Genomic Insights into Peanut Disease Resistance, Seed Traits, and Drought Water-Use Strategies

by

Jie Zhang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 8, 2026

Key words: Genome wide association study, MAPK, Drought Resistance, leaf spot disease, Water use strategy

Approved by

Charles Y. Chen, Chair, Professor of Crop, Soil and Environmental Sciences
Alvaro Sanz-Saez, Associate Professor of Crop, Soil and Environmental Sciences
Phat M. Dang, Research Biochemist of USDA-ARS National Peanut Research Laboratory, Dawson, GA
Alex Harkess, Faculty Investigator of HudsonAlpha Institute for Biotechnology, Huntsville, AL
Sushan Ru, Assistant Professor of Horticulture

Abstract

Peanut (*Arachis hypogaea L.*) is an important oilseed and food legume, but its improvement is constrained by narrow cultivated genetic diversity, allotetraploid genome complexity, foliar disease pressure, seed dormancy-related quality issues, and increasing drought risk. This dissertation integrates peanut genomics, association mapping, physiological phenotyping, transcriptomics, and gene co-expression network analysis to identify genetic and regulatory mechanisms underlying stress adaptation and agronomic trait variation. A genome-wide analysis of the mitogen-activated protein kinase (MAPK) gene family identified 30 MAPK genes in cultivated peanut and 16 and 15 MAPK genes in its diploid relatives A. duranensis and A. ipaensis, respectively. Comparative phylogenetic, structural, promoter, and synteny analyses showed conserved MAPK organization across peanut genomes. RNA-seq analysis under drought stress identified Ah_At_MAPK4 and Ah_Bt_MAPK4 as strongly drought-induced genes, with greater induction in drought-tolerant genotypes than in susceptible genotypes. Co-expression network analysis placed both genes in a drought-associated module enriched for calcium signaling and secondary metabolic pathways, supporting their roles as candidate regulators of drought-responsive signaling. To dissect genomic variation for disease resistance and seed traits, 87 accessions from the U.S. peanut mini-core collection were resequenced, generating 217 Gb of sequence data and 87,726 high-quality SNPs. Population structure analysis resolved three major genetic groups, and genome-wide association analysis identified significant loci for early leaf spot, late leaf spot, seed germination, and dormancy. Candidate-gene analysis highlighted Ah11g381400, homologous to ATE1, for early leaf spot resistance; Ah16g445600, homologous to ERF34, for late leaf spot resistance; and Ah19g214100, homologous to ICE1, as a candidate affecting germination and dormancy. Field-based physiological and transcriptomic analyses further examined drought adaptation in water-saver, water-spender, and drought-sensitive peanut genotypes. Progressive drought reduced photosynthesis, stomatal conductance, and electron transport rate while increasing intrinsic water-use efficiency. RNA-seq and WGCNA revealed a dominant

drought-stage response, a secondary transcriptomic axis associated with water-use strategy, and partial recovery after rewatering. Together, these findings provide candidate genes, molecular markers, and regulatory frameworks for functional validation, marker development, and genomics-assisted breeding of peanut cultivars with improved drought resilience, foliar disease resistance, and seed quality under water-limited and disease-prone production environments.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this thesis / dissertation, the following Artificial Intelligence (AI) tools were used: ChatGPT. These tools were used primarily to polish the language of the manuscript, including improving grammar, clarity, sentence structure, and overall readability. No AI tools were used to generate research data and produce original scientific content. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain scholarly integrity of this research.

Digital Accessibility Disclosure Statement

In the preparation of this thesis / dissertation, the following digital accessibility tools were used to ensure this document complies with federal requirements: Adobe Acrobat Pro. The author acknowledges full responsibility for the intellectual content of this work and has made a good faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, all content has been reviewed and revised to meet these requirements prior to final publication.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Charles Chen, for his invaluable guidance, encouragement, and support throughout my Ph.D. journey. His expertise, patience, and dedication to research have profoundly influenced my academic and professional development. I am sincerely grateful for the opportunities he has provided and for his continuous mentorship during my graduate studies.

I would also like to extend my sincere appreciation to my dissertation committee members, Dr. Phat Dang, Dr. Alvaro Sanz-Saez, and Dr. Alex Harkess, as well as my University Reader, Dr. Sushan Ru, for their valuable feedback, insightful suggestions, and constructive comments that significantly improved this dissertation. Their expertise and guidance have been greatly appreciated.

My sincere thanks also go to Sonettra for her assistance with administrative matters throughout my graduate studies. Her efficiency and willingness to help made many administrative processes much smoother.

I am grateful to my lab colleagues and friends, Bo Cheng, Sajid, and Fritzner, for their collaboration, support, and friendship. Working alongside them has made my research experience both enjoyable and rewarding. I also appreciate all current and former members of our research group for fostering a supportive and collaborative environment.

I would also like to thank my colleagues, and friends. Their encouragement, friendship, and the memorable experiences we shared have made my years at Auburn both meaningful and unforgettable.

Above all, I owe my deepest gratitude to my parents. Their unconditional love, unwavering support, and constant belief in me have been my greatest source of strength throughout this journey. No matter what pressures or challenges I faced, you have always been my strongest pillars, giving me the strength and courage to pursue my dreams.

Life goes on, and I will keep moving forward and embracing every moment.

Table of Contents

List of Tables

List of Figures

Chapter 1. Advances in Peanut (*Arachis hypogaea* L.) Genomics, Breeding, and Trait Dissection

1.1 Introduction and significance of peanut

The genus *Arachis*, a member of the legume family Fabaceae, is native to South America, where it occupies ecological zones ranging from tropical lowlands to subtropical highlands (Gregory et al., 1980). The genus is taxonomically diverse, with species assigned to sections on the basis of morphology, geographic distribution, genome relationships, and crossability (Krapovickas & Gregory, 2023). Within this genus, cultivated peanut (*Arachis hypogaea* L.) is the most economically important species and one of the major cultivated legumes. Archaeological and historical evidence indicates that peanut was cultivated in South America thousands of years ago, with archaeological records from Peru documenting early use by approximately 1200-1500 BC (Hammons et al., 2016). The species was formally named by Carl Linnaeus in 1753 (Bertioli et al., 2011).

Peanut is now grown widely in tropical, subtropical, and warm-temperate production systems and is valued as an oilseed, food legume, protein source, and cash crop. China and India remain major centers of peanut production, while large production areas also occur in Africa, the Americas, and other regions where the crop contributes to food security, farmer income, and crop diversification. Peanut kernels are especially important because they contain high concentrations of oil and protein, along with carbohydrates, minerals, vitamins, antioxidants, and other bioactive compounds. This nutritional and economic importance explains why peanut improvement programs have increasingly expanded beyond yield alone to include market quality, oil chemistry, stress tolerance, and disease resistance (Janila et al., 2013).

Cultivated peanut is morphologically diverse despite its narrow genetic background. It is generally classified into two subspecies, *hypogaea* and *fastigiata*, based largely on growth habit and the distribution of reproductive branches. These subspecies are further divided into botanical varieties that differ in inflorescence, pod, seed, and growth characteristics: *subsp. hypogaea* includes *var. hypogaea* and *var. hirsuta*, whereas *subsp. fastigiata* includes *var. fastigiata*, *var. vulgaris*, *var. peruviana*, and *var. aequatoriana* (Isleib et al., 2001). This botanical diversity is important because

market classes, seed traits, and breeding objectives often differ among the subspecies and botanical varieties.

As an allotetraploid crop with a large, repetitive genome and complex trait architecture, peanut presents substantial challenges for genetic improvement. These features explain why progress in peanut breeding has depended on the gradual integration of germplasm resources, conventional breeding, molecular marker development, high-density genetic maps, quantitative trait locus (QTL) discovery, genome-wide association study (GWAS), and more recent genomics-assisted breeding approaches.

1.2 Peanut genomics and genome evolution

Cultivated peanut (*Arachis hypogaea* L.) is an annual, predominantly self-pollinated allotetraploid with 2n = 4x = 40 chromosomes and an AABB genome constitution. It arose through hybridization between two diploid ancestors followed by genome duplication. The B subgenome is generally attributed to *Arachis ipaensis*, whereas *Arachis duranensis* is widely treated as the closest known representative of the A subgenome, although the precise identity of the direct A-genome donor has been discussed in comparative genomic studies (Bertioli et al., 2016; X. Chen et al., 2019). This allopolyploid origin is now reflected in the genomic architecture of cultivated peanut and remains central to studies of genome evolution, marker development, and trait dissection.

The genome of *A. hypogaea* is large, approximately 2.5 Gb, and contains a high proportion of repetitive DNA (Bertioli et al., 2016; Zhuang et al., 2019). Genome size in cultivated peanut is approximately comparable to the combined sizes of the A- and B-genome diploid progenitors, suggesting that large-scale genome expansion after polyploidization was limited (Samoluk et al., 2015). Cytogenetic studies using genomic in situ hybridization (GISH) have further confirmed that the A and B subgenomes remain distinguishable, with no extensive A-B chromosomal mosaicism (Bertioli et al., 2016; Ramos et al., 2006; Seijo et al., 2007). These observations are important for breeding because molecular markers and sequence reads must often distinguish between homologous and homeologous loci in a polyploid background.

The development of sequencing technologies has transformed peanut genomics. Early genetic studies relied on cytogenetics, low-throughput markers, and limited sequence information. Next-generation sequencing and long-read sequencing subsequently enabled the generation of reference genomes for the diploid progenitors *A. duranensis* and *A. ipaensis*, the cultivated tetraploid *A. hypogaea*, and the wild allotetraploid *A. monticola* (Bertioli et al., 2016; Bertioli et al., 2019; X. Chen et al., 2016; Lu et al., 2018; Yin et al., 2018; Zhuang et al., 2019). The cultivated peanut genome sequence revealed tens of thousands of predicted genes in each subgenome, with gene counts approximately 11% higher in the B subgenome (35,110) than in the A subgenome (31,359) (Bertioli et al., 2019).

These genome assemblies shifted peanut research from descriptive genetics toward genome-enabled trait dissection. They support genome-wide marker discovery, comparative genomics, synteny analysis, candidate-gene identification, pan-genomic comparisons, and marker-assisted introgression. However, the benefits of these resources are moderated by two biological challenges repeatedly emphasized in peanut genomics: the narrow genetic diversity of cultivated germplasm and the technical complexity created by polyploidy. Both factors reduce the frequency of informative polymorphisms and complicate marker development, QTL localization, and functional validation.

1.3 History and direction of peanut breeding

Peanut breeders have traditionally focused on yield, adaptation, seed size, oil content, high oleic acid, disease resistance, drought tolerance, and other traits relevant to producer profitability and market quality. In the United States, early breeding emphasized yield and adaptation, whereas recent breeding has increasingly incorporated quality, flavor, disease resistance, drought tolerance, and improved oil chemistry (Janila et al., 2013). Similar shifts have occurred globally as production systems face changing disease pressure, climate variability, and market demand for enhanced nutritional and processing quality.

Conventional peanut breeding has used introduction, mass selection, pedigree selection, bulk and modified bulk methods, single-seed descent, backcrossing, and

mutation breeding to improve agronomic performance and stress resistance. In early breeding, introduction and selection from landraces were essential. As breeding programs matured, artificial hybridization followed by selection in segregating generations became the dominant strategy for cultivar development. Pedigree selection is useful for traits with higher heritability, such as plant architecture, seed coat color, and certain pod or seed characteristics, whereas quantitative traits such as yield, seed composition, and stress tolerance usually require later-generation replicated testing across environments.

The historical development of peanut cultivars illustrates both the success and limitations of conventional breeding. Many U.S. cultivars were derived from planned crosses, with elite-by-elite hybridization and backcrossing used to combine yield, market type, disease resistance, and oil quality. However, yield improvement has become more difficult because modern elite cultivars share a narrow ancestry and breeding programs must simultaneously select for multiple traits, including resistance to root-knot nematode, tomato spotted wilt virus, leaf spots, rust, stem rot, aflatoxin contamination, drought tolerance, and high oleic acid content.

Conventional breeding remains indispensable, but it is less efficient when target traits are quantitatively inherited, strongly influenced by environment, expensive to phenotype, expressed late in development, or controlled by recessive alleles. Marker-assisted selection (MAS), QTL mapping, GWAS, and genomic selection do not replace field-based breeding; rather, they increase the precision and speed of selection when linked markers, diagnostic markers, high-quality phenotypes, and validated genomic regions are available. The central challenge for modern peanut improvement is therefore to integrate conventional selection with molecular and genomic tools so that genetic gain can be increased without sacrificing local adaptation and farmer- or market-preferred traits (Desmae et al., 2018).

1.4 Molecular marker development in peanut

DNA sequence variation is the basis of genetic diversity within and among species. Molecular markers directly reflect polymorphism at the DNA level and allow breeders to track genomic regions associated with traits of interest. Marker-assisted

selection is an indirect selection method in which individuals are selected based on markers linked to a target phenotype, such as productivity, disease resistance, abiotic stress tolerance, or seed quality, rather than by phenotype alone (Ribaut & Hoisington, 1998). In peanut, the first major applications of DNA markers included restriction fragment length polymorphism (RFLP) analysis to detect variation among cultivated peanut and wild Arachis species and to identify markers associated with root-knot nematode resistance (Kochert et al., 1991).

Early marker systems in peanut included RFLP, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence-characterized amplified regions (SCAR), sequence-related amplified polymorphisms (SRAP), simple sequence repeats (SSR), single nucleotide polymorphisms (SNPs), and Kompetitive Allele-Specific PCR (KASP) markers (Hong et al., 2015; Liang et al., 2009; Paran & Michelmore, 1993; Robarts & Wolfe, 2014; Semagn et al., 2014; Vos et al., 1995; Williams et al., 1990). RFLP, RAPD, and AFLP markers were valuable for early diversity analysis and map construction, especially in wild species and interspecific populations, but their use declined as more reproducible, transferable, and high-throughput marker systems became available.

SSR markers became widely used because they are co-dominant, PCR-based, relatively transferable, and often highly informative. Thousands of SSR markers have been developed in peanut and used for genetic diversity analysis, linkage mapping, QTL discovery, and association studies (Gupta et al., 2000; Pandey et al., 2012; Varshney et al., 2013). SSR-based maps were especially important before high-density SNP platforms became common, and integrated SSR consensus maps helped anchor linkage groups across populations and studies (Shirasawa et al., 2013; Tang et al., 2007; Varshney et al., 2009; H. Wang et al., 2012). Nevertheless, SSR discovery and genotyping are relatively labor-intensive, and polymorphism within cultivated peanut remains limited because of the narrow genetic base and repetitive nature of the genome.

SNP markers now dominate peanut genomics because they are abundant, scalable, compatible with high-throughput genotyping platforms, and suitable for dense

linkage maps, GWAS, genomic selection, and diagnostic marker development. SNP discovery has been accelerated by whole-genome resequencing, reduced-representation sequencing, genotyping-by-sequencing (GBS), SLAF-seq, RAD-seq, and SNP arrays. The 58K Axiom_Arachis SNP array provided a major platform for high-throughput peanut genotyping; evaluation across diverse accessions identified 44,424 polymorphic SNPs, supporting genetic mapping, population structure analysis, GWAS, QTL validation, and molecular breeding (Pandey et al., 2017).

Recent marker development has expanded beyond SSRs and SNPs to include insertion-deletion markers, DArT markers, transposable-element markers, epigenetic signatures, and functional markers derived from causal or candidate genes. KASP assays are especially useful for breeding because they convert sequence polymorphisms into robust, low-cost, allele-specific tests that can be deployed in large breeding populations (Pandey et al., 2024). The literature increasingly distinguishes random markers, which are linked to a trait but may not be causal, from functional markers, which target allelic variation in the gene directly responsible for the phenotype. For peanut, functional markers remain limited but are highly valuable, as illustrated by markers for FAD2 alleles affecting high oleic acid and markers linked to root-knot nematode resistance (Chu et al., 2007; Nawade et al., 2019).

1.5 Genetic linkage map construction and mapping populations

Molecular markers are further used to construct genetic linkage maps, which order markers according to recombination frequency in segregating populations. Linkage maps provide the framework for identifying QTLs, estimating genetic distances, examining recombination, comparing genome structure, and selecting markers for breeding. The earliest peanut linkage maps were built mainly from interspecific populations because wild Arachis species contain higher polymorphism than cultivated peanut. For example, an early RFLP map constructed from an F2 population derived from *A. stenosperma* x *A. cardenasii* assigned 117 segregating loci into 11 linkage groups (Halward et al., 1993).

As marker density increased, SSR- and SNP-based linkage maps improved resolution and genome coverage. An integrated consensus map of cultivated peanut

and wild relatives was anchored to 20 linkage groups, covered approximately 2651 cM, and contained 3693 marker loci (Shirasawa et al., 2013). More recent high-density maps built with SNP markers and sequencing-based genotyping have supported finer QTL localization and candidate-gene discovery (Khan et al., 2020). These developments are important because the cultivated peanut genome contains 20 chromosome pairs, and incomplete linkage maps reduce the power to identify marker-trait relationships across the A and B subgenomes.

Mapping population design determines the resolution and relevance of QTL discovery. F2, backcross, recombinant inbred line (RIL), and near-isogenic line populations are useful for linkage mapping and fine mapping, whereas nested association mapping (NAM), multi-parent advanced generation inter-cross (MAGIC), introgression lines, chromosome segment substitution lines, and diversity panels increase allele sampling and mapping power (Kassie et al., 2023; Gangurde et al., 2020).

Genotyping-by-sequencing and whole-genome resequencing now allow dense SNP discovery in mapping populations, but they introduce challenges such as missing data, genotyping error, and difficulty distinguishing homeologous variants. Bin-map approaches, sliding-window imputation, and high-throughput mapping software help reduce noise by collapsing adjacent markers with consistent genotypes into recombination bins. These methods increase mapping resolution and can improve QTL detection, but they still require accurate phenotyping, careful quality control, and validation in independent genetic backgrounds before marker deployment in breeding.

1.6 QTL mapping of important traits in peanut

QTL analysis links genotypic data from molecular markers with phenotypic data collected in a segregating population. It is particularly useful for traits showing continuous variation, such as yield, pod and seed morphology, oil content, disease resistance, drought tolerance, and other stress-related traits. QTL mapping estimates the number, position, effect size, and phenotypic variance explained by genomic regions associated with a trait. In peanut, QTL analysis has become a central approach

for dissecting complex traits and identifying genomic regions that can be validated for MAS and candidate-gene analysis.

1.6.1 Biotic stress resistance

Linkage mapping has been widely applied to resistance against foliar diseases, soilborne diseases, viruses, and nematodes. In one study, 28 QTLs associated with late leaf spot (LLS) resistance were detected, including a major QTLLLS01 region explaining 10.27-62.34% of phenotypic variance, and 15 QTLs associated with rust resistance were also identified (Sujay et al., 2012). Another study using an F5 population derived from Tifrunner x GT-C20 identified 23 QTLs, including one QTL for thrips resistance, 13 QTLs for leaf spot resistance, and nine QTLs for tomato spotted wilt virus resistance (H. Wang et al., 2013).

Major biotic stresses in peanut include foliar fungal diseases (early and late leaf spot, rust), tomato spotted wilt virus (TSWV), white mold/stem rot (caused by *Sclerotium rolfsii*), root-knot nematode, and *Aspergillus* contamination. For foliar diseases, 28 QTLs for late leaf spot (LLS) resistance have been detected, including a major *QTLLLS01* region explaining 10.27 - 62.34% of phenotypic variance, together with 15 QTLs for rust resistance (Sujay et al., 2012). An F5 population from *Tifrunner* × GT-C20 yielded 23 QTLs for leaf spot, TSWV, and thrips resistance (H. Wang et al., 2013). Subsequently, whole-genome resequencing was deployed to refine these regions, identifying 35 main-effect QTLs for early leaf spot (ELS), late leaf spot (LLS), and TSWV, with PVE ranging from 6.32% to 47.63%. Within these fine-mapped intervals, key candidate disease-resistance genes (such as R-genes and transcription factors) were discovered (Agarwal et al., 2018). A GWAS using the U.S. peanut mini-core and the Axiom_*Arachis* SNP array identified 46 QTLs for early and late leaf spot resistance, with most major loci on the B sub-genome and two regions on chromosome B09 conferring resistance to both diseases (H. Zhang et al., 2020); several of these and other foliar-disease loci have since been converted into KASP diagnostic markers for breeding (Pandey et al., 2024).

1.6.2 Yield and pod- or seed-related traits

Yield remains the primary target of peanut improvement, but it is a complex trait affected by pod number, seed size, plant architecture, harvest index, adaptation, disease pressure, and genotype-by-environment interaction. Consequently, many studies have focused on yield components such as hundred-pod weight, seed weight, pod length, pod width, seed length, seed width, pod constriction, and pod thickness. For example, QTL analysis of pod and seed dimensions identified 39 QTLs with phenotypic variance explained ranging from 1.25% to 26.11% (W. Chen et al., 2016). The use of *in silico* analysis contributed to the construction of a high-density linkage map which identified 23 QTLs for 15 agronomical traits, including 3 (*qPL05.1, qPL09.2, and qPL06.2*) for pod length, 2 (*qPW07.1 and qPW08.2*) for pod wideness, 1 (*qPT07.1*) for pod thickness, 2 (*qCP09.2 and qCP09.1*) for pod constriction, and one (*qWS08.2*) for seed weight (Shirasawa et al., 2012a).

From a breeding perspective, yield-component QTLs must be interpreted cautiously. A marker linked to seed size or pod shape may be valuable for a specific market class but less useful if it reduces total pod number, adaptation, or disease resistance. Multi-environment validation is therefore necessary before yield-related markers are used for selection. The most promising markers are those that are stable across environments, explain meaningful variance, and remain effective in elite breeding backgrounds.

1.6.3 Drought tolerance and abiotic stress traits

Drought stress is a major environmental constraint for peanut and is expected to become more important under increasing climate variability. Early QTL work used an extensive genetic map from the TAG 24 x ICGV 86031 population and identified 105 main-effect QTLs for drought component traits through composite interval mapping (Ravi et al., 2011). More recent studies identified main-effect QTLs for drought-related traits, including QTLs for pod yield per plant and specific dry weight, with phenotypic variance explained values reaching approximately 10.5% and 18.4%, respectively (Ghosh et al., 2022).

Drought-related QTLs are often small-effect, environment-specific, and influenced by plant developmental stage. Therefore, genomic selection, marker-assisted recurrent selection, and multi-environment genomic prediction may be more appropriate for drought tolerance than single-marker MAS alone. Modern peanut improvement will likely require integrating drought phenotyping, root architecture traits, canopy temperature, transpiration efficiency, metabolomics, transcriptomics, and genomic prediction to identify genotypes that maintain pod yield under water-limited conditions.

1.6.4 Seed quality, oil chemistry, and nutritional traits

Seed quality traits are increasingly important as peanut markets place greater emphasis on flavor, processing quality, nutritional value, oil stability, and human health. Oil content, protein content, fatty acid composition, and the oleic/linoleic acid ratio are central targets. QTLs related to oil content and fatty acid composition have been reported in multiple populations (Huang et al., 2015; N. Liu et al., 2019; Pandey, Wang, et al., 2014; Shasidhar et al., 2017; M. L. Wang et al., 2015). A major oil-content QTL, *qOCA08.1*, explained 10.14-27.19% of phenotypic variance (N. Liu et al., 2019). Whole-genome resequencing of a RIL population identified a major QTL, qA05.1, affecting oil content, protein content, and fatty acid composition, and associated SNPs were validated for possible marker-assisted breeding of high-oil cultivars (Sun et al., 2022).

High oleic acid breeding is one of the most successful examples of translation from molecular genetics to cultivar improvement in peanut. Mutations in FAD2 genes reduce conversion of oleic acid to linoleic acid, improving oil stability and quality. Because the target alleles can be tracked with functional or tightly linked markers, high-oleic introgression through marker-assisted backcrossing can be faster and more precise than selection by phenotyping alone.

1.7 Genome-wide association studies in peanut

Association mapping was first widely applied in human genetics to identify disease-related variants, and its successful use stimulated applications in plant genetics (Lander & Schork, 1994; Templeton et al., 1992). In plants, association mapping was exemplified by the identification of Dwarf8 polymorphisms associated with flowering

time in maize while accounting for population structure (Andersen et al., 2005; Thornsberry et al., 2001). This approach demonstrated that linkage disequilibrium (LD)-based mapping can identify functional polymorphisms and alleles relevant to quantitative traits.

In recent years, with the development of NGS and bioinformatics, more and more crop species genomes were constructed and the genomic information for various species is also continuously being enriched. GWAS has become a powerful strategy for dissecting complex traits in diverse populations. GWAS exploits historical recombination and LD to associate genome-wide marker variation with phenotypic variation. Compared with biparental QTL mapping, GWAS can survey broader allelic diversity and often achieves higher mapping resolution, although its power depends on marker density, sample size, LD decay, phenotyping accuracy, allele frequency, population structure correction, and statistical thresholds (Mackay & Powell, 2007; Tibbs Cortes et al., 2021).

In peanut, the development of high-density SNP arrays and reduced-representation sequencing methods has made GWAS increasingly feasible. The 58K Axiom_Arachis array, SLAF-seq, RAD-seq, and whole-genome resequencing have generated SNP resources for population structure analysis, LD estimation, trait mapping, and candidate-gene discovery (Pandey et al., 2017; X. Zhang et al., 2017). These resources allow GWAS to complement QTL mapping by interrogating genetically diverse panels, including mini-core collections, reference sets, wild species panels, and breeding germplasm.

Several peanut association studies demonstrate the utility of GWAS. Belamkar et al. (2011) assessed population structure and LD in 96 genotypes from the U.S. peanut mini-core collection, demonstrating the potential for LD mapping in cultivated peanut. Huang et al. (2012) evaluated wild Arachis accessions and identified alleles associated with higher oil content that were absent from cultivated peanut, highlighting the value of wild relatives as sources of novel variation. A GWAS of 195 peanut accessions using 13,435 high-quality SNPs identified 93 non-overlapping peak SNPs associated with

yield-related traits and revealed candidate genes, including arahy.RI9HIF, whose homolog has been implicated in yield improvement in rice (J. Wang et al., 2019).

A particularly influential GWAS used a reference set of 300 peanut accessions from 48 countries and combined high-resolution genotyping with multi-season phenotyping for 50 agronomic, disease-resistance, drought-related, and quality traits. This work identified 524 significant marker-trait associations for 36 traits, with phenotypic variance explained ranging from 5.81% to 90.09% (Pandey, Upadhyaya, et al., 2014). These associations provide starting points for marker validation, candidate-gene discovery, and breeding applications, but they require confirmation in independent populations and target environments before direct deployment.

The most important contribution of GWAS is not simply the number of significant markers detected. Its broader value is that it connects natural diversity, population history, high-density genotyping, and trait variation. In peanut, GWAS can identify alleles that were missed in biparental populations, prioritize candidate genes for functional analysis, guide allele mining in core collections, and provide targets for gene editing or functional marker development. However, GWAS should be interpreted as a discovery tool rather than a complete breeding solution; robust phenotyping, independent validation, and integration with QTL mapping, transcriptomics, and gene function studies remain essential.

1.8 Drought stress physiology and molecular responses in peanut

Drought is among the most widespread and economically damaging abiotic stresses affecting peanut production worldwide. Because peanut is often cultivated in rainfed and semi-arid environments, water deficit can reduce stand establishment, photosynthesis, reproductive development, pod filling, kernel quality, and final yield. With increasing climate variability, drought tolerance has become a priority trait for peanut improvement (Soba et al., 2024).

At the physiological level, peanut plants respond to soil drying through a coordinated set of adaptive processes. One of the earliest responses is abscisic acid (ABA)-mediated stomatal closure, which reduces water loss through transpiration but also restricts CO2 diffusion into leaves and lowers photosynthetic carbon fixation

(Farooq et al., 2009; Soba et al., 2024). Reproductive stages are especially vulnerable because flowering, pegging, and pod filling require sustained assimilate supply and cell turgor. Water stress during these stages can reduce peg penetration, pod set, kernel filling, and yield stability (Farooq et al., 2009; Kottapalli et al., 2009).

At the biochemical and molecular levels, drought tolerance is governed by interacting networks of hormone signaling, osmotic adjustment, antioxidant defense, transcriptional regulation, and metabolic reprogramming. ABA activates downstream signaling pathways that regulate stomatal conductance and stress-responsive gene expression (Bhogireddy et al., 2020; Kottapalli et al., 2009). Compatible solutes such as proline, glycine betaine, soluble sugars, and related metabolites help maintain cellular osmotic balance and protect proteins and membranes under dehydration (Farooq et al., 2009). Transcription factor families such as DREB/CBF, MYB, WRKY, NAC, and bZIP coordinate expression of many downstream drought-responsive genes (Bhogireddy et al., 2020; X. Wang et al., 2021).

Transcriptomic and genomic studies have substantially deepened understanding of drought response in peanut. Early microarray studies identified drought-responsive genes involved in ABA signaling, osmotic adjustment, reactive oxygen species scavenging, and cell-wall remodeling (Kottapalli et al., 2009). RNA-seq studies then provided higher-resolution expression profiles across tissues, developmental stages, and contrasting genotypes, revealing thousands of differentially expressed genes associated with drought adaptation (Bhogireddy et al., 2020; X. Wang et al., 2021). In parallel, QTL mapping and GWAS have targeted traits such as transpiration efficiency, specific leaf area, root architecture, chlorophyll status, and pod yield under stress (Dang et al., 2024; Pokhrel et al., 2024).

Among the genetic resources used for peanut GWAS, the U.S. peanut mini-core collection has emerged as a particularly valuable association panel. Developed by Holbrook and Dong (2005), the U.S. mini-core comprises 112 accessions selected from the larger U.S. core collection of 831 accessions using stratified sampling on morphological, geographic, and origin descriptors. It captures much of the genetic and phenotypic diversity of cultivated peanut while remaining tractable for replicated

phenotyping under managed-stress and field environments (Holbrook & Dong, 2005; Otyama et al., 2020). The mini-core has been widely used for association studies of seed quality, fatty acid composition, disease resistance, and other agronomically important traits (Belamkar et al., 2011; Kottapalli et al., 2007; H. Zhang et al., 2020), and is increasingly applied as a discovery panel for drought-related traits, making it an important resource for dissecting the genetic basis of drought tolerance in cultivated peanut.

Drought response is polygenic and strongly influenced by environmental conditions; therefore, the most effective breeding strategies will likely combine physiological screening, managed-stress field trials, high-throughput phenotyping, genomic prediction, QTL/GWAS validation, and multi-omics analyses. This integrated approach is more likely to produce cultivars that maintain yield stability under water limitation than selection based on a single drought index.

1.9 Conclusion

Cultivated peanut combines high agronomic value with substantial genomic complexity. As an allotetraploid with a large, repetitive genome and a relatively narrow domestication base, it has historically been challenging to dissect genetically. However, the past decade has produced reference genomes for the diploid progenitors and the cultivated allotetraploid, dense SNP genotyping platforms, well-characterized mapping and association panels, and improved phenotyping capacity. Together, these resources have transformed peanut research from descriptive genetics into genome-enabled trait dissection. Conventional breeding remains the foundation of cultivar development, but it is increasingly complemented by molecular markers, QTL mapping, GWAS, genomic selection, and multi-omics analyses, particularly for traits that are quantitatively inherited, environment-dependent, or expensive to phenotype, such as drought tolerance, disease resistance, and seed quality. The U.S. peanut mini-core collection is a particularly useful resource in this context, providing a tractable yet diverse panel for GWAS of complex traits including drought response. Building on these advances, the chapters that follow integrate physiological characterization, transcriptomic profiling, and genome-wide association mapping in the U.S. mini-core to dissect drought response in

cultivated peanut and to identify candidate loci and molecular resources that can support breeding for drought-tolerant cultivars.

References

Agarwal, G., Clevenger, J., Pandey, M. K., Wang, H., Shasidhar, Y., Chu, Y., Fountain, J. C., Choudhary, D., Culbreath, A. K., Liu, X., Huang, G., Wang, X., Deshmukh, R., Holbrook, C. C., Bertioli, D. J., Ozias-Akins, P., Jackson, S. A., Varshney, R. K., & Guo, B. (2018). High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. Plant Biotechnology Journal, 16(11), 1954–1967. https://doi.org/10.1111/pbi.12930

Andersen, J. R., Schrag, T., Melchinger, A. E., Zein, I., & Lübberstedt, T. (2005). Validation of Dwarf8 polymorphisms associated with flowering time in elite European inbred lines of maize (Zea mays L.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 111(2), 206-217. https://doi.org/10.1007/S00122-005-1996-6

Gregory, W. C., Krapovickas, A., & Gregory, M. P. (1980). Structure, variation, evolution, and classification in Arachis. In R. J. Summerfield & A. H. Bunting (Eds.), Advances in Legume Science (pp. 469-481). Royal Botanic Gardens, Kew.

Belamkar, V., Selvaraj, M. G., Ayers, J. L., Payton, P. R., Puppala, N., & Burow, M. D. (2011). A first insight into population structure and linkage disequilibrium in the US peanut minicore collection. Genetica, 139(4), 411-429. https://doi.org/10.1007/S10709-011-9556-2

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., Moretzsohn, M. C., Shirasawa, K., Huang, W., Vidigal, B., Abernathy, B., Chu, Y., Niederhuth, C. E., Umale, P., Araújo, A. C. G., Kozik, A., Kim, K., Burow, M. D., Varshney, R. K., Wang, X., Zhang, X., Barkley, N., Guimarães, P. M., Isobe, S., Guo, B., Liao, B., Stalker, H. T., Schmitz, R. J., Scheffler, B. E., Leal-Bertioli, S. C. M., Xun, X., Jackson, S. A., Michelmore, R., & Ozias-Akins, P. (2016). The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nature Genetics 2016 48:4, 48(4), 438-446. https://doi.org/10.1038/ng.3517

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy,

B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., Dash, S., Baidouri, M., Guo, B., Huang, W., Kim, K., Korani, W., Lanciano, S., Lui, C. G., Mirouze, M., Moretzsohn, M. C., Pham, M., Shin, J. H., Shirasawa, K., Sinharoy, S., Sreedasyam, A., Weeks, N. T., Zhang, X., Zheng, Z., Sun, Z., Froenicke, L., Aiden, E. L., Michelmore, R., Varshney, R. K., Holbrook, C. C., Cannon, E. K. S., Scheffler, B. E., Grimwood, J., Ozias-Akins, P., Cannon, S. B., Jackson, S. A., & Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nature Genetics 2019 51:5, 51(5), 877-884. https://doi.org/10.1038/s41588-019-0405-z

Bertioli, D. J., Seijo, G., Freitas, F. O., Valls, J. F. M., Leal-Bertioli, S. C. M., & Moretzsohn, M. C. (2011). An overview of peanut and its wild relatives. Plant Genetic Resources: Characterisation and Utilisation, 9(1), 134-149. https://doi.org/10.1017/S1479262110000444

Bhogireddy, S., Xavier, A., Garg, V., Layland, N., Arias, R., Payton, P., Nayak, S. N., Pandey, M. K., Puppala, N., & Varshney, R. K. (2020). Genome-wide transcriptome and physiological analyses provide new insights into peanut drought response mechanisms. Scientific Reports 2020 10:1, 10(1), 4071-. https://doi.org/10.1038/s41598-020-60187-z

Chen, W., Jiao, Y., Cheng, L., Huang, L., Liao, B., Tang, M., Ren, X., Zhou, X., Chen, Y., & Jiang, H. (2016). Quantitative trait locus analysis for pod- and kernel-related traits in the cultivated peanut (Arachis hypogaea L.). BMC genetics, 17(1). https://doi.org/10.1186/S12863-016-0337-X

Chen, X., Li, H., Pandey, M. K., Yang, Q., Wang, X., Garg, V., Li, H., Chi, X., Doddamani, D., Hong, Y., Upadhyaya, H., Guo, H., Khan, A. W., Zhu, F., Zhang, X., Pan, L., Pierce, G. J., Zhou, G., Krishnamohan, K. A. V. S., Chen, M., Zhong, N., Agarwal, G., Li, S., Chitikineni, A., Zhang, G. Q., Sharma, S., Chen, N., Liu, H., Janila, P., Li, S., Wang, M., Wang, T., Sun, J., Li, X., Li, C., Wang, M., Yu, L., Wen, S., Singh, S., Yang, Z., Zhao, J., Zhang, C., Yu, Y., Bi, J., Zhang, X., Liu, Z. J., Paterson, A. H., Wang, S., Liang, X., Varshney, R. K., & Yu, S. (2016). Draft genome of the peanut A-genome progenitor (Arachis duranensis) provides insights

into geocarpy, oil biosynthesis, and allergens. Proceedings of the National Academy of Sciences, 113(24), 6785-6790. https://doi.org/10.1073/PNAS.1600899113

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., Li, H., Wang, J., Liu, H., Li, S., Pandey, M. K., Zhang, Z., Zhou, G., Yu, J., Zhang, G., Yuan, J., Li, X., Wen, S., Meng, F., Yu, S., Wang, X., Siddique, K. H. M., Liu, Z. J., Paterson, A. H., Varshney, R. K., & Liang, X. (2019). Sequencing of Cultivated Peanut, Arachis hypogaea, Yields Insights into Genome Evolution and Oil Improvement. Molecular Plant, 12(7), 920-934. https://doi.org/10.1016/j.molp.2019.03.005

Chu, Y., Holbrook, C. C., Timper, P., & Ozias-Akins, P. (2007). Development of a PCR-based molecular marker to select for nematode resistance in peanut. Crop Science, 47(2), 841-847. https://doi.org/10.2135/cropsci2006.07.0474

Dang, P., Patel, J., Sorensen, R., Lamb, M., & Chen, C. Y. (2024). Genome-Wide Association Analysis Identified Quantitative Trait Loci (QTLs) Underlying Drought-Related Traits in Cultivated Peanut (Arachis hypogaea L.). Genes, 15(7), 868. https://doi.org/10.3390/GENES15070868

Desmae, H., Janila, P., Okori, P., Pandey, M. K., Motagi, B. N., Monyo, E., Mponda, O., Okello, D., Sako, D., Echeckwu, C., Oteng-Frimpong, R., Miningou, A., Ojiewo, C., & Varshney, R. K. (2018). Genetics, genomics and breeding of groundnut (Arachis hypogaea L.). Plant Breeding = Zeitschrift Fur Pflanzenzuchtung, 138(4), 425. https://doi.org/10.1111/PBR.12645

Dhillon, S. S., Rake, A., & Miksche, J. P. (1980). Reassociation Kinetics and Cytophotometric Characterization of Peanut (Arachis hypogaea L.) DNA. Plant Physiology, 65(6), 1121. https://doi.org/10.1104/PP.65.6.1121

Ersoz, E. S., Yu, J., & Buckler, E. S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. Genomics-Assisted Crop Improvement, 1, 97-119. https://doi.org/10.1007/978-1-4020-6295-7_5

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., & Basra, S. M. A. (2009). Plant drought stress: effects, mechanisms and management. Agronomy for Sustainable Development 2009 29:1, 29(1), 185-212. https://doi.org/10.1051/AGRO:2008021

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., Isleib, T. G., Holbrook, C. C., Xavier, A., Culbreath, A. K., Ozias-Akins, P., Varshney, R. K., & Guo, B. (2020). Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (Arachis hypogaea). Plant biotechnology journal, 18(6), 1457-1471. https://doi.org/10.1111/pbi.13311

Ghosh, S., Mahadevaiah, S. S., Gowda, S. A., Gangurde, S. S., Jadhav, M. P., Hake, A. A., Latha, P., Anitha, T., Chimmad, V. P., Mirajkar, K. K., Sharma, V., Pandey, M. K., Shirasawa, K., Nayak, S. N., Varshney, R. K., & Bhat, R. S. (2022). Genetic mapping of drought tolerance traits phenotyped under varying drought stress environments in peanut (Arachis hypogaea L.). Euphytica, 218(12). https://doi.org/10.1007/S10681-022-03120-X

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. Advances in Genetics, 104, 75-154. https://doi.org/10.1016/bs.adgen.2018.12.001

Gupta, P. K., Varshney, R. K., Varshney, R. K., & Gupta, P. K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, 113(3), 163-185. https://doi.org/10.1023/A:1003910819967

Halward, T., Stalker, H. T., & Kochert, G. (1993). Development of an RFLP linkage map in diploid peanut species. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 87(3), 379-384. https://doi.org/10.1007/BF01184927

Hammons, R. O., Herman, D., & Stalker, H. T. (2016). Origin and Early History of the Peanut. Peanuts: Genetics, Processing, and Utilization, 1-26. https://doi.org/10.1016/B978-1-63067-038-2.00001-0

Holbrook, C. C., & Dong, W. (2005). Development and evaluation of a mini core collection for the U.S. peanut germplasm collection. Crop Science, 45(4), 1540-1544. https://doi.org/10.2135/CROPSCI2004.0368

Hong, Y., Pandey, M. K., Liu, Y., Chen, X., Liu, H., Varshney, R. K., Liang, X., & Huang, S. (2015). Identification and Evaluation of Single-Nucleotide Polymorphisms in

Allotetraploid Peanut (Arachis hypogaea L.) Based on Amplicon Sequencing Combined with High Resolution Melting (HRM) Analysis. Frontiers in plant science, 6(DEC). https://doi.org/10.3389/FPLS.2015.01068

Huang, L., He, H., Chen, W., Ren, X., Chen, Y., Zhou, X., Xia, Y., Wang, X., Jiang, X., Liao, B., & Jiang, H. (2015). Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (Arachis hypogaea L.). TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 128(6), 1103. https://doi.org/10.1007/S00122-015-2493-1

Huang, L., Jiang, H., Ren, X., Chen, Y., Xiao, Y., Zhao, X., Tang, M., Huang, J., Upadhyaya, H. D., & Liao, B. (2012). Abundant Microsatellite Diversity and Oil Content in Wild Arachis Species. PLOS ONE, 7(11), e50002. https://doi.org/10.1371/JOURNAL.PONE.0050002

Isleib, T. G., Holbrook, C. C., & Gorbet, D. W. (2001). Use of Arachis spp. plant introductions in peanut cultivar development. Peanut Science, 28, 96-113.

