). Motifs 5 and 8 were only presented in SAtM88, while motif 9 only presented S20a and

10.3389/fpls.2023.1102174

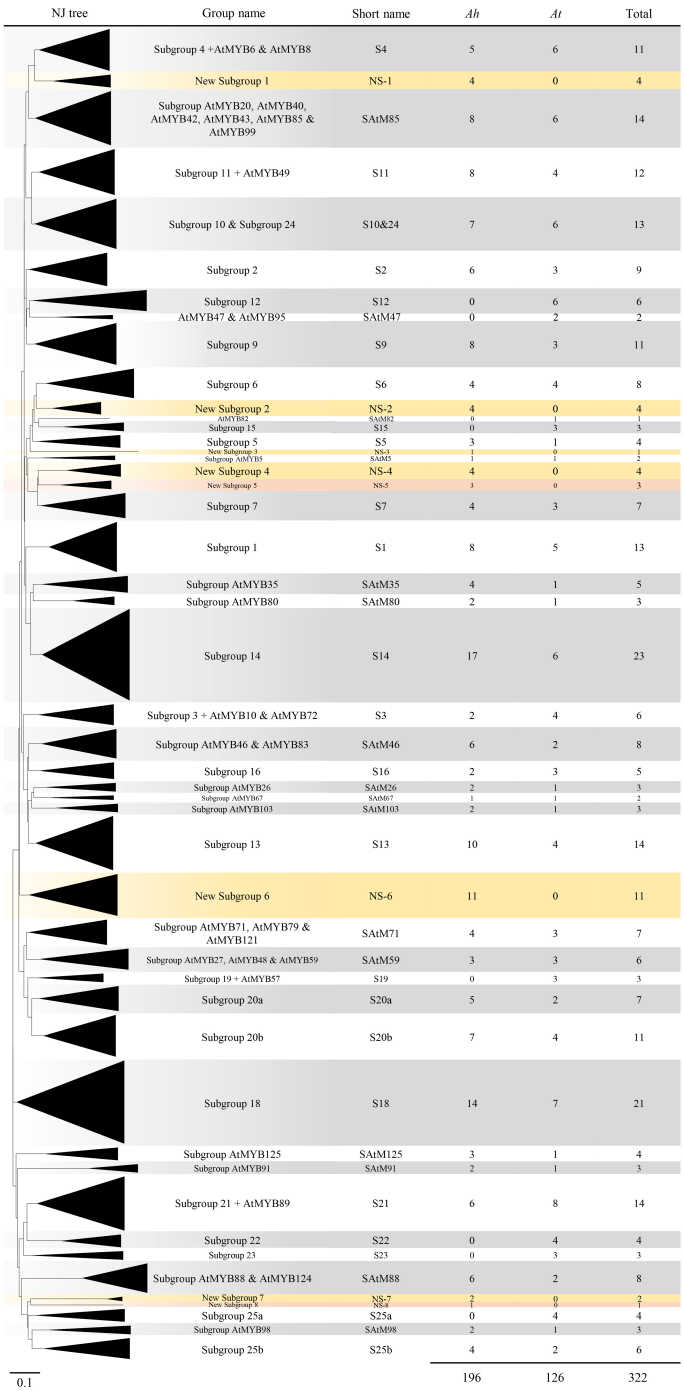


FIGURE 1
Neighbor-joining phylogenetic tree of R2R3-MYB proteins from peanut (*Arachis hypogaea* L.) and *Arabidopsis*. Each triangle represented an R2R3-MYB subgroup, defined based on the topology of the tree and the bootstrap values. Subgroup names were included next to each clade together with a short name to simplify the nomenclature. The number of genes of each species for each subgroup was also included. The eight new subgroups in peanut were marked in yellow or coral, while the other subgroups were in gray and white.

S20b, suggesting potential specific functions of these subgroups (Figures 2A, B). In general, most of the motif compositions of members in the same subgroup were similar at N-terminal, but differ at C-terminal, and the motif compositions of the members in different subgroups were not identical.

Exon and intron structure analysis showed that all *AhR2R3-MYB* genes possessed 1 to 21 exons (Figure 2C). In general, the size of introns was variable, but the locus and phase of introns were relatively conservative among subgroups (Jiang et al., 2004). Most of the *R2R3-MYB* genes (77%) contained two conserved introns, 14% contained one conserved intron, and the rest showed a different number. Genes in the same subgroup have similar gene structures and highly conserved in intron phasing (Figures 2A, C). The multiple introns of a gene provide the opportunity to selectively splice and provide variant proteins that may play different roles in biological processes (Min et al., 2015). Most members of the subgroups are intron-poor

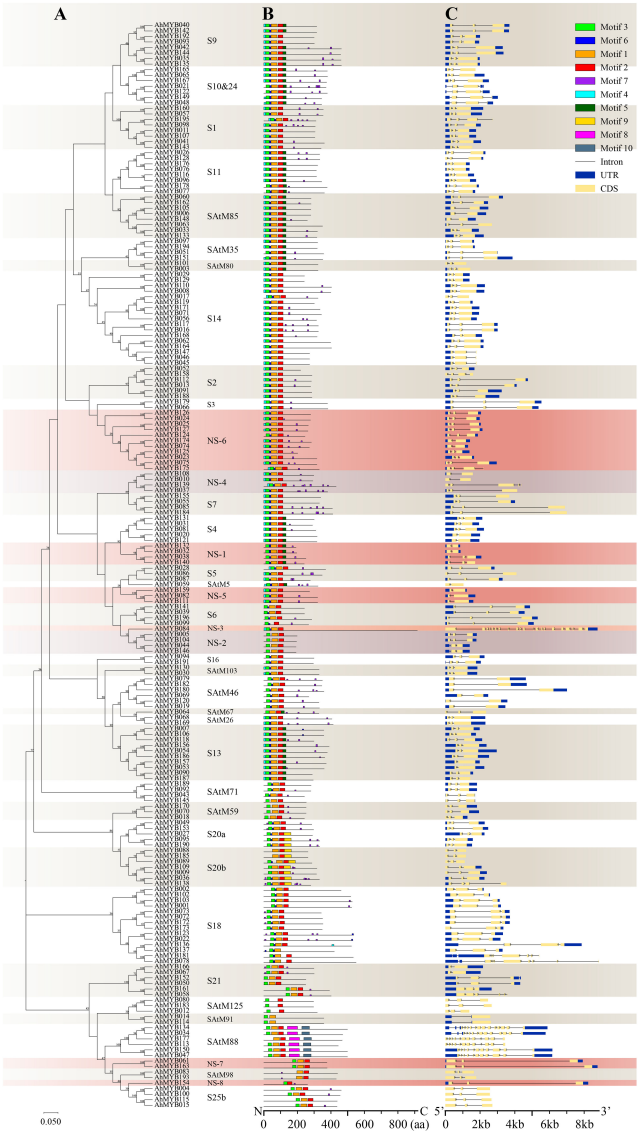


FIGURE 2
Phylogenetic relationships, conserved motifs, and gene structure analysis of peanut R2R3-MYBs. **(A)** The neighbor-joining phylogenetic tree was constructed by aligning the full-length amino acid sequences of 196 R2R3-MYBs in peanut. Coral and purple colors mark eight new subgroups, while beige and white indicate other subgroups. **(B)** The ten conserved motifs were shown in different colors and their specific sequence information was provided in Supplementary Figure S5. **(C)** The yellow box, blue box, and black line in the gene structure diagram represented CDS, UTR, and introns, respectively.

(Contain three or fewer) or intron-less genes. However, we found that genes in subgroups NS-3, S18, and SAtM88 possessed an abundant number of introns, and multiple transcripts were present in all three subgroups except for *AhMYB177*.

Chromosome localization, duplication, and evolution of the R2R3-MYB genes

Chromosomal localization showed that the A and B subgenomes contained 99 and 97 R2R3-MYB genes, respectively (Figure 3). This suggested that the distributions of *R2R3-MYB* genes between the two subgenomes were almost not biased. In addition, we found that the distribution of genes on the corresponding chromosomes was similar between A and B subgenomes, except for Chr07 and Chr17, Chr08 and Chr18. This was probably the result of the complex rearrangement event on chromosomes 7, and 8 of two diploid wild ancestors, and subsequent retention of this rearrangement after polyploidization (Bertioli et al., 2016; Bertioli et al., 2019). The *R2R3-MYB* genes were unevenly distributed among the 20 chromosomes. Chr03 (19), Chr08 (18), Chr13 (18), and Chr18 (15) contained a larger number of *R2R3-MYB* genes. Chr02 (5), Chr07 (4), Chr10 (4), Chr17 (4), and Chr20 (4) possessed fewer *R2R3-MYB* genes. The density of *R2R3-MYB* genes on Chr08 was significantly higher than that on the other 19 chromosomes.

The *R2R3-MYB* genes were far more abundant in *A. hypogaea* than in lower terrestrial plants, suggesting that a large-scale gene duplication event occurred during the evolution of the plants (Du et al., 2015). To explore the mechanism of *R2R3-MYB* gene expansion in the cultivated peanut, we further analyzed the syntenic

10.3389/fpls.2023.1102174

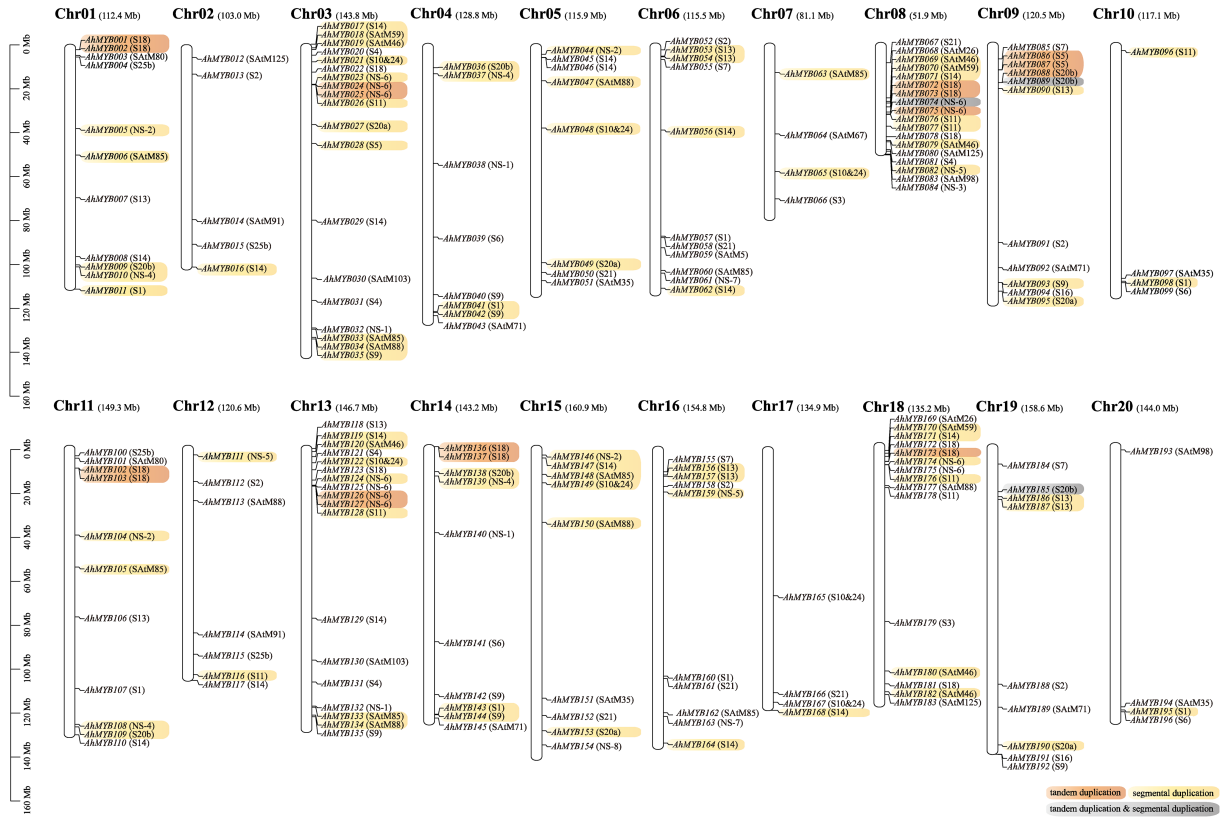


FIGURE 3
Distribution of 196 *R2R3-MYB* genes on 20 chromosomes. The name and length of the chromosome are displayed at the top of each chromosome. Yellow is for genes produced by segment duplication, orange is for genes from tandem duplication, and gray indicated genes that have experienced both types of duplication events. Subgroups were annotated to the right of the gene.

relationships between the peanut *R2R3-MYB* genes. Based on synteny analyses, 45 genes out of the 33 syntenic pairs in the A subgenome underwent segmental duplication events, while 39 genes out of the 24 syntenic pairs in the B subgenome were undergoing segmental duplication events (Figure 3 and Supplementary Table 5). 12 and eight genes in the A and B subgenomes, respectively, experienced tandem duplication episodes. In addition, significant numbers of orthologous genes were found between the A, and B subgenomes (Figure 4 and Supplementary Table 5). Based on synteny analyses, 173 of the 196 *AhR2R3-MYBs* had syntenic relationships between the two subgenomes. We found that some subgroups were expanded mainly by segment duplication, such as subgroups S11, NS-2, NS-4, SAtM85, S20a, S20b, S13, S1, S10&24, SAtM88, SAtM46, SAtM59, and NS-5, while tandem duplication occurred mainly in subgroups S5, NS-6, S20b, and S18 (Figure 3).

To further understand the evolution of R2R3-MYB duplication in peanut, we analyzed the rate of synonymous (Ks) and nonsynonymous substitutions (Ka) in gene duplication pairs (Supplementary Table 5). The Ks values for direct homologous gene pairs between A and B subgenomes range from 0 to 2.47. The frequency distribution peaks at Ks = 0.03, indicating a massive duplication event of *R2R3-MYB* genes 1.85 million years ago (Mya). Segment duplication and tandem duplication may occur in 147.75-35.91 and 163.87-2.20 Mya, respectively. Ka/Ks ratios for paralogous and orthologous ranged from 0-2.10 with an average of 0.27, whereas the ratios for tandem duplication ranged from 0.13-1.02 with an average of 0.36. The Ka/Ks analysis showed that the

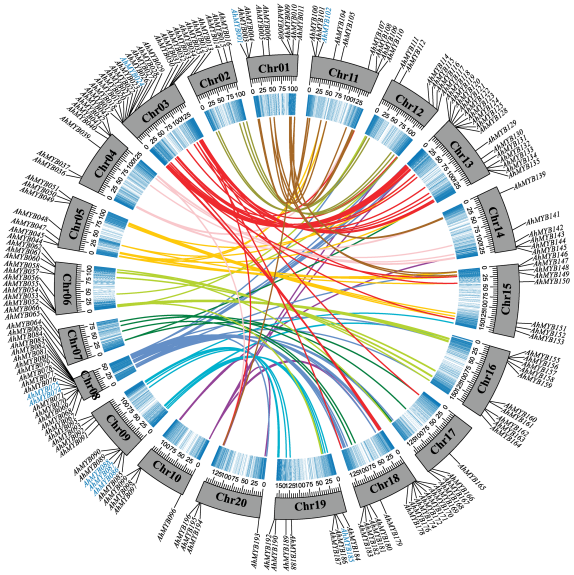


FIGURE 4
Circos diagram depicting the relationships of the chromosomes of *A. hypogaea*. Different colored connecting wires were used for different chromosome pairs. The blue color represented the density of genes. The scale for the gray bars was in megabases. Tandem duplication genes were highlighted in blue.

orthologous gene pair *AhMYB052-AhMYB158* was neutrally selected and the tandem duplication pair *AhMYB173-Arahy.QI53CA* and the orthologous gene pairs *AhMYB010-AhMYB108*, *Arahy.M1LASL-AhMYB118*, and *AhMYB031-AhMYB131* were subjected to positive selection, and all other synteny and tandem duplicated genes were subjected to purifying selection.

Tissue expression profiles of the *R2R3-MYBs*

To further study the expression pattern of the *R2R3-MYB* genes in different tissues and explore its function in peanut growth and development, the tissue expression profiles of the *R2R3-MYB* genes were analyzed by using the transcriptome data of 22 peanut tissues (*AhMYB015* and *AhMYB100* were not detectable in the dataset) (Figure 5 and Supplementary Table 6). The peanut R2R3-MYBs can be clustered into 10 groups according to their tissue expression pattern (Cluster 1-10). Most genes from same phylogenetic subgroup showed similar tissue expression patterns and were clustered into the same cluster, such as *AhMYB023*, *AhMYB024*, *AhMYB074, AhMYB125*, *AhMYB126*, *AhMYB174*, and *AhMYB175* in NS-6, which highly expressed in fruit and pericarp, were cluster in cluster 3 (Figure 5). All the members of subgroup SAtM35, for example, were clustered in cluster 9 (Figure 5). However, a certain proportion of the *AhMYBs* from the same phylogenetic subgroup showed differential tissue expression, such as *AhMYB028* and *AhMYB087* from S5, *AhMYB028* was highly expressed in the reproductive shoot, while *AhMYB087* was mostly expressed in seed (Figure 5), this might be because genes belonging to the same phylogenetic subgroup exercise similar functions, but in different tissues responses to different developmental processes or different environmental stimuli ((Dubos et al., 2010).

Interestingly, members of the eight peanut specific subgroups (NS-1 to 8) were distributed in all clusters except 2 and 4.The two NS-1 members (*AhMYB032* and *AhMYB132*) were strong expressed in perianth; the two collinear gene pairs (*AhMYB005* and *AhMYB104*, *AhMYB044* and *AhMYB146*) from NS-2 subgroup were highly expressed in seed and shoot, respectively. *AhMYB084* (NS-3) had significant expression both in leaf and shoot; the three members in NS-4 (*AhMYB010*, *AhMYB037* and *AhMYB139*) expressed mainly in leaf; *AhMYB082* and *AhMYB111 in* NS-5 showed higher expression level in nodule, but *AhMYB150* was enriched in peg tip; most of the members in NS-6 were clustered in cluster 3 and were highly expressed in pericarp; members in NS-7 (*AhMYB061* and *AhMYB163*) and NS-8 exhibited strong expression in seed. These subgroups contained totally 30 genes, 21 of which were expressed at high levels in reproductive organs, especially the 10 genes in NS-6 that had high transcript abundance in the early pod development stage after peg tip entry, suggesting an important role of these genes in peanut reproductive development.

Comparison of R2R3-MYB gene expression in subgenomes

The expression of homeologous gene pairs from A and B subgenomes was examined in various tissues and developmental phases (Supplementary Table 7 references from (Bertioli et al., 2019)). The total number of homeologous gene pairs with expression biased towards the A subgenome was similar to the previously reported it did not differ significantly from the number with expression biased towards the B subgenome (P = 0.16, binomial test; n = 47 and 42 for A and B, respectively) (Figure 6) (Chalhoub et al., 2014; Zhang et al., 2015). In nine tissues (lateral leaf, seeding leaf, vegetative shoot tip, reproductive shoot tip, perianth, gynoecium, pattee 6 seed, pattee 7 seed, and pattee 8 seed), of the homologous pairs, there were more A subgenome-highly expressed genes rather than B subgenome, whereas the other 10 tissues exhibited the reverse pattern. These differences were more significant in the four reproductive tissues ($P < 0.05$, binomial test). 14 homologous gene pairs had biased expression in just one tissue whereas 18 homologous gene pairs exhibited the same bias in several tissues (Supplementary Table 8). Interestingly, we identified four homologous gene pairs (*AhMYB006* and *AhMYB105*, *AhMYB011* and *AhMYB107*, *AhMYB053* and *AhMYB157*, *AhMYB074* and *AhMYB174*) showed opposite biased expression in specific tissues, such as homolog pair *AhMYB074* and *AhMYB174*, *AhMYB174* owned a higher expression level in lateral leaf {log2FoldChange (B.vs.A homeolog pair comparison):1.906081616}, while *AhMYB074 had a stronger expression level in perianth* {log2FoldChange (B.vs.A homeolog pair comparison): -2.204691792}.

Expression pattern of *R2R3-MYB* genes in peanut under waterlogging treatment

We examined the transcript abundance of the R2R3-MYB gene in peanut seedlings after water logging treatment to further investigate the function of the genes in water logging response. 90 genes from 24 subgroups (NS-6, S10&24, S11, S13, S14, S2, S20a, S20b, S21, SAtM85, SAtM88, etc.) were found response to waterlogging. According to the chronological order of the respond genes, they were divided into five categories (Figure 7). 26 genes showed a strong tendency to be down-regulated following the water logging treatment. Nine genes in the transitional response group showed strong elevated after 6 hours of treatment. Three genes (*AhMYB121*, *AhMYB008*, and *AhMYB089*) were considerably up-regulated among the 12 early responded genes, whereas the others showed a tendency of down- and subsequently up-regulation. At the time point of 3 and 5 days after treatment, a total of 27 and 16 genes were up-regulated, respectively. In addition, the results of qRT-PCR showed that the expression patterns of the 10 selected genes under water logging treatment were generally consistent with the transcriptome results (Supplementary Figure 6) which verified the results based on transcriptome data analysis.

Association analysis of R2R3-MYBs with pod size, total branch number and root-shoot ratio of peanut

To determine the role of *R2R3-MYB* genes in peanut, we conducted candidate gene association analysis using 59 single-nucleotide polymorphisms in *Ad* and *AiR2R3-MYBs* identified from transcriptome data of 146 peanut varieties (Supplementary Table 9)

10.3389/fpls.2023.1102174

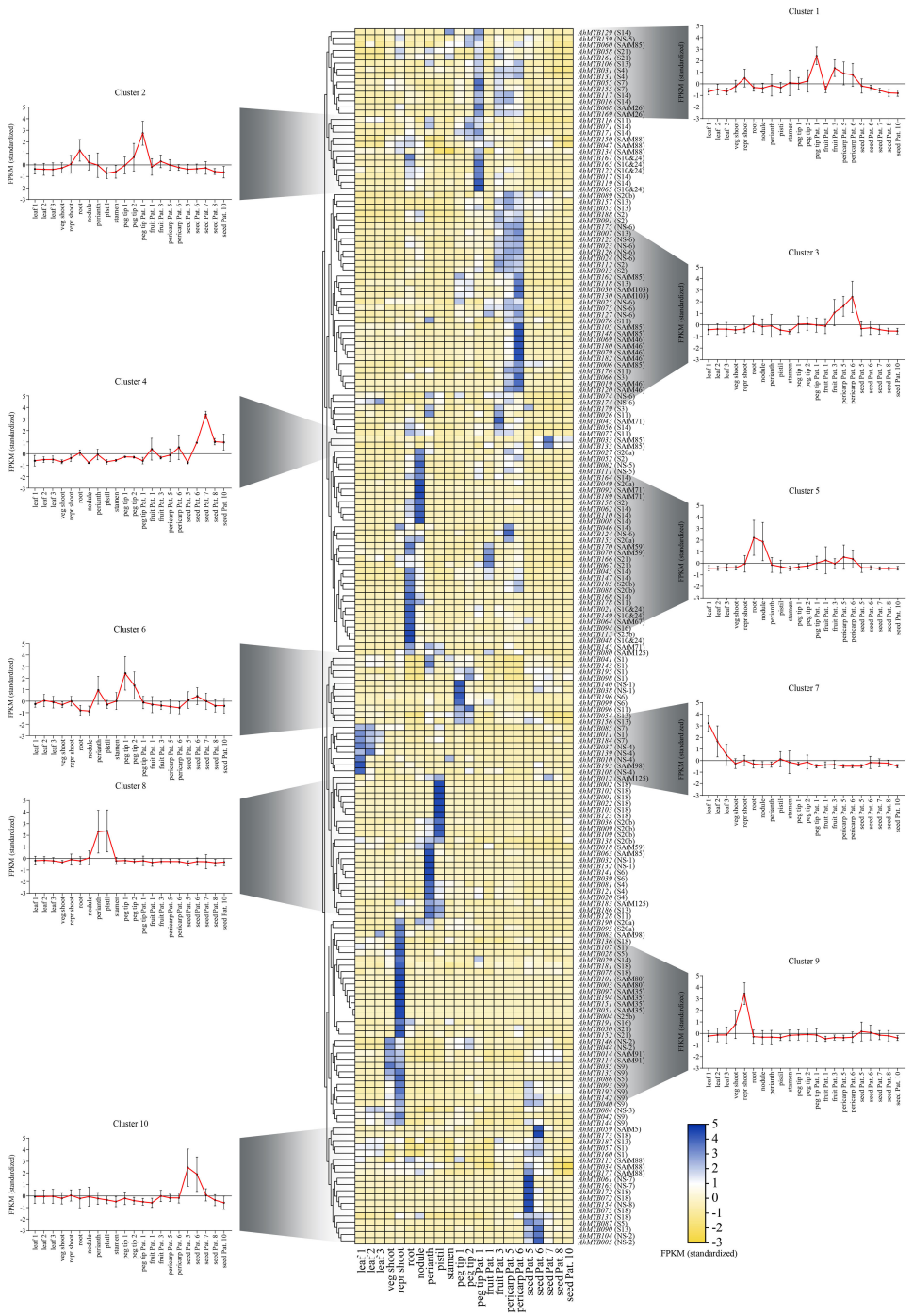


FIGURE 5
Heat map of the RNAseq transcript abundance pattern of the 194 R2R3-MYB genes in 22 different tissues in 10 expression clusters. For each gene, its name and the subgroup it belongs to were displayed on the right side of the heatmap. The expression pattern was generated based on the fragments per kilobase of exon per million fragments (FPKM) and analyzed by heatmap hierarchical clustering. The color scale (representing −2 to 5) was shown. The meanings of the abbreviations of the 22 tissues were as follows: seedling leaf 10 days post-emergence (leaf 1), main stem leaf (leaf 2), lateral stem leaf (leaf 3), vegetative shoot tip from the 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), Pattee 1 pod (fruit Pat. 1), Pattee 3 pod (fruit Pat. 3), Pattee 5 pericarp (pericarp Pat. 5), Pattee 6 pericarp (pericarp Pat. 6), Pat - tee 5 seed (seed Pat. 5), Pattee 6 seed (seed Pat. 6), Pattee 7 seed (seed Pat. 7), Pattee 8 seed (seed Pat. 8), and Pattee 10 seed (seed Pat. 10).

and phenotype of total branch number (TBN), pod length (PL) and root-shoot ratio (RS ratio) variation collected in five environments (Supplementary Table 10). One polymorphic site [A03_122181829 (C/S/G)] was uncovered highly associated with TBN, PL and RS ratio variation in five environments ($P < 0.01$) (Supplementary Table 11 and Figure 8A), located in the third exon region of *AdMYB03-18 (AdMYB03-18)* (Figure 8B). A03_122181829 mainly formed three haplotypes (A03_122181829 (C/S/G)) in the associated population

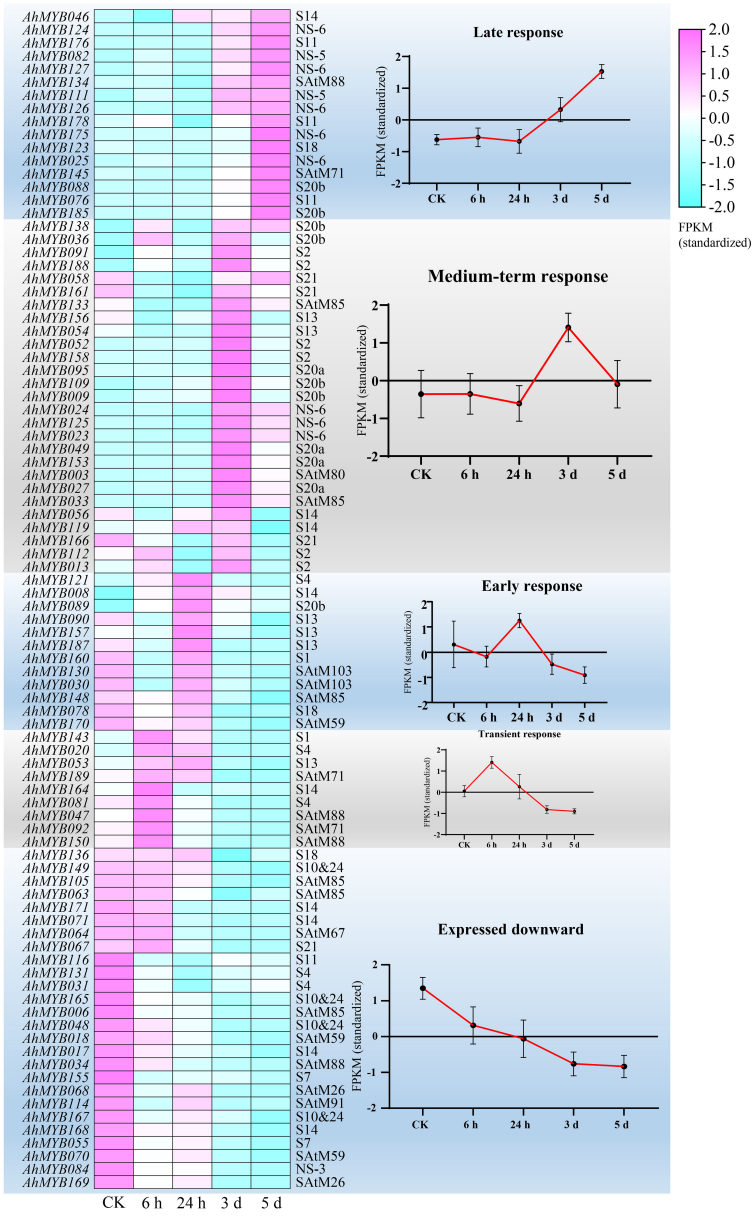


FIGURE 6
RNAseq transcript abundance patterns of 90 R2R3-MYB genes in root and stem tissues in five expression groups during water logging stress are shown as a heat map. The left and right sides of the heatmap, respectively, showed the names of each gene and the subgroup to which it belonged. Heatmap hierarchical clustering was used to construct the expression pattern based on the fragments per kilobase of exon per million fragments (FPKM). The color scale, which ranged from -2.0 to 2.0, was displayed. The labels at the bottom of the heat map indicate, from left to right, the control group, 6 h, 24 h, 3 days, and 5 days after water logging treatment.

(Figure 8C). Analysis results indicated that TBN in haplotype G was notably higher than those in haplotype C, while PL in haplotype C was higher than PL in haplotype G and S, RS ratio in haplotype S was higher than RS ratio in haplotype C and G (Figure 8D).

Discussion

The R2R3-MYB genes have been identified in various species, such as *Arabidopsis*, rice, maize, soybean, eucalyptus, tomato, and Chinese bayberry (Stracke et al., 2001; Jia et al., 2004; Chen et al., 2006; Du et al., 2012a; Du et al., 2012b; Soler et al., 2015; Li et al., 2016; Cao et al., 2021). In this study, 196 *R2R3-MYB* genes were identified from cultivated peanut (*Arachis hypogaea* L.), which were further characterized by co-phylogenetic analysis with the corresponding genes of *A. thaliana*. According to the topology of the resulting tree and the bootstrap values, a total of 322 R2R3-MYBs were divided into 48 subgroups. Among these subgroups, the *Arabidopsis* genes of the subgroups S20 and S25 were split into two subgroups (S20a and S20b; S25a and S25b), and the genes from subgroups S10 and S24 were merged (S10&24). The combination of the two subgroups into one group suggested that the genes in these subgroups may have close evolutionary relationships, and there is differentiation in function. In some subgroups (S12, S15, S19, S22, S23, S25a, SAtM47, AtMYB82), only genes in *Arabidopsis* were present,

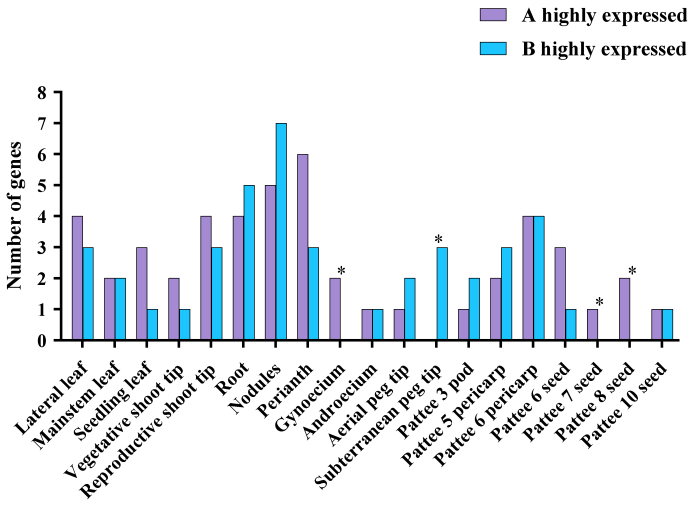


FIGURE 7
Differential expression of R2R3-MYB homologous genes in *A. hypogaea* cv. Tifrunner. To look for variations in the levels of gene expression in 19 distinct organs, homeologs were compared. Each subgenome was represented by the number of homeologous genes that were more strongly expressed (log2 fold change ≥ 1, Benjamini-Hochberg adjusted P < 0.05; Wald test) in each subgenome is represented. P-value correspond to binomial test with the odds of A genes being more highly expressed at 0.5 probability. *P < 0.05, others : not significant.

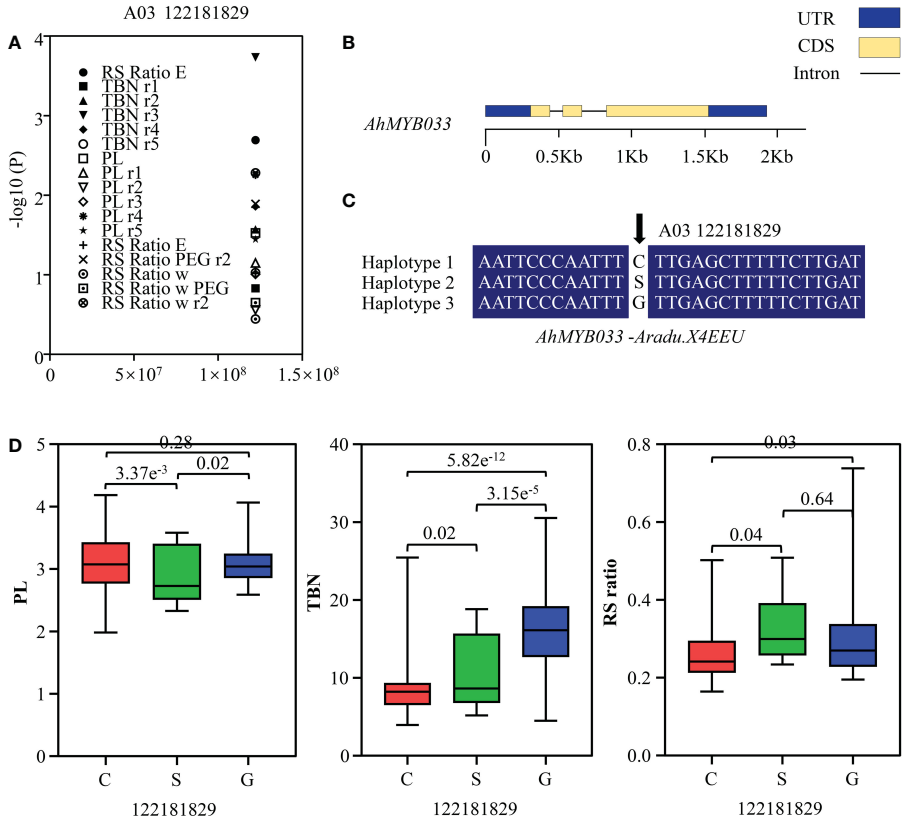


FIGURE 8
Association mapping results and the phenotypes of the polymorphic sites of peanut R2R3-MYBs associated with TBN, PL and RS ratio variation. **(A)** Association results between TBN/PL/RS ratio and the polymorphisms in *AhMYB033*. The vertical and horizontal axes of the scatter plot indicated the P value and the position of the SNP, respectively, and the phenotypes under different environmental treatments were represented by various patterns. **(B)** Gene structure of *AhMYB033*. The yellow box, blue box, and black line were CDS, UTR, and intron, respectively. **(C)** Sequences of three sites were significantly associated with TBN, PL and RS ratio variation. **(D)** Phenotypic comparison of haplotypes of the three associated sites with TBN, PL and RS ratio in five environments of the population. Three different colored boxes were used to indicate the phenotypes of the three haplotypes.

indicating that after peanut differentiated from *Arabidopsis*, genes were either lost in peanut or acquired in *Arabidopsis*. Moreover, in our results, there were eight subgroups (NS-1, 2, 3, 4, 5, 6, 7, 8) specific to peanut, indicating these genes might be newly expanded genes after differentiation from *Arabidopsis*. Several other genetic characteristics must be considered when performing subfamily classifications, including the presence of highly conserved intron patterns and motif distribution within each subfamily (Du et al., 2015). In this study, exon-intron pattern and motif distribution within each subgroup were also highly conserved, which independently supports our phylogeny analysis and classification results.

As genes in the same subgroup are believed to have relatively similar roles, sequence-based homology classification is crucial for developing hypotheses about the functions of R2R3-MYB genes that have not yet been studied in model species. It is also crucial for defining putative ortholog relationships with known genes in model species to ascertain the functions of genes in nonmodal species (Jiang et al., 2004). The function of novel genes in *Arabidopsis* and other species can be inferred by protein structure and expression patterns (Dubos et al., 2010). *Arabidopsis* members in subgroup S6 were involved in the phenylpropanoid pathway, which can activate the late biosynthetic genes (LBGs) leading to both anthocyanin and PA biosynthesis (Gonzalez et al., 2008), suggesting a potential regulatory function of the peanut *R2R3-MYB* genes in S6 in the phenylpropanoid metabolic pathway. (Zhao et al., 2019) identified the *AhTc1* gene encoding an R2R3-MYB transcription factor controlling peanut purple testa color by whole-genome resequencing-based QTL-seq. In this work, *AhTc1* was given the name *AhMYB099* and belongs to subgroup S6. The identification of *AhTc1* supports the above hypothesis, suggesting that these theories help to speculate on unknown gene functions. Similarly, in *Arabidopsis*, many studies have shown that members of S18 are involved in abscisic acid-mediated responses to environmental signals and in promoting anther and pollen development (Millar and Gubler, 2005; Reyes and Chua, 2007). *AtMYB37*, *AtMYB38*, and *AtMYB84*, members of subgroup S14, partly redundantly control axillary meristem tissue development in *Arabidopsis* (Keller et al., 2006; Muller et al., 2006). Under adverse circumstances, the root growth-specific regulator *AtMYB68* (subgroup 14) has an impact on the development of the entire plant. Members of subgroup S18 were grouped in gene expression clusters 8, 9, and 10. The *AhR2R3-MYB* genes in subgroup S14 were highly expressed in the subterranean peg tips, roots, and reproductive shoot tips. Therefore, it can be inferred that these genes perform similar functions as the same subgroup of *Arabidopsis* genes.

Gene expression profiles provided important threads for the study of gene function. In the present study, we also explored the role of the *R2R3-MYB* gene in water logging stress. A total of 90 genes from 24 subgroups responded to water logging stress in roots and stems. However, some of these 90 genes were not found to be highly expressed in roots and stems in the previous expression analysis of 22 tissues. For example, *AhMYB076* and *AhMYB176* were only expressed at high levels in the pericarp, and *AhMYB071* and *AhMYB171* were expressed at high levels in the underground part of the peg tip and the early developing pods. This mi ght be because these genes perform similar functions but have different expression patterns or all respond to certain environmental stimuli (Dubos et al., 2010). In addition, the peg tip of the underground parts becomes more root-like after reaching into the soil, which seems to explain the presence of these genes in the roots or stems in response to water logging stress (Kumar et al., 2019). The important roles of R2R3-MYBs in response to abiotic stress have been reported several times before in *Arabidopsis*, jatropha, sesame and maize (Dubos et al., 2010; Thirunavukkarasu et al., 2013; Juntawong et al., 2014; Mmadi et al., 2017; Yu et al., 2020). For example, AtMYB60 in subgroup S1 is involved in ABA-mediated control of stomatal opening and closing in response to drought (Cominelli et al., 2005). AtMYB15 in subgroup S2 is involved in cold stress response (Agarwal et al., 2006). AtMYB41 in subgroup S11 probably affects dehydration response after osmotic stress (Cominelli et al., 2008; Lippold et al., 2009). Several MYBs were significantly induced in jatropha, sesame and maize (Thirunavukkarasu et al., 2013; Juntawong et al., 2014; Mmadi et al., 2017; Yu et al., 2020). However, the expression pattern of R2R3-MYB genes under water logging stress exhibited significant temporal specificity. 18 genes showed a continuous down-regulation after treatment, while the remaining genes showed an up-regulation after 6 h, 24 h, 3 days, and 5 days after treatment, respectively. This stress response with a high number of participating genes and a certain temporal pattern disclosed a complex and ordered regulatory network involving *R2R3-MYB* genes in response to water logging stress.

Whole genome duplication or polyploidization occurs frequently in angiosperms and provides a great deal of material for plant evolution (Jiao, 2018). Most of the replicated genes from WGD are eventually lost (Lynch and Conery, 2000; Conant et al., 2014), and those that are retained are often biased toward certain functional gene taxa (Panchy et al., 2016). In this research, the *R2R3-MYB* gene also had a high homology ratio (88.3 %) between the two subgenomes of tetraploid peanut (*Arachis hypogaea*). As previously reported in *Arabidopsis* and Brassica, the genes that are typically retained (and therefore enriched) are kinases, transcriptional proteins, transcription factors, and genes functioning in transcriptional regulation (Maere et al., 2005; Liu et al., 2014). High retention of the *R2R3-MYB* homolog after polyploidization in peanut reveals functional conservation of the *R2R3-MYB* gene. Under natural selection, the ploidized genes experience different fates, such as partial copy loss and loss of function (pseudogenization), partial copy gaining new function, or each exercising part of the function of the ancestral gene (Conant and Wolfe, 2008; Panchy et al., 2016). The results showed that, in general, the homologous gene pairs were not significantly biased to be expressed between the two subgenomes. However, considering only one tissue, homologous pairs showed biased expression among subgenomes in the four reproductive tissues. A total of 32 homologous gene pairs exhibited biased expression between subgenomes in the same or different tissues, suggesting that some genes in these homologous pairs may be lost in function. More homologous genes did not have significantly biased expression and they may have been sub-functionalized, each exercising part of the function of the ancestral gene. Moreover, we identified four homologous pairs exhibiting different biases in different tissues, which reveal a novel functionalization of the *R2R3-MYB* gene. In conclusion, these findings revealed the fate of the *AhR2R3-MYB* genes after undergoing polyploidization and collectively maintaining a functional dosage balance.

Conclusion

In this study, 196 *R2R3-MYB* genes were identified in cultivated peanut genome. A phylogenetic study with *Arabidopsis* divided the

196 genes into 40 subgroups. Motif composition and gene structure analyses independently confirmed the subgroup delineation. According to the synteny analysis, polyploidization, facilitated in the expansion of the *AhR2R3-MYB* genes. Tissue expression pattern of *R2R3-MYB* genes, subgroup functional conservation, and diversification were discovered using gene expression analysis. The varied outcomes of homologous genes following polyploidization were revealed by the biased expression of homologous pairs in the two sub genomes. 90 genes exhibited a clearly time-specific expression pattern when stressed by waterlogging and AhMYB33 was identified by association analysis highly correlated with total branch number (TBN), pod length (PL) and root-shoot ratio (RS ratio). In conclusion, our research advances knowledge of the role of R2R3-MYB transcription factors in cultivated peanut, particularly in response to waterlogging stress.

Data availability statement

The raw data generated in this study have been deposited in the NCBI repository, accession number PRJNA291488.

Author contributions

LW conceived the idea of the paper, SW and ZX carried out all the experiments and data analyses. SW, YY, and WR prepared the figures and tables. SW wrote the manuscript, LW and JF made modifications to the article. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by grants from the National Natural Science Foundation of China (Nos. 31971820, 32160433, and 32160101) 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.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1102174/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Consistent sequence alignment of the R2R3-MYB domain in *Arachis hypogaea*. The alignment of 196 AhR2R3-MYB domains was performed using ClustalX and DNAMAN. The shading of the alignment represents different degrees of conservation among sequences, respectively. The sites of five highly conserved tryptophan residues (W) were indicated by pentacles.

SUPPLEMENTARY FIGURE 2
Consensus sequence and the level of conservation of R2R3-type MYB domains from peanut. The sequence logos of the R2 and R3 MYB repeats were based on multiple alignment analyses of 196 typical AhR2R3-MYB domains performed with ClustalX 2.1. The vertical axis indicated the degree of amino acid conservation, and the horizontal axis indicated the position of the amino acid on each repeat. The conserved tryptophan residues (Trp, W) in the MYB domain were marked with black asterisks. The replaced residues in the R3 repeat were shown by blue asterisks.

SUPPLEMENTARY FIGURE 3
Phylogenetic NJ tree constructed with 126 *Arabidopsis* and 196 peanut R2R3-MYB proteins. Bootstrap values were on the branch node. The subgroup labels were marked on the right side of the tree.

SUPPLEMENTARY FIGURE 4
Phylogenetic NJ tree constructed using 196 AhR2R3-MYB proteins. Bootstrap values were displayed at the branch nodes. The subgroup labels were labeled on the right side of the tree.

SUPPLEMENTARY FIGURE 5
10 MEME motif sequence logos in AhR2R3-MYBs.

SUPPLEMENTARY FIGURE 6
qRT-PCR verification of the expression of *AhR2R3-MYBs* after water logging stress.

