Legume Genomics and Genetics 2026, Vol.17, No.1, 32-48
http://cropscipublisher.com/index.php/lgg

Research Report Open Access

Transcriptomic Insights into Drought Stress Response in Soybean (*Glycine max* L.)

Weiguo Lu ✉, Lijun Qiu
Modern Agricultural Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China
✉ Corresponding email: lijun.qiu@cuixi.org

Legume Genomics and Genetics, 2026 Vol.17, No.1 doi: 10.5376/lgg.2026.17.0003
Received: 11 Feb., 2026
Accepted: 14 Feb., 2026
Published: 24 Mar., 2026

Copyright © 2026 Lu and Qiu, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:
Lu W.G., and Qiu L.J., 2026, Transcriptomic insights into drought stress response in soybean (*Glycine max* L.), Legume Genomics and Genetics, 17(1): 32-48 (doi: 10.5376/lgg.2026.17.0003)

Abstract Drought stress is a major abiotic stress factor that limits soybean growth and development, yield formation, and quality stability; its impact spans multiple critical stages, including germination, vegetative growth, reproductive development, and grain filling. With the rapid advancement of high-throughput sequencing technologies, transcriptomics has emerged as a crucial tool for deciphering the molecular mechanisms underlying drought tolerance in soybeans. Focusing on the theme of "transcriptomic analysis of soybean responses to drought stress," this review begins by outlining the physiological basis of soybean drought responses. It then systematically summarizes the workflow, data analysis methods, and strategies for identifying differentially expressed genes using RNA-Seq technology in drought tolerance research. Particular emphasis is placed on summarizing the regulatory roles of transcription factor families-such as WRKY, NAC, bZIP, and DREB-in drought signal transduction, osmotic adjustment, antioxidant defense, and hormonal responses. Furthermore, by integrating studies on non-coding RNAs, co-expression networks, and candidate gene functional validation, this review explores the structural characteristics of multi-layered gene regulatory networks involved in soybean drought responses. The review also utilizes comparative transcriptomic analyses of soybean accessions with varying degrees of drought tolerance to analyze the expression patterns of key drought-tolerance genes and assess their potential for application in molecular breeding. Overall, transcriptomics has not only deepened our understanding of the mechanisms underlying soybean responses to drought stress but has also provided a theoretical foundation for the discovery of drought-tolerance genes, the development of molecular markers, and precision breeding. Future research should prioritize the integration of multi-omics data with field-based phenotyping to enhance the depth of mechanistic analysis and improve the efficiency of breeding applications.

Keywords Soybean; Drought stress; Transcriptomics; Differentially expressed genes; Drought-tolerance breeding

1 Introduction

Soybean (*Glycine max* L.) is a major source of plant protein and oil, underpinning global food, feed, and industrial markets. However, its productivity is highly vulnerable to water deficits, especially in rainfed production systems and regions where climate change is intensifying the frequency and duration of drought events (Haghpanah et al., 2024). Drought stress disrupts plant water relations, reduces photosynthesis, alters carbon allocation, and ultimately depresses biomass accumulation and yield (Razi and Muneer, 2021). Controlled and field studies in soybean have shown that reductions in soil moisture below about 75% of field capacity lead to decreases in leaf area index, growth rate, and seed yield, with particularly strong sensitivity during reproductive stages (Wang et al., 2025). Yield penalties of 50–80% have been reported when severe drought occurs at flowering, pod setting, or seed filling, often associated with reductions in pod number, seed number, and 100-seed weight (Wei et al., 2018; Aziez et al., 2022;). In addition to quantitative yield losses, drought alters seed quality by modifying protein and oil content, which further compromises the economic value of the crop (Poudel et al., 2023).

Drought responses in soybean are stage-dependent and genotype-dependent, adding complexity to breeding and management. Experiments across multiple environments indicate that the flowering and grain-filling stages are particularly vulnerable, with grain-filling drought often causing the largest yield reductions and quality shifts (Demirtas et al., 2010). Even within modern commercial cultivars adapted to temperate regions such as Ontario, Canada, substantial genetic variation exists for drought tolerance indices based on yield under stress relative to well-watered conditions, reflecting differences in water use, biomass maintenance, and pod retention (Gebre et al., 2022). Physiologically, drought reduces leaf water potential and stomatal conductance, impairs chlorophyll content and gas exchange, and induces oxidative damage, as evidenced by elevated malondialdehyde levels and enhanced activities of antioxidant enzymes such as superoxide dismutase and peroxidase (Sekhurwane et al., 2025; Wang et al., 2025). While agronomic and physiological studies have clarified how and when drought damages soybean, the underlying molecular programs that orchestrate these responses-and that distinguish tolerant from sensitive genotypes-are only partially understood.

Over the past two decades, substantial progress has been made in elucidating the molecular mechanisms of plant drought stress responses in model and crop species. Drought is now recognized as a complex stimulus that triggers interconnected signaling networks involving Ca^{2+} fluxes, reactive oxygen species (ROS), and phytohormones, most prominently abscisic acid (ABA) (Razi and Muneer, 2021). ABA-dependent and ABA-independent pathways converge on a wide array of transcription factors (TFs), including members of the DREB/CBF, NAC, bZIP, MYB, and WRKY families, which regulate suites of stress-responsive genes involved in osmotic adjustment, cell protection, and growth modulation (Aslam et al., 2022; Geng et al., 2024; Yue et al., 2025). These downstream genes encode late embryogenesis abundant (LEA) proteins, dehydrins, heat shock proteins, antioxidant enzymes, and transporters that collectively mitigate dehydration, membrane damage, and metabolic perturbation (Haghpanah et al., 2024). Multi-level regulation further includes post-transcriptional and epigenetic controls, such as alternative splicing, microRNAs, and DNA or RNA methylation, which fine-tune transcript stability and translation under fluctuating stress conditions (Fan et al., 2025).

Insights from rice, sorghum, tomato, and other crops highlight the conservation and diversification of drought regulatory networks. In rice and other cereals, key drought-responsive genes and pathways have been extensively catalogued, including ABA signaling modules, MAPK cascades, and transcriptional networks that modulate root architecture, stomatal behavior, and chloroplast function (Fan et al., 2025; Yue et al., 2025). Reviews integrating physiological and molecular perspectives emphasize that drought tolerance is a polygenic, systems-level trait, shaped by signaling crosstalk among hormones and by networked gene modules acting at signaling, transcriptional, and effector levels (Razi and Muneer, 2021; Haghpanah et al., 2024). In soybean, functional studies such as CRISPR/Cas9 editing of specific transcription factors (e.g., GmHdz4) have demonstrated that targeted manipulation of drought-related regulatory genes can significantly enhance tolerance, underlining the translational potential of molecular knowledge for breeding (Ali et al., 2025). Despite these advances, many components of soybean-specific drought regulatory circuits and their dynamic coordination across development remain to be defined.