Janila, P., Nigam, S. N., Pandey, M. K., Nagesh, P., & Varshney, R. K. (2013). Groundnut improvement: Use of genetic and genomic tools. Frontiers in Plant Science, 4(FEB), 41544. https://doi.org/10.3389/FPLS.2013.00023

Kassie, F. C., Nguepjop, J. R., Ngalle, H. B., Assaha, D. V. M., Gessese, M. K., Abtew, W. G., Tossim, H.-A., Sambou, A., Seye, M., & Rami, J.-F. (2023). An Overview of Mapping Quantitative Trait Loci in Peanut (Arachis hypogaea L.). Genes, 14(6), 1176. https://doi.org/10.3390/genes14061176

Khan, S. A., Chen, H., Deng, Y., Chen, Y., Zhang, C., Cai, T., Ali, N., Mamadou, G., Xie, D., Guo, B., Varshney, R. K., & Zhuang, W. (2020). High-density SNP map facilitates fine mapping of QTLs and candidate genes discovery for Aspergillus flavus resistance in peanut (Arachis hypogaea). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 133(7), 2239-2257. https://doi.org/10.1007/S00122-020-03594-0

Kochert, G., Halward, T., Branch, W. D., & Simpson, C. E. (1991). RFLP variability in peanut (Arachis hypogaea L.) cultivars and wild species. TAG. Theoretical and

applied genetics. Theoretische und angewandte Genetik, 81(5), 565-570. https://doi.org/10.1007/BF00226719

Kottapalli, K. R., Burow, M. D., Burow, G., Burke, J., & Puppala, N. (2007). Molecular characterization of the U.S. peanut mini core collection using microsatellite markers. Crop Science, 47(4), 1718-1727. https://doi.org/10.2135/CROPSCI2006.06.0407

Kottapalli, K. R., Rakwal, R., Shibato, J., Burow, G., Tissue, D., Burke, J., Puppala, N., Burow, M., & Payton, P. (2009). Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. Plant, Cell and Environment, 32(4), 380-407. https://doi.org/10.1111/J.1365-3040.2009.01933.X

Krapovickas, A., & Gregory, W. C. (2023). Taxonomía del género Arachis (Leguminosae). Bonplandia, 8(1-4), 1-186. https://doi.org/10.30972/BON.81-43559

Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. Science (New York, N.Y.), 265(5181), 2037-2048. https://doi.org/10.1126/SCIENCE.8091226

Li, A., Zhou, M., Liao, G., Li, X., Wang, A., Xiao, D., He, L., & Zhan, J. (2023). CRISPR/Cas9 gene editing in peanut by Agrobacterium tumefaciens-mediated pollen tube transformation. Plant Cell, Tissue and Organ Culture, 155(3), 883-892. https://doi.org/10.1007/S11240-023-02607-2

Liang, X., Chen, X., Hong, Y., Liu, H., Zhou, G., Li, S., & Guo, B. (2009). Utility of EST-derived SSR in cultivated peanut (Arachis hypogaea L.) and Arachis wild species. BMC Plant Biology, 9, 35. https://doi.org/10.1186/1471-2229-9-35

Liu, H. J., & Yan, J. (2019). Crop genome-wide association study: a harvest of biological relevance. The Plant journal : for cell and molecular biology, 97(1), 8-18. https://doi.org/10.1111/TPJ.14139

Liu, N., Guo, J., Zhou, X., Wu, B., Huang, L., Luo, H., Chen, Y., Chen, W., Lei, Y., Huang, Y., Liao, B., & Jiang, H. (2019). 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.). TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 133(1), 37. https://doi.org/10.1007/S00122-019-03438-6

Lu, Q., Li, H., Hong, Y., Zhang, G., Wen, S., Li, X., Zhou, G., Li, S., Liu, H., Liu, H., Liu, Z., Varshney, R. K., Chen, X., & Liang, X. (2018). Genome sequencing and analysis of the peanut B-genome progenitor (Arachis ipaensis). Frontiers in Plant Science, 9, 361230. https://doi.org/10.3389/FPLS.2018.00604

Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. Trends in Plant Science, 12(2), 57-63. https://doi.org/10.1016/j.tplants.2006.12.001

Nawade, B., Mishra, G. P., Radhakrishnan, T., Sangh, C., Dobariya, J. R., & Kundu, R. (2019). Development of high oleic peanut lines through marker-assisted introgression of mutant ahFAD2 alleles and its fatty acid profiles under open-field and controlled conditions. 3 Biotech, 9(6), 243. https://doi.org/10.1007/s13205-019-1774-9

Otyama, P. I., Kulkarni, R., Chamberlin, K., Ozias-Akins, P., Chu, Y., Lincoln, L. M., MacDonald, G. E., Anglin, N. L., Dash, S., Bertioli, D. J., Fernández-Baca, D., Graham, M. A., Cannon, S. B., & Cannon, E. K. S. (2020). Genotypic Characterization of the U.S. Peanut Core Collection. G3: Genes|Genomes|Genetics, 10(11), 4013. https://doi.org/10.1534/G3.120.401306

Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., Chitikineni, A., Chavarro, C., Chen, X., Upadhyaya, H. D., Vishwakarma, M. K., Leal-Bertioli, S., Liang, X., Bertioli, D. J., Guo, B., Jackson, S. A., Ozias-Akins, P., & Varshney, R. K. (2017). Development and Evaluation of a High Density Genotyping 'Axiom_Arachis' Array with 58 K SNPs for Accelerating Genetics and Breeding in Groundnut. Scientific Reports 2017 7:1, 7(1), 40577-. https://doi.org/10.1038/srep40577

Pandey, M. K., Monyo, E., Ozias-Akins, P., Liang, X., Guimarães, P., Nigam, S. N., Upadhyaya, H. D., Janila, P., Zhang, X., Guo, B., Cook, D. R., Bertioli, D. J., Michelmore, R., & Varshney, R. K. (2012). Advances in Arachis genomics for peanut improvement. Biotechnology Advances, 30(3), 639-651. https://doi.org/10.1016/j.biotechadv.2011.11.001

Pandey, M. K., Upadhyaya, H. D., Rathore, A., Vadez, V., Sheshshayee, M. S., Sriswathi, M., Govil, M., Kumar, A., Gowda, M. V. C., Sharma, S., Hamidou, F.,

Kumar, V. A., Khera, P., Bhat, R. S., Khan, A. W., Singh, S., Li, H., Monyo, E., Nadaf, H. L., Mukri, G., Jackson, S. A., Guo, B., Liang, X., & Varshney, R. K. (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. https://doi.org/10.1371/JOURNAL.PONE.0105228

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., Tonnis, B., Barkley, N. A., Wang, J., Holbrook, C. C., Culbreath, A. K., Varshney, R. K., & Guo, B. (2014). Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (Arachis hypogaea L.). BMC Genetics, 15(1), 133. https://doi.org/10.1186/S12863-014-0133-4

Pandey, M. K., Gangurde, S. S., Shasidhar, Y., Sharma, V., Kale, S. M., Khan, A. W., Shah, P., Joshi, P., Bhat, R. S., Janila, P., Bera, S. K., & Varshney, R. K. (2024). High-throughput diagnostic markers for foliar fungal disease resistance and high oleic acid content in groundnut. BMC Plant Biology, 24(1), 262. https://doi.org/10.1186/s12870-024-04987-9

Paran, I., & Michelmore, R. W. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 85(8), 985-993. https://doi.org/10.1007/BF00215038

Pokhrel, S., Kharel, P., Pandey, S., Botton, S., Nugraha, G. T., Holbrook, C., & Ozias-Akins, P. (2024). Understanding the impacts of drought on peanuts (Arachis hypogaea L.): exploring physio-genetic mechanisms to develop drought-resilient peanut cultivars. Frontiers in Genetics, 15, 1492434. https://doi.org/10.3389/FGENE.2024.1492434

Ramos, M. L., Fleming, G., Chu, Y., Akiyama, Y., Gallo, M., & Ozias-Akins, P. (2006). Chromosomal and phylogenetic context for conglutin genes in Arachis based on genomic sequence. Molecular genetics and genomics : MGG, 275(6), 578-592. https://doi.org/10.1007/S00438-006-0114-Z

Ravi, K., Vadez, V., Isobe, S., Mir, R. R., Guo, Y., Nigam, S. N., Gowda, M. V. C., Radhakrishnan, T., Bertioli, D. J., Knapp, S. J., & Varshney, R. K. (2011). Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (Arachis hypogaea L.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 122(6), 1119-1132. https://doi.org/10.1007/S00122-010-1517-0

Ribaut, J. M., & Hoisington, D. (1998). Marker-assisted selection: New tools and strategies. Trends in Plant Science, 3(6), 236-239. https://doi.org/10.1016/S1360-1385(98)01240-0

Robarts, D. W. H., & Wolfe, A. D. (2014). Sequence-related amplified polymorphism (SRAP) markers: A potential resource for studies in plant molecular biology. Applications in Plant Sciences, 2(7), apps.1400017. https://doi.org/10.3732/APPS.1400017

Samoluk, S. S., Robledo, G., Podio, M., Chalup, L., Ortiz, J. P. A., Pessino, S. C., & Seijo, J. G. (2015). First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. Genetica, 143(1), 113-125. https://doi.org/10.1007/S10709-015-9820-Y

Seijo, G., Lavia, G. I., Fernández, A., Krapovickas, A., Ducasse, D. A., Bertioli, D. J., & Moscone, E. A. (2007). Genomic relationships between the cultivated peanut (Arachis hypogaea, Leguminosae) and its close relatives revealed by double GISH. American journal of botany, 94(12), 1963-1971. https://doi.org/10.3732/AJB.94.12.1963

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. Molecular Breeding, 33(1), 1-14. https://doi.org/10.1007/S11032-013-9917-X

Shasidhar, Y., Vishwakarma, M. K., Pandey, M. K., Janila, P., Variath, M. T., Manohar, S. S., Nigam, S. N., Guo, B., & Varshney, R. K. (2017). Molecular mapping of oil

content and fatty acids using dense genetic maps in groundnut (Arachis hypogaea L.). Frontiers in Plant Science, 8, 794. https://doi.org/10.3389/FPLS.2017.00794

Shirasawa, K., Bertioli, D. J., Varshney, R. K., Moretzsohn, M. C., Leal-Bertioli, S. C. M., Thudi, M., Pandey, M. K., Rami, J. F., Ka, D. F., Gowda, M. V. C., Qin, H., Guo, B., Hong, Y., Liang, X., Hirakawa, H., Tabata, S., & Isobe, S. (2013). Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of Arachis and divergence of the legume genomes. DNA research : an international journal for rapid publication of reports on genes and genomes, 20(2), 173-184. https://doi.org/10.1093/DNARES/DSS042

Shirasawa, K., Koilkonda, P., Aoki, K., Hirakawa, H., Tabata, S., Watanabe, M., Hasegawa, M., Kiyoshima, H., Suzuki, S., Kuwata, C., Naito, Y., Kuboyama, T., Nakaya, A., Sasamoto, S., Watanabe, A., Kato, M., Kawashima, K., Kishida, Y., Kohara, M., Kurabayashi, A., Takahashi, C., Tsuruoka, H., Wada, T., & Isobe, S. (2012). In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biology, 12, 80. https://doi.org/10.1186/1471-2229-12-80

Soba, D., Parker, S., Chen, C., Shekoofa, A., & Sanz-Saez, A. (2024). Peanut photosynthesis response to drought can include diffusive and biochemical limitations depending on cultivar. Physiologia Plantarum, 176(4), e14489. https://doi.org/10.1111/PPL.14489

Sujay, V., Gowda, M. V. C., Pandey, M. K., Bhat, R. S., Khedikar, Y. P., Nadaf, H. L., Gautami, B., Sarvamangala, C., Lingaraju, S., Radhakrishan, T., Knapp, S. J., & Varshney, R. K. (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.). Molecular breeding : new strategies in plant improvement, 30(2), 773-788. https://doi.org/10.1007/S11032-011-9661-Z

Sukumaran, S., & Yu, J. (2014). Association mapping of genetic resources: Achievements and future perspectives. Genomics of Plant Genetic Resources:

Volume 1. Managing, Sequencing and Mining Genetic Resources, 207-235. https://doi.org/10.1007/978-94-007-7572-5_9

Sun, Z., Qi, F., Liu, H., Qin, L., Xu, J., Shi, L., Zhang, Z., Miao, L., Huang, B., Dong, W., Wang, X., Tian, M., Feng, J., Zhao, R., Zheng, Z., & Zhang, X. (2022). QTL mapping of quality traits in peanut using whole-genome resequencing. The Crop Journal, 10(1), 177-184. https://doi.org/10.1016/J.CJ.2021.04.008

Tang, R., Gao, G., He, L., Han, Z., Shan, S., Zhong, R., Zhou, C., Jiang, J., Li, Y., & Zhuang, W. (2007). Genetic Diversity in Cultivated Groundnut Based on SSR Markers. Journal of Genetics and Genomics, 34(5), 449-459. https://doi.org/10.1016/S1673-8527(07)60049-6

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics, 132(2), 619-633. https://doi.org/10.1093/GENETICS/132.2.619

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., & Buckler, E. S. (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature genetics, 28(3), 286-289. https://doi.org/10.1038/90135

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. Plant Genome, 14(1), e20077. https://doi.org/10.1002/TPG2.20077

Varshney, R. K., Bertioli, D. J., Moretzsohn, M. C., Vadez, V., Krishnamurthy, L., Aruna, R., Nigam, S. N., Moss, B. J., Seetha, K., Ravi, K., He, G., Knapp, S. J., & Hoisington, D. A. (2009). The first SSR-based genetic linkage map for cultivated groundnut (Arachis hypogaea L.). TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 118(4), 729-739. https://doi.org/10.1007/S00122-008-0933-X

Varshney, R. K., Mohan, S. M., Gaur, P. M., Gangarao, N. V. P. R., Pandey, M. K., Bohra, A., Sawargaonkar, S. L., Chitikineni, A., Kimurto, P. K., Janila, P., Saxena, K. B., Fikre, A., Sharma, M., Rathore, A., Pratap, A., Tripathi, S., Datta, S., Chaturvedi,

S. K., Mallikarjuna, N., Anuradha, G., Babbar, A., Choudhary, A. K., Mhase, M. B., Bharadwaj, C., Mannur, D. M., Harer, P. N., Guo, B., Liang, X., Nadarajan, N., & Gowda, C. L. L. (2013). Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnology Advances, 31(8), 1120-1134. https://doi.org/10.1016/J.BIOTECHADV.2013.01.001

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic acids research, 23(21), 4407-4414. https://doi.org/10.1093/NAR/23.21.4407

Wang, H., Pandey, M. K., Qiao, L., Qin, H., Culbreath, A. K., He, G., Varshney, R. K., Scully, B. T., & Guo, B. (2013). Genetic Mapping and Quantitative Trait Loci Analysis for Disease Resistance Using F 2 and F 5 Generation-based Genetic Maps Derived from 'Tifrunner' × 'GT-C20' in Peanut. The Plant Genome, 6(3). https://doi.org/10.3835/PLANTGENOME2013.05.0018

Wang, H., Penmetsa, R. V., Yuan, M., Gong, L., Zhao, Y., Guo, B., Farmer, A. D., Rosen, B. D., Gao, J., Isobe, S., Bertioli, D. J., Varshney, R. K., Cook, D. R., & He, G. (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 Biology, 12, 10. https://doi.org/10.1186/1471-2229-12-10

Wang, J., Yan, C., Li, Y., Li, C., Zhao, X., Yuan, C., Sun, Q., & Shan, S. (2019). GWAS Discovery of Candidate Genes for Yield-Related Traits in Peanut and Support from Earlier QTL Mapping Studies. Genes, 10(10), 803. https://doi.org/10.3390/GENES10100803

Wang, M. L., Khera, P., Pandey, M. K., Wang, H., Qiao, L., Feng, S., Tonnis, B., Barkley, N. A., Pinnow, D., Holbrook, C. C., Culbreath, A. K., Varshney, R. K., & Guo, B. (2015). Genetic Mapping of QTLs Controlling Fatty Acids Provided Insights into the Genetic Control of Fatty Acid Synthesis Pathway in Peanut (Arachis hypogaea L.). PLOS ONE, 10(4), e0119454. https://doi.org/10.1371/JOURNAL.PONE.0119454

Wang, X., Yang, X., Feng, Y., Dang, P., Wang, W., Graze, R., Clevenger, J. P., Chu, Y., Ozias-Akins, P., Holbrook, C., & Chen, C. (2021). Transcriptome Profile Reveals Drought-Induced Genes Preferentially Expressed in Response to Water Deficit in Cultivated Peanut (Arachis hypogaea L.). Frontiers in Plant Science, 12, 645291. https://doi.org/10.3389/FPLS.2021.645291

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research, 18(22), 6531-6535. https://doi.org/10.1093/NAR/18.22.6531

Yin, D., Ji, C., Ma, X., Li, H., Zhang, W., Li, S., Liu, F., Zhao, K., Li, F., Li, K., Ning, L., He, J., Wang, Y., Zhao, F., Xie, Y., Zheng, H., Zhang, X., Zhang, Y., & Zhang, J. (2018). Genome of an allotetraploid wild peanut Arachis monticola: a de novo assembly. GigaScience, 7(6). https://doi.org/10.1093/GIGASCIENCE/GIY066

Zhang, H., Chu, Y., Dang, P., Tang, Y., Jiang, T., Clevenger, J. P., Ozias-Akins, P., Holbrook, C., Wang, M. L., Campbell, H., Hagan, A., & Chen, C. (2020). Identification of QTLs for resistance to leaf spots in cultivated peanut (Arachis hypogaea L.) through GWAS analysis. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 133(7), 2051-2061. https://doi.org/10.1007/S00122-020-03576-2

Zhang, X., Zhang, J., He, X., Wang, Y., Ma, X., & Yin, D. (2017). Genome-wide association study of major agronomic traits related to domestication in peanut. Frontiers in Plant Science, 8, 278584. https://doi.org/10.3389/FPLS.2017.01611

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., Chang, W. C., Zhang, L., Zhang, X., Tang, R., Garg, V., Wang, X., Tang, H., Chow, C. N., Wang, J., Deng, Y., Wang, D., Khan, A. W., Yang, Q., Cai, T., Bajaj, P., Wu, K., Guo, B., Zhang, X., Li, J., Liang, F., Hu, J., Liao, B., Liu, S., Chitikineni, A., Yan, H., Zheng, Y., Shan, S., Liu, Q., Xie, D., Wang, Z., Khan, S. A., Ali, N., Zhao, C., Li, X., Luo, Z., Zhang, S., Zhuang, R., Peng, Z., Wang, S., Mamadou, G., Zhuang, Y., Zhao, Z., Yu, W., Xiong, F., Quan, W., Yuan, M., Li, Y., Zou, H., Xia, H., Zha, L., Fan, J., Yu, J., Xie, W., Yuan, J., Chen, K., Zhao, S., Chu, W., Chen, Y., Sun, P., Meng, F., Zhuo, T., Zhao, Y., Li, C., He, G., Zhao, Y., Wang, C., Kavikishor, P. B., Pan, R. L.,

Paterson, A. H., Wang, X., Ming, R., & Varshney, R. K. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature Genetics, 51(5), 865-876. https://doi.org/10.1038/S41588-019-0402-2

Chapter 2: Genome-wide Identification and Expression Analysis Reveals the Drought-response MAPK Genes in Peanut

Abstract

Peanut (*Arachis hypogaea L.*) is one of most important oilseed and food crops, the drought stress remains the primary adverse environmental factor limiting its growth and productivity. Mitogen-activated protein kinase (MAPK) cascades play crucial roles in various signal transduction pathways, affecting a wide range of physiological processes and drought stress responses in plants, however the systematic analysis of the MAPK gene family in peanuts remains unexplored. In this study, we identified 30, 16, and 15 MAPK genes in *Arachis hypogaea*, *Arachis duranensis*, and *Arachis ipaensis*, respectively. RNA-seq analysis in drought-tolerant and drought-susceptible genotypes revealed that *Ah_At_MAPK4* and *Ah_Bt_MAPK4* were significantly upregulated under drought stress conditions, with substantially higher induction in drought-tolerant genotypes compared to drought-susceptible ones. Weighted Gene Co-expression Network Analysis (WGCNA) further identified a drought-responsive turquoise module highly correlated with drought tolerance traits, and both *Ah_At_MAPK4* and *Ah_Bt_MAPK4* were identified as core regulatory components within this module. Hub gene analysis revealed these MAPKs co-express with calmodulin-binding proteins, implicating calcium signaling in drought adaptation. Three-dimensional structural modeling confirmed both proteins possess canonical bilobed kinase architecture with properly positioned TEY motifs and intact catalytic machinery. This genome-to-structure analysis identifies *Ah_At_MAPK4* and *Ah_Bt_MAPK4* as key components in drought-responsive networks and provides molecular targets for enhancing drought resilience in peanut breeding.

2.1 Introduction

Peanut (*A. hypogaea*) is a globally important oilseed and food crop valued for its high nutritional content, including oil, protein, dietary fiber and vitamins (Arya et al. 2016). As an allotetraploid species (AABB, 2n=4x=40) derived from two wild diploid relatives, *A. duranensis* (A-genome) and *A. ipaensis* (B-genome), cultivated peanut possesses unique genomic complexity. Peanut production faces many challenges,

including biotic stresses such as pests and diseases, and abiotic stresses including salinity, heat, and particularly drought. Drought stress remains the primary adverse environmental factor limiting its growth and productivity (Puppala et al. 2023). With climate change intensifying the frequency and severity of drought events in major peanut-growing regions, understanding the molecular mechanisms underlying drought tolerance has become crucial for developing resilient peanut varieties and ensuring food security.

Plants utilize complicated signaling networks to adapt to complex stress conditions when exposed to changes in their natural environment (Hamel et al. 2006; Krasensky and Jonak 2012; Nawaz et al. 2023; Yang et al. 2020). In these complicated signaling networks, stress-activated molecular pathways play an important role in responding to various biotic and abiotic stresses (Kim et al. 2016; Zhu 2016; Sewelam et al. 2016; Suzuki et al. 2014). These may include injury from external forces, excessive sodium content, cold conditions, drought, and disease invasion (Wang et al. 2022). In previous studies, many gene families have been reported to regulate plant abiotic and biotic stresses. For example, the WRKY superfamily of plant transcription factors is involved in regulating response to various stresses (Wu et al. 2024), the MYB family plays important role in hormone signal transduction and abiotic stress tolerance (Ambawat et al. 2013), and mitogen-activated protein kinase (MAPK) are associated with abiotic and biotic stress adaptation (Manna et al. 2023). The MAPK family of transcription factors is particularly interesting, since they function within the pathway to transfer external stimuli to the cells, and are also critical in the developmental and physiological processes of plants. The MAPK cascade pathway transmits signals to downstream proteins through phosphorylation (Zhang and Zhang 2022; He et al. 2020). Additionally, MAPK cascades are involved in the regulation of growth and development in plants, acting downstream of receptor-like kinases (RLKs) and playing a role in processes like cell division and differentiation (Xu and Zhang 2015). MAPKs were reported to be categorized into four groups: A, B, C, and D. Members of A, B and C contain Thr-Glu-Tyr (TEY) motif while members of group D contains Thr-Asp-Tyr (TDY) motif (Group 2002). The MAPK signaling pathway is comprised of three sequential

protein kinases: MAPK, MAPKK, and MAPKKK. In response to external stimuli, MAPKKKs are activated first, then phosphorylate downstream MAPKKs through two serine/threonine residues in S/T-X3-5-S/T conserved motifs (Taj et al. 2010). Finally, the phosphorylated MAPKKs phosphorylate MAPKs, always on threonine and tyrosine residues in TXY motif, which then may phosphorylate other downstream transcription factors and signaling components (Chen et al. 2020b; Jimenez-Sanchez et al. 2007). The result of the MAPK cascade pathway is the activation of genes responsible for resistance.

Due to MAPK's key role in stress response and as a highly conserved sub-family of protein kinase (Meng and Zhang 2013), more research has been carried out to understand the mechanism of its actions in plants, alluding to its importance (Colcombet and Hirt 2008; Kong et al. 2012; Xu and Zhang 2015; Xu et al. 2018). In a recent study, researchers investigated the MAPK gene family within Tartary buckwheat (*Fagopyrum tataricum)*, identifying 16 distinct MAPK genes. Their analysis unveiled distinct groupings distinguished by specific motifs and promoter elements, particularly linked to light response, hormone regulation, and responses to abiotic stress. Notably, the expansion of MAPKs containing the TDY motif was observed, hinting at potential roles in developmental processes and stress resistance mechanisms. A total of 18 *AcMAPKs* were identified in kiwifruit (*Actinidia Chinensis*) (Wang et al. 2018), the expression profiles demonstrated that *AcMAPKs* were induced or repressed by various biotic and abiotic stresses and hormone treatments, suggesting their potential roles in the biotic and abiotic stress response and various hormone signal transduction pathways. A comprehensive analysis identified 17 *LsMAPKs* in lettuce (*Lactuca sativa L.*), and transgenic experiments reveal the silencing of *LsMAPK4* significantly inhibited the high-temperature-accelerated bolting in lettuce, indicating that *LsMPAK4* might be a potential regulator of lettuce bolting (Wang et al. 2022). A total of 5 MAPKK genes and 89 MAPKKK genes were identified in tomato, where the expression level of most of these genes changed significantly under multiple treatments, especially in response to salicylic acid and indole-3-acetic acid treatment, implying that these genes might have important roles in the plant hormone networks (Wu et al. 2014). Additionally, the

structurally non-canonical MAPKK, *MPKK10.2*, could enhance rice resistance to *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*) and increases rice tolerance to drought stress by phosphorylating and activating two MAPKs, *MPK6* and *MPK3* (Ma et al. 2017). In drought stress responses, MAPKs function as critical signaling components that integrate stress perception with adaptive responses. Overexpression of the Nicotiana protein kinase *NPK1* enhanced drought tolerance in transgenic maize (Shou et al. 2004), while in cotton, a group C MAP kinase gene, *GhMPK2*, positively regulated both salt and drought tolerance through abscisic acid (ABA) - dependent signaling pathways (Zhang et al. 2018). Comprehensive reviews have established that MAPK signaling pathways work in concert with ABA-dependent pathways to orchestrate plant abiotic stress responses (Danquah et al. 2014). Overall, the main functional categories of MAPKs include abiotic stress signaling, hormone regulation (notably ABA signaling), developmental processes, and biotic stress response. These findings contribute to a foundational understanding of the MAPK gene family and deepen our understanding of stress responses at the molecular level, laying the groundwork for future functional studies and presenting promising avenues for enhancing crop resilience and productivity through targeted interventions.

Despite the agricultural importance of peanut and its susceptibility to drought stress, systematic characterization of the MAPK gene family in peanut remains limited. Early studies have identified individual peanut MAPK genes through heterologous expression: *AhMPK3*, containing a TEY motif, enhanced defense responses when expressed in tobacco (Kumar et al. 2009), and *AhMPK6* induced hypersensitive response-like cell death and upregulated defense-related transcripts (Kumar and Kirti 2010). In a transcriptome analysis of leaves under well-watered and drought-stress conditions for the drought-tolerant cultivar “L422” of peanut, a total of 1,313 significantly expressed genes were identified at three-time points throughout the drought stress stage; six vital metabolic pathways, including the “MAPK signaling pathway-plant,” were enriched under severe drought stress (Zhao et al. 2021). These findings revealed the MAPK genes play a potential role in environmental stress response. Given the significance of the MAPK gene family in plants, it has been extensively studied and

identified across various plant species. However, these studies focused on isolated genes or global transcriptomic patterns rather than comprehensive family-wide characterization.

In this study, we conducted a genome-wide identification of MAPK gene family members in peanut species, including two diploid species *A. duranensis* and *A. ipaensis,* and the tetraploid cultivated species *A. hypogaea*. A total of 30 *Ah_MAPKs*, 16 *Du_MAPKs*, and 15 *Ip_MAPKs* were identified in *A. hypogaea, A. duranensis* and *A. ipaensis*, respectively. We performed a structural analysis along with an investigation into their potential functions under drought stress. The aim of the study was to provide a comprehensive understanding of MAPK signaling pathways in peanuts, which could contribute to future research on gene cloning, expression, and genetic improvement in peanut breeding.

2.2 Methods and materials

Gene resource

To identify MAPK genes within the peanut species, including the diploid ancestors *A. ipaensis* (aradu.V14167.gnm1) and *A. duranensis* (K30076.gnm2.1GWY), as well as the tetraploid cultivated species *A. hypogaea* (Tifrunner.gnm2.J5K5), the genome and protein sequences were downloaded from PeanutBase (https://www.peanutbase.org/). The MAPK protein sequences of *Arabidopsis* were obtained from TAIR (https://www.arabidopsis.org/). The protein kinase domain (PF00069) was downloaded from the Pfam database (https://www.ebi.ac.uk/interpro/entry/pfam/). For gene functional annotation of *A. hypogaea* , InterProScan v5.55-88.0 was applied to predict Gene Ontology (GO) Term with parameters "-goterms -iprlookup -appl Pfam"and eggnog v2.12 was applied to predict the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway with default parameters.

Genome-wide identification of MAPK genes in peanut

First, the *Arabidopsis thaliana* MAPK protein sequences were used as the search seed for a BLASTP (McGinnis and Madden 2004) homology search against the

predicted peanut protein models derived from the whole-genome data with an e-value of 1×10^{-5}, a minimum identity of 50% and a minimum coverage of 50% as a threshold. Then, the HMMER v3 (Mistry et al. 2021) program was used to remove redundant sequences by applying the serine/threonine protein kinase-like domain (PF00069) as a query for hidden Markov model (HMM) searches with a threshold of 1×10^{-5}. Finally, the SMART database (http://smart.embl-heidelberg.de/) was then used to detect potential mobile domains and domain architecture. WoLF PSORT (https://www.genscript.com/wolf-psort.html) was used to predict the subcellular localization of identified MAPK genes (Horton et al. 2007).

Sequence alignment and phylogenetic analysis

The protein sequences of MAPK genes from peanut and *Arabidopsis thaliana* were aligned using MUSCLE v3.8.31 (Edgar 2004). Maximum likelihood inference of phylogenetic relationships was performed using PhyML v3.3 (Guindon et al. 2010) with 1000 bootstrap replicates. The phylogenetic tree was visualized by iTOL (https://itol.embl.de/).

Motifs and gene structure analysis

Motif analysis was performed using the MEME Suite version 5.0.5 (Bailey et al. 2015). The tool was configured to identify conserved motifs within the protein sequences. The search was set to discover a total of 10 motifs, with each motif's width ranging from 6 to 100 amino acids. Based on the gff3 file, the gene structure was analyzed using the TBtools v2.008 (Chen et al. 2020a). Finally, TBtools was utilized to visualize the results of gene structure and conserved motif analysis.

Cis-regulatory element analysis

To further investigate cis-regulatory elements in the promoter regions of the identified MAPK genes, we obtained 2000 bp sequence upstream of the start codon for identified MAPK genes, and then these sequences were analyzed by the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The cis-acting elements of MAPK genes were identified and further visualized using TBtools software.

Chromosome distribution and collinearity analysis

The genomic locations of the identified MAPK genes were determined using annotation files. The chromosomal positions of the identified MAPK genes were then visualized using the TBtools. To identify syntenic genes, protein sequences were compared by all-versus-all BLASTP v2.10.0 with the parameters "-evalue 10e-5 -outfmt 6 -num_alignments 15" (Schaffer et al. 2001). Then the putatively homologous genes identified by BLASTP were analyzed for synteny by the MCScanX package using default settings (Wang et al. 2012). The syntenic genes were displayed with JCVI (https://github.com/tanghaibao/jcvi). The homologous gene pairs were used to calculate the Estimation of Nonsynonymous (Ka) and Synonymous (Ks) values by using KaKs_Calculator v2.0 (Wang et al. 2010).

Expression analysis of MAPK genes under drought stress

To investigate the expression patterns of the identified MAPK genes under drought stress conditions, we conducted an RNA-seq analysis. The RNA-seq data were obtained from a previous transcriptomic study performed in our laboratory on cultivated peanut (*A. hypogaea L.*) subjected to drought stress (PRJNA687542) (Wang et al. 2021). Four peanut genotypes were selected based on our previously characterized drought response phenotypes. The drought-tolerant genotypes (C76-16 and 587) exhibited higher pod yield under drought stress, lower specific leaf area, cooler canopy temperature, and less severe visual drought stress symptoms compared to the drought-susceptible genotypes (Tifrunner and 506), which showed opposite trends (Carter, 2015; Wang et al., 2021). Plants were grown under rainout shelters at the USDA Agricultural Research Service National Peanut Research Laboratory in Dawson, GA, United States. Two treatments were applied: full irrigation throughout the growing season (irrigated control) and middle-season drought stress initiated 61 days after planting. Drought stress was progressively applied over four weeks, with soil water potential reaching −1,050 to −1,200 kPa during the second week of treatment. Fully expanded leaves were collected from each genotype at the end of the drought period for RNA extraction. Total RNA was extracted using a modified cetyltrimethylammonium bromide method and purified using a Direct-Zol RNA MiniPrep Kit. RNA quality was

assessed using a NanoDrop ND-1000 spectrophotometer and Agilent 2100 Bioanalyzer. A total of 24 cDNA libraries (4 genotypes × 2 treatments × 3 replicates) were constructed and sequenced using an Illumina HiSeq 4000 platform. The RNA-seq data was trimmed of low-quality reads and adapter sequences using Fastp v. 0.23.4 with default parameters. Clean reads were mapped to the Tifrunner.gnm2.J5K5 reference genome using HISAT2 v2.1.0 (Kim et al. 2015; Bertioli et al. 2019). Only those reads with high mapping quality against the reference genome (-q 20) were kept. The mapped reads of each sample were assembled using StringTie v2.1.1. The gene expression level was normalized using the FPKM (Fragments per Kilobase of transcript per Million mapped reads) method. Principal Component Analysis (PCA) was performed on the normalized expression data using log2(count +1) transformed counts after filtering low-expression genes and retaining those with variance >0.5 across samples. The differentially expressed genes (DEGs) were identified using DESeq2 v 1.36.0 with the threshold of 4-fold expression changes | log2(fold-change) | ≥ 2 and p-value < 0.05.

Weighted gene co-expression network analysis

WGCNA was performed to identify gene modules and hub genes associated with drought stress responses in peanut. Raw RNA-seq counts from peanut samples under control and drought conditions were normalized using the regularized-logarithm transformation (rlog) from the DESeq2 package. The gene set for network construction consisted of all identified DEGs across four genotypes, combined with a curated list of 30 MAPK family genes. A soft-thresholding power was chosen using the scale-free topology criterion. An unsigned adjacency matrix was then transformed into a Topological Overlap Matrix (TOM). Gene modules were identified using dynamic tree cutting (minModuleSize = 30), and modules with high eigengene similarity were merged using a mergeCutHeight threshold of 0.25. Associations between module eigengenes and key phenotypic traits (drought, tolerant, tolerant × drought, and susceptible × drought) were assessed by Pearson's correlation, and the module showing the highest correlation with the tolerant × drought trait was selected. Hub genes in this module were defined by high module membership (MM) and gene significance (GS). A sub-network

of the top 20 hub genes and two key MAPK genes (*Ah_At_MAPK4* and *Ah_Bt_MAPK4*) was visualized in Cytoscape (v3.10.3), with edges weight > 0.78.

AlphaFold predicted structures analysis

The two targeted MAPK genes that we analyzed were obtained as pdb files from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/), and their structure models were visualized using PyMOL (3.1.6.1).

2.3 Results

Genome-wide identification of MAPK genes in peanut

A combination of BLASTp and HMM-based methods were used to identify MAPK family genes in three peanut species. Based on the intersection of results obtained from both BLASTP and HMM-based methods, we identified 30 MAPK genes in *A. hypogaea* renamed *Ah_At_MAPK1* - *Ah_At_MAPK14* and *Ah_Bt_MAPK1* - *Ah_Bt_MAPK16* based on the locations of subgenomes, 16 MAPK genes in *A. duranensis* renamed *Du_MAPK1* - *Du_MAPK16*, and 15 MAPK genes in *A. ipaensis* renamed *Ip_MAPK1* - *Ip_MAPK15*.