References

Agarwal, M., Hao, Y., Kapoor, A., Dong, C. H., Fujii, H., Zheng, X., et al. (2006). A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* 281, 37636–37645. doi: 10.1074/jbc.M605895200

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

Bedon, F., Ziolkowski, L., Walford, S. A., Dennis, E. S., and Llewellyn, D. J. (2014). Members of the MYBMIXTA-like transcription factors may orchestrate the initiation of fiber development in cotton seeds. *Front. Plant Sci.* 5, 179. doi: 10.3389/fpls.2014.00179

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517

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, 877–884. doi: 10.1038/s41588-019-0405-z

Cao, Y., Jia, H., Xing, M., Jin, R., Grierson, D., Gao, Z., et al. (2021). Genome-wide analysis of MYB gene family in Chinese bayberry (*Morella rubra*) and identification of members regulating flavonoid biosynthesis. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.691384

Cao, Y., Li, K., Li, Y., Zhao, X., and Wang, L. (2020). MYB transcription factors as regulators of secondary metabolism in plants. *Biol. (Basel)* 9 (3), 61. doi: 10.3390/biology9030061

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chen, L., Yang, H., Fang, Y., Guo, W., Chen, H., Zhang, X., et al. (2021). Overexpression of GmMYB14 improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnol. J.* 19, 702–716. doi: 10.1111/pbi.13496

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, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., et al. (2019). Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol. Plant* 12, 920–934. doi: 10.1016/j.molp.2019.03.005

Chen, H., Song, X., Shang, Q., Feng, S., and Ge, W. (2022). CFVisual: an interactive desktop platform for drawing gene structure and protein architecture. *BMC Bioinf.* 23. doi: 10.1186/s12859-022-04707-w

Chen, Y., Yang, X., He, K., Liu, M., Li, J., Gao, Z., et al. (2006). The MYB transcription factor superfamily of *Arabidopsis*: Expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol. Biol.* 60, 107–124. doi: 10.1007/s11103-005-2910-y

Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01446

Cominelli, E., Galbiati, M., Vavasseur, A., Conti, L., Sala, T., Vuylsteke, M., et al. (2005). A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Curr. Biol.* 15, 1196–1200. doi: 10.1016/j.cub.2005.05.048

Cominelli, E., Sala, T., Calvi, D., Gusmaroli, G., and Tonelli, C. (2008). Over-expression of the arabidopsis AtMYB41 gene alters cell expansion and leaf surface permeability. *Plant J.* 53, 53–64. doi: 10.1111/j.1365-313X.2007.03310.x

Conant, G. C., Birchler, J. A., and Pires, J. C. (2014). Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* 19, 91–98. doi: 10.1016/j.pbi.2014.05.008

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

Du, H., Feng, B., Yang, S., Huang, Y., Tang, Y., and Wu, K. (2012a). The R2R3-MYB transcription factor gene family in maize. *PloS One* 7, e37463–e37463. doi: 10.1371/journal.pone.0037463

Du, H., Liang, Z., Zhao, S., Nan, M., Tran, L. P., Lu, K., et al. (2015). The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: New insights into subfamily classification and expansion. *Sci. Rep.* 5. doi: 10.1038/srep11037

Du, H., Yang, S., Liang, Z., Feng, B., Liu, L., Huang, Y., et al. (2012b). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* 12, 106–106. doi: 10.1186/1471-2229-12-106

Goicoechea, M., Lacombe, E., Legay, S., Mihaljevic, S., Rech, P., Jauneau, A., et al. (2005). EgMYB2, a new transcriptional activator from Eucalyptus xylem, regulates secondary cell wall formation and lignin biosynthesis. *Plant J.* 43, 553–67. doi: 10.1111/j.1365-313X.2005.02480.x

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., and Shiu, S. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003. doi: 10.1104/pp.108.122457

He, J., Liu, Y., Yuan, D., Duan, M., Liu, Y., Shen, Z., et al. (2020). An R2R3 MYB transcription factor confers brown planthopper resistance by regulating the phenylalanine ammonia-lyase pathway in rice. *Proceedings of the National Academy of Sciences* 117, 271–277. doi: 10.1073/pnas.1902771116

Jia, L., Clegg, M. T., and Jiang, T. (2004). Evolutionary dynamics of the DNA-binding domains in putative *R2R3-MYB* genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.* 134, 575–585. doi: 10.1104/pp.103.027201

Jiang, C., Gu, X., and Peterson, T. (2004). Identification of conserved gene structures and carboxy-terminal motifs in the myb gene family of *Arabidopsis* and *Oryza sativa* l. ssp. indica. *Genome Biol.* 5, R46–R46. doi: 10.1186/gb-2004-5-7-r46

Jian, L., Kang, K., Choi, Y., Suh, M. C., and Paek, N. C. (2022). Mutation of OsMYB60 reduces rice resilience to drought stress by attenuating cuticular wax biosynthesis. *Plant J.* 112 (2), 339–351. doi: 10.1111/tpj.15947

Jiao, Y. (2018). Double the genome, double the fun: Genome duplications in angiosperms. *Mol. Plant* 11, 357–358. doi: 10.1016/j.molp.2018.02.009

Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant *MYB*-gene family. *Plant Mol. Biol.* 41, 577–585. doi: 10.1023/A:1006319732410

Juntawong, P., Sirikhachornkit, A., Pimjan, R., Sonthirod, C., Sangsrakru, D., Yoocha, T., et al. (2014). Elucidation of the molecular responses to waterlogging in jatropha roots by transcriptome profiling. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00658

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

Keller, T., Abbott, J., Moritz, T., and Doerner, P. (2006). *Arabidopsis* REGULATOR OF AXILLARY MERISTEMS1 controls a leaf axil stem cell niche and modulates vegetative development. *Plant Cell* 18, 598–611. doi: 10.1105/tpc.105.038588

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., et al. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *Plant J.* 16, 263–276. doi: 10.1046/j.1365-313x.1998.00278.x

Kumar, R., Pandey, M. K., Roychoudhry, S., Nayyar, H., Kepinski, S., and Varshney, R. K. (2019). Peg biology: Deciphering the molecular regulations involved during peanut peg development. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01289

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

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

Legay, S., Lacombe, E., Goicoechea, M., Brière, C., Séguin, A., Mackay, J., et al. (2007). Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Sci.* 173, 542–549. doi: 10.1016/j.plantsci.2007.08.007

Legay, S., Sivadon, P., Blervacq, A. S., Pavy, N., Baghdady, A., Tremblay, L., et al. (2010). EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in Arabidopsis and poplar. *New Phytol.* 188, 774–86. doi: 10.1111/j.1469-8137.2010.03432.x

Li, Z., Peng, R., Tian, Y., Han, H., Xu, J., and Yao, Q. (2016). Genome-wide identification and analysis of the MYB transcription factor superfamily in solanum lycopersicum. *Plant Cell Physiol.* 57, 1657–1677. doi: 10.1093/pcp/pcw091

Lippold, F., Sanchez, D. H., Musialak, M., Schlereth, A., Scheible, W. R., Hincha, D. K., et al. (2009). AtMyb41 regulates transcriptional and metabolic responses to osmotic stress in *Arabidopsis*. *Plant Physiol.* 149, 1761–1772. doi: 10.1104/pp.108.134874

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A., et al. (2014). The brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5, 3930. doi: 10.1038/ncomms4930

Liu, J., Osbourn, A., and Ma, P. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8, 689–708. doi: 10.1016/j.molp.2015.03.012

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., et al. (2005). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5454–5459. doi: 10.1073/pnas.0501102102

Martin, C., and Paz-Ares, J. (1997). MYB transcription factors in plants. *Trends Genet.* 13, 67–73. doi: 10.1016/s0168-9525(96)10049-4

Millar, A. A., and Gubler, F. (2005). The *Arabidopsis* GAMYB-like genes, *MYB33* and *MYB65*, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 17, 705–721. doi: 10.1105/tpc.104.027920

Min, X. J., Powell, B., Braessler, J., Meinken, J., Yu, F., and Sablok, G. (2015). Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. *BMC Genomics* 16. doi: 10.1186/s12864-015-1914-5

Mmadi, M. A., Dossa, K., Wang, L., Zhou, R., Wang, Y., Cisse, N., et al. (2017). Functional characterization of the versatile MYB gene family uncovered their important roles in plant development and responses to drought and waterlogging in sesame. *Genes (Basel).* 8 (12), 362. doi: 10.3390/genes8120362

Muller, D., Schmitz, G., and Theres, K. (2006). Blind homologous *R2R3 myb* genes control the pattern of lateral meristem initiation in *Arabidopsis*. *Plant Cell* 18, 586–597. doi: 10.1105/tpc.105.038745

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., et al. (1996). The cavity in the hydrophobic core of myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.* 3, 178–187. doi: 10.1038/nsb0296-178

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A., and Saedler, H. (1987). The regulatory cl locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO journal* 6, 3553–3558

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Reyes, J. L., and Chua, N. H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J.* 49, 592–606. doi: 10.1111/j.1365-313X.2006.02980.x

Saikumar, P., Murali, R., and Reddy, E. P. (1990). Role of tryptophan repeats and flanking amino acids in myb-DNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8452–8456. doi: 10.1073/pnas.87.21.8452

Sharma, K. K., and Bhatnagar-Mathur, P. (2006). Peanut (*Arachis hypogaea* l.). *Methods Mol. Biol.* 343, 347–358. doi: 10.1385/1-59745-130-4:347

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., and Shimizu, K. K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* 26, 193–207. doi: 10.1111/mec.13738

Soler, M., Camargo, E. L., Carocha, V., Cassan-Wang, H., San, C. H., Savelli, B., et al. (2015). The eucalyptus grandis R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. *New Phytol.* 206, 1364–1377. doi: 10.1111/nph.13039

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* 4, 447–456. doi: 10.1016/s1369-5266(00)00199-0

Su, L., Lv, A., Wen, W., Fan, N., Li, J., Gao, L., et al. (2022). MsMYB741 is involved in alfalfa resistance to aluminum stress by regulating flavonoid biosynthesis. *Plant J.* 112, 756–771. doi: 10.1111/tpj.15977

Thirunavukkarasu, N., Hossain, F., Mohan, S., Shiriga, K., Mittal, S., Sharma, R., et al. (2013). Genome-wide expression of transcriptomes and their co-expression pattern in subtropical maize (*Zea mays* l.) under waterlogging stress. *PloS One* 8 (8), e70433. doi: 10.1371/journal.pone.0070433

Toomer, O. T. (2018). Nutritional chemistry of the peanut (*Arachis hypogaea*). *Crit. Rev. Food Sci. Nutr.* 58, 3042–3053. doi: 10.1080/10408398.2017.1339015

Vimolmangkang, S., Han, Y., Wei, G., and Korban, S. S. (2013). An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* 13, 176. doi: 10.1186/1471-2229-13-176

Yu, F., Tan, Z., Fang, T., Tang, K., Liang, K., and Qiu, F. (2020). A comprehensive transcriptomics analysis reveals long non-coding RNA to be involved in the key metabolic pathway in response to waterlogging stress in maize. *Genes (Basel).* 11 (3), 267. doi: 10.3390/genes11030267

Zeng, R., Chen, T., Wang, X., Cao, J., Li, X., Xu, X., et al. (2021). Physiological and expressional regulation on photosynthesis, starch and sucrose metabolism response to waterlogging stress in peanut. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.601771

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* l. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207

Zhao, Y., Ma, J., Li, M., Deng, L., Li, G., Xia, H., et al. (2019). Whole-genome resequencing based QTL seq identified *AhTc1* gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. *Plant Biotechnol. J.* 18, 96–105. doi: 10.1111/pbi.13175

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876. doi: 10.1038/s41588-019-0402-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 March 2023
DOI 10.3389/fpls.2023.1123644

OPEN ACCESS

EDITED BY
Muhammad Ali Abid,
Agricultural Genomics Institute at Shenzhen (CAAS), China

REVIEWED BY
Pankaj Kumar,
Dr. Yashwant Singh Parmar University of Horticulture and Forestry, India
Phetole Mangena,
University of Limpopo, South Africa
Muhammad Azhar Nadeem,
Sivas University of Science and Technology, Türkiye

*CORRESPONDENCE
Yong Lei
leiyong@caas.cn
Boshou Liao
lboshou@hotmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 14 December 2022
ACCEPTED 14 February 2023
PUBLISHED 01 March 2023

CITATION
Huai D, Wu J, Xue X, Hu M, Zhi C, Pandey MK, Liu N, Huang L, Bai D, Yan L, Chen Y, Wang X, Kang Y, Wang Z, Jiang H, Lei Y, Varshney RK and Liao B (2023) Red fluorescence protein (DsRed2) promotes the screening efficiency in peanut genetic transformation. *Front. Plant Sci.* 14:1123644. doi: 10.3389/fpls.2023.1123644

Red fluorescence protein (DsRed2) promotes the screening efficiency in peanut genetic transformation

Dongxin Huai[1†], Jie Wu[1†], Xiaomeng Xue[1], Meiling Hu[1], Chenyang Zhi[1], Manish K. Pandey[2], Nian Liu[1], Li Huang[1], Dongmei Bai[3], Liying Yan[1], Yuning Chen[1], Xin Wang[1], Yanping Kang[1], Zhihui Wang[1], Huifang Jiang[1], Yong Lei[1*], Rajeev K. Varshney[2,4] and Boshou Liao[1*]

[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, China, [2]Center of Excellence in Genomics and Systems Biology (CEGSB), International Crops Research Institute of the Semi-Arid Tropics (ICRISAT), Hyderabad, India, [3]Institute of Industrial Crops, Shanxi Agricultural University, Taiyuan, China, [4]State Agricultural Biotechnology Centre, Crop Research Innovation Centre, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

Peanut (*Arachis hypogaea* L.), one of the leading oilseed crops worldwide, is an important source of vegetable oil, protein, minerals and vitamins. Peanut is widely cultivated in Asia, Africa and America, and China is the largest producer and consumer of peanut. Genetic engineering has shown great potential to alter the DNA makeup of an organism which is largely hindered by the low transformation and screening efficiency including in peanut. DsRed2 is a reporter gene widely utilized in genetic transformation to facilitate the screening of transformants, but never used in peanut genetic transformation. In this study, we have demonstrated the potential of the red fluorescence protein DsRed2 as a visual reporter to improve screening efficiency in peanut. DsRed2 was firstly expressed in protoplasts isolated from peanut cultivar Zhonhua 12 by PEG, and red fluorescence was successfully detected. Then, DsRed2 was expressed in peanut plants Zhonghua 12 driven by 35S promoter *via Agrobacterium tumefaciens*-mediated transformation. Red fluorescence was visually observed in calli and regenerated shoots, as well as in roots, leaves, flowers, fresh pod shells and mature seeds, suggesting that transgenic screening could be initiated at the early stage of transformation, and continued to the progeny. Upon screening with DsRed2, the positive plant rate was increased from 56.9% to 100%. The transgenic line was then used as the male parent to be crossed with Zhonghua 24, and the hybrid seeds showed red fluorescence as well, indicating that DsRed2 could be applied to hybrid plant identification very efficiently. DsRed2 was also expressed in hairy roots of Huayu 23 *via Agrobacterium rhizogenes*-mediated transformation, and the transgenic roots were easily selected by red fluorescence. In summary, the DsRed2 is an ideal reporter to achieve maximum screening efficiency and accuracy in peanut genetic transformation.

KEYWORDS

peanut, DsRed2, genetic transformation, agrobacterium, screening efficiency

Introduction

Peanut (*Arachis hypogaea* L.) is an important legume crop of the Fabaceae family, and is widely cultivated in tropical and subtropical regions (Sharma and Bhatnagar-Mathur, 2006). Peanut seed contains various important components with superior nutritional value, such as fat, protein, folate, tocopherol, phytosterols and polyphenolics (Jonnala et al., 2006; De Camargo and Canniatti-Brazaca, 2014). Apart from serving as an oil crop, there is a demand of peanut confectionary preparations such as desserts and snacks (Wang, 2018; Pandey et al., 2020). Therefore, various cultivars are required for the broad purpose of peanut, which can be hardly achieved through the narrow genetic base and conventional breeding methods (Liao, 2017). With the faster developments in the area of biotechnology, genetic engineering by plant transformation has shown great advantages in developing superior cultivars (Krishna et al., 2015). Hence, it is necessary to develop efficient and stable transformation systems of peanut as a foundation for its genetic engineering.

Several methods have been developed to generate highly stable transformed peanut plants, such as particle bombardment, *A. tumefaciens*-mediated transformation and pollen tube pathway system, and the former two methods have achieved relatively greater successes (Eapen and George, 1994; Deng et al., 2001; Zhou et al., 2023). For the particle bombardment method, genes are transferred to embryogenic callus in peanut, and transformed plants are regenerated from somatic embryogenesis (Krishna et al., 2015). For *A. tumefaciens*-mediated transformation, leaflet (Dolce et al., 2018), de-embryonated cotyledon (Hoa et al., 2021; Marka and Nanna, 2021), cotyledonary node (Lamboro et al., 2022) and hypocotyl (Yan et al., 2015) are the main explants, and transgenic plants are regenerated from somatic embryogenesis or organogenesis (Kundu and Gantait, 2018). However, the efficiency of transformation in peanut is still as low as 0.2%–3.3% (Rohini and Rao, 2000), which is determined by the regeneration ability of the explant, host genotype, vectors, screening efficiency and some other factors (Krishna et al., 2015). The low screening efficiency is mostly caused by insufficient selection pressure and improper identification method (Kumar et al., 2011). As a matter of fact, the selection pressure cannot be too strong to screen transformants due to the low regenerability of peanut. Since a few non-transgenic plants may escape and survive, all the regenerated plants need to be further identified (Tiwari and Tuli, 2012). Recently, the most commonly used test method is detection of the exogenous genes by PCR (Marka and Nanna, 2021). However, the accuracy is easily interfered by chimerism, *Agrobacterium*-contamination of the regenerant and aerosol. Therefore, it is significant to find an effective method to accurately distinguish transgenic and non-transgenic plants so as to reduce the workload.

Fluorescent proteins (FPs) are reporter genes widely utilized in genetic transformation of many plant species (Stewart, 2006). Due to their fluorescence characteristics, FPs can be transformed with target genes simultaneously to screen the positive transgenic events. The green fluorescent protein (GFP) isolated from jellyfish (*Aequorea Victoria*) is a frequently used reporter gene in the genetic transformation in monocot and dicot plants (Chung et al., 2000; Yang et al., 2005; Cui et al., 2020). However, its application is largely limited due to the overlapping of its spectral properties with those of several plant pigments (Yang et al., 2005). The red fluorescence protein DsRed2 is a modified form of DsRed from coral (*Discosoma* sp.), which can be easily distinguished from plant cell autofluorescence (Baird et al., 2000). DsRed2 has been successfully used as a visual reporter in numerous plant transformation studies, such as rice, soybean and cotton (Nishizawa et al., 2006; Sun et al., 2018; Toda et al., 2019). These studies have demonstrated that the utilization of FPs can contribute to rapid and accurate screening of stable transformants, which will significantly improve the screening efficiency.

In this study, DsRed2 was used as a visual reporter in peanut genetic transformation, which was delivered into peanut protoplasts by PEG and peanut plants by *A. tumefaciens*-mediated and *A. rhizogenes*-mediated transformation, respectively. The red fluorescence was monitored throughout the whole life of transgenic peanut plants. The screening efficiency with DsRed2 was also investigated. The performance of DsRed2 in transgenic peanut plants was evaluated to assess its ability to serve as a selective marker for peanut biotechnology.

Materials and methods

Plant materials

Three peanut cultivars Zhonghua 12 Zhonghua 24 and Huayu 23 were used in this study. Zhonghua 12 (var. *vulgaris*) and Zhonghua 24 (var. *hypogaea*) are developed by Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China in 2006 and 2015, respectively. Huayu 23 (var. *hypogaea*) is developed by Shandong Peanut Research Institute, Shandong Academy of Agricultural Sciences, Qingdao, China in 2004.

Vector construction

A cassette comprising a Nos promoter and 3' UTR flanking Basta selection marker gene was amplified by PCR from pBinGlyBar1 vector using the following primers with added *Asc*I restriction sites: 5'-CATGGGCGCGCCGCACGCTGCCGCAAGCAC-3' and 5'-CATGGGCGCGCCCGCGCCGATCTAGTAACA-3' (the added restriction sites are underlined). Then the *Asc*I digested fragment was inserted into the binary vector pBinGlyRed2 (Huai et al., 2020) containing a 35S promoter-driven DsRed2 gene to generate pBinBarRed (Figure S1).

Protoplast isolation and transfection

Protoplasts were isolated from leaves of 14 days old aseptic seedlings of peanut cultivar Zhonghua 12, which were obtained from seeds sterilized with 0.1% (w/v) $HgCl_2$ and cultured in dark.

The leaves were cut into 0.5 mm strips, and then dispersed into 30 mL of enzyme solution (2% cellulose R-10, 1% macerozyme R-10, 0.2% pectolase Y-23, 0.1% BSA, 0.6 M mannitol, 10 mM MES (pH 5.7) and 15 mM $CaCl_2$) with gently shaking at 40 rpm for 12 hours. After digestion, cells were filtered with a 70 μm nylon meshes, and then washed twice by 30 mL W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl and 2 mM MES, pH5.8). Protoplasts were resuspended in 2 mL W5 solution and stored on ice for 30 mins. Lastly, protoplasts were centrifuged at 100 g for 2 mins, and resuspended in 2 mL MMG solution (0.6 M mannitol, 15 mM $CaCl_2$ and 4 mM MES). Cell concentration was measured using a hemocytometer and a light microscope.

Totally $2x10^5$ isolated single protoplasts were suspended in 200 μL MMG solution, and stored on ice for 30 mins. Approximately 20 μg plasmid DNA and 220 μL PEG solution (40% PEG4000, 0.6 M mannitol and 100 mM $CaCl_2$) was added, mixed gently and stored at room temperature for 20 mins. After incubation, 1 mL W5 solution was added and stored on ice for 5 mins. Then cells were centrifuged at 100 g for 2 mins, and washed twice by 300 μL W5 solution. The protoplast pellets were resuspended with 1.5 mL W5 solution, and incubated in the dark at room temperature for 48-72 hours. The protoplast pellets were harvested for further analysis.

A. tumefaciens-mediated transformation

The vector pBinBarRed was introduced into *Agrobacterium tumefaciens* strain GV3101 by electro-transformation. A single transformed colony was inoculated into a flask (50 ml) containing sterile 10 ml LB liquid medium supplemented with 50 mg/L kanamycin. The flask was incubated at 28°C for 24 h in a shaker incubator set at 180 rpm until the OD_{600} reached 0.6-0.8. Bacterial suspension was pelleted by centrifugation for 10 min at 250 g and resuspended the cells in the *Agrobacterium* infection medium (2.2g/L MS-B5 and 20 g/L sucrose, OD_{600} = 0.6-0.8).

The *A. tumefaciens*-mediated transformation was performed as described by Sharma and Bhatnagar-Mathur (2006) with some modifications. Mature seeds from the peanut cultivar Zhonghua 12 were surface sterilized by rinsing in 75% ethanol for 1 min followed by treatment with 0.1% (w/v) $HgCl_2$ for 4 min and then washed 4-6 times with sterile-distilled water and soaked in sterile water overnight. After removing the seed coat, the embryo was removed and each cotyledon was cut into vertical halves which were used as explants. Explants were immersed in the *Agrobacterium* infection medium for 2-5 min, and then transferred onto the Co-cultivation Medium (4.4 g/L MS-B5, 20 g/L sucrose, 4 mg/L BAP, 1 mg/L 2,4-D and 8 g/L agar, pH=5.8) maintained in dark for 72 hours at 26°C (Figure 1A). Then explants were transferred onto the Shoot Induction Medium (4.4 g/L MS-B5, 20 g/L sucrose, 4 mg/L BAP, 1 mg/L 2,4-D, 300 mg/L Timentin, 1 mg/L Basta and 8 g/L agar, pH=5.8) and kept at 26°C under 16 h day/8 h dark for 14 days. The explants were transferred onto a fresh medium every 14 days until the shoots come out (Figures 1B–D). The shoots were transferred onto the Shoot Elongation Medium (4.4 g/L MS-B5, 20 g/L sucrose, 2 mg/L BAP, 300 mg/L Timentin and 8 g/L agar, pH=5.8) and kept at 26 °C under 16 h day/8 h dark for 14 days. The shoots were transferred onto a fresh medium every 14 days until the shoots grew to 3-4 cm high (Figures 1E, F). Subsequently, the elongated shoots were transferred onto the Root Induction Medium (4.4 g/L MS-B5, 20 g/L sucrose and 8 g/L agar, pH=5.8) until the roots grew to 4-5 cm long (Figures 1G, H). The plants were transplanted into autoclaved sand-soil (1:1) mixture in plastic pots and maintained in a growth cabinet at 26 °C under 16 h day/8 h dark until the seeds mature (Figure 1I).

A. rhizogenes-mediated transformation

The vector pBinBarRed was introduced into *Agrobacterium rhizogenes* strain K599 by electro-transformation. A single transformed colony was grown in 10 mL LB liquid medium containing kanamycin at 50 mg/L and incubated overnight at 28 °C with shaking at 180 rpm until the OD_{600} reached 0.6–0.8. The bacterial cells were collected by centrifugation for 10 min at 250 g and resuspended in the *A. rhizogenes* solutions (2.2 g/L MS, 20 g/L sucrose and 100 μM AS, OD_{600} = 0.6-0.8).

Mature seeds from the peanut cultivar Huayu 23 were surface sterilized and germinated on ½MS medium (Figures 2A–C). After 1 week, the radicle and hypocotyl were cut from each seedling and the remaining portion was used as explant. Explants were dipped in the *A. rhizogenes* solutions and incubated for 2-5 min, and then transferred onto the co-cultivation medium (4.4 g/L MS, 20 g/L sucrose, 50 μM AS and 8 g/L agar, pH=5.8) maintained in dark at 26°C for 3 days (Figure 2D). Subsequently, explants were transferred onto the hairy root induction medium (4.4 g/L MS, 20 g/L sucrose, 300 mg/L Timentin and 8 g/L agar, pH=5.8) and kept at 26 °C under 16 h day/8 h dark until the root grow well (Figures 2E, F).

Fluorescence observation

The fluorescence in protoplast cells was examined under a laser-scanning confocal microscope (Olympus FV 10-ASW). The fluorescence in callus and samples form transgenic plants were observed using a green-light hand-held lamp (LUYOR, China), with a red camera filter lens.

DNA extraction and PCR analysis

Total genomic DNA was isolated from young leaves of transgenic plantlets and wild-type peanut plants using a EasyScript Plant Genomic DNA Kit (Transgen, China). To detect the presence of *DsRed2* gene, the following primers were used for PCR amplification: 5'- TTCAAGGTGCGCATGGAG-3' and 5'-CGTTGTGGGAGGTGATGT-3'. The amplification cycle consisted of denaturation at 94 °C for 1 min, primer annealing at 58 °C for 1 min, and primer extension at 72 °C for 1 min. After 30 repeats of the thermal cycle and final extension 72 °C for 10 min, amplification products were analyzed on 1% agarose gels. The putative PCR product was 577 bp.

10.3389/fpls.2023.1123644

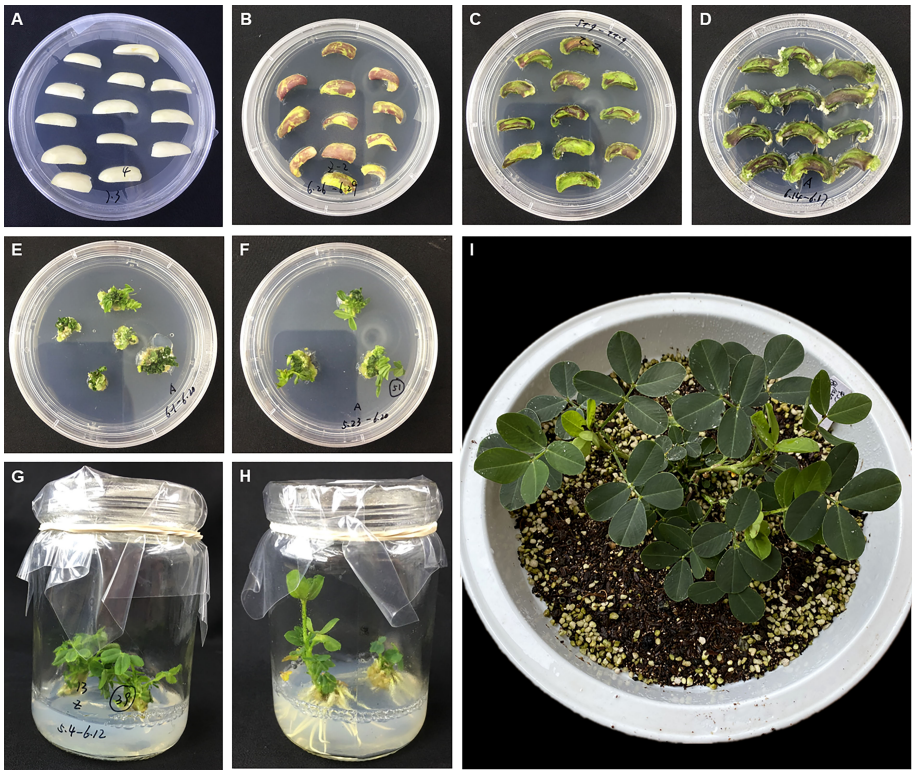


FIGURE 1
Agrobacterium tumefaciens-mediated peanut transformation. **(A)** Cotyledon explants on the Co-cultivation Medium. **(B–D)** Explants on the Shoot Induction Medium after one week **(B)**, two weeks **(C)** and three weeks **(D)**. (**E, F**). Induced adventitious shoots on the Shoot Elongation Medium after two weeks **(E)** and four weeks **(F)**. **(G, H)** Induced adventitious roots on the Root Induction Medium after two weeks **(G)** and four weeks **(H)**. **(I)** The plantlet was transplanted into soil and maintained in a growth cabinet.

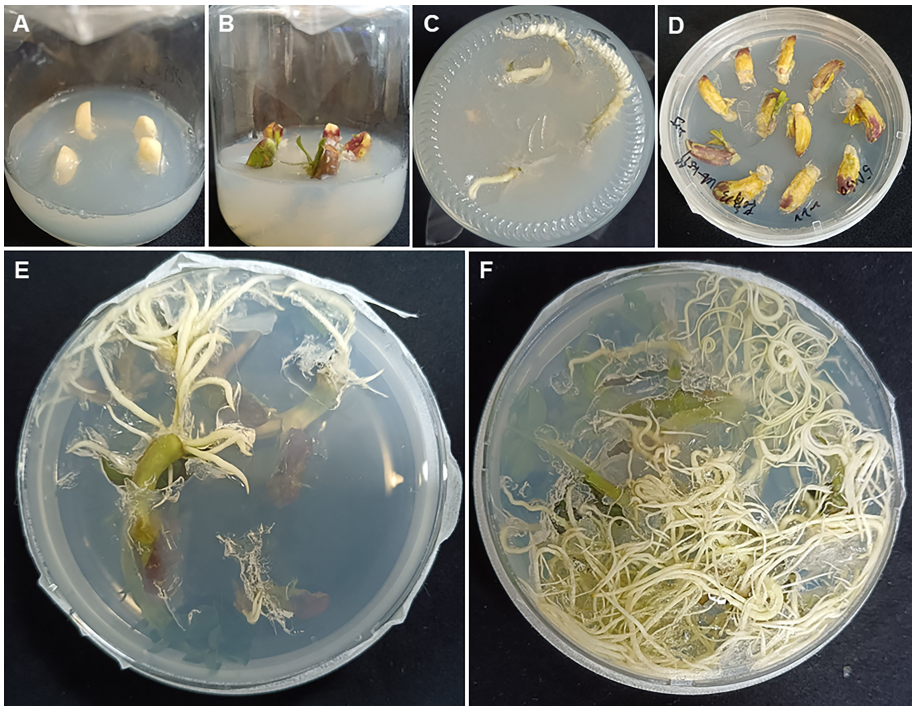


FIGURE 2
Agrobacterium rhizogenes-mediated peanut transformation. **(A)** Surface sterilized seeds on ½MS medium. **(B, C)** One-week old seedling for inoculation. **(D)** Explants on the co-cultivation medium. **(E, F)** Induced hairy roots appearing two weeks **(E)** and four weeks **(F)** after inoculation.

RNA extraction and real-time PCR analysis

Total RNA was extracted from young leaves of transgenic plantlets and wild-type peanut plants using TRIzol reagent (Sigma, USA). Reverse transcription was implemented using SuperScript IV First-Strand Synthesis System (Invitrogen, USA). Real time PCRs were performed on a Bio-Rad CFX96 Real-Time system using SYBR Green as fluorescent dye. The peanut *actin* gene was used as internal control using primers: 5'- TAAGAACAATGTTGCCATACAGA-3' and 5'-GTTGCCTTGGATTATGAGC-3'. The primers for *DsRed2* gene were: 5'-GTACGGCTCCAAGGTGTACG-3' and 5'-TAGATGAAGCAGCCGTCCTG-3'.

Crossing and hybrid identification

A high oleate peanut cultivar Zhonghua 24 was pollinated with the pollens from transgenic Line 1 (T_1), which were developed from normal oleate cultivar Zhonghua 12. As the difference between the oleic acid contents in parents were caused by mutations in *AhFad2* genes (Chu et al., 2009; Pandey et al., 2014), primers for *AhFad2* genes were used for hybrid identification: 5'- CACTAAGATTGAAGCTC-3' and 5'-CACTAAGATTGAAGCTC -3'. A 500 bp fragment was amplified by PCR and sequenced by Sanger sequencing.

Results

Evaluation of DsRed2 in peanut protoplast transformation

Firstly, the *DsRed2* gene was expressed in protoplasts isolated from peanut leaves. After transformation, the protoplasts were centrifuged to the bottom of tubes (Figure 3A). Under green light, the transformed protoplasts R-1 and R-2 showed bright red fluorescence compared with CK (Figure 3B). Under a laser confocal microscope, red fluorescence was detected in R-1 and R-2, but not in CK (Figure 3C). These results indicated that DsRed2 could be used as a reporter in peanut protoplast transformation.

Evaluation of DsRed2 in *A. tumefaciens*-mediated peanut transformation

The *DsRed2* gene was constitutively expressed in peanut *via A. tumefaciens*-mediated transformation. Under green light, red fluorescence could be observed at very early stage of callus formation (Figure 4B), and became increasingly pronounced with callus age. Upon the emergence of shoots, the transgenic shoots were selected by red fluorescence (Figure 4D and Figure S2B). The shoots showing red fluorescence developed into plantlets with red fluorescence (Figure 4H). The whole plant exhibited bright red fluorescence under green light at both vegetative and reproductive stages, including the root (Figure 4F), leaf (Figure 4K), flower (Figure 4M), pod shell (Figure 4O) and seed (Figure 4Q). Under white light, after removal of the seed coat, the color of embryo and cotyledon was red in transgenic seeds, but was white in CK (Figure 4R), allowing very easy discrimination of the transgenic seeds with naked eyes. The T_1 progeny of DsRed2 peanut exhibited the same morphological characteristics as T_0 generation, indicating that the fluorescence can be inherited by subsequent generations. Hence, DsRed2 can serve as a stable reporter in *A. tumefaciens*-mediated peanut transformation.

Molecular analysis of putative transgenic plants

A total of 52 T_0 transgenic plantlets were obtained, and 15 of which were tested by PCR. As shown in Figure 5A, all the 15 lines harbored the DsRed2 gene. Compared with the transgenic lines without the DsRed2 reporter, the positive rate increased from 56.9% (Figure S3) to 100% (Figure 5A). The expression levels of *DsRed2* were detected in leaves of the 15 transgenic lines by qRT-PCR. As expected, the expression of *DsRed2* was not detectable in the non-transformed control, while was detected in transgenic lines (Figure 5B). Hence, the application of DsRed2 in peanut transformation can greatly increase the screening efficiency and improve the positive rate.

Evaluation of DsRed2 in hybrid identification

The transgenic Line 1 with DsRed2 was used as the male parent (with normal oleate trait) to be crossed with the high-oleate peanut cultivar Zhonghua 24, and 28 F_1 hybrid seeds were obtained. The seed testa color was pink in the female parent Zhonghua 24, while red in the male parent transgenic Line 1. The seed testa color of hybrid was pink, suggesting that the F_1 seeds were obtained from the female parent instead of mixing with seeds from the male parent (Figure 6A). The color of embryo and cotyledon was white in the female parent, red in the male parent, while lightly red in the hybrid seeds, indicating that the DsRed2 gene was transferred into hybrid plants (Figure 6C). Furthermore, red fluorescence was detected in both hybrid seeds and male parent seeds, but not in female parent seeds under green light (Figures 6B and D). Sanger sequencing results revealed that the genotypes of *AhFAD2* genes controlling the oleic content in all the 28 red seeds were heterozygous, indicating that they were true hybrids (Figure S4). All these results suggested that DsRed2 could be used as a reliable reporter in peanut hybrid identification.

Evaluation of DsRed2 in *A. rhizogenes*-mediated peanut transformation

The selection effect of DsRed2 in *A. rhizogenes*-mediated peanut transformation was also evaluated (Figure 7a). As shown in Figure 7B, red fluorescence can only be observed in induced hairy roots but not

10.3389/fpls.2023.1123644

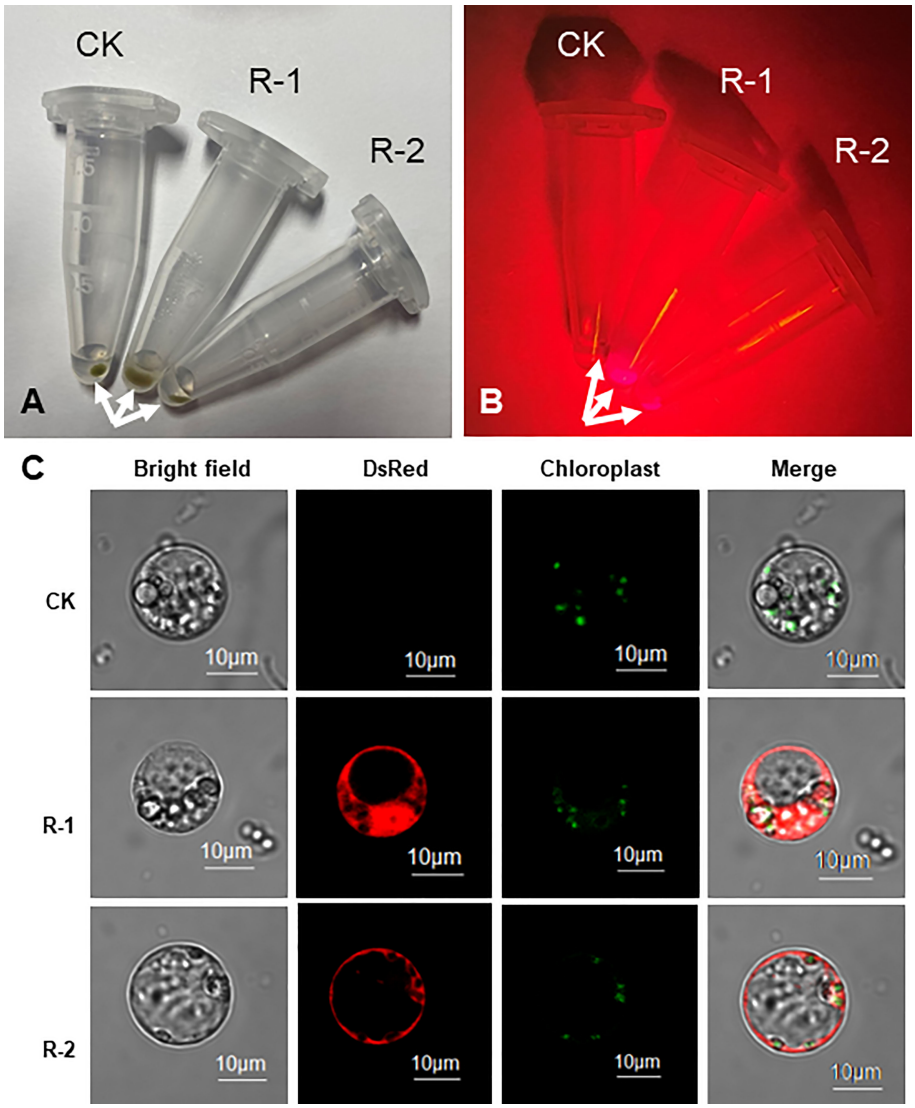


FIGURE 3
DsRed2 gene expression in peanut protoplast. **(A, B)** Centrifuged peanut protoplasts at the bottom of Eppendorf tubes under white light **(A)** and green light **(B)**. Arrows indicate the centrifuged protoplasts. **(C)** Characteristic features of organelles in peanut protoplasts. CK, peanut protoplast transformed with pBinGlyBar1; R-1 and R-2, peanut protoplast transformed with pBinBarRed.

in other parts under green light. Twenty-one plantlets with transgenes were selected by red fluorescence, and verified by PCR. The positive rate of transgenic hairy root was also reached to 100% (Figure 7C). These results also demonstrated that DsRed2 is also an ideal reporter in *A. rhizogenes*-mediated peanut transformation.

Discussion

A highly efficient transformation system is critical for the improvement of crops including peanut. Although several peanut transformation systems have been established, their transformation efficiency and reproducibility are still largely inadequate. Hence, it is necessary to improve the transformation efficiency by optimizing the transformation parameters such as plant genotype, inoculation, co-culture conditions and selective reporters (Krishna et al., 2015). GFP was used as a reporter in peanut genetic transformation by particle bombardment. Green fluorescence was observed in somatic embryo, root and leave, but green fluorescence in shoot tissues was confounded with fluorescence from chlorophyll (Joshi et al., 2005). DsRed2 is a red fluorescent protein from corals of the *Discosoma* genus, which has been used successfully in research on animals, fungi and plants (Baird et al., 2000). In this study, DsRed2 was used as a visible selective reporter and its performance in peanut genetic transformation was evaluated. DsRed2 was successfully expressed in peanut protoplasts and plants (Figures 3 and 4), suggesting that it

10.3389/fpls.2023.1123644

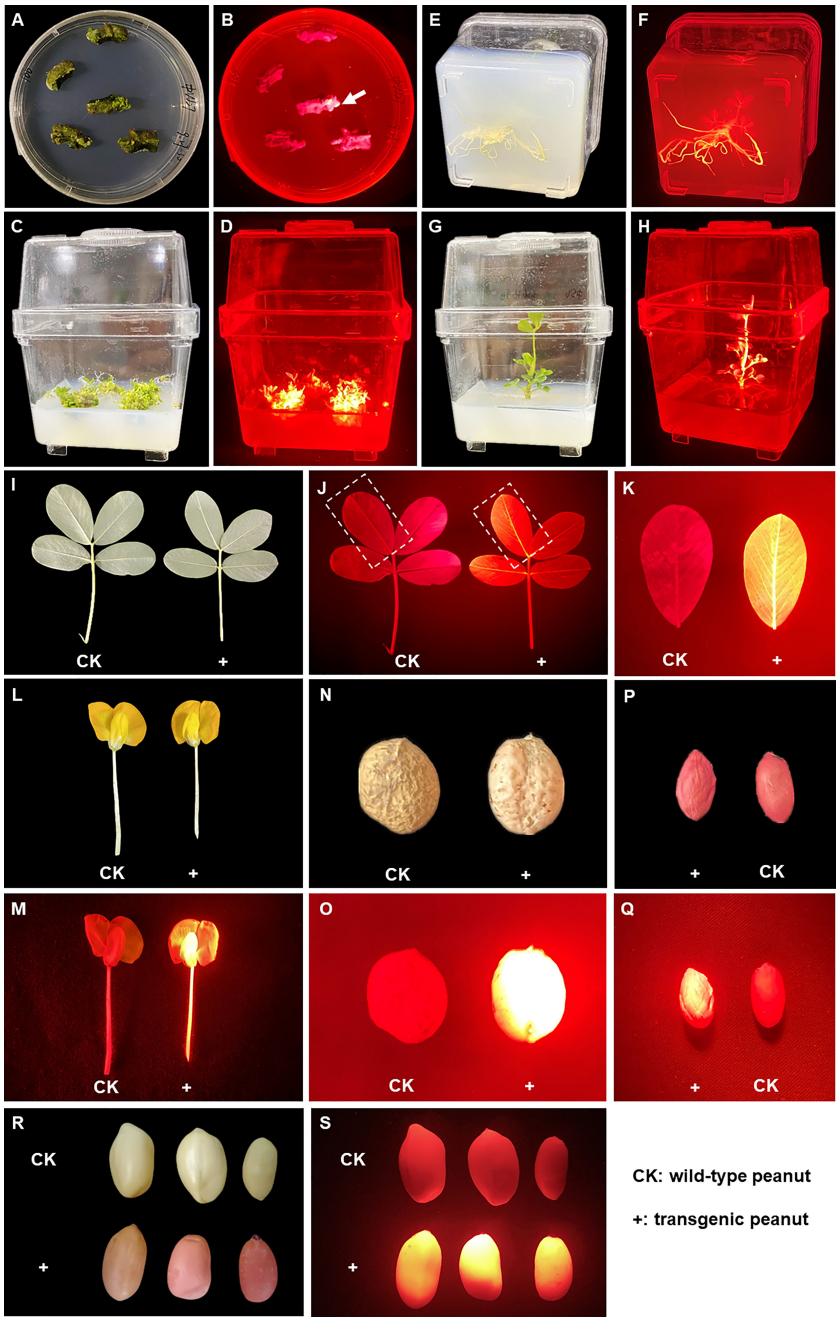


FIGURE 4
DsRed2 gene expression in callus, different tissues and organs *via A. tumefaciens*-mediated peanut transformation. **(A, B)** *DsRed2* expression in peanut callus under white light **(A)** and green light **(B)**. The arrow indicates the transformed callus with red fluorescent. **(C, D)** *DsRed2* expression in induced shoots under white light **(C)** and green light **(D)**. **(E, F)** *DsRed2* expression in induced roots under white light **(E)** and green light **(F)**. **(G, H)** *DsRed2* expression in transgenic plantlet under white light **(G)** and green light **(H)**. **(I–K)** *DsRed2* expression in mature leave under white light **(I)** and green light **(J, K)**. **(L, M)** *DsRed2* expression in flower under white light **(L)** and green light **(M)**. **(N, O)** *DsRed2* expression in pod under white light **(N)** and green light **(O)**. **(P, Q)** *DsRed2* expression in seed under white light **(P)** and green light **(Q)**. **(R, S)** *DsRed2* expression in seed without testa under white light **(R)** and green light **(S)**.

can be steadily expressed and inherited in peanut. Moreover, a red color could be clearly observed with naked eyes in embryo and cotyledon without the help of any instrument (Figures 4R and 6C), which can greatly facilitate the screening of transgenic seeds in the progeny and hybrids. Hence, DsRed2 can serve as an effective visual reporter gene for genetic transformation of peanut.

Compared with low identification accuracy of transgenic peanut plant is an important factor limiting the screening efficiency in

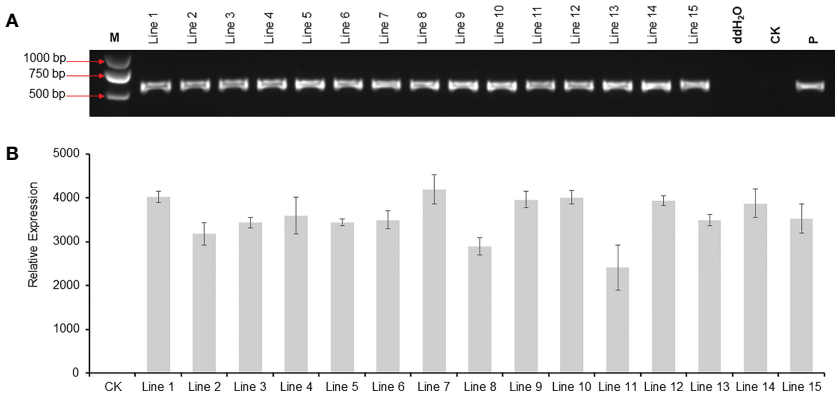


FIGURE 5
PCR and RT-PCR analysis of *DsRed2* gene in transgenic plants. **(A)** PCR amplification of *DsRed2* in transgenic plantlets. **(B)** The expression level of *DsRed2* in transgenic plantlets. Lines 1-15, transgenic lines with red fluorescent; CK, wild-type plant; P, plasmid of pBinBarRed.

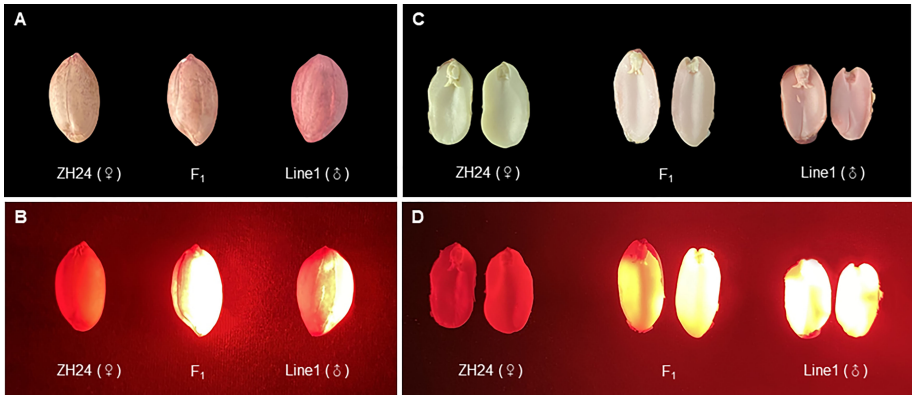


FIGURE 6
F_1 hybrids generated from a cross between *DsRed2* transgenic plants and Zhonghua 24 cultivar. **(A, B)** Seeds of F_1 hybrid, the female parent Zhonghua 24 and the male parent *DsRed2* transgenic Line1 under white light **(A)** and green light **(B)**. **(C, D)** The embryo and cotyledon of F_1 hybrid, the female parent Zhonghua 24 and the male parent *DsRed2* transgenic Line1 under white light **(C)** and green light **(D)**.

addition to resources wastage (Mallikarjuna and Varshney, 2014). In previous studies, the highest positive rate of peanut transformation was 90% (Krishna et al., 2015). In this study, application of DsRed2 significantly increased the screening efficiency and improved the positive rate from 56.9% to 100% (Figures 5 and S3), which is the highest score in peanut transformation. In previous studies, the most common identification method is detection of the exogenous genes by PCR and should be performed after acquired regenerated plants (Marka and Nanna, 2021). In this study, as red fluorescence could be observed at a very early stage of callus formation (Figure 4B), the screening could be started very early as well, reducing much labor and resource consumption. Furthermore, identification with PCR requires a special instrument and reagents, but screening with DsRed2 requires only a lamp and a filter, which is much cheaper and easy to operate. Therefore, screening with DsRed2 in peanut transformation can promote the efficiency and save resources to some extent.

In the present study, the transgenic peanut plants were generated *via* organogenesis in the *A. tumefaciens*-mediated transformation system. Red fluorescence was detected in all tissues of transgenic plants at both vegetative and reproductive stages, including the root (Figure 4F), leaf (Figure 4K), flower (Figure 4M), pod shell (Figure 4O) and seed (Figure 4Q). Interestingly, no chimera was detected in the transgenic plants. Relative to somatic embryogenesis, organogenesis protocol is easier

FIGURE 7
DsRed2 gene expression in hairy roots *via A. rhizogenes*-mediated peanut transformation.**(A, B)** *DsRed2* expression in hairy roots under white light **(A)** and green light **(B)**. **(C)** PCR amplification of *DsRed2* in transgenic hairy roots. HR-1 to HR-21, transgenic plantlets with red hairy roots; CK, wild-type plant; P, plasmid of pBinBarRed.

to handle and more time-saving, but may lead to the generation of chimera in some researches (Tiwari and Tuli, 2012). The red fluorescence screening can effectively avoid the occurrence of chimera. Therefore, application of the DsRed2 reporter in organogenesis protocol can greatly simplify the operation and increase the efficiency of peanut transformation.

Genome-editing technologies have revolutionized plant research and exhibit great potential in the improvement of crops (Kausch et al., 2019). In this study, an efficiently and stable transformation system was successfully established for peanut, which may be widely used in metabolic engineering and genome-editing of peanut and greatly facilitate future peanut improvement.

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

DH, YL and BL conceived and designed the experiments. HJ and DB supplied the peanut lines. JW, XX, MH, CZ, NL, LH, LY, YC, XW, YK and ZW performed the experiments. DH, JW, and XX analyzed the data. DH and JW wrote the manuscript. DH, PB, MKP, YL, RKV and BL contributed in data interpretation and revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Knowledge Innovation Program of Wuhan -Basic Research (2022020801010291), the Key R&D Program of China (2022YFD1200400), the Key R&D Program of Guangdong Province (2022B0202060004) and Innovation Program of the Chinese Academy of Agricultural Sciences (2022-2060299-089-031). The funders had no role in experiment design, data analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank Dr. Edgar B. Cahoon (Center for Plant Science Innovation and Department of Biochemistry, University of Nebraska-Lincoln) for supplies of vectors.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1123644/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Construct of pBinBarRed. *DsRed2* was inserted into pBinBarRed vector. Constitutive and seed-specific promoters, 3' UTR sequences and arrangements of cassettes are also shown.

SUPPLEMENTARY FIGURE 2
Selection of transgenic shoots using *DsRed2* gene. **(A)** Transgenic shoots and non-transgenic shoots under the white light. **(B)** Transgenic shoots and non-transgenic shoots under the green light.

SUPPLEMENTARY FIGURE 3
PCR analysis of *Bar* and *Luc* genes in transgenic plants without DsRed2 selection reporter. **(A)** PCR amplification of *Bar* gene in transgenic plantlets. **(B)** PCR amplification of *Luc* gene in transgenic plantlets. +, transgenic line which harboring both of *Bar* and *Luc* genes; -, non-transgenic line which possessing one or none of *Bar* and *Luc* genes. Among 65 putative plants, 37 plants were identified as transgenic plants, so the positive efficiency was 65.9%. 1-65, putative transgenic plants (T_0); CK, wild-type plant; P, plasmid of pBinGlyBar1.

SUPPLEMENTARY FIGURE 4
Sequence from a PCR product of *AhFAD2* genes in hybrids (F_1). The mutant allele of *AhFAD2A* had a 1-bp substitution (G:C→A:T) at position 448 after the start codon; the mutant allele of *AhFAD2B* had a 1-bp insertion (A:T) at position 442 after the start codon. The true hybrid (F_1) was identified by containing double peaks (G/A) at position 442 bp (*AhFAD2B*) and triple peaks (G/A/C) at position 448 bp (*AhFAD2A*) after the start codon. A, green; T, red; C, blue; G, black.

References

Baird, G., Zacharias, S., and David, A. (2000). Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. United States America* 97 (22), 11984–11989. doi: 10.1073/pnas.97.22.11984

Chung, B. C., Kim, J. K., Nahm, B. H., and Lee, C. H. (2000). In planta visual monitoring of green fluorescent protein in transgenic rice plants. *Molecules Cells* 10, 411–414.