Transcriptomics has emerged as a central tool to dissect plant drought responses at genome-wide scale. RNA sequencing (RNA-seq) enables quantitative profiling of gene expression changes across tissues, developmental stages, and contrasting genotypes, allowing the identification of differentially expressed genes (DEGs), coexpression modules, and enriched pathways under water deficit (Fracasso et al., 2016; Privitera et al., 2024; Yue et al., 2025). Comparative transcriptome studies in sorghum, tomato, Brassica, and other species have revealed that drought induces large transcriptional reprogramming in pathways such as phenylpropanoid and flavonoid biosynthesis, starch and sucrose metabolism, amino acid metabolism, ROS detoxification, and hormone signaling (Privitera et al., 2024; Yue et al., 2025). In soybean, integrated transcriptomic and metabolomic analyses under PEG-simulated drought at the seedling stage have demonstrated that tolerance involves coordinated adjustments in TCA cycle activity, glycolysis, secondary metabolism (especially flavonoids and isoflavones), and multiple hormone pathways, including ABA, auxin, gibberellin, and brassinosteroids (Wang et al., 2022). These studies not only identify candidate genes and metabolites associated with drought resistance, but also highlight cultivar-specific strategies for maintaining growth under water deficit (Wang et al., 2022).

As sequencing technologies advance and costs decline, transcriptomics is increasingly being combined with other omics layers-metabolomics, proteomics, and even chromatin accessibility assays such as ATAC-seq-to build systems-level models of drought adaptation (Privitera et al., 2024; Shen et al., 2025). Multi-omics integration has uncovered key regulatory hubs, such as TFs or signaling components whose expression patterns tightly correlate with protective metabolites or physiological traits, and has begun to link chromatin state to transcriptional activation of drought-responsive genes (Shen et al., 2025). For soybean, these approaches provide an unprecedented opportunity to connect field-relevant drought phenotypes with underlying gene networks, thereby accelerating the discovery of robust biomarkers and targets for breeding. The present study leverages transcriptomic analyses in soybean under drought stress to deepen understanding of its molecular response architecture, focusing on the identification of core regulatory pathways and genotype-specific expression signatures that can inform breeding strategies for enhanced drought resilience.

2 Physiological and Molecular Bases of Soybean Drought Stress Response

2.1 Physiological changes in soybean under drought stress

Drought stress triggers a cascade of physiological perturbations in soybean, affecting water status, photosynthesis, and growth. Water deficit reduces leaf water potential and relative water content, leading to stomatal closure that initially limits transpirational water loss but also constrains CO_2 uptake and photosynthetic carbon assimilation (Xu et al., 2023; Wang et al., 2025). Under progressive or prolonged drought, soybean exhibits marked declines in net photosynthetic rate, stomatal conductance, intercellular CO_2 concentration, and chlorophyll content, together with reduced leaf area, plant height, and biomass accumulation (Li et al., 2022; Wang et al., 2022). Chlorophyll fluorescence and OJIP analyses indicate impaired photosystem II efficiency, decreased electron transport, and enhanced non-photochemical quenching, reflecting both damage and photoprotective adjustments in the photosynthetic apparatus under severe water deficit (Falcioni et al., 2025). These physiological constraints are more pronounced at sensitive developmental stages, particularly flowering and grain filling, where decreases in photosynthesis and assimilate supply translate into reduced pod set, seed number, and seed weight, ultimately causing substantial yield losses (Li et al., 2022;).

Oxidative stress is a central component of the drought response. Water deficit promotes overproduction of reactive oxygen species (ROS) such as superoxide and hydrogen peroxide, inducing lipid peroxidation, protein oxidation, and membrane damage, which can be monitored via increased malondialdehyde (MDA) content and electrolyte leakage (Xuan et al., 2022; Shahriari et al., 2022). To mitigate ROS toxicity, soybean activates an antioxidant defense system, including enhanced activities of superoxide dismutase, catalase, peroxidases, and ascorbate peroxidase, alongside accumulation of non-enzymatic antioxidants and phenolic compounds (Li et al., 2022; Shahriari et al., 2022). Osmotic adjustment represents another key adaptive strategy: tolerant genotypes tend to maintain higher levels of compatible solutes such as proline, soluble sugars, and soluble proteins, helping to preserve cell turgor, stabilize proteins and membranes, and buffer redox status under drought (Figure 1) (Shahriari et al., 2022; Wang et al., 2022).

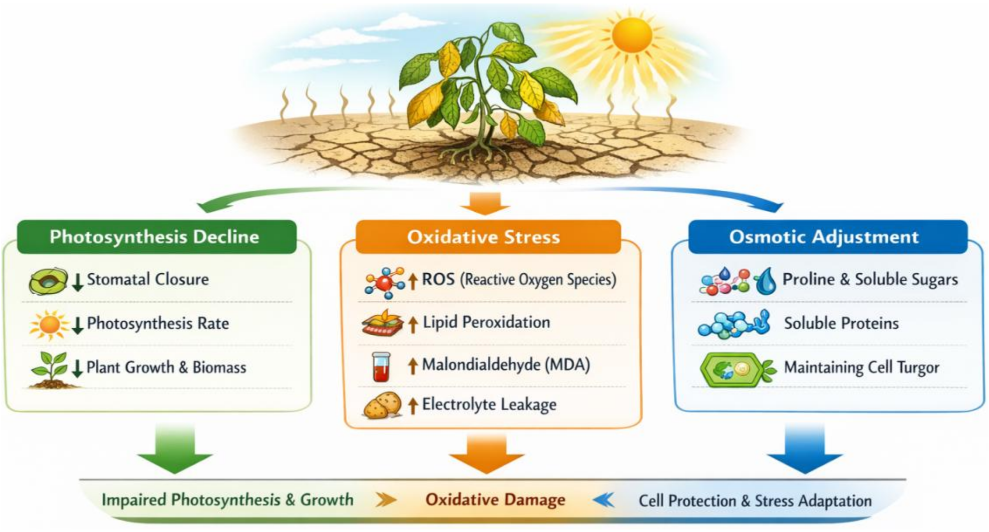

Figure 1 Physiological responses of soybean to drought stress

Comparative studies among cultivated and wild soybean germplasm show that drought-tolerant lines generally sustain better photosynthetic performance, less MDA accumulation, and more effective antioxidant and osmotic regulation than sensitive counterparts, underscoring the strong link between physiological plasticity and drought tolerance (Xu et al., 2023; Lin et al., 2024).

2.2 Plant drought resistance signal transduction pathways

The perception and transduction of drought signals in soybean rely on complex networks integrating cellular sensors, secondary messengers, and phytohormones. Water deficit is first perceived as changes in cell turgor and osmotic potential, leading to rapid calcium (Ca^{2+}) influx, ROS bursts, and alterations in membrane lipid-derived messengers such as phosphatidic acid (Soma et al., 2021). These early signals activate protein kinase cascades, including Ca^{2+}-dependent protein kinases and mitogen-activated protein kinase (MAPK) modules, which phosphorylate downstream targets to reprogram metabolism, ion fluxes, and gene expression (Soma et al., 2021; Li et al., 2022). In drought-tolerant soybean cultivars, transcriptome analyses have revealed preferential induction of signal transduction pathways such as Ca^{2+} signaling, MAPK signaling, and lipid signaling (e.g., phosphatidylinositol 4-phosphate 5-kinase), suggesting that more robust and diversified signaling underlies enhanced stress perception and response (Xu et al., 2023).