In addition, we performed molecular characterization, including analyses of protein length, molecular weight, isoelectric point, and subcellular localization, to infer the function of identified MAPK genes (Table 2.1). These putative *Ah_MAPK* genes were predicted to encode 368 to 612 amino acids in length, with putative molecular weights (MW) ranging from 42269.8 kDa to 69510.1 kDa, and protein isoelectric points (PI) ranging from 4.73 to 9.59. The putative *Du_MAPK* genes were predicted to encode 320 to 669 amino acids in length, with putative molecular weights (MW) ranging from 37009.9 kDa to 76405.3 kDa, and protein isoelectric points (PI) ranging from 4.69 to 9.47. The putative *Ip_MAPK* genes were predicted to encode 332 to 667 amino acids in length, with putative molecular weights (MW) ranging from 38276.3 kDa to 76019.9 kDa, and protein isoelectric points (PI) ranging from 4.58 to 9.5 (Table 2.1).

Subcellular localization prediction has revealed that within *A. hypogaea*, 4 putative *Ah_MAPKs* were located in the Cytoskeleton, 22 in the Cytoplasm, and 4 in the

Nucleus. In the case of *A. duranensis*, 2 *Du_MAPKs* were identified in the Chloroplast, 2 in the Cytoskeleton, 9 in the Cytoplasm, 1 in the Mitochondrion, and 2 in the Nucleus. As for *A. ipaensis*, 2 *Ip_MAPKs* were located in the Cytoskeleton, 9 in the Cytoplasm, 1 in the Endoplasmic Reticulum, 1 in the Mitochondrion, 1 in the Nucleus, and 1 in the Plastids (Table 2.1).

Phylogenetic analysis of MAPK genes

To better investigate the evolutionary relationship among the MAPK proteins, a Maximum Likelihood phylogenetic tree was constructed to determine the evolutionary relationship of MAPK genes by using amino acid sequences of identified 16,15 and 30 MAPK genes in *A. duranensis*, *A. ipaensis* and *A. hypogaea*, respectively, and with corresponding 20 *AtMAPKs* of *A. thaliana* (Yao et al. 2022). In plants, MAPK proteins have diverged into four major subfamilies (A-D), sequence comparison containing the amino acid motif TEY can be classified into three Groups, A, B and C, whereas sequence comparison containing the amino acid motif TDY forms a more distant Group D (Ichimura et al. 2002). This categorization is consistent with previous studies conducted on *Arabidopsis*, maize, and *Fagopyrum tataricum*. (Ichimura et al. 2002; Liu et al. 2013; Yao et al. 2022). Our phylogenetic analysis also showed the peanut MAPK genes were divided into four distinct groups (Groups A, B, C and D) specifically together with their MAPK orthologs in *Arabidopsis* (Figure 2.1). Group Acontained 4 *Ah_MAPKs*, 1 *Ip_MAPK*, and 2 *Du_MAPKs*, clustering with the well-characterized *AtMAPK3/6/10*. Group B comprised 8 *Ah_MAPKs*, 5 *Ip_MAPKs*, and 5 *Du_MAPKs*, grouping with *AtMAPK4/5/11/12/13*. Group C included 4 *Ah_MAPKs*, 2 *Ip_MAPKs*, and 2 *Du_MAPKs*, associated with *AtMAPK1/2/7/14*. Group D was the largest subfamily, containing 14 *Ah_MAPKs*, 7 *Ip_MAPKs*, and 7 *Du_MAPKs*, clustering with *AtMAPK8/9/15/16/18/19/20*. The distribution pattern showed Group D contained the most members, followed by Groups B and C, while Group A had the fewest genes.

Gene structure and motif composition of putative *MAPKs* in peanut

To further analyze the potential roles of putative MAPKs, we conducted exon-intron structures, motif distribution, and conserved domain analysis. The exon/intron

structures, including the number, length, and arrangement of exons and introns, were analyzed to provide insights into the gene structure diversity and phylogenetic relationship among members of the gene family. Structures and splice phases (0, 1, or 2) of introns/exons were determined by the alignment of genomic DNA with full-length cDNA of MAPKs. The exon/intron structures and motif location of putative MAPK genes can also be categorized into four groups, based on their phylogenetic relationships (Figure 2.2). Gene structure analysis showed that MAPK gene family structure was relatively complex and putative MAPK genes from different groups exhibit significantly different exon/intron structures, while the MAPK genes were highly conserved within the same group. The gene structures of putative MAPK genes in Groups A, B, and C exhibit more conservation compared to those in Group D, which display a complex arrangement of exons and introns.

The sequence characteristics of peanut MAPK genes were analyzed using MEME software, and a total of 12 distinct motifs were predicted. The motifs are more conserved for each of three peanut species. For *A. hypogaea*, those 10 conserved motifs with the same order and orient, including motif 11, 3, 4, 1, 7, 2, 8, 5, 6, 9, were identified in all D group, and 8 conserved motifs, including motif 12, 3, 4, 1, 7, 2, 8 and 5, were identified in all of A, B and C groups (Figure 2.2A). Furthermore, motif 3, 4, 1, 7, 2, 8 and 5 were identified in all MAPK genes of *A. hypogaea*. For *A. duranensis*, the combination of conserved motif (motif 4,2,1,6,3,11,8,10,5,7) was identified in nearly all MAPK genes (Figure 2.2B), expect *Du_MAPK15* (missing motif 4), *Du_MAPK10* (missing motif 11), *Du_MAPK2* (missing motif 6 and 3). For *A. ipaensis*, the conserved combination of motif (motif 3, 5, 1, 6, 2, 7, 4) was identified in nearly all MAPK genes expect *Ip_MAPK3* (missing motif 5) (Figure 2.2C). Conserved domain analysis revealed that groups A, B and C contain STKc_TEY_MAPK domain, while group D contains STKc_TDY_MAPK domain. The differences in motif composition between group D and groups A, B, and C could have significant effects on their biological functions. These potential effects need to be investigated further through biological experiments.

Chromosomal distribution and collinearity analysis of putative MAPKs

The majority of MAPK genes were located on the proximate or the distal ends of the chromosomes among three peanut species. In *A. hypogaea,* 14 MAPK genes were present on all chromosomes except chromosome 6, 9 and 10 for A subgenome and 16 MAPK genes were present on all chromosomes except chromosome 16 and 19 for B subgenome. For the two diploid peanut species, the 15 and 16 MAPK genes were present on all chromosomes except chromosome 6, 9, 10 for both *A. duranensis* and *A. ipaensis*. Chromosome 3 holds the most MAPK genes in both diploid species, and this finding also consistent with the At and Bt subgenomes of *A. hypogaea* (Figure 2.3). Collinearity analysis of the MAPK gene family among *A. ipaensis*, *A. duranensis*, and the subgenomes of *A.hypogaea* (At and Bt) shows that most of MAPK genes are largely conserved across these genomes. However, two notable homologous gene pairs located on chromosomes 7 and 8 exhibited divergence between the diploid species *A. duranensis* and *A. ipaensis* (Figure 2.4). This divergence was retained in the At and Bt subgenomes of *A. hypogaea* after hybridization and polyploidization, indicating that structural variations or evolutionary events in these gene pairs occurred prior to the formation of the tetraploid genome.

Estimation of Nonsynonymous (Ka) and Synonymous (Ks) substitutions per site and their ratios

To analyze the selection pressures of MAPK genes on two sub-genomes of allotetraploid peanut *A. hypogaea* and their corresponding two diploid peanut species *A. duranensis and A. ipaens*is, Ka/Ks per site and their ratios were calculated. As illustrated in Figure 2.5, the comparisons include the A subgenome of *A. hypogaea* vs. *A. duranensis* (A vs. At), the B subgenome of *A. hypogaea* and *A. ipaensis* (B vs. Bt), and the internal comparison within *A. hypogaea* between the At and Bt subgenomes (At vs. Bt). The majority of the Ka/Ks ratio comparisons exhibited values below 1 indicating that purifying selection of MAPK genes occurred in both diploid and polyploid peanut species. This pattern suggests a strong conservation of essential genetic functions, critical for maintaining stability and functionality within the plant. However, an exception was observed in the pair *Du_MAPK13* and *Ah_At_MAPK9*, which showed a ratio above 1, as detailed in the supplemental table S2.1. This indicates possible positive selection,

highlighting a potential adaptive genetic evolution in this gene pair. Additionally, we observed that compared to the A vs. At, the MAPK genes in the B vs. Bt are more conserved, with fewer mutations occurring.

Cis-Acting element analysis of putative MAPKs

Cis-acting element analysis was conducted to further explore the potential function transcriptional regulation of the identified MAPKs in peanuts. We extracted 2000 bp upstream sequences of the start codon of identified MAPKs to analyze cis-acting elements in the promoter regions. Cis-element analysis revealed the presence of various potential regulatory elements in the promoter regions of MAPK genes from *A. hypogaea*, its potential progenitor *A. duranensis*, and the other wild relative *A. ipaensis*. A total of 1002, 545 and 531 cis-acting elements were identified in *A. hypogaea*, *A. duranensis*, and *A. ipaensis*, respectively. These cis-acting elements could be classified into three major types, stress responsive (abiotic/biotic), phytohormone responsive and plant growth and developmental (Supplemental Table S2.2 and Figure 2.6). In the first category related to stress responses, 415, 235 and 228 relevant elements were identified in *A. hypogaea*, *A. duranensis*, and *A. ipaensis*, respectively. Notably, the MYB and MYC elements have been previously reported to be involved in drought stress responses (Abe et al. 1997), and the MYB and MYC accounted for more 55% of stress responsive elements across all three species. In the second category of phytohormone-responsive elements, a total of 269, 147 and 149 cis-acting elements were identified in *A. hypogaea*, *A. duranensis*, and *A. ipaensis*, respectively. Within this category, two important stress-responsive cis-regulatory elements ABRE and ERE were identified and they accounted for more 54% of phytohormone-responsive elements across all three species. The third category comprised elements associated with plant growth and development, with 318, 163 and 154 elements detected in *A. hypogaea, A. duranensis*, and *A. ipaensis*. Among these, two crucial light-responsive elements Box-4 and G-Box were found, which accounted for 61% of plant growth and development elements across all species. The occurrence frequencies of these key cis-regulatory elements, such as MYB, ABRE, and Box-4, were highly consistent among the three peanut species. This suggests a high degree of conservation of MAPK genes in peanut species

and their close association with stress responses. These findings lay the foundation for further elucidating the upstream regulatory mechanisms governing MAPK gene expression in peanut.

Expression analysis of MAPK genes response to drought stress in *A. hypogaea*

To investigate the expression patterns of identified MAPK genes under drought stress conditions, we conducted RNA-seq analysis of four peanut genotypes (two drought-tolerant: C76-16 and 587; two drought-susceptible: Tifrunner and 506) under control and drought treatment conditions. PCA of RNA-seq data revealed that drought treatment and genotype-specific responses are the main drivers of transcriptional variation among peanut samples (Figure S2.1). The first principal component (5.6), explaining 29.63% of variance, distinguished drought-treated samples from controls. The second principal component (PC2), accounting for 11.06% of variance, reflected genotypic differences that were distinct under drought conditions, with drought-tolerant genotypes (C76-16 and 587) separating from susceptible genotypes (506 and Tifrunner), whereas no significant genotypic separation under control conditions.

Hierarchical clustering of MAPK gene expression (Figure 2.7) revealed distinct expression patterns in response to drought stress. Notably, *Ah_At_MAPK4* and *Ah_Bt_MAPK4* showed remarkable upregulation under drought conditions, particularly in drought-tolerant genotypes. These genes displayed high expression levels under drought conditions compared to low expression under control conditions, suggesting their potential role in drought response mechanisms.

To quantitatively assess the expression patterns observed in the hierarchical clustering heatmap, we examined the FPKM values of *Ah_At_MAPK4* and *Ah_Bt_MAPK4* across individual biological replicates (Figures 2.8 and 2.9). This quantitative analysis allowed us to determine fold-change values and statistical significance of differential expression between drought-treated and control samples within each genotype, as well as to compare induction levels between drought-tolerant and drought-susceptible genotypes. The results demonstrated that under control conditions, expression levels of these two genes were similar between drought-tolerant

and drought-susceptible genotypes. *Ah_At_MAPK4* showed baseline expression values of 79.82 ± 1.88, 90.20 ± 7.21, 93.21 ± 10.99, and 86.90 ± 15.37 FPKM in 587, C76-16, 506, and Tifrunner, respectively, under control conditions. Similarly, *Ah_Bt_MAPK4* exhibited comparable expression levels across all genotypes under control conditions: 69.92 ± 2.37, 69.16 ± 4.87, 70.42 ± 7.21, and 71.17 ± 13.07 FPKM, respectively. However, under drought stress, both genes showed significant upregulation in all genotypes, with substantially higher induction in drought-tolerant genotypes compared to drought-susceptible ones. *Ah_At_MAPK4* expression increased to 447.74 ± 56.64 and 510.05 ± 49.89 FPKM in drought-tolerant genotypes 587 and C76-16, representing 5.61-fold ($p = 0.0228$) and 5.65-fold ($p = 0.0125$) increases, respectively. In contrast, in drought-susceptible genotypes 506 and Tifrunner, expression levels reached only 239.86 ± 95.62 and 267.75 ± 103.33 FPKM, corresponding to 2.57-fold ($p = 0.2640$) and 3.08-fold ($p = 0.2202$) increases (not statistically significant). Similarly, *Ah_Bt_MAPK4* showed significant upregulation of 6.56-fold ($p = 0.0226$) and 7.17-fold ($p = 0.0108$) in drought-tolerant genotypes 587 and C76-16, with FPKM values of 458.61 ± 59.67 and 495.60 ± 46.15, respectively. In drought-susceptible genotypes 506 and Tifrunner, with FPKM values of 213.30 ± 83.90 and 240.66 ± 93.90, the expression increases were more moderate at 3.03-fold ($p = 0.2300$) and 3.38-fold ($p = 0.2110$).

Combining the results from the heatmap (Figure 2.7) and FPKM analysis (Figures 2.8 and 2.9), we observed that *Ah_At_MAPK4* and *Ah_Bt_MAPK4* exhibit significant drought-induced expression patterns. These homologous genes showed similar expression levels in both drought-tolerant and drought-susceptible genotypes under normal conditions. However, under drought stress, they were upregulated in all genotypes, with significantly higher induction in drought-tolerant genotypes compared to drought-susceptible ones. These findings indicate that these MAPK genes not only respond to drought stress but also exhibit enhanced responsiveness in drought-tolerant materials, suggesting their important role in peanut drought tolerance mechanisms.

WGCNA analysis reveals drought-responsive gene modules

To investigate the regulatory networks associated with *Ah_At_MAPK4* and *Ah_Bt_MAPK4*, WGCNA was conducted. Hierarchical clustering of normalized

expression profiles identified gene modules, grouping genes with similar co-expression patterns. Dynamic tree cutting with a minimum module size of 30 was performed, followed by merging of modules with highly similar eigengenes at a cut height of 0.25, generating three distinct modules colored blue, grey, and turquoise (Figure 2.10A), with the turquoise module encompassing 2099 genes, followed by blue with 1717 genes, and grey containing 253 genes (Figure 2.10B, Table S2.3-S2.5). Soft-thresholding power of 12 was selected to ensure scale-free network topology during module construction (Figure S2.2). Notably, both drought-responsive MAPK genes, *Ah_At_MAPK4* and *Ah_Bt_MAPK4*, were localized within the turquoise module, suggesting this module's important role in mediating drought stress responses. Correlation analysis between module eigengenes and drought-related phenotypic traits revealed the turquoise module as the most strongly positively correlated with the tolerant × drought interaction ($r = 0.89$, $p < 0.001$), indicating its importance in the drought tolerance response. The blue module also showed moderate correlation with drought traits, while the grey module lacked significant associations (Figure 2.10C).

To demonstrate the biological significance of the turquoise module, we performed comprehensive GO and KEGG analysis. GO analysis revealed significant overrepresentation of molecular functions associated with drought stress responses and secondary metabolite biosynthesis (Figure 2.11A and Table S2.6). The turquoise module was enriched in terpene synthase activity, lyase activity, and magnesium ion binding, suggesting roles in terpenoid biosynthesis and cofactor-dependent enzymatic processes related to stress tolerance. KEGG pathway analysis identified enrichment of key metabolic and signaling pathways (Figure 2.11B). Terpenoid backbone biosynthesis was among the enriched pathways, aligning directly with the GO enrichment for terpene synthase activity. In addition, the simultaneous enrichment of cysteine and methionine metabolism alongside key signaling networks such as Ras, cAMP, and cGMP-PKG pathways suggests a coordinated regulatory architecture.

Hub gene network analysis in the target modules

To contextualize *Ah_At_MAPK4* and *Ah_Bt_MAPK4* within the drought-associated turquoise module, 36 hub candidates were selected based on high

connectivity and used to build a co-expression network together with these two MAPK genes (Figure 2.12 and Table S2.7). GO analysis of the hubs identified calmodulin binding as significantly enriched (p = 6.99 × 10^-4; FDR = 0.007), implicating a calcium-dependent signaling as a key component downstream of MAPK activation (Table S2.7). KEGG results revealed enrichment for diterpenoid and gibberellin biosynthetic pathways, consistent with stress-induced secondary metabolism (Table S2.7). Co-expression network analysis showed that *Ah_At_MAPK4* and *Ah_Bt_MAPK4* are strongly connected to multiple hub genes within the turquoise module. The enrichment of calmodulin binding among hub genes highlights the potential importance of calcium-dependent signaling in this drought-responsive network.

Structural modeling of MAPK proteins

To further investigate the functional architecture of the key drought-responsive genes, the 3D protein structures of *Ah_At_MAPK4* and *Ah_Bt_MAPK4* were modeled using the AlphaFold database. The predicted models for both proteins were highly reliable, with average pLDDT scores of 89.5 and 89.69, respectively. Both proteins displayed the canonical bilobed kinase fold characteristic of MAPKs (Figure 2.13). Key functional features were clearly resolved, including an ATP-binding pocket in the cleft between the N- and C-terminal lobes and an accessible activation loop containing the conserved TEY motif. These predicted structural features are consistent with the catalytic and regulatory architecture of MAPKs and provide a structural basis for their role as active signal-transducing kinases in response to drought stress.

2.4 Discussion

In this study, we conducted a whole genome wide identification of MAPK in three peanut species, two wild diploid species *A. duranensis*, *A. ipaensis*, and one cultivated tetraploid specie, *A. hypogaea*. A total of 61 MAPK genes were identified in the three peanut genomes, with 30 in *A. hypogaea*, 16 in *A. duranensis*, and 15 in *A. ipaensis*.

Characterization analysis of the identified MAPK genes revealed differences in parameters such as protein length, molecular weights, and protein isoelectric point among the three species, reflecting the diversification of MAPKs during evolution.

Subcellular localization prediction showed that peanut MAPKs are primarily localized in the cytoplasm, specifically 22 *Ah_MAPK*, 9 *Du_MAPKs*, and 9 *Ip_MAPKs*. This is consistent with their roles in intracellular signal transduction.

We conducted phylogenetic analysis to understand the evolutionary relationship among the identified MAPK genes in peanut species and *A. thaliana*. In plants, MAPK genes have diverged into four subfamilies based on the conserved TEY/TDY motifs in the activation loop region (Ichimura et al. 2002; Yao et al. 2022; Liu et al. 2013). In this study, we classified the 30 *Ah_MAPKs*, 16 *Du_MAPKs*, and 15 *Ip_MAPKs* into four major subfamilies (A, B, C, and D), consistent with the classification of MAPKs in previous reports. In group A, 4 *AhMAPKs*, 1 *IpMAPK*, and 2 *DuMAPKs* clustered together with the well-studied *Arabidopsis AtMAPK3*, *AtMAPK6*, and *AtMAPK10*. *AtMAPK3* and *AtMAPK6* have been shown to be involved in plant responses to biotic stress, such as against Botrytis cinerea (Han et al. 2010), and abiotic stress, like in salt stress (Liu et al. 2020a), while *AtMAPK10* is related to cell division and development. A recent study found that *AtMAPK10* negatively regulates seed growth by inhibiting the transcriptional activity of *WRKY10* during endosperm development in *A. thaliana* (Xi et al. 2021). Therefore, group A may play similar regulatory roles in peanuts. Group B includes 8 *AhMAPKs*, 5 *IpMAPKs*, 5 *DuMAPKs*, and *Arabidopsis AtMAPK4/5/11/12/13*. *AtMAPK4/11* are key players in the formation of the cell plate during cytokinesis (Kosetsu et al. 2010), whereas *AtMAPK12* participates in regulating auxin signaling (Lee et al. 2009). We speculate that group B MAPKs may be involved in peanut growth/development processes. The 4 *AhMAPK*s, 2 *IpMAPKs*, and 2 *DuMAPKs* were classed into group C which closely relates to with *AtMAPK1/2/7/14* that were reported in their response to ABA (de Zelicourt et al. 2016). Thus, this group of MAPKs may have similar functions with group B in peanut development. Group D is the largest MAPK subfamily, containing 14 *AhMAPKs*, 7 *IpMAPKs*, and 7 *DuMAPKs*. Members in this group are highly homologous to *Arabidopsis AtMAPK8/9/15/16/18/19/20*, which are characterized by the presence of the TDY motif in their T-loop and an extended C-terminal region, compared to other MAPK groups (A, B, and C). In conclusion, the peanut MAPK family represents high diversity, with different subfamily members that

potentially playing distinct regulatory roles in plant growth, development, and stress responses. While this study provides a foundation for understanding peanut MAPKs, further research, including gene modification through knockout or overexpression, is crucial to definitively determine the specific functions of individual MAPK family members in peanut.

Cis-acting elements play a crucial role in governing the spatial-temporal patterns of gene expression by serving as the binding sites where transcription factors can bind and implement their regulatory effects (Li et al. 2020). The cis-element analysis unveils the putative cis-regulatory landscape governing MAPK gene expression in peanut. In our results, we identified important stress/phytohormone-responsive cis-elements like MYB, MYC, ABRE and ERE motifs. These elements suggest potential roles for the MAPK genes in mediating plant responses to various environmental stimuli. The prevalence of conserved stress-responsive elements emphasizes the importance of MAPKs in stress adaptation, while the identification of hormone and development-related motifs expands our understanding of their potential regulatory mechanisms. These findings help us for future investigations into the intricate transcriptional networks orchestrating MAPK activity in response to environmental stimuli and developmental programs in peanut.

Our expression analysis revealed that *Ah_At_MAPK4* and *Ah_Bt_MAPK4* are significantly upregulated in response to drought stress, with higher induction in drought-tolerant peanut genotypes compared to drought-susceptible ones. Interestingly, our previous phylogenetic analysis placed these two homologous genes in Group A, clustering with *AtMAPK3, AtMAPK6*, and *AtMAPK10* from *Arabidopsis*. In *Arabidopsis*, Group A MAPKs have been extensively characterized for their critical roles in abiotic stress responses (Ichimura et al. 2002; Taj et al. 2010). Particularly, *AtMAPK3* and *AtMAPK6* function as positive regulators in response to various environmental stresses including drought, salt, cold, and oxidative stress (Moustafa et al. 2014; Liu et al. 2020b). The significant drought-induced expression of *Ah_At_MAPK4* and *Ah_Bt_MAPK4* observed in our study aligns with the known functions of their *Arabidopsis* orthologs, suggesting evolutionary conservation of MAPK signaling in

drought response across plant species. The substantially higher upregulation of these genes in drought-tolerant genotypes (5.61-7.17 fold) compared to drought-susceptible genotypes (2.57-3.38 fold) further indicates their potential involvement in enhanced drought tolerance mechanisms. This differential expression pattern might be attributed to more efficient activation of MAPK cascades in tolerant genotypes, potentially leading to improved regulation of downstream drought-responsive genes and metabolic. To explore this hypothesis further, we employed WGCNA to identify co-expression networks associated with these MAPK genes. Both *Ah_At_MAPK4* and *Ah_Bt_MAPK4* were assigned to the turquoise module, which showed the strongest correlation with drought tolerance traits ($r = 0.89$, $p < 0.001$). Functional enrichment analysis of this module revealed significant overrepresentation of processes central to drought adaptation, including terpene synthase activity, defense responses, and terpenoid biosynthesis pathways. Terpenoids are well-established osmoprotectants that help maintain cellular osmotic balance during water deficit, and their enrichment in the MAPK-associated module suggests these kinases may regulate osmoprotectant production as part of the drought response. Within this module, co-expression network analysis identified 20 hub genes, many of which were annotated with calmodulin binding activity. This finding highlights the potential importance of calcium-dependent signaling, a well-established upstream activator of MAPK cascades, in orchestrating drought-responsive transcriptional programs. Moreover, enrichment of diterpenoid and gibberellin biosynthesis pathways among hub genes suggests that MAPK activation may feed into secondary metabolism and hormone-mediated growth regulation during stress conditions. Consistently, both proteins exhibited highly conserved kinase domains with canonical ATP-binding and activation motifs, consistent with the structural architecture reported for Group A MAPKs in *Arabidopsis*. In conclusion, our findings establish *Ah_At_MAPK4* and *Ah_Bt_MAPK4* as central regulators of drought tolerance in peanut, with conserved structural features that support their catalytic activity and identify them as promising targets for functional validation and genetic improvement. While our integrated multi-omics approach provides compelling evidence for the drought-responsive function of *Ah_At_MAPK4* and *Ah_Bt_MAPK4*, we acknowledge a

limitation of the current study is the lack of qRT-PCR validation for these expression patterns. Future studies should validate these findings through qRT-PCR and conduct functional characterization via gene knockout or overexpression experiments. Additionally, investigating upstream regulatory mechanisms, including miRNA-mediated post-transcriptional regulation, would provide a more comprehensive understanding of MAPK signaling in peanut drought responses.

2.5 Conclusion

In summary, genome-wide analysis of MAPK gene families was first conducted in peanut. We identified 30 MAPK genes in *A. hypogaea*, 16 MAPK genes in *A. duranensis* and 15 MAPK genes in *A. ipaensis*. Phylogenetic analysis clustered all the identified MAPK genes into four distinct groups, which is consistent with the previous studies. RNA-seq expression profiling under drought stress conditions demonstrated that *Ah_At_MAPK4* and *Ah_Bt_MAPK4* were significantly upregulated in response to drought stress, with substantially higher induction in drought-tolerant genotypes compared to drought-susceptible. WGCNA further confirmed their central positions in a drought-associated co-expression module enriched in calcium signaling and secondary metabolic pathways, underscoring their regulatory importance in stress adaptation. In addition, AlphaFold-based structural modeling confirmed that both proteins possess the canonical kinase architecture required for catalytic function. These findings provide valuable information for understanding the MAPK gene family in peanut and identify promising candidates for future research into their roles in stress tolerance and enhancing abiotic stress resilience in peanuts through crop improvement.

Data Availability Statement

The genome and protein sequences of peanut species (*A. hypogaea, A. duranensis, and A. ipaensis*) used in this study were obtained from PeanutBase (https://www.peanutbase.org/). The specific cultivated peanut reference genome was *A. hypogaea* cv. Tifrunner, assembly Tifrunner.gnm2.J5K5. The *Arabidopsis thaliana* MAPK protein sequences were obtained from TAIR (https://www.arabidopsis.org/). The protein kinase domain (PF00069) was downloaded from the Pfam database

(https://www.ebi.ac.uk/interpro/entry/pfam/). The RNA-seq data used for expression analysis under drought stress conditions are publicly available in the NCBI Sequence Read Archive (SRA) database under the accession number PRJNA687542. All other data generated during this study are included in the manuscript and supplementary materials.

List of Figures

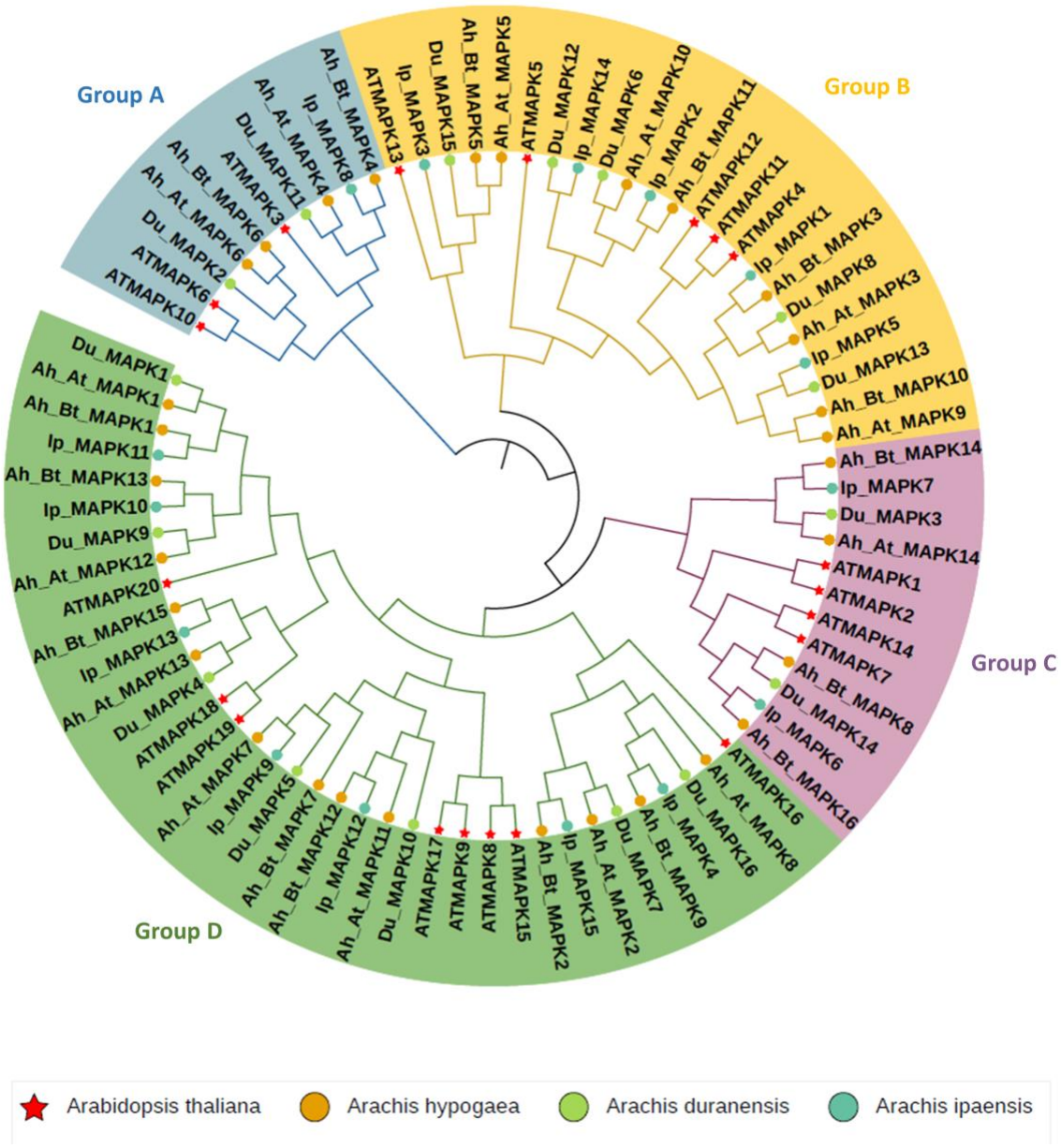

Figure 2.1. Phylogenetic relationship of putative MAPK genes in A. ipaensis, A. duranensis, A. hypogaea, and Arabidopsis thaliana. Different color blocks within the tree denote distinct groups, five-points stars represent. Different shapes preceding the MAPK gene represent different species.

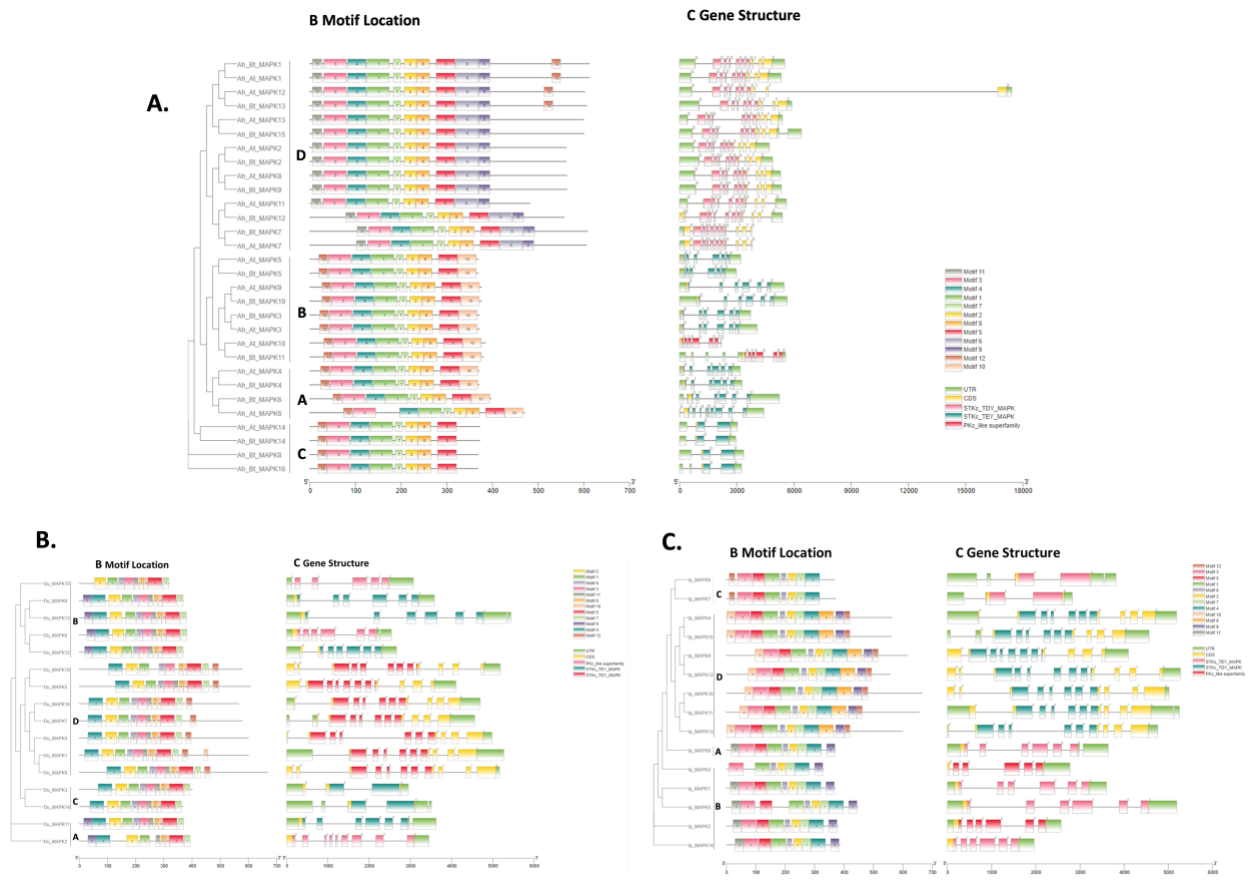

Figure 2.2. (A) Analysis of the structure and conserved motifs of MAPK genes in A. hypogaea; (B) Analysis of the structure and conserved motifs of MAPK genes in A. duranensis; (C) Analysis of the structure and conserved motifs of MAPK genes in A. ipaensis

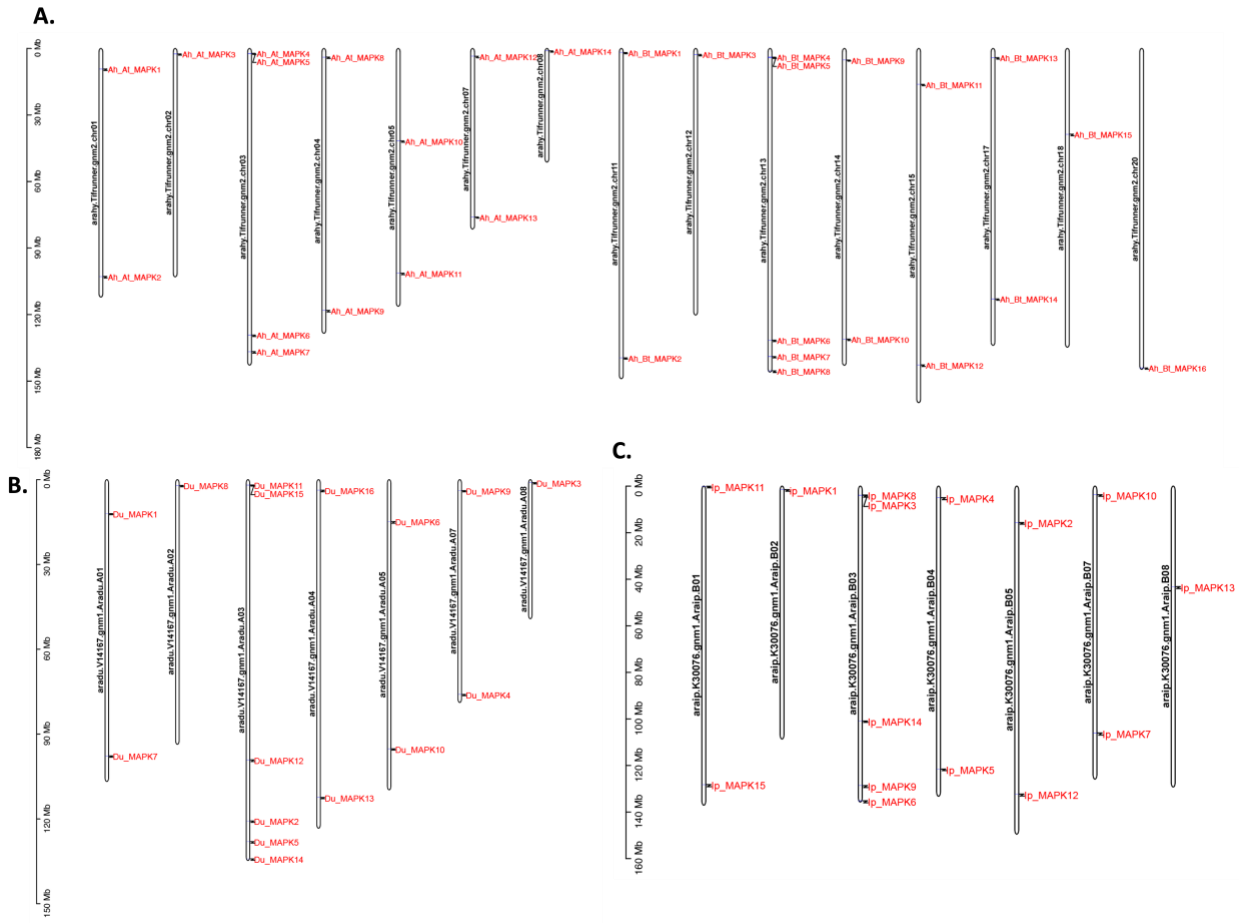

Figure 2.3. (A) Chromosomal distribution of identified MAPK genes in A. hypogaea; (B) Chromosomal distribution of identified MAPK genes in A. duranensis; (C) Chromosomal distribution of identified MAPK genes in A. ipaensis