Chu, Y., Holbrook, C. C., and Ozias-Akins, P. (2009). Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. *Crop Sci.* 49 (6), 2029–2036. doi: 10.2135/cropsci2009.01.0021

Cui, M. L., Liu, C., Piao, C. L., and Liu, C. L. (2020). A stable *Agrobacterium rhizogenes*-mediated transformation of cotton (*Gossypium hirsutum* l.) and plant regeneration from transformed hairy root *via* embryogenesis. *Front. Plant Sci.* 11, 604255. doi: 10.3389/fpls.2020.604255

De Camargo, A. C., and Canniatti-Brazaca, S. G. (2014). Peanuts as a source of protein, unsaturated fatty acids, tocopherol and polyphenols. In *Peanuts: Production, Nutritional Content and Health Implications*; R. W. Cook, Ed.; Nova Science Publishers: New York, NY, USA. 69–80.

Deng, X. Y., Wei, Z. M., and An, H. L. (2001). Transgenic peanut plants obtained by particle bombardment *via* somatic embryogenesis regeneration system. *Cell Res.* 11 (2), 156–160. doi: 10.1038/sj.cr.7290081

Dolce, N. R., Faloci, M. M., and Gonzalez, A. M. (2018). *In vitro* plant regeneration and cryopreservation of *Arachis glabrata* (Fabaceae) using leaflet explants. *In Vitro Cell. Dev. Biology-Plant* 54 (2), 133–144. doi: 10.1007/s11627-017-9865-y

Eapen, S., and George, L. (1994). *Agrobacterium tumefaciens* mediated gene transfer in peanut (*Arachis hypogaea* l.). *Plant Cell Rep.* 13, 582–586. doi: 10.1007/BF00234516

Hoa, P. T. B., Tue, N. H., Trang, P. T. Q., Hang, L. T., Tien, N. Q. D., and Loc, N. H. (2021). An efficient protocol for *in vitro* regeneration of peanut (*Arachis hypogaea* l.) cultivar L14. *Bioscience J.* 37, 1981–3163, e37019. doi: 10.14393/BJ-v37n0a2021-56949

Huai, D., Xue, X., Li, Y., Wang, P., Li, J., Yan, L., et al. (2020). Genome-wide identification of peanut *KCS* genes reveals that *AhKCS1* and *AhKCS28* are involved in regulating VLCFA contents in seeds. *Front. Plant Sci.* 11, 406. doi: 10.3389/fpls.2020.00406

Jonnala, R. S., Dunford, N. T., and Chenault, K. (2006). Tocopherol, phytosterol and phospholipid compositions of genetically modified peanut varieties. *J. Sci. Food Agric.* 86 (3), 473–476. doi: 10.1002/jsfa.2372

Joshi, M., Niu, C., Fleming, G., Hazra, S., Chu, Y., Nairn, C. J., et al. (2005). Use of green fluorescent protein as a non-destructive marker for peanut genetic transformation. *In Vitro Cell. Dev. Biology-Plant* 41, 437–445. doi: 10.1079/IVP2005676

Kausch, A. P., Nelson-Vasilchik, K., Hague, J., Mookkan, M., Quemada, H., Dellaporta, S., et al. (2019). Edit at will: Genotype independent plant transformation in the era of advanced genomics and genome editing. *Plant Sci.* 281, 186–205. doi: 10.1016/j.plantsci.2019.01.006

Krishna, G., Singh, B. K., Kim, E. K., and Morya, V. K. (2015). Progress in genetic engineering of peanut (*Arachis hypogaea* l.)-a review. *Plant Biotechnol. J.* 13 (2), 147–162. doi: 10.1111/pbi.12339

Kumar, V., Campbell, L., and Rathore, K. S. (2011). Rapid recovery-and characterization of transformants following *Agrobacterium*-mediated T-DNA transfer to sorghum. *Plant Cell Tissue Organ Culture* 104 (2), 137–146. doi: 10.1007/s11240-010-9809-2

Kundu, S., and Gantait, S. (2018). Fundamental facets of somatic embryogenesis and its applications for advancement of peanut biotechnology. *Biotechnologies Crop Improvement* 1, 267–298. doi: 10.1007/978-3-319-78283-6_8

Lamboro, A., Han, X., Yang, S., Li, X., Yao, D., Song, B., et al. (2022). Combination of 6-benzylaminopurine and thidiazuron promotes highly efficient shoot regeneration from cotyledonary node of mature peanut (*Arachis hypogaea* l.) cultivars. *Phyton-International J. Exp. Bot.* 91 (12), 2619–2631. doi: 10.32604/phyton.2022.021404

Liao, B. S. (2017). "Germplasm characterization and trait discovery in peanut," in *The peanut genome* (Cham: Springer), 53–68.

Mallikarjuna, N., and Varshney, R. K. (2014). "Genetics, genomics and breeding of peanut: An introduction," in *Genetics, genomics and breeding of crop plants* (CRC Press), 1–12.

Marka, R., and Nanna, R. S. (2021). Expression of *Tcchitinase-i* gene in transgenic peanut (*Arachis hypogaea* l.) confers enhanced resistance against leaf spot and rust diseases. *Plant Growth Regul.* 93 (1), 53–63. doi: 10.1007/s10725-020-00663-8

Nishizawa, K., Kita, Y., Kitayama, M., and Ishimoto, M. (2006). A red fluorescent protein, DsRed2, as a visual reporter for transient expression and stable transformation in soybean. *Plant Cell Rep.* 25 (12), 1355–1361. doi: 10.1007/s00299-006-0210-x

Pandey, M. K., Ming, L. W., Qiao, L., Feng, S., Khera, P., Hui, W., et al. (2014). Identification of QTLs associated with oil content and mapping genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* l.). *BMC Genet.* 15 (1), 1–14. doi: 10.1186/s12863-014-0133-4

Pandey, M. K., Pandey, A. K., Kumar, R., Nwosu, C. V., Guo, B., Wright, G. C., et al. (2020). Translational genomics for achieving higher genetic gains in groundnut. *Theor. Appl. Genet.* 133 (5), 1679–1702. doi: 10.1007/s00122-020-03592-2

Rohini, V. K., and Rao, K. S. (2000). Transformation of peanut (*Arachis hypogaea* l.): a non-tissue culture based approach for generating transgenic plants. *Plant Sci.* 150 (1), 41–49. doi: 10.1016/S0168-9452(99)00160-0

Sharma, K. K., and Bhatnagar-Mathur, P. (2006). Peanut (*Arachis hypogaea* l.). *Agrobacterium Protoc.* 343, 347–358. doi: 10.1385/1-59745-130-4:347

Stewart, C. N.Jr (2006). Go with the glow: fluorescent proteins to light transgenic organisms. *Trends Biotechnol.* 24 (4), 155–162. doi: 10.1016/j.tibtech.2006.02.002

Sun, L., Alariqi, M., Zhu, Y., Li, J., Li, Z., Wang, Q., et al. (2018). Red fluorescent protein (DsRed2), an ideal reporter for cotton genetic transformation and molecular breeding. *Crop J.* 6 (4), 366–376. doi: 10.1016/j.cj.2018.05.002

Tiwari, S., and Tuli, R. (2012). Optimization of factors for efficient recovery of transgenic peanut (*Arachis hypogaea* l.). *Plant Cell Tissue Organ Culture* 109 (1), 111–121. doi: 10.1007/s11240-011-0079-4

Toda, E., Koiso, N., Takebayashi, A., Ichikawa, M., Kiba, T., Osakabe, K., et al. (2019). An efficient DNA- and selectable-marker-free genome-editing system using zygotes in rice. *Nat. Plants* 5 (4), 363–368. doi: 10.1038/s41477-019-0386-z

Wang, Q. (2018). "Peanut processing characteristics and quality evaluation," in *Peanut processing characteristics and quality evaluation* (Singapore: Springer), 97–98.

Yan, M., Huang, B., Miao, L., Yi, M., Li, W., Wei, G., et al. (2015). Induction *in vitro* and plant regeneration of embryo axe of peanut (*Arachis hypogaea* l.). *Chin. Agric. Sci. Bull.* 31 (21), 149–156. doi: 10.11924/j.issn.1000-6850.casb15010168

Yang, J., Yang, X. L., Cheng, Z., Shan, L. I., and Zhu, Y. G. (2005) in *International Agro-Biotechnology Conference* (China: Tongji University Shanghai 200092).

Zhou, M., Luo, J., Xiao, D., Wang, A., He, L., and Zhan, J. (2023). An efficient method for the production of transgenic peanut plants by pollen tube transformation mediated by agrobacterium tumefaciens. *Plant Cell Tissue Organ Culture* 152 (1), 207–214. doi: 10.1007/s11240-022-02388-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fpls.2023.1145624

OPEN ACCESS

EDITED BY
Yihua Wang,
Nanjing Agricultural University, China

REVIEWED BY
Muhammad Aamir Manzoor,
Anhui Agricultural University, China
Garima Kushwaha,
Guardanthealth, Inc., United States

*CORRESPONDENCE
Weijian Zhuang
weijianz@fafu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 16 January 2023
ACCEPTED 10 March 2023
PUBLISHED 31 March 2023

CITATION
Cai T, Sharif Y, Zhuang Y, Yang Q, Chen X, Chen K, Chen Y, Gao M, Dang H, Pan Y, Raza A, Zhang C, Chen H and Zhuang W (2023) *In-silico* identification and characterization of *O-methyltransferase* gene family in peanut (*Arachis hypogaea* L.) reveals their putative roles in development and stress tolerance. *Front. Plant Sci.* 14:1145624. doi: 10.3389/fpls.2023.1145624

In-silico identification and characterization of *O-methyltransferase* gene family in peanut (*Arachis hypogaea* L.) reveals their putative roles in development and stress tolerance

Tiecheng Cai[1†], Yasir Sharif[1†], Yuhui Zhuang[2], Qiang Yang[1], Xiangyu Chen[1,3], Kun Chen[4], Yuting Chen[1], Meijia Gao[4], Hao Dang[1], Yijing Pan[1], Ali Raza[1], Chong Zhang[1], Hua Chen[1] and Weijian Zhuang[1*]

[1]Center of Legume Plant Genetics and System Biology, College of Agronomy, Fujian Agriculture and Forestry University (FAFU), Fuzhou, Fujian, China, [2]College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China, [3]Crops Research Institute, Fujian Academy of Agricultural Science, Fuzhou, Fujian, China, [4]College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

Cultivated peanut (*Arachis hypogaea*) is a leading protein and oil-providing crop and food source in many countries. At the same time, it is affected by a number of biotic and abiotic stresses. O-methyltransferases (*OMTs*) play important roles in secondary metabolism, biotic and abiotic stress tolerance. However, the *OMT* genes have not been comprehensively analyzed in peanut. In this study, we performed a genome-wide investigation of *A. hypogaea OMT* genes (*AhOMTs*). Gene structure, motifs distribution, phylogenetic history, genome collinearity and duplication of *AhOMTs* were studied in detail. Promoter *cis*-elements, protein-protein interactions, and micro-RNAs targeting *AhOMTs* were also predicted. We also comprehensively studied their expression in different tissues and under different stresses. We identified 116 *OMT* genes in the genome of cultivated peanut. Phylogenetically, *AhOMTs* were divided into three groups. Tandem and segmental duplication events played a role in the evolution of *AhOMTs*, and purifying selection pressure drove the duplication process. *AhOMT* promoters were enriched in several key *cis*-elements involved in growth and development, hormones, light, and defense-related activities. Micro-RNAs from 12 different families targeted 35 *AhOMTs*. GO enrichment analysis indicated that *AhOMTs* are highly enriched in transferase and catalytic activities, cellular metabolic and biosynthesis processes. Transcriptome datasets revealed that *AhOMTs* possessed varying expression levels in different tissues and under hormones, water, and temperature stress. Expression profiling based on qRT-PCR results also supported the transcriptome results. This study provides

the theoretical basis for further work on the biological roles of *AhOMT* genes for developmental and stress responses.

KEYWORDS

bioinformatics, environmental stress, functional annotation, gene duplication, micro-RNAs, peanut genomics, phylogenetic tree

Introduction

In *Arabidopsis thaliana*, *O-methyltransferases* (*OMTs*) are heterogeneous enzymes involved in the flavonoid and lignin production pathways (Guo et al., 2001). There are three classes of plant *methyltransferases*: *C- methyltransferases*, *N-methyltransferases*, and *O-methyltransferases* (Roje, 2006). In plants, *OMTs* assist the transfer of the methyl group of *S-adenosyl-L-methionine* (SAM) to the hydroxyl group of numerous organic chemical compounds, ultimately synthesizing the methyl ether variants of these substances (Struck et al., 2012). Based on the molecular weight and bivalent ion dependence, OMTs are divided into *Caffeoyl-CoA OMT* (*CCoAOMT*) and *Caffeic acid OMT* (*COMT*). *COMTs* are the main representative of type I, and *CCoAOMTs* are of type II (Davin and Lewis, 1992). Depending upon the resemblance in sequence and protein motifs, *OMT* genes are further classified into two separate categories: PL-OMT I and PL-OMT II (*CCoAOMT* and *COMT*, respectively) (Joshi and Chiang, 1998). *COMT*-type proteins bind to a variety of substrates, including caffeoyl CoA ester, caffeic acid, chalcones, myoinositol, scoulerine, 5-hydroxyferuloylester, and 5-hydroxyferulic acid (Ye et al., 1994; Roje, 2006). *CCoAOMT*-type enzymes use a pair of substrates, caffeoyl CoA and 5-hydroxyferuloyl CoA, to function (Davin and Lewis, 1992). *COMT* and *CCoAOMT* both mediate the lignin biosynthesis process. The *CCoAOMT* enzyme catalyzes an early step in the pathway by converting *caffeoyl CoA* to *feruloyl CoA* (Dudareva and Pichersky, 2008), despite the fact that sinapyl alcohol, a key component of S-type lignin, is mostly biosynthesized by *COMT* proteins at the end of the biosynthetic pathway (Ye et al., 1994; Buer et al., 2010).

Lignin is the second most prevalent biopolymer on the planet and is an essential element of cell walls in certain higher plants (Ralph et al., 2004). It offers mechanical strength to plants and assists water movement throughout whole plant tissues (Liu et al., 2018), and also an excellent barricade for pathogens, fungi, and insects (Peng et al., 2014), so it helps to improve plant response toward environmental calamities (Moura et al., 2010). To understand their significance, *OMT* genes have been extensively studied in various plants, such as *Arabidopsis* and rice (Hamberger et al., 2007), citrus and sorghum (Liu et al., 2016b; Rakoczy et al., 2018), switchgrass and dove tree (LIU et al., 2016a), tea plant (Lin et al., 2021) etc. Concerning wheat, Nguyen and his team analyzed the expression profiles of lignin biosynthesis-related genes, including a number of *CCoAOMTs*, to determine the likely mechanisms behind their expression patterns. They discovered that lignin content was directly linked with lodging resistance, tolerance to various biotic and abiotic stresses, and quality of feedstock biomass (Nguyen et al., 2016). *TaCCoAOMT1* regulates lignin biosynthesis (Ma and Luo, 2015); previously, this gene has been reported as a key stem cell growth regulator (Bi et al., 2011). Due to their significant roles in secondary metabolism, intensive work has been done on *OMT* genes throughout the years (Bout and Vermerris, 2003; Goujon et al., 2003; Kota et al., 2004; Li et al., 2006; Lin et al., 2006; Yoshihara et al., 2008; Ma, 2009; Zhou et al., 2009). A detailed evaluation of the *OMT* genes in peanut has yet to be performed, despite the fact that the genes' well-established role offers a good foundation for our research.

Therefore, *OMT* genes were studied at a genome-wide scale in *A. hypogaea* and its wild progenitors. One hundred and sixteen *OMT* genes were found in the cultivated peanut genome. Further, we looked into the evolutionary connections of these *AhOMT* genes, their conserved domains and motifs, gene structure, and genomic position. We likewise investigated the *AhOMT* promoters; similarly, expression in different organs under various stress conditions was investigated as well. This study will provide a base for further research on individual genes in peanut and will aid in exploring the biological roles of the *OMT* genes.

Materials and methods

Identification and characterization of *OMT* genes in *A. hypogaea*

OMT genes in the genome of *A. hypogaea* were comprehensively searched. The protein sequences of *AtOMTs* were acquired from the TAIR database (https://www.arabidopsis.org/) (Lamesch et al., 2012) and soybean *OMTs* from Legume Information System (https://legumeinfo.org/) (Gonzales et al., 2005). *A. ipaensis* and *A. duranensis OMT* sequences were obtained from the PeanutBase database (https://www.peanutbase.org/home) (Bertioli et al., 2016). The sequences of whole-genome proteins of *A. hypogaea* were obtained from the Peanut Genome Resource database (PGR) (http://peanutgr.fafu.edu.cn/) (Zhuang et al., 2019). The protein sequences of *OMTs* from *A. duranensis*, *A. ipaensis*, *A. thaliana*, and *G. max* were used to search the *AhOMTs* by BLASTP search with TBtools software (Chen et al., 2020). Further, the HMM search method was also used to search the *OMT* proteins from *A. hypogaea* genome. The Pfam database was searched to obtain the HMM files for the *OMT* family

(PF08100 and PF00891) (http://pfam.xfam.org/). The identified proteins were scanned at NCBI and Pfam databases to verify the *OMT* domain. ProtParam tool (http://web.expasy.org/protparam/) determined the physicochemical characteristics of *AhOMTs* (Gasteiger et al., 2005). The subcellular localizations of *AhOMT* proteins in different cell organelles were predicted by the CELLO version v2.5 (http://cello.life.nctu.edu.tw/) (Yu et al., 2006). General Feature Format (GFF3) files were used to view the exon-intron distribution pattern of *AhOMTs* through TBtools software. Conserved motifs of *AhOMT* proteins were determined by the MEME database (https://meme-db.org/motifs/) (Bailey et al., 2015).

Phylogenetic and gene duplication analysis of *AhOMTs*

A phylogenetic tree comprising *A. ipaensis, A. duranensis, G. max*, *A. hypogaea*, and *A. thaliana* proteins was constructed to investigate their phylogenetic connections. Protein sequences were subjected to multiple sequence alignment by MUSCLE method with the help of MEGAX software (https://megasoftware.net/home) (Kumar et al., 2018). A neighbor-joining tree was generated through 1,000 bootstraps with the poisson model. MCScanX was run to identify the duplicated genes. The KaKs Calculator 2.0 program with the MYN approach was used to determine the rates of synonymous and nonsynonymous substitution (Wang et al., 2010). T = ks/2r was used to compute the divergence time with the neutral substitution coefficient $r=8.12\times10^{-9}$ (Bertioli et al., 2016).

Analysis of *AhOMT* promoters and miRNAs prediction

Promoter sequences up to 2 kb were used to find different binding cites and cis-elements through the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). Coding sequences of *AhOMTs* were used to identify putative miRNAs targeting the *AhOMT* genes through the psRNATarget database (https://www.zhaolab.org/psRNATarget/home) (Dai et al., 2018).

Genome collinearity and orthologous gene clusters

Comparative synteny was analyzed to examine evolutionary genome conservations between three peanut species and *Arabidopsis*. The genome and GFF3 files of all these species were subjected to McScanX in TBtools software, and the resulting files were used for multiple synteny analysis. The orthologous *OMT* proteins were identified in *A. hypogea*, *A. duranensis*, *A. ipaensis*, and *A. thaliana* through OthroVenn2 (https://orthovenn2.bioinfotoolkits.net/home) (Xu et al., 2019). Protein sequences of *Arabidopsis*, soybean, and three peanut species were used to identify orthologous genes. The peanut species were assessed individually with each other and with *Arabidopsis* and soybean to identify orthologous gene clusters.

Functional annotation and prediction of protein-protein interactions

For functional annotation prediction (GO and KEGG), *AhOMT* proteins were scanned at the EggNOG database (http://eggnog-mapper.embl.de/) (Huerta-Cepas et al., 2019). Enrichment analyses were executed in TBtools software from predicted GO and KEGG annotations.

Protein-protein interactions were predicted based on studied *AtOMTs*. STRING 11.5 tool (https://www.string-db.org/cgi/) (Szklarczyk et al., 2019) was used to construct the interaction network between peanut and *Arabidopsis OMTs*. The top 10 interactions were predicted with a medium threshold level (0.4). MCL clustering with inflation parameter 10 was used, and dotted lines were used between cluster edges.

Expression profiling of *AhOMT* Gsenes

Transcriptome expression data were accessed to view the expression levels of *AhOMTs* in various organs, phytohormones, water, and temperature treatments. Transcriptome expression data for different tissues (leaf, stem, stem tip, fluorescence, root, root and stem, root tip, root nodule, gynophore/peg, pericarp, testa, cotyledons, and embryo), hormones (ABA, SA, brassinolide, paclobutrazol, ethephon, and ddH_2O as control), water (drought and normal irrigation) and temperature treatments (low temperature and room temperature) were accessed from the PGR database. The log2 normalization Fragments per kilobase million (FPKM) of *AhOMTs* were used to construct the heatmaps.

Stress treatments and qRT-PCR analysis

Seedlings of peanut cultivar Minhua 6 (M-6) were grown in the greenhouse for stress treatments. Four-leaf old M-6 plants were subjected to abscisic acid stress (ABA 10 μg/mL) and low temperature (4°C). Samples were collected before treatment (0h, CK) for both ABA and low temperature and 3, 6, 9, and 12 hours after treatment. RNA was extracted by the CTAB method with some modifications (Sharif et al., 2022). cDNA was synthesized by Evo M-MLV RT Kit (Accurate Biotechnology, Hunan, Co., Ltd. China) following the manufacturer's protocol. qRT-PCR was performed following our previous study (Sharif et al., 2022), while peanut *Actin* gene was used as the internal control. Data were analyzed by the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen, 2001). Expression levels at different time points were subjected to analysis of variance (ANOVA) and LSD test at $\alpha=0.05$. Primers used for qRT-PCR are given in Supplementary Table 1.

Results

Identification and characterization of *OMT* genes in *A. hypogaea*

BLASTP and HMM searches were performed to find out the *AhOMT* family genes. Twenty-four genes were found in

Arabidopsis, 55 in *G. max*, 58 in *A. duranensis*, and 68 in *A. ipaensis* through a comprehensive search in their respective genome databases. BLASTP search using these proteins and HMM search identified 116 *OMT* genes in the *A. hypogaea* genome. Table 1 shows the details of all 116 *AhOMT* genes. Briefly, *AhOMT* genes varied in size, ranging from 57aa (*AhOMT84* and *AhOMT110*) to 449aa (*AhOMT63*). The same genes possessed the shortest and longest CDS lengths: (*AhOMT84*, *AhOMT110*) with 174bp and *AhOMT63* with 1350bp. The physicochemical properties of these genes also varied accordingly. The molecular weights were from 6.537 kDa (*AhOMT84* and *AhOMT110*) to 502.99 kDa (*AhOMT63*), and theoretical isoelectric points varied from 4.5 (*AhOMT84*, *AhOMT110*) to 9.06 (*AhOMT108*). The differences in isoelectric point (pI) and molecular weights (MW) are attributable to post-translational modifications and a high concentration of basic amino acids.

The subcellular localization prediction of *AhOMT* proteins showed a diverse kind of localization. The main organelle where all *OMTs* were localized was the cytoplasm, while some *AhOMTs* were also localized in more than one cell compartment, including the nucleus, mitochondria, chloroplast, plasma membrane, and extracellular spaces. The physicochemical properties of *AhOMTs* are given in detail in Table 1. Similar patterns of genomic and physicochemical properties were found in the *AdOMTs* and *AiOMTs*. The shortest of *AdOMTs* was *AdOMT25* and *AdOMT41*, with a protein and CDS length of 104 aa and 312 bp, respectively. While the longest *AdOMT* was *AdOMT57*, with a protein and CDS length of 1760 aa and 5280 bp, respectively. The other physiochemical properties also varied, as the molecular weight ranged from 11.78 kDa for *AdOMT41* to 194.78 kDa for *AdOMT57*. The theoretical isoelectric points varied from 4.86 for *AdOMT43* to 8.51 for *AdOMT46*. The protein, CDS lengths, and physiochemical properties of *AdOMTs* are given in Supplementary Table 2. *OMTs* of *A. ipaensis* also possessed similar protein, CDS lengths and other properties. Proteins varied from 68 aa (*AiOMT43*) to 707 aa (*AiOMT63*), while CDS lengths from 204 bp (*AiOMT43*) to 2121 bp (*AiOMT63*). The expected molecular weight for *AiOMTs* ranged from 7.83 kDa (*AiOMT43*) to 78.87 kDa (*AiOMT63*), while the pI varied from 4.56 (*AiOMT19*) to 9.08 (*AiOMT57*). Most *AiOMTs* were located in the cytoplasm, while others were located in mitochondria, endoplasmic reticulum, and nucleus. Supplementary Table 3 shows detailed information about *AiOMTs*.

Phylogenetic relations of *AhOMT* genes

The phylogenetic tree containing *A. ipaensis*, *A. duranensis*, *G. max*, *A. thaliana*, and *A. hypogaea OMT*s divided them into three main groups (Figure 1). *OMTs* of all five species were dispersed in all clades of the phylogenetic tree, indicating that the *OMTs* genes diverged before the divergence of ancestral species. The phylogenetic results revealed that Group I comprised 14 *OMT* members (two *GmOMTs*, one *AtOMT*, four *AiOMTs*, six *AhOMTs*, and one *AdOMT*). Group II comprises 146 *OMT* members (20 *GmOMTs*, 21 *AtOMT*, 31 *AiOMTs*, 50 *AhOMTs*, and 24 *AdOMTs*). Group III contains 160 *OMTs* members (32 from *G. max*, two from *A. thaliana*, 31 from *A. ipaensis*, 62 from *A. hypogaea*, and 33 from *A. duranensis*). In summary, it can be hypothesized from the phylogenetic groupings that *OMTs* from different species with falling in a similar clade will probably perform similar functions. The greater number of *OMTs* in cultivated peanut than in its diploid progenitors and other model plants represent a high evolutionary rate in *A. hypogaea*.

Chromosomal locations and gene duplication

Chromosomal location results revealed that all 116 *AhOMT* genes were dispersed on 18 chromosomes. Chromosomes Chr04 and Chr06 did not possess any *OMT* gene, while one gene was present on the unassembled genome region (Chr00). Chromosomes Chr00, Chr08, and Chr16 possessed one *OMT* each, while Chr07 possessed the highest genes in the A subgenome (15 genes) and in the B subgenome on Chr14 (28 genes) and Chr17 (16 genes), and all other chromosomes possessed varying numbers of *OMT* genes (Figure 2). Chromosomes Chr03, Chr09, and Chr19 had two genes each. Chr01, Chr05, Chr11, Chr12, and Chr18 possessed three genes each, Chr02 possessed four, and Chr15 possessed five *AhOMTs*. Chr20 is next with six genes, Chr10 with eight genes, and Chr13 with ten genes (Figure 2). The *A. duranensis* genome possessed 58 *OMTs* (*AdOMTs*) unevenly distributed on all ten chromosomes. Only chromosome A09 possessed a single *OMT*; all other chromosomes contained multiple copies of *AdOMTs* ranging from 2-19. Chromosome A08 possessed two *AdOMTs*, while the highest number was present on chromosome A07, which had 19 *AdOMTs* (Supplementary Figure 1). The genome of *A. ipaensis* contained 68 copies of *OMT* genes (*AiOMTs*) ranging from 2-16 genes. Chromosome B06 had the least number of *AiOMTs* (two), while chromosomes B04 and B07 possessed the highest number of *AiOMTs* (16 genes each) (Supplementary Figure 2).

Gene duplication analysis revealed 32 duplicated pairs of *AhOMTs*. To estimate the molecular evolution rate, the synonymous (Ks) and nonsynonymous (Ka) substitutions were computed for duplicated genes. Positive selection pressure was assumed when Ka/Ks>1, purifying selection when Ka/Ks<1, and neutral selection when Ka/Ks=1 (Yang and Bielawski, 2000). Results showed that mainly purifying selection drove the genome duplication. Furthermore, the duplicated gene pair divergence timeframe was estimated as t=ks/2r. The expected divergence time varied from 1.078 million years ago (mya) for *AhOMT10*:*AhOMT50* to 185.317 MYA for *AhOMT10*:*AhOMT32* (Table 2). Most genes were segmentally duplicated, but some were tandemly duplicated (Figure 3).

Gene structure and motifs analysis

To better understand the gene structure of *AhOMTs*, we viewed their exon-intron distribution patterns. According to the findings, the introns in *AhOMT* genes varied from 0 to 5, and exons from 1 to

10.3389/fpls.2023.1145624

TABLE 1 Identified OMT genes in *Arachis hypogaea* genome and their physicochemical properties.

mRNA ID	Renamed	Genomic position	Protein (aa)	CDS (bp)	Exons	MW (Da)	pI	Subcellular localization
AH00G01370.1	*AhOMT1*	Chr00, 1743491…1746458, +	367	1104	2	41630.02	5.82	Cytoplasmic
AH01G10670.1	*AhOMT2*	Chr01, 14636118… 14641001, -	252	759	4	28709.05	5.85	Cytoplasmic/Nuclear
AH01G10690.1	*AhOMT3*	Chr01, 14941212… 14944770, +	243	732	4	27654.91	5.29	Cytoplasmic
AH01G14360.1	*AhOMT4*	Chr01, 35441302… 35442524, -	367	1104	1	40114.28	5.02	Cytoplasmic
AH02G04460.1	*AhOMT5*	Chr02, 5569873…5573618, -	386	1161	4	42470.84	5.44	Cytoplasmic
AH02G04490.1	*AhOMT6*	Chr02, 5589981…5593675, -	385	1158	4	42375.64	5.44	Cytoplasmic
AH02G12590.1	*AhOMT7*	Chr02, 32788056… 32790140, -	136	411	3	15838.53	8.7	Extracellular
AH02G16370.1	*AhOMT8*	Chr02, 64332037… 64333326, -	229	690	3	25721.55	5.21	Cytoplasmic
AH03G14330.1	*AhOMT9*	Chr03, 20628097… 20629889, -	353	1062	2	39373.4	5.62	Cytoplasmic
AH03G37380.1	*AhOMT10*	Chr03, 129298299… 129299491, -	365	1098	1	41012.78	5.73	Cytoplasmic
AH05G20880.1	*AhOMT11*	Chr05, 86349327… 86352613, +	370	1113	2	41370.39	5.34	Cytoplasmic
AH05G25050.1	*AhOMT12*	Chr05, 93242846… 93246353, +	238	717	5	26615.79	4.88	Cytoplasmic
AH05G37230.1	*AhOMT13*	Chr05, 113557257… 113558753, +	344	1035	2	38242.97	5.78	Cytoplasmic
AH07G11630.1	*AhOMT14*	Chr07, 16284062… 16284768, +	121	366	2	13465.62	4.95	PlasmaMembrane
AH07G11650.1	*AhOMT15*	Chr07, 16301351… 16313771, -	366	1101	4	40534.71	5.24	Cytoplasmic
AH07G11670.1	*AhOMT16*	Chr07, 16369775… 16372965, +	200	603	3	22486.09	5.96	Cytoplasmic
AH07G11680.1	*AhOMT17*	Chr07, 16456841… 16459427, +	367	1104	4	40435.58	5.91	Cytoplasmic
AH07G11850.1	*AhOMT18*	Chr07, 16692680… 16694596, -	365	1098	2	41236.48	5.37	Cytoplasmic
AH07G12680.1	*AhOMT19*	Chr07, 18937780… 18939658, +	280	843	2	31266.38	6.14	Cytoplasmic/Mitochondrial
AH07G12700.1	*AhOMT20*	Chr07, 19043156… 19045086, +	360	1083	2	40509.67	5.2	Cytoplasmic
AH07G12730.1	*AhOMT21*	Chr07, 19167296… 19169130, +	359	1080	2	40469.82	5.23	Cytoplasmic
AH07G12760.1	*AhOMT22*	Chr07, 19308477… 19309711, +	192	579	2	21730.12	7.84	Nuclear
AH07G12770.1	*AhOMT23*	Chr07, 19319340… 19319576, +	78	237	1	8935.4	4.86	Cytoplasmic/Nuclear
AH07G12810.1	*AhOMT24*	Chr07, 19440775… 19449768, -	365	1098	2	40817.08	5.55	Cytoplasmic
AH07G12840.1	*AhOMT25*	Chr07, 19516967… 19518887, -	367	1104	2	41222.72	5.9	Cytoplasmic

(Continued)

10.3389/fpls.2023.1145624

TABLE 1 Continued

mRNA ID	Renamed	Genomic position	Protein (aa)	CDS (bp)	Exons	MW (Da)	pI	Subcellular localization
AH07G12900.1	*AhOMT26*	Chr07, 19764268… 19768113, -	428	1287	3	47972.23	5.98	Cytoplasmic
AH07G23120.1	*AhOMT27*	Chr07, 76086770… 76090019, -	310	933	5	34361.73	8.64	Chloroplast
AH07G23750.1	*AhOMT28*	Chr07, 78005200… 78008162, -	237	714	4	26289.16	7.06	Mitochondrial/Cytoplasmic
AH08G27130.1	*AhOMT29*	Chr08, 47394140… 47396474, -	377	1134	2	42882.73	6.52	Cytoplasmic
AH09G01720.1	*AhOMT30*	Chr09, 2012805…2015264, -	367	1104	3	41450.75	5.57	Cytoplasmic
AH09G34670.1	*AhOMT31*	Chr09, 120176827… 120179854, +	205	618	6	22778.35	4.72	PlasmaMembrane/Cytoplasmic
AH10G02290.1	*AhOMT32*	Chr10, 1961715…1962889, +	361	1086	1	40872.35	5.6	Cytoplasmic
AH10G15020.1	*AhOMT33*	Chr10, 54238111… 54240140, +	248	747	3	27918	5.54	Cytoplasmic
AH10G16660.1	*AhOMT34*	Chr10, 74583515… 74589326, +	365	1098	4	40023.61	5.67	Cytoplasmic
AH10G18790.1	*AhOMT35*	Chr10, 88278618… 88281277, -	360	1083	4	40453.18	5.59	Cytoplasmic
AH10G18800.1	*AhOMT36*	Chr10, 88297763… 88305007, -	366	1101	4	40511.62	6.09	Cytoplasmic/Chloroplast/ Mitochondrial
AH10G32230.1	*AhOMT37*	Chr10, 114148160… 114150510, -	366	1101	2	40547.15	5.92	Cytoplasmic
AH10G32240.1	*AhOMT38*	Chr10, 114152916… 114155305, -	365	1098	3	40542.23	5.66	Cytoplasmic
AH10G32250.1	*AhOMT39*	Chr10, 114157814… 114159860, -	361	1086	2	40403.42	5.97	Cytoplasmic
AH11G10250.1	*AhOMT40*	Chr11, 18606924… 18610941, -	252	759	4	28733.08	5.72	Cytoplasmic/Nuclear
AH11G10290.1	*AhOMT41*	Chr11, 18873867… 18877209, +	243	732	4	27629.84	5.19	Cytoplasmic
AH11G14590.1	*AhOMT42*	Chr11, 41545738… 41546978, +	368	1107	1	40200.38	5.07	Cytoplasmic
AH12G04920.1	*AhOMT43*	Chr12, 6605031…6609257, -	386	1161	5	42419.84	5.45	Cytoplasmic
AH12G04930.1	*AhOMT44*	Chr12, 6637640…6658360, -	385	1158	4	42338.6	5.5	Cytoplasmic
AH12G19430.1	*AhOMT45*	Chr12, 87354252… 87355625, -	229	690	3	25723.56	5.21	Cytoplasmic
AH13G16990.1	*AhOMT46*	Chr13, 21399995… 21403296, -	360	1083	3	40162.33	5.62	Cytoplasmic
AH13G18140.1	*AhOMT47*	Chr13, 23575873… 23692960, -	362	1089	2	40791.14	6.01	Cytoplasmic/Mitochondrial
AH13G18150.1	*AhOMT48*	Chr13, 23582222… 23583001, -	259	780	1	28803.22	5.71	Cytoplasmic
AH13G18180.1	*AhOMT49*	Chr13, 23740970… 23743323, -	362	1089	2	40836.95	5.62	Cytoplasmic
AH13G40550.1	*AhOMT50*	Chr13, 130194454… 130195643, -	365	1098	1	41041.81	5.66	Cytoplasmic
AH13G54850.1	*AhOMT51*	Chr13, 146141983… 146144481, +	361	1086	4	40440.53	6.12	Cytoplasmic

(Continued)

10.3389/fpls.2023.1145624

TABLE 1 Continued

mRNA ID	Renamed	Genomic position	Protein (aa)	CDS (bp)	Exons	MW (Da)	pI	Subcellular localization
AH13G54860.1	*AhOMT52*	Chr13, 146149238… 146151105, +	367	1104	2	40773.52	5.81	Cytoplasmic
AH13G54880.1	*AhOMT53*	Chr13, 146162045… 146163928, +	370	1113	2	41220.02	6.01	Cytoplasmic
AH13G54900.1	*AhOMT54*	Chr13, 146172622… 146175803, +	367	1104	3	40778.33	5.67	Cytoplasmic
AH13G54910.1	*AhOMT55*	Chr13, 146190691… 146192373, +	367	1104	3	40907.59	5.92	Cytoplasmic
AH14G35680.1	*AhOMT56*	Chr14, 125806126… 125811662, +	369	1110	3	41468.64	5.61	Cytoplasmic
AH14G35740.1	*AhOMT57*	Chr14, 125872352… 125874795, +	369	1110	3	41683.11	5.79	Cytoplasmic
AH14G35970.1	*AhOMT58*	Chr14, 126140382… 126143911, -	367	1104	3	42115.6	5.55	Cytoplasmic
AH14G35990.1	*AhOMT59*	Chr14, 126193089… 126196501, -	367	1104	3	42007.48	5.4	Cytoplasmic/PlasmaMembrane
AH14G36310.1	*AhOMT60*	Chr14, 126627551… 126635090, -	293	882	3	32983.34	5.69	Cytoplasmic
AH14G36320.1	*AhOMT61*	Chr14, 126649959… 126651981, +	363	1092	2	41410.16	5.6	Cytoplasmic
AH14G36340.1	*AhOMT62*	Chr14, 126673606… 126675256, -	266	801	2	29884.76	5.21	Cytoplasmic/PlasmaMembrane/ Chloroplast
AH14G36350.1	*AhOMT63*	Chr14, 126701052… 126704074, -	449	1350	4	50299.16	5.3	Cytoplasmic
AH14G37140.1	*AhOMT64*	Chr14, 127461985… 127463907, -	357	1074	2	40448.73	5.97	Cytoplasmic
AH14G37150.1	*AhOMT65*	Chr14, 127470799… 127472796, -	362	1089	2	40849.08	5.21	Cytoplasmic
AH14G37180.1	*AhOMT66*	Chr14, 127485031… 127487122, -	362	1089	2	40883.95	5.03	Cytoplasmic
AH14G37190.1	*AhOMT67*	Chr14, 127510636… 127512447, -	362	1089	2	40819.03	5.04	Cytoplasmic
AH14G37200.1	*AhOMT68*	Chr14, 127525946… 127528357, +	359	1080	3	40298.99	6.38	Cytoplasmic
AH14G39080.1	*AhOMT69*	Chr14, 129224323… 129226281, +	212	639	3	24003.78	6.51	Cytoplasmic
AH14G39130.1	*AhOMT70*	Chr14, 129291044… 129293057, -	265	798	2	29330.54	6.07	Cytoplasmic
AH14G39140.1	*AhOMT71*	Chr14, 129294783… 129302994, -	311	936	4	35192.57	5.31	Cytoplasmic
AH14G39150.1	*AhOMT72*	Chr14, 129304986… 129307172, -	311	936	3	35207.56	5.83	Cytoplasmic
AH14G43190.1	*AhOMT73*	Chr14, 132761740… 132764130, -	359	1080	3	40177.73	6.37	Cytoplasmic
AH14G43200.1	*AhOMT74*	Chr14, 132775330… 132777397, +	327	984	2	36928.61	5.51	Cytoplasmic
AH14G43220.1	*AhOMT75*	Chr14, 132793894… 132796027, +	362	1089	2	41090.29	4.86	Cytoplasmic

(Continued)

TABLE 1 Continued

mRNA ID	Renamed	Genomic position	Protein (aa)	CDS (bp)	Exons	MW (Da)	pI	Subcellular localization
AH14G43240.1	*AhOMT76*	Chr14, 132813746… 132815906, +	379	1140	2	42884.6	5.28	Cytoplasmic
AH14G43250.1	*AhOMT77*	Chr14, 132824810… 132826734, +	289	870	2	32424.1	4.85	Cytoplasmic
AH14G43260.1	*AhOMT78*	Chr14, 132831217… 132831999, +	260	783	1	29079.49	6.03	Extracellular/Cytoplasmic/ PlasmaMembrane
AH14G44010.1	*AhOMT79*	Chr14, 133523795… 133526394, +	369	1110	3	41613.79	5.55	Cytoplasmic
AH14G44020.1	*AhOMT80*	Chr14, 133543553… 133545010, +	263	792	2	29659.24	5.27	Cytoplasmic
AH14G44040.1	*AhOMT81*	Chr14, 133578635… 133581542, +	363	1092	3	41374.11	5.83	Cytoplasmic/PlasmaMembrane
AH14G44050.1	*AhOMT82*	Chr14, 133600044… 133603205, +	367	1104	3	41974.4	5.39	Cytoplasmic
AH14G44230.1	*AhOMT83*	Chr14, 133827191… 133829559, -	369	1110	3	41590	5.52	Cytoplasmic
AH15G03640.1	*AhOMT84*	Chr15, 5904808…5904981, +	57	174	1	6537.49	4.5	Cytoplasmic/Nuclear
AH15G09730.1	*AhOMT85*	Chr15, 17072781… 17073111, -	80	243	1	8857.37	6.38	Cytoplasmic
AH15G09740.1	*AhOMT86*	Chr15, 17085732… 17086870, -	232	699	2	25463.92	5.59	Cytoplasmic
AH15G30330.1	*AhOMT87*	Chr15, 143877154… 143880313, -	231	696	5	25743.82	5.1	Cytoplasmic
AH15G34850.1	*AhOMT88*	Chr15, 149516849… 149520694, -	372	1119	2	41633.72	5.41	Cytoplasmic
AH16G14480.1	*AhOMT89*	Chr16, 24990946… 24992363, -	283	852	3	30936.49	5.33	Cytoplasmic
AH17G11080.1	*AhOMT90*	Chr17, 17572788… 17574790, +	230	693	4	25641.62	6.7	Cytoplasmic
AH17G11130.1	*AhOMT91*	Chr17, 17599988… 17615368, -	367	1104	4	40666.91	5.31	Cytoplasmic
AH17G11160.1	*AhOMT92*	Chr17, 17675441… 17680659, +	373	1122	3	41174.51	5.31	Cytoplasmic
AH17G11170.1	*AhOMT93*	Chr17, 17720906… 17725507, +	373	1122	4	41157.32	5.16	Cytoplasmic
AH17G11190.1	*AhOMT94*	Chr17, 17820754… 17838864, +	374	1125	4	41252.7	5.62	Cytoplasmic
AH17G11220.1	*AhOMT95*	Chr17, 17982468… 17985690, +	367	1104	4	40481.65	5.71	Cytoplasmic
AH17G11350.1	*AhOMT96*	Chr17, 18600583… 18601386, -	267	804	1	29461.77	5.43	Cytoplasmic/Chloroplast
AH17G12150.1	*AhOMT97*	Chr17, 21159366… 21161312, +	363	1092	2	40940.59	5.75	Cytoplasmic
AH17G12180.1	*AhOMT98*	Chr17, 21347129… 21349187, +	349	1050	2	39358.41	5.19	Cytoplasmic
AH17G12210.1	*AhOMT99*	Chr17, 21390326… 21393847, +	363	1092	2	40924.45	5.53	Cytoplasmic

(Continued)

TABLE 1 Continued

mRNA ID	Renamed	Genomic position	Protein (aa)	CDS (bp)	Exons	MW (Da)	pI	Subcellular localization
AH17G12230.1	*AhOMT100*	Chr17, 21541743… 21545356, -	259	780	2	29208.88	5.3	Cytoplasmic
AH17G12310.1	*AhOMT101*	Chr17, 21840489… 21842905, -	288	867	2	32635.6	4.87	Cytoplasmic
AH17G12370.1	*AhOMT102*	Chr17, 21929588… 21931865, -	376	1131	2	42715.59	5.29	Cytoplasmic
AH17G12380.1	*AhOMT103*	Chr17, 21978612… 21984311, -	384	1155	2	42688.46	6.38	PlasmaMembrane/Cytoplasmic
AH17G12420.1	*AhOMT104*	Chr17, 22160786… 22163384, -	352	1059	2	39265.49	5.16	Cytoplasmic
AH17G12450.1	*AhOMT105*	Chr17, 22249751… 22251901, -	364	1095	2	40704.96	5.3	Cytoplasmic
AH18G08980.1	*AhOMT106*	Chr18, 10575268… 10579544, -	290	873	3	32434.37	5.94	Cytoplasmic
AH18G18630.1	*AhOMT107*	Chr18, 42350728… 42358246, -	357	1074	4	39647.35	6.08	Cytoplasmic
AH18G19640.1	*AhOMT108*	Chr18, 53503322… 53506529, +	311	936	4	34338.74	9.06	Chloroplast/Mitochondrial
AH19G00660.1	*AhOMT109*	Chr19, 514524…516286, -	362	1089	3	40364.32	5.66	Cytoplasmic
AH19G24900.1	*AhOMT110*	Chr19, 113810989… 113811162, -	57	174	1	6537.49	4.5	Cytoplasmic/Nuclear
AH20G07220.1	*AhOMT111*	Chr20, 9290110…9291284, -	361	1086	1	40833.25	5.6	Cytoplasmic
AH20G13670.1	*AhOMT112*	Chr20, 21608191… 21610458, +	248	747	5	27947.09	5.83	Cytoplasmic
AH20G19630.1	*AhOMT113*	Chr20, 57791932… 57794102, -	248	747	3	27931	5.53	Cytoplasmic
AH20G22290.1	*AhOMT114*	Chr20, 99007639… 99013440, -	403	1212	4	44371.69	6.89	Cytoplasmic
AH20G24820.1	*AhOMT115*	Chr20, 114186636… 114189506, -	360	1083	3	40458.19	5.68	Cytoplasmic
AH20G24830.1	*AhOMT116*	Chr20, 114193140… 114198405, -	360	1083	4	39987.05	6.16	Cytoplasmic

The + and - represents the positive and negative DNA strands.

6. Many *AhOMT* genes were composed of a single intron and two exons. Forty-two out of 116 *AhOMTs* possessed two exons. Three and four exons were also common, as 30 genes possessed three exons while 25 genes had four exons. Thirteen genes were composed of a single exon, and only *AhOMT31* comprised six exons (Figure 4). EME server identified conserved motifs inside the full-length protein sequences of *AhOMT* genes in order to determine structural diversification and functional assessment. Ten conserved motifs were predicted in *AhOMT* genes (Figure 4). Conserved motifs varied in length as motif 1 was the most extended motif with 39 amino acids, while 4^{th}-6^{th} and 8^{th}-10^{th} motifs were the shortest with 21 amino acid residues (Supplementary Table 4). In a nutshell, conserved motif, phylogenetic, and gene structure analysis indicated that *AhOMT* proteins comprise extremely well-sustained members of amino acids that remain inside a group. Proteins with similar motifs and structures can therefore be functionally related. The motif distribution patterns and gene structure of *OMTs* of wild progenitors were as per *A. hypogaea OMTs*. Information on motifs and structure of *AdOMTs* are given in Supplementary Figure 3, and on *AiOMTs* is given in Supplementary Figure 4.

Promoter analysis of *AhOMTs* genes

The *cis*-elements of any genes' promoter are responsible for controlling its expression and functions. We examined *cis*-acting regions in the *AhOMT* promoters to know their functional and regulatory roles. Predicted *cis*-elements showed that aside from the CAAT- and TATA-Box (core promoter elements), a large number of other key elements were also present (Figure 5). We classified these *cis*-regulatory elements into four groups according to their functions: development and growth-related, hormones-responsive,

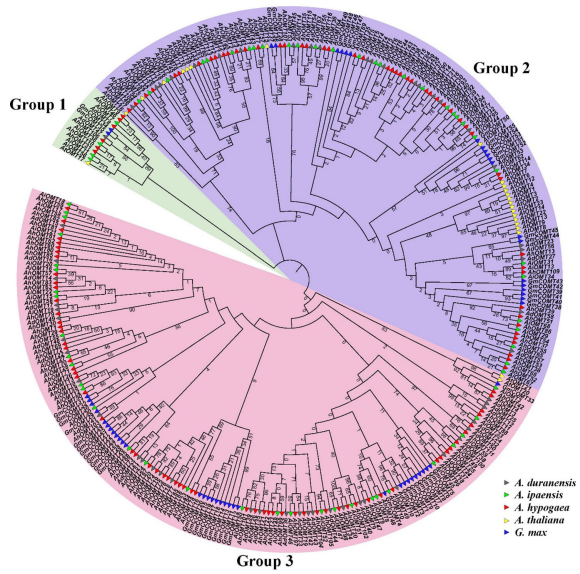


FIGURE 1
Phylogenetic relationships of O-Methyltransferase genes of *A. hypogea*, *A. duranensis*, *A. ipaensis*, *A. thaliana*, and *G. max*. The phylogenetic tree classified all *OMTs* into three groups, *OMTs* of all species are present in all clades. Group one is the smallest group as compared to other groups.

light-responsive, and stress-related elements. All 116 *AhOMTs* were enriched with hormones- and light-responsive elements, 108 genes were enriched with growth and development-related elements, and 94 genes were enriched with stress-responsive elements (Figure 6).