Abscisic acid (ABA) is the central hormonal regulator of drought resistance. Under water deficit, ABA accumulates in leaves and roots and binds to PYR/PYL/RCAR receptors, leading to inhibition of PP2C phosphatases and activation of subclass III SnRK2 kinases (Soma et al., 2021). Activated SnRK2s phosphorylate ion channels in guard cells to promote stomatal closure, as well as downstream transcription factors such as ABA-responsive element binding proteins/factors (AREB/ABFs), thereby coupling rapid physiological responses with transcriptional reprogramming. ABA-independent pathways function in parallel, involving subclass I SnRK2s, which modulate gene expression stability via mRNA decay and growth regulation during drought. Crosstalk between ABA and other hormones-including jasmonic acid, brassinosteroids, and ethylene-further shapes the drought response, with comparative transcriptomics in soybean showing that jasmonate- and brassinosteroid-related signaling components are particularly enriched in drought-tolerant genotypes (Xu et al., 2023). Recent work in soybean has also uncovered a circadian-linked module in which GmPRR3b modulates ABA signaling and drought responses via repression of the ABA-responsive transcription factor GmABF3, illustrating how core clock components integrate environmental water signals with temporal regulation of stress defenses (Li et al., 2024).

2.3 Regulatory mechanisms of drought stress-related gene expression

Drought-induced transcriptional reprogramming in soybean is governed by multilayered regulatory networks centered on transcription factors (TFs) that act as molecular switches. ABA-dependent regulation largely proceeds through ABRE-binding bZIP TFs (AREB/ABFs), which bind ABA-responsive elements (ABREs) in target promoters to activate genes involved in osmotic adjustment, LEA proteins, and antioxidant systems (Soma et al., 2021; Wei et al., 2024). ABA-independent regulation involves dehydration-responsive element (DRE)–binding proteins (DREB/CBF) and the broader AP2/ERF family, as well as NAC, MYB, bHLH, and WRKY TFs, which interact with DRE and other cis-elements to control suites of drought-responsive genes (Soma et al., 2021; Wei et al., 2024). Integrative systems analyses of soybean transcriptomes have identified co-expression modules enriched for these TF families, with hub genes predicted to coordinate pathways such as photosynthesis, redox homeostasis, and systemic acquired resistance under water deficit (Shahriari et al., 2022). Comparative RNA-seq studies consistently report that drought-tolerant soybean genotypes exhibit a greater number and magnitude of differentially expressed TF genes than sensitive lines, particularly in NAC, WRKY, bZIP, ERF, and NF-Y families, indicating stronger or earlier transcriptional activation of protective pathways (Shahriari et al., 2022; Xu et al., 2023; Wei et al., 2024).

Functional characterization of individual TFs has begun to clarify how these regulators confer drought tolerance. Overexpression of NAC-type TFs such as GmNAC19 and GmNAC3 enhances drought resistance by improving root development, maintaining chlorophyll content, and modulating key physiological indices, including soluble sugar, soluble protein, proline, and antioxidant enzyme activities, while reducing ROS accumulation and MDA levels (Cui et al., 2024). Similarly, genome-wide and expression analyses of soybean ERF TFs have identified GmERF205 as a positive regulator of drought tolerance, whose overexpression improves growth, root architecture, and yield under water deficit in both controlled and field conditions (Abdullah et al., 2025). At the network level, modules like GmPRR3b–GmABF3 integrate circadian, ABA, and drought signals, where GmPRR3b negatively regulates drought responses by suppressing GmABF3, and GmABF3 overexpression restores tolerance (Li et al., 2024). Beyond transcription initiation, post-transcriptional mechanisms-including microRNAs, mRNA decay pathways mediated by subclass I SnRK2s, and alternative splicing-as well as chromatin-level regulation and cis-element architecture in DEG promoters, further fine-tune drought-responsive gene expression. Together, these regulatory layers define a dynamic and interconnected gene network that enables soybean to sense drought, reconfigure its transcriptome, and deploy coordinated physiological and metabolic defenses.

3 Soybean Transcriptome Sequencing Technologies and Data Analysis Methods

3.1 Principles and experimental workflow of RNA-Seq Technology

RNA sequencing (RNA-seq) is a next-generation sequencing (NGS)–based technology that profiles the complete set of RNA molecules in a cell or tissue, providing quantitative and qualitative information on gene expression, alternative splicing, and novel transcripts. In contrast to microarrays, RNA-seq does not rely on pre-designed probes and therefore offers a broader dynamic range and the ability to detect previously unannotated genes and isoforms (Severin et al., 2010). In soybean and other plants, most drought-stress studies currently use short-read Illumina platforms, which generate millions of 50–150 bp reads that can be accurately aligned to the reference genome or transcriptome to estimate expression levels for each gene or transcript (Min et al., 2021). The core principle is straightforward: RNA molecules present in the sample at the time of extraction are converted into complementary DNA (cDNA), fragmented, and sequenced, and the number of reads mapping to each feature is taken as a proxy for its abundance, subject to normalization for gene length and sequencing depth (Stark et al., 2019).

A typical RNA-seq workflow for soybean drought studies begins with careful experimental design, including definition of stress treatments, time points, tissues, and biological replication, followed by high-quality RNA extraction from control and stressed samples (Machado et al., 2019; Min et al., 2021). After assessing RNA integrity (e.g., using Bioanalyzer RIN values), mRNA is usually enriched by poly(A) selection, though rRNA depletion or total RNA protocols are alternatives for non-polyadenylated transcripts (Conesa et al., 2016; Stark et al., 2019). The mRNA is then fragmented, reverse-transcribed into cDNA, ligated to platform-specific adapters, and amplified to construct sequencing libraries, which are subjected to high-throughput sequencing on an Illumina instrument (Chen et al., 2016). Emerging long-read platforms such as PacBio and Oxford Nanopore are increasingly used to capture full-length transcripts and complex isoforms, complementing short-read RNA-seq by improving transcriptome annotation and isoform resolution in soybean (Stark et al., 2019; Tyagi et al., 2022; Monzó et al., 2025). Together, these technologies underpin a diverse range of soybean transcriptomic applications, from drought response profiling to expression atlas construction (Severin et al., 2010).

3.2 Transcriptome data quality control and sequence alignment

Robust downstream inference from RNA-seq requires stringent quality control (QC) at both the raw read and alignment levels. Initial QC typically uses tools such as FastQC and MultiQC to assess per-base sequence quality scores, GC content, adapter contamination, and sequence duplication rates, identifying potential technical problems introduced during library preparation or sequencing (Chen et al., 2023; Lindlöf, 2025). When necessary, adapters and low-quality bases are trimmed, and overly short reads are removed to minimize mapping errors. Post-trimming metrics such as the proportion of reads retained, updated quality score distributions, and base composition profiles are then re-evaluated to ensure that preprocessing has improved, rather than degraded, data quality (Li et al., 2014). In soybean drought studies, additional QC often includes checking for rRNA contamination, evaluating sequencing depth sufficiency for differential expression analysis, and confirming sample clustering via principal component analysis (PCA) to detect outliers or mislabeled samples (Kong et al., 2025).