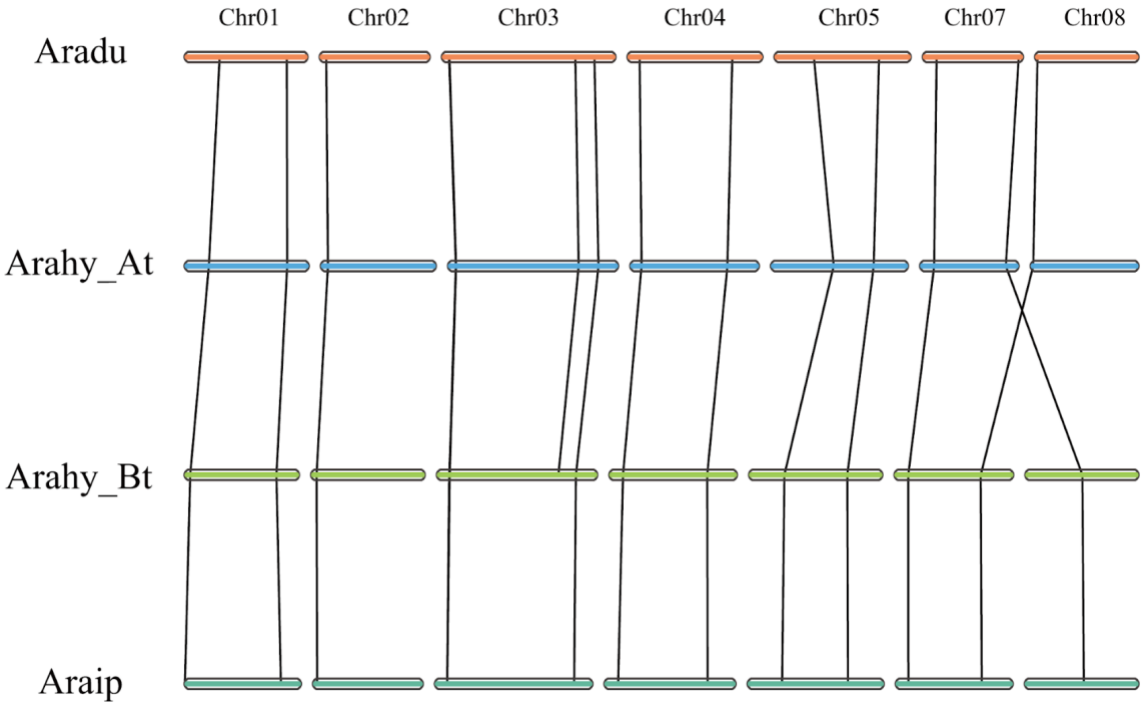

Figure 2.4. Collinearity analysis of identified MAPK genes between A. duranensis, A. ipaensis and A, B subgenomes in A. hypogaea

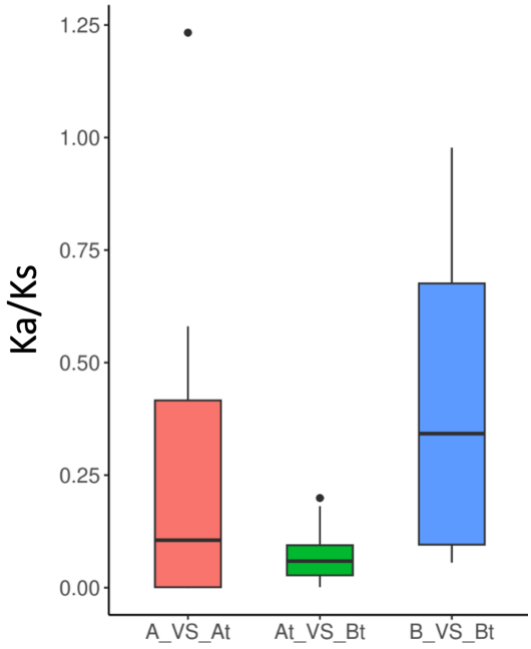

Figure 2.5. Selection pressure analysis. Ka/Ks ratio comparisons represented by box plots in different colors: the red box represents the comparison between the A subgenome of A. hypogaea and A. duranensis, the green box represents the

comparison within the subgenomes (At vs. Bt) of A. hypogaea, and the blue box represents the comparison between the B subgenome of A. hypogaea and A. ipaensis

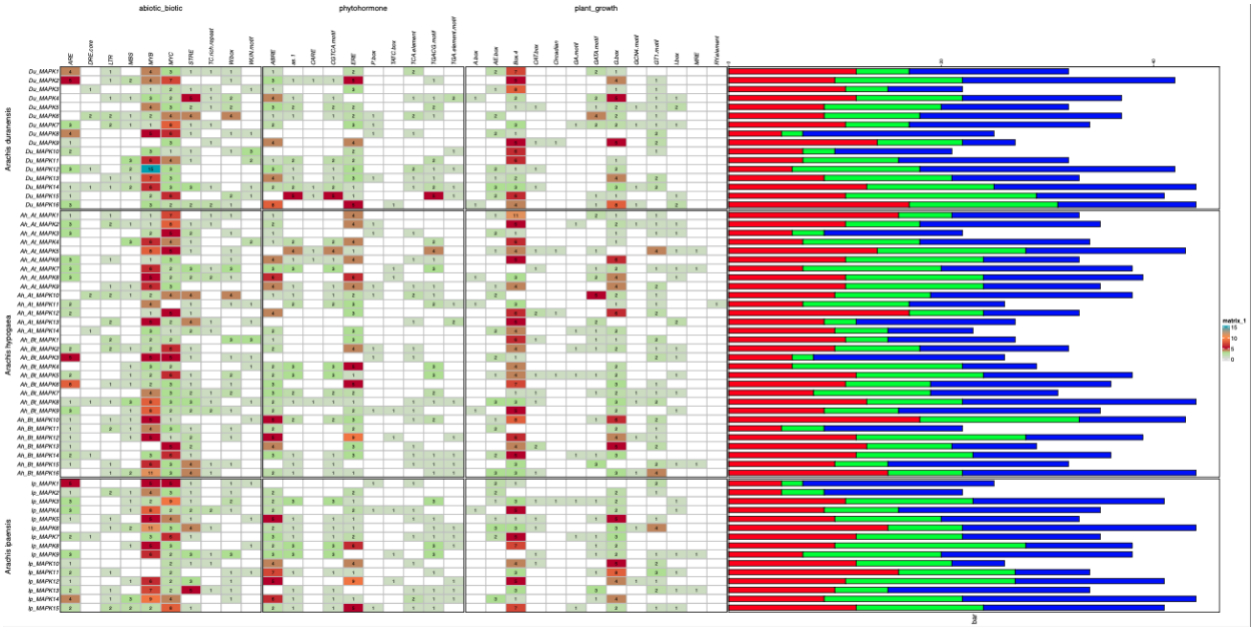

Figure 2.6. Cis-acting element prediction analysis in identified MAPK genes. The heatmap represents the presence (colored boxes) or absence (white boxes) of various cis-acting elements in the promoter regions of MAPK genes. The stacked bar chart on the right depicts the distribution of cis-acting elements classified into three categories: stress-responsive (abiotic/biotic) in blue, plant growth and developmental in red, and phytohormone responsive in green.

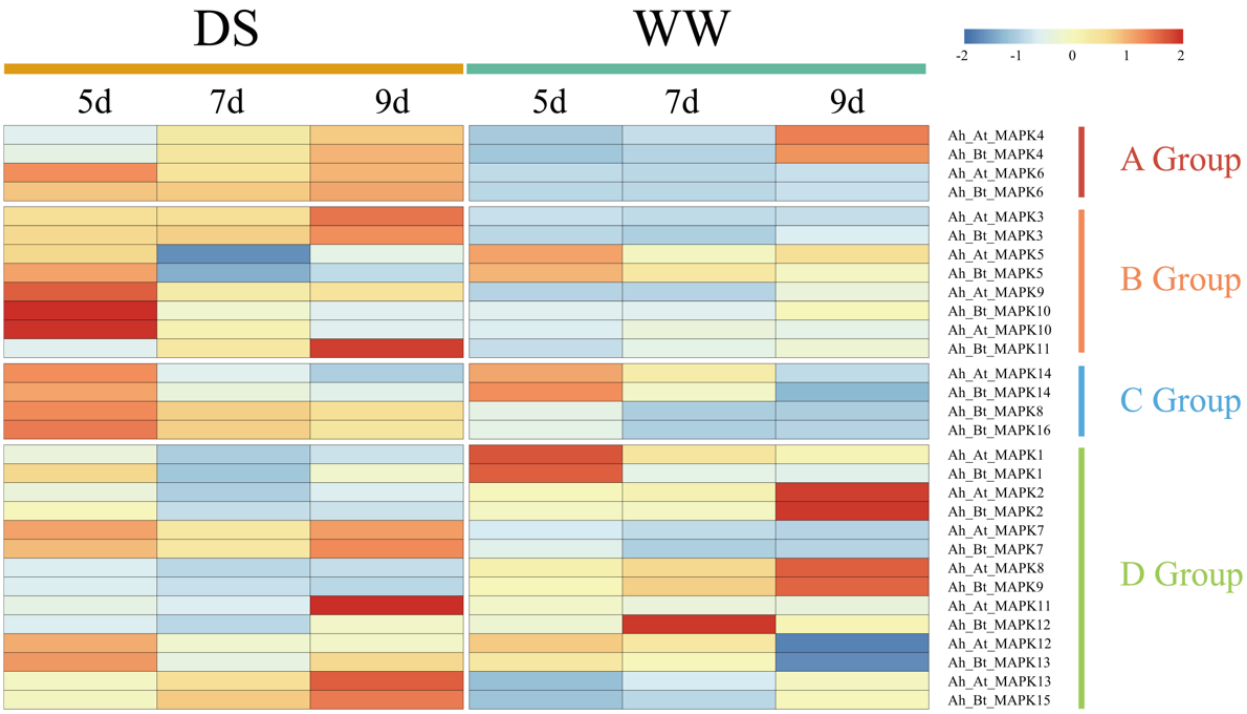

Figure 2.7. Heatmap of MAPK gene expression patterns in drought-tolerant and drought-susceptible peanut genotypes under control and drought conditions.

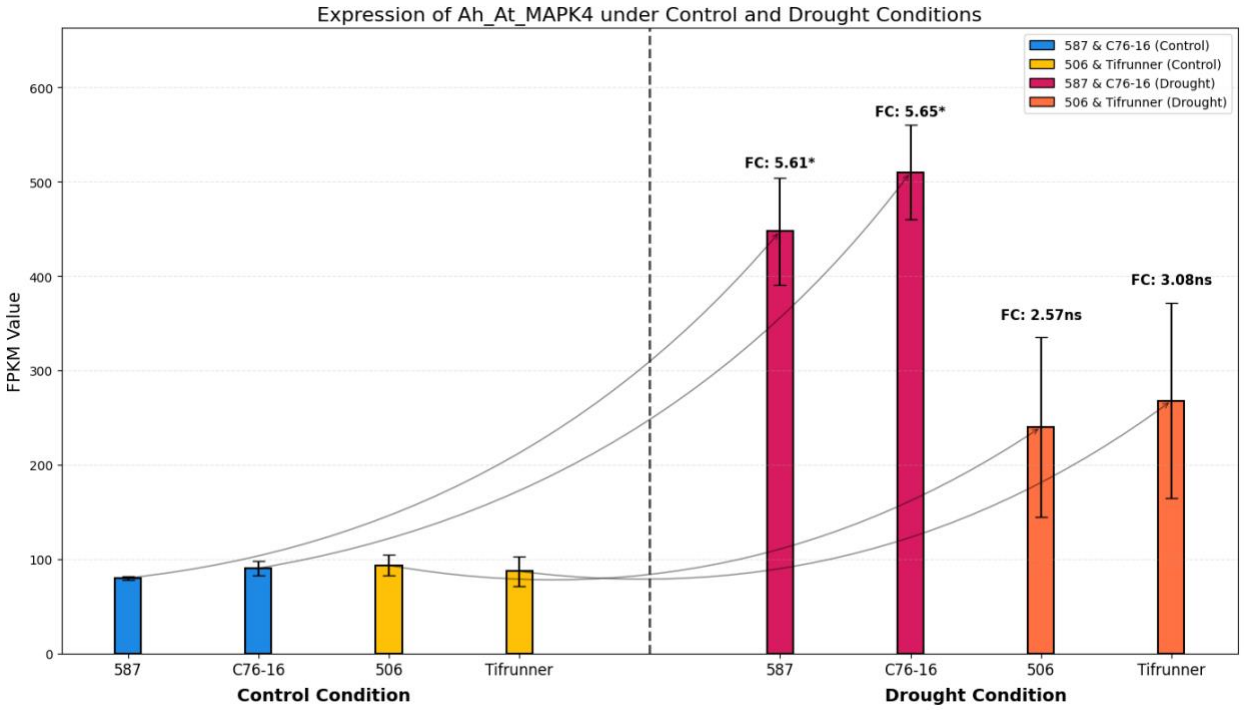

Figure 2.8. Expression analysis of Ah_At_MAPK4 under control and drought conditions. Error bars indicate standard error of the mean (SEM). Fold change (FC) values with

significance levels are shown above each drought treatment bar (* p < 0.05, ** p < 0.01, *** p < 0.001, ns = not significant).

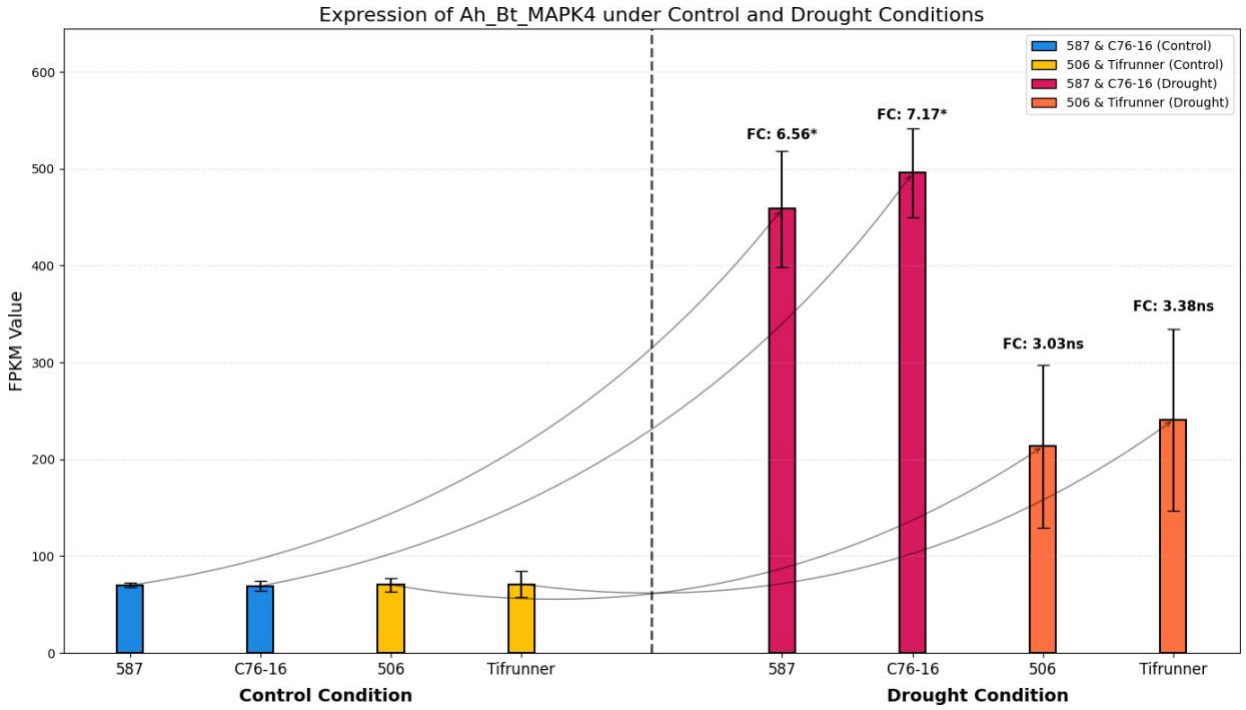

Figure 2.9. Expression analysis of Ah_Bt_MAPK4 under control and drought conditions. Error bars indicate standard error of the mean (SEM). Fold change (FC) values with significance levels are shown above each drought treatment bar (* p < 0.05, ** p < 0.01, *** p < 0.001, ns = not significant).

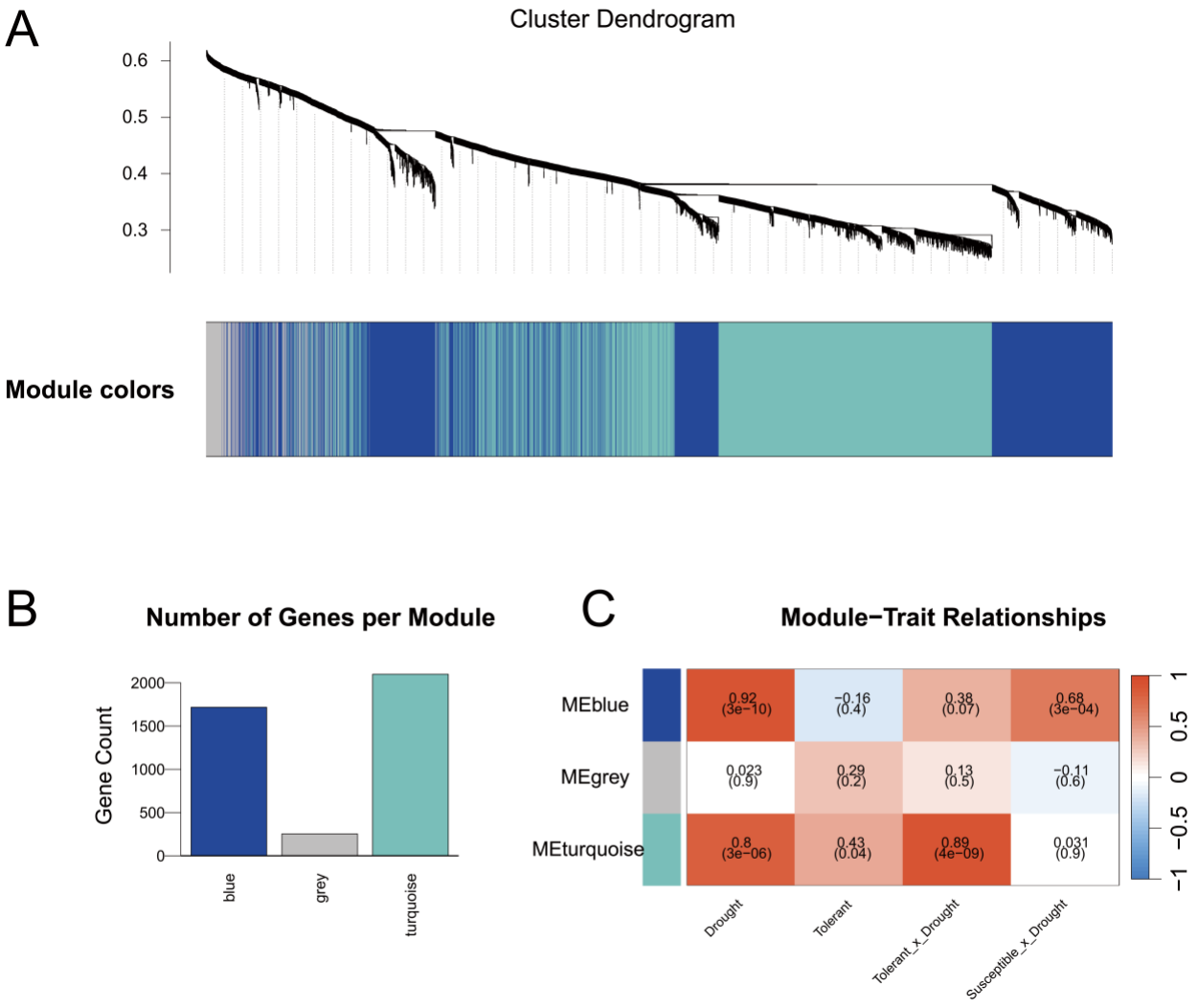

Figure 2.10. Weighted gene co-expression network analysis in drought-stress response of peanut leaves. (A) Hierarchical clustering dendrogram and module colors of DEGs and MAPKs. (B) Distribution of gene counts in modules (C) Module-trait relationship heatmap of 3 modules

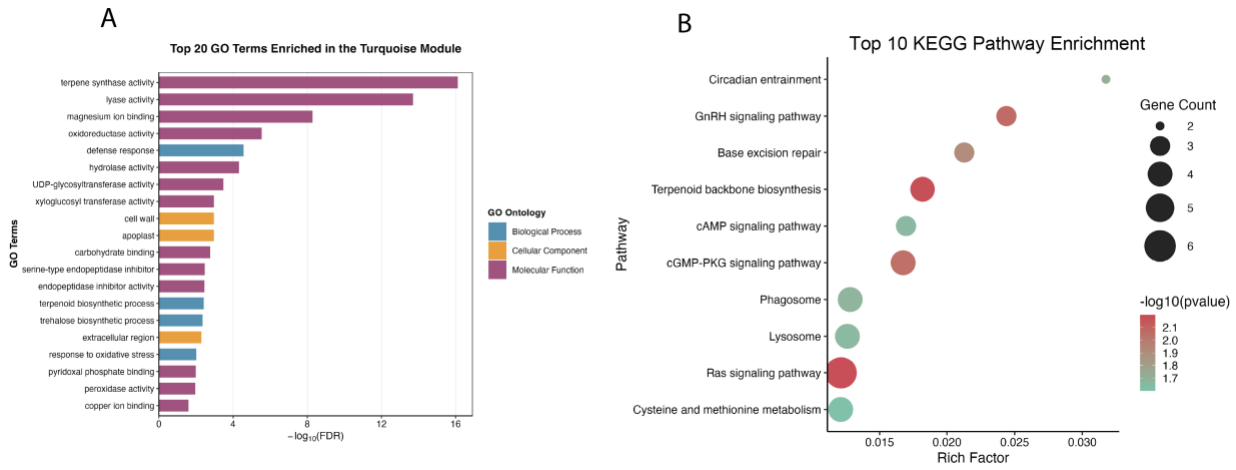

Figure 2.11. Functional enrichment analysis of genes in the turquoise module. (A) Top 20 GO terms enriched in the turquoise module. (B) Top 10 KEGG pathways enriched in the turquoise module

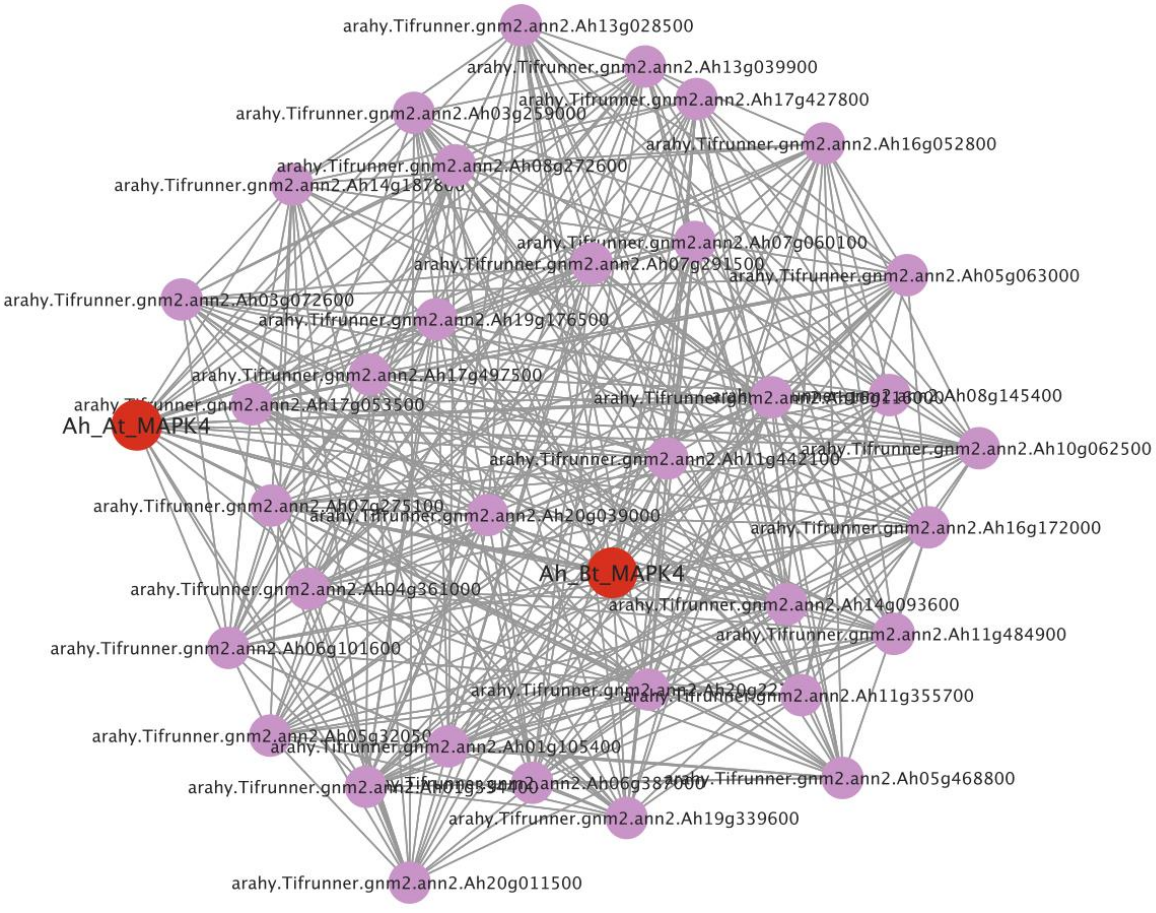

Figure 2.12. Interaction networks of Hub genes in the turquoise module

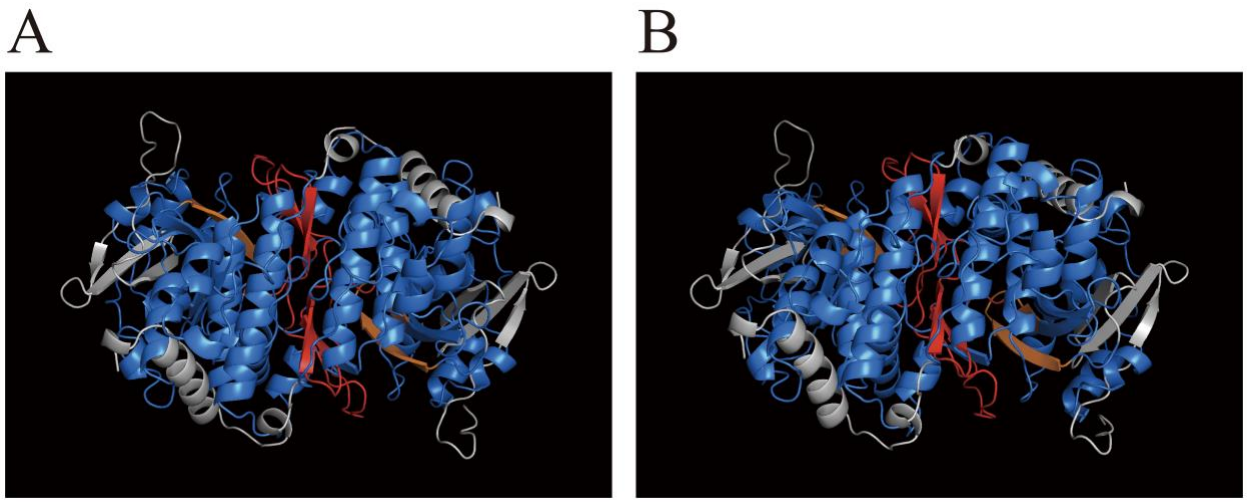

Figure 2.13. Predicted 3D structures of Ah_At_MAPK4 (A) and Ah_Bt_MAPK4 (B). The kinase domain is shown in marine blue, the predicted ATP-binding site in orange, and the activation loop containing the TEY motif in red

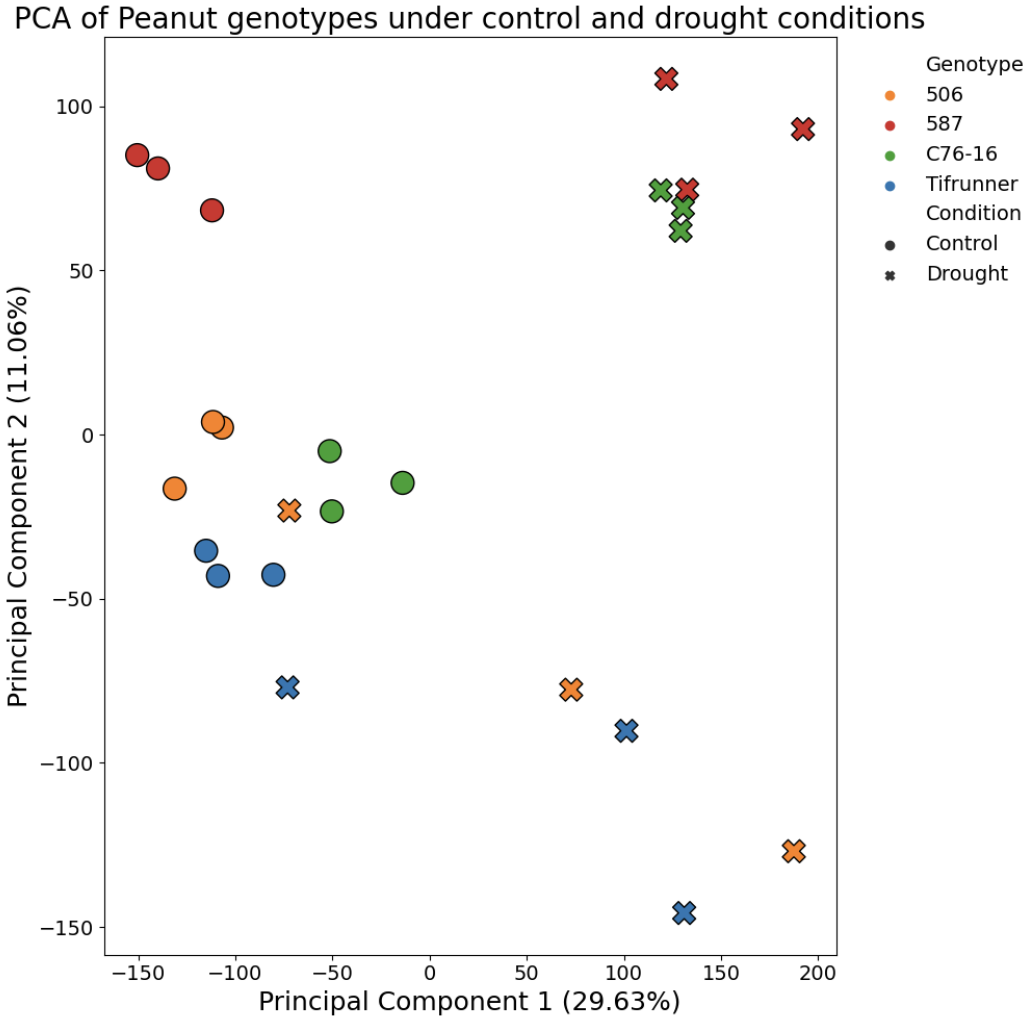

Figure S2.1. PCA of RNA-seq transcriptome data from four peanut genotypes (C76-16, 587, Tifrunner, and 506) under control and drought conditions. Each point represents a biological replicate colored by genotype and shaped by treatment (control or drought)

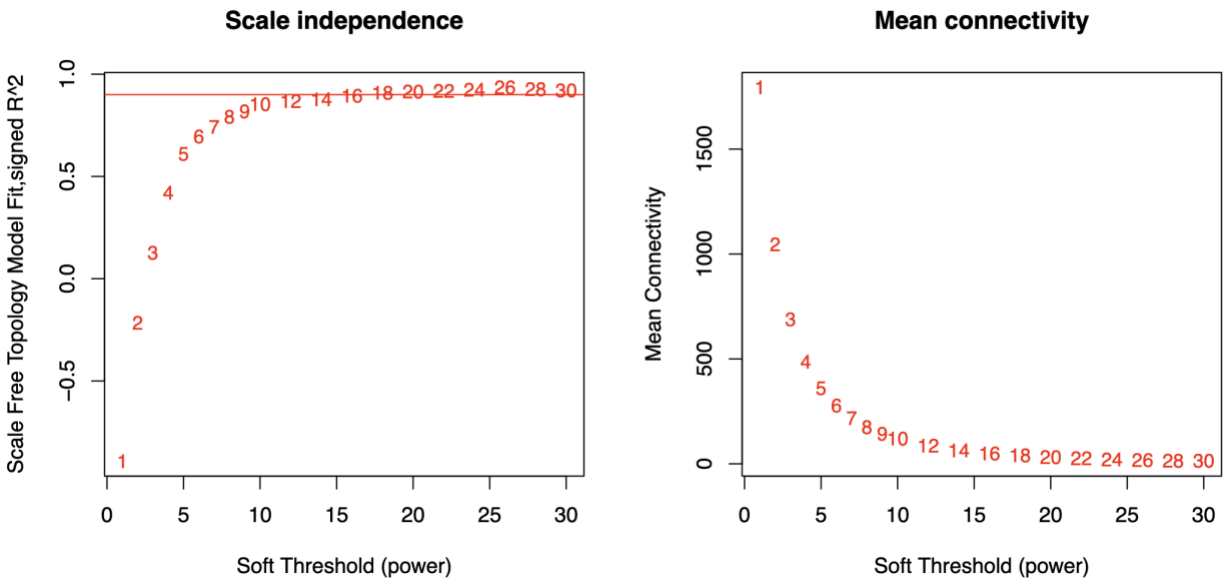

Figure S2.2. Selection of the soft-thresholding power (β) for gene co-expression network construction

List of Tables

Table 2.1. The characteristics of putative MAPK genes in A. hypogaea, A. duranensis and A. ipaensis

Table S2.1. MAPK_Kaks

Table S2.2. cis-acting element analysis of MAPK genes in peanut

Tables S2.3–S2.5. Genes assigned to the turquoise, blue, and grey modules identified by WGCNA.

Table S2.6. GO and KEGG pathway enrichment results for genes in the turquoise co-expression module identified by WGCNA

Table S2.7. GO and KEGG pathway enrichment results for hubs

References

Abe, H., K. Yamaguchi-Shinozaki, T. Urao, T. Iwasaki, D. Hosokawa et al., 1997 Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9 (10):1859-1868.

Ambawat, S., P. Sharma, N.R. Yadav, and R.C. Yadav, 2013 MYB transcription factor genes as regulators for plant responses: an overview. Physiology and Molecular Biology of Plants 19 (3):307-321.

Arya, S.S., A.R. Salve, and S. Chauhan, 2016 Peanuts as functional food: a review. Journal of Food Science and Technology-Mysore 53 (1):31-41.

Bailey, T.L., J. Johnson, C.E. Grant, and W.S. Noble, 2015 The MEME Suite. Nucleic Acids Res 43 (W1):W39-49.

Bertioli, D.J., J. Jenkins, J. Clevenger, O. Dudchenko, D. Gao et al., 2019 The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nat Genet 51 (5):877-884.

Carter, J.W., 2015. Middle-Season Drought Tolerance in a RIL Population of Cultivated Peanut. M.S. thesis. Auburn Univ., Auburn, AL.

Chen, C., H. Chen, Y. Zhang, H.R. Thomas, M.H. Frank et al., 2020a TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant 13 (8):1194-1202.

Chen, L., H. Sun, F. Wang, D. Yue, X. Shen et al., 2020b Genome-wide identification of MAPK cascade genes reveals the GhMAP3K14-GhMKK11-GhMPK31 pathway is involved in the drought response in cotton. Plant Mol Biol 103 (1-2):211-223.

Colcombet, J., and H. Hirt, 2008 Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. Biochem J 413 (2):217-226.

Danquah, A., A. de Zelicourt, J. Colcombet, and H. Hirt, 2014 The role of ABA and MAPK signaling pathways in plant abiotic stress responses. Biotechnol Adv 32(1):40-52.

de Zelicourt, A., J. Colcombet, and H. Hirt, 2016 The Role of MAPK Modules and ABA during Abiotic Stress Signaling. Trends Plant Sci 21 (8):677-685.

Edgar, R.C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32 (5):1792-1797.

Group, M., 2002 Mitogen-activated protein kinase cascades in plants: a new nomenclature. Trends Plant Sci 7 (7):301-308.

Guindon, S., J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk et al., 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59 (3):307-321.

Hamel, L.P., M.C. Nicole, S. Sritubtim, M.J. Morency, M. Ellis et al., 2006 Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. Trends Plant Sci 11 (4):192-198.

Han, L., G.J. Li, K.Y. Yang, G. Mao, R. Wang et al., 2010 Mitogen-activated protein kinase 3 and 6 regulate Botrytis cinerea-induced ethylene production in Arabidopsis. Plant J 64 (1):114-127.

He, X.W., C.Z. Wang, H.B. Wang, L.G. Li, and C. Wang, 2020 The Function of MAPK Cascades in Response to Various Stresses in Horticultural Plants. Frontiers in Plant Science 11.

Horton, P., K.J. Park, T. Obayashi, N. Fujita, H. Harada et al., 2007 WoLF PSORT: protein localization predictor. Nucleic Acids Res 35 (Web Server issue):W585-587.

Ichimura, K., K. Shinozaki, G. Tena, J. Sheen, Y. Henry et al., 2002 Mitogen-activated protein kinase cascades in plants: a new nomenclature. Trends in Plant Science 7 (7):301-308.

Jimenez-Sanchez, M., V.J. Cid, and M. Molina, 2007 Retrophosphorylation of Mkk1 and Mkk2 MAPKKs by the Slt2 MAPK in the yeast cell integrity pathway. J Biol Chem 282 (43):31174-31185.

Kim, D., B. Langmead, and S.L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12 (4):357-360.

Kim, H.S., S.C. Park, C.Y. Ji, S. Park, J.C. Jeong et al., 2016 Molecular characterization of biotic and abiotic stress-responsive MAP kinase genes, IbMPK3 and IbMPK6, in sweetpotato. Plant Physiol Biochem 108:37-48.

Kong, Q., N. Qu, M. Gao, Z. Zhang, X. Ding et al., 2012 The MEKK1-MKK1/MKK2-MPK4 kinase cascade negatively regulates immunity mediated by a mitogen-activated protein kinase kinase kinase in Arabidopsis. Plant Cell 24 (5):2225-2236.

Kosetsu, K., S. Matsunaga, H. Nakagami, J. Colcombet, M. Sasabe et al., 2010 The MAP kinase MPK4 is required for cytokinesis in Arabidopsis thaliana. Plant Cell 22 (11):3778-3790.

Krasensky, J., and C. Jonak, 2012 Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J Exp Bot 63 (4):1593-1608.