Elements responsive to light mainly include TCT-motif, GATA-motif, G-box, Box-4, GT1-motif, GA-motif, chs-CMA element, I-box, and AT-1 motif. Other light-responsive elements include 3-AF1 binding site, ATC-motif, AE-box, MRE element, Box II, CAG-motif, CGTCA-motif ATCT-motif, ACE element, Gap-box, TCCC-motif, GTGGC-motif, LAMP-element, LS7 element, and Sp1 element were also present. Hormones responsive class includes ABA-responsive (ABRE), auxin-responsive (AuxRE, AuxRR-core, CGTCA-motif, TGA-box), gibberellins responsive (GARE motif, P- and TATC-box), MeJA-responsive (CGTCA-motif, TGACG-motif), SA-responsive (SARE, TCA-element), and ethylene-responsive (ERE) elements. The growth and development category contained anaerobic induction responsive (ARE), meristem expression responsive (CAT-box), endosperm expression related (GCN4-motif, AACA-motif), circadian control (CAAAGATATC), and zein metabolism-related (O2-site) elements. The stress-responsive class further includes defense and stress response (TC-rich repeats), drought-responsive (MBS), low-temperature responsive (LTR), and wound-related (WUN-motif) elements (Figure 6).

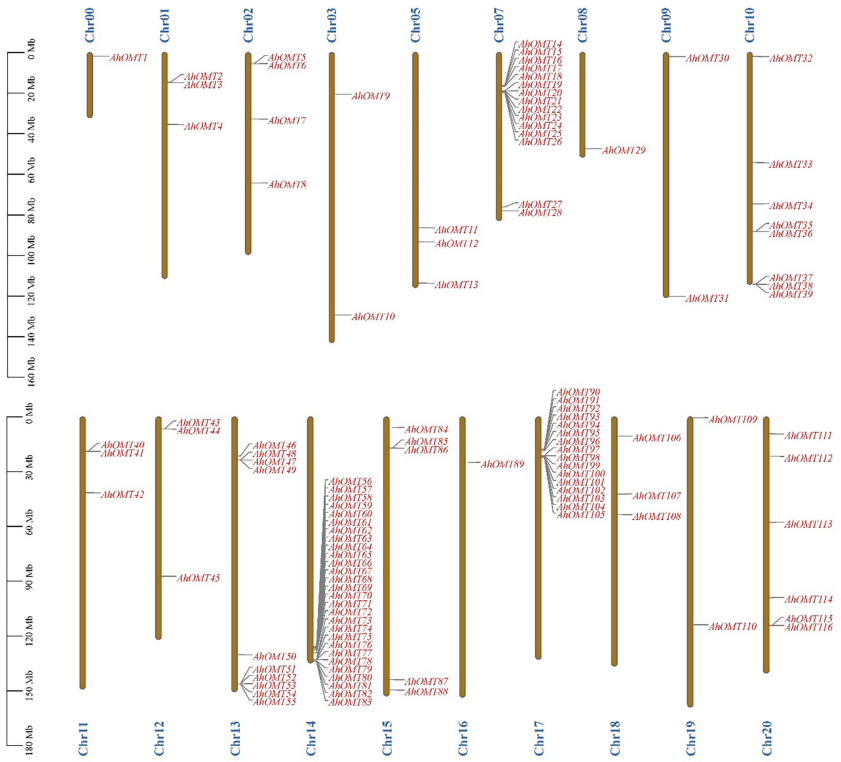


FIGURE 2
Chromosomal distribution of *AhOMTs* genes. The highest number of *AhOMTs* on A-subgenome were present on Chr07 (15 genes) and in B-subgenome on Chr14 (28 genes) and Chr17 (16 genes), and all other chromosomes possessed varying numbers of OMT genes.

TABLE 2 Calculation of Ka/Ks values and divergence time of duplicated genes.

Seq_1	Seq_2	Ka	Ks	Ka_Ks	Selection Pressure	Divergence Time
AhOMT2	*AhOMT40*	0.006822	0.030554	0.223279	Purifying	1.881
AhOMT5	*AhOMT43*	0.022377	0.036767	0.608621	Purifying	2.264
AhOMT8	*AhOMT45*	0.003812	0.051412	0.074145	Purifying	3.166
AhOMT10	*AhOMT32*	0.418478	3.009543	0.13905	Purifying	185.317
AhOMT10	*AhOMT50*	0.006978	0.017506	0.398609	Purifying	1.078
AhOMT11	*AhOMT88*	0.006941	0.05584	0.124305	Purifying	3.438
AhOMT12	*AhOMT87*	0.011542	0.030155	0.382771	Purifying	1.857
AhOMT13	*AhOMT109*	0.016391	0.026532	0.617772	Purifying	1.634
AhOMT18	*AhOMT96*	0.021194	0.046102	0.459725	Purifying	2.839
AhOMT19	*AhOMT97*	0.02151	0.107486	0.200115	Purifying	6.619
AhOMT19	*AhOMT102*	0.125147	0.476578	0.262594	Purifying	29.346
AhOMT21	*AhOMT99*	0.0749	0.191072	0.391998	Purifying	11.766
AhOMT21	*AhOMT104*	0.134389	0.50998	0.263517	Purifying	31.403
AhOMT24	*AhOMT105*	0.186122	0.728876	0.255355	Purifying	44.882
AhOMT25	*AhOMT100*	0.026577	0.097381	0.272919	Purifying	5.996
AhOMT26	*AhOMT102*	0.141688	0.460419	0.307738	Purifying	28.351
AhOMT27	*AhOMT108*	0.008499	0.037275	0.228001	Purifying	2.295
AhOMT29	*AhOMT1*	0.0036	0.045451	0.079214	Purifying	2.799
AhOMT32	*AhOMT50*	0.412074	2.671299	0.15426	Purifying	164.489
AhOMT32	*AhOMT111*	0.005889	0.030978	0.190114	Purifying	1.908
AhOMT33	*AhOMT113*	0.005231	0.024123	0.216852	Purifying	1.485
AhOMT34	*AhOMT114*	0.002375	0.024228	0.098039	Purifying	1.492
AhOMT35	*AhOMT115*	0.004795	0.046675	0.102727	Purifying	2.874
AhOMT37	*AhOMT52*	0.124998	0.287727	0.434431	Purifying	17.717
AhOMT39	*AhOMT51*	0.005973	0.029436	0.202899	Purifying	1.813
AhOMT46	*AhOMT106*	0.125462	1.269437	0.098833	Purifying	78.167
AhOMT57	*AhOMT83*	0.018462	0.054159	0.340895	Purifying	3.335
AhOMT58	*AhOMT80*	0.061586	0.172839	0.356321	Purifying	10.643
AhOMT60	*AhOMT81*	0.022063	0.060556	0.364345	Purifying	3.729
AhOMT62	*AhOMT80*	0.052874	0.129797	0.407362	Purifying	7.992
AhOMT63	*AhOMT79*	0.019671	0.086954	0.226226	Purifying	5.354
AhOMT65	*AhOMT74*	0.034348	0.040983	0.838105	Purifying	2.524

Prediction of miRNAs and synteny analysis

Numerous studies in the last few years have revealed that micro-RNAs regulate the expression of genes under developmental processes and stress responses (Chen et al., 2019; Wani et al., 2020; Raza et al., 2021a). For this reason, we predicted *miRNAs* targeting *AhOMT* genes sequentially to get more understanding of miRNA-mediated post-transcriptional regulations of *AhOMT* genes. Micro RNAs from 12 different families targeted 35 *AhOMTs*. Supplementary Table 5 contains the complete information on all *miRNAs*. Two members of the *miR156* family targeted *AhOMT34*, *AhOMT37*, *AhOMT38*, *AhOMT52-AhOMT55*, *AhOMT87*, and *AhOMT114*. *miR160-3p* was found to target four *OMTs*. Some of the *miRNAs* targeting the *AhOMTs* with their target sites are shown in Figure 7. More research for their expression levels and the genes they target is needed to establish their biological involvement in the peanut genome.

Comparative synteny analysis among *A. hypogaea*, diploid peanut species, and *A. thaliana* represented remarkable

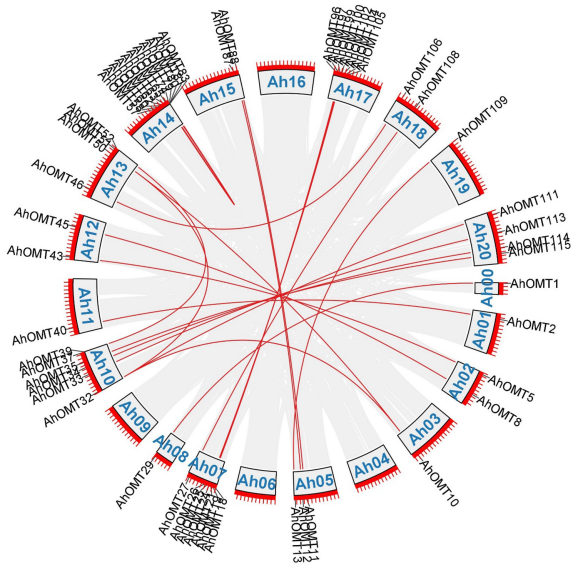


FIGURE 3
Duplicated gene pairs among *AhOMTs*. Out of 116 genes, 42 are duplicated genes. Red lines show the duplicated *AhOMT* pairs, while the gray lines in the background show the syntenic blocks (duplicated pairs) among different chromosomes.

evolutionary, duplication, expression, and functional relationships. *AhOMTs* mainly showed significant syntenic relationships with its wild progenitors and *Arabidopsis*; however, the syntenic relationships of *A. hypogaea* were closer to wild peanut species than *Arabidopsis*. A total of 56 syntenic relationships of *A. hypogaea* were found in the genome of *A. duranensis* and 60 in *A. ipaensis*. In contrast, only four syntenic relationships were found among *AhOMTs* and *AtOMTs*. The synteny analysis showed that *A. hypogaea* is closer to its wild progenitors than *Arabidopsis*. The syntenic relations of these species are shown in Figure 8.

Identification of orthologous gene clusters

Identifying orthologous gene clusters is important to assess the polyploidization events during a gene family's evolution. A relative assessment was developed to identify orthologous gene clusters shared by *A. hypogea, A. duranensis, A. ipaensis, G. max, and A. thaliana*. The detected gene clusters and their respective overlapping regions are presented in greater detail in Figure 9. A. *hypogea* recorded maximum clusters, followed by *A. ipaensis*, *A. duranensis*, *G. max*, and *A. thaliana*. Results showed that three gene clusters are shared among all these species, while 18 gene clusters are solely composed of *OMTs*

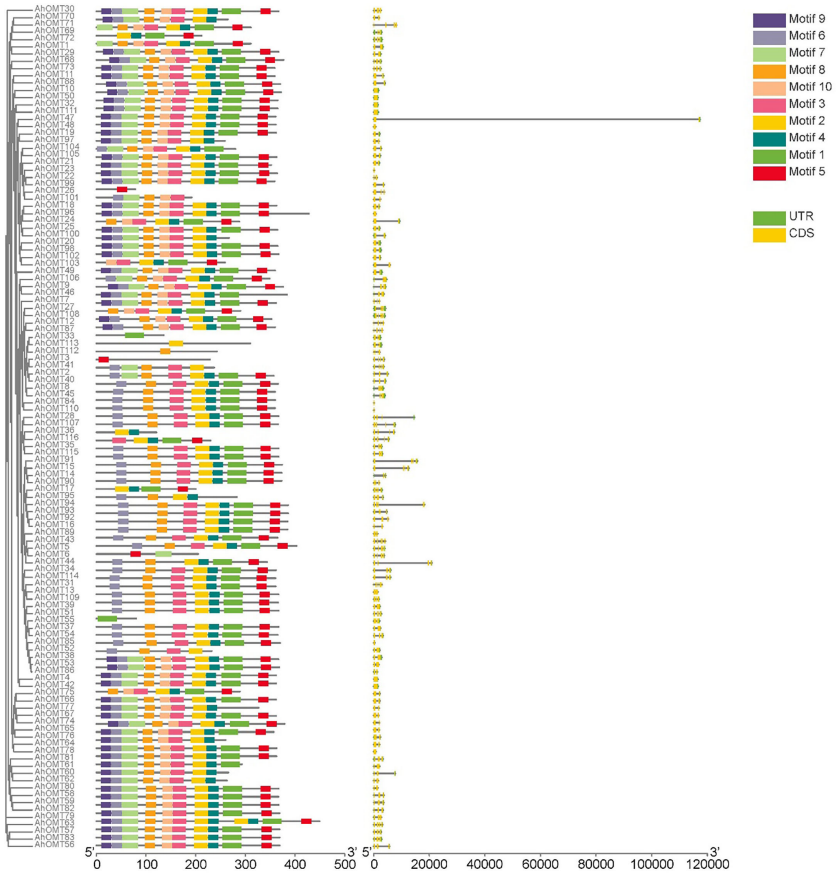


FIGURE 4
Conserved motifs distribution patterns and gene structure (exons-introns distribution) of *AhOMTs* genes. Commonly shared motifs among genes tend to cluster in the same groups, referring to their similar functions.

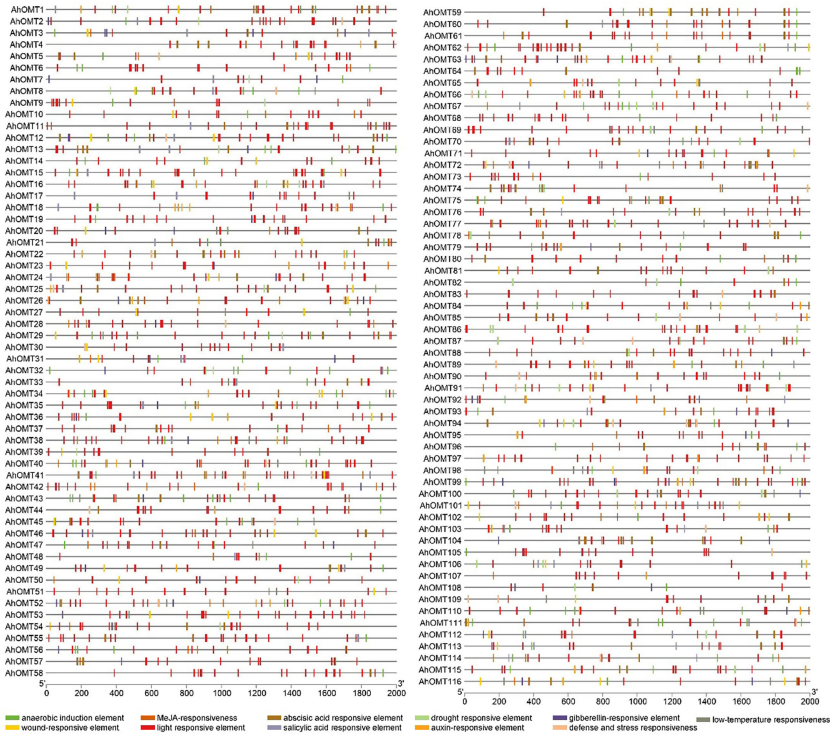


FIGURE 5
Cis-regulatory elements of *AhOMT* promoters. *Cis*-elements analysis revealed important elements responsive to light, hormones, growth and development, and stress responsiveness.

found in peanut diploid and tetraploid species, which indicates that polyploidization has evolved new peanut-specific orthologous *OMT* clusters. We also identified orthologous gene clusters among three peanut species (Supplementary Figure 5). Comparatively, 100, 89, 94, 36, and 21 orthologous *OMTs* were found in *A. hypogea*, *A. duranensis*, *A. ipaensis*, *G. max*, and *A. thaliana*, respectively. Thirty in-paralogs were identified in *A. hypogea*, and only two were found in *A. ipaensis*. *A duranensis* did not show any in-paralogous gene. Surprisingly 32, 14, and 20 singletons were also found in *A. hypogea*, *A. duranensis*, and *A. ipaensis*, respectively (Supplementary Table 6). Results demonstrated that identified orthologous genes decrease with increased phylogenetic distances.

Prediction of protein-protein interaction network

The Functions of *AhOMTs* could be speculated based on well-studied *Arabidopsis OMTs*. Using the STRING database, we performed the interaction network analysis of cultivated peanut *OMT* proteins relative to orthologues in *Arabidopsis* to understand their functions. Protein interaction network prediction showed that *AhOMT116* has functions related to C4H that regulate carbon flux to essential pigments for pollination or UV protection. *AhOMT7* and *AhOMT111* may function as Cinnamoyl-CoA reductase 1 (IRX4) involved in lignin biosynthesis at the latter stages. *AhOMT87* has *CCOAMT*-like functions, a putative caffeoyl-CoA O-methyltransferase of *Arabidopsis* that helps in the biosynthesis of feruloylated polysaccharides. *AhOMT77* has 4CL1-related functions (4-coumarate-CoA ligase 1), involved in the later phase of the general phenylpropanoid pathway. *AhOMT31* may function as SNC1, a putative disease-resistance protein of the TIR-NB-LRR-type. The interaction network of *AhOMTs* with well-studied *Arabidopsis* proteins is given in Figure 10. Some *OMTs* did not show interactions with reported *Arabidopsis* proteins, and there is a possibility that these proteins have some other functions yet to be reported.

Functional annotation analysis of *AhOMTs*

GO annotation analysis of *AhOMTs* was performed to view their possible roles in biological processes (BP), molecular functions (MF), and cellular components (CC). GO enrichment results provided highly enriched terms related to BP, MF, and CC (Figure 11). *AhOMTs* were mainly involved in MF and BP categories. *AhOMTs* were highly enriched in transferase activity (GO:0016740), catalytic activities (GO:0003824), methyltransferase activity (GO:0008168, GO:0008171, GO:0042409), and S-adenosylmethionine-dependent methyltransferase activity (GO:0008757) in MF category. In the BP category, *AhOMTs* were highly enriched in methylation (GO:0032259), biosynthetic process (GO:0009058, GO:0044249), cellular metabolic processes (GO:0044237, GO:0008152), and aromatic compound metabolism (GO:0006725). The KEGG enrichment analysis showed that *AhOMTs* are mainly involved in metabolic processes, including

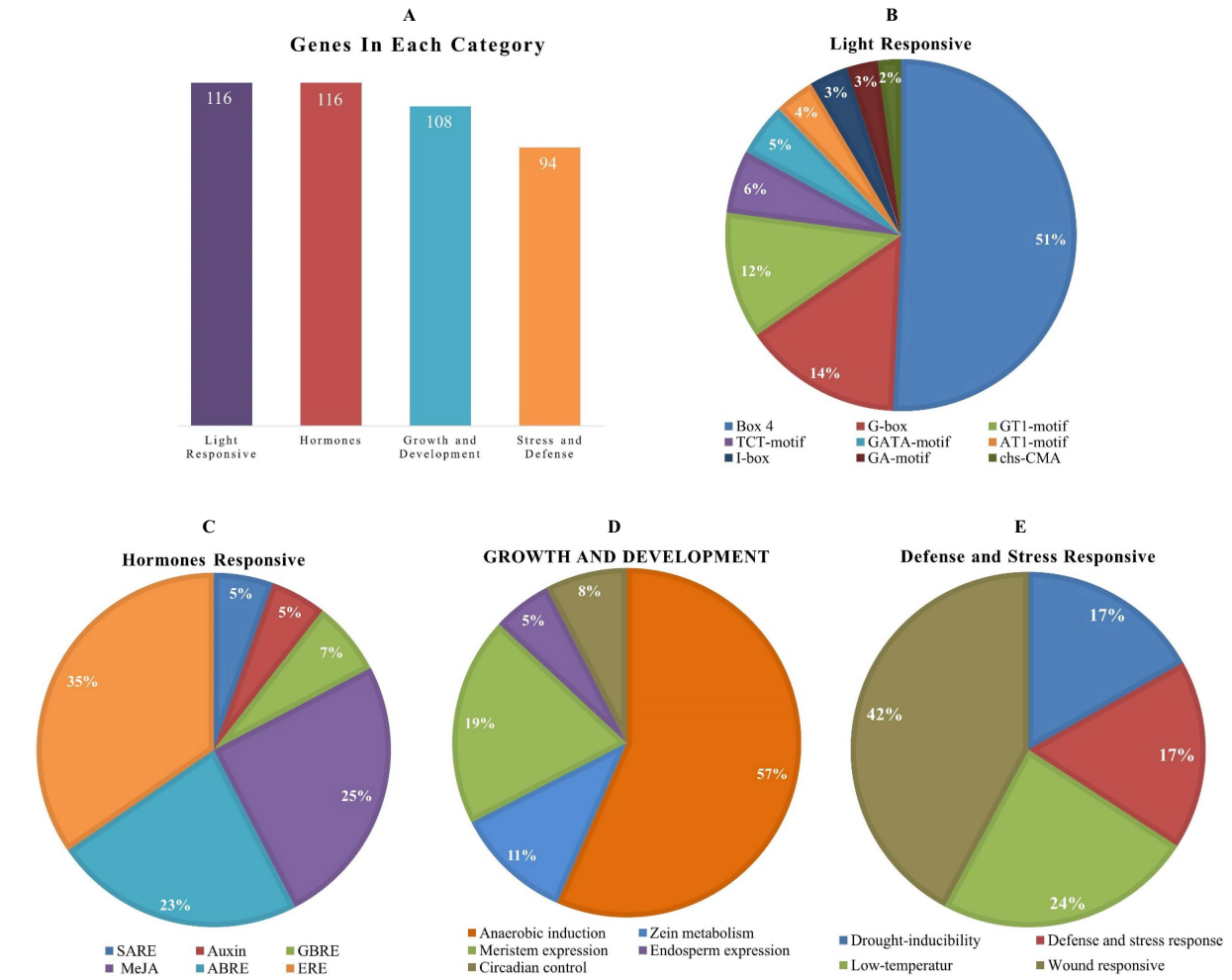


FIGURE 6
Cis-regulatory elements of *AhOMT* promoters. **(A)** The number of genes in different categories of elements. **(B)** Composition of light-responsive, **(C)** hormone-responsive, **(D)** growth and development responsive, and **(E)** stress-responsive factors.

01058 acridone alkaloid biosynthesis, 00943 isoflavonoid biosynthesis, B 09110 secondary metabolites biosynthesis, 00380 tryptophan metabolism, 00941 flavonoid production, and amino acid B 09105 metabolism (Figure 11). Collectively, it is evident from functional annotation analysis that *AhOMTs* play key roles in several cellular, biological, and molecular functions.

Expression profiling of *AhOMTs* in different organs

AhOMT genes' expression levels in different organs/tissues, containing leaf, stem, flower, root, root nodule, peg, pericarp, testa, cotyledon, embryo, etc., was determined using the peanut RNA-seq datasets. According to the expression profiling results, there was a noticeable variance in the expression of various tissues. Transcriptome expression results showed that *AhOMT32-AhOMT35, AhOMT45, AhOMT71, AhOMT106, AhOMT113, AhOMT114*, and *AhOMT116* genes showed relatively higher levels of transcriptional abundance in the leaf, stem, flower, root, root nodule, peg, pericarp, testa, cotyledon, and embryo. These genes can be suitable candidates for improving peanut growth and yield. *AhOMT9* and *AhOMT46* specifically showed high expression in root nodules (Figure 12). It can be speculated that these two genes are good targets to improve nitrogen fixation that can provide good crops by effectively fixing the soil nitrogen. FPKM values of transcriptome expression of *AhOMTs* are given in Supplementary Table 6.

Expression profiling of *AhOMTs* under hormones, drought, and temperature stress

Transcriptome data provided the expression patterns of 116 *AhOMTs* for different phytohormones (ABA, SA, Brassinolide, Paclobutrazol, and Ethephon) treatment, water stress (drought and regular irrigation), temperature stress (4°C and 28°C). Under temperature stress, the *AhOMT106* gene was highly active, while *AhOMT35*, *AhOMT71*, and *AhOMT113* were also expressed in most cases, but *AhOMT35* did not show expression under drought stress. Almost 16 genes showed expression in response to ABA and

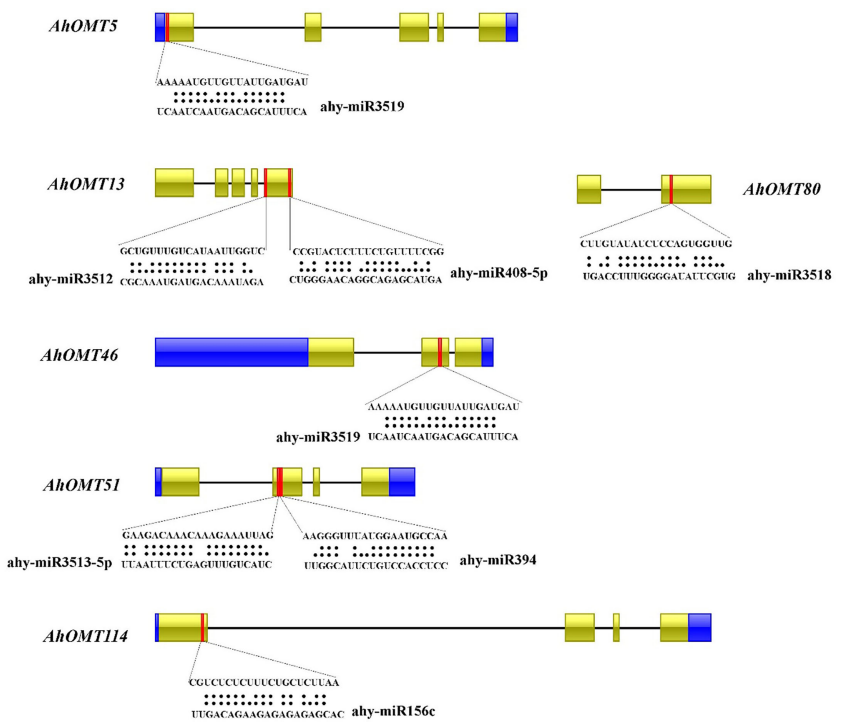


FIGURE 7
Predicted miRNAs targeting *AhOMTs*. Schematic representation of some miRNAs targeting *AhOMTs*, and their target sites.

SA, 14 genes responded to brassinolide, and 12 genes were responsive to ethephone. Thirteen genes were expressed under decreased temperature, and almost 11 genes were responsive to drought stress (Figure 13). Many genes were non-responsive to the hormones, water and temperature treatments.

Quantitative expression profiling under ABA and low-temperature treatment

For real-time expression profiling by qRT-PCR, 12 *AhOMT* genes were randomly selected. These genes included *AhOMT-7, AhOMT-18, AhOMT-33, AhOMT-34, AhOMT-35, AhOMT-46, AhMT-61, AhOMT-71, AhOMT-93, AhOMT-106, AhOMT-113,* and *AhOMT-116*. These genes were selected based on their response to hormones, water and temperature stress, while genes with higher and lower expression were considered. Under ABA treatment, the expression of all selected genes corresponds to their transcriptome expression. For instance, *AhOMT-7, AhOMT-33, AhOMT-34, AhOMT-35, AhOMT-71, AhOMT-93, AhOMT-106, AhOMT-113*, and *AhOMT-116* were upregulated under ABA stress, while *AhOMT-18*, *AhOMT-46*, and *AhOMT-61* were downregulated (Figure 14). Under low temperature, a similar expression was found as of ABA treatment. Although there were some deviations in transcriptome expression and qRT-PCR expression, overall, the expression pattern of all selected genes is in accordance with transcriptome expression (Figure 15). The results of qRT-PCR represent the reliability of transcriptome datasets.

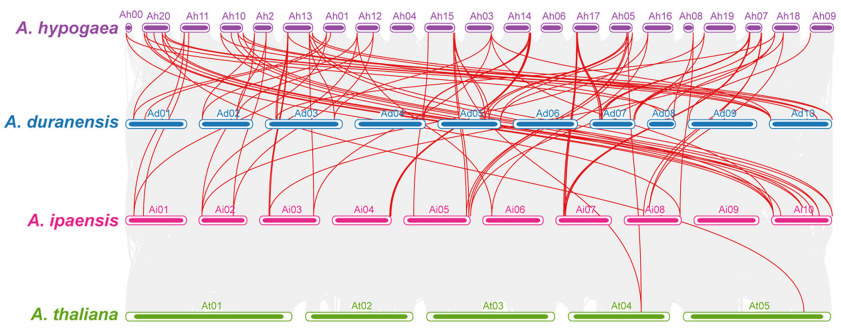


FIGURE 8
Synteny analysis among *A. hypogea*, *A. duranensis*, *A. ipaensis*, and *A. thaliana*. Synteny analysis showed key evolutionary relationships of *OMTs* in diploid and tetraploid peanut species. *AhOMTs* possessed highly conserved syntenic relationships with other peanut species as compared to *Arabidopsis*.

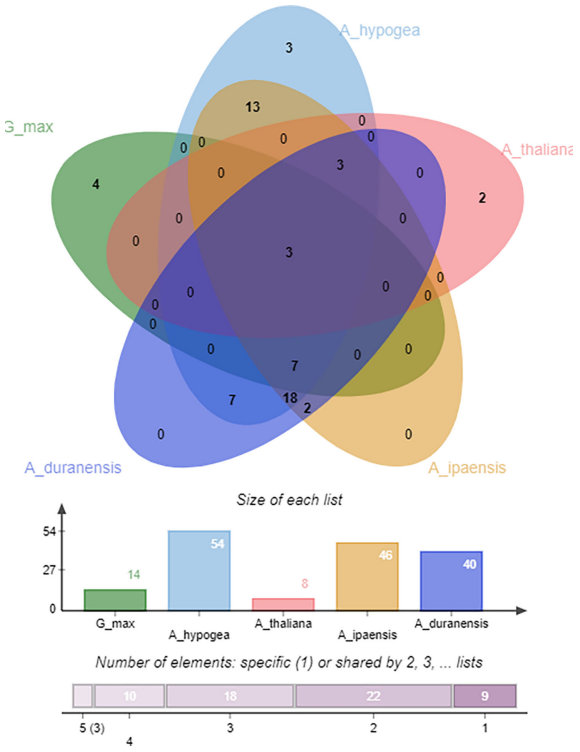


FIGURE 9
Orthologous genes' clusters among *A. hypogea*, *A. duranensis*, *A. ipaensis*, *A. thaliana*, and *G. max*. *A. hypogea* recorded maximum clustered. Eighteen gene clusters are solely composed of OMTs found in peanut diploid and tetraploid species, indicating that polyploidization has evolved new peanut-specific orthologous OMT gene clusters.

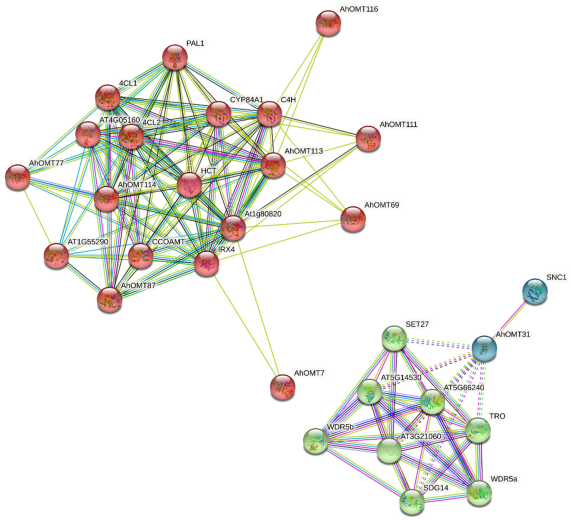


FIGURE 10
The predicted protein–protein interaction network of AhOMTs using STRING database. Putative protein functions of AhOMTs are predicted on well-studied OMT orthologues in *Arabidopsis*.

Discussion

Several plants, including *A. thaliana*, *B. distachyon*, *B. napus*, *P. trichocarpa*, *O. sativa*, and others, have been studied at the whole-genome level to determine the presence and possible roles of *OMT* family genes. Because of their importance for synthesizing S-type lignin, the roles of *OMTs* have been well established. Lignin is the cell wall's most important component to cope with environmental and biological stress (Boerjan et al., 2003). Reduced lignin production poses the plant to a lodging state (Hu et al., 2015). Reduced lignin concentration in legumes reduces stalk strength which ultimately reduces diseases and pathogens resistance (Bellaloui, 2012). Genome size, genome duplication, and gene distribution all have a significant influence on genetic diversity. Genetic duplication has been recognized for years as a source of the expression, originality, and variety found in gene families across species (Wang et al., 2012). Additionally, some *AhOMT* duplications may be crucial to their multiplication as they can bring neofunctionalization and diversity in gene families (Lavin et al., 2005; Chapman et al., 2008).

Some gene families have originated and extended due to tandem or segmental duplications. Gene family's evolution in this manner is crucial for their diversification (Cannon et al., 2004). The opposite is also true: gene function may have an impact on copy number and genome structure, resulting in widely disparate patterns of segmental or tandem duplication (Cannon et al., 2004). After tandem duplication, genes occur in clusters (Savard et al., 2011). It is important to understand the evolution of gene clusters to provide updated information on evolutionary history. Previously occurrence of tandem duplication was confirmed in pomegranates by gene mapping by Yuan and coworkers. They identified three *OMT* genes (*PgOMT01* to *PgOMT03*). Relatively large scale duplication of the pomegranate genome resulted in forming the *PgOMT* tandem duplications (Yuan et al., 2018). To a certain extent, tandem duplication has evolved the *PgAOMT* family.

Exon numbers and distribution patterns have a key role in the expression of any gene (Kolkman and Stemmer, 2001). In our investigation, most *AhOMTs* had fewer introns, and members of the same evolutionary group tended to have exon–intron patterns comparable. For instance, the presence of two or more introns in *AhOMT* genes demonstrates that the *OMT* gene development may be directly tied to the diversity of gene architectures. A similar set of findings has also been observed for the *OMT* gene family in Chinese jujube (Song et al., 2017). Several studies have found that genes with lesser introns expressed rapidly as introns can influence expression by delayed transcript synthesis in three different means, by (1) splicing, (2) increasing the length of the growing transcript, or (3) increasing the energy requirement of the transcript of lengthy transcripts (Jeffares et al., 2008). Less number of introns in most *AhOMTs* than its progenitors indicates a possible quicker response to induction; however, additional research is required to confirm this hypothesis. *OMT* proteins from five species used in this study were clustered into three phylogenetic groups. Conferring to the phylogenetic tree, three unique groups represents substrate specificity according to their functional traits (Joshi and Chiang, 1998).

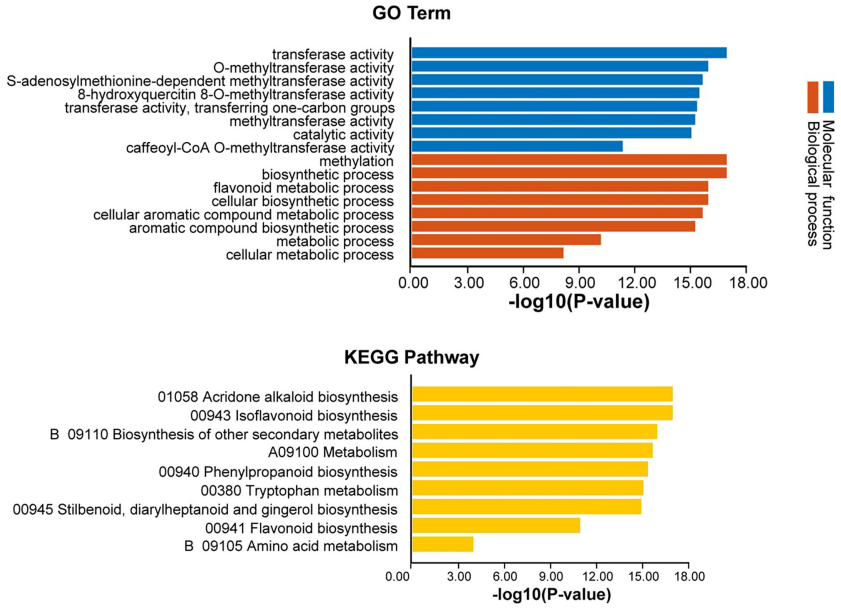


FIGURE 11
Functional annotation (GO, KEGG) analysis of *AhOMTs*. The gene ontology analysis showed that *AhOMTs* are involved in molecular function (MF), and biological processes (BP). The KEGG enrichment analysis showed that *AhOMTs* are mainly involved in metabolic processes.

To control gene transcription, various proteins must bind to *cis*-regions of the promoter. GT1-motif (Gao et al., 2004), GATA-motif (Argüello-Astorga and Herrera-Estrella, 1998), I-box (Donald and Cashmore, 1990), and G-box (Giuliano et al., 1988), are *cis*-regions needed for light-mediated transcription. According to our findings, S-type lignin may be controlled by *AhOMT* genes, which may interact with light-induced proteins and have circadian patterns in their gene promoters. The circadian rhythm regulates many genes in higher plants, including those involved in photosynthesis and starch mobilization. Hormones highly influence plant growth and development. According to Kim and coworkers, the kenaf *OMT* gene (*Hibiscus cannabinus*) is expressed after six hours of SA, ABA, auxin, ethylene, and GA treatment (Kim et al., 2013). Their findings also support our results, as *AhOMT* genes were generally influenced by hormone treatments. With this, SA-related factors were discovered in the *AhOMT* promoters, implying their key function in the hormonal regulation of *AhOMT*. When they studied the *OMT* gene, they observed that it could be stimulated by H_2O_2, cold, and salt, which showed that hormonal and abiotic stimuli might affect the *OMT* genes' transcription. We also found similar findings for cold stress, as *AhOMTs* were highly influenced (up- and down-regulated). Another study indicated that *Brassica napus OMT* family genes were more highly expressed under drought-stressed circumstances than in regular irrigation (Li et al., 2016). The cold and drought have been shown to significantly increase the expression of an *OMT* gene in *Ligusticum chuanxiong* (Li et al., 2015). Some *OMT* promoter sequences included stress-related motifs such as ARE, LTR, and MBS. *OMT* genes are influenced by salt, and cold stress (Kim et al., 2013) and the presence of stress-responsive elements suggests that *OMT* genes might play a role in neutralizing the abiotic stresses. Some *AhOMT* gene promoters were revealed to have heat-responsive and MBS sights that can collectively induce drought tolerance. In addition to these CREs, stress response involves TC-rich repeats, W1-BOx, ARE, and LTR (Zhang et al., 2015). In light of these studies, it could be speculated that abiotic stress may promote *AhOMT* genes's expression, although more work is required for its confirmation. Micro-RNAs have got wide attention for their developmental and stress-tolerance roles. We identified miRNAs ahy-miR156a, ahy-miR167-3p, ahy-miR3513-5p, ahy-miR3521, ahy-miR156a, ahy-miR160-3p, ahy-miR3508, ahy-miR3513-3p, ahy-miR3518, ahy-miR3519 etc., targeting *AhOMTs* (Supplementary Table 5). ahy-miR3521 have been reported to target the *AhOPT3.2*, this gene is also targeted by ahy-miR156a. additionally ahy-miR156a also targets and down-regulates the *AhOPT3.3* and *AhOPT3.4*. ahy-miR167-3p targets *AhYSL3.2*, *AhYSL3.4*, and *AhYSL3.7*. all of these miRNAs downregulates their corresponding genes by mRNA cleavage (Wang et al., 2022). Our miRNAs prediction results also revealed their cleaving activity.

The *OMT* genes in plants have earlier been shown to be vital genes that regulate the expression of a protein necessary for development and growth (Zhang et al., 2021). Gene expression in different organs and tissues was investigated in this research. *AhOMTs* demonstrated diverse expressions in time- and space-defined manners. The expression differences in different tissues indicate the functional differences between *OMT* genes (Zhang et al., 2021). This research also demonstrated that the expression of these genes might be triggered by a certain environment or may be highly unique to a particular organ or developmental stage. Among various abiotic stresses, low temperature and drought stress significantly impair the plant growth and production (Raza et al., 2021b; Raza et al., 2022a; Raza et al., 2022b; Raza et al., 2023) Owing

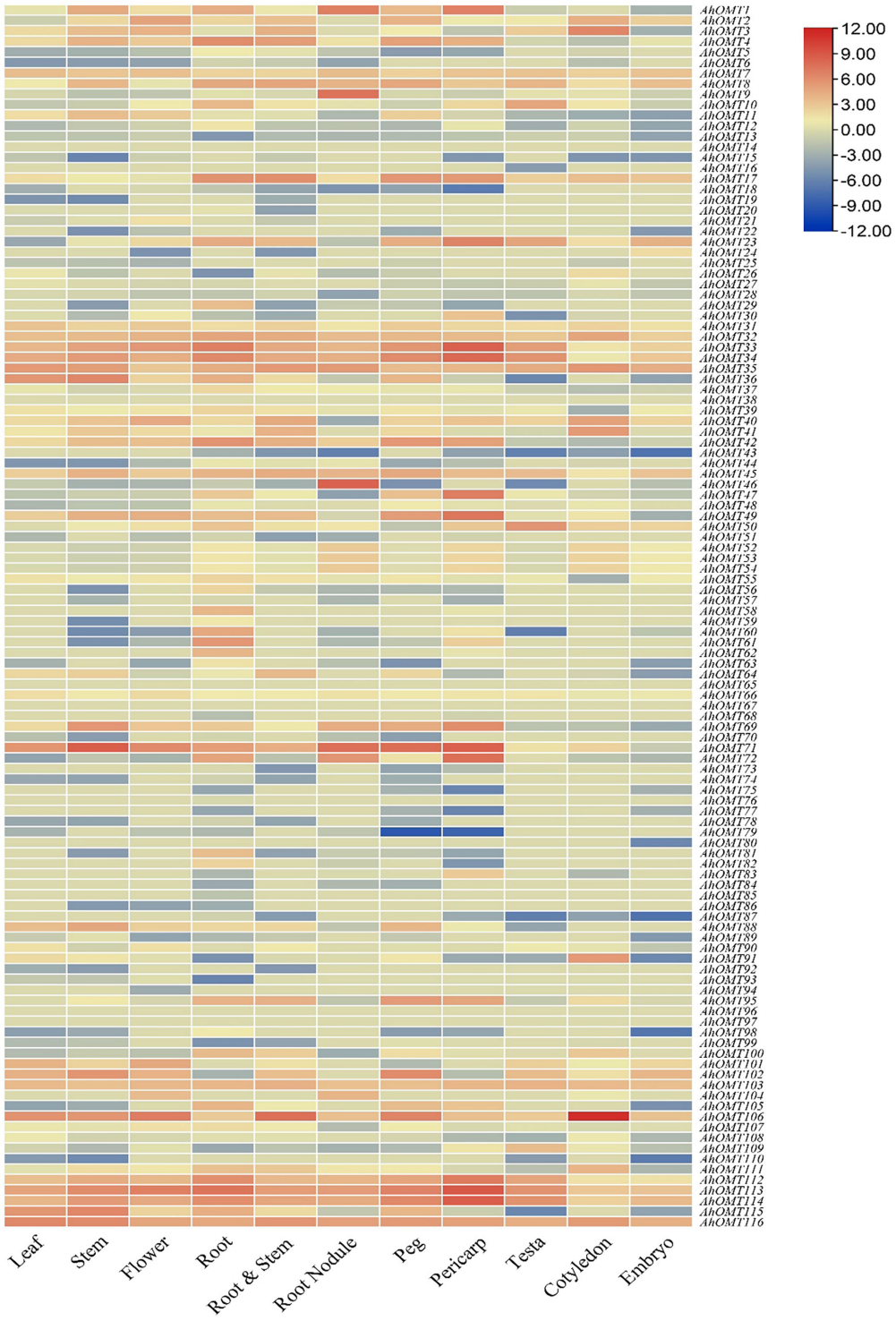


FIGURE 12
Transcriptome expression of *AhOMTs* in different tissues. *AhOMTs* possessed varying expression matrix in different tissues. All genes possessed varying levels of expression in different tissues.

to this, the *OMT* expression under these stressful environments was investigated. According to our findings, the expression of *AhOMT-7, AhOMT-33, AhOMT-34, AhOM-35, AhOMT-71, AhOM-93, AhOMT106, AhOMT-113*, and *AhOMT116* increased when exposed to low temperatures and hormones treatment. Under drought stress, some *AhOMTs* were up-regulated, and others were down-regulated. Our findings are in agreement with previous reports such as *OMTs* were upregulated in response to drought

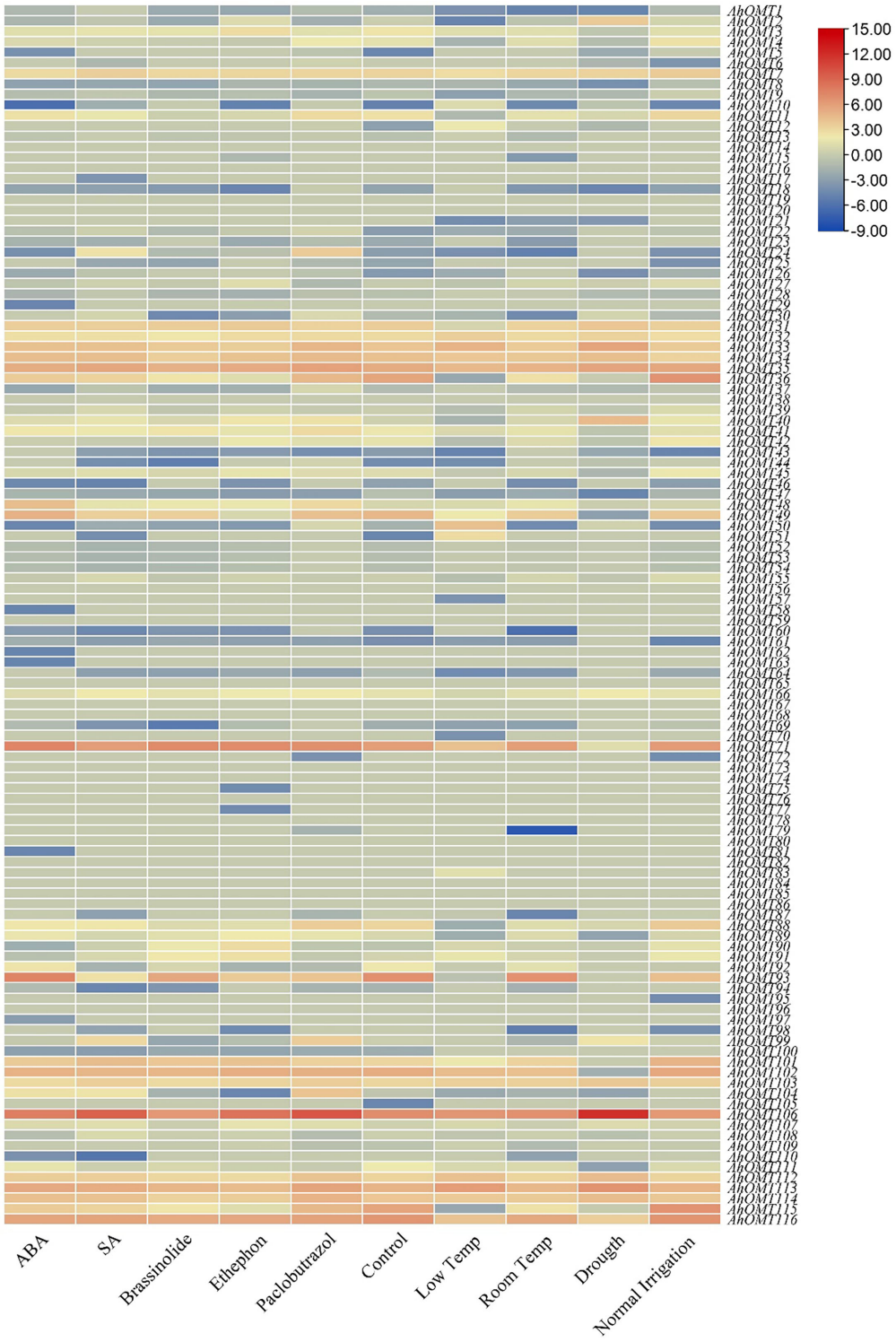


FIGURE 13
Transcriptome expression of *AhOMTs* under different hormones and stress conditions. Under the stress conditions, the *AhOMT106* gene was most active, while *AhOMT35*, *AhOMT71*, *AhOMT113* were also expressed in most of cases.

stress in grape barriers (Giordano et al., 2016) and down-regulated in *Brassica napus* (Li et al., 2016). In terms of the mechanism of this event, further research is needed in this area as well. In the near future, the integration of genomics and genome editing technologies could be coupled to improve the production of orphan crops including peanut (Yaqoob et al., 2023). As a result, evolutionary links, structure, and expression of *AhOMT* genes were thoroughly investigated in this work, revealing that these genes

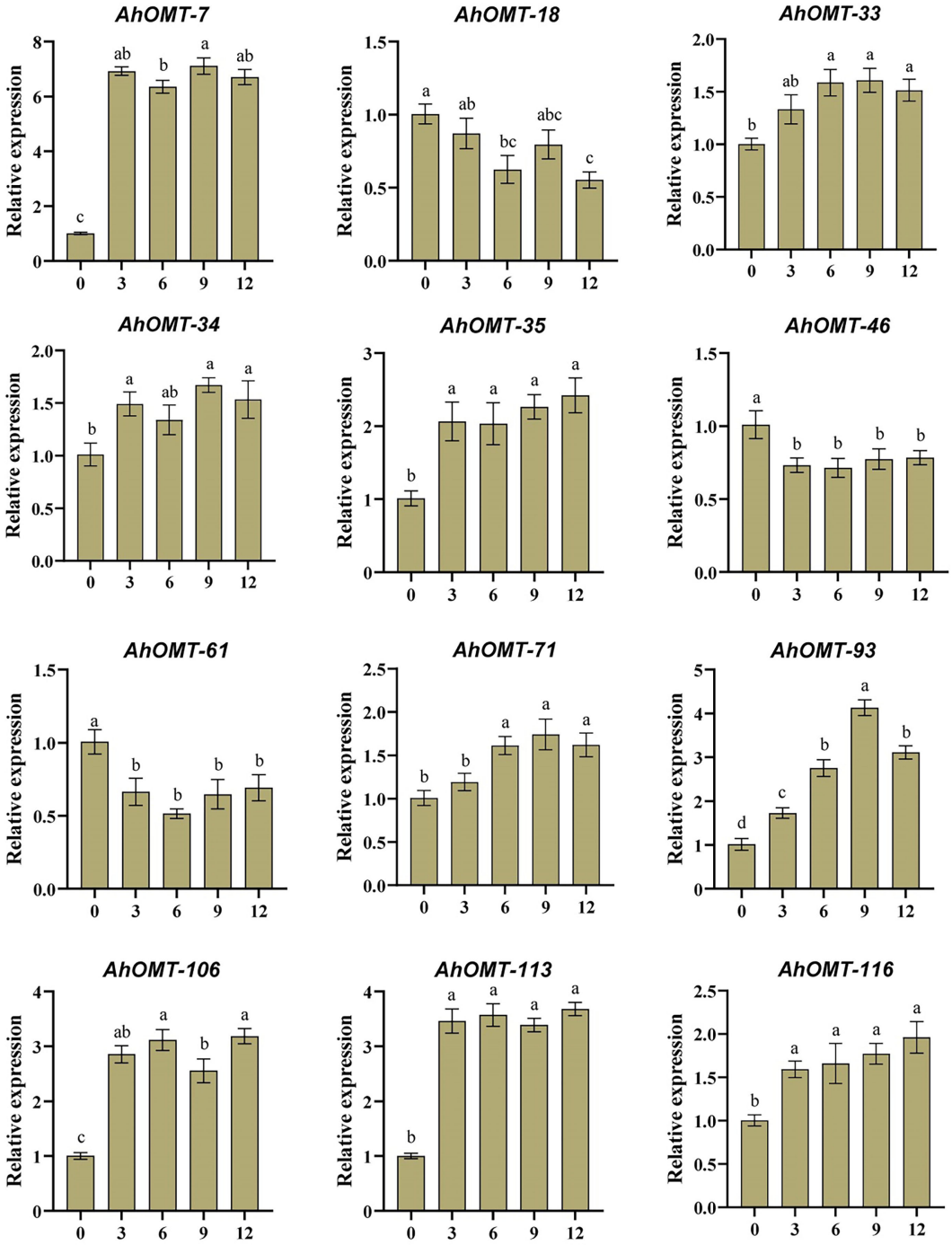


FIGURE 14
Expression profiling of *AhOMT* genes in response to ABA treatment. Mainly *AhOMT* genes recorded increased expression under ABA stress, while some genes were down-regulated. a, b, and c represents the significance levels among expression at different time points.

played a critical role in peanut stress tolerance and offered a theoretical basis for peanut breeding efforts.