High-quality reads are subsequently aligned to the soybean reference genome using splice-aware aligners such as HISAT2, STAR, or similar tools, which can correctly map reads spanning exon–exon junctions (Machado et al., 2019; Kong et al., 2025). Mapping statistics-including overall alignment rate, proportion of uniquely mapped reads, distribution across exons, introns, and intergenic regions, and coverage uniformity across gene bodies-provide essential indicators of library complexity and technical artifacts (Lindlöf, 2025). For example, soybean drought-response datasets routinely report >90% mapping rates and high Q30 scores, reflecting well-constructed libraries and appropriate reference annotations (Figure 2) (Wang et al., 2022; Xuan et al., 2022; Kong et al., 2025). Specialized QC suites such as RSeQC or related packages further evaluate GC bias, strand specificity, 5′–3′ coverage bias, and read distribution over genomic features, helping to flag over-amplification, degradation, or sample-preparation biases that might confound differential expression and functional analyses. These quality-assured alignments then serve as the basis for accurate quantification and downstream statistical modeling.

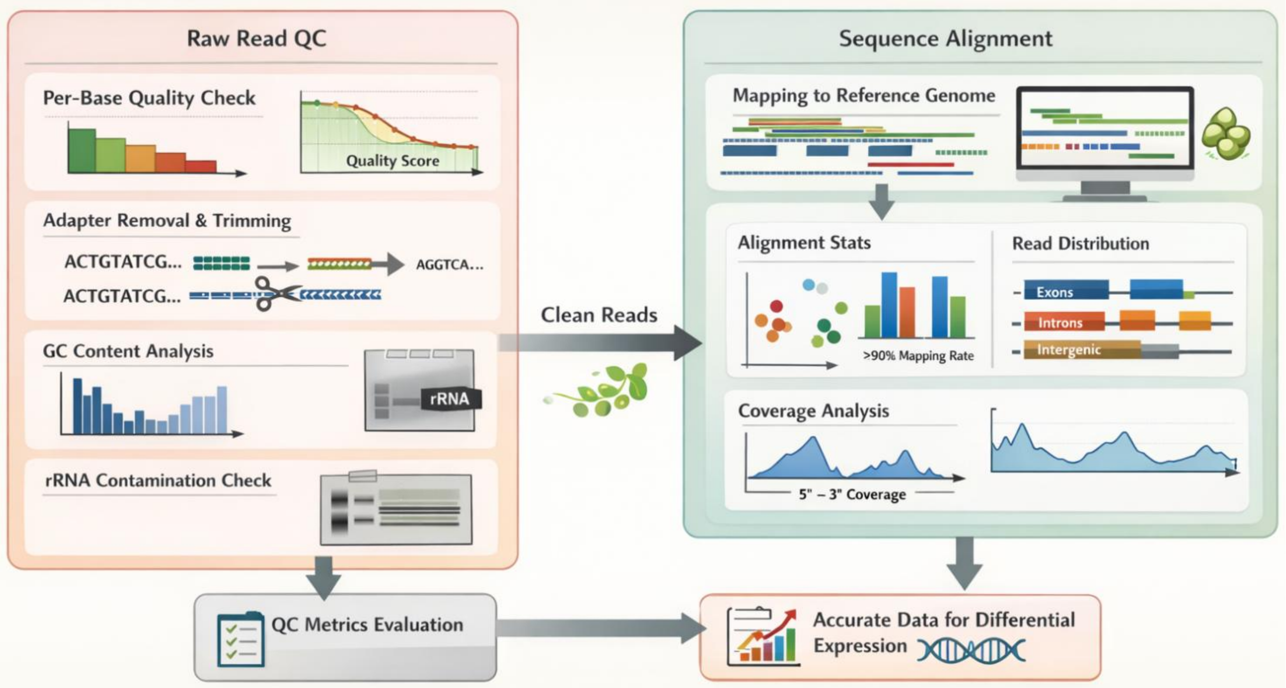

Figure 2 RNA-seq quality control and alignment workflow in soybean research (Adopted from Kong et al., 2025)

3.3 Identification and functional annotation of differentially expressed genes

Quantification of gene or transcript abundance from aligned reads is typically performed using counting tools such as featureCounts or HTSeq, or via pseudo-alignment and transcript-level methods like Salmon and Kallisto, generating a matrix of raw counts across all samples (Conesa et al., 2016; Stark et al., 2019; Chen et al., 2023). Because raw counts are influenced by library size and compositional effects, differential expression analysis relies on statistical frameworks such as DESeq2, edgeR, or related negative binomial–based models, which estimate dispersion and fit generalized linear models to identify differentially expressed genes (DEGs) between drought and control conditions (Li et al., 2022; Ludt et al., 2022). Stringent thresholds on adjusted p-values (e.g., $\text{FDR} < 0.05$) and fold-change (e.g., $|\text{log2FC}| \geq 1$) are commonly applied to define biologically meaningful DEGs, while PCA and clustering of normalized expression values help assess sample separation by treatment and genotype (Wang et al., 2022; Kong et al., 2025). In soybean drought studies, thousands of DEGs are often detected across time points and tissues, reflecting extensive transcriptional reprogramming under water deficit.

Functional interpretation of DEGs relies on integrating expression data with gene annotation resources and pathway databases. Gene Ontology (GO) enrichment analysis is routinely used to identify over-represented biological processes, molecular functions, and cellular components among up- and down-regulated genes, highlighting pathways such as hormone signaling, reactive oxygen species detoxification, osmolyte metabolism, and photosynthesis in drought responses (Wang et al., 2022; Kong et al., 2025). KEGG pathway and MapMan analyses further reveal coordinated changes in primary and secondary metabolism, including the TCA cycle, glycolysis, flavonoid and isoflavone biosynthesis, and cell wall remodeling, which have been repeatedly implicated in soybean drought tolerance (Wang et al., 2022; Kong et al., 2025). Network-based approaches, such as weighted gene co-expression network analysis (WGCNA), can identify co-expressed modules and hub genes associated with drought-tolerance traits or time-dependent responses (Li et al., 2022). By combining DEG lists with functional enrichment and network analyses, soybean RNA-seq studies have pinpointed key transcription factors, signaling components, and metabolic enzymes as candidate regulators of drought resistance, providing targets for functional validation and molecular breeding.

4 Identification and Functional Analysis of Differentially Expressed Genes in Soybean Under Drought Stress

4.1 Key differentially expressed genes induced by drought stress

Transcriptomic profiling under water deficit has revealed thousands of differentially expressed genes (DEGs) in soybean, with both drought intensity and genotype strongly shaping expression patterns. In the drought-sensitive cultivar Liaodou 15, 2214, 3684 and 2985 DEGs were identified at mild, moderate and severe drought at flowering, respectively, indicating progressive transcriptional reprogramming as stress intensifies (Li et al., 2022). Many of these DEGs are involved in antioxidant defenses, secondary metabolism and hormone signaling, and their induction is associated with increased activities of superoxide dismutase (SOD) and catalase (CAT) and elevated malondialdehyde (MDA), reflecting active but only partially effective stress mitigation (Li et al., 2022). Comparative analyses between tolerant and sensitive genotypes generally show a larger number of DEGs and stronger fold-changes in tolerant lines, suggesting that robust and timely transcriptional activation is a hallmark of drought resilience (Xuan et al., 2022).