Kumar, K.R., and P.B. Kirti, 2010 A mitogen-activated protein kinase, AhMPK6 from peanut localizes to the nucleus and also induces defense responses upon transient expression in tobacco. Plant Physiol Biochem 48 (6):481-486.

Kumar, K.R., T. Srinivasan, and P.B. Kirti, 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 (1):65-81.

Lee, J.S., S. Wang, S. Sritubtim, J.G. Chen, and B.E. Ellis, 2009 Arabidopsis mitogen-activated protein kinase MPK12 interacts with the MAPK phosphatase IBR5 and regulates auxin signaling. Plant J 57 (6):975-985.

Li, R., F. Zhu, and D. Duan, 2020 Function analysis and stress-mediated cis-element identification in the promoter region of VqMYB15. Plant Signaling & Behavior 15 (7):1773664.

Liu, J., X. Wang, L. Yang, W. Nan, M. Ruan et al., 2020a Involvement of active MKK9-MAPK3/MAPK6 in increasing respiration in salt-treated Arabidopsis callus. Protoplasma 257 (3):965-977.

Liu, J., X.M. Wang, L. Yang, W.B. Nan, M.J. Ruan et al., 2020b Involvement of active MKK9-MAPK3/MAPK6 in increasing respiration in salt-treated callus. Protoplasma 257 (3):965-977.

Liu, Y.K., D. Zhang, L. Wang, and D.Q. Li, 2013 Genome-Wide Analysis of Mitogen-Activated Protein Kinase Gene Family in Maize. Plant Molecular Biology Reporter 31 (6):1446-1460.

Ma, H., J. Chen, Z. Zhang, L. Ma, Z. Yang et al., 2017 MAPK kinase 10.2 promotes disease resistance and drought tolerance by activating different MAPKs in rice. The Plant Journal 92 (4):557-570.

Manna, M., B. Rengasamy, and A.K. Sinha, 2023 Revisiting the role of MAPK signalling pathway in plants and its manipulation for crop improvement. Plant Cell Environ 46 (8):2277-2295.

McGinnis, S., and T.L. Madden, 2004 BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research 32:W20-W25.

Meng, X., and S. Zhang, 2013 MAPK cascades in plant disease resistance signaling. Annu Rev Phytopathol 51:245-266.

Mistry, J., S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar et al., 2021 Pfam: The protein families database in 2021. Nucleic Acids Res 49 (D1):D412-D419.

Moustafa, K., S. AbuQamar, M. Jarrar, A.J. Al-Rajab, and J. Trémouillaux-Guiller, 2014 MAPK cascades and major abiotic stresses. Plant Cell Reports 33 (8):1217-1225.

Nawaz, M., J. Sun, S. Shabbir, W.A. Khattak, G. Ren et al., 2023 A review of plants strategies to resist biotic and abiotic environmental stressors. Sci Total Environ 900:165832.

Puppala, N., S.N. Nayak, A. Sanz-Saez, C. Chen, M.J. Devi et al., 2023 Sustaining yield and nutritional quality of peanuts in harsh environments: Physiological and molecular basis of drought and heat stress tolerance. Front Genet 14:1121462.

Schaffer, A.A., L. Aravind, T.L. Madden, S. Shavirin, J.L. Spouge et al., 2001 Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res 29 (14):2994-3005.

Sewelam, N., K. Kazan, and P.M. Schenk, 2016 Global Plant Stress Signaling: Reactive Oxygen Species at the Cross-Road. Frontiers in Plant Science 7.

Shou, H., P. Bordallo, and K. Wang, 2004 Expression of the Nicotiana protein kinase (NPK1) enhanced drought tolerance in transgenic maize. J Exp Bot 55(399):1013-1019.

Suzuki, N., R.M. Rivero, V. Shulaev, E. Blumwald, and R. Mittler, 2014 Abiotic and biotic stress combinations. New Phytol 203 (1):32-43.

Taj, G., P. Agarwal, M. Grant, and A. Kumar, 2010 MAPK machinery in plants Recognition and response to different stresses through multiple signal transduction pathways. Plant Signaling & Behavior 5 (11):1370-1378.

Wang, D., Y. Zhang, Z. Zhang, J. Zhu, and J. Yu, 2010 KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics 8 (1):77-80.

Wang, G., T. Wang, Z.H. Jia, J.P. Xuan, D.L. Pan et al., 2018 Genome-Wide Bioinformatics Analysis of MAPK Gene Family in Kiwifruit (Actinidia Chinensis). Int J Mol Sci 19 (9).

Wang, T., M. Liu, Y. Wu, Y. Tian, Y. Han et al., 2022 Genome-Wide Identification and Expression Analysis of MAPK Gene Family in Lettuce (Lactuca sativa L.) and Functional Analysis of LsMAPK4 in High- Temperature-Induced Bolting. Int J Mol Sci 23 (19).

Wang, X., X.L. Yang, Y.C. Feng, P. Dang, W.W. Wang et al., 2021 Transcriptome Profile Reveals Drought-Induced Genes Preferentially Expressed in Response to Water Deficit in Cultivated Peanut (L.). Frontiers in Plant Science 12.

Wang, Y., H. Tang, J.D. Debarry, X. Tan, J. Li et al., 2012 MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40 (7):e49.

Wu, J., J. Wang, C. Pan, X. Guan, Y. Wang et al., 2014 Genome-wide identification of MAPKK and MAPKKK gene families in tomato and transcriptional profiling analysis during development and stress response. PLoS One 9 (7):e103032.

Wu, W., J. Yang, N. Yu, R. Li, Z. Yuan et al., 2024 Evolution of the WRKY Family in Angiosperms and Functional Diversity under Environmental Stress. Int J Mol Sci 25 (6).

Xi, X., Z. Hu, X. Nie, M. Meng, H. Xu et al., 2021 Cross Inhibition of MPK10 and WRKY10 Participating in the Growth of Endosperm in Arabidopsis thaliana. Frontiers in Plant Science 12:640346.

Xu, J., and S. Zhang, 2015 Mitogen-activated protein kinase cascades in signaling plant growth and development. Trends Plant Sci 20 (1):56-64.

Xu, R., P. Duan, H. Yu, Z. Zhou, B. Zhang et al., 2018 Control of Grain Size and Weight by the OsMKKK10-OsMKK4-OsMAPK6 Signaling Pathway in Rice. Mol Plant 11 (6):860-873.

Yang, X., Y. Li, H. Chen, J. Huang, Y. Zhang et al., 2020 Photosynthetic Response Mechanism of Soil Salinity-Induced Cross-Tolerance to Subsequent Drought Stress in Tomato Plants. Plants (Basel) 9 (3).

Yao, Y., H. Zhao, L. Sun, W. Wu, C. Li et al., 2022 Genome-wide identification of MAPK gene family members in Fagopyrum tataricum and their expression during development and stress responses. Bmc Genomics 23 (1):96.

Zhang, L., Y. Xi, Y. Kong, D. Guo, L. Li, et al., 2018 A cotton group C MAP kinase gene, GhMPK2, positively regulates salt and drought tolerance in tobacco. Plant Mol Biol 97(1-2):17-31.

Zhang, M., and S. Zhang, 2022 Mitogen-activated protein kinase cascades in plant signaling. J Integr Plant Biol 64 (2):301-341.

Zhao, N., S. Cui, X. Li, B. Liu, H. Deng et al., 2021 Transcriptome and Co-expression Network Analyses Reveal Differential Gene Expression and Pathways in Response to Severe Drought Stress in Peanut (Arachis hypogaea L.). Front Genet 12:672884.

Zhu, J.K., 2016 Abiotic Stress Signaling and Responses in Plants. Cell 167 (2):313-324.

Chapter 3: Whole Genome-wide Association Study Reveals Genetic Insights into Leaf Spot Disease Resistances and Seed Germination/Dormancy in Peanut

Abstract

Peanut (*Arachis hypogaea* L.) is an important crop in the world, serving as a key source of edible oil and protein. Comprehensive genomic and phenotypic analyses were conducted on 87 accessions from the U.S. peanut mini-core collection using 217 Gb of high-quality resequencing data to identify the candidate genes and markers that underlie the leaf spot resistance and seed dormancy in peanuts. A total of 87,726 SNPs were identified and mapped across 20 chromosomes, revealing a higher SNP density in the B subgenome (35.55 SNPs/Mb) compared to the A subgenome (33.26 SNPs/Mb). Phylogenetic, population structure, and principal component analyses consistently partitioned the accessions into three distinct gene pools designated as Group 1, 2, and 3. Group 1, comprising primarily *Arachis hypogaea*, included 28 genotypes; Group 2, mainly *fastigiata* types, comprised 18 accessions; while Group 3, displaying the highest diversity, contained mixed genotypes from the other groups. Linkage disequilibrium analysis indicated an LD decay distance of approximately 63.1 kb, confirming that the marker density was sufficient for GWAS. Significant SNP associations at a suggestive threshold of $p < 1.14 \times 10^{-5}$ were identified for leaf spot, seed germination and dormancy agronomic traits. As a result, three candidate genes were identified: *Ah11g381400*, homologous to *Arabidopsis* ATE1, was associated with early leaf spot resistance; *Ah16g445600*, a homolog of ERF34, was linked to late leaf spot resistance; and *Ah19g214100*, homologous to ICE1, emerged as a central regulator affecting both germination and dormancy. These findings provide actionable targets for marker-assisted selection to enhance disease resilience and seed quality in breeding programs.

Keywords: *Arachis hypogaea L.*; genome wide association study (GWAS); leaf spot resistance; seed dormancy; seed germination

3.1 Introduction

Peanut is an important oil and protein crop that originated in South America. Although the genus *Arachis* contains about 80 species, only one species, *Arachis*

hypogaea has been domesticated as cultivated peanut. It is now widely grown, particularly in China, India, African countries, and the United States (X. Zhao et al., 2011). The cultivated peanut is an allotetraploid with 20 chromosome pairs (2n = 4x = 40), arising from the genome doubling of an interspecific hybrid within the section *Arachis*, which resulted from the crossbreeding of two older diploid ancestor species *Arachis duranensis* and *Arachis ipaensis* genomes (Chen et al. 2019). For its high nutritional value, *Arachis hypogaea* is extensively utilized in food products, oils, and animal feed (X. Zhao et al., 2011). Studies have shown that peanut kernels contain approximately 25% crude protein and 46% lipid content, making them a rich source of protein and oil for human consumption and industrial applications (Satish Ingale & S. K. Shrivastava, 2011). Additionally, peanut meal, a by-product of oil extraction, is widely used as animal feed due to its high protein content (T. Zhao et al., 2023).

The economic importance *of Arachis hypogaea* has driven extensive research efforts to understand its origin, domestication and genetic architecture, particularly in yield, quality and resistance to biotic and abiotic stresses. Large-scale genomic resources, including variation mapping, evolutionary analyses, and pangenome studies, have substantially improved the resolution of trait dissection and accelerated locus discovery for agronomic traits related to yield, quality, and stress adaptation (Lu et al. 2024; Guo et al. 2024; Zheng et al. 2024; Zhao et al. 2025). Among these traits, resistance to leaf spot and seed dormancy are of critical importance for sustainable peanut production and quality improvement (Clevenger et al. 2018; Han et al. 2018; Liang et al. 2018; Chu et al. 2019; Kumar et al. 2020; Chavarro et al. 2020; Wang et al. 2022). Understanding the genetic basis of these complex traits is key to improving peanut productivity and resistance of new cultivars.

Leaf spot, caused by fungal pathogens *Cercospora arachidicola* (early leaf spot) and *Cercosporidium personatum* (late leaf spot), is one of the most devastating diseases in peanut cultivation (P. M. Dang et al., 2021), leading to significant yield losses and increased reliance on chemical fungicides. Extensive research has been studied in this field. Earlier studies developed a comprehensive genetic linkage map comprising 248 marker loci, leading to the identification of 48 quantitative trait loci

(QTLs) associated with disease resistance. These included 6 QTLs for tomato spotted wilt virus (TSWV), 22 for early leaf spot (ELS), and 20 for late leaf spot (LLS) resistance (Khera et al., 2016). More recently, three consistent ELS resistance QTLs (*qELS.A03_1.1, qELS.A03_1.2, and qELS.B03*) were identified, with *qELS.A03_1.1* corresponds to a previously reported LLS resistance region in GPBD4 (Chu et al., 2019). Additionally, major QTLs associated with resistance to LLS have been mapped on chromosome A03, including a common quantitative trait locus (QTL) shared with rust disease (LLS_{QTL1}/$Rust_{QTL}$) (Ahmad et al., 2020).

Seed dormancy is an important trait in crop domestication, playing a key role in preventing pre-harvest sprouting (PHS) and minimizing yield losses in various crops (Bewley & Black, 1985; Nonogaki et al., 2018; Shelar et al., 2014). Research has demonstrated significant variability in seed dormancy among different peanut accessions and botanical varieties. The botanical variety *hypogaea* exhibited stronger dormancy compared to other botanical varieties (Silva et al., 2017; M. L. Wang et al., 2012). Further study showed that seed dormancy in peanut has been linked to two major QTLs on chromosomes A04 and A05, with a candidate gene identified (*Arahy.KB746A*, ethylene-responsive transcription factor), providing opportunities for marker-assisted breeding to improve dormancy and reduce preharvest sprouting (Wang et al. 2022). Our previous study identified a significant QTL on chromosome A05, explaining 20% of the phenotypic variation in seed dormancy (Patel et al., 2022). A recent study using multi-locus and single-locus GWAS analysis with the groundnut mini-core collection and 58K SNP array data identified 47 significant SNP-trait associations (9 from ML-GWAS and 38 from SL-GWAS), including a validated QTL on chromosome B02 (*qFSD-B02-1*) and 9 key candidate genes regulating ABA/GA pathways for fresh seed dormancy improvement (Bomireddy et al., 2024).

Genome-wide association study (GWAS) is a powerful tool that employs linkage disequilibrium (LD) to identify genetic variants associated with phenotypic traits in large populations, enabling the detection of loci contributing to specific traits (Mackay & Powell, 2007). In peanut, most previous studies have focused on mapping quantitative trait loci associated with yield-related parameters, including seed weight, seed size, pod

length, pod weight, and other productivity-related traits (W. Chen et al., 2016; Shirasawa et al., 2012b; S. Zhang et al., 2019). Building on this foundation, recent high-impact studies have substantially advanced peanut GWAS resources. A large-scale genomic variation map constructed from 390 accessions identified 22,309 significant associations across 28 agronomic traits, including loci governing plant architecture and oil biosynthesis (Lu et al., 2024). Combining chloroplast and whole-genome sequencing with linkage mapping and GWAS revealed candidate genes and genomic regions associated with flowering pattern, pod and seed weight, and oil content, while clarifying evolutionary diversification within cultivated peanut (Zheng et al., 2024). Alongside these comprehensive genomic analyses, targeted studies continue to find important regions for key traits. For instance, recent work has successfully identified five loci on chromosome B06 influencing yield components, and two loci on chromosome A08 related to protein and oil content (Guo et al. 2024).

In our previous leaf spot resistance study, 46 QTLs were identified with phenotypic variation explained (PVE) from 10.19 to 24.11%, including 18 QTLs for resistance to ELS and 28 QTLs for LLS (Zhang et al. 2020). Additionally, using the same genetic diversity panel, we also investigated seed germination and dormancy traits and revealed genomic regions associated with these critical seed characteristics (Patel et al. 2022). However, these previous studies were conducted using SNP array-based genotyping platforms, which provided limited genomic resolution for precise mapping of trait-associated regions and more accurate identification of candidate genes.

To further reveal the genetic architecture in yield, quality and resistance to biotic and abiotic stresses, we performed whole genome association study for 2 disease-related traits, and 4 seed germination and dormancy related traits in 87 peanut accessions from the U.S. mini-core collection and identified 87,726 SNPs. Our objectives were to: (1) investigate the population structure and genetic diversity, (2) identify genomic regions associated with multiple important traits, including early and late leaf spot resistance and seed dormancy related traits through GWAS analysis, and (3) determine candidate genes within significant loci and analyze their haplotype

variations to provide insights into the genetic architecture of these agronomically important traits.

3.2 Methods

Phenotype evaluation

In this study, 87 accessions of the US mini-core collection were phenotyped. These genotypes were comprised of three botanical varieties: *hypogaea, fastigiata, vulgaris* and *peruviana* (Table S3.1) (Holbrook & Dong, 2005). The geographic origins of these accessions are diverse, such as the United States, China, India, Nigeria, Argentina, Brazil, South Africa, Sudan, and Zambia, which are the main peanut-growing areas worldwide.

The evaluation procedures for Early Leaf Spot (ELS) and Late Leaf Spot (LLS) phenotypes were conducted as described in our previous study (H. Zhang et al., 2020). ELS assessments were carried out in 2013 and 2014 at the E.V. Smith Research Center of Auburn University, located in Shorter, AL (32°29′N, 85°53′W). LLS evaluations were conducted in 2013, 2014, and 2015 at the Wiregrass Research and Extension Center of Auburn University in Headland, AL (31°22′N, 85°19′W). For both diseases, severity was visually assessed on a plot basis using the Florida scale, which ranges from 1 (no disease) to 10 (plant death) (Chiteka et al., 1988).

The procedures for evaluating seed dormancy, germination, and grading followed previously published protocols (Patel et al., 2022). In 2010, peanut seeds were planted in Dawson, GA, USA, using a randomized complete block design with three replications in two-row, 10-foot plots. Irrigation ensured optimal soil moisture, and crop management adhered to standard practices, including nutrient application (N:P:K = 120:26:33 kg ha−1) and pest control. Mature peanut pods were harvested, and pods were dried to 10% moisture for further evaluation. For germination testing, fungicide-treated seeds were evaluated using duplicate sets of 25 seeds on moistened germination paper in controlled growth chambers (30 °C for 12 h with 8 h light, 20 °C for 12 h dark, 85% relative humidity). Germination was recorded at 7, 14, and 21 days, with non-germinated and decayed seeds classified as non-viable.

DNA isolation, genome resequencing and SNP detection

For each peanut accession, we collected young leaf tissue for DNA extraction using the CTAB method, and we constructed paired-end whole genome sequencing libraries with insert sizes ranging from 300 bp to 500 bp. The libraries were sequenced using the Illumina NovaSeq 6000 platform, generating a total of 217 Gb of high-quality sequences, with approximately 2.5 Gb per sample. *The Arachis hypogaea* reference genome was utilized for mapping (Bertioli et al., 2019a), and genomic variations were identified using the Khufu platform from HudsonAlpha institute.

Phylogenetic tree and population structure

To clarify the phylogenetic relationship of mini-core accessions, we used VCF2Dis (https://github.com/BGI-shenzhen/VCF2Dis) to calculate P-distance matrix. Subsequently, FastMe2.0 (Lefort et al., 2015) used to infer the phylogenetic relationship between individuals using the Neighbor-Joining (NJ) method. Finally, the phylogenetic tree was visualized by iTOL (https://itol.embl.de/). Population structure was calculated with Admixture, and the number of assumed genetic clusters (K) ranged from 1 to 10, with 10,000 iterations performed for each run. Additionally, we conducted the Principal Component Analysis (PCA) to assess the genetic structure using GCTA 1.24.2 (http://cnsgenomics.com/software/gcta/pca.html) (Yang et al., 2011).

Based on the results of the population genetic structure analysis, we conducted ANOVA to assess the phenotypic differences among the three groups. Subsequently, multiple comparisons were performed using the Honestly Significant Difference (HSD) method.

Population genetics analysis

We performed population genetics analyses to measure genetic differentiation and variation across the genome, fixation statistics (F_{ST}) to assess population differentiation and nucleotide diversity (θ_{π}) to quantify genetic variation within populations. Both were calculated in VCFtools v0.1.17 (Danecek et al., 2011), with sliding windows of 100 Kb and step size of 20 Kb.

Linkage-disequilibrium analysis

To analyze the decay of linkage disequilibrium (LD), we used PLINK software (Purcell et al., 2007) to calculate the LD coefficient (r^2) between pairwise high-quality SNPs. The following parameters were applied: --ld-window-r2 0, --ld-window 99999, and --ld-window-kb 1000. The calculated r^2 values were then used to estimate LD decay by plotting r^2 against physical distance and determining the point at which LD decreased to a baseline level.

GWAS analysis

Prior to GWAS, SNPs were quality-filtered, and only variants with minor allele frequency (MAF) ≥ 0.05 and missing rate ≤ 0.2 were retained. The final filtered dataset comprised 87,726 SNPs and was used for association analyses of leaf spot disease and seed germination and dormancy traits. To account for accession imbalance due to geographical distribution, we applied the genome-wide efficient mixed-model association (GEMMA v0.98.3) software package (Zhou & Stephens, 2012). The mixed linear model (MLM) analysis was performed using the following equation:

$$y = X\alpha + S\beta + K\mu + e$$

Where y is the vector of phenotype, α and β are fixed effects representing marker and non-marker effects respectively; μ is random effects. X, S and K are the incidence matrices corresponding to α, β and μ, respectively, and e is the vector of random residual effects. To correct for population structure, the top three principal components (PCs) were included in the S matrix, and a simple matching coefficient matrix was used to construct the K matrix. All analyses were conducted using GEMMA. A genome-wide significance threshold was determined using a uniform threshold of 1/n, where n was the effective number of genomic variations (calculated as 87,726), following previous peanut GWAS that applied the same empirical threshold (Li et al., 2022; Xu et al., 2025). Marker effects (β) were tested for significance using the Wald test in GEMMA. From the Manhattan plots, all SNPs passing the significance threshold were considered significant and retained for downstream analyses. To evaluate the fit between observed

and expected values of quantitative traits, Manhattan plot results were validated using Q-Q plots.

Identification of the candidate genes in the GWAS associated loci.

To detect the candidate genes, we defined the candidate region as spanning 1 megabase (1 Mb) upstream and downstream of each significant SNP. Annotated genes and SNPs within these regions were extracted and analyzed. For each gene of candidate region, which was functionally annotated using TAIR genomic databases website (https://www.arabidopsis.org/), and their potential functions relevance to the trait of interest were retained.

Haplotype Analysis

For each SNP within the candidate region, haplotype-specific phenotypic associations were evaluated using a non-parametric framework. The Kruskal–Wallis test (significance set at $p < 0.05$) was used to assess overall differences in phenotype distributions among haplotype groups. When significant differences were detected, Dunn's post hoc test was performed for pairwise comparisons.

3.3 Results

Population properties of mini-core peanut collection

To further reveal the genomic variation of mini-core peanut accessions, we generated a total of 217 Gb high-quality resequencing data from 87 accessions collected worldwide (Table S3.1). We subsequently identified a total of 87,726 SNPs, and they widely distributed on 20 chromosomes. To investigate the SNP distribution between the A and B subgenomes, we calculated SNP densities, finding that the A subgenome exhibited 33.26 SNPs/Mb, while the B subgenome had 35.55 SNPs/Mb. This indicates that the B subgenome has a higher SNP density compared to the A subgenome, with chromosome 19 showing particularly high SNP density (Figure S3.1). Furthermore, the nucleotide diversity (θ_π) within the two subgenomes was calculated, revealing that the nucleotide diversity of B-subgenome (1.54×10^{-5}) is higher than A-subgenome (1.43×10^{-5}).

In order to uncover the phylogenetic relationships of mini-core collection, we performed phylogenetic tree construction, population structure, and PCA analysis (Figure 1). Three distinct groups were identified from the 87 mini-core accessions based on the NJ analysis designated here as G1, G2 and G3 (Figure 1A). These divisions were also supported by population structure analysis (Figure 1B) and PCA (Figure 1C). Group 1 is represented by *Arachis hypogaea* and contains 28 genotypes. Group 2 mainly consists of *fastigiata* types and includes 18 accessions. Group3, which has more diversity compared to the other two groups, spans a larger portion and includes other two group memberships, indicating a mixture of genotypes from both Group1 and Group 2.

The nucleotide diversity (θ_π) values across the three subpopulations ranged from 1.291 to 1.402 × 10^{-5} (Figure 1D). Group 2 exhibited the highest nucleotide diversity, reaching 1.402 × 10^{-5}, followed closely by Group 1 with 1.399 × 10^{-5}. In contrast, Group 3 showed the lowest diversity at 1.291 × 10^{-5}. These values indicate a comparable level of genetic variation within the groups. The fixation index values (F_{ST}) among the three subpopulations ranged from 0.280 to 0.373, which revealed moderate to high levels of genetic differentiation. Specifically, the F_{ST} between Group 1 and Group 2 was 0.371, between Group 2 and Group 3 was 0.373, and between Group 1 and Group 3 was 0.280. These results suggest that Groups 1 and 3 are relatively less differentiated compared to other pairwise comparisons, but overall, the populations exhibit substantial genetic structuring, which may reflect historical isolation or limited gene flow.

Phenotypic analysis

To investigate how these genetic groups correspond to phenotypic variation, we performed ANOVA for key traits, followed by multiple comparisons using the HSD method. For disease resistance traits ELS and LLS, the results revealed significant differences ($p < 0.001$) for both traits. Tukey's HSD test further confirmed these differences, highlighting clear phenotypic distinctions among the genetic subpopulations (Table S3.2 and Figure 3.2A and 3.2B). For seed dormancy-related traits, including seed dormancy and germination, significant differences were observed among the groups based on the number of germinated seeds recorded at 7, 14, and 21 days.

Tukey's HSD test showed that Group 3 consistently exhibited lower germination rates and higher dormancy compared to Groups 1 and 2 (Table S3.2 and Figure 3.2C-F).

Analysis of linkage disequilibrium (LD)

The decay rate of LD was calculated as the pairwise correlation coefficient (r^2) from the maximum value to the threshold value of 0.1. The LD decay distance was approximately 63.1 kb for all the accessions (Figure S3.2). The availability of 87,726 SNPs exceeds the minimum requirement of 44,000 SNPs for GWAS, ensuring sufficient marker density for reliable association mapping.

Genome wide association studies with quantitative traits and candidate genes identification and haplotype analysis

To dissect the genetic basis of agronomically important traits, we focused on the 6 traits, including 2 disease-related traits, and 4 seed germination and dormancy related traits. The multiple environmental traits phenotypic values (for multiple environmental traits) were used to perform GWAS with 87,726 SNPs. A threshold of p-value < 1.14×10^{-5} was employed to identify significant SNP markers associated with the target traits. In terms of disease resistance, we found 8 significant SNPs for early leaf spot (ELS) and 10 for late leaf spot (LLS). In the seed dormancy related traits, we identified 65 SNPs associated with 7-day germination, of which 22 were consistently validated in two independent phenotypic analyses. Likewise, 79 SNPs were associated with 14-day germination, with 29 validated across two separate evaluations. Similarly, for 21-day germination, we detected 70 SNPs, 32 of which were repeatedly confirmed in both phenotypic analyses. In addition, dormancy exhibited the highest number of significant SNPs (112), and notably, 40 of these SNPs were detected in all four dormancy-related traits (7-day, 14-day, 21-day, and dormancy). Among these 40, Chr10 and Chr20 contained the most associations, with 9 and 11 SNPs, respectively, underscoring their potential importance in controlling peanut seed dormancy (Table S3.3). These findings also show that most of these SNPs are distributed in the B sub-genome. Our results provide insights into the genetic architecture of complex traits in peanut, highlighting both unique and overlapping loci associated seed dormancy, and disease resistance

traits. These findings offer a valuable resource for further exploration of candidate genes and the development of marker-assisted selection strategies for peanut breeding programs.

Leaf spot

Four significant SNPs associated with ELS were identified on chromosome 11 (Figure 3.3A), and candidate regions were defined by extending 1 Mb upstream and downstream of each SNP. Within this candidate region, we identified *Ah11g381400*, which is homologous to the *Arabidopsis ATE1*(*AT5G05700*) gene. To further investigate the genetic basis underlying the ELS phenotype, we performed a haplotype analysis focusing on the SNP loci within the candidate region. Accessions were grouped based on the reference (C) or alternate (T) allele, resulting in 63 accessions with the C haplotype and 21 with the T haplotype. The average ELS value was 3.70 in the C group and 3.26 in the T group (Figure 3.3C), with a statistically significant difference ($p = 7.526 \times 10^{-3}$). Accessions with the T allele exhibited a 13.5% reduction in susceptibility relative to those with the C allele. Previous studies have shown that *Arabidopsis ATE1* mutants exhibit delayed leaf senescence and altered pathogen responses (De Marchi et al., 2016; Graciet et al., 2009), suggesting that *Ah11g381400* may represent an important candidate gene for further investigation into the ELS regulatory pathway, with the T allele conferring enhanced resistance.

For the LLS phenotype, a significant SNP was identified on chromosome 16 (Figure 3.4A). Extending 1 Mb upstream and downstream of this SNP defined a candidate region that included the *Ah16g445600* gene, which is homologous to *Arabidopsis ERF34* (*AT2G44940*), a gene known to play critical roles in stress responses and senescence regulation (Park et al. 2022). Among the analyzed accessions, the reference (C) haplotype was detected in four individuals, the alternative (G) haplotype in 12, and the heterozygous (C/G) genotype in 69. Haplotype-based phenotypic analysis revealed that individuals with the C haplotype exhibited an average LLS value of 4.0, compared with 5.3 for those with the G haplotype, while heterozygous (C/G) accessions displayed an intermediate value of 4.6 (Figure 3.4C). The disease index for the C haplotype was 32.5% lower than that for the G haplotype, indicating that

the C allele confers enhanced resistance to LLS. These findings underscore the significant association of *Ah16g445600* with LLS and highlight its potential role, possibly involving *ERF34*-mediated pathways, in regulating leaf senescence and stress responses.

Seed germination and dormancy

A total of 65, 79, and 70 significant SNPs were associated with germination at 7, 14, and 21 days, respectively, while 112 significant SNPs were detected for seed dormancy. Among these SNPs, an important candidate region common to all germination stages and dormancy traits contained the gene *Ah19g214100* (Figure 3.5A-D), a homolog of *Arabidopsis ICE1* (*AT3G26744*). This finding is consistent with recent studies demonstrating *ICE1*'s pivotal role in regulating ABA signaling and seed dormancy. For instance, Hu et al. (Hu et al., 2019) reported that *ICE1* negatively modulates ABA-responsive genes to influence germination, while MacGregor et al. (MacGregor et al. 2019) established *ICE1* as a key determinant of primary seed dormancy, with *ice1* mutants exhibiting enhanced dormancy phenotypes reversible by stratification. These functional parallels strongly implicate *Ah19g214100* as a central regulator of germination and dormancy.

Haplotype analysis of the *ICE1* homolog region revealed a consistent distribution pattern across all examined traits. The population comprised 52 individuals with the alternative (C) haplotype, 28 with the reference (T) haplotype, and 3 with the heterozygous (T/C) genotype. Statistical analysis using the Kruskal–Wallis test and Dunn’s post hoc test demonstrated significant phenotypic differences between the reference and alternative haplotypes across all traits ($p < 0.05$). Interestingly, we observed an inverse relationship between germination and dormancy phenotypes associated with these haplotypes (Figure 3.5E-H). For germination traits (7-day, 14-day, and 21-day), accessions carrying the alternative (C) haplotype exhibited significantly higher germination rates compared to those with the reference (T) haplotype. Conversely, in the dormancy assessment, reference (T) haplotype carriers showed higher dormancy levels, while alternative (C) haplotype carriers demonstrated reduced dormancy. This inverse relationship between germination efficiency and dormancy

intensity mirrors the antagonistic roles of ABA and gibberellin signaling pathways in seed physiology. Overall, these results reinforce the notion that this locus may exert pleiotropic effects on both germination and dormancy, underscoring its potential significance for seed-related traits.

3.4 Discussion

This study generated 217 Gb of high-quality resequencing data from 87 collected U.S. mini core peanut accessions and identified 87,726 SNPs distributed across 20 chromosomes. Additionally, we observed differential SNP distribution between subgenomes, with the B-subgenome exhibiting higher nucleotide diversity (1.54×10^{-5}) compared to the A-subgenome (1.43×10^{-5}). This aligns with previous studies suggesting that the B genome progenitor (*A. ipaensis*) possesses greater genetic diversity than the A genome progenitor (*A. duranensis*) (Bertioli et al. 2016; Chen et al. 2019).

The population structure analysis revealed three major genetic groups, representing different gene pools within cultivated peanut. Group 1, predominantly represented by *Arachis hypogaea*, and Group 2, mainly comprising *fastigiata* types, displayed distinct genetic backgrounds, while Group 3 appeared to be a mixture of these two pools.

The moderate to high F_{ST} values (0.280 - 0.373) between these groups indicate substantial genetic differentiation, suggesting historically limited gene flow or potential adaptation to different environments or breeding objectives. Group 3 exhibited the lowest nucleotide diversity (1.291×10^{-5}) compared to Groups 1 and 2, potentially indicating a narrower genetic base, which could be the result of more intensive selection during breeding or a founder effect. Importantly, our phenotypic analysis further validates these genetic groupings, with significant phenotypic differentiation among the three groups for multiple traits, particularly in disease resistance (ELS and LLS) and seed dormancy characteristics, with Group 3 exhibiting consistently higher dormancy and lower germination rates.

The GWAS analysis identified numerous significant SNPs associated with disease resistance, seed grade, and seed dormancy traits, providing insights into their genetic architecture. The identification of a higher number of significant SNPs in the B subgenome for seed dormancy-related traits suggests the predominant role of B subgenome in controlling these characteristics. This aligns with our observation of higher genetic diversity in this subgenome.

To identify candidate genes underlying significant GWAS signals, we adopted a strategy of examining regions extending 1 Mb upstream and downstream from each significant SNP (Fang et al., 2017; Yano et al., 2016). We acknowledge that this 2 Mb window is broad relative to the LD decay estimated in our dataset (~63.1 kb) and therefore may include non-causal genes. However, previous analysis of the U.S. peanut mini-core also reported long-range LD persistence (genome-wide half-decay distance ~3.78 Mb) (Otyama et al., 2019), indicating that LD extent can vary substantially depending on marker platform and population structure. Thus, we used the ±1 Mb window as a conservative discovery strategy to reduce false negatives. Future fine-mapping with larger populations and denser marker panels, combined with the integration of expression QTL (eQTL) data, would substantially narrow these intervals and facilitate the identification of causal variants with greater precision.

Our GWAS analysis identified key candidate genes for both disease resistance and seed dormancy traits. For ELS resistance, *Ah11g381400*, homologous to *Arabidopsis ATE1*, emerged as a promising candidate. *ATE1* regulates protein arginylation, affecting leaf senescence and pathogen responses (Yoshida et al. 2002; Graciet et al. 2010). In *Arabidopsis*, loss-of-function mutations in AtATE1 produce a delayed leaf senescence phenotype, with senescence progressing significantly more slowly than in wild-type plants under both age-dependent and dark-induced conditions (Yoshida et al. 2002). Beyond senescence, the N-end rule pathway mediated by ATE1 regulates diverse processes including seed germination, leaf and shoot development, and stress responses (Tasaki & Kwon, 2007). The connection between senescence regulation and disease resistance is mechanistically plausible: fungal pathogens such as *Cercospora arachidicola* actively exploit host senescence programs to facilitate

colonization and sporulation, and plants with attenuated senescence frequently show enhanced resistance to necrotrophic and hemibiotrophic pathogens (Häffner et al., 2015). The associated SNP at *Ah11g381400* is intronic (Figure 3.3B), suggesting that the observed allelic effects may be mediated through regulatory mechanisms such as altered splicing or transcriptional modulation rather than direct amino acid changes. In our haplotype analysis, accessions carrying the T allele at the lead SNP exhibited a 13.5% reduction in ELS susceptibility relative to those carrying the C allele. These results suggest that this allelic variation may influence *Ah11g381400* expression or splicing, thereby modulating the rate of pathogen-accelerated senescence in inoculated leaves.

For LLS resistance, we identified *Ah16g445600*, homologous to *ERF34*, an AP2/ERF family transcription factor. It functions within ethylene signaling pathways that regulate both stress responses and senescence (Iqbal et al., 2017). *ERF34* belongs to a set of ERF transcription factors whose expression is altered under both age-dependent and dark-induced senescence, and all identified members of this group have been previously implicated in biotic and abiotic stress responses (Park et al. 2022). This functional background provides a mechanistic explanation for the LLS phenotype, because late leaf spot progression is accompanied by senescence-associated tissue damage. Accessions carrying the C haplotype at the candidate gene showed a 32.5% lower disease index compared to G haplotype carriers in our analysis. The lead SNP at this locus is located in the CDS region and is annotated as a synonymous variant (Figure 3.4B), with codon substitution from GTC to GTG (both encoding valine). Although it does not change amino acid identity, synonymous variants may still influence codon usage, translation efficiency, mRNA stability, or may be in linkage disequilibrium with nearby causal variants. Therefore, functional validation is still required to determine the causal mechanism.

For seed dormancy, *Ah19g214100*, homologous to *Arabidopsis ICE1*, was consistently identified across all germination timepoints. *AtICE1* is a *bHLH* transcription factor originally characterized for its role in cold stress signaling but subsequently identified as a key regulator of seed dormancy (Chinnusamy et al. 2003; MacGregor et

al. 2019). In *Arabidopsis*, *ICE1* is expressed in the endosperm and contributes to determining primary dormancy depth, in part through negative regulation of ABA signaling during germination (MacGregor et al. 2019; Hu et al. 2019). In our population, the C haplotype was associated with higher germination rates and reduced dormancy compared with the T haplotype, consistent with ABA-gibberellin antagonism in seed physiology. The lead SNP at this locus is intronic (Figure 3.5l), suggesting a potential regulatory effect rather than a direct coding change. The pleiotropic association of this locus across all four dormancy-related traits (7-, 14-, and 21-day germination, and dormancy score) further supports *Ah19g214100* as a central candidate gene for dormancy regulation in peanut.

These findings provide valuable targets for marker-assisted selection to optimize both disease resistance and seed characteristics in peanut breeding programs. Overall, this study advances our understanding of genetic diversity and trait architecture in peanut, offering both fundamental insights into crop evolution and practical tools for crop improvement. The identified genetic candidate genes provide a foundation for developing peanut varieties with enhanced disease resistance and optimized seed characteristics to meet the challenges of sustainable agriculture. Further research will be essential to validate the specific genes responsible for these key agronomic traits.