Conclusion

This study identified 116 *OMT* genes in cultivated peanut. Sequentially to get well perceptive of the *AhOMT* genes, we conducted a wide range of genomic analyses, including evolutionary and genomic characterization, genes structural analysis, *cis*-acting regions, prediction of *miRNAs*, and conserved motifs analysis. A combination of gene structure and phylogenetic analysis revealed three main groups of *AhOMTs*. In addition, these genes' expression was profiled across different tissues against low temperature, hormones, and drought stress. Furthermore, the *AhOMT* genes expression demonstrated

10.3389/fpls.2023.1145624

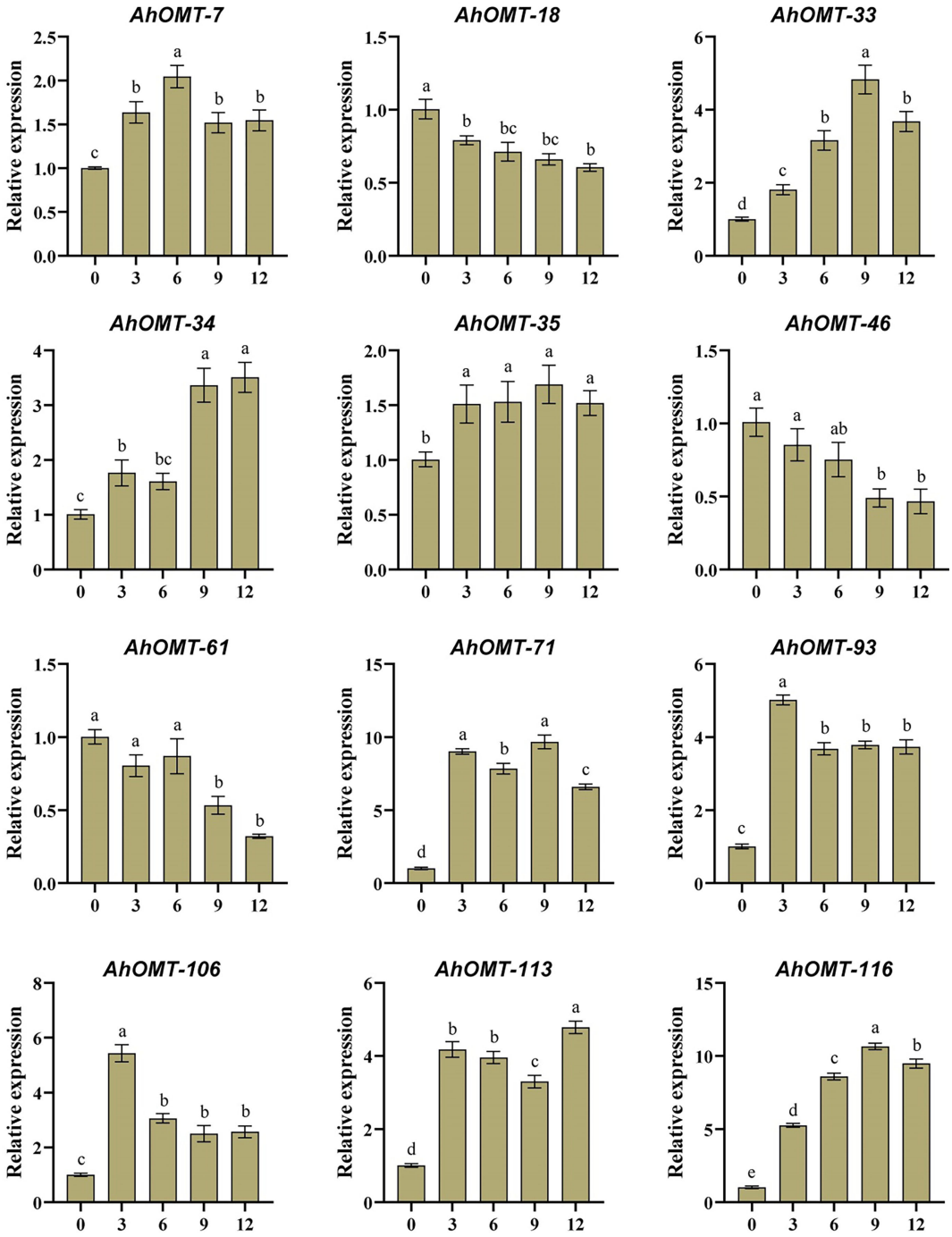


FIGURE 15
Expression profiling of *AhOMT* genes in response to cold stress. Mainly *AhOMT* genes recorded increased expression under cold stress, while some genes were down-regulated. a, b, and c represents the significance levels among expression at different time points.

that *AhOMT-7, AhOMT-33, AhOMT-34, AhOM-35, AhOMT-71, AhOM-93, AhOMT106, AhOMT-113*, and *AhOMT116* played a vital role against low temperature, hormones, and drought treatments. This study establishes the framework for future work into the functional study of *AhOMT* in peanut breeding programs.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA480120.

Author contributions

WZ and HC conceived the idea and designed the study. TC, YS, YZ, QY, and XC analyzed the data and wrote the manuscript. KC, YC, MG, HD, YP, AR, and CZ helped in literature search, revision, and provided technical guidance. WZ, HC, and YZ supervised the work and edited the final version. TC and YS equally contributed to the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation (NSFC) of China (U1705233, 32072103, 32272155, 31701463, and 31601337), the Science and Technology Foundation of Fujian Province of China (2021N5007 and 2017N0006), the Special Fund for Scientific and Technological Innovation of Fujian Agriculture and Forestry University (KFb22010XA and KFb22011XA), and Foreign Cooperation Projects for Fujian Academy of Agricultural Sciences (DWHZ2021-20).

Acknowledgments

Authors are thankful to Center of Legume Plant Genetics and system Biology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China for providing favorable conditions for this work and technical guidance.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1145624/full#supplementary-material

References

Argüello-Astorga, G., and Herrera-Estrella, L. (1998). Evolution of light-regulated plant promoters. *Annual Review of Plant Physiology and Plant Molecular Biology* 49, 525–555. doi: 10.1146/annurev.arplant.49.1.525

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research* 43, W39–W49. doi: 10.1093/nar/gkv416

Bellaloui, N. (2012). Soybean seed phenol, lignin, and isoflavones partitioning as affected by seed node position and genotype differences. *Food Nutr. Sci.* 3, 447. doi: 10.4236/fns.2012.34064

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of arachis duranensis and arachis ipaensis, the diploid ancestors of cultivated peanut. *Nature Genetics* 48, 438–446. doi: 10.1038/ng.3517

Bi, C., Chen, F., Jackson, L., Gill, B. S., and Li, W. J. (2011). Expression of lignin biosynthetic genes in wheat during development and upon infection by fungal pathogens. *Plant Molecular Biology Reporter* 29, 149–161. doi: 10.1007/s11105-010-0219-8

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi: 10.1146/annurev.arplant.54.031902.134938

Bout, S., and Vermerris, W. (2003). A candidate-gene approach to clone the sorghum brown midrib gene encoding caffeic acid O-methyltransferase. *Molecular Genetics and Genomics* 269, 205–214. doi: 10.1007/s00438-003-0824-4

Buer, C. S., Imin, N., and Djordjevic, M. A. (2010). Flavonoids: new roles for old molecules. *Journal of Intigrative Plant Biology* 52, 98–111. doi: 10.1111/j.1744-7909.2010.00905.x

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. J. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in arabidopsis thaliana. *BMC Plant Biology* 4, 1–21. doi: 10.1186/1471-2229-4-10

Chapman, M. A., Leebens-Mack, J. H., and Burke, J. M. (2008). Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Mol. Biol. Evol.* 25, 1260–1273. doi: 10.1093/molbev/msn001

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. *Molecular Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, H., Yang, Q., Chen, K., Zhao, S., Zhang, C., Pan, R., et al. (2019). Integrated microRNA and transcriptome profiling reveals a miRNA-mediated regulatory network of embryo abortion under calcium deficiency in peanut (Arachis hypogaea l.). *BMC Genomics* 20, 1–17. doi: 10.1186/s12864-019-5770-6

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

Davin, L. B., and Lewis, N. G. (1992). "Phenylpropanoid metabolism: Biosynthesis of monolignols, lignans and neolignans, lignins and suberins," in *Phenolic metabolism in plants* (New York: Springer), 325–375.

Donald, R., and Cashmore, A. R. (1990). Mutation of either G box or I box sequences profoundly affects expression from the arabidopsis rbcS-1A promoter. *The EMBO Journal* 9, 1717–1726. doi: 10.1002/j.1460-2075.1990.tb08295.x

Dudareva, N., and Pichersky, E. J. (2008). Metabolic engineering of plant volatiles. *Current Opinion in Biotechnology* 19, 181–189. doi: 10.1016/j.copbio.2008.02.011

Gao, Y., Li, J., Strickland, E., Hua, S., Zhao, H., Chen, Z., et al. (2004). An arabidopsis promoter microarray and its initial usage in the identification of HY5 binding targets *in vitro*. *Plant Molecular Biology* 54, 683–699. doi: 10.1023/B:PLAN.0000040898.86788.59

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. J. (2005). Protein identification and analysis tools on the ExPASy server. 571–607. doi: 10.1385/1-59259-890-0:571

Giordano, D., Provenzano, S., Ferrandino, A., Vitali, M., Pagliarani, C., Roman, F., et al. (2016). Characterization of a multifunctional caffeoyl-CoA O-methyltransferase activated in grape berries upon drought stress. *Plant Physiol. Biochem.* 101, 23–32. doi: 10.1016/j.plaphy.2016.01.015

Giuliano, G., Pichersky, E., Malik, V., Timko, M., Scolnik, P., and Cashmore, A. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *PNAS* 85, 7089–7093. doi: 10.1073/pnas.85.19.7089

Gonzales, M. D., Archuleta, E., Farmer, A., Gajendran, K., Grant, D., Shoemaker, R., et al. (2005). The legume information system (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Res.* 33, D660–D665. doi: 10.1093/nar/gki128

Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., et al. (2003). A new arabidopsis thaliana mutant deficient in the expression of O-methyltransferase

impacts lignins and sinapoyl esters. *Plant Molecular Biology* 51, 973–989. doi: 10.1023/A:1023022825098

Guo, D., Chen, F., Inoue, K., Blount, J. W., and Dixon, R. A. (2001). Downregulation of caffeic acid 3-o-methyltransferase and caffeoyl CoA 3-o-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and s lignin. *The Plant Cell* 13, 73–88. doi: 10.1105/tpc.13.1.73

Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B., and Douglas, C. (2007). Genome-wide analyses of phenylpropanoid-related genes in populus trichocarpa, arabidopsis thaliana, and oryza sativa: The populus lignin toolbox and conservation and diversification of angiosperm gene families. *Canadian Journal of Botany* 85, 1182–1201. doi: 10.1139/B07-098

Hu, D., Liu, X., Wang, C., Yang, H., Li, H., Ruan, R., et al. (2015). Expression analysis of key enzyme genes in lignin synthesis of culm among different lodging resistances of common buckwheat (Fagopyrum esculentum moench.). *Sci. Agric. Sin.* 48, 1864–1872. doi: 10.3864/j.issn.0578-1752.2015.09.20

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47, D309–D314. doi: 10.1093/nar/gky1085

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends in Genetics* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Joshi, C. P., and Chiang, V. L. (1998). Conserved sequence motifs in plant s-adenosyl-L-methionine-dependent methyltransferases. *Plant Molecular Biology* 37, 663–674. doi: 10.1023/A:1006035210889

Kim, J., Choi, B., Cho, B.-K., Lim, H.-S., Kim, J. B., Natarajan, S., et al. (2013). Molecular cloning, characterization and expression of the caffeic acid O-methyltransferase (COMT) ortholog from kenaf (Hibiscus cannabinus). *Plant Omics* 6, 246–253.

Kolkman, J. A., and Stemmer, W. P. (2001). Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* 19, 423–428. doi: 10.1038/88084

Kota, P., Guo, D., Zubieta, C., Noel, J., and Dixon, R. A. (2004). O-Methylation of benzaldehyde derivatives by "lignin specific" caffeic acid 3-o-methyltransferase. *Phytochemistry* 65, 837–846. doi: 10.1016/j.phytochem.2004.01.017

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. J. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35, 1547. 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, D1202–D1210. doi: 10.1093/nar/gkr1090

Lavin, M., Herendeen, P. S., and Wojciechowski, M. F. (2005). Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology* 54, 575–594. doi: 10.1080/10635150590947131

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

Li, W., Lu, J., Lu, K., Yuan, J., Huang, J., Du, H., et al. (2016). Cloning and phylogenetic analysis of brassica napus l. caffeic acid O-methyltransferase 1 gene family and its expression pattern under drought stress. *PloS One* 11, e0165975. doi: 10.1007/s11103-006-0029-4

Li, H. M., Rotter, D., Hartman, T. G., Pak, F. E., Havkin-Frenkel, D., and Belanger, F. C. (2006). Evolution of novel O-methyltransferases from the vanilla planifolia caffeic acid O-methyltransferase. *Plant Molecular Biology* 61, 537–552. doi: 10.1584/jpestics.31.47

Li, J.-J., Zhang, G., Yu, J.-h., Li, Y.-y., Huang, X.-h., Wang, W.-J., et al. (2015). Molecular cloning and characterization of caffeic acid 3-o-methyltransferase from the rhizome of ligusticum chuanxiong. *Biotechnology Letters* 37, 2295–2302. doi: 10.1007/s10529-015-1917-y

Lin, F., Yamano, G., Hasegawa, M., Anzai, H., Kawasaki, S., and Kodama, O. (2006). Cloning and functional analysis of caffeic acid 3-o-methyltransferase from rice (Oryza sativa). *Journal of Pesticide Science* 31, 47–53. doi: 10.1584/jpestics.31.47

Lin, S.-J., Yang, Y.-Z., Teng, R.-M., Liu, H., Li, H., and Zhuang, J. (2021). Identification and expression analysis of caffeoyl-coenzyme a O-methyltransferase family genes related to lignin biosynthesis in tea plant (Camellia sinensis). *Protoplasma* 258, 115–127. doi: 10.1007/s00709-020-01555-4

LIU, S.-j., HUANG, Y.-h., HE, C.-j., Cheng, F., and ZHANG, Y.-w. (2016a). Cloning, bioinformatics and transcriptional analysis of caffeoyl-coenzyme a 3-o-methyltransferase in switchgrass under abiotic stress. *Journal of Integrative Agriculture* 15, 636–649. doi: 10.1016/S2095-3119(16)61363-1

Liu, X., Luo, Y., Wu, H., Xi, W., Yu, J., Zhang, Q., et al. (2016b). Systematic analysis of O-methyltransferase gene family and identification of potential members involved in the formation of O-methylated flavonoids in citrus. *Gene* 575, 458–472. doi: 10.1016/j.gene.2015.09.048

Liu, Q., Luo, L., and Zheng, L. (2018). Lignins: biosynthesis and biological functions in plants. *International Journal of Molecular Sciences* 19, 335. doi: 10.3390/ijms19020335

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

Ma, Q.-H. (2009). The expression of caffeic acid 3-o-methyltransferase in two wheat genotypes differing in lodging resistance. *Journal of Experimental Botany* 60, 2763–2771. doi: 10.1093/jxb/erp132

Ma, Q.-H., and Luo, H.-R. (2015). Biochemical characterization of caffeoyl coenzyme a 3-o-methyltransferase from wheat. *Planta* 242, 113–122. doi: 10.1007/s00425-015-2295-3

Moura, J. C. M. S., Bonine, C. A. V., de Oliveira Fernandes Viana, J., Dornelas, M. C., and Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of Integrative Plant Biology* 52, 360–376. doi: 10.1111/j.1744-7909.2010.00892.x

Nguyen, T.-N., Son, S., Jordan, M. C., Levin, D. B., and Ayele, B. T. (2016). Lignin biosynthesis in wheat (Triticum aestivum l.): Its response to waterlogging and association with hormonal levels. *BMC Plant Biol.* 16, 1–16. doi: 10.1186/s12870-016-0717-4

Peng, D., Chen, X., Yin, Y., Lu, K., Yang, W., Tang, Y., et al. (2014). Lodging resistance of winter wheat (Triticum aestivum l.): Lignin accumulation and its related enzymes activities due to the application of paclobutrazol or gibberellin acid. *Field Crops Research* 157, 1–7. doi: 10.1016/j.fcr.2013.11.015

Rakoczy, M., Femiak, I., Alejska, M., Figlerowicz, M., and Podkowinski, (2018). Sorghum CCoAOMT and CCoAOMT-like gene evolution, structure, expression and the role of conserved amino acids in protein activity. *Mol. Genet. Genomics* 293, 1077–1089. doi: 10.1007/s00438-018-1441-6

Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., et al. (2004). Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochemistry Review* 3, 29–60. doi: 10.1023/B:PHYT.0000047809.65444.a4

Raza, A., Charagh, S., Abbas, S., Hassan, M. U., Saeed, F., Haider, S., et al. (2023). Assessment of proline function in higher plants under extreme temperatures. *Plant Biol.* doi: 10.1111/plb.13510

Raza, A., Charagh, S., García-Caparrós, P., Rahman, M. A., Ogwugwa, V. H., Saeed, F., et al. (2022a). Melatonin-mediated temperature stress tolerance in plants. *GM Crops Food* 13, 196–217. doi: 10.1080/21645698.2022.2106111

Raza, A., Mubarik, M. S., Sharif, R., Habib, M., Jabeen, W., Zhang, C., et al. (2022b). Developing drought-smart, ready-to-grow future crops. *Plant Genome*, e20279. doi: 10.1002/tpg2.20279

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. *Int. J. Mol. Sci.* 22, 4281. doi: 10.3390/ijms22084281

Raza, A., Tabassum, J., Kudapa, H., and Varshney, R. K. (2021b). Can omics deliver temperature resilient ready-to-grow crops? *Crit. Rev. Biotechnol.* 41, 1209–1232. doi: 10.1080/07388551.2021.1898332

Roje, S. (2006). S-Adenosyl-L-methionine: beyond the universal methyl group donor. *Phytochemistry* 67, 1686–1698. doi: 10.1016/j.phytochem.2006.04.019

Savard, O. T., Bertrand, D., and El-Mabrouk, N. (2011). "Evolution of orthologous tandemly arrayed gene clusters," in *BMC bioinformatics*, 12, 1–12. (Galway, Ireland: BioMed Central)

Sharif, Y., Chen, H., Deng, Y., Ali, N., Khan, S., Zhang, C., et al. (2022). Cloning and functional characterization of a pericarp abundant expression promoter (AhGLP17-1P) from peanut (Arachis hypogaea l.). *Front. Genet.* 12. doi: 10.3389/fgene.2021.821281

Song, S., Zhou, H., Sheng, S., Cao, M., Li, Y., and Pang, X. (2017). Genome-wide organization and expression profiling of the SBP-box gene family in Chinese jujube (Ziziphus jujuba mill.). 18, 1734. doi: 10.3390/ijms18081734

Struck, A. W., Thompson, M. L., Wong, L. S., and Micklefield, J. (2012). S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* 13, 2642–2655. doi: 10.1002/cbic.201200556

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

Wang, C., Wang, X., Li, J., Guan, J., Tan, Z., Zhang, Z., et al. (2022). Genome-wide identification and transcript analysis reveal potential roles of oligopeptide transporter genes in iron deficiency induced cadmium accumulation in peanut. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.894848

Wang, Y., Wang, X., and Paterson, A. H. (2012). Genome and gene duplications and gene expression divergence: a view from plants. *Ann. New York Acad. Sci.* 1256, 1–14. doi: 10.1111/j.1749-6632.2011.06384.x

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

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, 1063–1081. doi: 10.2478/s11756-019-00397-7

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 47, W52–W58. doi: 10.1093/nar/gkz333

Yang, Z., and Bielawski, J. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* 15, 496–503. doi: 10.1016/S0169-5347(00)01994-7

Yaqoob, H., Tariq, A., Bhat, B. A., Bhat, K. A., Nehvi, I. B., Raza, A., et al. (2023). Integrating genomics and genome editing for orphan crop improvement: a bridge between orphan crops and modern agriculture system. *GM Crops Food* 14, 1–20. doi: 10.1080/21645698.2022.2146952

Ye, Z.-H., Kneusel, R. E., Matern, U., and Varner, J. (1994). An alternative methylation pathway in lignin biosynthesis in zinnia. *The Plant Cell* 6, 1427–1439. doi: 10.1105/tpc.6.10.1427

Yoshihara, N., Fukuchi-Mizutani, M., Okuhara, H., Tanaka, Y., and Yabuya, T. (2008). Molecular cloning and characterization of O-methyltransferases from the flower buds of iris hollandica. *Journal of Plant Physiology* 165, 415–422. doi: 10.1016/j.jplph.2006.12.002

Yu, C. S., Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics* 64, 643–651. doi: 10.1002/prot.21018

Yuan, Z., Fang, Y., Zhang, T., Fei, Z., Han, F., Liu, C., et al. (2018). The pomegranate (Punica granatum l.) genome provides insights into fruit quality and ovule developmental biology. *Plant Biotechnology Journal* 16, 1363–1374. doi: 10.1111/pbi.12875

Zhang, X., Chen, B., Wang, L., Ali, S., Guo, Y., Liu, J., et al. (2021). Genome-wide identification and characterization of caffeic acid O-methyltransferase gene family in soybean. *Plants* 10, 2816. doi: 10.3390/plants10122816

Zhang, J., Li, Y., Jia, H.-X., Li, J.-B., Huang, J., Lu, M.-Z., et al. (2015). The heat shock factor gene family in salix suchowensis: a genome-wide survey and expression profiling during development and abiotic stresses. *Frontiers in Plant Science* 6, 748. doi: 10.3389/fpls.2015.00748

Zhou, J.-M., Seo, Y. W., and Ibrahim, R. K. (2009). Biochemical characterization of a putative wheat caffeic acid O-methyltransferase. *Plant Physiology and Biochemistry* 47, 322–326. doi: 10.1016/j.plaphy.2008.11.011

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics* 51, 865–876. doi: 10.1038/s41588-019-0402-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 May 2023
DOI 10.3389/fpls.2023.1153293

OPEN ACCESS

EDITED BY
Rajeev K. Varshney,
Murdoch University, Australia

REVIEWED BY
Sunil S. Gangurde,
University of Georgia, United States
Aamir W. Khan,
University of Missouri, United States

*CORRESPONDENCE
Dongmei Bai
baidm1221@163.com
Boshou Liao
lboshou@hotmail.com

RECEIVED 29 January 2023
ACCEPTED 28 March 2023
PUBLISHED 08 May 2023

CITATION
Zhang X, Zhang X, Wang L, Liu Q, Liang Y, Zhang J, Xue Y, Tian Y, Zhang H, Li N, Sheng C, Nie P, Feng S, Liao B and Bai D (2023) Fine mapping of a QTL and identification of candidate genes associated with cold tolerance during germination in peanut (*Arachis hypogaea* L.) on chromosome B09 using whole genome re-sequencing.
Front. Plant Sci. 14:1153293.
doi: 10.3389/fpls.2023.1153293

Fine mapping of a QTL and identification of candidate genes associated with cold tolerance during germination in peanut (*Arachis hypogaea* L.) on chromosome B09 using whole genome re-sequencing

Xin Zhang[1,2], Xiaoji Zhang[3], Luhuan Wang[3], Qimei Liu[4], Yuying Liang[3], Jiayu Zhang[3], Yunyun Xue[1], Yuexia Tian[1], Huiqi Zhang[1], Na Li[1], Cong Sheng[5], Pingping Nie[6], Suping Feng[7], Boshou Liao[8]* and Dongmei Bai[1]*

[1]Institute of Industrial Crops, Shanxi Agricultural University, Taiyuan, China, [2]State Key Laboratory of Sustainable Dryland Agriculture, Shanxi Agricultural University, Taiyuan, China, [3]College of Agronomy, Shanxi Agricultural University, Taigu, China, [4]College of Plant Protection, Shanxi Agricultural University, Taigu, China, [5]Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [6]College of Life Sciences, Zaozhuang University, Zaozhuang, China, [7]College of Food Science and Engineering, Hainan Tropical Ocean College, Hainan, China, [8]The Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China

Low temperatures significantly affect the growth and yield of peanuts. Temperatures lower than 12 °C are generally detrimental for the germination of peanuts. To date, there has been no report on precise information on the quantitative trait loci (QTL) for cold tolerance during the germination in peanuts. In this study, we developed a recombinant inbred line (RIL) population comprising 807 RILs by tolerant and sensitive parents. Phenotypic frequencies of germination rate low-temperature conditions among RIL population showed normally distributed in five environments. Then, we constructed a high density SNP-based genetic linkage map through whole genome re-sequencing (WGRS) technique and identified a major quantitative trait locus (QTL), qRGRB09, on chromosome B09. The cold tolerance-related QTLs were repeatedly detected in all five environments, and the genetic distance was 6.01 cM (46.74 cM - 61.75 cM) after taking a union set. To further confirm that qRGRB09 was located on chromosome B09, we developed Kompetitive Allele Specific PCR (KASP) markers for the corresponding QTL regions. A regional QTL mapping analysis, which was conducted after taking the intersection of QTL intervals of all environments into account, confirmed that qRGRB09 was between the KASP markers, G22096 and G220967 (chrB09:155637831–155854093), and this

region was 216.26 kb in size, wherein a total of 15 annotated genes were detected. This study illustrates the relevance of WGRS-based genetic maps for QTL mapping and KASP genotyping that facilitated QTL fine mapping of peanuts. The results of our study also provided useful information on the genetic architecture underlying cold tolerance during germination in peanuts, which in turn may be useful for those engaged in molecular studies as well as crop improvement in the cold-stressed environment.

KEYWORDS

peanut, whole genome re-sequencing, cold tolerance, germination, QTL, candidate genes

1 Introduction

Peanut crops (*Arachis hypogaea* L.), which are cultivated for their usefulness as oil plants and cash crops, play an essential role as a source of edible vegetable oil as well as a leisure food. However, in high-latitude or high-altitude areas, also termed "cold areas," coldness is an important stressor that limits the growth as well as yield of crops, including peanuts. Reportedly, global agricultural production, with the exception of typically tropical areas, may be affected by various cold-induced adverse effects, that may limit the planting range of crops, reduce crop yield and quality, and even cause crop failure. Annually estimated worldwide crop losses caused by cold stress amount to hundreds of billions of dollars (Jeon and Kim, 2013; Barrero-Gil and Salinas, 2018). Therefore, improving cold tolerance in peanuts for the betterment of peanut crops deserves to be considered as an urgent issue that requires a rapid resolution.

As thermophilic crops, peanuts require relatively high temperatures throughout their development (Wang et al., 2003). Cold injury causes different degrees of damage to peanuts during germination, seedling emergence, flowering, maturity, and other key growth stages (Zhang X, et al., 2022). Among these, damage at the germination stage is known to be a common occurrence. Shorter durations of cold stress delay the emergence of seedlings, while slightly longer durations may seriously damage the viability of germinated seeds, causing mold invasion, necrosis, and seed rotting, resulting in a lack of seedlings and ridge cutting, which seriously affects yield and quality (Bai et al., 2018). In view of the harmful effects exerted by cold stress on peanuts, with particular reference to peanut cultivation in cold regions, developing high-yielding peanut varieties with strong cold tolerance may help broaden planting areas, improve yield per unit area, and ensure product quality.

Cold tolerance of crops refers to the adaptability of crops to low-temperature environments, including tolerance to cold stress and the ability to quickly resume growth, once cold stress is removed. At present, studies investigating cold tolerance in germinating peanuts are mainly focused on the identification and screening of different germplasms for cold tolerance, differential physio-biochemical changes, and the metabolic response of peanuts to cold stress (Upadhyaya et al., 2009; Bai et al., 2018; Patel et al., 2022). The research methods used by these studies are mainly centered on phenotype identification as well as physiological and biochemical analysis. However, the advances in molecular biology-related technologies have caused some studies to shift their focus onto preliminary research on peanut cold tolerance at the molecular level. Wang et al. (2013) analyzed differentially expressed genes (DEGs) in peanuts subjected to normal and low-temperature treatments and obtained two genes that are specifically expressed in relation to cold tolerance at the germination stage. Bai et al. (2018) used 90 pairs of simple sequence repeat (SSR) polymorphic primers to evaluate the cold tolerance of 72 peanut varieties at the germination stage and found a wide variety of variations. Chang et al. (2019) screened four cold-tolerant and four cold-susceptible cultivars, for indicators of seed germination and physiological indices of seedlings at 4 °C and found that the contents of soluble sugar, soluble protein, and free proline were mildly reduced in cultivars with strong cold tolerance. However, elucidation of peanut cold tolerance is easily affected by identification methods, environment, season, climate, and other factors. In general, existing research, which is centered on germplasm exploration and genetic traits associated with cold tolerance at the germination stage, appears to lack systematism, and has failed to either produce any breakthroughs, or fine map quantitative trait locus (QTLs), or identify candidate genes, thereby severely limiting the progress of cold tolerance breeding in peanuts.

Effective molecular marker-assisted breeding for cold tolerance in peanuts involves identifying QTLs closely linked to target traits. At present, some progress has been made in QTL mapping of cold tolerance during the germination stage, in crops such as rice (Shakiba et al., 2017; Yang et al., 2020), maize (Hu et al., 2016; Li et al., 2018), and sorghum (Knoll et al., 2008; Upadhyaya et al., 2016). Continuous and rapid improvements in high-throughput sequencing technology have resulted in peanut genomic research making rapid progress. Genome sequence analysis of two wild diploid peanuts (*A. duranensis* and *A. ipaensis*) was completed in 2016 (Bertioli et al., 2016; Chen et al., 2016). In 2018, the genome sequencing analysis of *A. monticola*, an allotetraploid wild peanut, was completed (Yin et al., 2018); In 2019, a major breakthrough was made in the genome sequencing of cultivated heterotetraploid

peanuts. Genomic data of the Chinese peanut varieties, *A. hypogaea* cv. Shitouqi (Zhuang et al., 2019) and *A. hypogaea* cv. Fuhuasheng (Chen et al., 2019), and the American peanut variety, *A. hypogaea* cv. Tifrinner (Bertioli et al., 2019) have been published, and these are widely used in relevant research. These advances have greatly enriched genomic information pertaining to peanuts, and moreover, various high-throughput genotyping techniques (Semagn et al., 2014; Clevenger et al., 2017; Guo et al., 2019) have made continuous progress by effectively promoting the discovery and localization of QTLs of several peanut traits. Liu et al. (2020) identified three main QTLs that regulate oil content. Khedikar et al. (2017) obtained the main QTL related to the fruit-setting number per plant, productivity, 100 fruit weight and other traits. Luo et al. (2018) successively detected major QTLs with stable pod size and kernel yield using a recombinant inbred line (RIL) population, and subsequently developed efficient Kompetitive Allele Specific PCR (KASP) markers. Huang et al. (2016) used SSR and transposon markers to map the main QTL related to aflatoxin resistance. Luo et al. (2019) identified the main QTL associated with bacterial wilt resistance in multiple environments. Han S, et al. (2018) successfully identified QTLs regulating resistance to early and late leaf spot diseases, while Agarwal et al. (2019) identified a major QTL associated with resistance to the tomato spotted wilt virus in peanuts. Thus, current progress in the development of molecular markers for important peanut traits, mentioned above, has provided a theoretical basis as well as a technical guarantee for those exploring QTLs associated with cold tolerance in peanuts. Further, RIL populations may be a better option for generating replicated and multi-environmental phenotyping data in variable space and time, that may help decipher environmental effects on target traits. Such phenotyping data may enable the detection of stable and consistent QTLs, as well as the subsequent detection of linked markers that may be useful for breeding purposes. Based on the above considerations, this study describes the development of a RIL population that can be utilized for genotyping and multi-environment phenotyping data linked to cold tolerance during germination in peanuts. The results of our study may expectedly help develop a better understanding of trait locus mapping, leading to the successful identification of genomic regions regulating cold tolerance.

2 Materials and methods

2.1 Plant materials and development of RIL population for cold tolerance

The RIL population developed from DF12 (cold-susceptible, male parents) × Huayu 44 (cold-resistant, female parents) at the Institute of Industrial Crops, Shanxi Agricultural University, Taiyuan, China, comprised 807 individuals RILs. Huayu 44, an inter-group hybrid cultivar of the peanut genus, was bred by Shandong Peanut Research Institute (CAAS) using the incompatible wild species *Arachis glablata* Benth. DF12 is a breeding line with high oleic acid content and derived from Kainongxuan 01-6 and Baisha 1016 cross and was bred by the Industrial Crops Research Institute of the Henan Academy of Agricultural Sciences. The single seed descent (SSD) method was used until the F_8 generation to develop the RIL population. The entire RIL population and their parents were planted in five different environments. Those grown at the Le-Dong (LD) experimental station (N1843′ and E10900′) in 2019 and 2021 were referred to as E1 and E2, respectively. In 2020, those used for the experiment conducted at the Fen-Yang (FY) experimental station (N3742′ and E11179′), were referred to as E3. In 2021, the experiments were conducted at the FY and Nan-Bin (NB) experimental stations (N1838′ and E10921′), and those grown there were referred to as E4 and E5, respectively (Figure 1).

2.2 Phenotyping for cold tolerance-related traits

Based on previous identification techniques and standards for peanut cold tolerance at the germination stage established by us (Bai et al., 2022), surface-sterilized seeds were incubated in water-soaked filter paper in Petri dishes, each of which contained 20 seeds. Germination experiments were performed in a growth chamber at a relative humidity of 70% and a temperature of 12°C for 3 d, then at 2°C for 3 d, and finally at 25°C for 3 d to recovery treatment. Another set was grown in a growth chamber at 25°C, as a control. The seeds were tested three times each time for replication purposes. The average value of three replicates was used as the statistical unit. The factors measured were defined as follows:

$$\textit{Germination standard} = \textit{radicle length equal to or greater than seed length}; \quad (1)$$

$$\textit{Germination rate (GR)} = \textit{the number of normally germinated seeds on day 4 at 25 °C} / \textit{number of tested seeds} \times 100\,\%; \quad (2)$$

$$\textit{Relative germination rate (RGR)} = \textit{germination rate after low temperature treatment} / \textit{control germination rate} \times 100\,\%. \quad (3)$$

SPSS 25.0 was used to analyze each variation coefficient and descriptive statistics.

2.3 DNA extraction and whole genome re-sequencing

Seed buds of the RIL population were collected and used to extract high-quality genomic DNA using the CTAB method with some modifications (Tamari et al., 2013). DNA quality and quantity were checked on 0.8% agarose using a Qubit 2.0 fluorometer (Thermo Fisher, USA). Whole genome re-sequencing (WGRS) was performed for two hundred RILs randomly selected from the 807 individuals for simultaneous single nucleotide polymorphisms (SNPs) discovery and

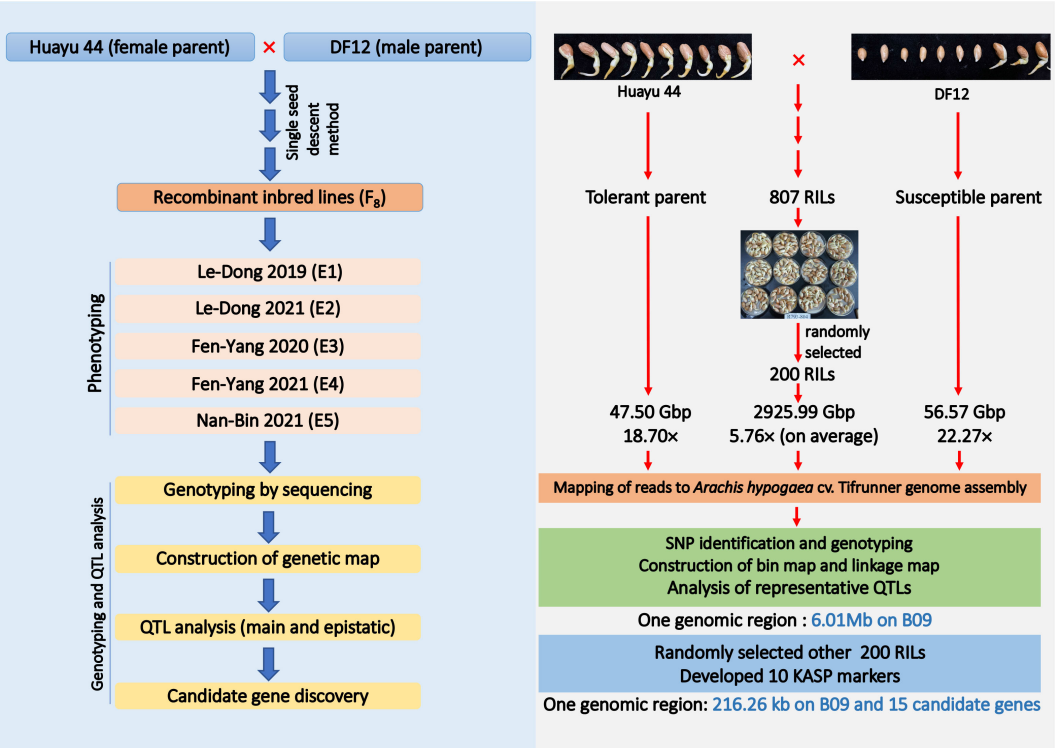


FIGURE 1
Flow chart for population development, high-density genotyping and multi-environments phenotyping.

genotyping of the mapping population. For this purpose, the DNA of each sample was fragmented into approximately 350 bp DNA fragments using a Covaris S2/E210 focused ultrasonicator (Covaris, Woburn, MA, United States). Sheared DNA was end-repaired and a single nucleotide (A) overhang was subsequently added to the repaired fragments using Klenow Fragment (3′→ 5′ exo–) (NEB, United States) and dATP at 37°C, following which barcodes and Illumina sequencing adapters were ligated to the A-tailed fragments using T4 DNA ligase (Thermo Fisher Scientific, Inc., USA). Sequence depth of the two parents was approximately 20X, while that of 200 RIL was approximately 5X on average. Polymerase chain reaction (PCR) was performed using diluted shearing-ligation DNA samples, dNTP, Q5® High-Fidelity DNA Polymerase, and PCR primers. The PCR products were then purified using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) and pooled. Next, the pooled samples were separated using 2% agarose gel electrophoresis. Fragments of 350 bp (with barcodes and adaptors) were excised and purified using a QIAquick gel extraction kit (Qiagen, Hilden, Germany). The gel-purified products were then diluted for pair-end sequencing (each end 150 bp) on an Illumina Novaseq 6000 system using a standard protocol (Illumina, Inc., San Diego, CA, USA).

2.4 SNP identification and genotyping

Low-quality reads (quality score< 20e) were filtered out, and raw reads were sorted into each progeny according to barcode sequences. After the barcodes were trimmed from each high-quality read, clean reads from the same sample were mapped to the cultivated peanut genome sequence (*A. hypogaea* cv. Tifrunner; https://peanutbase.org/) using Burrows-Wheeler Aligner software v.0.6.1 (Li et al., 2009). Samtools v.1.15.1 (Li and Durbin, 2009) was used for mark duplicates, while GATK v.4.1.9 (McKenna et al., 2010) was used for local realignment and base recalibration. An SNP set was formed by combining GATK and Samtools (Li and Durbin, 2009) SNP calling analysis with default parameters. SNPs identified between parents were homozygous, and only biallelic SNPs were regarded as polymorphic, and the marker integrity over 80% was maintained for bin calling later. Genotypes of the RIL were recognized at this SNP site.

2.5 Bin map and linkage map construction

Based on the SNP VCF matrix, SNPBinner v.0.1.1 software (Gonda et al., 2019) which uses the Hidden Markov Model (HMM) method to calculate recombination breakpoints, was employed to build co-separation bin markers. Based on their location in the peanut genome, marker loci were partitioned primarily into linkage groups (LGs). The linkage map was constructed from recombination bins serving as genetic markers using Icimapping v.4.1 using model 2-opt, and window size=5 for rippling (Meng et al., 2015). Map distances were estimated using the Kosambi mapping function (Kosambi, 2012). The genetic linkage map, based on the reference genome, was visualized using the R package in LinkageMapView v.2.1.2 (Ouellette et al., 2018). All sequences of

bin markers that were constructed in the linkage map were aligned back to the physical genome sequence of Tifrunner using BLAT software v.36x5 (Kent, 2002) to confirm their physical positions in the genome. The Spearman correlation coefficient was calculated in order to evaluate the collinearity between the genetic map and the reference genome.

2.6 QTL analysis and fine mapping *via* KASP-based genotyping

QTLs were identified using an interval mapping model implemented *via* the Complex interval mapping (CIM) in R/qtl package (Arends et al., 2010). The logarithm of odds (LOD) threshold was determined by applying 1000 permutation tests with 1% probability for each trait. Subsequently, KASP genotyping was performed, and 10 SNP markers were developed inside the main QTL as well as on its sides. Two hundred RILs from 807 lines that did not participate in WGRS were randomly selected and genotyped. Based on genotyping results and their corresponding RGR phenotypes, a regional genetic map and QTL analysis were performed using IciMapping v.4.1 (https://www.isbreeding.net/), according to Liu's method (Liu et al., 2019).

3 Results

3.1 Phenotypic variation of cold tolerance-related traits in the RILs and their parents

The relative germination rates of two hundred RILs and their parents from five environments were tested, respectively. The parents "Huayu 44" and "DF12" showed distinct differences in cold tolerance (Figure 2A; Table S1). The radicle lengths of almost all tested Huayu 44's seeds were greater than their seed lengths. By contrast, the radicle lengths of most DF12's seeds were either equal to or less than their seed lengths, while some seeds failed to germinate. The average relative germination rates of Huayu 44 and DF12 were 97.42%, and 38.63%, respectively. The range of variation in the relative germination rates of the RIL populations was large (0–96.55%) (Table S1). The types of population variation were relatively rich, with a coefficient of variation of more than 33%. Instrument-measured traits in all five environments followed normally distributed (Figure 2B; Table S2).

3.2 Variation calling and annotation

Whole genome re-sequencing of the female parent, male parent and all the two hundred individuals generated a total 47.50, 56.57, and 2925.99 Gb of clean data with Q30 ≥ 84%, representing, on average, sequencing depths of 18.70×, 22.27×, and 5.76×, respectively (Figure 1; Table S3). All clean reads were mapped to the genome of the cultivated peanut Tifrunner, where the mapping rates were 99.07% for Huayu 44, 98.80% for DF12 and 99.35% for the offspring on average. A total of 447,528 SNPs and 220,443 InDels were called between the parents (Figure 3A). Most SNPs (88.57%) were annotated in the intergenic regions, whereas 1.52% were annotated in the exon regions (Figure 3B). With respect to InDels, 77.41% were annotated in intergenic regions while 0.61% were located in exon regions (Figure 3C). This map lays a foundation for further bin establishment and candidate genes speculation.

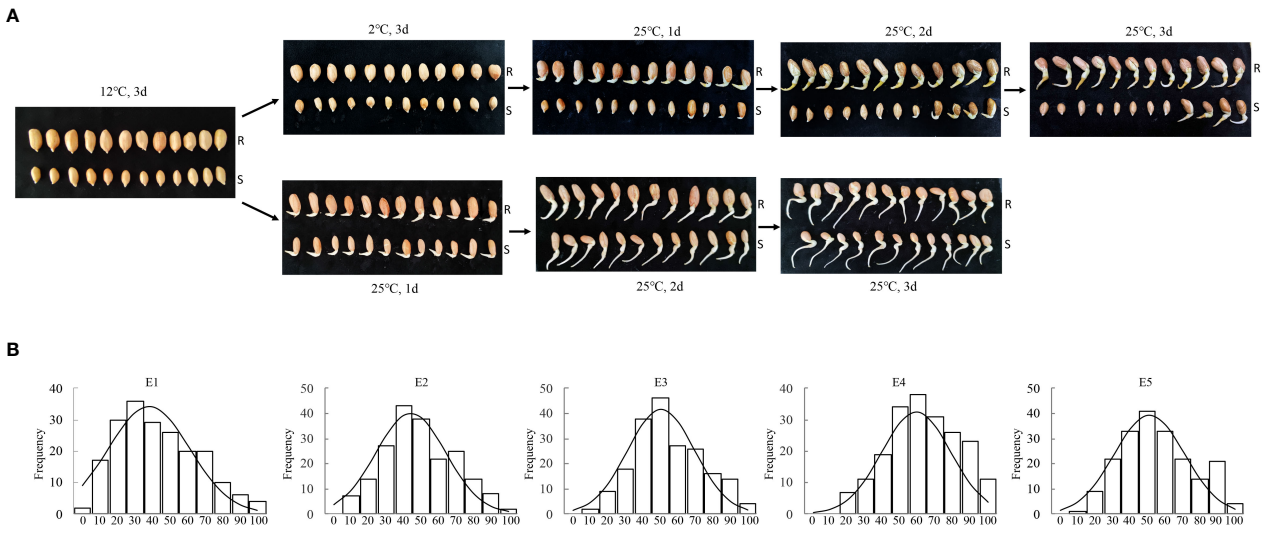


FIGURE 2
Phenotypic variation of cold tolerance-related Traits in RILs and their parents. **(A)** Germination experiments were performed in a growth chamber at a relative humidity of 70% and a temperature of 12 °C for 3 d, then at 2 °C for 3 d, and finally at 25 °C for 3 d to recovery treatment. The parents showed distinct differences in cold tolerance. The radicle lengths of almost all tested Huayu 44's seeds were greater than their seed lengths. By contrast, the radicle lengths of most DF12's seeds were either equal to or less than their seed lengths, while some seeds failed to germinate. R, female parent Huayu 44; S, male parent DF12. **(B)** Frequency distribution for relative germination rate in RIL population at E1, E2, E3, E4 and E5. Instrument-measured traits in all five environments followed normally distributed.

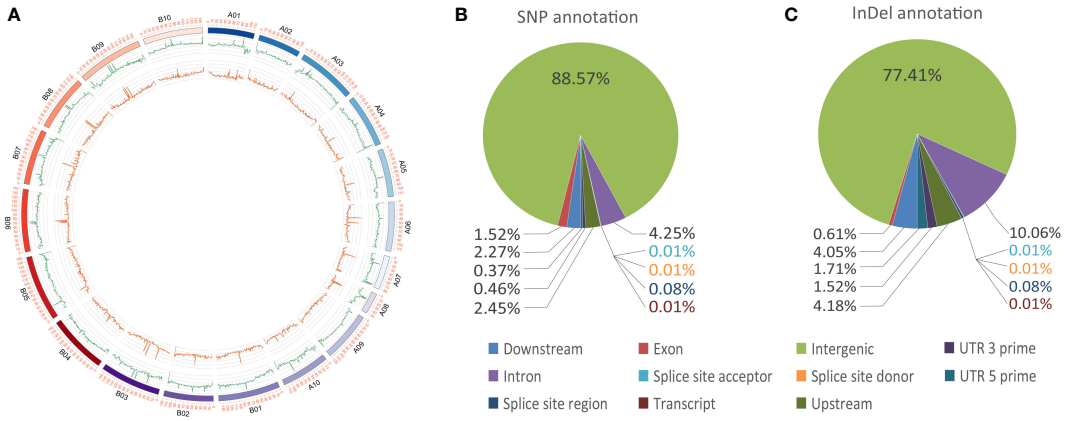


FIGURE 3
Genome variations and annotations. **(A)** Circos plot of SNP and InDel distributions. The outer ring indicates the SNP distribution, whereas the inner ring indicates the InDel distribution. **(B)** Pie charts of SNP annotations. **(C)** Pie charts of InDel annotation information.

3.3 Bin map and genetic map construction

Among all the SNPs, 166,274 are parental polymorphic and homozygous (aa×bb), and these SNPs are used for bin calling. Based on the NMM method, a bin map representing recombination intervals was generated, and 2494 bins were obtained on 20 chromosomes (Figure 4A; Table S4). The mean bin number was approximately 125, with the highest number being approximately 194 on chromosome B03 and the lowest number being approximately 34 on chromosome A04. The basic information statistics of the number of bins, distance, MaxGap, and Spearman of each LG are shown (Tables S5 and S6). The largest and minimum bin MaxGap belonged to chromosomes B07 (18.9 cM) and B06 (1.99 cM), respectively. On average, there were 67 SNPs within a bin. This set of bin markers was used to construct a high-density genetic map including 20 linkage groups (Figure 5). LG8 had the largest genetic distance (110.90 cM), while the smallest genetic distance (21.08 cM) was observed for LG4. For LG8, the largest gap stretched across 36 cM. The ratio of the genetic distances between adjacent markers< 5 cM across all LGs approached 93% (Table S5). Moreover, collinearity was high between the genetic map and the reference genome (Table S6; Figure S1). The recombination hotspot (RH) analysis revealed most RHs were unequally distributed across all 20 LGs, most of which were located on the arm of corresponding chromosomes (Figure 4B). The sources of the larger segments in each individual were consistent, indicating that the quality of the genetic map was very high. The evaluation of the linkage relationship showed that the linkage relationship between adjacent markers on each LG was very strong. With the increase of distance, the linkage relationship between markers as well as that between distant markers gradually weakened, indicating that the sequence of markers was correct (Figure S1). These genomic and genetic indicators indicated that a high-resolution and high-quality genetic map had been constructed for QTL identification.