Across cultivars and developmental stages, a subset of drought-responsive genes appears conserved. Meta-analysis and co-expression network studies integrating multiple datasets identified 2168 robust DEGs and hub genes associated with photosynthesis, cytokinin dehydrogenase activity, glyceraldehyde-3-phosphate dehydrogenase, and systemic acquired resistance, pointing to central roles for carbon metabolism and stress signaling in drought tolerance (Shahriari et al., 2022). Specific metabolic genes, such as NCED1 and an SDR enzyme in abscisic acid (ABA) biosynthesis, and CHS, C4H, CAD, F3H, VTC and GST in flavonoid and ascorbate–glutathione pathways, have been repeatedly highlighted as key nodes linking ROS detoxification, secondary metabolism and yield maintenance under drought (Li et al., 2022). Lipid metabolism genes, including ADH and GPAT3, show consistent differential expression across several tolerant and sensitive cultivars and are proposed as universal drought-responsive markers (Yang et al., 2023). Integrative transcriptome–metabolome work further implicates genes in the tricarboxylic acid (TCA) cycle, glycolysis, and isoflavone biosynthesis as central regulators of energy supply and protective metabolites during water deficit (Wang et al., 2022).

4.2 Analysis of drought resistance-related transcription factor families

Transcription factors (TFs) represent a major fraction of drought-induced DEGs and act as key switches in stress-responsive networks. Comparative RNA-seq of tolerant SS2-2 and sensitive Taekwang revealed strong SS2-2-specific upregulation of TFs from multiple families, including bZIP (GmbZIP), DREB, ERF, HD-ZIP, MYB, NAC and WRKY; in particular, one GmbZIP, one GmDREB, one GmHDZIP, five GmMYBs, six GmNACs, and 26 GmWRKYs were markedly induced only in the tolerant cultivar under severe water deficit (Yang et al., 2023). Weighted gene co-expression network analysis (WGCNA) in Liaodou 15 highlighted WRKY (e.g., Glyma.15G021900, Glyma.15G006800), MYB (Glyma.15G190100, Glyma.15G237900), and bZIP (Glyma.15G114800) TFs as central drought-responsive regulators (Li et al., 2022). A meta-network analysis across datasets likewise identified TF-rich co-expression modules and linked specific cis-regulatory motifs in DEG promoters to TF families such as bZIP, AP2/ERF, NAC and MYB, underscoring their pervasive roles in coordinating drought responses (Shahriari et al., 2022).

Functional validation studies have confirmed that individual TFs can markedly enhance drought tolerance. Overexpression of NAC TFs such as GmNAC19 and GmNAC3 improves root growth, chlorophyll retention, osmolyte accumulation and antioxidant enzyme activities, while reducing ROS and MDA, thereby conferring stronger drought resistance in soybean and yeast models (Amin et al., 2025). ERF family members like GmERF205 and AP2/ERF144 enhance survival, root development and yield under drought, and activate downstream stress-responsive genes in transgenic soybean (Wang et al., 2022; Wu et al., 2025). bZIP TFs, including GmbZIP2 and GmTRAB1, increase tolerance by boosting proline and antioxidant systems, enhancing ABA sensitivity and stomatal regulation, and upregulating LEA, GST and other stress-related genes (Yang et al., 2020; Li et al., 2024). Additional families such as HD-ZIP (GmHdz4), NF-YB (GmNFYB17) and TGA-bZIPs also positively regulate drought tolerance, often through improved root architecture, ABA-related signaling, and fine-tuned redox homeostasis (Zhong et al., 2022). Collectively, these findings establish a multi-family TF hierarchy that integrates ABA, ROS, and developmental cues into coherent transcriptional programs under drought.

4.3 Metabolic pathways and drought resistance regulatory networks

Pathway enrichment and network analyses of soybean drought DEGs consistently reveal extensive remodeling of primary and secondary metabolism. Key pathways include photosynthesis, carbon fixation, glycolysis/gluconeogenesis, the TCA cycle, pentose phosphate pathway, amino acid and sugar metabolism, and hormone biosynthesis and signaling (Wang et al., 2022). Under drought, many photosynthesis-related genes are downregulated, reflecting reduced carbon assimilation, whereas genes involved in carbohydrate metabolism and TCA cycle are reprogrammed to support energy balance and osmotic adjustment (Shahriari et al., 2022). Secondary metabolite pathways-particularly phenylpropanoid, flavonoid, and isoflavonoid biosynthesis-are strongly enriched and associated with accumulation of flavonoids and related compounds that contribute to ROS scavenging and membrane protection. Integrative transcriptome–metabolome studies at the seedling and grain-filling stages show concordant upregulation of flavonoid/isoflavone pathway genes and metabolites in tolerant genotypes or melatonin-treated plants, highlighting these pathways as central biochemical buffers against drought damage.

Regulatory network reconstruction, using WGCNA, protein–protein interaction (PPI) and promoter motif analysis, has begun to map how TFs and signaling components control these metabolic shifts. Co-expression modules associated with drought contain hub genes linked to photosynthesis, cytokinin dehydrogenase activity, systemic acquired resistance, secondary metabolite biosynthesis, and antioxidant systems, with TFs from NAC, WRKY, bZIP, AP2/ERF, NF-Y and TGA families occupying central positions (Shahriari et al., 2022; Sun et al., 2022). In tolerant cultivars, hormone-related pathways, notably ABA, jasmonic acid and brassinosteroid signaling, as well as Ca^{2+} and MAPK cascades, are more strongly represented among DEGs, suggesting tighter coupling of environmental perception, transcriptional control and downstream metabolism . Newly proposed data-driven feature engineering pipelines that integrate multi-omics and non-omics datasets further refine candidate drought-tolerance genes and pathways beyond classical WGCNA, offering a scalable framework to prioritize regulatory nodes for breeding and engineering (Kao et al., 2025). Together, these studies depict drought resistance in soybean as an emergent property of interconnected transcriptional, hormonal and metabolic networks that reallocate resources, enhance cellular protection, and stabilize growth under water-limited conditions.

5 Expression Regulation and Functional Validation of Drought Resistance-Related Genes in Soybean

5.1 Construction of transcriptional regulatory networks

Transcriptome-scale analyses in soybean under drought stress have enabled the reconstruction of transcriptional regulatory networks that connect sensing, signaling, and downstream defenses. Co-expression and systems biology approaches, such as weighted gene co-expression network analysis (WGCNA), cluster thousands of differentially expressed genes into modules associated with physiological traits or drought treatments, revealing hub genes and central transcription factors (TFs) that coordinate pathway-level responses (Shahriari et al., 2022).

In an integrative analysis of multiple drought transcriptome datasets, 2168 robust DEGs were grouped into eight core modules, several of which were positively correlated with water-deficit stress and enriched for photosynthesis, cytokinin dehydrogenase activity, and systemic acquired resistance. Promoter motif mining across these DEGs uncovered over-represented cis-acting regulatory elements, including ABRE, DRE, W-box, and MYB-binding sites, implicating bZIP, AP2/ERF, NAC, WRKY, and MYB TF families as key upstream regulators . Protein–protein interaction (PPI) networks further refined these modules by highlighting central nodes such as GLYMA_04G209700 and GLYMA_06G030500 with high connectivity, which may integrate signals from multiple pathways and represent promising breeding targets.