List of Figures

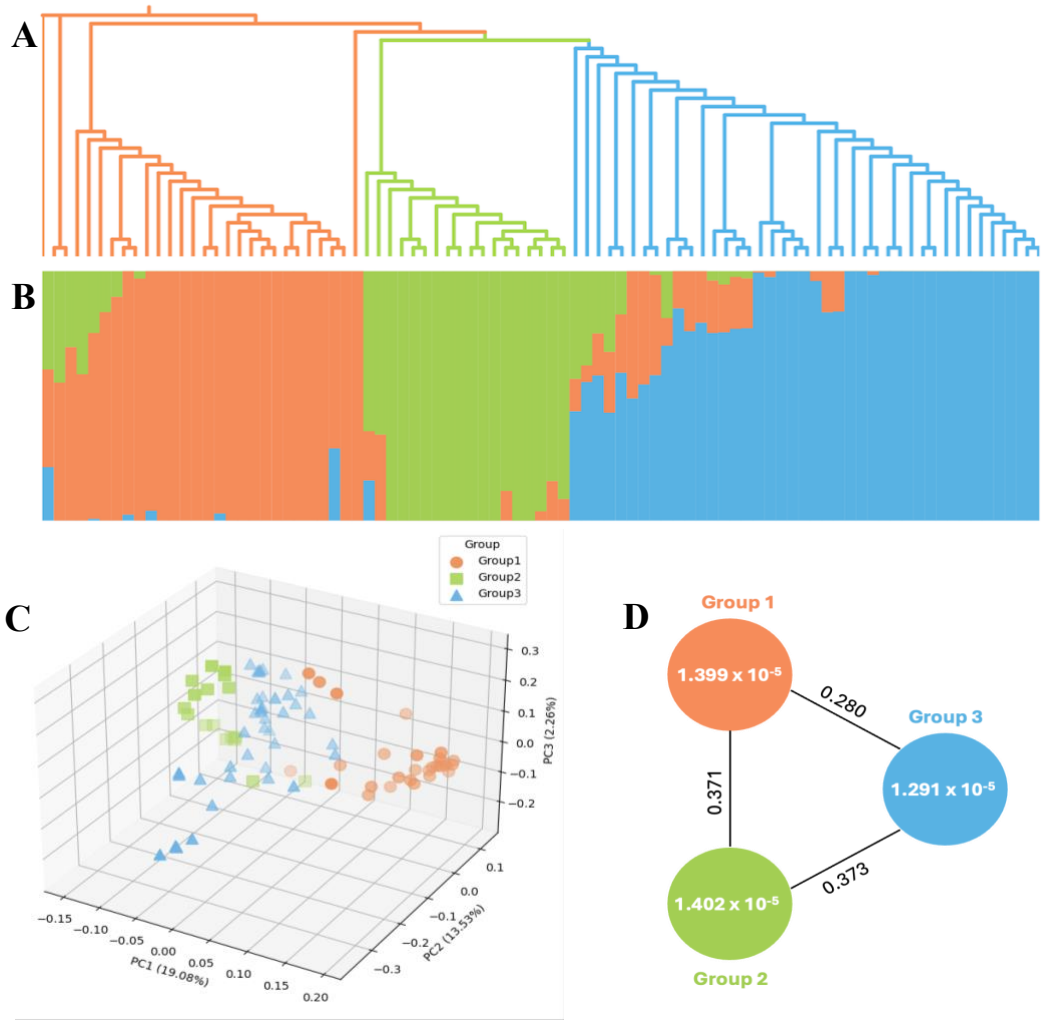

Figure 3.1: Phylogenetic tree, genetic structure, PCA. **A.** Phylogenetic tree of the population, generated with 87,726 high-quality SNPs. **B.** Structure analysis with K = 3. The y-axis quantifies cluster membership, and the x-axis represents different accessions. The orders and positions of accessions on the x axis are consistent with those in the phylogenetic tree. **C.** PCA plot of the first three components (PC1, PC2 and PC3). **D.** Genetic diversity and population differentiation across the three groups. The values in the circles represent the genetic diversity ($\Theta\pi$) of the groups, and the values between the groups indicate population differentiation (F_{ST}).

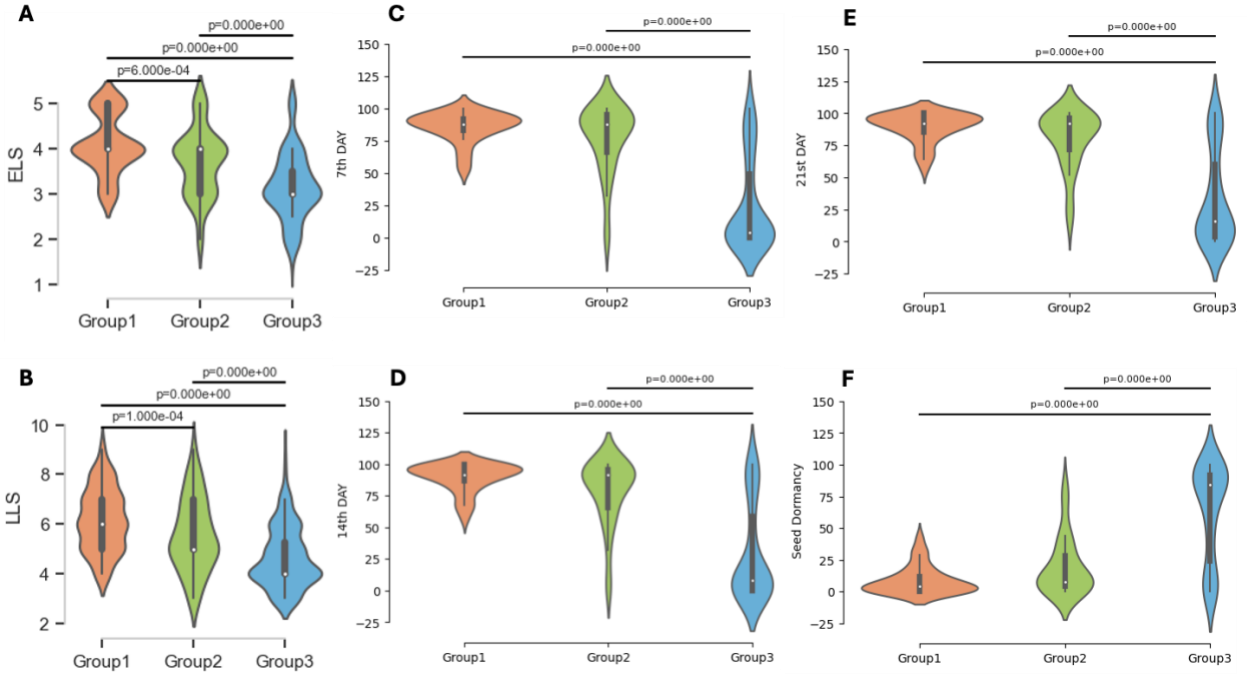

Figure 3.2: Phenotypic variation among three genetic groups for disease resistance and seed dormancy traits. A. ELS resistance scores across three genetic groups (p<0.001). **B.** LLS resistance scores among genetic groups showing significant differences (p<0.001). **C.** Seed germination rates at 7 days post-planting across three genetic groups (p<0.001). **D.** Seed germination rates at 14 days post-planting across three genetic groups (p<0.001). **E.** Seed germination rates at 21 days post-planting across three genetic groups (p<0.001). **F.** Overall seed dormancy levels among the three genetic groups (p<0.001).

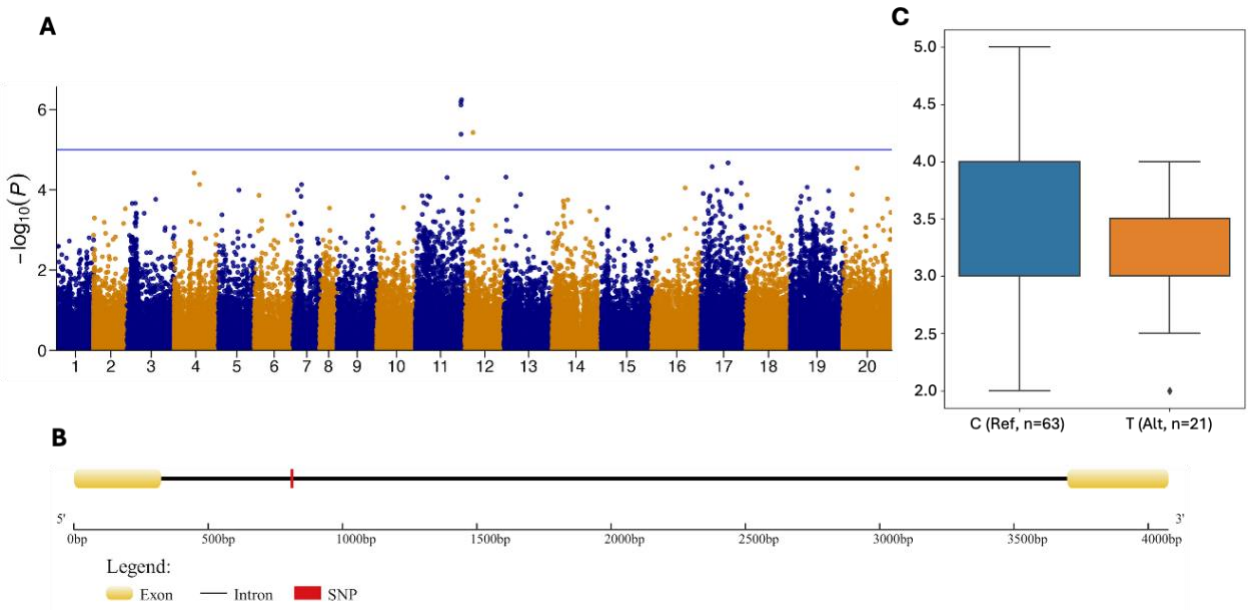

Figure 3.3: A. Manhattan plot for ELS. The blue horizontal line indicates the suggestive significance threshold ($-\log_{10}P = 4.94$). **B.** Gene structure of Ah11g381400, highlighting exons (yellow boxes), introns (black lines), and the significant SNP (red). **C.** Haplotype analysis for ELS

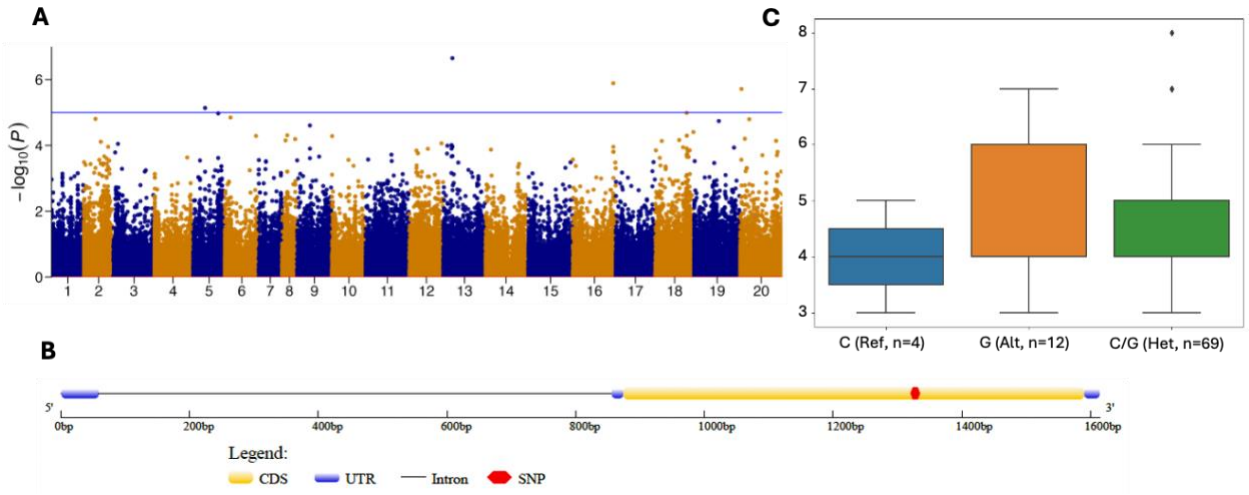

Figure 3.4: A. Manhattan plot for LLS. The blue horizontal line indicates the suggestive significance threshold ($-\log_{10}P = 4.94$). **B.** Gene structure of Ah16g445600, highlighting exons (yellow boxes), introns (black lines), and the significant SNP (red). **C.** Haplotype analysis for LLS.

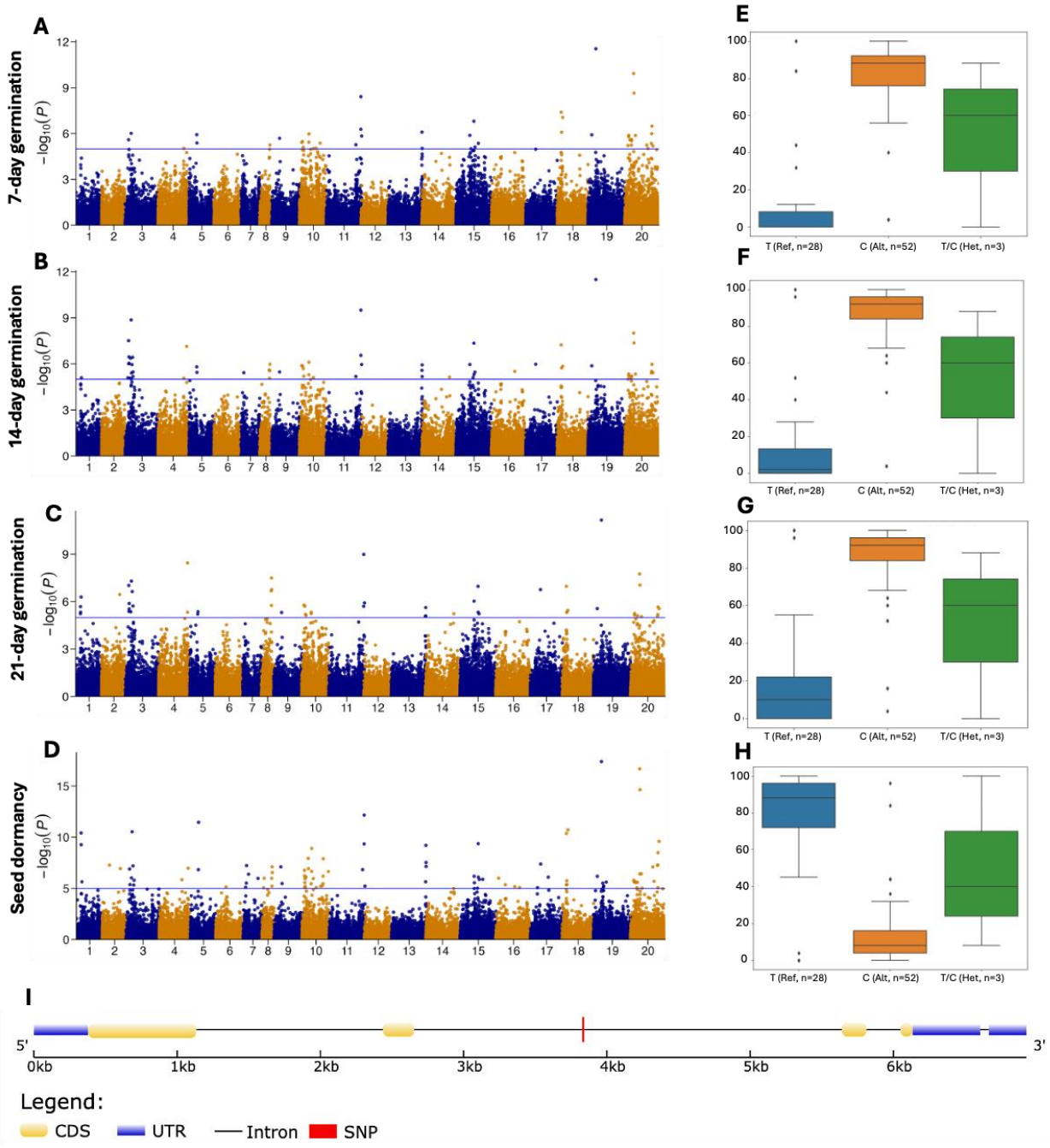

Figure 3.5: Genome-wide association study and haplotype analysis of seed germination and dormancy traits. A. Manhattan plot of SNP associations with 7-day germination. B. Manhattan plot of SNP associations with 14-day germination. C. Manhattan plot of SNP associations with 21-day germination. D. Manhattan plot of SNP associations with seed dormancy. E. Box plot showing phenotypic differences among haplotypes for 7-day germination. F. Box plot showing phenotypic differences among haplotypes for 14-day germination. G. Box plot showing phenotypic differences among haplotypes for 21-day germination. **H.** Box plot showing phenotypic differences among

haplotypes for seed dormancy. **I.** Gene structure of Ah19g214100 (ICE1 homolog) showing the functional SNP position and genomic features (CDS, UTR, intron).

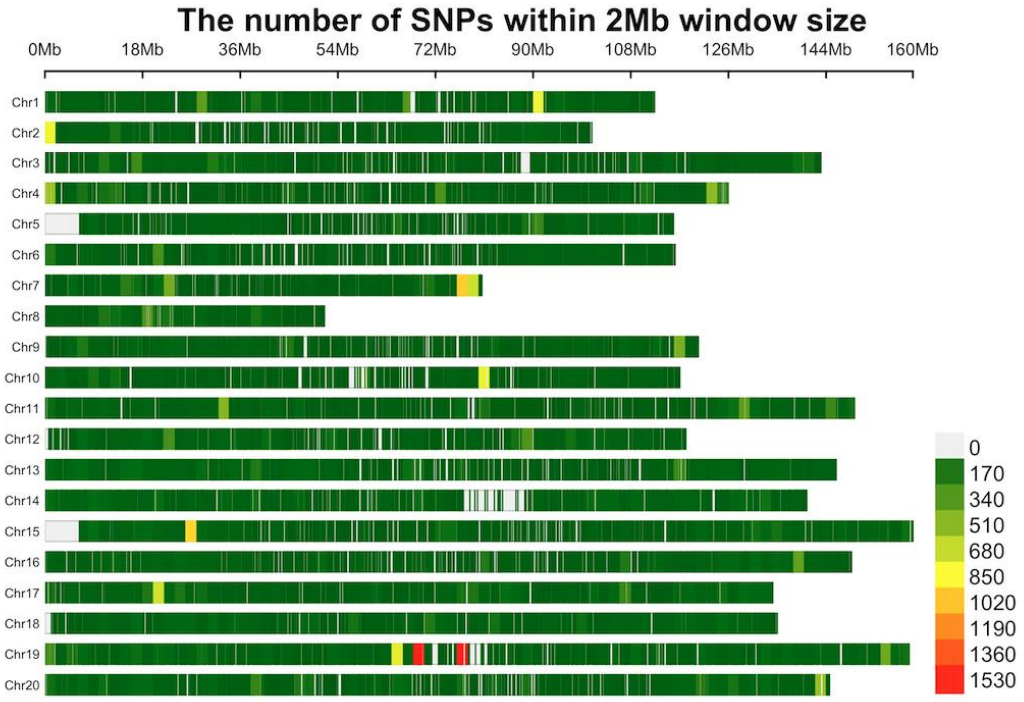

Figure S3.1: Distribution of SNP markers in the 20 chromosomes. Horizontal axis represents physical distance along each chromosome. Vertical axis represents the 20 chromosomes of *Arachis hypogaea*; The color index represents the number of SNPs in 2.0 Mb window.

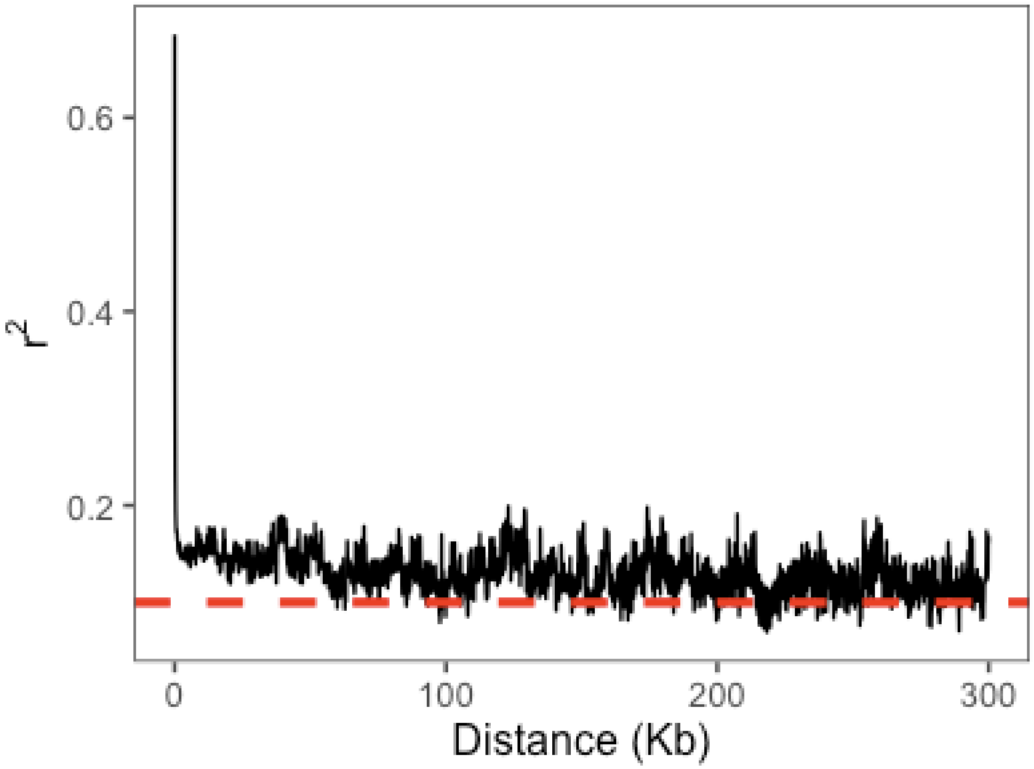

Figure S3.2: Linkage disequilibrium (LD) decay over distance.

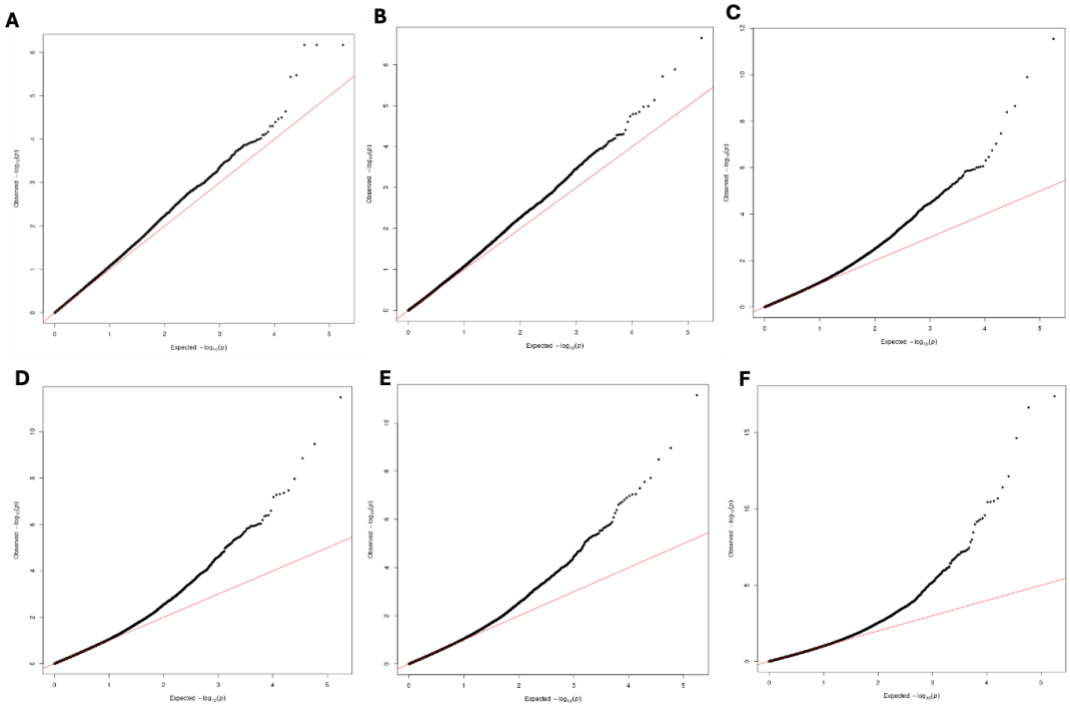

Figure S3.3: Q–Q plot of observed vs. expected log10P values. A. ELS. B. LLS. C. 7-day germination. D. 14-day germination E. 21-day germination F. seed dormancy.

References

Ahmad S, Nawade B, Sangh C, 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:. https://doi.org/10.1007/S13205-020-02446-4

Bertioli DJ, Cannon SB, Froenicke L, et al (2016) The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nature Genetics 2016 48:4 48:438–446. https://doi.org/10.1038/ng.3517

Bertioli DJ, Jenkins J, Clevenger J, et al (2019) The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nature Genetics 2019 51:5 51:877–884. https://doi.org/10.1038/s41588-019-0405-z

Bewley JD, Black M (1985) Seeds. https://doi.org/10.1007/978-1-4615-1747-4

Bomireddy D, Sharma V, Gangurde SS, et al (2024) Multi-locus genome wide association study uncovers genetics of fresh seed dormancy in groundnut. BMC Plant Biol 24:1–14. https://doi.org/10.1186/S12870-024-05897-6/FIGURES/6

Chavarro C, Chu Y, Holbrook C, et al (2020) Pod and Seed Trait QTL Identification To Assist Breeding for Peanut Market Preferences. G3 Genes|Genomes|Genetics 10:2297–2315. https://doi.org/10.1534/G3.120.401147

Chen W, Jiao Y, Cheng L, et al (2016) Quantitative trait locus analysis for pod- and kernel-related traits in the cultivated peanut (Arachis hypogaea L.). BMC Genet 17:1–9. https://doi.org/10.1186/S12863-016-0337-X/FIGURES/2

Chen X, Lu Q, Liu H, et al (2019) Sequencing of Cultivated Peanut, Arachis hypogaea, Yields Insights into Genome Evolution and Oil Improvement. Mol Plant 12:920–934. https://doi.org/10.1016/J.MOLP.2019.03.005

Chinnusamy V, Ohta M, Kanrar S, et al (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev 17:1043. https://doi.org/10.1101/GAD.1077503

Chiteka ZA, Gorbet DW, Knauft DA, et al (1988) Components of Resistance to Late Leafspot in Peanut. II. Correlations Among Components and Their Significance in Breeding for Resistance. Peanut Science 15:76–81. https://doi.org/10.3146/I0095-3679-15-2-9

Chu Y, Chee P, Culbreath A, et al (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 10:393682. https://doi.org/10.3389/FPLS.2019.00883/BIBTEX

Clevenger J, Chu Y, Chavarro C, et al (2018) Mapping late leaf spot resistance in peanut (Arachis hypogaea) using QTL-seq reveals markers for marker-assisted selection. Front Plant Sci 9:318811. https://doi.org/10.3389/FPLS.2018.00083/BIBTEX

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156. https://doi.org/10.1093/BIOINFORMATICS/BTR330

Dang PM, Lamb MC, Chen CY (2021) Association of differentially expressed R-gene candidates with leaf spot resistance in peanut (Arachis hypogaea L.). Mol Biol Rep 48:323. https://doi.org/10.1007/S11033-020-06049-3

De Marchi R, Sorel M, Mooney B, et al (2016) The N-end rule pathway regulates pathogen responses in plants. Scientific Reports 2016 6:1 6:1–15. https://doi.org/10.1038/srep26020

Fang L, Wang Q, Hu Y, et al (2017) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. Nature Genetics 2017 49:7 49:1089–1098. https://doi.org/10.1038/ng.3887

Graciet E, Mesiti F, Wellmer F (2010) Structure and evolutionary conservation of the plant N-end rule pathway. Plant J 61:741–751. https://doi.org/10.1111/J.1365-313X.2009.04099.X

Graciet E, Walter F, Maoiléidigh DÓ, et al (2009) The N-end rule pathway controls multiple functions during Arabidopsis shoot and leaf development. Proc Natl Acad Sci U S A 106:13618–13623. https://doi.org/10.1073/PNAS.0906404106

Guo M, Deng L, Gu J, et al (2024a) Genome-wide association study and development of molecular markers for yield and quality traits in peanut (Arachis hypogaea L.). BMC Plant Biol 24:1–16. https://doi.org/10.1186/S12870-024-04937-5/FIGURES/9

Häffner E, Konietzki S, Diederichsen E (2015) Keeping Control: The Role of Senescence and Development in Plant Pathogenesis and Defense. Plants 4:449. https://doi.org/10.3390/PLANTS4030449

Han S, Yuan M, Clevenger JP, 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 9:371666. https://doi.org/10.3389/FPLS.2018.01012/BIBTEX

Holbrook CC, Dong W (2005) Development and Evaluation of a Mini Core Collection for the U.S. Peanut Germplasm Collection. Crop Sci 45:1540–1544. https://doi.org/10.2135/CROPSCI2004.0368

Hu Y, Han X, Yang M, et al (2019) The Transcription Factor INDUCER OF CBF EXPRESSION1 Interacts with ABSCISIC ACID INSENSITIVE5 and DELLA Proteins to Fine-Tune Abscisic Acid Signaling during Seed Germination in Arabidopsis. Plant Cell 31:1520–1538. https://doi.org/10.1105/TPC.18.00825

Huang R, Li H, Gao C, et al (2023) Advances in omics research on peanut response to biotic stresses. Front Plant Sci 14:1101994. https://doi.org/10.3389/FPLS.2023.1101994/PDF

Iqbal N, Khan NA, Ferrante A, et al (2017) Ethylene role in plant growth, development and senescence: interaction with other phytohormones. Front Plant Sci 8:235913. https://doi.org/10.3389/FPLS.2017.00475/PDF

Khera P, Pandey MK, Wang H, 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. https://doi.org/10.1371/JOURNAL.PONE.0158452

Kumar R, Janila P, Vishwakarma MK, et al (2020) Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. Plant Biotechnol J 18:992–1003. https://doi.org/10.1111/PBI.13266

Lefort V, Desper R, Gascuel O (2015) FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Mol Biol Evol 32:2798–2800. https://doi.org/10.1093/MOLBEV/MSV150

Li L, Cui S, Dang P, et al (2022) GWAS and bulked segregant analysis reveal the Loci controlling growth habit-related traits in cultivated Peanut (Arachis hypogaea L.). BMC Genomics 2022 23:1 23:403-. https://doi.org/10.1186/S12864-022-08640-3

Liang Y, Baring MR, Septiningsih EM (2018) Mapping of Quantitative Trait Loci for Yield and Grade Related Traits in Peanut (Arachis hypogaea L.) Using High-Resolution SNP Markers. Plant Breed Biotechnol 6:454–462. https://doi.org/10.9787/PBB.2018.6.4.454

Lu Q, Huang L, Liu H, et al (2024) A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits. Nature Genetics 2024 56:3 56:530–540. https://doi.org/10.1038/s41588-024-01660-7

MacGregor DR, Zhang N, Iwasaki M, et al (2019) ICE1 and ZOU determine the depth of primary seed dormancy in Arabidopsis independently of their role in endosperm development. Plant J 98:277–290. https://doi.org/10.1111/TPJ.14211

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63. https://doi.org/10.1016/J.TPLANTS.2006.12.001

Nonogaki H, Barrero JM, Li C (2018) Editorial: Seed dormancy, germination, and pre-harvest sprouting. Front Plant Sci 871:428588. https://doi.org/10.3389/FPLS.2018.01783/BIBTEX

Otyama PI, Wilkey A, Kulkarni R, et al (2019) Evaluation of linkage disequilibrium, population structure, and genetic diversity in the U.S. peanut mini core collection. BMC Genomics 2019 20:1 20:481-. https://doi.org/10.1186/S12864-019-5824-9

Park SJ, Park S, Kim Y, et al (2022) Ethylene responsive factor34 mediates stress-induced leaf senescence by regulating salt stress-responsive genes. Plant Cell Environ 45:1719–1733. https://doi.org/10.1111/PCE.14317

Patel JD, Wang ML, Dang P, et al (2022) Insights into the Genomic Architecture of Seed and Pod Quality Traits in the U.S. Peanut Mini-Core Diversity Panel. Plants (Basel) 11:. https://doi.org/10.3390/PLANTS11070837

Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81:559. https://doi.org/10.1086/519795

Satish Ingale, S. K. Shrivastava (2011) Nutritional study of new variety of groundnut(Arachis hypogaea L.) JL-24 seeds. African Journal of Food Science 2:1–2

Shelar VR, Karjule AP, Jayadeva B (2014) Induction of dormancy in groundnut -A review. Agricultural Reviews 35:216. https://doi.org/10.5958/0976-0741.2014.00908.8

Shirasawa K, Koilkonda P, Aoki K, et al (2012) In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol 12:1–13. https://doi.org/10.1186/1471-2229-12-80/FIGURES/2

Silva MFC, Silva CRC, Lima LM, Santos RC (2017) Differential expression of dormancy-Associated genes in fastigiata and hypogaea peanut. Genetics and Molecular Research 16:. https://doi.org/10.4238/GMR16039820

Tasaki T, Kwon YT (2007) The mammalian N-end rule pathway: new insights into its components and physiological roles. Trends Biochem Sci 32:520–528. https://doi.org/10.1016/J.TIBS.2007.08.010

Wang ML, Pinnow CYC, Barkley NA, et al (2012) Seed Dormancy Variability in the U.S. Peanut Mini-Core Collection. Research Journal of Seed Science 5:84–95. https://doi.org/10.3923/RJSS.2012.84.95

Wang ML, Wang H, Zhao C, et al (2022) Identification of QTLs for Seed Dormancy in Cultivated Peanut Using a Recombinant Inbred Line Mapping Population. Plant Mol Biol Report 40:208–217. https://doi.org/10.1007/S11105-021-01315-5

Xu J, Jiang X, Yin X, et al (2025) Genome-wide association analysis in peanut accessions uncovers the genetic basis regulating oil and fatty acid variation. BMC Plant Biology 2025 25:1 25:651-. https://doi.org/10.1186/S12870-025-06690-9

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 88:76. https://doi.org/10.1016/J.AJHG.2010.11.011

Yano K, Yamamoto E, Aya K, et al (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat Genet 48:927–934. https://doi.org/10.1038/NG.3596

Yoshida S, Ito M, Callis J, et al (2002) A delayed leaf senescence mutant is defective in arginyl-tRNA:protein arginyltransferase, a component of the N-end rule pathway in Arabidopsis. Plant J 32:129–137. https://doi.org/10.1046/J.1365-313X.2002.01407.X

Zhang H, Chu Y, Dang P, et al (2020) Identification of QTLs for resistance to leaf spots in cultivated peanut (Arachis hypogaea L.) through GWAS analysis. Theoretical and Applied Genetics 133:2051–2061. https://doi.org/10.1007/S00122-020-03576-2/TABLES/5

Zhang S, Hu X, Miao H, et al (2019) QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (Arachis hypogaea L.). BMC Plant Biol 19:1–15. https://doi.org/10.1186/S12870-019-2164-5/FIGURES/5

Zhao K, Xue H, Li G, et al (2025) Pangenome analysis reveals structural variation associated with seed size and weight traits in peanut. Nature Genetics 2025 57:5 57:1250–1261. https://doi.org/10.1038/s41588-025-02170-w

Zhao T, Ying P, Zhang Y, et al (2023) Research Advances in the High-Value Utilization of Peanut Meal Resources and Its Hydrolysates: A Review. Molecules 2023, Vol 28, Page 6862 28:6862. https://doi.org/10.3390/MOLECULES28196862

Zhao X, Chen J, Du F (2011) Potential use of peanut by-products in food processing: a review. J Food Sci Technol 49:521. https://doi.org/10.1007/S13197-011-0449-2

Zheng Z, Sun Z, Qi F, et al (2024) Chloroplast and whole-genome sequencing shed light on the evolutionary history and phenotypic diversification of peanuts. Nature Genetics 2024 56:9 56:1975–1984. https://doi.org/10.1038/s41588-024-01876-7

Zhong C, Liu Y, Li Z, et al (2023) Genome-wide analysis reveals regulatory mechanisms and expression patterns of TGA genes in peanut under abiotic stress

and hormone treatments. Front Plant Sci 14:1269200. https://doi.org/10.3389/FPLS.2023.1269200/PDF

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 2012 44:7 44:821–824. https://doi.org/10.1038/ng.2310

Chapter 4: Integrated Physiology and Transcriptomics Reveal Divergent Drought Adaptation in Water-Saver and Water-Spender Peanut Cultivars

Abstract

Peanut drought resilience is critical for production stability in the southeastern United States, where irrigation capacity is limited. We integrated field physiology and transcriptomics to dissect contrasting water-use strategies in five cultivars: two water savers (AU16, Line8), two water spenders (AU17, PI50), and one drought-sensitive genotype (PI39). Plants were evaluated at week 1 (W1), week 3 (W3), week 5 (W5), and post-rewatering recovery (RC), generating 120 RNA-seq libraries. Progressive drought reduced photosynthesis, stomatal conductance, and electron transport rate, while intrinsic water-use efficiency increased. In a principal component analysis (PCA) of these traits, drought-stressed and well-watered plants separated progressively from W1 to W5 along the main axis of physiological variation and then partially converged again after rewatering (RC); this indicates that the physiological gap between stressed and control plants widened as drought intensified and was only partly reversed during recovery. WGCNA of 6,000 highly variable genes from 45 drought-treated W1/W3/W5 samples identified 22 non-grey modules. Most transcriptomic variation followed drought progression, with key stage modules showing concordant associations with photosynthesis and stomatal conductance. A secondary, time-independent axis captured water-use strategy differences, represented by modules enriched for translation/ribosome functions and proteolysis-related processes, respectively. After rewatering, the major drought-stage modules did not return to their pre-drought state, indicating a persistent transcriptional imprint of stress, whereas a photosynthesis-associated module partially rebounded toward baseline. The strategy modules, by contrast, changed little across drought and recovery, suggesting that they reflect stable water-use-strategy-associated regulation rather than transient drought-stage responses. Within the two strategy modules, we prioritized ten candidate hub genes, eight of which were independently confirmed as differentially expressed in at least one genotype-by-timepoint drought-versus-control comparison. Together, these results show that peanut drought response was driven mainly by drought progression, while water-use strategy

contributed a smaller but consistent transcriptomic signature that persisted across the drought-recovery cycle.

Keywords: Peanut (*Arachis hypogaea*); drought stress; water-use strategy; water saver and water spender; RNA-seq; WGCNA; stomatal conductance; rewatering recovery

4.1 Introduction

Peanut (*Arachis hypogaea L.*), also known as groundnut, is a globally important legume and oilseed crop valued for its edible seeds, favorable protein and oil composition. It is a primary source of cooking oil in Asian markets, while in the United States, it is most extensively used in the production of peanut butter. Peanut production heavily relies on rainfed systems in semi-arid and subtropical regions, leaving the crop vulnerable to frequent water deficits (Pokhrel et al., 2024). Similarly, much of the U.S. crop is grown with limited irrigation, increasing its risk of seasonal drought. Therefore, understanding how peanut responds to drought, and how genotypes differ in water use and recovery, is essential for stabilizing yield and quality where irrigation is limited (Bhogireddy et al., 2020; Reddy et al., 2003.).

Drought stress progresses as a multiscale process that begins with declining soil water availability and propagates through the soil-plant-atmosphere continuum. As soil dries, root water uptake becomes increasingly constrained, leaf water potential declines, and turgor-dependent processes such as cell expansion and leaf area development are reduced (Chaves et al., 2003; Flexas & Medrano, 2002). At the canopy level, stomata respond to hydraulic and hormonal signals, including abscisic acid-mediated signaling, by inducing stomatal closure, which reduces transpirational water loss but simultaneously limits CO_2 diffusion and carbon assimilation (Lawlor, 2002). Photosynthetic inhibition under drought is mediated by both stomatal and non-stomatal limitations, the relative contributions of which are governed by stress severity, duration, and developmental stage (Chaves et al., 2003; Flexas & Medrano, 2002; Lawlor, 2002). As drought continues, the disruption of redox homeostasis and overaccumulation of reactive oxygen species (ROS) can damage membranes, proteins, and photosystems, particularly when antioxidant defense systems are overwhelmed (Farooq et al., 2009). Carbon metabolism and source-sink relations are also altered, with consequences for

partitioning to vegetative and reproductive organs. In legumes, symbiotic dinitrogen (N_2) fixation is highly susceptible to early-stage water deficits due to the acute sensitivity of nodule metabolism to host hydration status and carbon assimilate partitioning, which ultimately imposes a significant constraint on reproductive growth (Sinclair et al., 2007). At whole-plant and field scales, these interacting effects translate into reduced biomass accumulation, compromised reproductive development, and lower yield stability, making drought one of the most consistent constraints on crop productivity worldwide (Boyer, 1982).