3.4 QTL identification for cold tolerance-related traits

QTL mapping was performed using CIM in R/qtl software, while ggplot^2 was used to map LOD linkage distribution. The LOD significance threshold ($p< 0.01$) was calculated using a Permutation Test (1000 times), and *qRGRB09* on chromosome B09 was stably detected in all five environments after taking a union set, with a LOD ranging from 4.99 to 11.96, a phenotypic variation explained (PVE) ranging from 10.85 to 24.07%, and an additive effect value ranging from 6.26 to 10.27. The maximum phenotypic variation contribution rate and additive effect value were 24.07% and 10.27, respectively. According to the regional linkage analysis, the five QTLs were located between 49.70–61.75 cM (4.23 Mb), 54.23–61.75 cM (3.81 Mb), 54.77–56.37 cM (755.85 kb), 50.78–61.75 cM (4.07 Mb), and 46.74–61.75 cM (6.01 Mb) in E1, E2, E3, E4, and E5, respectively (Figure 6; Table 1). The physical distance of *qRGRB09* was 6.01 Mb (46.74–61.75 cM) after taking a union set. This indicated that, *qRGRB09* was a major QTL site that was not sensitive to the environment and regulated cold tolerance during germination in peanuts. These findings suggested that *qRGRB09* may be useful for determining the KASP mark for interval design.

3.5 Fine mapping and prediction of candidate genes on chromosome B09

We found 64 SNP sites suitable for KASP and then developed 10 evenly distributed KASP markers (Table S7) covering the QTL regions (chrB09:152644698–156822195) detected in the RIL population, to further confirm and narrow the *qRGRB09* on chromosome B09. Eight of the 10 markers were successfully genotyped (Figure S2). When intersection of QTL intervals of all environments was taken into account, the regional QTL mapping analysis placed *qRGRB09*

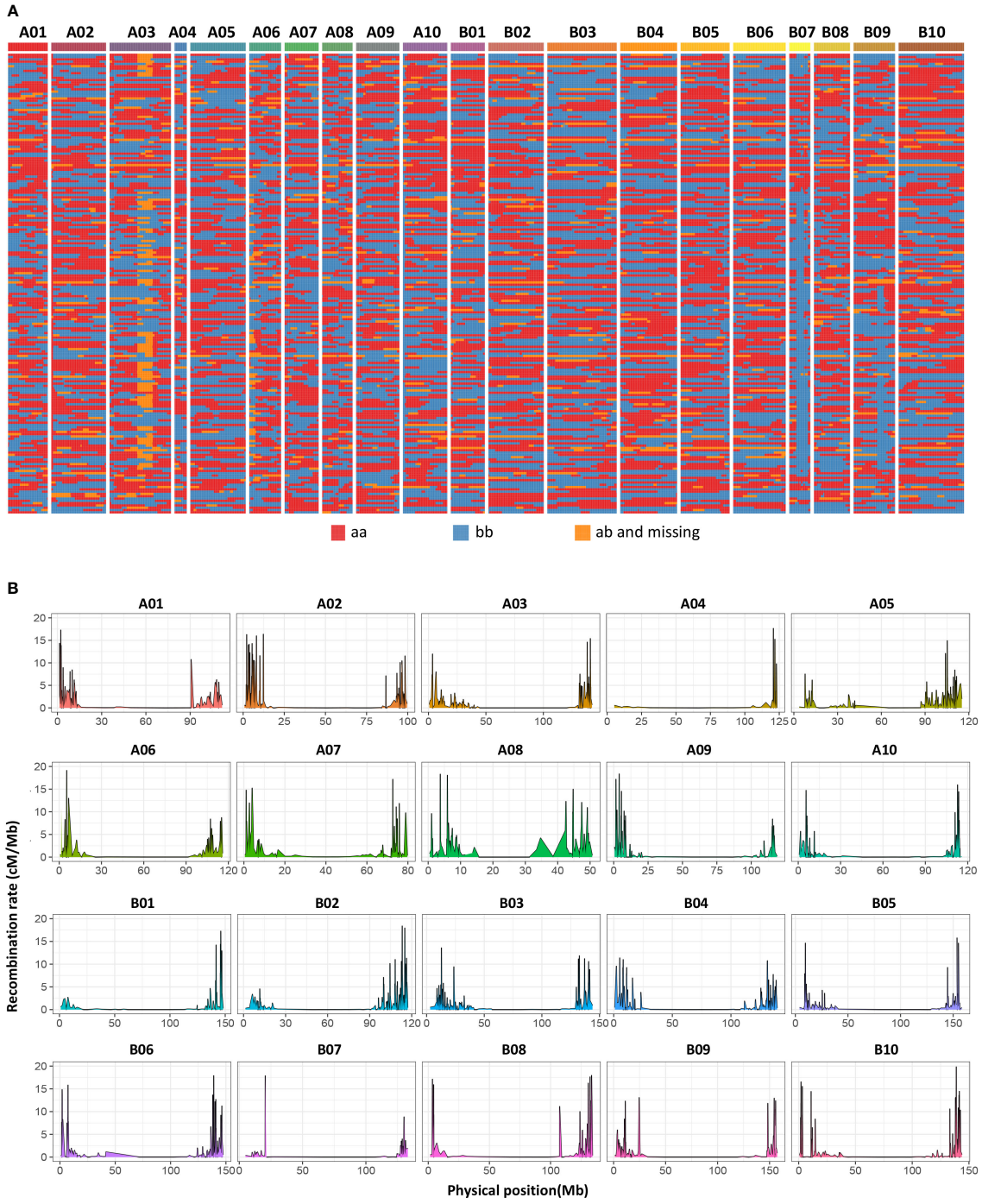


FIGURE 4
Bin map and heatmap of the RIL population. **(A)** A total of 2494 bins were inferred from resequencing-based high-quality SNPs. Red indicates DF12's background, and blue indicates Huayu 44's background. Yellow indicates heterozygous genotypes and missed genotypes. **(B)** Genome-wide recombination hotspots in the 20 peanut chromosomes.

between KASP markers G22096 (P16) and G22097 (P17) (chrB09:155637831–155854093) spaced 14.57–16.35 cM apart, with a 6.10 LOD and 11.55% PVE and a physical length of 216.26 kb (Figure 7A; Table S8). Annotation information was obtained to mine causal genes in the above mentioned 216.26 kb region. Fifteen annotated genes, including *Arahy.GJ31ZB*, *Arahy.RK8HN4*, *Arahy.VKD07K*, *Arahy.YV67XB*, *Arahy.9KP105*, *Arahy.A245TJ*, *Arahy.28TJ89*, *Arahy.NA36PT*, *Arahy.MQZQ11*, *Arahy.12LLDS*, *Arahy.Z64X3Y*, *Arahy.GW3QL7*, *Arahy.HNK57V*, *Arahy.H41NY2*, and *Arahy.2U2D57*, were observed (Figure 7B; Table S9). These candidate genes encoded receptor-like protein kinases (RLKs), tRNA ligase, MYB transcription factor, BAX inhibitor motif-containing protein (BI1) -like protein, Chaperone DnaJ-domain superfamily protein (DnaJ), serine/threonine-protein kinase, and some

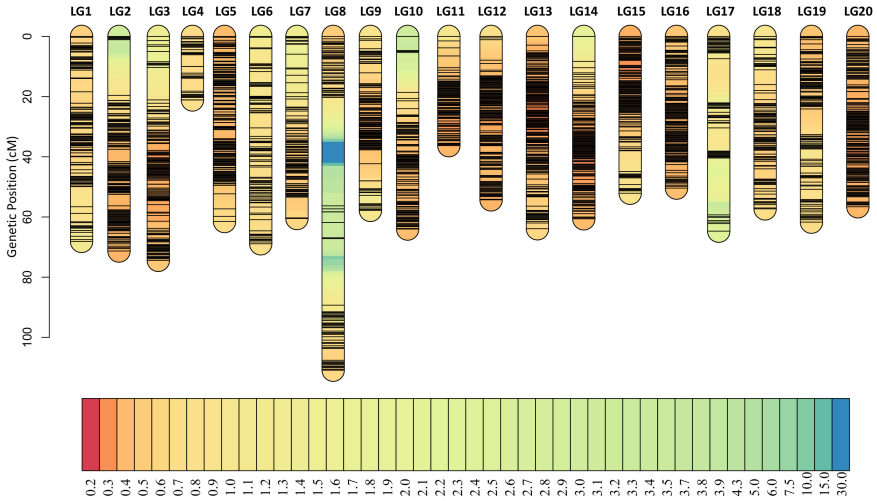


FIGURE 5
High-density genetic map. A high-resolution and high-quality genetic map had been constructed for QTL mapping. The short black lines on linkage groups mean the locus of bin markers.

uncharacterized proteins. Importantly, homologs of some candidate genes have been shown to be involved in cold tolerance (Shakiba et al., 2017; Davik et al., 2021; Han et al., 2022; Li et al., 2022). Then, we examined the expression levels of *RLK*, *MYB* and *DnaJ* in two parents, respectively. The three candidate genes demonstrated uniformly increased expression level in Huayu44, while hardly any noticeable change was observed in DF12 (Figure S3). Further exploitation and functional investigation of these genes may contribute to a better understanding of the cold-tolerance domestication process in peanuts.

4 Discussion

Cold tolerance is vital for the sustainability of peanut production in cold areas. However, dedicated trait mapping or marker discovery studies aimed at elucidating the genetic basis underlying cold tolerance during peanut germination appear to be lacking. The current study, which was conducted in the above context, generated data pertaining to multi-environment phenotyping, high-quality genetic maps, QTLs, and linked markers, all of which were linked to cold tolerance during germination. The information generated by this study may provide the groundwork for the establishment of projects aimed at facilitating fine mapping, gene discovery, and marker development associated with cold tolerance in peanuts.

4.1 Ideal simulation condition was constructed for cold stress at the germination stage

To simulate the field conditions, the alternating low temperature regime (12°C, 3 d; 2°C, 3 d; and 25°C, 3 d recovery) was adopted during seed germination in our study. The parents showed distinct differences in cold tolerance across all five environments. In the RIL populations,

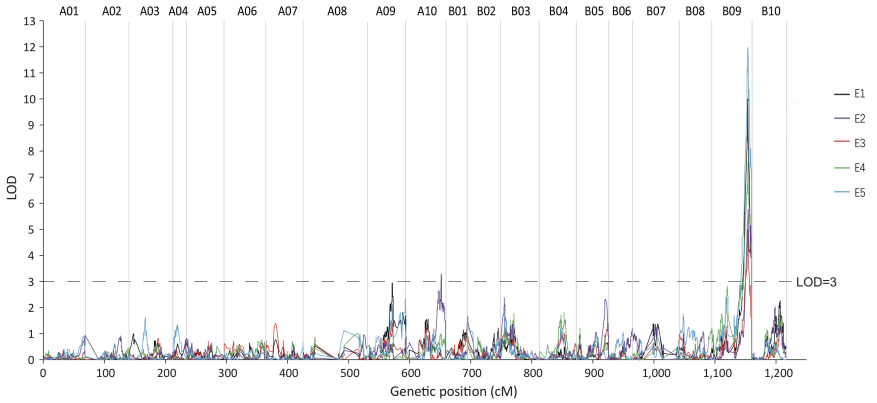


FIGURE 6
The germination stage cold-related QTL *qRGRB09* is located on chromosome B09. *qRGRB09* on chromosome B09 was stably detected in all five environments.

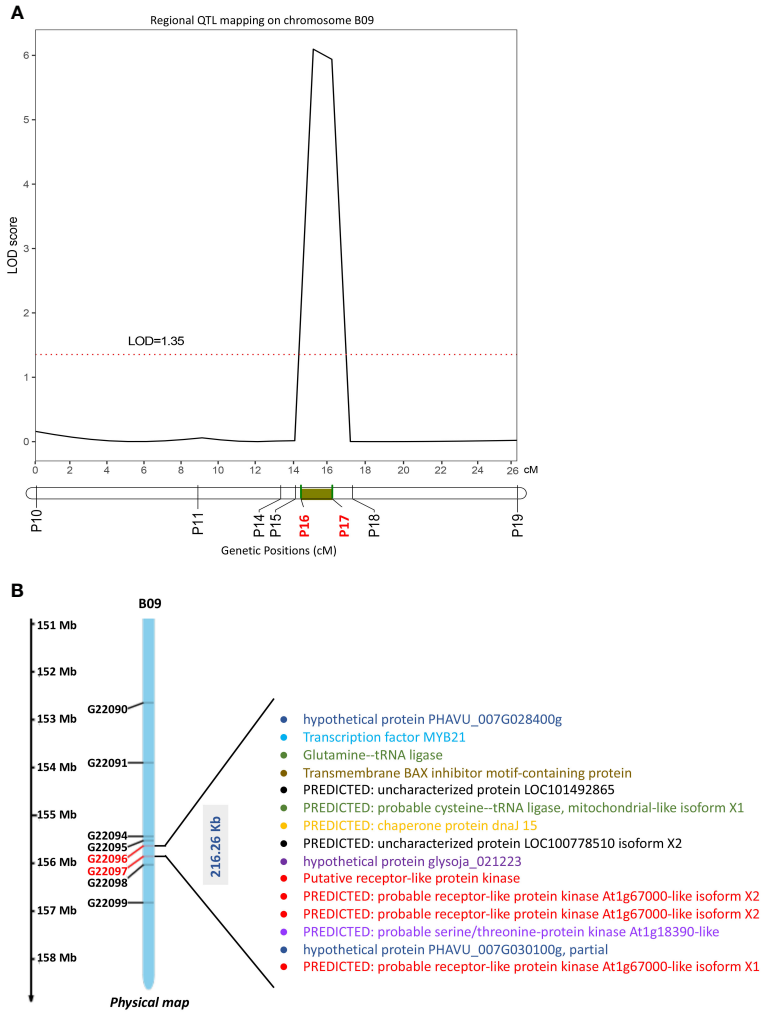


FIGURE 7
Fine mapping of cold-related QTLs and candidate genes screen. **(A)** KASP-based regional linkage analysis to narrow the region QTL (*qRGRB09*) using IciMapping V4.1 in the RIL population (black curve). P16, KASP marker G22096; P17, KASP marker G22097. **(B)** Candidate genes identified in 216.26 Kb QTL region mapped on chromosome B09. This region was rich in RLKs, tRNA ligase, MYB, BI1-like protein, *DnaJ*, serine/threonine-protein kinase, and other uncharacterized proteins.

the variation range of relative germination rate was large, 0.00–94.83% in E1(2019LD), 3.77–92.45% in E2 (2021LD), 7.27–94.83% in E3 (2020FY), 12.07–96.55% in E4 (2021FY), and 8.93–94.83% in E5 (2021NB) (Table S1), all of which were normally distributed. Types of population variation were relatively rich, with a coefficient of variation of more than 33%. The current experiments confirmed the feasibility of simulating conditions as well as constructing successful RIL populations. Therefore, the above-stated simulation conditions at the germination stage should provide the ideal stage for phenotyping for the purpose of genetic mapping and for making reliable selections in breeding programs aimed at developing tolerant varieties.

4.2 Re-sequencing-based bin map strategy facilitated the mapping resolution

WGRS has accelerated the process of gene mapping and cloning in crops. The high-density genetic maps based on re-sequencing constructed by bin markers have shown that continuous SNPs without recombination in one bin marker have been realized in many crops (Li et al., 2016; Lu et al., 2017; Han K, et al., 2018; Tong et al., 2020), including leguminous species (Agarwal et al., 2018; Agarwal et al., 2019; Garg et al., 2022). In General, a population containing 100–400 individuals harbors 1000–4000 bins (Zhang H, et al., 2022), indicating that the QTL mapping resolution for a species with a genome size above 2 Gb may approximate hundreds of kilobases (Zhang et al., 2022). We obtained 2494 bins in total with 1008 kb per bin on average (Figure 4A; Table S4), where four out of five QTL regions were under 5 Mb (Table 1), indicating that a high mapping resolution had been achieved. Moreover, we believe that sequencing-based genotyping also enhances mapping resolution and facilitates the development of high-quality genetic maps. The present study developed a high-quality genetic map with 447,528 SNP loci. The tetraploid genome of cultivated peanuts with a narrow genetic base poses a challenge when attempting to achieve optimum genetic density. High-marker density genetic maps are

10.3389/fpls.2023.1153293

essential for high-resolution genetic mapping and marker discovery.

4.3 A novel and consistent fundamental QTL for further fine mapping

Following the release of peanut genome assemblies, many studies have reported quantitative variations in a diverse range of traits (Agarwal et al., 2018; Luo et al., 2018; Agarwal et al., 2019; Luo et al., 2019). Cold tolerance during germination is important for many plants. At present, many studies focused on the QTL mapping of this trait have been conducted. In rice, a high-density genetic map detected six QTLs, which explained 5.13–9.42% of the total PVE during the germination stage and different genetic loci that regulated cold tolerance at germination and bud stages (Yang et al., 2020). Thapa et al. (2020) identified 31 distinct QTL regions and 13 QTL regions in the japonica subset, and 7 distinct QTL regions in the indica subset, underlying cold tolerance during germination in rice. Hu et al. (2016) performed QTL analysis using 243 IBM Syn4 RILs and detected six QTLs associated with a low-temperature germination rate. Li et al. (2018) generated three connected $F_{2:3}$ populations to detect QTLs related to seed germination ability at low temperatures and found 43 QTLs and three mQTL regions. Interestingly, our study revealed only one average region QTL (*qRGRB09*) with a high LOD (13.44) and PVE (26.61%) for cold tolerance traits, which were continuously mapped using the phenotypic data from five environments. The different results comparing the previous studies and ours might be caused by different testing conditions, such as type and number of molecular markers, mapping populations, as well as temperature simulation conditions. Anyway, our results clearly identified the B09 genome as the source of tolerance that hosted the main-effect QTL for cold tolerance, indicating that the cold tolerance mechanism for peanuts may be different from that of other crops.

4.4 Identification of candidate genes for cold tolerance

Using SNP-derived KASP markers, a regional genetic map rapidly mapped *qRGRB09* to a shared region of 216.26kb (chr:155637831-155854093) on chromosome B09 (Figure 7A; Table S8). The physical length of the region was slightly smaller than that of the preliminary mapping. In preliminary mapping, the physical length of QTLs was 4.23 Mb, 3.81 Mb, 755.85 kb, 4.07 Mb, and 6.01 Mb in E1, E2, E3, E4, and E5, respectively. This shows that we had successfully performed fine mapping using KASP. Based on the results of QTL analysis and the evaluation of this region with the peanut genome assembly, followed by genome annotation, identified a total of 15 genes (Figure 7B; Table S9). This region was rich in RLKs, tRNA ligase, MYB, BI1-like protein, *DnaJ*, serine/threonine-protein kinase, and other uncharacterized proteins. The leucine-rich repeat receptor-like protein kinase (LRR-RLK) gene family, which is the largest family of the receptor-like protein kinase (RLKs) superfamily in higher plants, regulates plant growth, stress responses, and signal transduction (Cheng et al., 2021). Interestingly, we identified four isoforms of RLK proteins in the

TABLE 1 QTLs detected in the RILs population using the high-resolution genetic map.

Environment	QTL	Chromosome	LOD Threshold	Max LOD	Low Marker	High Marker	Low Position (cM)	High Position (cM)	Physical Interval	Additive effect	PVE (%)
E1 (2019LD)	*qRGRB09*	B09	3.97	10.00	C19P153976706	C19P157324174	49.70	61.75	4.23 Mb	10.27	20.57
E2 (2021LD)	*qRGRB09*	B09	4.06	5.76	C19P154454417	C19P157324174	54.23	61.75	3.81 Mb	7.03	12.43
E3 (2020FY)	*qRGRB09*	B09	3.93	4.99	C19P154855110	C19P155555990	54.77	56.37	755.85 Kb	6.26	10.85
E4 (2021FY)	*qRGRB09*	B09	4.24	7.57	C19P154187296	C19P157324174	50.78	61.75	4.07 Mb	7.77	15.99
E5 (2021NB)	*qRGRB09*	B09	3.92	11.96	C19P152409354	C19P157324174	46.74	61.75	6.01 Mb	9.77	24.07
Average	*qRGRB09*	B09	4.01	13.44	C19P153065061	C19P157324174	47.01	61.75	5.43 Mb	8.07	26.61

PVE, phenotypic variation explained; LOD, logarithm of odds.

216.26Kb region which may be key candidate genes for cold tolerance. In addition, the region also contained a serine/threonine protein kinase, the homolog of which is *LRK10L1.2*, reported to be a good candidate for regulating the effects of freezing tolerance at the QTL identified in *Fragaria vesca* L. (Davik et al., 2021). MYB transcription factors are associated with the mechanism of cold tolerance in other plants (Shakiba et al., 2017; Han et al., 2022; Li et al., 2022). Reportedly, *DnaJ* co-chaperons also play a vital role in stress response and have been found to be involved in the maintenance of PSII under chilling stress and the enhancement of drought tolerance in tomato (Kong et al., 2014; Wang et al., 2014). Bax inhibitor is a conserved protein that suppresses the pro-apoptotic protein Bax and eventually inhibits cell death in plants (Huckelhoven, 2004). Some studies have shown that Bax inhibitor, a cytoprotective protein, is involved in the response to heat stress by plants and fungi (Chen et al., 2015; Lu et al., 2018). Heat-stress-induced upregulation of the Bax inhibitor leads to the upregulation of heat-response genes, such as *sHSP*, *HSP70B*, and *HSP90.1*, which enhance thermotolerance in wheat (Lu et al., 2018). These studies indicate that the BI1-like protein deserves to be considered as an important temperature-response protein, which may contribute to cold tolerance. that is, comparable expression in wild type and dcl1, but measurably reduced level in dcl4 mutant (Figure 4A). Our qRT results also confirm that these candidate gene, such as *RLK*, *MYB* and *DnaJ*, may be key candidate genes for future genomics-assisted improvement for cold tolerance.

5 Conclusion

In summary, this study successfully performed WGRS-based genotyping, leading to the creation of a high-quality genetic map for cold tolerance during peanut germination for the first time. This data was used in multi-environment phenotyping, which helped dissect the polymorphic nature of this important stress factor and facilitated the identification of one pivotal genomic region corresponding to cold tolerance in peanuts. Most importantly, a 216.26 kb QTL region on chromosome B09 comprising 15 genes, including potential cold-tolerant genes, such as RLKs, MYB and *DnaJ*, among others, was also discovered. However, further exploration aimed at dissecting the 216.26 kb region *via* constructing secondary mapping populations and assessing the potential of identified candidate genes and genetic markers that can be used in breeding *via* next-generation sequencing technologies, are felt to be warranted.

Data availability statement

The data presented in the study are deposited in the NCBI Sequence Read Archive (SRA) repository, accession number PRJNA931845.

Author contributions

DB and BL designed experiments. XZ, XJZ, LW, QL, YL, JZ, YX, YT, HZ, NL, CS, PN and SF performed experiments. XZ and DB performed the data analysis. XZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (31871662), the National Peanut Industry Technology System Construction (CARS-13), the Open Project of Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, P. R. China (KF2021004), the Fundamental Research Program of Shanxi Province (202203021221178), the Science and Technology Major Project of Shanxi Province (202201140601025), the Research Program Sponsored by the State Key Laboratory of Sustainable Dryland Agriculture (in preparation), Shanxi Agricultural University (202105D121008), the earmarked fund for Modern Agro-industry Technology Research System (2023CYJSTX05), and the Key Research and Development Program of Hainan Province (ZDYF2022XDNY217).

Acknowledgments

We are grateful to Mr Yuhui Xu (Adsen Biotechnology Co., Ltd., Urumchi 830022, China; genetics_2010@163.com) for kindly providing the QTL mapping analysis and Editage (www.editage.cn) for English language editing.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1153293/full#supplementary-material

References

Agarwal, G., Clevenger, J., Kale, S. M., Wang, H., Pandey, M. K., Choudhary, D., et al. (2019). A recombination bin-map identified a major QTL for resistance to tomato spotted wilt virus in peanut (*Arachis hypogaea*). *Sci. Rep.* 9, 18246. doi: 10.1038/S41598-019-54747-1

Agarwal, G., Clevenger, J., Pandey, M. K., Wang, H., Shasidhar, Y., Chu, Y., et al. (2018). High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnol. J.* 16 (11), 1954–1967. doi: 10.1111/pbi.12930

Arends, D., Prins, P., Jansen, R. C., and Broman, K. W. (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* 26 (23), 2990–2992. doi: 10.1093/bioinformatics/btq565

Bai, D. M., Xue, Y. Y., Huang, L., Huai, D. X., Tian, Y. X., Wang, P. D., et al. (2022). Assessment of cold tolerance of different peanut varieties and screening of evaluation indexes at germination stage. *Acta Agronomica Sin.* 48 (8), 2066–2079. doi: 10.3724/SP.J.1006.2022.14163

Bai, D. M., Xue, Y. Y., Zhao, J. J., Huang, L., and Jiang, H. F. (2018). Identification of cold-tolerance during germination stage and genetic diversity of SSR markers in peanut landraces of shanxi province. *Acta Agronomica Sin.* 44 (10), 1459. doi: 10.3724/SP.J.1006.2018.01459

Barrero-Gil, J., and Salinas, J. (2018). Gene regulatory networks mediating cold acclimation: the CBF pathway. *Adv. Exp. Med. Biol.* 1081, 3–22. doi: 10.1007/978-981-13-1244-1_1

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48 (4), 438–446. doi: 10.1038/ng.3517

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

Chang, B. W., Zhong, P., Liu, J., Tang, Z. H., and Guo, W. (2019). Effect of low-temperature stress and gibberellin on seed germination and seedling physiological responses in peanut. *Acta Agronomica Sin.* 45 (1), 118. doi: 10.3724/SP.J.1006.2019.84043

Chen, Y. X., Duan, Z. B., Chen, P. L., Shang, Y. F., and Wang, C. S. (2015). The bax inhibitor MrBI-1 regulates heat tolerance, apoptotic-like cell death, and virulence in *Metarhizium robertsii*. *Sci. Rep.* 5, 10625. doi: 10.1038/srep10625

Chen, X. P., Li, H., Pandey, M. K., Yang, Q. L., Wang, X. Y., Garg, V., et al. (2016). Draft genome of the peanut a-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc. Natl. Acad. Sci. U.S.A.* 113 (24), 6785–6890. doi: 10.1073/pnas.1600899113

Chen, X. P., Lu, Q., Liu, H., Zhang, J. N., Hong, Y. B., Lan, H. F., et al. (2019). Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol. Plant* 12 (7), 920–934. doi: 10.1016/j.molp.2019.03.005

Cheng, W., Wang, Z., Xu, F., Ahmad, W., Lu, G., Su, Y., et al. (2021). Genome-wide identification of LRR-RLK family in *saccharum* and expression analysis in response to biotic and abiotic stress. *Curr. Issues Mol. Biol.* 43 (3), 1632–1651. doi: 10.3390/cimb43030116

Clevenger, J., Chu, Y., Chavarro, C., Agarwal, G., Bertioli, D. J., Leal-Bertioli, S. C. M., et al. (2017). Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Mol. Plant* 10 (2), 309–322. doi: 10.1016/j.molp.2016.11.015

Davik, J., Wilson, R. C., Njah, R. G., Grini, P. E., Randall, S. K., Alsheik, M. K., et al. (2021). Genetic mapping and identification of a QTL determining tolerance to freezing stress in *Fragaria vesca* l. *PLoS One* 16 (5), e0248089. doi: 10.1371/journal.pone.0248089

Garg, V., Dudchenko, O., Wang, J., Khan, A. W., Gupta, S., Kaur, P., et al. (2022). Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement. *J. Adv. Res.* 42, 315–329. doi: 10.1016/j.jare.2021.10.009

Gonda, I., Ashrafi, H., Lyon, D. A., Strickler, S. R., Hulse-Kemp, A. M., Ma, Q., et al. (2019). Sequencing-based bin map construction of a tomato mapping population, facilitating high-resolution quantitative trait loci detection. *Plant Genome* 12 (1), 180010. doi: 10.3835/plantgenome2018.02.0010

Guo, Z., Wang, H., Tao, J., Ren, Y., Xu, C., Wu, K., et al. (2019). Development of multiple SNP marker panels affordable to breeders through genotyping by target sequencing (GBTS) in maize. *Mol. Breed.* 39 (3), 37. doi: 10.1007/s11032-019-0940-4

Han, K., Lee, H. Y., Ro, N. Y., Hur, O. S., Lee, J. H., Kwon, J. K., et al. (2018). QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnol. J.* 16 (9), 1546–1558. doi: 10.1111/pbi.12894

Han, S., Yuan, M., Clevenger, J. P., Li, C., Hagan, A., Zhang, X., et al. (2018). A SNP-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (*Arachis hypogaea* l.). *Front. Plant Sci.* 9, 1012. doi: 10.3389/fpls.2018.01012

Han, Z., Zhang, C., Zhang, H., Duan, Y., Zou, Z., Zhou, L., et al. (2022). CsMYB transcription factors participate in jasmonic acid signal transduction in response to cold stress in tea plant (*Camellia sinensis*). *Plants (Basel)* 11 (21), 2869. doi: 10.3390/plants11212869

Hu, S., Lubberstedt, T., Zhao, G., and Lee, M. (2016). QTL mapping of low-temperature germination ability in the maize IBM Syn4 RIL population. *PLoS One* 11 (3), e0152795. doi: 10.1371/journal.pone.0152795

Huang, L., Ren, X., Wu, B., Li, X., Chen, W., Zhou, X., et al. (2016). Development and deployment of a high-density linkage map identified quantitative trait loci for plant height in peanut (*Arachis hypogaea* l.). *Sci. Rep.* 6, 39478. doi: 10.1038/srep39478

Huckelhoven, R. (2004). BAX inhibitor-1, an ancient cell death suppressor in animals and plants with prokaryotic relatives. *Apoptosis* 9 (3), 299–307. doi: 10.1023/b:appt.0000025806.71000.1c

Jeon, J., and Kim, J. (2013). Cold stress signaling networks in *Arabidopsis*. *J. Plant Biol.* 56 (2), 69–76. doi: 10.1007/s12374-013-0903-y

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome Res.* 12 (4), 656–664. doi: 10.1101/gr.229202

Khedikar, Y., Pandey, M. K., Sujay, V., Singh, S., Nayak, S. N., Klein-Gebbinck, H. W., et al. (2017). Identification of main effect and epistatic quantitative trait loci for morphological and yield-related traits in peanut (*Arachis hypogaea* l.). *Mol. Breed.* 38 (1), 7. doi: 10.1007/s11032-017-0764-z

Knoll, J., Gunaratna, N., and Ejeta, G. (2008). QTL analysis of early-season cold tolerance in sorghum. *Theor. Appl. Genet.* 116 (4), 577–587. doi: 10.1007/s00122-007-0692-0

Kong, F., Deng, Y., Zhou, B., Wang, G., Wang, Y., and Meng, Q. (2014). A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. *J. Exp. Bot.* 65 (1), 143–158. doi: 10.1093/jxb/ert357

Kosambi, D. D. (2012). The estimation of map distances from recombination values. *Ann. Hum. Genet.* 12, 125–130. doi: 10.1111/j.1469-1809.1943.tb02321.x

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci. U.S.A.* 113 (41), E6026–E6035. doi: 10.1073/pnas.1610115113

Li, X., Wang, G., Fu, J., Li, L., Jia, G., Ren, L., et al. (2018). QTL mapping in three connected populations reveals a set of consensus genomic regions for low temperature germination ability in *Zea mays* l. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00065

Li, W., Zhong, J., Zhang, L., Wang, Y., Song, P., Liu, W., et al. (2022). Overexpression of a *Fragaria vesca* MYB transcription factor gene (*FvMYB82*) increases salt and cold tolerance in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 23 (18), 10538. doi: 10.3390/ijms231810538

Liu, N., Guo, J., Zhou, X., Wu, B., Huang, L., Luo, H., et al. (2020). High-resolution mapping of a major and consensus quantitative trait locus for oil content to a ~ 0.8-Mb region on chromosome A08 in peanut (*Arachis hypogaea* l.). *Theor. Appl. Genet.* 133 (1), 37–49. doi: 10.1007/s00122-019-03438-6

Liu, G., Zhao, T., You, X., Jiang, J., Li, J., and Xu, X. (2019). Molecular mapping of the *Cf-10* gene by combining SNP/InDel-index and linkage analysis in tomato (*Solanum lycopersicum*). *BMC Plant Biol.* 19 (1), 15. doi: 10.1186/s12870-018-1616-7

Lu, X., Xiong, Q., Cheng, T., Li, Q. T., Liu, X. L., Bi, Y. D., et al. (2017). A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10 (5), 670–684. doi: 10.1016/j.molp.2017.03.006

Lu, P. P., Zheng, W. J., Wang, C. T., Shi, W. Y., Fu, J. D., Chen, M., et al. (2018). Wheat bax inhibitor-1 interacts with TaFKBP62 and mediates response to heat stress. *BMC Plant Biol.* 18 (1), 259. doi: 10.1186/s12870-018-1485-0

Luo, H., Guo, J., Ren, X., Chen, W., Huang, L., Zhou, X., et al. (2018). Chromosomes A07 and A05 associated with stable and major QTLs for pod weight and size in cultivated peanut (*Arachis hypogaea* l.). *Theor. Appl. Genet.* 131 (2), 267–282. doi: 10.1007/s00122-017-3000-7

Luo, H., Pandey, M. K., Khan, A. W., Wu, B., Guo, J., Ren, X., et al. (2019). Next-generation sequencing identified genomic region and diagnostic markers for resistance to bacterial wilt on chromosome B02 in peanut (*Arachis hypogaea* l.). *Plant Biotechnol. J.* 17 (12), 2356–2369. doi: 10.1111/pbi.13153

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi: 10.1101/gr.107524.110

Meng, L., Li, H. H., Zhang, L. Y., and Wang, J. K. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3 (3), 269–283. doi: 10.1016/j.cj.2015.01.001

Ouellette, L. A., Reid, R. W., Blanchard, S. G., and Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics* 34 (2), 306–307. doi: 10.1093/bioinformatics/btx576

Patel, J., Khandwal, D., Choudhary, B., Ardeshana, D., Jha, R. K., Tanna, B., et al. (2022). Differential physio-biochemical and metabolic responses of peanut (*Arachis hypogaea* l.) under multiple abiotic stress conditions. *Int. J. Mol. Sci.* 23 (2), 660. doi: 10.3390/ijms23020660

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33 (1), 1–14. doi: 10.1007/s11032-013-9917-x

Shakiba, E., Edwards, J. D., Jodari, F., Duke, S. E., Baldo, A. M., Korniliev, P., et al. (2017). Genetic architecture of cold tolerance in rice (*Oryza sativa*) determined through high resolution genome-wide analysis. *PLoS One* 12 (3), e0172133. doi: 10.1371/journal.pone.0172133

Tamari, F., Hinkley, C. S., and Ramprashad, N. (2013). A comparison of DNA extraction methods using *Petunia hybrida* tissues. *J. Biomol Tech* 24 (3), 113–118. doi: 10.7171/jbt.13-2403-001

Thapa, R., Tabien, R. E., Thomson, M. J., and Septiningsih, E. M. (2020). Genome-wide association mapping to identify genetic loci for cold tolerance and cold recovery during germination in rice. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00022

Tong, Z., Zhou, J., Xiu, Z., Jiao, F., Hu, Y., Zheng, F., et al. (2020). Construction of a high-density genetic map with whole genome sequencing in *Nicotiana tabacum* l. *Genomics* 112 (2), 2028–2033. doi: 10.1016/j.ygeno.2019.11.015

Upadhyaya, H. D., Reddy, L. J., Dwivedi, S. L., Gowda, C., and Singh, S. (2009). Phenotypic diversity in cold-tolerant peanut (*Arachis hypogaea* l.) germplasm. *Euphytica* 165 (2), 279–291. doi: 10.1007/s10681-008-9786-2

Upadhyaya, H. D., Wang, Y. H., Sastry, D. V., Dwivedi, S. L., Prasad, P. V., Burrell, A. M., et al. (2016). Association mapping of germinability and seedling vigor in sorghum under controlled low-temperature conditions. *Genome* 59 (2), 137–145. doi: 10.1139/gen-2015-0122

Wang, G., Cai, G., Kong, F., Deng, Y., Ma, N., and Meng, Q. (2014). Overexpression of tomato chloroplast-targeted DnaJ protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco. *Plant Physiol. Biochem.* 82, 95–104. doi: 10.1016/j.plaphy.2014.05.011

Wang, C. B., Cheng, B., Zheng, Y. P., Sha, J. F., Li, A. D., and Sun, X. S. (2003). Effects of temperature to seed emergence, seedling growth and anthesis of peanut. *J. Peanut Sci.* 32 (4), 5. doi: 10.3969/j.issn.1002-4093.2003.04.001

Wang, X. Z., Tang, Y. Y., Wu, Q., Gao, H. Y., Hu, Q. D., and Zhang, S. W. (2013). Cloning and analysis of differentially expressed genes from peanut in response to chilling stress. *J. Nucl. Agric. Sci.* 27 (2), 152–157. doi: 10.11869/hnxb.2013.02.0152

Yang, J., Li, D., Liu, H., Liu, Y., Huang, M., Wang, H., et al. (2020). Identification of QTLs involved in cold tolerance during the germination and bud stages of rice (*Oryza sativa* l.) *via* a high-density genetic map. *Breed Sci.* 70 (3), 292–302. doi: 10.1270/jsbbs.19127

Yin, D., Ji, C., Ma, X., Li, H., Zhang, W., Li, S., et al. (2018). Genome of an allotetraploid wild peanut *Arachis monticola*: a *de novo* assembly. *Gigascience* 7 (6), giy066. doi: 10.1093/gigascience/giy066

Zhang, X., Ren, C., Xue, Y., Tian, Y., Zhang, H., Li, N., et al. (2022). Small RNA and degradome deep sequencing reveals the roles of microRNAs in peanut (*Aarachis hypogaea* l.) cold response. *Front. Plant Sci.* 13, 920195. doi: 10.3389/fpls.2022.920195

Zhang, H., Zhang, X., Li, M., Yang, Y., Li, Z., Xu, Y., et al. (2022). Molecular mapping for fruit-related traits, and joint identification of candidate genes and selective sweeps for seed size in melon. *Genomics* 114 (2), 110306. doi: 10.1016/j.ygeno.2022.110306

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51 (5), 865–876. doi: 10.1038/s41588-019-0402-2

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 22 May 2023
DOI 10.3389/fpls.2023.1101994

OPEN ACCESS

EDITED BY
Weijian Zhuang,
Fujian Agriculture and Forestry University, China

REVIEWED BY
Charles Y. Chen,
Auburn University, United States
Cheng Liu,
University of Florida, United States
Yuanyuan Li,
University of California, Davis, United States
Carolina Ballen-Taborda,
Clemson University, United States

*CORRESPONDENCE
Shengchun Zhang
sczhang@scnu.edu.cn
Weichang Yu
wyu@szu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 18 November 2022
ACCEPTED 18 April 2023
PUBLISHED 22 May 2023

CITATION
Huang R, Li H, Gao C, Yu W and Zhang S (2023) Advances in omics research on peanut response to biotic stresses. *Front. Plant Sci.* 14:1101994. doi: 10.3389/fpls.2023.1101994

Advances in omics research on peanut response to biotic stresses

Ruihua Huang[1†], Hongqing Li[1†], Caiji Gao[1], Weichang Yu[2,3,4*] and Shengchun Zhang[1*]

[1]Guangdong Key Laboratory of Biotechnology for Plant Development, College of Life Sciences, South China Normal University, Guangzhou, China, [2]Guangdong Key Laboratory of Plant Epigenetics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, [3]Liaoning Peanut Research Institute, Liaoning Academy of Agricultural Sciences, Fuxing, China, [4]China Good Crop Company (Shenzhen) Limited, Shenzhen, China

Peanut growth, development, and eventual production are constrained by biotic and abiotic stresses resulting in serious economic losses. To understand the response and tolerance mechanism of peanut to biotic and abiotic stresses, high-throughput Omics approaches have been applied in peanut research. Integrated Omics approaches are essential for elucidating the temporal and spatial changes that occur in peanut facing different stresses. The integration of functional genomics with other Omics highlights the relationships between peanut genomes and phenotypes under specific stress conditions. In this review, we focus on research on peanut biotic stresses. Here we review the primary types of biotic stresses that threaten sustainable peanut production, the multi-Omics technologies for peanut research and breeding, and the recent advances in various peanut Omics under biotic stresses, including genomics, transcriptomics, proteomics, metabolomics, miRNAomics, epigenomics and phenomics, for identification of biotic stress-related genes, proteins, metabolites and their networks as well as the development of potential traits. We also discuss the challenges, opportunities, and future directions for peanut Omics under biotic stresses, aiming sustainable food production. The Omics knowledge is instrumental for improving peanut tolerance to cope with various biotic stresses and for meeting the food demands of the exponentially growing global population.

KEYWORDS

Arachis, omics, pest, pathogen, fungi, virus, bacterial

1 Introduction

Arachis hypogaea (peanut or groundnut) is among the most important oil and food legumes with annual production of ~46 million tons (http://www.fao.org/faostat/en/#home). It is cultivated in more than 100 countries around the world in tropical and subtropical regions, and is the principal source of digestible protein, cooking oil and

vitamins in development and developing regions of Asia, Africa and America for fighting malnutrition and ensuring food security (Arya et al., 2015). Productivity levels of peanut in most of the developing countries have remained low due to several production constraints which include biotic and abiotic stresses. Breeding new cultivars to improve productivity is the best way to meet the needs of the producers, consumers and industry. As an allotetraploid species in the *Arachis* genus, peanut has extremely low genetic diversity because most of the other species in the genus are diploid (Bertioli et al., 2019). Peanut is particularly susceptible to a number of pest and pathogens due in part to the lack of gene exchange with its diploid wild ancestors that have resistance genes (Bertioli et al., 2016; Moretzsohn et al., 2013). The limited genetic diversity and the tetraploid complexity of cultivated gene pool is a barrier and challenge to create cultivars with broad resistance, excellent quality and high yield (Pandey et al., 2020). On the other hand, diploid wild relatives (*Arachis* spp.) with a larger genetic diversity evolving in a variety of habitats and biotic challenges are significant sources of resistance genes and a rich source of novel alleles that can be introduced into the cultivated species by unconventional method (Leal-Bertioli et al., 2009). Therefore, there is an urgent need to exploit gene resources in diploid species by using Omics methods.

Plant genome research has facilitated gene discovery and gene functional elucidation. With Omics, scientists can manage the intricate global biological systems based on advances in Omics technology (Mochida and Shinozaki, 2010). Recent advances in DNA sequencing technology have promoted the rapid development of science and made any other new applications beyond genome sequencing possible (Lister et al., 2009). Particularly, the emergence of next-generation sequencing makes whole-genome resequencing for variant discovery, transcriptional regulatory networks analysis, RNA sequencing analysis (RNA-seq) for transcriptome and noncoding RNAome, quantitative detection analysis (Chip-seq) for epigenome dynamics and DNA-protein interactions become viable applications (Lister et al., 2009). Other techniques, including interactomic analysis for protein-protein interactions, hormonomic analysis for plant hormone signaling, and metabolomic analysis of metabolic products, have been developed (Kojima et al., 2009; Saito and Matsuda, 2010). The omics technologies will help researchers to mine and screen specific genes involved in crop improvement. In addition, integrated network analysis reveals molecular connections between genes and metabolites, boots our understanding the relationships between phenotypic and genotype (Shinozaki and Sakakibara, 2009; Vadivel, 2015; Kumar et al., 2017). Over the past few decades, advances in genomics, transcriptomics, metabolomics, and proteomic analysis with the development of cutting-edge technologies have greatly facilitated the increase in the study of molecular aspects of peanut-biotic factor interactions. Therefore, different omics-based studies have attempted to decipher the molecular pathways that contribute to crop defenses against diseases and pests. In this review, we mainly retrospect the studies on the basic of vary Omics analyses concentrating mainly on those with relevant data on peanut defense responses and resistance to biotic stresses including insect pests, pathogen and bacteria.

2 Biotic stresses on peanut

2.1 Insect pests of peanut

In the semi-arid tropical regions, peanut is a significant crop and is a key component of the diets of both developed and emerging nations. Despite having a high potential for output, farmer's fields typically yield very little due to insect pests and diseases pressure. The peanut crop is infringed by a large number of insects, which lead to disastrous consequences ranging from incidental feeding to almost whole plant destruction and finally yield loss (Wightman and Rao, 1994). According to Stalker and Campbell (1983), peanut is harmed by more than 350 kinds of insects, the most harmful of which are root-knot nematodes, *Aphis craccivora*, *Helicoverpa armigera*, and *Spodoptera litura* (Table 1).

The root-knot nematode (RKN) *Meloidogyne arenaria* is a significant danger to peanut yield particularly in India, China, and the United States (Dong et al., 2008). The RKNs are obligate endoparasites of the *Meloidogyne* genus, with about 100 species described (Decraemer and Hunt, 2006), and the most destructive plant-parasitic nematodes worldwide (Jones et al., 2013), which can infest almost all cultivated plant species (Trudgill and Blok, 2001). The four most common RKN species causing most yield losses in crops are *Meloidogyne incognita*, *M. arenaria*, *M. javanica*, and *M. hapla* (Agrios, 2005). Plants infected by nematodes exhibit symptoms like reduced growth, withering, as well as increased sensitivity to other infections (Mota et al., 2018).

TABLE 1 Impact of insect pests on peanut.

Name	Distribution	Symptom	Severity	Yield loss	Reference
Root-Knot Nematode	US、Africa and Asia	reduced growth and withering	most damaging plant-parasitic nematodes	annual billion dollar	Mota et al., 2018
Aphis craccivora Koch	worldwide	wilt, become yellow or brown, and eventually die	serious	around 20%	Blackman and Eastop, 2007
Helicoverpa armigera	Asia, Africa, southern Europe, and Australia	Feeding on plant's flowering and fruiting bodies	destructive	over $US 2 billion annually	McGahan et al., 1991; Sharma, 2005; Tay et al., 2013
Spodoptera litura	Asia	severely defoliating	destructive	35–55%	Prasad and Gowda, 2006; Srinivasa et al., 2012

Aphis craccivora is an important group of insects with worldwide distribution. Most aphid species comprise a group of closely related populations which may have genetic divergence so that they could be considered as host races, nascent or sister species (or subspecies) (Blackman and Eastop, 2007). Aphid makes approximately 20% peanut yield loss, and causes damage on peanut from seedling to whole mature green plants (Blackman and Eastop, 2007). The Aphid causes both direct and indirect harm to peanut by removing the sap, causing irritation and toxicity, depositing honeydew, growing sooty mold, and spreading the rosette virus, e.g. at least seven viruses utilized Aphid as their vector to damage groundnuts, of which the *Peanut Stripe Virus* (PStV) and the *Groundnut Rosette Virus* (GRV) are the most significant (Blount et al., 2002). When Aphid infestation is severe, the plants may begin to wilt, become yellow or brown, and eventually die and peanut yields are significantly decreased (Blount et al., 2002).

Helicoverpa armigera, one of the most destructive agricultural pests, is thought to cost the US economy $2 billion annually (Tay et al., 2013). Asia, Africa, southern Europe, and Australia all have a significant population of *H. armigera* (Sharma, 2005). More than 200 plant species are impacted, including peanut (Pratissoli et al., 2015). Peanut yield is substantially impacted by *H. armigera*. The direct feeding behavior of *H. armigera* larvae on the plant's flowering and fruiting bodies is one of the main explanations for a significant decline in agricultural productivity (McGahan et al., 1991).

Spodoptera litura is one of the most destructive species that larvae eat voraciously on leaves, severely defoliating the plant and only leaving the midrib veins, which can result in yield losses of 35–55% (Srinivasa et al., 2012). *S. litura* causes maximum damage at the stages during flowering and fruiting (Prasad and Gowda, 2006).

2.2 Microbial pathogen on peanut

2.2.1 Fungi

Peanut growth and development is threatened by a variety of biotic stresses, of which the four fungal diseases leaf spot, rust, stem rot and *Aspergillus flavus* are predominate (Table 2).

Leaf spot includes the *Cercospora arachidicola*-caused early leaf spot (ELS) and the *Phaeoisariopsis personata*-caused late leaf spot (LLS). Both leaf spot diseases can occur on the leaves, petioles, stems, and pegs of peanut and produce lesions up to 1 centimeter in diameter (McDonald et al., 1985). Leaves infected with ELS generally show brown lesions around with a yellow ring on the upper side (Tshilenge-Lukanda et al., 2012), while LLS fungal disease usually exhibit dark brown to black spots (Tshilenge-Lukanda et al., 2012). ELS and LLS are destructive fungal diseases of cultivated peanut, causing yield losses of up to 70% under favorable conditions in the United States and around the world (Anco et al., 2020).

Rust, caused by *Puccinia arachidis* Speg. (Subrahmanyam et al., 1993), is another major fungal diseases restricting peanut yield in countries with warm, tropical climates, with losses as high as 50% reported in India (Varshney et al., 2014). Due to the tendency of rust-infected leaves to stay attached to the plant and the pathogen's short life cycle, the disease can spread quickly and prodigiously. More seriously, rust-infected peanut reduce agricultural productivity, affects the seed oil quality, the haulm and the odder yield (Leal-Bertioli et al., 2015).