At the single-gene level, regulatory networks are being constructed around specific TFs and structural genes through a combination of transcriptomics, in silico network prediction, and targeted validation. For example, genome-wide analysis of the AP2/ERF family identified GmAP2/ERF144 as strongly induced by drought and salt stress; co-expression-based network prediction followed by RT-qPCR showed that several interacting genes were upregulated 3–30-fold in GmAP2/ERF144-overexpressing lines, suggesting that this TF acts as an upstream node activating a suite of stress-responsive genes (Wang et al., 2022). Similar strategies have mapped downstream targets of NAC TFs such as GmNAC3 and GmNAC19, which modulate genes involved in osmolyte metabolism, antioxidant defenses, and root development (Cui et al., 2024). More recently, data-driven feature-engineering pipelines have integrated multi-omics and non-omics datasets to prioritize candidate drought-tolerance genes and regulatory hubs with greater robustness than traditional WGCNA, providing a scalable framework for refining soybean drought regulatory networks. Together, these approaches are transforming lists of DEGs into structured gene regulatory networks that clarify how diverse signals converge on a relatively small number of master regulators.

5.2 The role of non-coding rnas in drought resistance regulation

Non-coding RNAs (ncRNAs) add an additional, transcriptome-wide layer of control to soybean drought responses, complementing transcriptional regulation by TFs. In soybean, integrated small RNA, degradome, and mRNA sequencing under drought identified hundreds of miRNA–mRNA pairs with inverse expression, indicating widespread post-transcriptional repression in stress signaling networks. A notable example is gma-miR398c, which is strongly induced by drought and targets multiple peroxisome-related copper/zinc superoxide dismutase and copper chaperone genes (GmCSD1a/b, GmCSD2a/b/c, GmCCS) (Zhou et al., 2019). Overexpression of gma-miR398c in Arabidopsis and soybean reduced expression of these targets, impaired ROS scavenging capacity, increased electrolyte leakage, and promoted stomatal opening, leading to higher drought sensitivity (Zhou et al., 2019). Conversely, knockout of miR398c enhanced tolerance, demonstrating that this miRNA acts as a negative regulator of drought resistance by constraining antioxidant defenses. Alternative splicing of GmCSD1a/b was also observed, suggesting that isoform switching can partially bypass miRNA regulation and fine-tune stress responses.

Beyond miR398, broader plant studies show that drought-responsive miRNAs commonly target TFs and hormone signaling components, embedding ncRNAs within both ABA-dependent and ABA-independent pathways (Singroha et al., 2021). For instance, conserved modules such as miR159–MYB and miR169–NF-YA modulate ABA signaling, while miR156–SPL and miR393–TIR1 regulate development and auxin responses, balancing growth and survival under water limitation (Singroha et al., 2021). Long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) further diversify regulatory possibilities by acting as scaffolds, decoys, or precursors in small RNA pathways, and by interfacing with chromatin states (Bolc et al., 2025). Although lncRNA and circRNA research in soybean drought responses is still in its infancy, evidence from other crops indicates that ncRNAs orchestrate hormone crosstalk, oxidative stress defenses, and root architecture, positioning them as tractable entry points for engineering drought resilience while managing yield trade-offs (Gelaw and Sanan-Mishra, 2021). Future integration of ncRNA catalogs with soybean TF–target networks and epigenomic maps will be essential to capture the full regulatory landscape of drought adaptation.

5.3 Methods for functional validation of candidate drought resistance genes

Functional validation is indispensable for confirming the roles of candidate drought-responsive genes identified by transcriptomics, GWAS, or network analysis. In soybean, Agrobacterium-mediated transformation of composite plants with transgenic hairy roots is widely used for rapid gain- and loss-of-function assays. Overexpression of TF genes such as GmNAC3, GmNAC19, and GmHdz4 in hairy roots improved root system architecture, increased fresh and dry root biomass, and enhanced PEG-simulated drought tolerance, often accompanied by reduced hydrogen peroxide and superoxide accumulation and elevated antioxidant enzyme activities (Cui et al., 2024). Conversely, CRISPR/Cas9-mediated editing of GmHdz4 and other regulators demonstrated that disrupting negative regulators or fine-tuning TF dosage can significantly alter osmolyte accumulation, ROS detoxification, and survival under water deficit (Zhong et al., 2022; Khatamov et al., 2025). These root-focused systems are complemented by stable whole-plant transformants, in which phenotypes such as germination rate, plant height, leaf water status, and yield can be assessed across developmental stages and in field conditions. For example, overexpression of GmAP2/ERF144, GmERF205, and the LEA gene GmPM35 in soybean conferred higher relative water content, improved photosynthetic parameters, reduced MDA and ROS accumulation, and increased yield or biomass under drought (Wang et al., 2025).

A wide range of physiological, biochemical, and molecular assays is used to quantify drought tolerance in validated lines. Standard measurements include relative water content, electrolyte leakage, chlorophyll content, proline and soluble sugar levels, MDA content, and activities of SOD, CAT, and POD, alongside NBT/DAB staining and gas-exchange parameters to monitor ROS status and photosynthetic efficiency (Wu et al., 2025). RT-qPCR is routinely employed to confirm transgene expression and to assess downstream targets predicted from regulatory networks, as in the case of GmAP2/ERF144 and GmNAC3 (Wang et al., 2022). Multi-omics validation is emerging, with some studies combining transcriptome–proteome integration or sRNA–degradome–RNA-seq to link gene perturbation with global network changes (Shahriari et al., 2022). New pipelines that integrate feature-engineering with co-functional networks (e.g., SoyNet) further prioritize candidates and pathways for experimental follow-up (Kao et al., 2025). Together, these methodological advances-ranging from rapid hairy-root assays and genome editing to field-level phenotyping and multi-omics integration-are progressively converting computational predictions into mechanistically grounded targets for breeding drought-resilient soybean.

6 Case Study: Transcriptomic Analysis of Key Genes Involved in Soybean Drought Stress Response

6.1 Transcriptome comparison of soybean varieties with varying drought tolerance

Comparative transcriptomic studies using contrasting soybean genotypes have clarified how tolerant and sensitive varieties deploy distinct molecular strategies under drought. At the seedling stage, drought-tolerant Jindou 21 (JD) and drought-sensitive Tianlong No.1 (N1) displayed markedly different DEG profiles: 6038 DEGs were detected in JD versus 4112 in N1, indicating a broader and more dynamic transcriptional reprogramming in the tolerant cultivar (Xuan et al., 2022). KEGG enrichment showed that JD preferentially activated plant hormone (JA, brassinosteroid) signaling, calcium and MAPK cascades, and stress-related TFs and cell-wall remodeling genes, whereas these pathways were weaker or absent in N1 (Xuan et al., 2022). A similar pattern was found in wild soybean, where tolerant genotypes (DTP, DTL) and sensitive genotypes (DSP, DSL) showed 4850 and 6272 DEGs, respectively, but only 547 DEGs had consistent opposite expression between tolerant and sensitive groups, highlighting a core set of genes whose genotype-dependent regulation underpins drought tolerance at germination.