Given the complex and quantitatively inherited nature of drought response, current research increasingly combines physiological phenotyping with transcriptomic and genetic analyses to resolve its underlying mechanisms and heritable variation. Physiological studies have provided frameworks for drought adaptation, including traits related to water acquisition, water conservation, and water-use efficiency, and have emphasized that trait value depends on environment and plant phenology (Araus et al., 2002; Blum, 2005). Building on this foundation, RNA-seq has enabled high-resolution identification of stress-responsive pathways, while GWAS has mapped loci associated with drought-related phenotypes in diverse germplasm panels. Across crops, these approaches consistently implicate hormonal signaling, osmotic adjustment, ROS detoxification, membrane protection, and carbon remobilization as key components of adaptive response. Representative examples include maize studies in which natural variation at ZmNAC111 and ZmVPP1 was associated with seedling drought performance (Mao et al., 2015; X. Wang et al., 2016), and a recent integrative maize analysis combining GWAS and transcriptome information that identified ZmGRAS15 as a candidate regulator linked to drought tolerance and root elongation (D. Wang et al., 2025). In soybean, the integration of genome-wide association studies and transcriptomic profiling has similarly accelerated the discovery of novel candidate genes underlying drought tolerance during the critical germination stage (Kong et al., 2025). In summary, these studies show that drought resilience is rarely controlled by single genes alone; instead, it arises from coordinated networks whose phenotypic effects depend on developmental stage, water availability, and genetic background.

In peanut, these adaptive mechanisms are essential, as the crop is exceptionally vulnerable when water deficit occurs during reproductive stages such as flowering, peg penetration, and pod filling, which severely limits kernel development (Pokhrel et al., 2024; Reddy et al., 2003). Previous studies have shown reductions in canopy growth, gas exchange, and partitioning efficiency under water deficit, and severe stress events may elevate risks of quality defects such as pre-harvest aflatoxin contamination (Pokhrel et al., 2024). Large-scale evaluations of groundnut germplasm confirm a synergistic interaction between heat and drought stress, which significantly exacerbates yield loss. This combined environmental stress drives a complex genotype-by-environment (G×E) architecture, demonstrating that breeding selections made in environments failing to capture target stress conditions are unreliable predictors of actual field performance (Hamidou et al., 2013). Recent peanut molecular studies have strengthened these observations by identifying drought-responsive genes and expression patterns, as well as genomic regions associated with drought-related traits (Bhogireddy et al., 2020; Gundaraniya et al., 2023; Kottapalli et al., 2009; X. Wang et al., 2021). Transcriptome analyses of contrasting genotypes have reported differential regulation of signaling, protective metabolism, and cellular maintenance pathways under progressive stress, while GWAS has identified multiple loci and candidate genes associated with variation in drought-relevant traits (P. Dang et al., 2024). Within these regulatory networks, our previous genome-wide characterization study highlighted the important role of kinase cascades, identifying specific MAPK gene family members whose expression is strongly induced by drought (J. Zhang et al., 2025). In parallel, physiological studies have shown repeatable genotypic differences in transpiration efficiency, canopy conductance behavior, and the soil moisture thresholds for transpiration declines (Balota et al., 2012; Jyostna Devi et al., 2009; Ratnakumar et al., 2009). These findings align with the water-use strategy concept formalized in which some genotypes conserve water earlier during soil drying (water savers) while others like common bean and durum wheat can maintain higher conductance and gas exchange for longer periods (water spenders) (Blum, 2009; Nakhforoosh et al., 2016; Polania et al., 2016; Polania et al., 2022). In previous study, similar phenomenon was

identified in peanut where cultivars with relatively higher stomatal conductance functioned as water spenders, while those exhibiting more conservative conductance and higher intrinsic water-use efficiency acted as water savers (Zhang et al., 2022). However, yield difference between drought-tolerant cultivars and drought-sensitive genotype was more significant than its difference among water savers and water spenders (Zhang et al., 2022). This indicates that both water-use strategies can contribute to drought tolerance, while yield also depends on factors such as carbon allocation, the timing and duration of stress exposure, and post-drought recovery capacity; these factors cannot be resolved from gas-exchange physiology alone.

Based on this background, the present study was designed to integrate physiological and transcriptomic analyses to clarify how water-use strategies in peanut respond during drought imposition and recovery. We compared water saver and water spender cultivars under drought conditions and included a drought-sensitive cultivar as a reference. Specific objectives of this study were (1) To quantify temporal changes in leaf gas-exchange traits (A, gs, ETR, and WUEi) across drought progression (W1, W3, W5) and post-rewatering recovery (RC). (2) Resolved transcriptome-scale co-expression modules associated with drought stage versus water-use strategy and characterized their functional enrichment. (3) To evaluate the reversibility of key modules after rewatering and prioritized candidate hub genes linked to strategy-associated regulation. We hypothesized that water savers and water spenders would show distinct gs and WUEi trajectories during both drought and recovery, and that these physiological differences would map to strategy-associated transcriptional programs superimposed on a dominant stage-driven drought response.

4.2 Materials and Methods

Plant Materials and Experimental Design

The experiment was conducted at the National Peanut Research Laboratory (NPRL), USDA-ARS, Dawson, Georgia, USA. In the third week of May, we planted five peanut cultivars representing distinct physiological traits based on prior breeding programs: two water spenders (AU17 and PI50), two water savers (AU16 and Line 8),

and one drought-sensitive cultivar (PI39) (Zhang et al., 2022). The cultivars were planted across four plots. Three plots were assigned to the drought treatment, which utilized automated rainout shelters triggered by rain sensors to exclude precipitation. The remaining plot served as a well-watered control.

Drought stress was initiated 60 days after planting (DAP) and continued for five weeks. Following the five-week stress period, plots were irrigated for two weeks to restore soil water availability and to evaluate physiological resilience during recovery relative to peak drought.

Leaf gas exchange measurements

Leaf gas exchange was measured using a LI-6800 Portable Photosynthesis System (LI-COR Biosciences, Lincoln, NE, USA) during the drought cycle at weeks 1, 3, and 5 and after rewatering recovery (W1, W3, W5, and RC). Measurements were taken on two terminal leaflets that were consistently selected across cultivars, sampling dates, and replications. During each measurement, chamber environmental conditions, including light, relative humidity, and air temperature, were adjusted to approximate ambient field conditions at the time of sampling. Leaves were allowed to stabilize in the chamber before data were logged. Stomatal conductance (gs), net photosynthetic rate (A), and electron transport rate (ETR) were recorded using the LI-6800 system, and intrinsic water-use efficiency (WUEi) was calculated as A/gs.

Physiological data analysis

To evaluate the multivariate structure of physiological variation across cultivars, treatments, and drought stages, PCA was performed on the four core traits (A, gs, WUEi, ETR). PCA biplots were constructed to visualize sample clustering and trait loading directions, with panels faceted by drought stage (W1, W3, W5, RC).

Strategy-focused physiological testing was restricted to net photosynthesis and stomatal conductance, the two traits previously used to define water-use behavior. Analyses were performed on drought-treated samples only, using measurements from W1, W3, and W5. For each cultivar, trait values were averaged across the three drought stages, and pairwise contrasts were tested between each water spender (AU17, PI50)

and each water saver (AU16, Line8) using ANOVA with Tukey's HSD post hoc comparisons. Because this was a field experiment, pairwise contrasts were interpreted at a threshold of $P < 0.10$.

Leaf sampling, RNA sequencing, and bioinformatics

Leaf tissues were collected for RNA sequencing from five cultivars at four developmental stages: predrought baseline (W1), progressive drought (W3 and W5), and post-rewatering recovery (RC), with three independent biological replicates per treatment combination. This design generated 120 samples (5 genotypes x 2 treatmentss x 4 timepoints x 3 biological replicates). Total RNA was extracted using the Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. RNA quality and integrity were assessed prior to downstream library preparation (P. M. Dang et al., 2009). Strand-specific cDNA libraries were then prepared and sequenced on a NovaSeq X Plus platform to generate paired-end reads.

Raw reads were trimmed of adapters and low-quality bases using fastp (v0.23.4) (S. Chen et al., 2018). Clean reads were aligned to the *Arachis hypogaea* reference genome (Tifrunner.gnm2.J5K5) using HISAT2 (v2.2.1) (Bertioli et al., 2019b; Kim et al., 2015). Transcripts were assembled for each sample and merged into a non-redundant set using StringTie (v2.1.1) (Pertea et al., 2015). Gene expression levels were normalized as Transcripts Per Million (TPM), and raw counts were extracted using the prepDE.py script for downstream analysis. Differential expression analysis was performed using DESeq2 (v1.36.0, Love et al., 2014). Low-abundance genes (counts < 10 in fewer than three samples) were pre-filtered. Differentially expressed genes (DEGs) were identified between drought and control treatments within each genotype and time point. Significant DEGs were defined by an FDR-adjusted p-value < 0.05 and an absolute $\log_2$(fold-change) > 1.

Weighted gene co-expression network analysis

To identify gene co-expression modules associated with drought stage and water-use strategy, we performed weighted gene co-expression network analysis (WGCNA) using the R package WGCNA (v1.74) (Langfelder & Horvath, 2008).

Genotypes were classified into three water-use strategies: drought-sensitive (ds), water saver (ws), and water spender (wsp). Gene expression levels were normalized using a variance-stabilizing transformation (VST). For the primary drought-progression network, only drought-treated samples from W1, W3, and W5 were included. After removing zero-variance genes, the top 6,000 most variable genes were retained. Candidate soft-thresholding powers were evaluated as 1-10 and even integers from 12 to 30. A signed network was constructed using signed adjacency and signed TOM. Power selection followed a predefined fallback rule: when no candidate reached the target scale-free topology fit ($R^2 \geq 0.80$), the final power was chosen as the value ≥ 6 with mean connectivity ≥ 20 that maximized the scale-free fit index. This yielded softPower = 30 for the primary analysis. Modules were detected using blockwiseModules with deepSplit = 3, minModuleSize = 30, and mergeCutHeight = 0.20. A sensitivity analysis was performed using softPower = 14 with all other parameters fixed. Module eigengenes were correlated with time_numeric (encoded as W1=1, W3=3, W5=5), genotype class indicators (class_ds, class_ws, class_wsp), and genotype indicators. Nominal module-trait p-values were Benjamini-Hochberg (BH) adjusted separately within each trait column. For selected key modules, stage contrasts (W1 vs W3, W1 vs W5, W3 vs W5) were tested by pairwise t-tests with BH correction.

Physiology linkage analysis

To evaluate whether co-expression modules were associated with physiological response, module eigengenes from the drought-only WGCNA network (W1, W3, W5) were correlated with physiological traits, including photosynthesis (A) and stomatal conductance (gs), across 45 drought-treated samples. Pearson correlations were performed for each module-trait pair, and p-values were adjusted using the Benjamini-Hochberg procedure within each trait column.

Module eigengene projection and recovery testing

To assess whether drought-defined co-expression modules recovered upon rewatering, module eigengenes were recalculated for all 60 drought-treated samples spanning W1, W3, W5, and RC using the gene-to-module assignments established in

the W1/W3/W5 network. For each of the 22 non-grey modules, a linear mixed model was fitted:

$$ME \sim timepoint \times class + (1|genotype)$$

where timepoint (W1, W3, W5, RC) and class (ws, wsp, ds) were treated as fixed factors with their interaction, and genotype was modeled as a random intercept to account for within-class genetic correlation. Three planned contrasts were extracted per module and class: RC versus W5 (post-drought recovery), RC versus W1 (return to baseline), and W5 versus W1 (cumulative drought effect). Degrees of freedom were approximated using Satterthwaite's method (lmerTest v3.1) (Kuznetsova et al., 2017), and p-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure across all module-contrast-class combinations (198 tests). Class-level pairwise comparisons at the RC timepoint were performed separately and FDR-corrected. All mixed-model analyses were conducted in R using lme4 and emmeans (Bates et al., 2015).

DEG-based recovery gene classification

As a complementary, module-independent validation, gene-level recovery status was classified following the previous study ((Ritchie et al., 2015).Using the VST-transformed expression matrix, differential expression among drought timepoints was tested with on the 45 drought-treated samples at W1, W5, and RC (Ritchie et al., 2015). The model included timepoint and genotype as fixed effects (no class term), and timepoint contrasts were adjusted for genotype. Differential expression was summarized with two contrasts (W5 versus W1 and RC versus W1) using FDR < 0.05 and |log2FC| ≥ 1: Recovered (significant in W5 vs W1 only), Persistent (significant in both contrasts), Recovery-specific (significant in RC vs W1 only), and Unchanged (significant in neither contrast). Gene-level classes were then intersected with WGCNA module membership to obtain module-level recovery composition profiles.

Transcriptomic Recovery Analysis

Transcriptomic recovery was evaluated by projecting the recovery stage (RC) samples, alongside W1 and W5 samples, onto the established module structure. Mean

MEs were computed for each water-use class. The degree of recovery for each module was quantified using a Recovery fraction, calculated as follows:

$$Recovery\ fraction = \frac{|\mathrm{W5}\ -\ \mathrm{W1}|\ -\ |\mathrm{RC}\ -\ \mathrm{W1}|}{|\mathrm{W5}\ -\ \mathrm{W1}|}$$

where W1, W5, and RC represent the mean module eigengene values at the baseline, severe drought, and post-rewatering stages, respectively. Based on the residual distance to the predrought baseline (W1), expression trajectories were categorized into four patterns: Recovered, partially recovered, Persistent stress, or Overshoot/Rewired.

Hub gene identification

To prioritize candidate hub genes within the strategy-associated modules MEcyan and MEtan, two gene-level metrics were computed across the 45 drought-treated samples used in WGCNA construction. Module membership (kME) was defined as the Pearson correlation between each gene's expression profile and its module eigengene. Gene significance for water-spender status (GS_wsp) was defined as the Pearson correlation between each gene's expression and the binary class_wsp trait indicator, which was selected because class_wsp showed the strongest module-level association among all class indicators for both MEcyan ($r = -0.687$, FDR = 4.43×10^{-6}) and MEtan ($r = 0.580$, FDR = 3.47×10^{-4}). Candidate genes were first filtered to the top 30% by |kME| within each module, then ranked by a composite score: score = $0.6 \times$ scaled(|kME|) + $0.4 \times$ scaled(|GS_wsp|), where scaled means min-max normalization within the candidate pool. The top five annotated genes per module were retained. Hub candidates were additionally cross-referenced DEG sets that identified between drought and control treatments.

4.3 Results

RNA-seq data quality and alignment

A total of 120 RNA-seq libraries (5 cultivars × 2 treatments × 4 time points × 3 biological replicates) were sequenced. Properly paired alignment rates were uniformly high, ranging from 91.3% to 99.0% (mean = 98.1%), indicating consistent mapping

quality across all libraries (Figure 4.1A). Library size distributions were broadly consistent across samples and showed no obvious batch-specific outliers (Figure 4.1B).

Physiological response to progressive drought

Across all five cultivars, progressive drought reduced photosynthesis, stomatal conductance, and electron transport rate, whereas intrinsic water-use efficiency (WUEi) increased from W1 to W5 (Figure 4.2). In the strategy-focused comparison based on drought-stage means, AU17 showed consistently higher photosynthesis and stomatal conductance than both water savers (AU16 and Line8), with pairwise differences meeting the pre-specified $P < 0.10$ threshold. In contrast, PI50 did not differ significantly from either water saver for photosynthesis or stomatal conductance (all $P > 0.10$). Thus, in this field season, AU17 provided the clearest water-spender physiological contrast, whereas PI50 showed weaker separation from saver phenotypes.

PCA of photosynthesis, stomatal conductance, WUEi, and electron transport rate showed that PC1 explained 81.0% of variance and PC2 explained 14.1% (Figure 4.3). The loading vectors indicated that photosynthesis and stomatal conductance contributed in the same direction (negative PC1), whereas WUEi loaded in the opposite direction (positive PC1), defining the major physiological trade-off captured by PC1. Across stages, drought and control samples were already separable at W1, diverged further at W3, reached the largest displacement at W5, and showed partial reconvergence at RC.

Differential Expression Dynamics Across Drought Stages and Cultivars

Differential expression analysis (drought vs. control; 20 genotype-by-stage contrasts) revealed distinct genotype-dependent temporal trajectories (Figure 4.4). At W1, DEG counts were relatively low across cultivars (273 to 1,330) (Figure 4.4A). By W3, DEG responses increased in all genotypes, with the strongest responses in line8 (10,568), PI50 (8,573), and AU17 (7,220), compared with PI39 (4,057) and AU16 (1,137). At W5, transcriptomic responses diverged among genotypes. AU16 and PI50 reached peak DEG levels (16,984 and 14,940, respectively), PI39 and line8 remained elevated (10,924 and 9,162), whereas AU17 declined to 820. At the recovery stage,

PI50 exhibited 241 DEGs, AU16 and AU17 showed similarly low DEG counts (2,814 and 2,522), whereas line8 and PI39 displayed higher DEG counts (11,493 and 17,546, respectively). Across genotypes and stages, up-regulated DEGs generally exceeded down-regulated DEGs (Figure 4.4B).

WGCNA identifies 22 co-expression modules spanning drought stage

To further characterize the structure of DEG temporal variation across genotypes, co-expression analysis was performed to partition expression patterns into stage-associated and genotype-associated components. WGCNA of 6,000 highly variable genes across 45 drought-treated samples spanning three drought stages (W1, W3, and W5) identified 22 non-grey co-expression modules, ranging in size from 31 genes (darkgreen) to 2,558 genes (turquoise), with an additional 651 genes assigned to the grey set (Supplementary Table S4.1).

Drought stage dominates the transcriptomic landscape

Module-trait correlations indicated that co-expression patterns were dominated by drought stage (time_numeric), while a smaller subset of modules showed associations with water-use strategy class (class_wsp), suggesting an additional, biologically meaningful dimension of transcriptomic variation (Figure 4.5). MEpink showed the strongest positive association with drought stage ($r = 0.979$, FDR = 8.510×10^{-30}), while MEyellow ($r = -0.899$, FDR = 5.85×10^{-16}) and MEblue ($r = -0.890$, FDR = 2.05×10^{-15}) showed similarly strong inverse associations. Overall, 12 modules were associated with drought stage at FDR < 0.001, indicating that progressive drought is the dominant driver of transcriptome remodeling in this drought-only design (Figure 4.5, Supplementary Table S4.2).

Two strategy modules distinguish water saver and water spender independently of drought stage

In contrast to the strong effects of drought stage, a limited number of modules were specifically associated with variation in water-use strategy independent of time. Notably, two modules exhibited significant correlations with the water spender classification (class_wsp) but showed no detectable association with drought stage.

MEcyan (63 genes) was negatively correlated with class_wsp (r = -0.687, FDR = 4.43 × 10^{-6}) and not associated with time (r = 0.110, FDR = 0.531), whereas MEtan (81 genes) was positively correlated with class_wsp (r = 0.580, FDR = 3.47 × 10^{-4}) and also showed no association with time (r = 0.180, FDR = 0.286) (Figure 4.5B, Supplementary Table S4.2). These findings indicate that, despite the dominant influence of drought stage on global co-expression structure, specific modules capture transcriptional variation linked to intrinsic water-use strategies rather than drought duration itself.

Physiology linkage analysis

Physiology linkage analysis showed that drought-stage modules were significantly associated with these traits. MEpink was negatively correlated with both photosynthesis and stomatal conductance, whereas MEblue and MEyellow were positively correlated with both traits. In contrast, strategy-associated modules (MECyan and METan) were not significantly correlated with the physiological traits (Figure 4.6).

GO/KEGG analysis

To better understand the biological functions of the key WGCNA modules, we performed GO and KEGG enrichment analyses for both drought-stage associated modules (MEpink, MEblue, and MEyellow) and strategy associated modules (MEcyan and MEtan). MEpink was enriched for membrane-associated transport and signaling functions, including transmembrane transport, transmembrane transporter activity, and protein kinase/phosphorylation terms, and corresponding genes showed progressive increases from W1 to W5. MEblue was enriched for photosynthesis-related functions, including channel activity, photosynthesis, photosystem I, and photosynthesis-antenna proteins, with genes showing consistent downregulation across drought stages. MEyellow was enriched for catalytic and oxidoreductase-related GO functions, including monooxygenase activity and iron ion binding, and its associated genes predominantly showed decreasing trajectories; no KEGG pathways were significantly enriched for this module. Together, these results indicate coordinated activation of stress-responsive transport/signaling programs alongside repression of photosynthetic and metabolic programs as drought intensified (Figure 4.7; Figure 4.8; Supplementary Table S4.3).

In contrast to the strong drought stage modules, MEcyan and MEtan captured transcriptional variation associated with water-use strategy rather than progression time. MEcyan was negatively correlated with class_wsp (r = -0.687, FDR = 4.43×10^{-6}) and showed no significant association with time_numeric (r = 0.110, FDR = 0.531), whereas MEtan was positively correlated with class_wsp (r = 0.580, FDR = 3.47×10^{-4}) and was not significantly associated with time (r = 0.180, FDR = 0.286) (Figure 4.5; Supplementary Table S4.2). Class-stratified eigengene profiles indicated that MEcyan was lowest in the water-spender class (wsp) relative to ds and ws, whereas MEtan showed an increase from drought-sensitive to water-spender classes (Figure 4.9). Functionally, MEcyan was enriched for translation-related processes (including ribosome-associated terms), whereas MEtan was enriched for proteolysis-related functions, particularly cysteine-type peptidase activity (Figure 4.10). Together, these results indicate strategy-linked transcriptional programs that are distinct from the shared drought-stage response.

Recovery Contrasts of Key Drought-Stage and Strategy Modules

Module eigengenes from the W1/W3/W5 network were projected onto RC samples to test reversibility after rewatering (Figure 4.11). The drought-stage modules showed distinct recovery behavior. MEpink (drought-induced; r = 0.979 with time) showed no significant RC_vs_W5 change in any class (ws: -0.028, FDR = 0.145; wsp: +0.008, FDR = 0.720; ds: -0.041, FDR = 0.133), while RC_vs_W1 remained strongly positive in all classes (all FDR < 10^{-13}), indicating persistent activation after rewatering. MEyellow (drought-repressed; r = -0.899) behaved similarly: RC_vs_W1 remained significantly negative in all classes (all FDR < 10^{-11}), with only a modest ds rebound from W5 to RC (RC_vs_W5 = +0.077, FDR = 0.012). In contrast, MEblue (photosynthesis-associated; r = -0.890) showed significant positive RC_vs_W5 shifts in all classes (ws: +0.161, FDR = 8.1×10^{-6}; wsp: +0.162, FDR = 7.8×10^{-6}; ds: +0.135, FDR = 0.006), indicating a post-rewatering rebound of this photosynthesis-associated module.

Strategy modules showed limited drought-recovery fluctuation. For MEcyan, only ws displayed a transient drought increase (W5_vs_W1 = +0.086, FDR = 0.004) followed

by reversal at RC (RC_vs_W5 = -0.126, FDR = 4.8×10^{-5}), while RC_vs_W1 was non-significant in all classes (all FDR > 0.21). For MEtan, only wsp showed a modest drought increase (W5_vs_W1 = +0.094, FDR = 0.015) with reversal after rewatering (RC_vs_W5 = -0.104, FDR = 0.008), and no residual RC_vs_W1 difference (FDR = 0.841). Together, these results indicate persistent drought imprinting in MEpink/MEyellow, partial reversibility of the photosynthesis-associated MEblue program, and largely time-independent behavior of class-associated strategy modules, with only limited transient class-specific shifts during drought and recovery.

DEG-Based Recovery Classification Across WGCNA Modules

To provide module-independent support for eigengene-based recovery patterns, we performed DEG-based recovery classification across all tested genes. This analysis identified 3,094 Recovered genes, 1,823 Persistent genes, and 2,097 Recovery-specific genes among 49,952 tested genes (Figure 4.12). Mapping these classes to WGCNA modules was concordant with the module-level mixed-model results: MEpink and MEyellow were strongly enriched for Persistent genes (82.9% and 82.8%), whereas MEblue showed a mixed composition (47.0% Persistent, 39.4% Recovered), consistent with partial reversibility after rewatering.

Strategy-associated modules remained largely static at the gene level. MEcyan and MEtan were clustered by Unchanged genes (92.1% and 91.4%), with no Persistent genes in MEcyan and only four in MEtan, supporting limited drought-recovery dynamics in these modules. Beyond the five primary modules, MEroyalblue and MElightgreen showed high Recovered fractions, whereas MElightcyan and MEgreenyellow showed high Recovery-specific fractions (Figure 4.12).

Gene significance polarization confirms module coherence

Gene-level GS_wsp profiles showed strong sign consistency within both strategy modules. In MEcyan, all 63 genes had negative GS_wsp values (60/63 with $p < 0.05$), whereas in MEtan all 81 genes had positive GS_wsp values (79/81 with $p < 0.05$) (Supplementary Table S4.4, Supplementary Table S4.5). This same-sign pattern is

consistent with MEcyan and MEtan representing internally coherent, strategy-associated expression modules.

Hub genes of MEcyan and MEtan

Joint ranking by |kME| and |GS_wsp| identified five annotated hub candidates in each strategy module (Supplementary Table S4.6). MEcyan candidates showed high module membership (kME = 0.893-0.920) with negative GS_wsp (-0.815 to -0.539), whereas MEtan candidates showed similarly high module membership (kME = 0.924-0.969) with positive GS_wsp (0.495-0.583). Ah09g269200 (MEcyan; score = 0.814) and Ah11g354500 (MEtan; score = 0.806) were the top-ranked candidates. Of the ten candidates, eight were detected as significant DEGs in at least one genotype-by-timepoint drought-versus-control comparison. Heatmap profiling of these eight genes across drought-treated samples showed that MEtan candidates were generally elevated at later drought stages, while MEcyan candidates displayed the opposite tendency with greater genotype dependence (Figure 4.13). Together, these genes define a focused set of high-priority targets for downstream validation.

4.4 Discussion

Drought adaptation in peanut is agronomically critical in the southeastern United States, where rainfall variability and limited irrigation infrastructure constrain production stability. In Alabama, where only 10% farms have access to irrigation, genotype-level differences in water-use behavior can directly determine yield stability under mid-season stress. Our previous work found that both water saver and water spender cultivars can express drought tolerance, but through distinct physiological strategies (Zhang et al., 2022). This study integrates physiology and transcriptomics to identify the molecular programs underlying drought progression, water-use strategy divergence, and post-drought recovery in peanut.

Across genotypes, drought progression was accompanied by decreases in photosynthesis and stomatal conductance, together with an increase in WUEi from W1 to W5. This pattern is consistent with established drought physiology in peanut and other W3 crops (Soba et al., 2024). PCA supported this finding that photosynthesis and

stomatal conductance loaded in the opposite direction to WUEi along PC1, and sample separation along this axis primarily reflected drought severity rather than cultivar. Differential expression analysis further highlighted strong temporal and genotypic heterogeneity. DEG counts were low at W1, increased significantly by W3, and diverged at W5 and RC depending on cultivar. AU16 and PI50 showed very high DEG burdens at W5, AU17 showed a dramatic reduction at W5 despite high W3 response, and PI39 showed the highest DEG burden at RC. These patterns indicate that the temporal dynamics of drought response and recovery differ substantially among genotypes. However, relying solely on DEG abundance is insufficient to separate drought stage responses from genotype-associated regulatory structures, necessitating subsequent co-expression analysis to further these complex networks.

WGCNA identified 22 non-grey modules and showed that drought stage was the dominant axis of co-expression organization. MEpink was positively associated with drought stage, whereas MEyellow and MEblue were negatively associated. Functional enrichment matched these trajectories, with MEpink enriched for transport/signaling functions and MEblue enriched for photosynthesis-related terms, while MEyellow was enriched for oxidoreductase-associated processes. Physiology linkage analysis provided independent support, as MEpink correlated negatively with photosynthesis and stomatal conductance, whereas MEblue and MEyellow correlated positively with both traits. Together, these results indicate that major stage-associated modules capture biologically meaningful physiological decline during progressive drought [3,5,6,17,18,26].

A second signal was captured by strategy-associated modules. MEcyan and MEtan were strongly associated with class_wsp but not with drought stage, indicating strategy-linked variation that was largely independent of progression time. Functional enrichment distinguished these modules from drought-stage modules, with translation/ribosome terms in MEcyan and proteolysis, including cysteine-type peptidase activity, in MEtan. Gene-level GS patterns showed coherent sign direction within both modules, supporting internal consistency of module behavior.

Recovery analyses further supported these WGCNA findings. Stage-associated modules showed differential reversibility after rewatering: MEpink and MEyellow remained significantly divergent from W1, consistent with persistent drought imprinting, whereas MEblue showed a significant rebound from W5 toward W1, indicating partial recovery. In contrast, strategy-associated modules (MEcyan and MEtan) showed limited W5-to-RC fluctuation and were enriched for Unchanged genes in DEG-based recovery classification. These results indicate that drought-stage modules are dynamically responsive across stress and recovery, while strategy-associated modules remain comparatively stable within the timeframe of this experiment.

Hub prioritization within MEcyan and MEtan translated module-level findings into tractable candidate genes. Joint ranking by |kME| and |GS_wsp| identified five candidates per module, and most were also detected in at least one genotype-by-timepoint drought versus control DEG contrast. These candidates are suitable for targeted validation because they combine network centrality with trait association. At the same time, hub status is correlative. Functional validation remains necessary before mechanistic or breeding deployment. Integration with QTL/GWAS evidence and multi-environment phenotyping will be important for translational use.

While these results provide critical insights, certain limitations must be acknowledged. Primarily, as W1 serves as an early drought reference rather than a true well-watered baseline, recovery (RC) dynamics reflect a partial biological reset toward the W1 state rather than total stress alleviation. Furthermore, functional studies are partially limited by current annotation bottlenecks in peanut. Nevertheless, the combined evidence supports a two-component model in which a dominant stage-driven transcriptional program is superimposed by strategy-associated regulatory variation. This framework provides a practical basis for combining dynamic stress-response traits with stable strategy-linked molecular markers in peanut improvement programs.

List of Figures

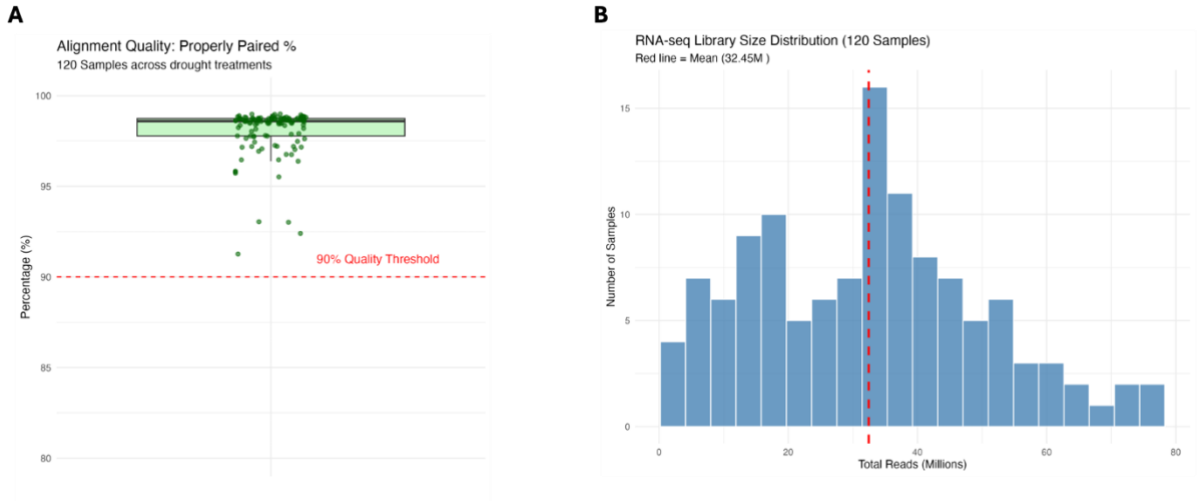

Figure 4.1: A. Alignment quality: distribution of properly paired read percentages across 120 samples. B. Library size distribution (total reads in millions) across 120 RNA-seq libraries

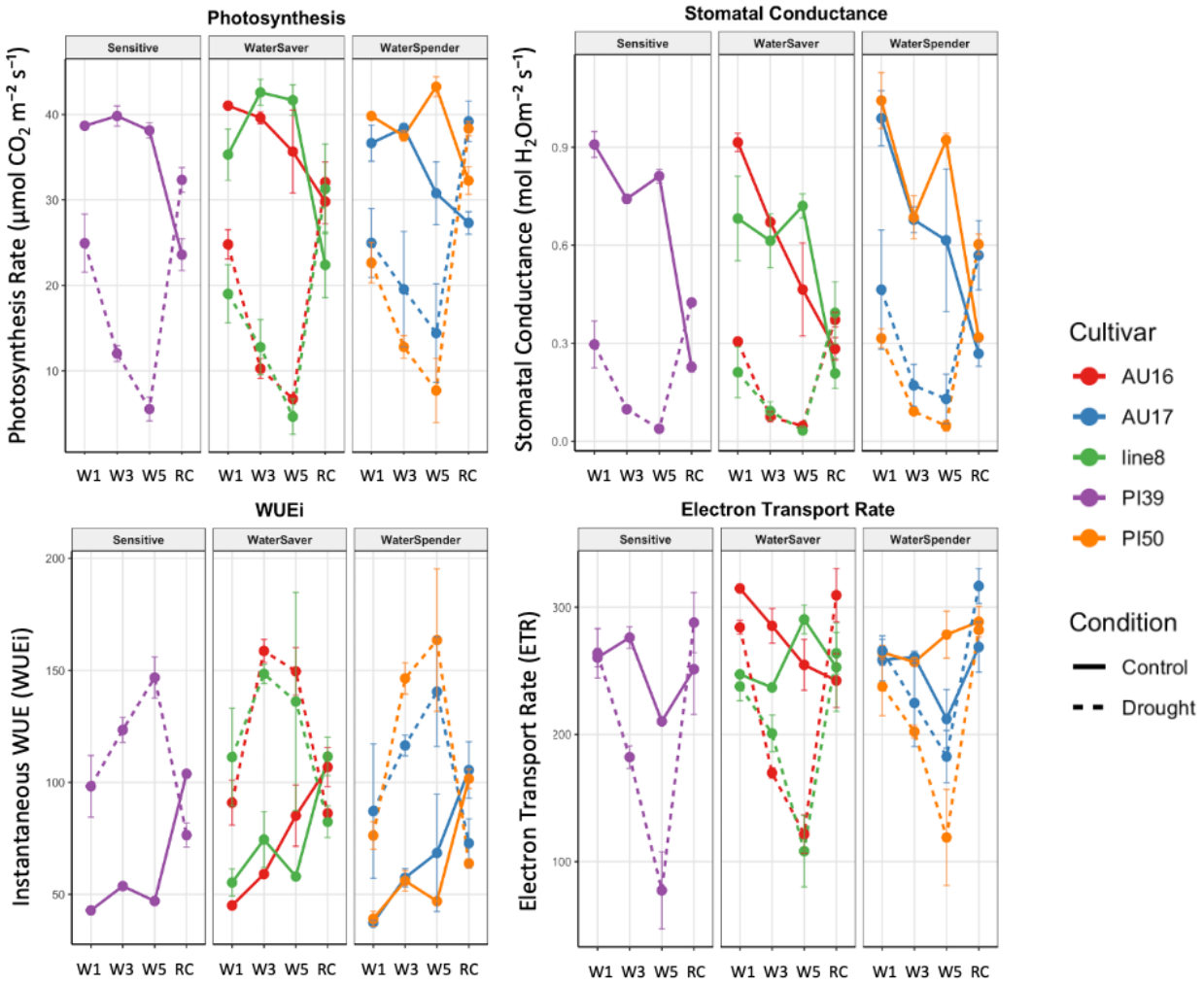

Figure 4.2. Time-series of four core gas-exchange traits under control and drought across five peanut cultivars. Panels are faceted by water-use strategy group. Error bars represent ± SE (n = 3).

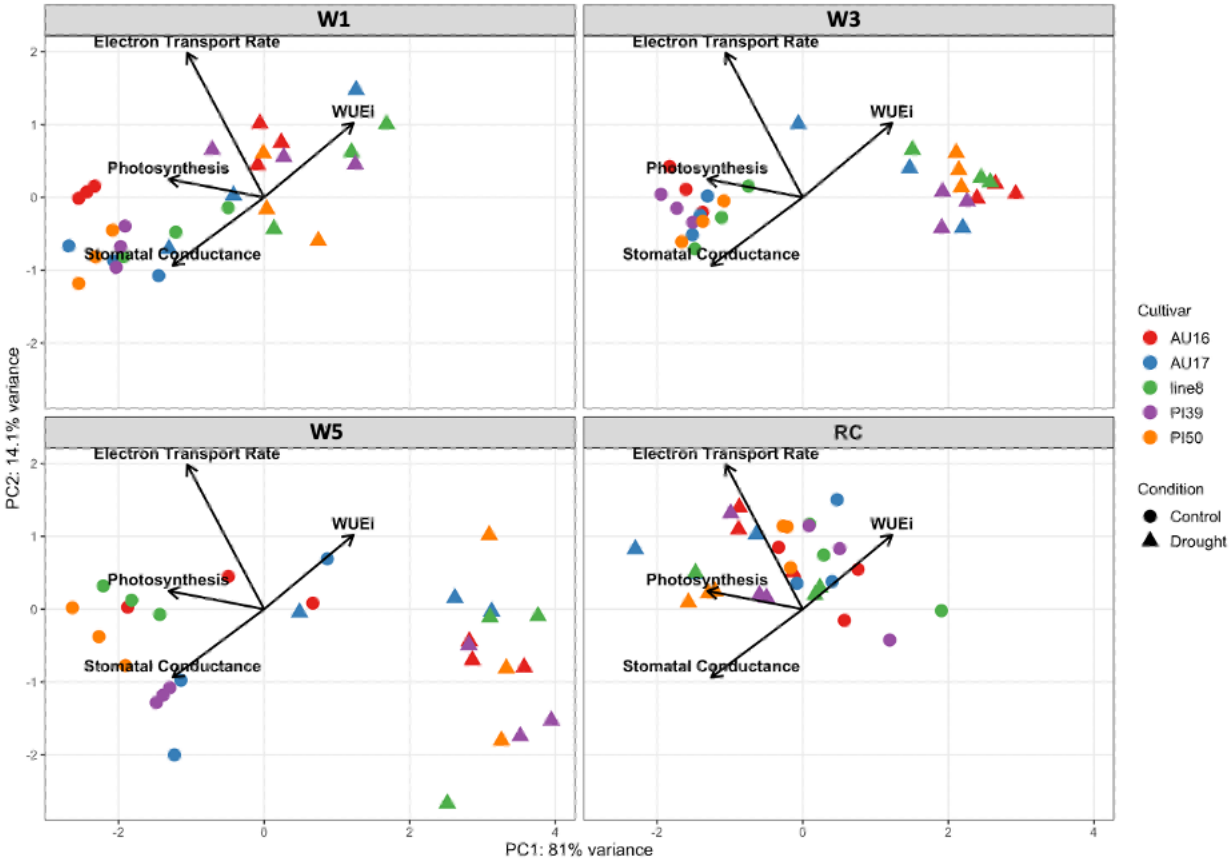

Figure 4.3. Stage-faceted PCA biplot of physiological traits under control and drought treatments.