Stem rot is the deadliest disease in peanuts and produce markedly yields loss to peanut (Punja, 1985). Four mycelial compatibility groups (MCG) *S. rolfsii* were found among a total of 132 isolates from peanut fields in Ibaraki (Japan) and many isolates were clonal (Okabe and Matsumoto, 2000). This disease is widespread in peanut-growing areas (Thiessen and Woodward, 2012), and caused by *Sclerotium rolfsii* with thick, white hyphae that resemble silk in growth (Ma et al., 2022). Peanut infected with *S. rolfsii* generally exhibits the dark-brown lesions on the stem at soil surface or below the soil surface, followed by gradually yellowing and wilting of leaves (Termorshuizen, 2007). Peanut infect with *S. rolfsii* can produce rot on stem, peg and pod, and up to 80% yield loss (Mehan et al., 1994). Pessimistically, *S. rolfsii* is hard to control because sclerotia derived from *S. rolfsii* overwinter in the soil and attack peanut in the following season (Mayee et al., 1988; Le et al., 2012). After being infected with *S. rolfsii*, peanut plants may experience branch withering and perhaps complete plant wilting (Mayee et al., 1988; Le et al., 2012).

Aspergillus flavus fungus can produce Aflatoxin that threatens to the peanut industry (Krishna et al., 2015). Aflatoxin contaminated peanuts affect human and animal health when consumed (Pittet, 1998; Kew, 2013). Agonizingly, peanut pods and seeds can be infected and Aflatoxin is produced before harvest as well as during the steps of drying, storing, and transportation after harvest (Torres et al., 2014).

TABLE 2 Impact of microbial pathogen on peanut.

Name	Distribution	Symptom	Severity	Yield loss	Reference
Leaf spot	worldwide	brown lesions around with a yellow ring (ELS),dark brown to black spots(LLS)	destructive disease	up to 70%	Tshilenge-Lukanda et al., 2012; Grichar et al., 1998
Rust	worldwide	reduced seed size, and low seed oil content	major fungal diseases	up to 70%	Subrahmanyam et al., 1993
Stem rot	widespread	Feeding on plant's flowering and fruiting bodies	destructivedisease	80% yield loss	Thiessen and Woodward, 2012; Mehan et al., 1994; Punja, 1985
Aspergillus flavus	globally	induce aflatoxins	The major yield limiting biotic stress	Does not reduce yield directly	Krishna et al., 2015

2.2.2 Viruses

Peanuts are infected by various viruses, including *tomato spotted wilt virus* (TSWV), *cucumber mosaic virus* (CMV), *peanut stripe virus* (PStV), *peanut stunt virus* (PSV), *peanut bud necrosis virus* (PBNV), *peanut mottle virus* (PeMoV), *peanut ringspot virus* (PRSV) in the growth and development (Table 3).

TSWV, a propagative and single-stranded RNA virus, is one of the most important pathogenic virus that attacks peanut in the southeastern United States (Culbreath and Srinivasan, 2011; Culbreath, et al., 2003; Garcia et al., 2000). TSWV is transmitted by at least 10 thrips species with a sustained and reproductive manner (Pappu et al., 2009; Riley et al., 2011), of which *Frankliniella fusca* and *F. occidentalis* are predominant (Riley et al., 2011). Peanut attacked by TSWV generally exhibit stunting phenotype particularly when TSWV infects peanut plant at an early stage of growth and development (Culbreath et al., 2003). Beyond that, peanut infected by TSWV also exhibits chlorosis, necrosis or ring spots in peanut leaves (Culbreath et al., 2003). It was reported that TSWV disease alone is estimated to cost US $12.3 million annually loss (Riley et al., 2011).

PStV is one of the most prevalent plant-infecting viruses, a member of the genus Potyvirus, and one of the largest groups of viruses that infect plants (Singh et al., 2009). PStV viruses have a 350-kD polyprotein that is translated by a single open reading frame, which are roughly 10 kb in length and carry a single positive-strand RNA (Urcuqui-Inchima et al., 2001; Wei et al., 2010). PStV is one of the most universal distributed peanut viruses limiting peanut yield by losing around 20%. A number of nations, including China, the US, the Philippines, Thailand, Indonesia, Malaysia, and Korea have reported PStV (Xu et al., 1983; Demski and Lovell, 1985; Saleh et al., 1989; Choi et al., 2001; Choi et al., 2006). PStV can be spread by aphids in a non-sustained manner. In addition, PStV was 10–100% prevalent in the fields and 1–50% in peanut seeds (Chen et al., 1990; Xu et al., 1991; Bi et al., 1999; Xu, 2002). The principal infection source in the field is the infected peanut seeds. On peanut, PStV can induce a number of symptoms, including stripes, light mottle, and blotches that is occasionally encircled by necrotic or chlorotic ringspots (Middleton and Saleh, 1988).

2.2.3 Bacterial

Many bacterial diseases occur in peanuts grown in tropical and subtropical areas because of the warm and wet weather, of which bacterial wilt is predominant (Jiang et al., 2017).

Peanut bacterial wilt, caused by the soil-borne bacterium *Ralstonia solanacearum*, is one of the most devastating diseases in peanut (Salanoubat et al., 2002). *R. solanacearum* ranked second among the top 10 pathogenic bacteria in plant pathology because it spread worldwide and could survive many years in soils (Hayward, 1991; Mansfield et al., 2012). *R. solanacearum* attacks peanut generally through the root system and then spreads to the aboveground parts through the vascular system. If the bacteria breed up to high levels, the plant will show signs of wilting and die (Genin, 2010). In addition, Bacterial wilt disease can lead to 10–30% yield losses and 50–100% in severe circumstances (Jiang et al., 2017) (Table 3).

3 Importance and types of omics approaches for peanut science

The advancement of biotechnology to address plant productivity and stress tolerance has been sparked by the emergence of contemporary genetic engineering methodologies and high throughput biological research tools. Plant biotechnology combined with Omics has the potential to solve a number of issues that currently hinder agriculture, such as diseases and pests, pressures from the environment and climate change (Pérez-Clemente et al., 2013). Omics include but are not limited to genomics, transcriptomics, proteomics, epigenetics, metabolomics, miRNAomics, epigenomics and phenomics (Haas et al., 2017), all were used to improve the peanut cultivars (Figure 1). Genomics-assisted breeding (GAB) has demonstrated great potential for improving peanut varieties. High-quality genome assembly and well-annotated genome are very crucial for GAB. The succeed genome sequence assemblies of wild diploid progenitors, wild tetraploid and both the subspecies of cultivated tetraploids (Bertioli et al., 2016; Bertioli et al., 2019; Zhuang et al., 2019), providing a cornerstone for functional genomics and peanut improvement. Based on the availability of reference genome for both the diploid progenitors, genome-wide simple sequence repeat (SSR) markers were discovered (Zhao et al., 2017). Moreover, whole-genome resequencing (WGRS) of mapping populations has facilitated development of high-density single nucleotide polymorphism (SNP)–based genetic map and genome-wide SNP genotyping array, which were developed for fine mapping and candidate gene discovery for disease resistance in peanut (Clevenger et al., 2017; Pandey et al., 2017c; Agarwal et al., 2018). Although marker-assisted selection approaches have been used to develop superior peanut lines, technological advancements in sequencing and high-throughput genotyping can enhance genetic diversity and forward generation

TABLE 3 Impact of Viruses and Bacterial on peanut.

Name	Distribution	Symptom	Severity	Yield loss	Reference
Tomato Spotted Wilt Virus	southeastern United States	stunting phenotype	destructive disease	annual $12.3 million	Garcia et al., 2000; Culbreath et al., 2003; Riley et al., 2011
Stripe Virus	Asian country and United States	stripes, light mottle, and blotches	most prevalent plant-infecting viruses	around 20%	Choi et al., 2001; Singh et al., 2009; Middleton and Saleh, 1988
Bacterial wilt	globally	wilting and die	the most devastating diseases	10–100%	Mansfield et al., 2012; Genin, 2010; Salanoubat et al., 2002; Jiang et al., 2017

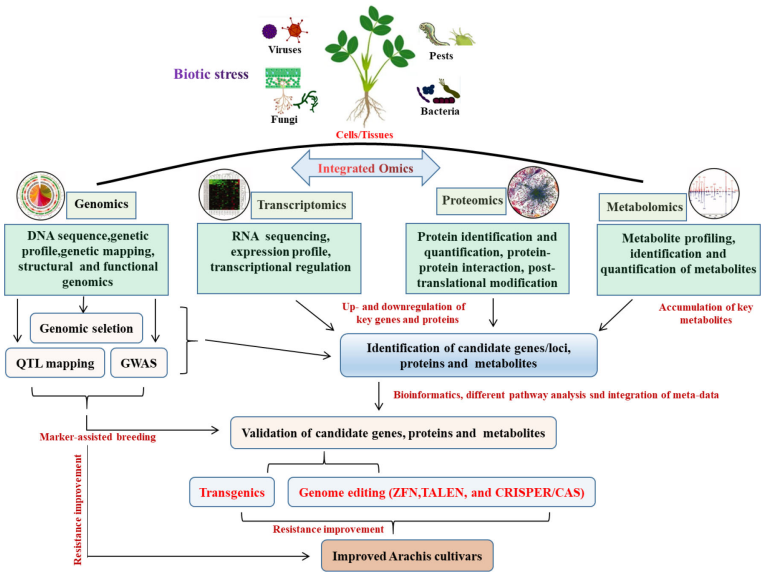


FIGURE 1
Flowchart for improvement of peanut abiotic stress resistance using Omics technologies.

and genomic selection, as well as faster candidate gene discovery in the peanut breeding program (Varshney et al., 2019). For example, RNA sequencing (RNA-seq), which belongs to the transcriptomic approach, can improve the genome annotation and gene discovery especially for the genes which encodes for proteins and non-coding RNAs. Understanding the full metabolic networks involving genes, transcripts, proteins, and metabolites in biological systems is currently crucial because it is extremely difficult to succeed with the strategy of expression of a few single genes in peanut. In this regard, it becomes important to conduct a comprehensive analysis using functional genomics tools such as transcriptomics, proteomics, and metabolomics to characterize plant-pathogen interactions in order to unveil the genetic and metabolic responses of a specific plant species to infection (Pandey et al., 2020).

4 Omics advances in understanding peanut responses to biotic stress

4.1 Peanut responses to pests

4.1.1 Root-knot nematodes

Most peanut cultivars are susceptible to the root-knot nematode (RKN) *M. arenaria*, while the wild diploid *Arachis* species exhibit high resistance (Holbrook and Stalker, 2003). Thus, many gene(s)/QTLs sources linked to RKN resistance were identified in different wild *Arachis* germplasm (Table 4). The interspecific *Arachis* hybrid TxAG-6 ([*A. batizocoi* × (*A. cardenasii* × *A. diogoi*)]^{4x}) was the source of ELS and LLS resistance and the donor parent to introgress resistance into commercial cultivar. In order to improve peanut resistance to RKN, gene segments from wild TxAG-6 were introduced into peanut cultivars through an interspecific hybridization backcrossing scheme, and the root-knot nematode-resistant varieties, COAN and NemaTAM were developed with marker-assisted backcrossing is in the USA (Garcia et al., 1995; Simpson and Starr, 2001; Stalker et al., 2002; Simpson et al., 2003). Previously, the RKN resistance of the cultivated peanut is derived from introgression of a large segment on chromosome A09 from the wild species *A. cardenasii* (Nagy et al., 2010). *A. stenosperma* has high potential in peanut breeding as it owns strong RKN resistance (Ballén-Taborda et al., 2019). Three quantitative trait loci (QTL) were genetically mapped to strongly influence nematode root galling and egg production by using 93 recombinant inbred lines (RILs) developed from a cross between *A. duranensis* and *A. stenosperma* (Leal-Bertioli et al., 2016). Two loci controlling the resistance on chromosomes A02 and A09 have been validated in cultivar peanut to reduce nematode reproduction by up to 98%, and the large-effect QTL on A02 is enriched for genes encoding TIR-

TABLE 4 QTL associated with Peanut responses to Root-knot nematodes.

	Location	Population	Reference
Root-knot nematodes	chromosomes 02, 04 and 09	*A. duranensis* × *A. stenosperma*	Bertioli et al., 2016
	chromosomes A02 and A09	(*A.batizocoi* × *A.stenorperma*) ×*A.hypogaea*	Ballén-Taborda et al., 2019; Ballén-Taborda et al., 2021; Ballén-Taborda et al., 2022
	chromosomes A09	*A.cardnesii*	Nagy et al., 2010

NBS-LRR proteins (Ballén-Taborda et al., 2019; Ballén-Taborda et al., 2021; Ballén-Taborda et al., 2022).

To understand the molecular components underlying RKN resistance, researchers took advantage of genomics and transcriptomics to demonstrated that wild *Arachis* species *A. stenosperma* harbors high levels of resistance to RKN infection through the onset of hypersensitive response (HR), which is usually caused by resistance genes (R) (Proite et al., 2010; Dang et al., 2013; Guimaraes et al., 2015). As we all know, the majority of plant R genes are the NBS-LRR type genes (Meyers et al., 1999). Transcriptome analysis of two peanut wild relatives, *A. stenosperma* representing the highly RKN resistant and *A. duranensis* as the moderately susceptible, during early stages of RKN infection, found that resistance genes against root-knot nematode infection were NBS-LRR class of plant disease resistance (R) genes. Two decades ago, 78 NBS-LRR coding sequences with unknown functions were identified in wild *Arachis* species *A. stenosperma* by resistance gene analogues (RGAs) targeting degenerate primers in the NBS domain (Bertioli et al., 2003). Furthermore, over 300 representative genes segmented into four NBS-LRR family types were isolated from the genome-wide analysis in the wild peanut species (Bertioli et al., 2016; Song et al., 2017). Similarly, hundreds of RGAs were identified from several peanut cultivars (Yuksel et al., 2005; Ratnaparkhe et al., 2011; Wang et al., 2012). Suppression subtractive hybridization (SSH) revealed that pathogenesis related (PR) protein, patatin-like protein and universal stress related protein (USPs) genes, which related to the resistance operative against invading nematodes, were expressed in the early stages of RKN-infected NemaTAM roots (Tirumalaraju et al., 2011). In addition, seven genes including one gene encoding a resistance protein MG13, were differentially expressed in RKN-infected wild *Arachis* (Morgante et al., 2013). Comparative genomics combining with differential gene expression analysis in 22 plant species including peanuts revealed the conserved immune response genes triggered by RKN infection. The core genes include plant defense and secondary metabolite production (Mota et al., 2020). In addition, genes involved in hormone signaling and secondary metabolites production may be involved in RKN resistance (Mota et al., 2018). Consistently, genes engaged in salicylic and jasmonic acids signaling pathways as well as genes in auxin balance regulation were found in the transcriptome analysis of RKN-resistant *Arachis* genotypes (Guimaraes et al., 2015). These results suggest the role of phytohormones in root-knot nematode resistance.

In addition to transcriptomics and genomics, metabolomics and miRNAomics are also making important contributions to peanut RKN research. The miRNAomics with whole-transcriptome RNA-seq revealed that 430 mRNAs, 111 miRNAs, 4453 lncRNAs, and 123 circRNAs were differentially expressed upon RKN infection, among which a total number of 10 lncRNAs, 4 circRNAs, 5 miRNAs, and 13 mRNAs involved in the oxidation reduction process and biological metabolism processes in RKN infected peanuts (Xu et al., 2022). Furthermore, proteome combining with transcriptome analysis identified differentially expressed proteins and genes during root-knot nematode infection (Martins et al., 2020). Most of the differentially expressed proteins are related to plant responses to pathogens. And the plant defense related genes encoding the ADH (alcohol dehydrogenase), CCR1 (cinnamoyl-CoA reductase 1), ENO (enolase), eIF5A (eukaryotic translation initiation factor 5A) and MLP34 (MLP-like protein 34) were found during peanut RKN infection, and the AsMLP34 was considered as a candidate in peanut RKN resistance (Martins et al., 2020).

4.1.2 Aphis craccivora

To prevent *Aphis craccivora* infestation in peanut, resistance lines were first identified using phenomics (Padgham et al., 1990). Genomic analysis by Merwe et al. (2001) revealed that *Aphis craccivora* resistance was regulated by a single recessive gene. DNA markers linked to aphid resistance and a partial genetic linkage map were developed by bulk segregate analysis (BSA) and amplified fragment length polymorphism (AFLP) (Herselman et al., 2004). The F2:3 population derived from a cross with the aphid-resistant parent ICG 12991 were examined using a total of 308 AFLP primer combinations to find markers linked to aphid resistance. Twelve of the twenty putative markers were mapped to five linkage groups, spanned a map distance of 139.4 cM (Herselman et al., 2004). Recently, metabolomics with high performance liquid chromatography (HPLC) was employed to analyze the phenols fingerprinting of peanut plants with different resistance levels to aphid infestation, and common compounds such as the chlorogenic, syringic, quercetin, and ferulic acids were identified during aphid infestation (War et al., 2016). The quantities of the identified compounds are depending on genotypes and modes of aphids feeding.

4.1.3 Helicoverpa armigera

In an effort to identify potential defense strategy, the biochemical basis of *H. armigera* infestation in peanut was analyzed with protein electrophoretic analysis and enzymatic assays (Harsulkar et al., 1999). Non-host peanut containing proteinase inhibitors (PIs) effectively inhibited the gut proteinases (HGPs) activity of *H. armigera* and larval growth (Harsulkar et al., 1999). Morphological traits were linked to resistance to *H. armigera* and can be used as markers of resistance selection. Significant correlations were found between main stem thickness, leaflet shape and length, hypanthium length, number of hairs on leaves, standard petal length and petal pattern, basal leaflet width, number of hairs, adherent length and width of stipules, plug length and *H. armigera* infestation (Sharma et al., 2003). Twelve resistance lines were selected under field and greenhouse conditions by screening 30 *Arachis* spp. (Sharma et al., 2003).

Salicylic acid (SA) and Jasmonic acid (JA) were also identified to induce defensive responses in peanut against *H. armigera* infestation. JA pretreatment markedly increases peroxidase (POD) and polyphenol oxidase (PPO) activities and high level total phenols, hydrogen peroxide (H_2O_2) and malondialdehyde (MDA) in peanut, and different levels of *H. armigera* resistance were recorded (War et al., 2011). Exogenous application of JA and SA induced resistance to *H. armigera*. Susceptible peanut and genotypes with different levels of *H. armigera* resistance showed higher amounts of secondary metabolites and levels enzymatic

activities, and reduced *H. armigera* growth and development when pretreated with JA and SA in the green house (War et al., 2011). Recently, metabolomics with high performance liquid chromatography (HPLC) were employed to analyze the phenols fingerprinting of peanut plants with different resistance levels to *H. armigera* infestation (War et al., 2016). The observed common compounds were chlorogen, clove, quercetin and ferulic acid during *H. armigera* infestation (War et al., 2016).

4.1.4 Spodoptera litura

Scientists have been trying to study the resistance of peanut to *S. litura* over the past few decades (Sharma et al., 2003). Some peanut morphological characteristics showed markedly correlation and/or regression coefficients with *S. litura* damage under field and greenhouse conditions, and were used as markers of selection for resistance to *S. litura* infection (Sharma et al., 2003). In addition, peanut grown under elevated carbon dioxide (CO_2) level showed higher level carbon and polyphenols content, which reduced insect digestion efficiency, slowed down the growth of individual pest and reduced the *S. litura* pupation (Srinivasa et al., 2012). Furthermore, JA application can also boost the resistance to *S. litura* (War et al., 2011).

Genetic engineering has been proven to be effective in controlling insect pest. Ectopically expressing AhMPK3A, an *Arabidopsis* AtMPK3 gene homology in peanut, showed resistance to the first and second instar larvae of *S. litura* and generated higher expression levels of defense response genes such as PR1a, PR1b and LOX1 (Kumar et al., 2009). A chimeric Bt toxin protein cry1AcF with cry1Ac (domain I & II) and cry1F (domain III) was also employed to develop resistance peanut to *S. litura*. The cry1AcF transgenic peanut showed that cry1AcF significantly increased the mortality of *S. litura* larvae, and was effective against *S. litura* (Keshavareddy et al., 2013).

4.2 Peanut responses to microbial pathogens

4.2.1 Fungi

4.2.1.1 Leaf spot

For the past decades, many works have been done focusing on uncovering major gene for peanut leaf spot resistance, and masses of QTLs were identified (Table 5). The wild species accession, *A. cardenasii* GKP10017, is an important donor of leaf spot resistance to the peanut crop, because the segments from chromosome A02 and A03 correspond to some very strong QTLs that confer resistance to rust and LLS (Bertioli et al., 2021). Chu et al. (2019) discovered four leaf spot diseases resistance QTLs on both chromosome 3 and 5 in the Florida-07× GP-NC WS16 population. Zhang et al. (2020) identified two QTLs closely associating with resistance to ELS and LLS on chromosome B09 in the US mini-core peanut collection. In peanuts, many markers were recently found to be associated with QTLs for leaf spot disease resistance. Recent developments in mapping technologies for peanuts have identified of large numbers of QTL-associated polymorphic markers involved in peanuts ELS and LLS resistance. Eleven QTLs were found on a genetic map containing 56 microsatellites, or simple sequence repeat (SSR) marker loci through genetic mapping of GPBD4 (resistant varieties) derived a recombinant inbred line (RIL) population for LLS resistance (Khedikar et al., 2010), and 28 QTLs were identified on an improved genetic map with 260 SSR loci in the same RIL (Sujay et al., 2012). Two major QTLs for LLS resistance on chromosomes B10 and A03 were characterized based on mapping with 139 additional SSR and transposable element markers (Pandey et al., 2016). Nine QTLs involved in ELS resistance and 22 QTLs involved in LLS resistance were discovered and used in marker-assisted breeding (Pandey et al., 2017a). Khera et al. (2016) found 22 and

TABLE 5 QTL associated with Peanut responses to Leaf Spot.

	Location	Population	Reference
Leaf Spot	chromosomes B10 and A03	Zhonghua 5 x ICGV 86699	Zhang et al., 2017
	chromosome B06	Zhonghua 5 x ICGV 86699	Zhang et al., 2017
	chromosome 3 and 5	Florida-07× GP-NC WS16	Khera et al., 2016
	chromosome B06	Zhonghua 5 ICGV 86699	Zhou et al., 2016
	chromosome B09	GJG17 ×GPBD4	Zhang et al., 2017
	chromosome 3 and 5	Florida-07× GP-NC WS16	Chu et al. (2019)
	chromosome B09	US mini-core peanut collection	Zhang et al. (2020)
	A02 and A03	TAG 24× GPBD 4	Shirasawa et al., 2018
	chromosome A03	GJG17 × GPBD4	Ahmad et al., 2020
	on linkage group AhXV	TAG 24 x GPBD 4	Sujay et al., 2012
	A sub-genome	Tifrunner × GT-C20	Pandey et al., 2017a
	chromosomeA02, B04, B06, B09, and B10	Tamrun OL07 × Tx964117	Liang et al., 2017
	chromosomes A03, B04 and B05	Yuanza 9102	Han et al. (2017)

20 QTLs for ELS and LLS respectively, with a SSR-based map containing 248 loci in a population derived from SunOleic-97R × NC94022. QTLs for ELS and LLS were also recovered by utilizing single nucleotide polymorphism (SNP) linkage map, and identified a major QTL for late leaf spot resistance *via* analysis of interval sequences in peanut (Han et al., 2018). Zhou et al. (2016) used the Zhonghua 5 ICGV 86699 population for genetic mapping with 1685 SNPs from double-digested restriction-site associated DNA (ddRAD) sequences, and detected 20 LLS QTLs, among which 5 of the 6 major QTLs were located on chromosome B06. Liang et al. (2017) reported six QTLs on different chromosomes of ELS resistant parent Tx964117 were found using ddRAD-seq markers with 1211 loci developed from Tamrun OL07 × Tx964117 population. In a region harboring major QTLs for LLS and rust diseases, seven novel candidate expressed sequence tag-derived simple sequence repeat markers (EST-SSRs) related to stress were mapped using the F2 mapping population (GJG17 × GPBD4), and two major QTLs for LLS were found (Ahmad et al., 2020). In marker-assisted selection, the consensus QTLs across different genetic backgrounds are important and necessary. Shirasawa et al. (2018) pinpointed QTL candidates for LLS in a 1.4-Mb genome regions on A02, and selected four resistance-related genes as candidates for LLS in this region. Lu et al. (2018) identified one major Meta-QTL harboring 26 candidate genes for LLS in a region of about 0.38 cM. Moreover, QTL-seq approach was used to identify diagnostic markers and genomic regions for LLS resistance in peanut, and nine candidate genes with 17 SNPs were identified and one of these SNPs could serve as an allele-specific diagnostic marker (Pandey et al., 2017b). These delimited candidate genes-containing genomic regions will be valuable in uncovering the key resistant genes and in the development of LLS disease resistance in peanut breeding.

Transcriptome analysis was also performed in peanut leaf spot diseases. Cyclophilin gene with potential roles in peanut first line of defense are characterized using differential gene expression analysis following infection with the peanut late spot pathogen (Kumar and Kirti, 2011). Han et al. (2017) developed a highly susceptible M14 mutant to LLS derived from Yuanza 9102 cultivar. RNA-Seq analysis in M14 and the wild type Yuanza 9102 (resistant to several fungal diseases) leaf tissues under LLS pathogen challenge showed 2219 differentially expressed genes including 1317 up- and 902 down-regulated genes, including up-regulated pathogenesis-related (PR) protein genes, WRKY transcription factor genes, down-regulated chloroplast genes and depressed plant hormones in the M14 mutant. Furthermore, genes possibly involved in recognition events and early signaling responses to the pathogen, including resistance related proteins, hypersensitive cell death, cell wall strengthening and metabolism and signal transduction, were identified by transcriptomic and proteomic analyze (Kumar and Kirti, 2015). Dang et al. (2019) verified a group of 214 R-genes expressed in peanut leaves induced with leaf spot pathogen infection. Gong et al. (2020) identified 133 differentially expressed genes (DEGs) between ELS resistant and ELS susceptible peanut lines by transcriptome analysis, including leucine rich repeat (NLR) type resistance genes on the chromosome B2, peanut phytoalexin deficient 4 (PAD4) gene involved in NLR resistance proteins mediated immunity, and polyphenol oxidase (PPO) genes crucial to early leaf spot resistance in peanut. In a comparative transcriptome analysis from resistant line (GPBD 4), resistant introgression line (ICGV 13208) and a susceptible line (TAG 24), the resistance genes for LLS resistance, Aradu.P20JR and Aradu.Z87JB, were revealed on chromosome A02 and A03, respectively (Gangurde et al., 2021). Dang et al. (2021) reported 36 R-genes were markedly correlated with and differentially expressed in resistant lines. Most of the R-genes are receptor like kinases (RLKs) and receptor like proteins (RLPs) that function in precepting the presence of pathogen at the cell surface and initiating protection response.

In addition, metabolomics has also been applied to explore the mechanism of peanut response to LLS. LLS resistant peanut genotypes has higher levels secondary metabolites including but not limited to phenolic acid, flavanols, stilbenes and terpenoids (Mahatma et al., 2021).

4.2.1.2 Rust

Mapping efforts have been concentrated on genomic sequence-based analysis and SNP enrichment mapping so as to define interested regions and identify candidate genes. Twelve QTLs for rust were identified by QTL analysis of 268 recombinant inbred lines from the rust segregation mapping population TAG 24 x GPBD 4. Interestingly a major QTL associated with rust resistance, as well as a candidate SSR marker (IPAHM 103) linked to this QTL, was identified and validated by both composite interval mapping and single-marker analysis using a broad range of resistant/susceptible breeding lines and progeny lines from another mapping population (TG 26 x GPBD 4) (Khedikar et al., 2010). Pasupuleti et al. (2016) used a marker-assisted backcross (MABC) method to import the major QTL accounting for 80% of the phenotypic variation (PV) in rust resistance. Nine candidate genes for rust resistance on chromosomes A03 spreading on a 3.06-Mb region were discovered with a QTL-Seq approach employed numerous resistance and susceptible lines from TAG 24 × GPBD 4 population (Pandey et al., 2017a). Five candidate genes for rust resistance within a 1.2 cM fragment on chromosomes A03 flanked by two SSR markers SSR_GO340445 and FRS 72 were also found with the help of mapping on the VG 9514 × TAG 24 population (Mondal and Badigannavar, 2018). A 2.7-Mb genome location of the rust resistance genes in the same genomic region of chromosome A03 was confirmed with the help of ddRAD-Seq and whole-genome resequencing for the population derived from hybrid between TAG 24 and GPBD 4 (Shirasawa et al., 2018). Recently, a major rust QTL containing resistance-related genes and R-genes functioning in inducing hypersensitive response (HR) during rust infection were validated by mapping with seven novel stress-related candidate EST-SSRs using 328 individuals from the F2 (GJG17×GPBD4) mapping population (Ahmad et al., 2020) (Table 6). In addition, nine rust resistant genotypes showed a 77% to 120% increase in pod yield under rust disease pressure over control, revealing significant environment (E) and genotype × environment (G×E) interactions (Chaudhari et al., 2019). QTL-Seq approach has been deployed to identify genomic regions and diagnostic markers for rust resistance in peanut, and 30

TABLE 6 QTL associated with Peanut responses to Rust.

	Location	Population	Reference
Rust	chromosomes A03	TAG 24 × GPBD 4	Pandey et al., 2017a
	chromosomes A03	VG 9514 × TAG 24	Mondal and Badigannavar, 2018
	chromosomes A03	TAG 24 and GPBD 4	Shirasawa et al., 2018
	chromosomes A03	GJG17×GPBD4	Ahmad et al., 2020
	chromosomes A06	TG 26 x GPBD 4	Khedikar et al., 2010

nonsynonymous SNPs affecting 25 candidate genes, and three allele-specific SNP diagnostic markers for rust resistance were identified (Pandey et al., 2017b). QTLs for rust resistance in the peanut wild species *A. magna* were developed using single-nucleotide polymorphism competitive allele-specific polymerase chain reaction markers and the marker function was validated in both diploid and tetraploid peanuts (Leal-Bertioli et al., 2015).

RNA-Seq data from susceptible peanut genotype JL-24 and resistant peanut genotype GPBD-4 revealed differentially expressed genes included pathogenesis-related (PR) proteins, ethylene-responsive factor, thaumatin, and F-box as well as R genes such as NBS-LRR upregulated in resistant genotype, whereas transcription factors such as WRKY, MYB, bZIP family down-regulated in susceptible genotype (Rathod et al., 2020). Using map-based cloning, a dominant rust resistance gene VG 9514-R located between FRS 72 and SSR_GO340445 markers in arahy03 chromosome was isolated and shown non-synonymous mutations in different protein domains (Mondal et al., 2022).

4.2.1.3 Stem rot

By far, the QTLs and markers for have not discovered enough for peanut resistance to *S. rolfsii*. Utilizing QTL analysis with multi-season phenotyping and genotyping data from a TG37A×NRCG-CS85 population, 44 major epistatic QTLs explained phenotypic variation ranging from 14.32 to 67.95% were found (Dodia et al., 2019). Molecular mechanisms of peanuts resistant to *S. rolfsii* have been studied mostly with transcriptomic tools. The salicylic acid, defense-related signal molecule, peroxidase, marker enzymes, polymer lignin as well as the phenylalanine ammonia lyase-1,3-glucanase, all which related to the systemic acquired resistance and were observed could be induced by *S. rolfsii* derived elicitors (Nandini et al., 2010). RNA-Seq from infected peanut tissue found 12 differentially expressed genes including 7 genes related to defense in the plants and 3 genes related to virulence in the fungi (Jogi et al., 2016). Bosamia et al. (2020) unraveled genes encoding jasmonic acid pathway enzymes, receptor-like kinases, and transcription factors (TFs) including Zinc finger protein, WRKY, and C2-H2 zinc finger with high level expression in resistant peanut genotypes by RNA sequencing approaches. The pathogen-associated molecular patterns (PAMP)-triggered immunity was considered as a potential mechanism of stem rot resistance, while the jasmonic acid signaling pathway was deemed to a potential defense mechanism in peanut. There is also evidence of different enzymes activity in the crosstalk between peanut and *S. rolfsii* (Bosamia et al., 2020). *De-novo* genome sequencing of two distinct pathogen strains ZY and GP3 with high and weak aggressiveness respectively revealed the genomic basis for vary aggressiveness of *S. rolfsii* (Yan et al., 2021). The poll of pathogenicity-associated gene relate to aggressiveness were differed between GP3 and ZY based on comparative genomic analysis. GP3 and ZY possessed 58 and 45 unique pathogen-host interaction (PHI) genes, respectively. In addition, compared with GP3, ZY strain had more carbohydrateactive enzymes (CAZymes) in its secretome, especially the carbohydrate esterase (CBM), the polysaccharide lyase (PL) and the glycoside hydrolase (GH) family (Yan et al., 2021).

4.2.1.4 Aspergillus flavus

Recently, series of QTLs in peanuts about resistance to *A. flavus* infection were successfully identified by QTL mapping (Table 7). Yu et al. (2019) identified two QTLs with 7.96 and 12.16% phenotypic variation explained (PVE) on chromosomes A03 and A10, respectively by constructing a genetic map with 1,219 SSR loci and a recombinant inbred line (RIL) population resulted from crossing ICG12625 with susceptible cultivar Zhonghua 10. Khan et al. (2020) identified a major QTL with a PVE of 18.11% on A03 and a minor QTL with a PVE of 4.4% on B04 by utilizing SNP based genetic map using specific length amplified fragment sequencing (SLAF-Seq). Based on 200 recombinant inbred lines (RILs) mapping population from a hybrid of a susceptible variety Zhonghua 16 with resistant germplasm J11; Jiang et al. (2021) reported six novel resistant QTLs on chromosomes A05, A08, B01, B03, and B10 with 5.03–10.87% PVE, respectively.

The molecular mechanisms of peanut–*A. flavus* interactions and peanut resistance to aflatoxin generation need to be studied to create effective countermeasures against postharvest aflatoxin contamination. A great number of transcriptome analysis gave hints and information to this aspect. Guo et al. (2008) constructed cDNA libraries from the seeds of peanut resistant cultivars (GT-C20) and susceptible cultivars (Tifrunner) with 21777 EST sequences. Using two-dimensional electrophoresis, mass spectrometry and real-time RT-PCR; Wang et al. (2010) reported the identification of twelve potentially differentially expressed proteins between resistant peanut variety YJ-1 and susceptible peanut variety Yueyou 7 under conditions of sufficient water, drought stress and flavus infection with drought stress. According to a peanut oligonucleotide microarray chip analysis, 62 genes with upregulated expression in resistant cultivar and 22 putative Aspergillus-resistance genes with high level expression in the resistant cultivar were determined (Guo et al., 2011). In

TABLE 7 QTL associated with Peanut responses to *Aspergillus flavus*..

	Location	Population	Reference
Aspergillus flavus	chromosomes A03 and A10	ICG12625× Zhonghua 10	Yu et al., 2019
	chromosomes A03 and B04	Zhonghua 16 × J11	Khan et al., 2020
	chromosomes A05, A08, B01, B03, and B10	Zhonghua 16 × J11	Jiang et al., 2021

addition, a series of aflatoxins-responsive proteins, including those involved in immune signaling and innate immunity, induction of defense, PAMP perception, penetration resistance, hypersensitive response, DNA and RNA stabilization, biosynthesis of phytoalexins, condensed tannin synthesis, cell wall responses, peptidoglycan assembly, detoxification and metabolic regulation, were identified in peanut cotyledons infected with aflatoxin-producing (toxigenic) but not non-aflatoxinogenic (toxigenic) *A. flavus* strains, using a differential proteomics approach (Wang et al., 2012). Global transcriptome analysis of post-harvest peanut seeds of susceptible (Zhonghua 12) and resistant (Zhonghua 6) peanut genotypes undergoing fungal infection and aflatoxin production revealed 128, 725 unigenes, of which 30, 143 were differentially expressed, and 842 are potential defense-related genes, including pathogenesis-related proteins, leucine-rich repeat receptor-like kinases, transcription factors, mitogen-activated protein kinase, nucleotide binding site-leucine-rich repeat proteins, polygalacturonase inhibitor proteins, ADP-ribosylation factors and other defense-related crucial factors (Wang et al., 2016). Transcriptomic and proteomic analyses identified 663 DEGs and 314 differentially expressed proteins during the infection of J11 peanut by *A. flavus* (Zhao et al., 2019). Transcriptomic network analysis from the publically available RNA-seq datasets of resistant and susceptible peanut cultivars infected with *A. flavus* revealed a series of candidate genes involved in resistance response against *A. flavus*, including genes encoding R proteins, pattern recognition receptor genes, protein P21, laccase, thaumatin-like protein 1b and pectinesterases (Jayaprakash et al., 2021). Core genes positively associated with peanut resistance to *A. flavus* were determined by weighted gene coexpression network analysis (WGCNA) and comparative transcriptome approach (Cui et al., 2022). About 18 genes encoding pattern recognition receptors (PRRs), MAPK kinase, serine/threonine kinase (STK), 1 aminocyclopropane-1-carboxylate oxidase (ACO1), pathogenesis related proteins (PR10), phosphatidylinositol transfer protein, SNARE protein SYP121, cytochrome P450, pentatricopeptide repeat (PPR) protein and pectinesterase, might contribute to peanut resistance to *A. flavus* (Cui et al., 2022).

Metabolite and miRNA profiling work for *A. flavus* resistance were also reported in peanut. Sharma et al. (2021) found that the pipecolic acid (Pip) was a key component of peanut resistance to *A. flavus* by performing untargeted metabolite profiling. And the function of Pip against *A. flavus* was validated by employing multiple resistant and susceptible peanut cultivars. Correlation analysis of small RNAs, transcriptomes and degradomes revealed a total number of 447 genes, 30 miRNAs and 21 potential miRNA/mRNA pairs showing significantly differential expression when resistant cultivar (GT-C20) and susceptible cultivar (Tifrunner) were treated with *A. flavus*. The accumulation of flavonoids in resistant and susceptible genotypes might be regulated by miR156/SPL pairs and the NBS-LRR gene expression level in resistant genotype might be regulated by miR482/2118 family (Zhao et al., 2020).

4.2.2 Viruses

TSWV, a single-stranded RNA virus, is one of the most pathogenic viruses in peanut. TSWV not only caused spotted wilt disease but also constrain peanut yield (Garcia et al., 2000). To better understand the mechanisms of peanut-TSWV interactions and peanut resistance to TSWV, QTL mapping was used to locate QTL for TSWV resistance (Table 8). The first genetic linkage map based on NC94022-derived population was constructed and a substantial QTL with a PVE of 35.8% associated with resistance to TSWV on linkage group A01 was identified (Qin et al., 2012). An enhanced genetic map from the same population was constructed and the QTLs related to multi-year TSWV phenotypic data on chromosome A01 were identified (Khera et al., 2016). About 48 QTLs with phenotypic variance explained (PVE) ranging from 3.88 to 29.14% and six QTLs associated with spotted wilt resistance were identified, among which five QTLs were found on the A01 and the other one located on A09 chromosome (Khera et al., 2016). A major QTL on chromosome A01 flanked by marker AHGS4584 and GM672, associated with spotted wilt disease with up to 22.7% PVE in a spotted wilt resistant cultivar Florida-EP^{TM} 113 was identified based on phenotypic data, and 2,431 SSR markers were screened from the two parental lines whole peanut genome (Tseng et al., 2016). Three QTLs on chromosome A01 of RIL derived from peanut lines of SunOleic 97R and NC94022 were identified using the whole genome re-sequencing approach, among which one QTL had the greatest impact on phenotypic variation, reaching 36.51%, including one 89.5 Kb genomic region with a set of genes coding for NBS-LRR proteins, strictosidine synthase-like, and chitinases (Agarwal et al., 2019). Recently, 11 QTLs for TSWV resistance were discovered using the recombinant inbred line (RIL) mapping population of "Tifrunner × GT-C20" (Pandey et al., 2017a). Zhao et al. (2018) refined the resistance QTL to a 0.8 Mb region on A01 chromosome with SSR markers.

PStV and other viruses may also infect peanut (Singh et al., 2009). However, studies on the mechanism of peanut response to various viruses using Omics technology are very few. Some reports indicate that transgenic peanuts carrying viral gene fragments can improve their antiviral ability. For example, peanut lines carrying viral coat protein gene sequences, exhibited high resistance levels to PStV (Higgins et al., 2004). Genetic engineering of peanut using genes encoding the nucleocapsid protein (N gene) of peanut bud necrosis virus was also tested against bud necrosis disease in peanut,

TABLE 8 QTL associated with Peanut responses to TSWV.

	Location	Population	Reference
tomato spotted wilt virus(TSWV)	chromosome A01	NC94022	Qin et al., 2012
	chromosome A01	NC94022	Khera et al., 2016
	chromosomeA09 and A01	SunOleic 97R × NC94022	Khera et al., 2016
	chromosome A01	Florida-EP^{TM}	Tseng et al., 2016
	chromosome A01	SunOleic 97R×NC94022	Agarwal et al., 2019
	chromosome A01	Tifrunner × GT-C20	Pandey et al., 2017a

a disease for which no persistent resistance in the existing germplasm (Rao et al., 2013). Recent research found that PeaeIF4E and PeaeIF(iso)4E, the eukaryotic translation initiation factors played important roles in PStV infection, and the silencing of PeaeIF4E and PeaeIF(iso)4E genes significantly weakened PStV accumulation in peanut (Xu et al., 2017).

4.2.3 Bacterial

R. solanacearum has a wide host range and strong long-term survival ability, making it very difficult to eradicate. In the past decade, progress has been made in the genetic behavior, trait mapping, gene discovery and diagnostic markers of peanut bacterial wilt resistance. Based on SSR and AFLP analyses, the genetic relationships of 31 peanut genotypes with different resistance levels to *R. solanacearum* were assessed, and four SSR primers and one AFLP primer were found to be effective (Jiang et al., 2007). Linkage map analysis using SSR and SNP markers found two major QTLs associated with *R. solanacearum* resistance on linkage groups LG01 and LG10 with PVE of 12 to 21% (Zhao et al., 2016). A major and stable QTL for bacterial wilt resistance (qBWB02.1) on chromosome B02 was validated by a high-density SNP map with a RIL population from the hybrid Yuanza YZ9012 with Xuzhou68-4 (Wang et al., 2018). Meanwhile, a QTL in the same linkage region was confirmed by BSA based on sequencing-based trait mapping approach and QTL-seq for recombinant inbred line (RIL) derived from a cross between cultivars Yuanza 9102 and Xuzhou 68-4 (Luo et al., 2019). In addition, two major QTLs on chromosome B02 were verified through linkage mapping and QTL-seq on the basic of a RIL population produced from the hybrid between peanut cultivars Zhonghua6 and Xuhua13 and were further validated by two diagnostic markers (Luo et al., 2020). This same major QTL on chromosome B02 was repeatedly identified with different methods and RIL populations (Wang et al., 2018; Luo et al., 2020), suggesting that it is the main QTL for peanut resistance to bacterial wilt. By applying KASP markers that were polymorphic between the two parents, the major QTL qBWA12, for resistance of peanut against bacterial, was fine mapped to a 216.7 kb region based on whole-genome resequencing data (Qi et al., 2022) (Table 9).

Beside the QTLs and markers, a number of peanut genes associated with *R. solanacearum* interactions were also reported. Using complementary DNA amplified length polymorphism (cDNA-AFLP) technique; Peng et al. (2011) studied a BW-sensitive cultivar, 'Zhonghua 12', and a BW-resistant one, 'Yuanza 9102' upon *R. solanacearum* infection, analyzed differential expression of genes associated to BW-resistance, and found 40 transcript-derived fragments (TDFs) closely related to BW resistance, which encode proteins associated with cell structure or/and protein synthesis, defense, energy, signal transduction, metabolism, cell growth and transcription,. Huang et al. (2012) screened differentially expressed genes, including those involved in jasmonic acid and ethylene signal transduction, from peanut cDNA libraries of *R. solanacearum* challenged roots and leaves. Chen et al. (2014) performed global transcriptome profiling on the *R. solanacearum*-infected roots of peanut susceptible (S) and resistant (R) genotypes, and found that the down-regulation of primary metabolism and the genotype-specific expression pattern of defense related DEGs (R gene, cell wall genes, LRR-RLK, etc.) contributing to the resistance difference between S and R genotypes. Recently, 174 WRKY genes (AhWRKY) were identified from the cultivated peanut genome, their differential expression patterns were analyzed in sensitive and resistant peanut genotypes infected with the *R. solanacearum*, and the possible roles of candidate WRKY genes were identified in peanut resistance against *R. solanacearum* infection (Yan et al., 2022). A series of candidate genes with possible bacterial wilt resistance were directly cloned from peanut and functionally studied. Overexpression of the peanut AhRRS5 (a novel peanut NBS-LRR gene) or the AhRLK1 (CLAVATA1-like leucine-rich repeat receptor-like kinase) or the AhGLK1b (GOLDEN2-like Transcription Factor) enhanced plant disease resistance to *R. solanacearum* (Zhang et al., 2017; Zhang et al., 2019; Ali et al., 2020).

TABLE 9 QTL associated with Peanut responses to Bacterial wilt.

	Location	Population	Reference
Bacterial wilt	chromosome B02	Yuanza YZ9012 × Xuzhou68-4	Wang et al., 2018
	chromosome B02	Zhonghua6 and Xuhua13	Luo et al., 2020
	mapped to a 216.68 kb physical region	*A. hypogaea*	Qi et al., 2022

5 Application of omics research in breeding program

In the last two decades, using the most successful approach namely marker-assisted selection (MAS) or marker-assisted backcrossing (MABC), diagnostic markers have successfully been developed in groundnut for resistance to nematode, rust and LLS. The first excellent example is the marker-assisted improvement of popular cultivar for nematode resistance in USA. The discovery of resistance to *M. arenaria* in wild *Arachis* species and the interspecific hybrid, TxAG-6, allowed the development of the first peanut cultivar resistant to *M. arenaria* (Simpson and Starr, 2001). The interspecific *Arachis* hybrid TxAG-6 was the source of this resistance and the donor parent in a backcross breeding program to introgress resistance into cultivated peanut. Gene segments with resistance from TxAG-6 were introduced into peanut cultivars through an interspecific hybridization backcrossing, two root-knot nematode-resistant varieties, COAN and NemaTAM were first developed with marker-assisted backcrossing is in the USA (Simpson and Starr, 2001; Simpson et al., 2003; Church et al., 2005). By far, at least other four commercial nematode-resistant cultivars (Tifguard, Webb, TifN/V OL, and Georgia 14N) resulting from this cross have been released (Denwar et al., 2021).

The major peanut rust resistance-related QTLs and markers were almost mapped from two recombinant inbred line (RIL) mapping populations, namely 'TAG 24' (susceptible) × 'GPBD 4' (resistant) and 'TG 26' (susceptible) × 'GPBD 4' (Khedikar et al., 2010; Sujay et al., 2012). The QTL region controlling rust resistance, including the disease resistance-linked markers (IPAHM103, GM2079, GM1536 and GM2301) from the disease-resistant donor GPBD 4, were introgressed into two elite peanut varieties ('TAG 24' and 'ICGV 91114') and one old but popular variety ('JL 24') through MABC in India. The backcrossed homozygous lines (BC3F2) were obtained in just three years of time and shown significant increase in rust resistance (Varshney et al., 2014). In addition, TMV2 is a very popular groundnut variety among the Indian farmers but is highly susceptible to LLS and rust. The LLS and rust resistance linked markers (GM2009, GM2079, GM2301, GM1839 and IPAHM103) from the disease-resistant donor GPBD 4 were introgressed into TMV2 using MABC approach, and two homozygous backcross lines, namely TMG-29 and TMG-46, were obtained which showed resistance to rust and LLS (Kolekar et al., 2017). Recently, the above linked and validated markers for resistance to rust and LLS were used to improve three popular Indian cultivars (GJG 9, GG 20, and GJGHPS 1) using MABC, and the Phenotyping of the ILs, using the 58 K SNP array for assessing background genome recovery across the chromosomes (Pandey et al., 2017c), revealed their disease resistance scores comparable to the resistant parent GPBD 4 (Shasidhar et al., 2020).

6 Conclusion and perspective

This review summarizes various omics studies in peanut biotic stress in the past two decades emphasizing on peanut resistance to various biotic pests and pathogens. Progresses in the field of peanut responding to biotic stress are accumulating, and large amount of omics data have been produced to decipher the molecule clues between peanut and the infesting agents. In the implementation of marker-assisted selection, the first step is the identification of genomic regions conferring disease resistance in wild species, which can speed up the introgression of wild disease resistance genes with less linkage drag. Also, the identified disease resistance genes can be deployed in biotechnology and genomics-assisted breeding for the development of disease resistant cultivars to reduce the yield loss in peanut production. However, the accumulated data over the past years are intertwined and disorganized, and have not been rationally sorted out. The technology that produces large amount of Omics data has surpassed our ability to analyze and utilize them. Big data is a common phenomenon of this era, but it is extremely urgent to develop technology in dealing with the Omics data, and most importantly to use the Omics data in crop breeding for better resistance to biotic and abiotic stresses.

Due to the peanut genome complexity, research on peanut Omics are relatively lagged as compared with other crops as well as with the model plant species. Not many results are obtained from peanut biotic stresses by using the Omics tools such as transcriptomics, proteomics and other technologies. For the major peanut pests and pathogens, efforts are mostly on the most important pests and diseases such as the root-knot nematodes, insects and fungi, while less attention is paid to the viral and bacterial pathogens. Thus, more efforts are needed in future Omics research on biotic stresses affecting peanuts.

In addition, results from Omics data including putative QTLs, molecular markers, candidate genes, metabolites, proteins, etc., need to be cross tested with wet lab research before they can be reliably used in scientific research and commercial breeding. For those coarsely mapped QTLs, other technologies may be needed to precisely map them to a smaller region on the corresponding chromosome, and eventually clone them. Candidate genes that may play a role in peanut disease and insect resistance, functional genetics are needed to overexpress them in transgenic research or their mutations need to be created by CRISPR genome editing technology to verify their related functions. On the other hand, Omics data are valuable in genome editing to design efficient CRISPR targets, with reduced off-target mutations in crop breeding. With the success of peanut genome editing with the CRISPR technology, it is possible to use Omics data in peanut molecular breeding for pest and disease resistance (Zhang et al., 2021).

Author contributions

RH, SZ, HL, CG and WY write this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Open Competition Program of Top Ten Critical Priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (2022SDZG05), the National Natural Science Foundation of China (322703553 for SZ; U22A20475 for WY) and the Shenzhen Commission of Science and Technology Innovation Projects (JCYJ20190808143207457, JCYJ20180305124101630, JCYJ20170818094958663) to WY; A research grant from the China Good Crop Company (Shenzhen) Limited to HL.