At later stages, comparative analyses of elite cultivars under gradual water deficit or PEG-induced drought further revealed variety-specific strategies. The drought-tolerant Heinong 44 (HN44) maintained higher ABA levels, smaller stomata, lower stomatal conductance, and higher instantaneous water-use efficiency than the sensitive Suinong 14 (SN14), supported by transcriptomic enrichment of ABA signaling and glutathione metabolism in HN44 under severe stress (Xu et al., 2023). In another seedling study, drought-resistant L14 exhibited a “drought-avoidance” strategy with slow-wilting phenotype, fewer DEGs and relatively stable photosynthesis, carbohydrate and lipid metabolism compared with drought-sensitive L21, which showed stronger transcriptional perturbation and downregulation of key leucine-rich repeat receptor-like kinases (LRR-RLKs) (Figure 3) (Li et al., 2023). In L14, these LRR-RLKs were upregulated and proposed to enhance ABA-induced stomatal closure and drought avoidance. Collectively, these comparative datasets show that tolerant varieties either mount earlier and more coordinated activation of signaling and protective pathways or minimize transcriptional and physiological disruption through drought-avoidance traits.

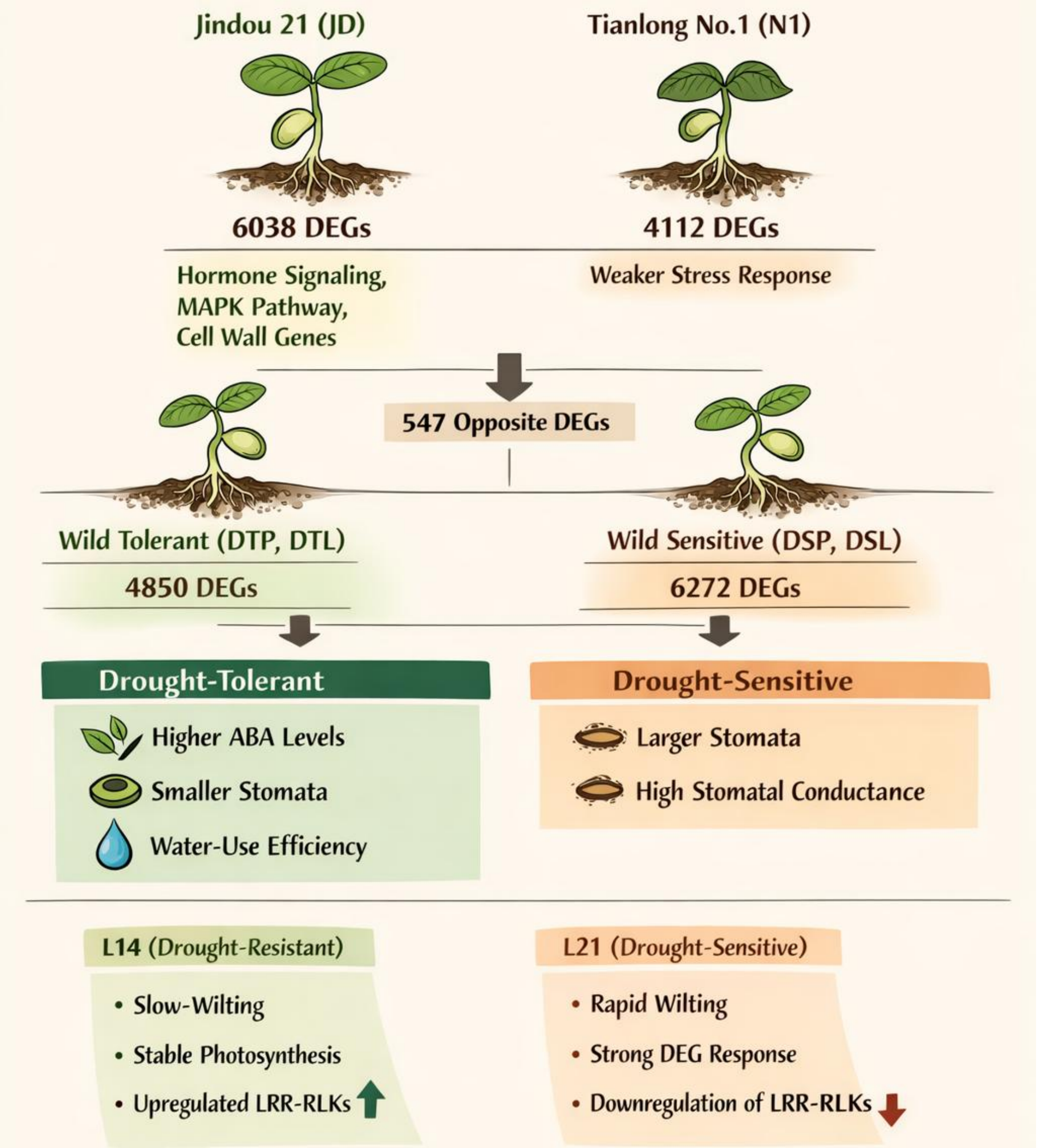

Figure 3 Comparative transcrictome analysis of soybean drought tolerance

6.2 Analysis of expression patterns of key drought resistance genes

Within these comparative datasets, several classes of key drought resistance genes display characteristic expression patterns that separate tolerant from sensitive genotypes. In JD, DEGs in JA and brassinosteroid signaling, calcium channels, MAPKs, cell wall remodeling, and stress-related proteins (e.g., peroxidases, LEA-like proteins) were preferentially induced under drought, whereas their orthologs were weakly or inconsistently regulated in N1 (Xuan et al., 2022). Numerous TFs, including NAC, bZIP, MYB, WRKY, ERF, and HD-ZIP family members, were more abundant in JD than in N1, and some TFs (such as specific NACs, WRKYs, and MYBs) were exclusively differentially expressed in the tolerant cultivar, suggesting a tolerant genotype-specific regulatory layer (Xuan et al., 2022). In wild soybean, DEGs associated with water and auxin transport, cell wall/membrane modification, antioxidant activity, and secondary metabolism were enriched in tolerant genotypes, and 10 genes on a drought-QTL-rich region of chromosome 8-including two hub genes with non-synonymous SNPs-were prioritized as key regulators of root-based drought tolerance (Aleem et al., 2020).

Other case studies highlight specific candidate genes and modules whose expression tracks drought tolerance across tissues and stages. In seedling leaves and roots of tolerant versus sensitive varieties (HN44/HN65, drought-resistant L14/drought-sensitive L21), core pathways such as phenylpropanoid and isoflavonoid biosynthesis, TCA-cycle-linked energy metabolism, ABA-mediated signaling, and glutathione and ROS detoxification are consistently more strongly induced in tolerant materials (Wang et al., 2022; Xu et al., 2023; Wang et al., 2024). Integrative systems-biology analysis across multiple datasets identified 2168 robust DEGs grouped into modules enriched in photosynthesis, cytokinin dehydrogenase, and systemic acquired resistance, with hub genes such as GLYMA_04G209700 and GLYMA_06G030500 showing high connectivity and proposed roles as central coordinators of drought tolerance (Shahriari et al., 2022). At the single-gene scale, expression analyses repeatedly show strong drought induction of TFs like GmAP2/ERF144, GmERF205, GmNAC3, and GmNAC19, as well as structural genes in ABA, cell wall, and secondary metabolism pathways in tolerant backgrounds, aligning their transcriptional patterns with observed physiological resilience (Park et al., 2025).