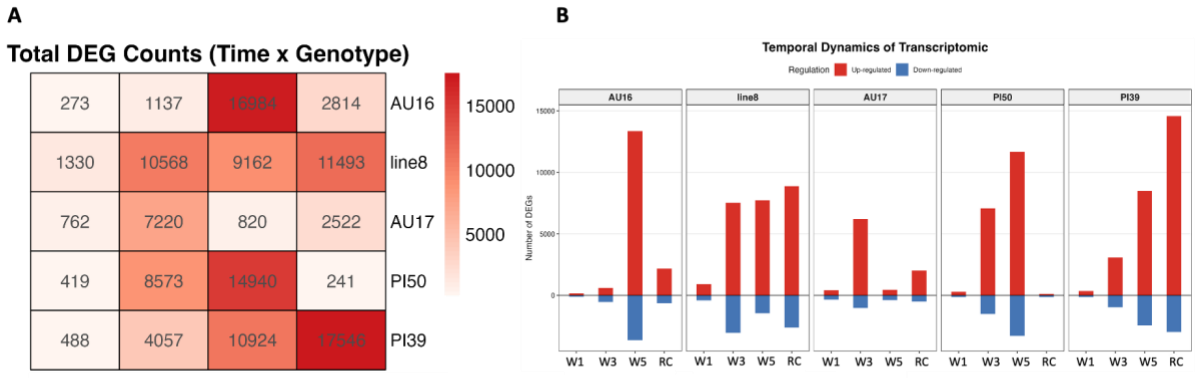

Figure 4.4. Differential expression genes (DEGs) across drought versus control (20 contrasts: 5 cultivars × 4 time points), FDR < 0.05 and $|\log_2$ fold change$| > 1$.
(A) Heatmap of total DEGs per contrast. (B) Temporal pattern of transcriptomic response for each cultivar. (B) Temporal pattern of transcriptomic response for each

cultivar. Bars show up-regulated DEGs (red, above zero) and down-regulated DEGs (blue, below zero).

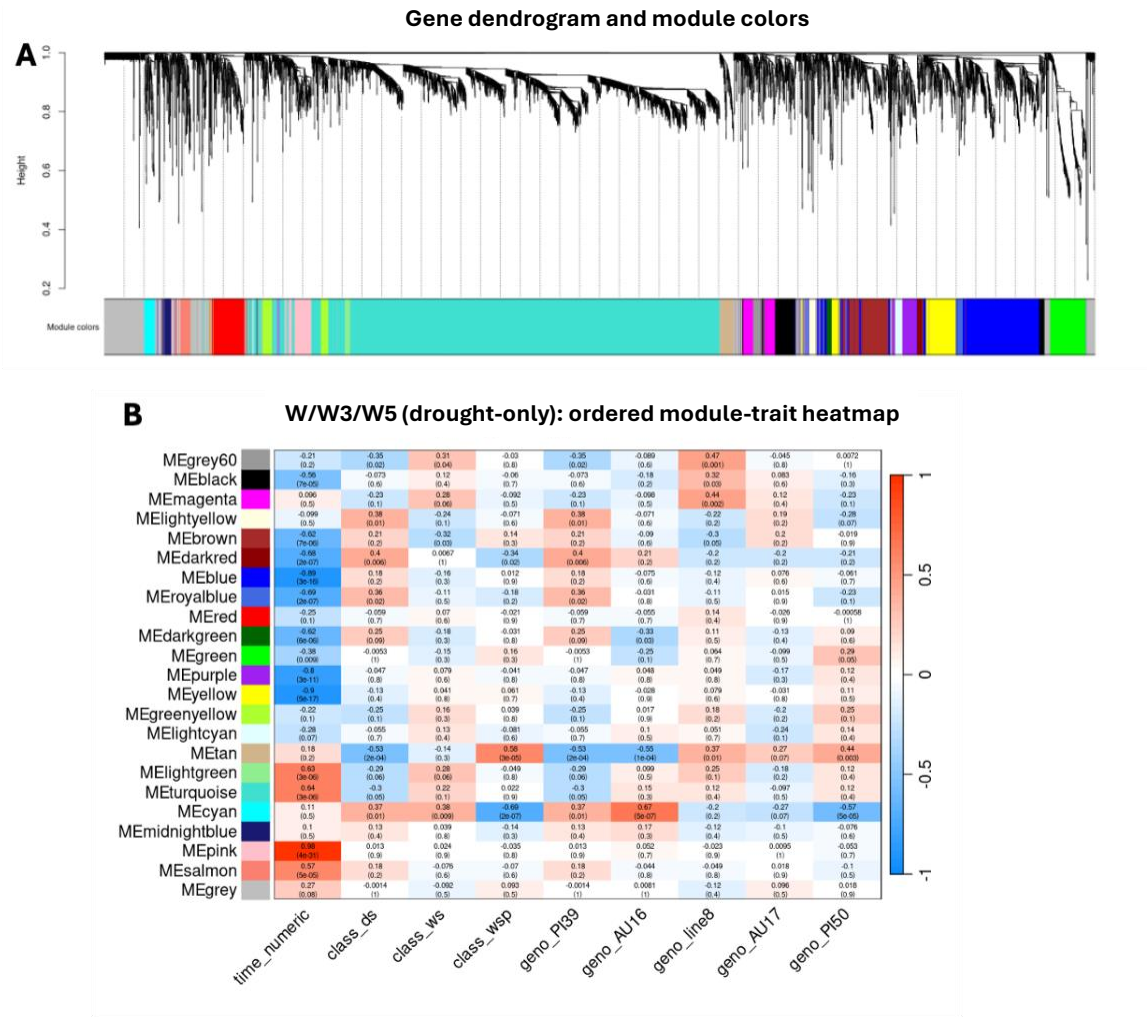

Figure 4.5. A. Gene co-expression dendrogram and module assignment in the drought-only network. 5B. Ordered module-trait correlation heatmap for drought stage and water-use strategy traits.

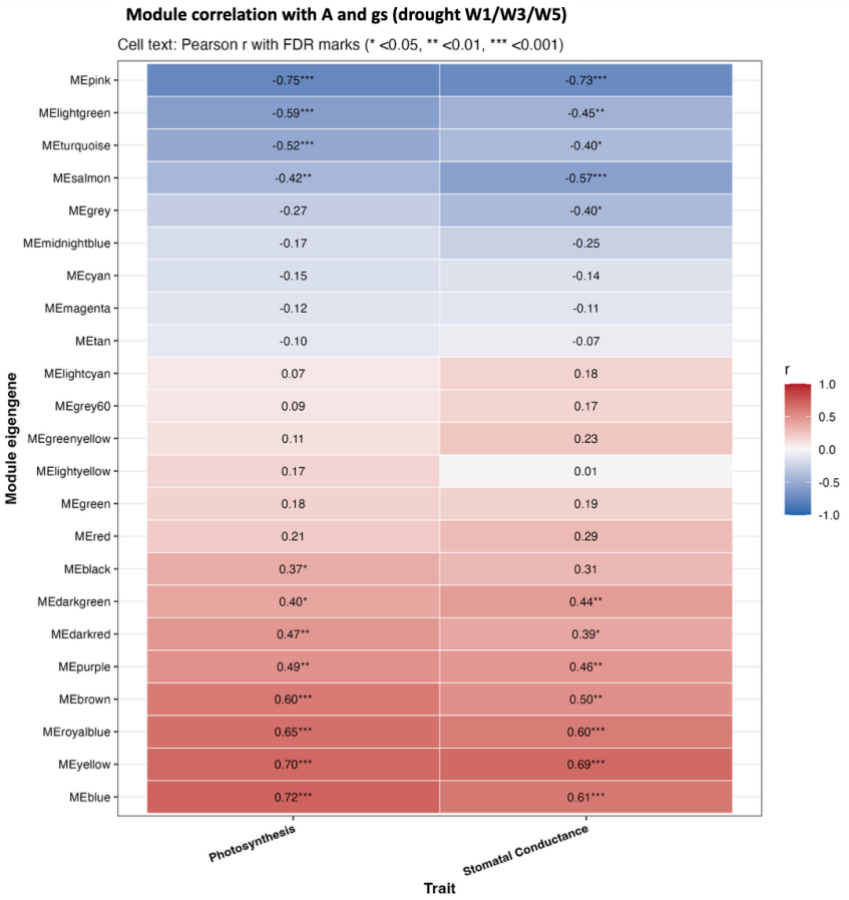

Figure 4.6. Module-trait correlation heatmap linking WGCNA modules to gas-exchange traits under drought. Heatmap showing Pearson correlations between module eigengenes (rows) and photosynthesis (A) and stomatal conductance (gs) (columns) across drought-treated samples at W1, W3, and W5. Cell colors represent correlation direction and magnitude (blue, negative; red, positive), and cell labels show correlation coefficients (r). Asterisks indicate Benjamini-Hochberg FDR significance within each trait column (*FDR < 0.05, ** FDR < 0.01, *** FDR < 0.001).

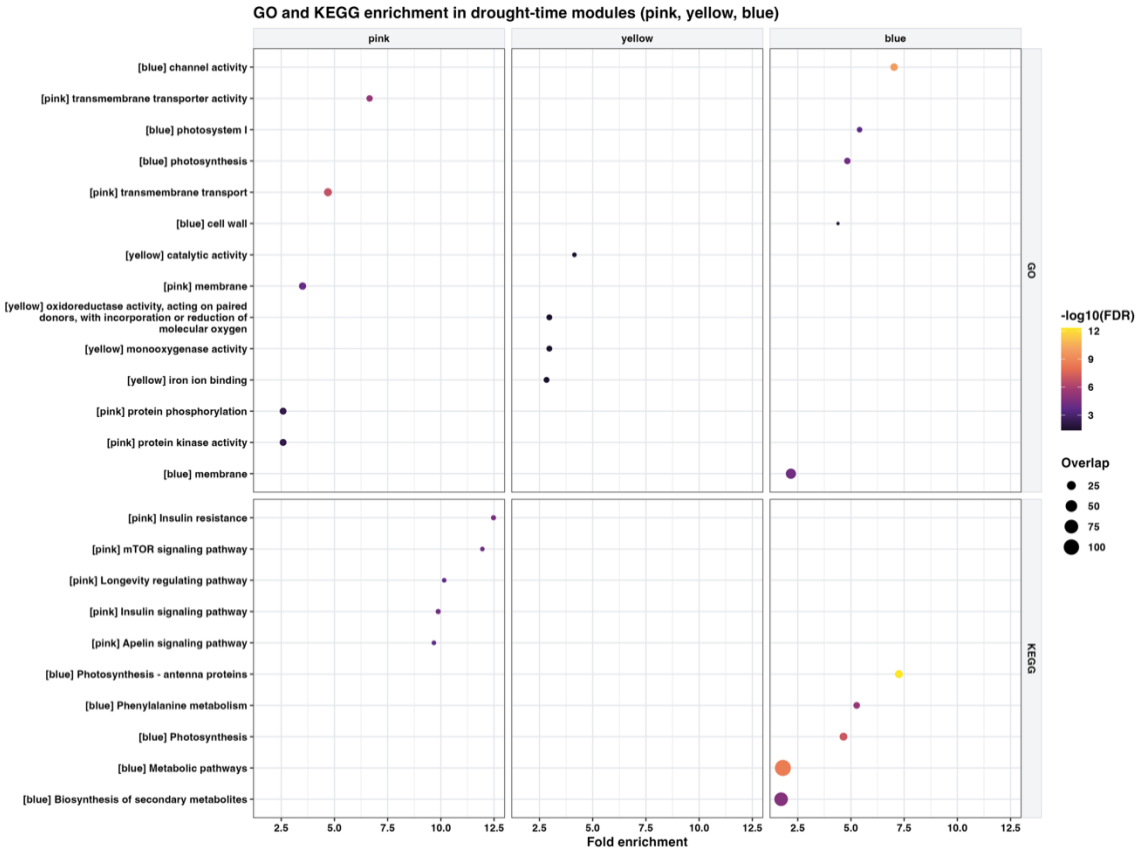

Figure 4.7. GO and KEGG enrichment of drought-progression modules (MEpink, MEyellow, and MEblue) identified by WGCNA. Dot position indicates fold enrichment, dot size indicates overlap gene count, and dot color indicates enrichment significance (-log10(FDR)). MEpink is enriched for transport/signaling-related functions, MEblue for photosynthesis-related processes, and MEyellow for catalytic and oxidoreductase-associated functions.

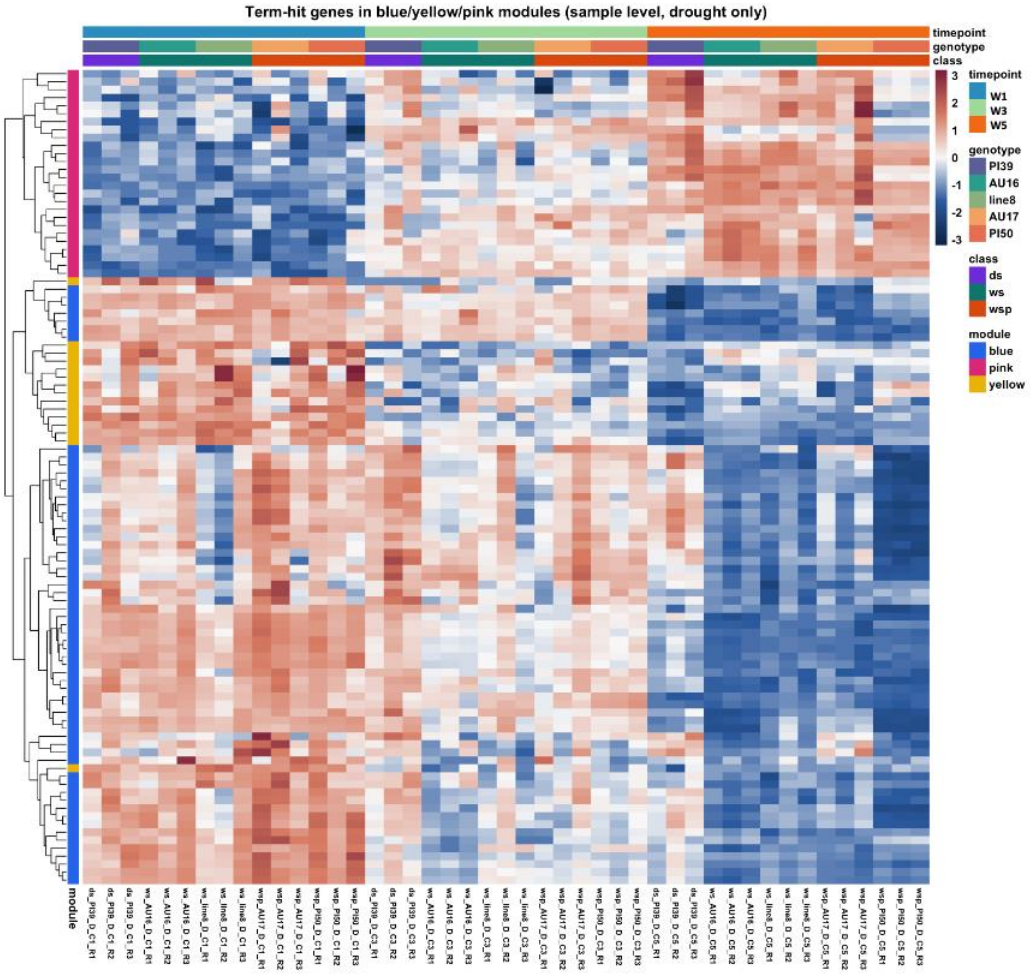

Figure 4.8. Expression profiling of drought-responsive module genes across developmental stages. > Heatmap showing the row-scaled expression (z-scores) of representative genes from the MEpink, MEyellow, and MEblue modules across drought stages (W1, W3, and W5). Genes were selected based on top-ranked GO and KEGG enrichment terms. Columns represent individual biological replicates, clustered by sample metadata (class, genotype, and timepoint). Rows are hierarchically clustered using Euclidean distance to reflect expression similarity.

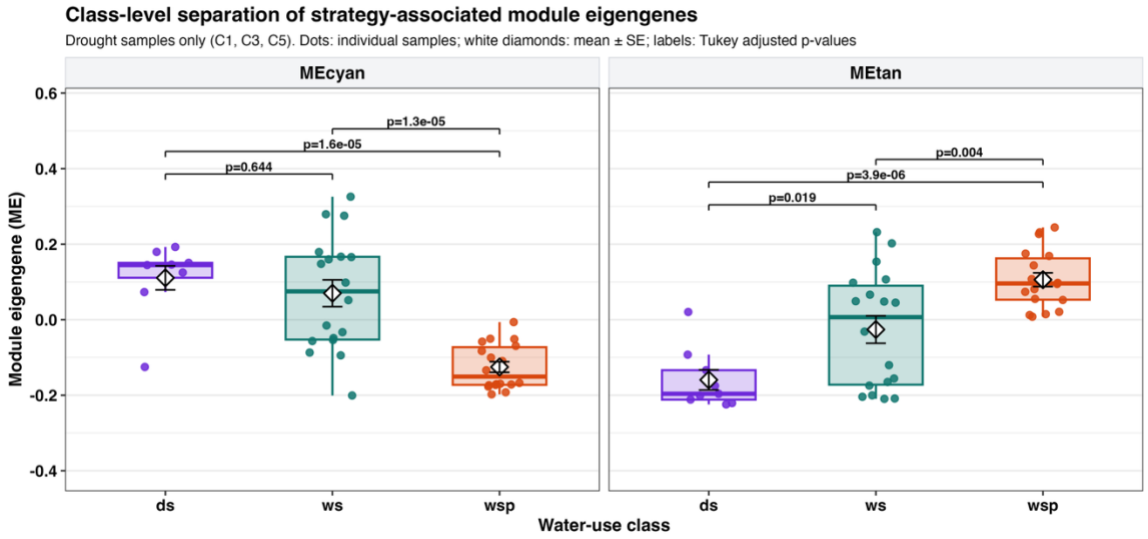

Figure 4.9 Class-level separation of strategy-associated module eigengenes under drought. Boxplots show module eigengene (ME) distributions for MEcyan and MEtan across drought-sensitive (ds), water-saver (ws), and water-spender (wsp) classes using drought-treated samples from W1, W3, and W5. Colored dots represent individual samples, boxes indicate the interquartile range with median lines, and white diamonds indicate class means ± SE. Horizontal brackets show Tukey-adjusted pairwise p-values.

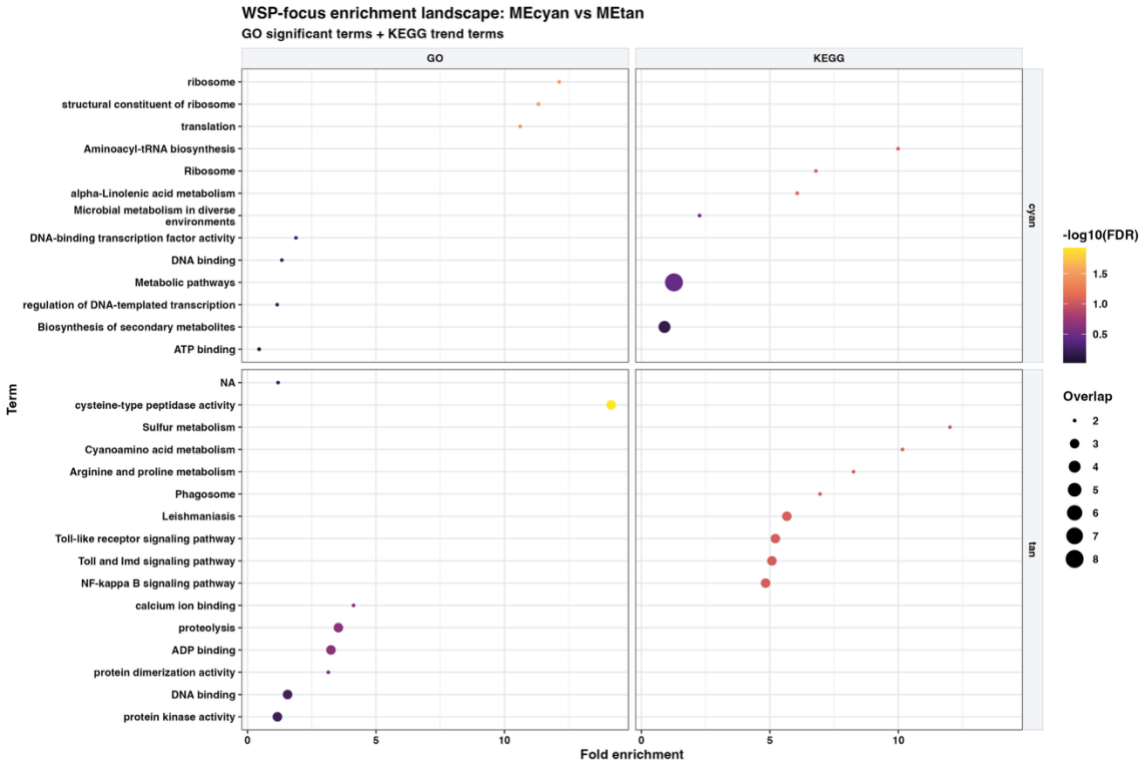

Figure 4.10. GO and KEGG enrichment landscape of strategy-associated modules (MEcyan and MEtan) in drought-treated samples. Dot plots show top enriched terms for each module, faceted by module (cyan, tan) and database (GO, KEGG). The x-axis

indicates fold enrichment, dot size indicates overlap gene count, and dot color indicates enrichment strength (-log10(p)).

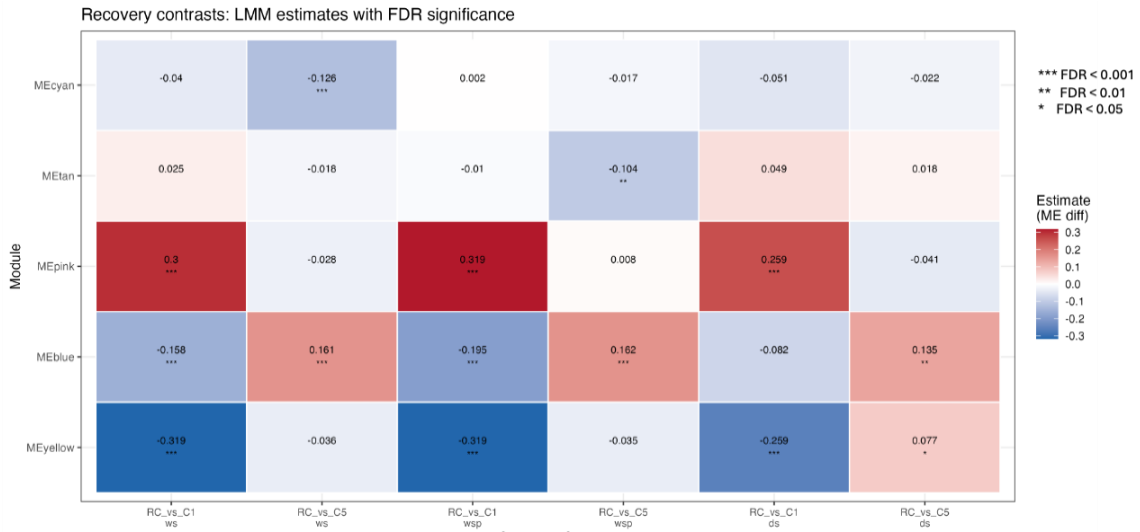

Figure 4.11. Recovery contrasts of key drought-stage and strategy modules after rewatering. Heatmap of linear mixed-model contrast estimates for module eigengenes (MEs) across water-use classes (ws, wsp, ds).

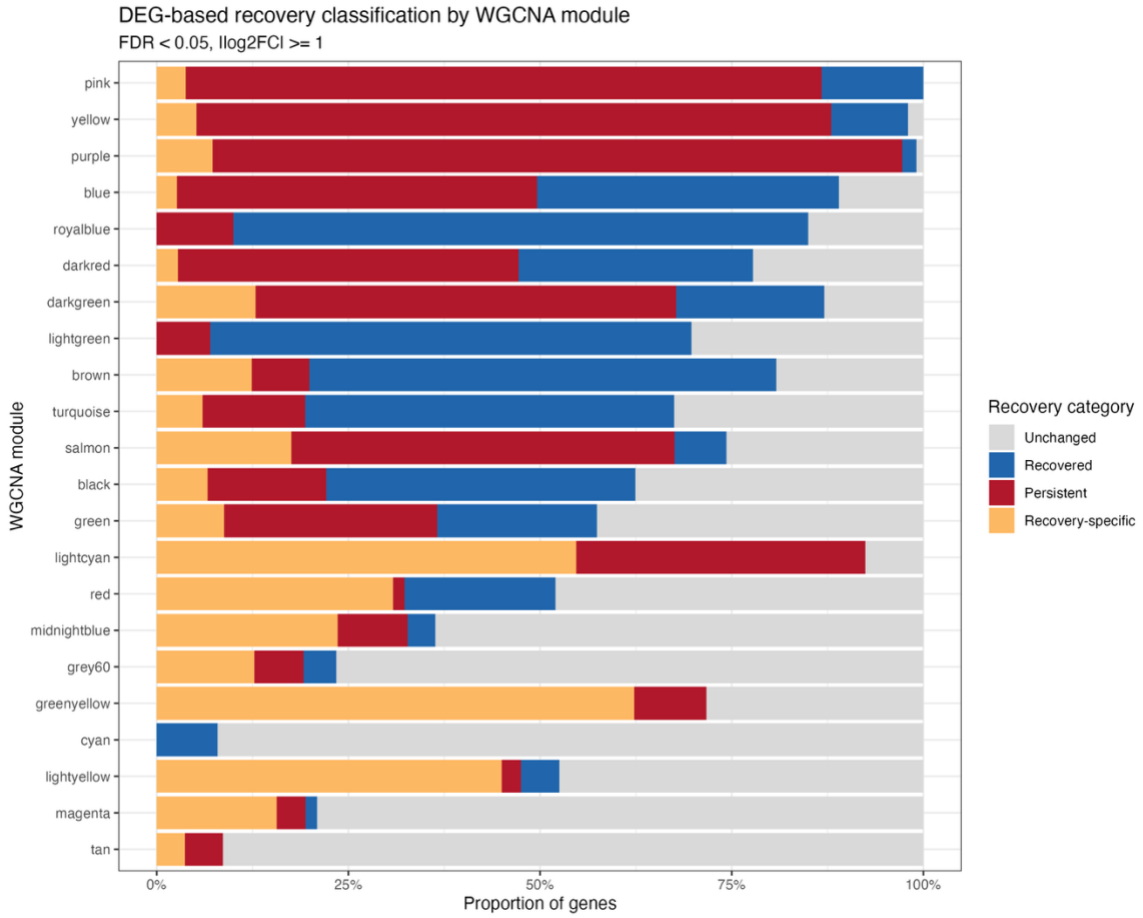

Figure 4.12. DEG-based recovery classification by WGCNA module. Stacked bar plot showing, for each non-grey WGCNA module, the proportion of genes classified as Unchanged (gray), Recovered (blue), Persistent (red), or Recovery-specific (orange). Gene classes were defined from limma contrasts in drought-treated samples (W1, W5, RC; n = 45) using FDR < 0.05 and |log2FC| ≥ 1: Recovered, significant in W5 vs W1 only; Persistent, significant in both W5 vs W1 and RC vs W1; Recovery-specific, significant in RC vs W1 only; Unchanged, significant in neither contrast.

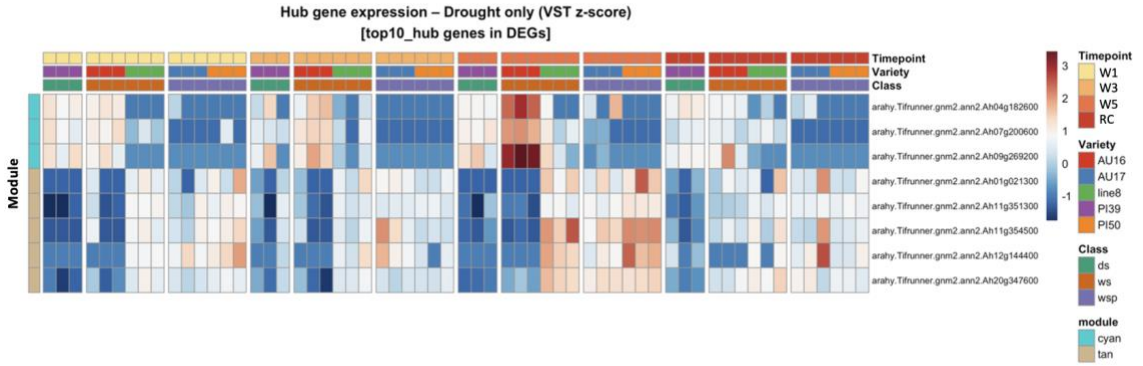

Figure 4.13. Expression heatmap of hub gene candidates under drought.

Reference

Araus, J. L., Slafer, G. A., Reynolds, M. P., & Royo, C. (2002). Plant Breeding and Drought in W3 Cereals: What Should We Breed For? Annals of Botany, 89(7), 925. https://doi.org/10.1093/aob/mcf049

Balota, M., McGrath, S., Isleib, T. G., & Tallury, S. (2012). Transpiration Response to Vapor Pressure Deficit in Field Grown Peanut. Peanut Science, 39(1), 53–61. https://doi.org/10.3146/PS11-13.1

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/JSS.V067.I01

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., … Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nature Genetics 2019 51:5, 51(5), 877–884. https://doi.org/10.1038/s41588-019-0405-z

Bhogireddy, S., Xavier, A., Garg, V., Layland, N., Arias, R., Payton, P., Nayak, S. N., Pandey, M. K., Puppala, N., & Varshney, R. K. (2020). Genome-wide transcriptome and physiological analyses provide new insights into peanut drought response mechanisms. Scientific Reports, 10(1), 4071. https://doi.org/10.1038/s41598-020-60187-z

Blum, A. (2005). Drought resistance, water-use efficiency, and yield potential - Are they compatible, dissonant, or mutually exclusive? Australian Journal of Agricultural Research, 56(11), 1159–1168. https://doi.org/10.1071/AR05069

Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. Field Crops Research, 112(2-3), 119-123. https://doi.org/10.1016/j.fcr.2009.03.009

Boyer, J. S. (1982). Plant productivity and environment. Science (New York, N.Y.), 218(4571), 443–448. https://doi.org/10.1126/science.218.4571.443

Chaves, M. M., Maroco, J. P., & Pereira, J. S. (2003). Understanding plant responses to drought - from genes to the whole plant. Functional Plant Biology : FPB, 30(3), 239–264. https://doi.org/10.1071/FP02076

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics (Oxford, England), 34(17), i884–i890. https://doi.org/10.1093/BIOINFORMATICS/BTY560

Dang, P. M., Rowland, D. L., Faircloth, W. H., Dang, P. M., Rowland, D. L., & Faircloth, W. H. (2009). Comparison of ELISA and RT-PCR assays for the detection of Tomato spotted wilt virus in peanut. Peanut Science, 36(2), 133–137. https://doi.org/10.3146/PS08-024.1

Dang, P., Patel, J., Sorensen, R., Lamb, M., & Chen, C. Y. (2024). Genome-Wide Association Analysis Identified Quantitative Trait Loci (QTLs) Underlying Drought-Related Traits in Cultivated Peanut (Arachis hypogaea L.). Genes, 15(7). https://doi.org/10.3390/GENES15070868

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., & Basra, S. M. A. (2009). Plant drought stress: effects, mechanisms and management. Agronomy for Sustainable Development 2009 29:1, 29(1), 185–212. https://doi.org/10.1051/agro:2008021

Flexas, J., & Medrano, H. (2002). Drought-inhibition of photosynthesis in W3 plants: stomatal and non-stomatal limitations revisited. Annals of Botany, 89(2), 183–189. https://doi.org/10.1093/aob/mcf027

Gundaraniya, S. A., Ambalam, P. S., Budhwar, R., Padhiyar, S. M., & Tomar, R. S. (2023). Transcriptome analysis provides insights into the stress response in cultivated peanut (Arachis hypogaea L.) subjected to drought-stress. Molecular Biology Reports, 50(8), 6691–6701. https://doi.org/10.1007/S11033-023-08563-6

Hamidou, F., Halilou, O., & Vadez, V. (2013). Assessment of Groundnut under Combined Heat and Drought Stress. Journal of Agronomy and Crop Science, 199(1), 1–11. https://doi.org/10.1111/j.1439-037X.2012.00518.x

Jyostna Devi, M., Sinclair, T. R., Vadez, V., & Krishnamurthy, L. (2009). Peanut genotypic variation in transpiration efficiency and decreased transpiration during

progressive soil drying. Field Crops Research, 114(2), 280–285. https://doi.org/10.1016/J.FCR.2009.08.012

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods 2015 12:4, 12(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kong, K., Xu, M., Wu, L., Zhou, H., Wang, R., Zhao, T., Wang, C., & Song, Y. (2025). Genome-wide association study and transcriptome analysis reveal candidate genes related to drought stress in the germination stage of soybean. Frontiers in Plant Science, 16, 1621869. https://doi.org/10.3389/fpls.2025.1621869

Kottapalli, K. R., Rakwal, R., Shibato, J., Burow, G., Tissue, D., Burke, J., Puppala, N., Burow, M., & Payton, P. (2009). Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. Plant, Cell & Environment, 32(4), 380–407. https://doi.org/10.1111/J.1365-3040.2009.01933.X

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559. https://doi.org/10.1186/1471-2105-9-559

Lawlor, D. W. (2002). Limitation to photosynthesis in water-stressed leaves: stomata vs. metabolism and the role of ATP. Annals of Botany, 89 Spec No(7), 871–885. https://doi.org/10.1093/aob/mcf110

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L. S. P., & Qin, F. (2015). A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nature Communications, 6. https://doi.org/10.1038/ncomms9326

Nakhforoosh, A., Bodewein, T., Fiorani, F., & Bodner, G. (2016). Identification of water use strategies at early growth stages in durum wheat from shoot phenotyping and physiological measurements. Frontiers in Plant Science, 7(AUG2016), 211067. https://doi.org/10.3389/FPLS.2016.01155/TEXT

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 2015 33:3, 33(3), 290–295. https://doi.org/10.1038/nbt.3122

Pokhrel, S., Kharel, P., Pandey, S., Botton, S., Nugraha, G. T., Holbrook, C., & Ozias-Akins, P. (2024). Understanding the impacts of drought on peanuts (Arachis hypogaea L.): exploring physio-genetic mechanisms to develop drought-resilient peanut cultivars. Frontiers in Genetics, 15, 1492434. https://doi.org/10.3389/FGENE.2024.1492434/FULL

Polania, J.A., Poschenrieder, C., Beebe, S. and Rao, I.M., 2016. Effective use of water and increased dry matter partitioned to grain contribute to yield of common bean improved for drought resistance. Frontiers in plant science, 7, p.660. https://doi.org/10.3389/fpls.2016.00660

Polania, J. A., Salazar-Chavarría, V., Gonzalez-Lemes, I., Acosta-Maspons, A., Chater, C. C. C., & Covarrubias, A. A. (2022). Contrasting Phaseolus Crop Water Use Patterns and Stomatal Dynamics in Response to Terminal Drought. Frontiers in Plant Science, 13, 894657. https://doi.org/10.3389/FPLS.2022.894657/TEXT

Ratnakumar, P., Vadez, V., Nigam, S. N., & Krishnamurthy, L. (2009). Assessment of transpiration efficiency in peanut (Arachis hypogaea L.) under drought using a lysimetric system. Plant Biology (Stuttgart, Germany), 11 Suppl 1(SUPPL.1), 124–130. https://doi.org/10.1111/J.1438-8677.2009.00260.X

Reddy, T. Y., Reddy, V. R., & Anbumozhi, V. (n.d.). Physiological responses of groundnut (Arachis hypogea L.) to drought stress and its amelioration: a critical review.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47–e47. https://doi.org/10.1093/NAR/GKV007

Sinclair, T. R., Purcell, L. C., King, C. A., Sneller, C. H., Chen, P., & Vadez, V. (2007). Drought tolerance and yield increase of soybean resulting from improved symbiotic

N2 fixation. Field Crops Research, 101(1), 68–71. https://doi.org/10.1016/j.fcr.2006.09.010

Soba, D., Parker, S., Chen, C., Shekoofa, A., & Sanz-Saez, A. (2024). Peanut photosynthesis response to drought can include diffusive and biochemical limitations depending on cultivar. Physiologia Plantarum, 176(4). https://doi.org/10.1111/PPL.14489

Wang, D., Liu, X., He, G., Wang, K., Li, Y., Guan, H., Wang, T., Zhang, D., Li, C., & Li, Y. (2025). GWAS and transcriptome analyses unravel ZmGRAS15 regulates drought tolerance and root elongation in maize. BMC Genomics, 26(1). https://doi.org/10.1186/s12864-025-11435-x

Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., Yang, X., & Qin, F. (2016). Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. Nature Genetics 2016 48:10, 48(10), 1233–1241. https://doi.org/10.1038/ng.3636

Wang, X., Yang, X., Feng, Y., Dang, P., Wang, W., Graze, R., Clevenger, J. P., Chu, Y., Ozias-Akins, P., Holbrook, C., & Chen, C. (2021). Transcriptome Profile Reveals Drought-Induced Genes Preferentially Expressed in Response to Water Deficit in Cultivated Peanut (Arachis hypogaea L.). Frontiers in Plant Science, 12, 645291. https://doi.org/10.3389/FPLS.2021.645291/TEXT

Zhang, J., Meng, Q., Sanz-Saez, A., & Chen, C. (2025). Genome-wide identification and expression analysis reveals the drought-response MAPK genes in peanut (Arachis hypogaea L.). The Plant Genome, 18(4). https://doi.org/10.1002/TPG2.70166

Zhang, Q., Dang, P., Chen, C., Feng, Y., Batchelor, W., Lamb, M., & Sanz-Saez, A. (2022). Tolerance to mid-season drought in peanut can be achieved by high water use efficiency or high efficient use of water. Crop Science, 62(5), 1948–1966. https://doi.org/10.1002/CSC2.20806;WEBSITE:WEBSITE:ACSESS.ONLINELIBRARY.WILEY.COM;WGROUP:STRING:PUBLICATION