Conflict of interest

Author WY was employed by China Good Crop Company Shenzhen Limited.

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

References

Agarwal, G., Clevenger, J., Kale, S. M., Wang, H., Pandey, M. K., Choudhary, D., et al. (2019). A recombination bin-map identified a major QTL for resistance to tomato spotted wilt virus in peanut (*Arachis hypogaea*). *Sci. Rep.* 9, 18246. doi: 10.1038/s41598-019-54747-1

Agarwal, G., Clevenger, J., Pandey, M. K., Wang, H., Shasidhar, Y., Chu, Y., et al. (2018). High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnol. J.* 16 (11), 1954–1967. doi: 10.1111/pbi.12930

Agrios, G. N. (2005). "Plant diseases caused by nematodes," in *Introduction to plant pathology*, *fifth ed* (London: Elsevier Academic Press), 826–872.

Ahmad, S., Nawade, B., Sangh, C., Mishra, G. P., Bosamia, T. C., Kumar., N., et al. (2020). Identification of novel QTLs for late leaf spot resistance and validation of a major rust QTL in peanut (*Arachis hypogaea l.*). *3 Biotech.* 10, 458. doi: 10.1007/s13205-020-02446-4. T. R.

Ali, N., Chen, H., Zhang, C., Khan, S. A., Gandeka, M., Xie, D., et al. (2020). Ectopic expression of AhGLK1b (GOLDEN2-like transcription factor) in arabidopsis confers dual resistance to fungal and bacterial pathogens. *Genes (Basel)* 11, 343. doi: 10.3390/genes11030343

Anco, D. J., Thomas, J. S., Jordan, D. L., Shew, B. B., Monfort, W. S., Mehl, H. L., et al. (2020). Peanut yield loss in the presence of defoliation caused by late or early leaf spot. *Plant Dis.* 104 (5), 1390–1399. doi: 10.1094/PDIS-11-19-2286-RE

Arya, S. S., Salve, A. R., and Chauhan, S. (2015). Peanuts as functional food: a review. *J. Food Sci. Technol.* 53 (1), 31–41. doi: 10.1007/s13197-015-2007-9

Ballén-Taborda, C., Chu, Y., Ozias-Akins, P., Holbrook, C. C., Timper, P., Jackson, S. A., et al. (2022). Development and genetic characterization of peanut advanced backcross lines that incorporate root-knot nematode resistance from *Arachis* stenosperma. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.785358

Ballén-Taborda, C., Chu, Y., Ozias-Akins, P., Timper, P., Holbrook, C. C., Jackson, S. A., et al. (2019). A new source of root-knot nematode resistance from *Arachis* stenosperma incorporated into allotetraploid peanut (*Arachis hypogaea*). *Sci. Rep.* 9, 17702. doi: 10.1038/s41598-019-54183-1

Ballén-Taborda, C., Chu, Y., Ozias-Akins, P., Timper, P., Jackson, S., Bertioli, D., et al. (2021). Validation of resistance to root-knot nematode incorporated in peanut from the wild relative *Arachis stenosperma*. *Agron. J.* 113, 2293–2230. doi: 10.1002/agj2.20654

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517

Bertioli, D. J., Clevenger, J., Godoy, I. J., Stalker, H. T., Wood, S., Santos, J. F., et al. (2021). Legacy genetics of *Arachis* cardenasii in the peanut crop shows the profound benefits of international seed exchange. *Proc. Natl. Acad. Sci. U.S.A.* 118 (38), e2104899118. doi: 10.1073/pnas.2104899118

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, 877–884. doi: 10.1038/s41588-019-0405-z

Bertioli, D. J., Leal-Bertioli, S. C., Lion, M. B., Santos, V. L., Pappas, G., Cannon, S. B., et al. (2003). A large scale analysis of resistance gene homologues in *Arachis*. *Mol. Genet. Genomics* 270, 34–45. doi: 10.1007/s00438-003-0893-4

Bi, Y. P., Li, G. C., Wang, X. L., Li, J. W., Shan, L., Guo, B. T., et al. (1999). Cloning and sequencing of peanut stripe virus coat protein gene. *J. Agric. Biotechnol.* 7, 211–214.

Blackman, R. L., and Eastop, V. F. (2007). "Taxonomic issues," in *Aphids as crop pests*. Eds. H. F. van Emden and R. Harrington (U. K: CAB International), 1–29.

Blount, A. R., Pittman, R. N., Smith, B. A., Morgan, R. N., Dankers, W., Sprenkel, R. K., et al. (2002). First report of peanut stunt virus in perennial peanut in north Florida and southern Georgia. *Plant Dis.* 86, 326. doi: 10.1094/PDIS.2002.86.3.326C

Bosamia, T. C., Dodia, S. M., Mishra, G. P., Ahmad, S., Joshi, B., Thirumalaisamy, P. P., et al. (2020). Unraveling the mechanisms of resistance to sclerotium rolfsii in peanut (*Arachis hypogaea l.*) using comparative RNA-seq analysis of resistant and susceptible genotypes. *PloS One* 15, e0236823. doi: 10.1371/journal.pone.0236823

Chaudhari, S., Khare, D., Patil, S. C., Sundravadana, S., Variath, M. T., Sudini, H. K., et al. (2019). Genotype × environment studies on resistance to late leaf spot and rust in genomic selection training population of peanut (*Arachis hypogaea l.*). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01338

Chen, K. R., Xu, Z. Y., Zhang, Z. Y., and Chen, J. X. (1990). Seed transmission of peanut stripe virus (PStV) in peanut III. the changes of the rate of PStV infected peanut seeds between high and low seed-borne lines at the peanut pods development. *Chin. J. Oil Crop Sci.* 4, 79–82.

Chen, Y., Ren, X., Zhou, X., Huang, L., Yan, L., Lei, Y., et al. (2014). Dynamics in the resistant and susceptible peanut (Arachis hypogaea l.) root transcriptome on infection with the ralstonia solanacearum. *BMC Genomics* 15. doi: 10.1186/1471-2164-15-1078

Choi, H. S., Kim, J. S., Cheon, J. U., Choi, J. K., Pappu, S. S., and Pappu, H. R. (2001). First report of peanut stripe virus (Family potyviridae) in south Korea. *Plant Dis.* 85, 679. doi: 10.5423/PPJ.2006.22.1.097

Choi, H. S., Kim, M., Park, J. W., Cheon, J. U., Kim, K. H., Kim, J. S., et al. (2006). Occurrence of bean common mosaic virus (BCMV) infecting peanut in Korea. *Plant Pathol. J.* 22, 97–102. doi: 10.5423/PPJ.2006.22.1.097

Chu, Y., Chee, P., Culbreath, A., Isleib, T. G., Holbrook, C. C., and Ozias-Akins, P. (2019). Major QTLs for resistance to early and late leaf spot diseases are identified on chromosomes 3 and 5 in peanut (*Arachis hypogaea*). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00883

Church, G. T., Starr, J. L., and Simpson, C. E. (2005). A recessive gene for resistance to meloidogyne arenaria in interspecific *Arachis* spp. hybrids. *J. Nematol* 37 (2), 178–184.

Clevenger, J., Chu, Y., Chavarro, C., Agarwal, G., Bertioli, D. J., Leal-Bertioli, S. C. M., et al. (2017). Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Mol. Plant* 10 (2), 309–322. doi: 10.1016/j.molp.2016.11.015

Cui, M., Han, S., Wang, D., Haider, M. S., Guo, J., Zhao, Q., et al. (2022). Gene Co-expression network analysis of the comparative transcriptome identifies hub genes associated with resistance to aspergillus flavus l. @ in cultivated peanut (*Arachis hypogaea l.*). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.899177

Culbreath, A. K., and Srinivasan, R. (2011). Epidemiology of spotted wilt disease of peanut caused by tomato spotted wilt virus in the southeastern u. s. *Virus Res.* 159, 101–109. doi: 10.1016/j.virusres.2011.04.014

Culbreath, A. K., Todd, J. W., and Brown, S. L. (2003). Epidemiology and management of tomato spotted wilt in peanut. *Annu. Rev. Phytopathol.* 41, 53–75. doi: 10.1016/j.virusres.2009.01.009

Dang, J. L., Horvath, D. M., and Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science* 341, 746–751. doi: 10.1126/science.1236011

Dang, P. M., Lamb, M. C., Bowen, K. L., and Chen, C. Y. (2019). Identifcation of expressed r-genes associated with leaf spot diseases in cultivated peanut. *Mol. Biol. Rep.* 46, 225–239. doi: 10.1007/s11033-018-4464-5

Dang, P. M., Lamb, M. C., and Chen, C. Y. (2021). Association of differentially expressed r-gene candidates with leaf spot resistance in peanut (*Arachis hypogaea l.*). *Mol. Biol. Rep.* 48, 323–334. doi: 10.1007/s11033-020-06049-3

Decraemer, W., and Hunt, D. (2006). "Structure and classifcation," In R. N. Perry and M. Moens (Eds.), *Plant Nematology*, pp. 3–32.

Demski, J. W., and Lovell, G. R. (1985). Peanut stripe virus and the distribution of peanut seed. *Plant Dis.* 69, 734–738.

Denwar, N. N., Simpson, C. S., Starr, J. L., Wheeler, T. A., and Burow, M. D. (2021). Evaluation and selection of interspecific lines of groundnut (*Arachis hypogaea* l.) for resistance to leaf spot disease and for yield improvement. *Plants (Basel)* 10 (5), 873. doi: 10.3390/plants10050873

Dodia, S. M., Joshi, B., Gangurde, S. S., Thirumalaisamy, P. P., Mishra, G. P., Narandrakumar, D., et al. (2019). Genotyping-by-sequencing based genetic mapping reveals large number of epistatic interactions for stem rot resistance in groundnut. *Theor. Appl. Genet.* 132, 1001–1016. doi: 10.1007/s00122-018-3255-7

Dong, W. B., Holbrook, C. C., Timper, P., Brenneman, T. B., Chu, Y., and Ozias-Akins, P. (2008). Resistance in peanut cultivars and breeding lines to three root-knot nematode species. *Plant Dis.* 92, 631–638. doi: 10.1094/PDIS-92-4-0631

Gangurde, S. S., Nayak, S. N., Joshi, P., Purohit, S., Sudini, H. K., Chitikineni, A., et al. (2021). Comparative transcriptome analysis identified candidate genes for late leaf spot resistance and cause of defoliation in groundnut. *Int. J. Mol. Sci.* 22, 4491. doi: 10.3390/ijms22094491

Garcia, L. E., Brandenburg, R. L., and Bailey, J. E. (2000). Incidence of tomato spotted wilt virus (Bunyaviridae) and tobacco thrips in virginiatype peanuts in north Carolina. *Plant Dis.* 84, 459–464. doi: 10.1094/PDIS.2000.84.4.459

Garcia, G. M., Stalker, H. T., and Kochert, G. (1995). Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea l.*) using RFLP and RAPD markers. *Genome* 38, 166–176. doi: 10.1139/g95-021

Genin, S. (2010). Molecular traits controlling host range and adaptation to plants in ralstonia solanacearum. *New Phytol.* 187, 920–928. doi: 10.1111/j.1469-8137.2010.03397.x

Gong, L., Han, S., Yuan, M., Ma, X., Hagan, A., and He, G. (2020). Transcriptomic analyses reveal the expression and regulation of genes associated with resistance to early leaf spot in peanut. *BMC Res. Notes* 13, 381. doi: 10.1186/s13104-020-05225-9

Grichar, W. J., Besler, B. A., and Jaks, A. J. (1998). Groundnut (Arachis hypogaea l.) cultivar response to leaf spot disease development under four disease management programs. *Peanut Sci.* 25, 35–39. doi: 10.3146/i0095-3679-25-1-9

Guimaraes, P. M., Guimaraes, L. A., Morgante, C. V., Silva, O., Araujo, A. C., Martins, A. C., et al. (2015). Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. *PloS One* 10, e0140937. doi: 10.1371/journal.pone.0140937

Guo, B., Chen, Z.-Y., Lee, R. D., and Scully, B. T. (2008). Drought stress and preharvest aflatoxincontamination in agricultural commodity: genetics, genomics and proteomics. *J. Integr. Plant Biol.* 50, 1281–1291. doi: 10.1111/j.1744-7909.2008.00739.x

Guo, B., Fedorova, N. D., Chen, X., Wan, C. H., Wang, W., Nierman, W. C., et al. (2011). Gene expression profiling and identification of resistance genes to aspergillus flavus infection in peanut through EST and microarray strategies. *Toxins (Basel)* 3, 737–753. doi: 10.3390/toxins3070737

Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., Ralser, M., et al. (2017). Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. *Curr. Opin. Syst. Biol.* 6, 37–45. doi: 10.1016/j.coisb.2017.08.009

Han, S., Liu, H., Yan, M., Qi, F., Wang, Y., Sun, Z., et al. (2017). Differential gene expression in leaf tissues between mutant and wild-type genotypes response to late leaf spot in peanut (*Arachis hypogaea l.*). *PloS One* 12, e0183428. doi: 10.1371/journal.pone.0183428

Han, S., Yuan, M., Clevenger, J. P., Li, C., Hagan, A., Zhang, X., et al. (2018). A SNP-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (*Arachis hypogaea l.*). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01012

Harsulkar, A. M., Giri, A. P., Patankar, A. G., Gupta, V. S., Sainani, M. N., and Ranjekar, P. K. (1999). Deshpande VV. successive use of non-host plant proteinase inhibitors required for effective inhibition of helicoverpa armigera gut proteinases and larval growth. *Plant Physiol.* 121, 497–506. doi: 10.1104/pp.121.2.497

Hayward, A. C. (1991). Biology and epidemiology of bacterial wilt caused by pseudomonas solanacearum. *Annu. Rev. Phytopathol.* 29, 65–87. doi: 10.1146/annurev.py.29.090191.000433

Herselman, L., Thwaites, R., Kimmins, F. M., Courtois, B., van der, Merwe., P. J., and Seal, S. E. (2004). Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea l.*) resistance to the aphid vector of groundnut rosette disease. *Theor. Appl. Genet.* 109, 1426–1433. doi: 10.1007/s00122-004-1756-z

Higgins, C. M., Hall, R. M., Mitter, N., Cruickshank, A., and Dietzgen, R. G. (2004). Peanut stripe potyvirus resistance in peanut (*Arachis hypogaea l.*) plants carrying viral coat protein gene sequences. *Transgenic Res.* 13, 59–67. doi: 10.1023/b:trag.0000017166.29458.74

Holbrook, C., and Stalker, H. (2003). Peanut breeding and genetic resources. *Plant Breed Rev.* 22, 297–356. doi: 10.1002/9780470650202.ch6

Huang, J., Yan, L., Lei, Y., Jiang, H., Ren, X., and Liao, B. (2012). Expressed sequence tags in cultivated peanut (*Arachis hypogaea*): discovery of genes in seed development and response to ralstonia solanacearum challenge. *J. Plant Res.* 125, 755–769. doi: 10.1007/s10265-012-0491-9

Jayaprakash, A., Roy, A., Thanmalagan, R. R., Arunachalam, A., and Ptv, L. (2021). Immune response gene coexpression network analysis of *Arachis hypogaea* infected with aspergillus flavus. *Genomics* 113, 2977–2988. doi: 10.1016/j.ygeno.2021.06.027

Jiang, H., Liao, B., Ren, X., Lei, Y., Mace, E., Fu, T., et al. (2007). Comparative assessment of genetic diversity of peanut (*Arachis hypogaea l.*) genotypes with various levels of resistance to bacterial wilt through SSR and AFLP analyses. *J. Genet. Genomics* 34, 544–554. doi: 10.1016/S1673-8527(07)60060-5

Jiang, Y., Luo, H., Yu, B., Ding, Y., Kang, Y., Huang, L., et al. (2021). High-density genetic linkage map construction using whole-genome resequencing for mapping QTLs of resistance to aspergillus flavus infection in peanut. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.745408

Jiang, G., Wei, Z., Xu, J., Chen, H., Zhang, Y., She, X., et al. (2017). Bacterial wilt in China: history, current status, and future perspectives. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01549

Jogi, A., Kerry, J. W., Brenneman, T. B., Leebens-Mack, J. H., and Gold, S. E. (2016). Identification of genes differentially expressed during early interactions between the stem rot fungus (Sclerotium rolfsii) and peanut (*Arachis hypogaea*) cultivars with increasing disease resistance levels. *Microbiol. Res.* 184, 1–12. doi: 10.1016/j.micres

Jones, J. T., Haegeman, A., Danchin, E. G., Gaur, H. S., Helder, J., Jones, M. G., et al. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* 14, 946–961. doi: 10.1111/mpp.12057

Keshavareddy, G., Rohini, S., Ramu, S. V., Sundaresha, S., Kumar, A. R., Kumar, P. A., et al. (2013). Transgenics in groundnut (*Arachis hypogaea l.*) expressing cry1AcF gene for resistance to spodoptera litura (F.). *Physiol. Mol. Biol. Plants* 19, 343–352. doi: 10.1007/s12298-013-0182-6

Kew, M. C. (2013). Aflatoxins as a cause of hepatocellular carcinoma. j. gastrointestin. *Liver Dis.* 22, 305–310. doi: 10.1586/17474124.2013.832500

Khan, S. A., Chen, H., Deng, Y., Chen, Y., Zhang, C., Cai, T., et al. (2020). High-density SNP map facilitates fine mapping of QTLs and candidate genes discovery for aspergillus flavus resistance in peanut (*Arachis hypogaea*). *Theor. Appl. Genet.* 133, 2239–2257. doi: 10.1007/s00122-020-03594-0

Khedikar, Y. P., Gowda, M. V., Sarvamangala, C., Patgar, K. V., Upadhyaya, H. D., Varshney, R. K., et al. (2010). A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea l.*). *Theor. Appl. Genet.* 121 (5), 971–984. doi: 10.1007/s00122-010-1366-x

Khera, P., Pandey, M. K., Wang, H., Feng, S., Qiao, L., Culbreath, A. K., et al. (2016). Mapping quantitative trait loci of resistance to tomato spotted wilt virus and leaf spots in a recombinant inbred line population of peanut (*Arachis hypogaea l.*) from SunOleic 97R and NC94022. *PloS One* 11, e0158452. doi: 10.1371/journal.pone.0158452

Kojima, M., Kamada-Nobusada, T., Komatsu, H., Takei, K., Kuroha, T., Mizutani, M., et al. (2009). Highly sensitive and high-throughput analysis of plant hormones using MS-probe modification and liquid chromatography-tandem mass spectrometry: an application for hormone profiling in oryza sativa. *Plant Cell Physiol.* 50, 1201–1214. doi: 10.1093/pcp/pcp057

Kolekar, R. M., Sukruth, M., Shirasawa, K., Nadaf, H. L., Motagi, B. N., Lingaraju, S., et al. (2017). Marker-assisted backcrossing to develop foliar disease-resistant genotypes in TMV 2 variety of peanut (*Arachis hypogaea l.*). *Plant Breed* 136, 948–953. doi: 10.1111/pbr.12549

Krishna, G., Singh, B. K., Kim, E. K., Morya, V. K., and Ramteke, P. W. (2015). Progress in genetic engineering of peanut (*Arachis hypogaea l.*)–a review. *Plant Biotechnol. J.* 13, 147–162. doi: 10.1111/pbi.12339

Kumar, R., Bohra, A., Pandey, A. K., Pandey, M. K., and Kumar, A. (2017). Metabolomics for plant improvement: status and prospects. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01302

Kumar, K. R., and Kirti, P. B. (2011). Differential gene expression in *Arachis diogoi* upon interaction with peanut late leaf spot pathogen, phaeoisariopsis personata and characterization of a pathogen induced cyclophilin. *Plant Mol. Biol.* 75, 497–513. doi: 10.1007/s11103-011-9747-3

Kumar, D., and Kirti, P. B. (2015). Transcriptomic and proteomic analyses of resistant host responses in *Arachis diogoi* challenged with late leaf spot pathogen, phaeoisariopsis personata. *PloS One* 10, e0117559. doi: 10.1371/journal.pone.0117559

Kumar, K. R., Srinivasan, T., and Kirti, P. B. (2009). A mitogen-activated protein kinase gene, AhMPK3 of peanut: molecular cloning, genomic organization, and heterologous expression conferring resistance against spodoptera litura in tobacco. *Mol. Genet. Genomics* 282, 65–81. doi: 10.1007/s00438-009-0446-6

Le, C. N., Mendes, R., Kruijt, M., and Raaijmakers, J. M. (2012). Genetic and phenotypic diversity of sclerotium rolfsii in groundnut felds in central Vietnam. *Plant Dis.* 96, 389–397. doi: 10.1094/PDIS-06-11-0468

Leal-Bertioli, S. C., Cavalcante, U., Gouvea, E. G., Ballén-Taborda, C., Shirasawa, K., Guimarães, P. M., et al. (2015). Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker-assisted selection. *G3 (Bethesda)* 5, 1403–1413. doi: 10.1534/g3.115.018796

Leal-Bertioli, S. C., José, A. C., Alves-Freitas, D. M., Moretzsohn, M. C., Guimarães, P. M., Nielen, S., et al. (2009). Identification of candidate genome regions controlling disease resistance in *Arachis*. *BMC Plant Biol.* 9, 112. doi: 10.1186/1471-2229-9-112

Leal-Bertioli, S. C., Moretzsohn, M. C., Roberts, P. A., Ballén-Taborda, C., Borba, T. C., Valdisser, P. A., et al. (2016). Genetic mapping of resistance to meloidogyne arenaria in *Arachis stenosperma*: a new source of nematode resistance for peanut. *G3 (Bethesda)* 6, 377–390. doi: 10.1534/g3.115.023044

Liang, Y., Barring, M., Wang, S., and Septiningsih, E. M. (2017). Mapping QTLs for leafspot resistance in peanut using SNP-based next-generation sequencing markers. *Plant Breed Biotech.* 5, 115–122. doi: 10.9787/pbb.2017.5.2.115

Lister, R., Gregory, B. D., and Ecker, J. R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.* 12, 107–118. doi: 10.1016/j.pbi.2008.11.004

Lu, Q., Liu, H., Hong, Y., Li, H., Liu, H., Li, X., et al. (2018). Consensus map integration and QTL meta-analysis narrowed a locus for yield traits to 0.7 cM and refined a region for late leaf spot resistance traits to 0.38 cM on linkage group A05 in peanut (*Arachis hypogaea l.*). *BMC Genomics* 19, 887. doi: 10.1186/s12864-018-5288-3

Luo, H., Pandey, M. K., Khan, A. W., Wu, B., Guo, J., Ren, X., et al. (2019). Next-generation sequencing identified genomic region and diagnostic markers for resistance to bacterial wilt on chromosome B02 in peanut (*Arachis hypogaea l.*). *Plant Biotechnol. J.* 17, 2356–2369. doi: 10.1111/pbi.13153

Luo, H., Pandey, M. K., Zhi, Y., Zhang, H., Xu, S., Guo, J., et al. (2020). Discovery of two novel and adjacent QTLs on chromosome B02 controlling resistance against bacterial wilt in peanut variety zhonghua 6. *Theor. Appl. Genet.* 133, 1133–1148. doi: 10.1007/s00122-020-03537-9

Ma, Y., Yang, J., Yang, D., Qin, G., and Zu, J. (2022). Development of 1,5-Diaryl-Pyrazole-3-Formate analogs as antifungal pesticides and their application in controlling peanut stem rot disease. *Front. Microbiol.* 12, 728173. doi: 10.3389/fmicb.2021.728173

Mahatma, M. K., Thawait, L. K., Jadon, K. S., Thirumalaisamy, P. P., Bishi, S. K., Rathod, K. J., et al. (2021). Metabolic profiling for dissection of late leaf spot disease resistance mechanism in groundnut. *Physiol. Mol. Biol. Plants* 27, 1027–1041. doi: 10.1007/s12298-021-00985-5

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x

Martins, A. C. Q., Mehta, A., Murad, A. M., Mota, A. P. Z., Saraiva, M. A. P., Araújo, A. C. G., et al. (2020). Proteomics unravels new candidate genes for meloidogyne resistance in wild *Arachis*. *J. Proteomics* 217, 103690. doi: 10.1016/j.jprot.2020

Mayee, C. D., Datar, V. V., Raychaudhuri, S. P., and Verma, J. P. (1988). Diseases of groundnut in the tropics. *Rev. Trop. Plant Pathol.* 5, 85–118.

McDonald, D., Subrahmanyam, P., Gibbons, R. W., and Smith, D. H. (1985). Early and late leaf spots of groundnut. *Patancheru: ICRISAT Inf. Bulletin*. 12, 1–19.

McGahan, P., Lloyd, R. J., and Rynne, K. P. (1991). "The cost of helicoverpa in Queensland crops," in *A review of heliothis research in Australia conference and workshop series QC91006*. Eds. H. Twine and M. P. Zalucki (Brisbane: Queensland Department of PrimaryIndustries), 11–28.

Mehan, V. K., Mayee, C. D., and McDonald, D. (1994). Management of sclerotium rolfsiicaused stem and pod rots of groundnut – a critical review. *Int. J. Pest Manage* 40, 313–320. doi: 10.1080/09670879409371906

Merwe, P., Subrahmanyan, P., Kimmins, F. M., and Willekens, J. (2001). Mechanisms of resistance to groundnut rosette. *Int. Arachis Newsl* 21, 43–46.

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-313x.1999.t01-1-00606.x

Middleton, K. J., and Saleh, N. (1988). "Peanut stripe virus disease in Indonesia and the ACIAR project". in *in Coordination of Research on Peanut Stripe Virus*. (Patancheru: ICRISAT), pp. 4–6.

Mochida, K., and Shinozaki, K. (2010). Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.* 51, 497–523. doi: 10.1093/pcp/pcq027

Mondal, S., and Badigannavar, A. M. (2018). Mapping of a dominant rust resistance gene revealed two r genes around the major Rust_QTL in cultivated peanut (*Arachis hypogaea l.*). *Theor. Appl. Genet.* 131, 1–11. doi: 10.1007/s00122-018-3106-6

Mondal, S., Mohamed Shafi, K., Raizada, A., Song, H., Badigannavar, A. M., and Sowdhamini, R. (2022). Development of candidate gene-based markers and map-based cloning of a dominant rust resistance gene in cultivated groundnut (*Arachis hypogaea* l.). *Gene* 827, 146474. doi: 10.1016/j.gene.2022.146474

Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C., Valls, J. F., and Bertioli, D. J. (2013). A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* 111, 113–126. doi: 10.1093/aob/mcs237

Morgante, C. V., Brasileiro, A. C. M., Roberts, P. A., Guimaraes, L. A., Araujo, A. C. G., Fonseca, L. N., et al. (2013). A survey of genes involved in *Arachis stenosperma* resistance to meloidogyne arenaria race 1. *Funct. Plant Biol.* 40, 1298–1309. doi: 10.1071/FP13096

Mota, A. P. Z., Fernandez, D., Arraes, F. B. M., Petitot, A. S., de Melo, B. P., de Sa, M. E. L., et al. (2020). Evolutionarily conserved plant genes responsive to root-knot nematodes identified by comparative genomics. *Mol. Genet. Genomics* 295, 1063–1078. doi: 10.1007/s00438-020-01677-7

Mota, A. P. Z., Vidigal, B., Danchin, E. G. J., Togawa, R. C., Leal-Bertioli, S. C. M., Bertioli, D. J., et al. (2018). Comparative root transcriptome of wild *Arachis* reveals NBS-LRR genes related to nematode resistance. *BMC Plant Bio* l18, 159. doi: 10.1186/s12870-018-1373-7

Nagy, E. D., Chu, Y., Guo, Y., Khanal, S., Tang, S., Li, Y., et al. (2010). Recombination is suppressed in an alien introgression in peanut harboring rma, a dominant root-knot nematode resistance gene. *Mol. Breed.* 26, 357–370. doi: 10.1007/s11032-010-9430-4

Nandini, D., Mohan, J. S. S., and Singh, G. (2010). Induction of systemic acquired resistance in *Arachis hypogaea l.* by sclerotium rolfsii derived elicitors. *J. Phytopathol.* 158, 594–600. doi: 10.1111/j.1439-0434.2009.01666.x

Okabe, I., and Matsumoto, N. (2000). Population structure of sclerotium rolfsii in peanut fields. *Mycoscience* 41, 145–148. doi: 10.1007/BF02464323

Padgham, D. E., Kimmins, F. M., and Rao, G. V. R. (1990). Resistance in groundnut (*Arachis hypogaea* l.) to aphis craccivora (Koch). *Ann. Appl. Biol.* 117, 285–294. doi: 10.1111/j.1744-7348.1990.tb04214.x

Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., et al. (2017c). Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7, 40577. doi: 10.1038/srep40577

Pandey, M. K., Khan, A. W., Singh, V. K., Vishwakarma, M. K., Shasidhar, Y., Kumar, V., et al. (2017b). QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea l.*). *Plant Biotechnol. J.* 15, 927–941. doi: 10.1111/pbi.12686

Pandey, M. K., Kolekar, R. M., Khedikar, Y. P., and Varshney, (2016). QTL mapping for late leaf spot and rust resistance using an improved genetic map and extensive phenotypic data on a recombinant inbred line population in peanut (*Arachis hypogaea l.*). *Euphytica Int. J. Plant Breed* 209 (1), 147–156. doi: 10.1007/s10681-016-1651-0

Pandey, M. K., Pandey, A. K., Kumar, R., Nwosu, C. V., Guo, B., Wright, G. C., et al. (2020). Translational genomics for achieving higher genetic gains in groundnut. *Theor. Appl. Genet.* 133 (5), 1679–1702. doi: 10.1007/s00122-020-03592-2

Pandey, M. K., Wang, H., Khera, P., Vishwakarma, M. K., Kale, S. M., Culbreath, A. K., et al. (2017a). Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (*Arachis hypogaea l.*). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00025

Pappu, H. R., Jones, R. A. C., and Jain, R. K. (2009). Global status of tospovirus epidemics in diverse cropping systems: successes achieved and challenges ahead. *Virus Res.* 141, 219–236. doi: 10.1016/j.virusres.2009.01.009

Pasupuleti, J., Pandey, M. K., Manohar, S. S., Variath, M. T., Nallathambi, P., Nadaf, H. L., et al. (2016). Foliar fungal disease-resistant introgression lines of groundnut (*Arachis hypogaea l.*) record higher pod and haulm yield in multilocation testing. *Plant Breed* 135, 355–366. doi: 10.1111/pbr.12358

Peng, W. F., Lv, J. W., Ren, X. P., Huang, L., Zhao, X. Y., Wen, Q. G., et al. (2011). Differential expression of genes related to bacterial wilt resistance in peanut (*Arachis hypogaea l.*). *Yi Chuan* 33, 389–396. doi: 10.3724/sp.j.1005.2011.00389

Pérez-Clemente, R. M., Vives, V., Zandalinas, S. I., López-Climent, M. F., Muñoz, V., and Gómez-Cadenas, A. (2013). Biotechnological approaches to study plant responses to stress. *BioMed. Res. Int.* 2013, 654120. doi: 10.1155/2013/654120

Pittet, A. (1998). Natural occurrence of mycotoxins in foods and feeds: an update review. *Rev. Med. Vet.* 149, 479–492. doi: 10.1016/j.fgb.2014.02.005

Prasad, M. N. R., and Gowda, M. V. C. (2006). Mechanisms of resistance to tobacco cutworm (Spodoptera litura f.) and their implications to screening for resistance in groundnut. *Euphytica* 149, 387–399. doi: 10.1007/s10681-006-9088-5

Pratissoli, D., Pirovani, V. D., and Lima, W. L. (2015). Occurrence of helicoverpa armigera (Lepidoptera: noctuidae) on tomato in the espírito Santo state. *Hortic. Bras.* 33, 101–105. doi: 10.1590/S0102-053620150000100016

Proite, K., Carneiro, R., Falco, R., Gomes, A., and Bertioli, D. (2010). Post-infection development and histopathology of meloidogyne arenaria race 1 on *Arachis* spp. *Plant Pathol.* 57, 974–980. doi: 10.1111/j.1365-3059.2008.01861.x

Punja, Z. K. (1985). The biology, ecology, and control of sclerotium rolfsii. *Annu. Rev. Phytopathol.* 23, 97–127. doi: 10.1146/annurev.py.23.090185.000525

Qi, F., Sun, Z., Liu, H., Zheng, Z., Qin, L., Shi, L., et al. (2022). QTL identification, fine mapping, and marker development for breeding peanut (*Arachis hypogaea l.*) resistant to bacterial wilt. *Theor. Appl. Genet.* 135, 1319–1330. doi: 10.1007/s00122-022-04033-y

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., et al. (2012). An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea l.*) constructed from two RIL populations. *Theor. Appl. Genet.* 124, 653–664. doi: 10.1007/s00122-011-1737-y

Rao, S. C., Bhatnagar-Mathur, P., Kumar, P. L., Reddy, A. S., and Sharma, K. K. (2013). Pathogen-derived resistance using a viral nucleocapsid gene confers only partial non-durable protection in peanut against peanut bud necrosis virus. *Arch. Virol.* 158, 133–143. doi: 10.1007/s00705-012-1483-8

Rathod, V., Hamid, R., Tomar, R. S., Patel, R., Padhiyar, S., Kheni, J., et al. (2020). Comparative RNA-seq profiling of a resistant and susceptible peanut (*Arachis* hypogaea) genotypes in response to leaf rust infection caused by puccinia arachidis. *3 Biotech.* 10, 284. doi: 10.1007/s13205-020-02270-w

Ratnaparkhe, M. B., Wang, X., Li, J., Compton, R. O., Rainville, L. K., Lemke, C., et al. (2011). Comparative analysis of peanut NBS-LRR gene clusters suggests evolutionary innovation among duplicated domains and erosion of gene microsynteny. *New Phytol.* 192, 164–178. doi: 10.1111/j.1469-8137.2011.03800.x

Riley, D. G., Joseph, S. V., Srinivasan, R., and Diffiffiffie, S. (2011). Thrips vectors of tospoviruses. *J. Integ. Pest. Manage.* 2, 1–10. doi: 10.1603/IPM10020

Saito, K., and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61, 463–489. doi: 10.1146/annurev.arplant.043008.092035

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., et al. (2002). Genome sequence of the plant pathogen ralstonia solanacearum. *Nature* 415, 497–502. doi: 10.1038/415497a

Saleh, N., Horn, N. M., Reddy, D. V. R., and Middleton, K. J. (1989). Peanut stripe virus in Indonesia. *Eur. J. Plant Pathol.* 99, 123–127. doi: 10.1007/bf01997480

Sharma, H. C. (2005). *Heliothis/Helicoverpa management: emerging trends and strategies for future research* (New Delhi, India: Oxford and IBH Publishing Co.Pvt. Ltd), 469 pp.

Sharma, S., Choudhary, B., Yadav, S., Mishra, A., Mishra, V. K., Chand, R., et al. (2021). Metabolite profiling identified pipecolic acid as an important component of peanut seed resistance against aspergillus flavus infection. *J. Hazard Mater* 404, 124155. doi: 10.1016/j.jhazmat.2020.124155

Sharma, H. C., Pampapathy, G., Dwivedi, S. L., and Reddy, L. J. (2003). Mechanisms and diversity of resistance to insect pests in wild relatives of groundnut. *J. Econ Entomol* 96, 1886–1897. doi: 10.1093/jee/96.6.1886

Shasidhar, Y., Variath, M. T., Vishwakarma, M. K., Manohar, S. S., Gangurde, S. S., Sriswathi, M., et al. (2020). Improvement of three popular Indian groundnut varieties for foliar disease resistance and high oleic acid using SSR markers and SNP array in marker-assisted backcrossing. *Crop J.* 8, 1–15.

Shinozaki, K., and Sakakibara, H. (2009). Omics and bioinformatics: an essential toolbox for systems analyses of plant functions beyond 2010. *Plant Cell Physiol.* 50, 1177–1180. doi: 10.1093/pcp/pcp085

Shirasawa, K., Bhat, R. S., Khedikar, Y. P., Sujay, V., Kolekar, R. M., Yeri, S. B., et al. (2018). Sequencing analysis of genetic loci for resistance for late leaf spot and rust in peanut (*Arachis hypogaea l.*). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01727

Simpson, C. E., and Starr, J. L. (2001). Registration of 'COAN' peanut. *Crop Sci.* 41, 918–918. doi: 10.2135/cropsci2001.413918x

Simpson, C. E., Starr, J. L., Church, G. T., Burow, M. D., and Paterson, A. H. (2003). Registration of 'NemaTAM' peanut. *Crop Sci.* 43, 1561–1561. doi: 10.2135/cropsci2003.1561

Singh, M. K., Chandel, V., Hallan, V., Ram, R., and Zaidi, A. A. (2009). Occurrence of peanut stripe virus on patchouli and raising of virus-free patchouli plants by meristem tip culture. *J. Plant Dis. Prot* 116, 2–6. doi: 10.1007/BF03356278

Song, H., Wang, P., Li, C., Han, S., Zhao, C., Xia, H., et al. (2017). Comparative analysis of NBS-LRR genes and their response to aspergillus flavus in *Arachis*. *PloS One* 12, e0171181. doi: 10.1371/journal.pone.0171181

Srinivasa, Rao., M., Manimanjari, D., Vanaja, M., Rama, Rao., C. A., Srinivas, K., Rao, V. U., et al. (2012). Impact of elevated CO_2 on tobacco caterpillar, spodoptera litura on peanut, *Arachis hypogea*. *J. Insect Sci.* 12, 103. doi: 10.1673/031.012.10301

Stalker, H. T., Beute, M. K., Shew, B. B., and Barker, K. R. (2002). Registration of two root-knot nematode-resistant peanut germplasm lines. *Crop Sci.* 42, 312–313. doi: 10.2135/cropsci2002.312a

Stalker, H. T., and Campbell, W. V. (1983). Resistance of wild species of peanut to an insect complex. *Peanut Sci.* 10, 30–33. doi: 10.3146/i0095-3679-10-1-9

Subrahmanyam, P., McDonald., D., Reddy., U., Nigam., S. N., and Smith., D. H. (1993). Origin and utilization of rust resistance in groundnut. *Curr. Plant Sci. Biotechnol. Agr.* 18, 147–158. doi: 10.1007/978-94-011-2004-3_12

Sujay, V., Gowda, M. V., Pandey, M. K., Bhat, R. S., Khedikar, Y. P., Nadaf, H. L., et al. (2012). Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea l.*). *Mol. Breed* 30 (2), 773–788. doi: 10.1007/s11032-011-9661-z

Tay, W. T., Soria, M. F., Walsh, T., Thomazoni, D., Silvie, P., Behere, G. T., et al. (2013). A brave new world for an old world pest: helicoverpa armigera (Lepidoptera: noctuidae) in Brazil. *PloS One* 8, e80134. doi: 10.1371/journal.pone.0080134

Termorshuizen, A. J. (2007). Fungal and fungus-like pathogens of potato. *Potato Biol. Biotechnol.*, 643–665. doi: 10.1016/b978-044451018-1/50071-3

Thiessen, L. D., and Woodward, J. E. (2012). Diseases of peanut caused by soilborne pathogens in the southwestern united states. *ISRN Agron.*, 517905. doi: 10.5402/2012/517905

Tirumalaraju, S. V., Jain, M., and Gallo, M. (2011). Differential gene expression in roots of nematode-resistant and -susceptible peanut (*Arachis hypogaea*) cultivars in response to early stages of peanut root-knot nematode (Meloidogyne arenaria) parasitization. *J. Plant Physiol.* 168, 481–492. doi: 10.1016/j.jplph.2010.08.006

Torres, A. M., Barros, G. G., Palacios, S. A., Chulze, S. N., and Battilani, P. (2014). Review on pre and post-harvest management of peanuts to minimize aflatoxin contamination. *Food Res. Int.* 62, 11–19. doi: 10.1016/j.foodres.2014.02.023

Trudgill, D. L., and Blok, V. C. (2001). Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. *Annu. Rev. Phytopathol.* 39, 53–77. doi: 10.1146/annurev.phyto.39.1.53

Tseng, Y. C., Tillman, B. L., Peng, Z., and Wang, J. (2016). Identifcation of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EP '113'. *BMC Genet.* 17, 128. doi: 10.1186/s12863-016-0435-9

Tshilenge-Lukanda, L., Nkongolo, K. K. C., Kalonji-Mbuyi, A., and Kizungu, R. V. (2012). Epidemiology of the groundnut (*Arachis hypogaea l.*) leaf spot disease: genetic analysis and developmental cycles. *Am. J. Plant Sci.* 3, 582–588. doi: 10.4236/ajps.2012.35070

Urcuqui-Inchima, S., Haenni, A. L., and Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Res.* 74, 157–175. doi: 10.1016/S0168-1702(01)00220-9

Vadivel, A. K. (2015). Gel-based proteomics in plants: time to move on from the tradition. front. *Plant Sci.* 6. doi: 10.3389/fpls.2015.00369

Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* 132 (3), 797–816. doi: 10.1007/s00122-018-3252-x

Varshney, R. K., Pandey, M. K., Janila, P., Nigam, S. N., Sudini, H., Gowda, M. V., et al. (2014). Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea l.*). *Theor. Appl. Genet.* 127, 1771–1781. doi: 10.1007/2Fs00122-014-2338-3

Wang, Z., Huai, D., Zhang, Z., Cheng, K., Kang, Y., Wan, L., et al. (2018). Development of a high-density genetic map based on specific length amplified fragment sequencing and its application in quantitative trait loci analysis for yield-related traits in cultivated peanut. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00827

Wang, H., Lei, Y., Wan, L., Yan, L., Lv, J., Dai, X., et al. (2016). Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by aspergillus flavus. *BMC Plant Biol.* 16, 54. doi: 10.1186/s12870-016-0738-z

Wang, H., Penmetsa, R. V., Yuan, M., Gong, L., Zhao, Y., Guo, B., et al. (2012). Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea l.*). *BMC Plant Biol.* 12, 10. doi: 10.1186/1471-2229-12-10

Wang, T., Zhang, E., Chen, X., Li, L., and Liang, X. (2010). Identification of seed proteins associated with resistance to pre-harvested aflatoxin contamination in peanut (*Arachis hypogaea l.*). *BMC Plant Biol.* 10, 267. doi: 10.1186/1471-2229-10-267

War, A. R., Paulraj, M. G., War, M. Y., and Ignacimuthu, S. (2011). Jasmonic acid-mediated-induced resistance in groundnut (*Arachis hypogaea l.*) against helicoverpa armigera (hubner) (lepidoptera: noctuidae). *J. Plant Growth Regul.* 30, 512–523. doi: 10.1007/s00344-011-9213-0

War, A. R., Sharma, S. P., and Sharma, H. C. (2016). Differential induction of flavonoids in groundnut in response to helicoverpa armigera and aphis craccivora infestation. *Int. J. Insect Sci.* 8, 55–64. doi: 10.4137/IJIS.S39619

Wei, T., Zhang, C., Hong, J., Xiong, R., Kasschau, K. D., Zhou, X., et al. (2010). Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PloS Pathog.* 6, e1000962. doi: 10.1371/journal.ppat.1000962

Wightman, J. A., and Rao, G. V. (1994). Groundnut pests. *World Crop*, 395–479. doi: 10.1007/978-94-011-0733-4_11

Xu, Z. Y. (2002). Progress on research to control peanut virus disease. *Plant Protect. Technol. Ext.* 22, 37–39.

Xu, Z. Y., Chen, K. R., Zhang, Z. Y., and Chen, J. X. (1991). Seed transmission of peanut stripe virus in peanut. *Plant Dis.* 75, 723–726. doi: 10.1094/PD-75-0723

Xu, Z. Y., Yu, Z., Liu, J. L., and Barnett, O. W. (1983). A virus causing peanut mild mottle in hubei province, China. *Plant Dis.* 67, 1029–1032. doi: 10.1094/PD-67-1029

Xu, P., Li, H., Wang, X., Zhao, G., Lu, X., Dai, S., et al. (2022). Integrated analysis of the lncRNA/circRNA-miRNA-mRNA expression profiles reveals novel insights into potential mechanisms in response to root-knot nematodes in peanut. *BMC Genomics* 23, 239. doi: 10.1186/s12864-022-08470-3

Xu, M., Xie, H., Wu, J., Xie, L., Yang, J., and Chi, Y. (2017). Translation initiation factor eIF4E and eIFiso4E are both required for peanut stripe virus infection in peanut (*Arachis hypogaea l.*). *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00338

Yan, L., Jin, H., Raza, A., Huang, Y., Gu, D., and Zou, X. (2022). WRKY genes provide novel insights into their role against ralstonia solanacearum infection in cultivated peanut (*Arachis hypogaea l.*). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.986673

Yan, L., Wang, Z., Song, W., Fan, P., Kang, Y., Lei, Y., et al. (2021). Genome sequencing and comparative genomic analysis of highly and weakly aggressive strains of sclerotium rolfsii, the causal agent of peanut stem rot. *BMC Genomics* 22, 276. doi: 10.1186/s12864-021-07534-0

Yu, B., Huai, D., Huang, L., Kang, Y., Ren, X., Chen, Y., et al. (2019). Identification of genomic regions and diagnostic markers for resistance to aflatoxin contamination in peanut (*Arachis hypogaea l.*). *BMC Genet.* 20, 32. doi: 10.1186/s12863-019-0734-z

Yuksel, B., Estill, J., Schulze, S., and Paterson, A. (2005). Organization and evolution of resistance gene analogs in peanut. *Mol. Gen. Genomics* 274, 248–263. doi: 10.1007/s00438-005-0022-7

Zhang, C., Chen, H., Cai, T., Deng, Y., Zhuang, R., Zhang, N., et al. (2017). Overexpression of a novel peanut NBS-LRR gene AhRRS5 enhances disease resistance to ralstonia solanacearum in tobacco. *Plant Biotechnol. J.* 15, 39–55. doi: 10.1111/pbi.12589

Zhang, C., Chen, H., Zhuang, R. R., Chen, Y. T., Deng, Y., Cai, T. C., et al. (2019). Overexpression of the peanut CLAVATA1-like leucine-rich repeat receptor-like kinase AhRLK1 confers increased resistance to bacterial wilt in tobacco. *J. Exp. Bot.* 70, 5407–5421. doi: 10.1093/jxb/erz274

Zhang, H., Chu, Y., Dang, P., Tang, Y., Jiang, T., Clevenger, J. P., et al. (2020). Identifcation of QTLs for resistance to leaf spots in cultivated peanut (*Arachis hypogaea l.*) through GWAS analysis. *Theor. Appl. Genet.* 133, 2051–2061. doi: 10.1007/s00122-020-03576-2

Zhang, W., Xian, J., Sun, C., Wang, C., Shi, L., and Yu, W. (2021). Preliminary study of genome editing of peanut FAD2 genes by CRISPR/Cas9. *Acta Agronomica Sin.* 47, 1481–1490.

Zhao, X., Li, C., Yan, C., Wang, J., and Shan, S. (2019). Transcriptome and proteome analyses of resistant preharvest peanut seed coat in response to aspergillus flavus infection. *Electron J. Biotechn* 39, 82–90. doi: 10.1016/j.ejbt.2019.03.003

Zhao, C., Li, T., Zhao, Y., Zhang, B., Li, A., Zhao, S., et al. (2020). Integrated small RNA and mRNA expression profiles reveal miRNAs and their target genes in response to aspergillus flavus growth in peanut seeds. *BMC Plant Biol.* 20, 215. doi: 10.1186/s12870-020-02426-z

Zhao, C., Qiu, J., Agarwal, G., Wang, J., Ren, X., Xia, H., et al. (2017). Genome-wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaensis*, and their application in cultivated peanut (*A. hypogaea*). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01209

Zhao, Z., Tseng, Y. C., Peng, Z., Lopez, Y., Chen, C. Y., Tillman, B. L., et al. (2018). Refining a major QTL controlling spotted wilt disease resistance in cultivated peanut (*Arachis hypogaea l.*) and evaluating its contribution to the resistance variations in peanut germplasm. *BMC Genet.* 19, 17. doi: 10.1186/s12863-018-0601-3

Zhao, Y., Zhang, C., Chen, H., Yuan, M., Nipper, R., Prakash, C. S., et al. (2016). QTL mapping for bacterial wilt resistance in peanut (*Arachis hypogaea l.*). *Mol. Breed* 36, 13. doi: 10.1007/s11032-015-0432-0

Zhou, X., Xia, Y., Liao, J., Liu, K., Li, Q., Dong, Y., et al. (2016). Quantitative trait locus analysis of late leaf spot resistance and plant-Type-Related traits in cultivated peanut (*Arachis hypogaea l.*) under multi-environments. *PloS One* 11, e0166873. doi: 10.1371/journal.pone.0166873

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51 (5), 865–876. doi: 10.1038/s41588-019-0402-2

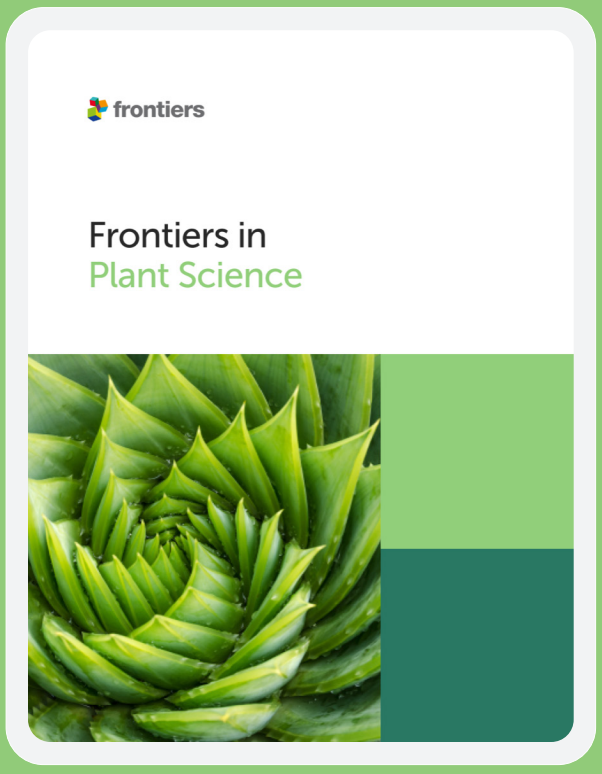
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

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