6.3 Potential applications of drought resistance genes in molecular breeding

The case-study genes and modules identified by transcriptomics offer multiple entry points for molecular breeding of drought-resistant soybean. GWAS-integrated transcriptomics at germination detected 58 QTLs and defined 22 candidate genes within large-effect QTLs; functional annotations highlighted several as strong regulators of drought tolerance, providing markers that can be directly used in marker-assisted selection (MAS) and genomic prediction for early-stage vigor under water deficit (Kong et al., 2025). Similar QTL–expression integration in a PI416937 × Cheongsang RIL population identified five candidate genes, including two involved in ion homeostasis and plasma membrane ATPase regulation and heat-shock protein synthesis, which could be introgressed to improve cell protection under dehydration (Park et al., 2025). Comparative root transcriptomes of wild and cultivated soybean, and multi-omics pipelines that pinpoint hub genes on QTL-rich regions, enable the systematic mining of alleles from wild relatives for pre-breeding and broadening of the genetic base (Aleem et al., 2020).

Validated TFs and structural genes can also be deployed through transgenic or genome-editing approaches to create drought-tolerant germplasm. Overexpression of GmAP2/ERF144 and GmERF205 significantly enhanced leaf relative water content, reduced membrane damage, improved root growth, and increased yield under drought in field or simulated conditions, demonstrating their utility as “plug-in” tolerance modules (Wu et al., 2025; Cui et al., 2024). NAC TFs such as GmNAC3 and GmNAC19 improved root architecture, antioxidant capacity, and osmolyte balance when overexpressed in soybean or heterologous systems, suggesting that stacking multiple TFs with complementary modes of action could generate more robust tolerance (Wang et al., 2022; Amin et al., 2025). Integrative data-driven feature-engineering pipelines that combine co-expression networks, co-functional resources (e.g., SoyNet) and multi-omics evidence help prioritize the most promising drought-tolerance genes for such engineering, reducing false positives and focusing breeding resources on high-value targets (Kao et al., 2025). As genomic selection and CRISPR-based editing become routine in soybean improvement, these transcriptome-defined candidate genes and networks will form the backbone of molecular designs for cultivars that maintain yield and water-use efficiency under increasingly frequent drought episodes.

7 Prospects for the Application of Transcriptomics in Soybean Drought Resistance Breeding

7.1 Mining of drought resistance gene resources

Transcriptomics has become a central tool for systematically mining drought resistance gene resources in soybean, especially when combined with diverse genetic backgrounds and developmental stages. Comparative RNA-seq of wild and cultivated soybeans, as well as tolerant and sensitive genotypes, has identified thousands of drought-responsive DEGs across roots, leaves, and seeds, which are then filtered by GO/KEGG enrichment and network analysis to pinpoint genes involved in hormone signaling, antioxidant defense, secondary metabolism, and cell wall remodeling (Shahriari et al., 2022; Tang et al., 2023). At the germination stage, integration of RTM-GWAS with root transcriptomes led to the identification of 58 QTLs and 22 candidate genes, including transcription factors and transporters whose drought-induced expression patterns align with tolerant phenotypes. Similar transcriptome-driven surveys across seed developmental stages have revealed stage-specific DEGs related to heat shock proteins, LEA proteins, and regulatory TFs, providing novel candidates for protecting yield under reproductive-stage drought. These studies underscore the value of transcriptomics for cataloguing stress-induced genes and connecting them to specific tissues, developmental windows, and physiological traits.

New computational pipelines are significantly improving the efficiency and precision of this gene mining. An integrative data-driven feature-engineering (DFE) framework aggregates cloud-based transcriptomic, genomic, and non-omics datasets and applies robust feature prioritization to identify key drought-tolerant genes (DT genes) while reducing noise and false positives relative to traditional WGCNA (Kao et al., 2025). Multi-omics-based alternative splicing prediction in wild soybean has further expanded the candidate space by revealing 139 genes co-expressed at transcript and protein levels and subject to drought-induced alternative splicing, with isoform-specific regulation of genes such as FT1, CCR1L, and RPL18 linked to drought adaptation (Kim et al., 2024). Joint transcriptome–proteome and transcriptome–metabolome analyses in roots and seedlings have highlighted phenylpropanoid, flavonoid, and TCA-cycle pathways as core drought-resistance modules and nominated both structural genes and pathway regulators as breeding targets (Zhao et al., 2021; Wang et al., 2022). Together, these advances are transforming transcriptomics from a descriptive tool into a discovery engine that yields prioritized, experimentally tractable drought-resistance gene resources.

7.2 Molecular markers and genomic selection in breeding

The increasing availability of high-density SNP data and transcriptome-defined candidate genes has accelerated the development of molecular markers for drought tolerance in soybean. GWAS using tens of thousands to millions of SNPs, coupled with PEG-based germination assays and multi-trait drought indices, have detected numerous QTLs controlling germination rate, root traits, and seedling vigor under water deficit (Kong et al., 2025). For instance, RTM-GWAS with 95,043 SNPs identified 58 QTLs at germination, including 10 large-effect loci; by intersecting these regions with drought-responsive DEGs and co-expression modules, 22 high-confidence candidate genes were defined as valuable genetic resources for breeding (Kong et al., 2025). A separate GWAS of 264 Chinese accessions with 2,597,425 SNPs detected 92 significant SNPs and nine candidate genes, and led to the development of two Kompetitive Allele Specific PCR (KASP) markers tightly linked to drought tolerance at germination, providing low-cost, high-throughput tools for marker-assisted selection (MAS) (Jia et al., 2024). Traditional QTL mapping using high-density SLAF-seq maps has also identified genomic regions controlling plant height and seed weight per plant under drought, with several major and common QTLs proposed for deployment in MAS (Ren et al., 2020).

Transcriptomics enhances these marker efforts by providing functional context and enabling candidate-gene-based markers and genomic prediction models. Expression profiling within QTL intervals helps prioritize functional genes, such as those regulating ion homeostasis, plasma membrane ATPase activity, or heat-shock protein synthesis, which can then be converted into diagnostic markers for drought tolerance (Park et al., 2025). Integrating expression-based co-expression modules and hub genes into genomic selection frameworks allows breeders to weight markers near regulatory hubs or pathway bottlenecks more heavily, potentially increasing prediction accuracy for complex traits like drought tolerance (Shahriari et al., 2022; Valliyodan et al., 2016). As sequencing costs fall, genomic selection models trained on whole-genome SNP data, multi-environment drought phenotypes, and transcriptomic signatures from key tissues and stages are expected to become standard tools. These models can accelerate the identification of superior lines and support pyramiding of multiple drought-tolerance loci while maintaining yield and quality under variable climates (Valliyodan et al., 2016).

7.3 Application of multi-omics integration in drought resistance research

Multi-omics integration builds on transcriptomics to provide a more complete view of soybean responses to drought and to reveal pathways and regulators that would be invisible at a single molecular layer. Combined transcriptome–metabolome analyses at the seedling and root levels have shown that drought-tolerant genotypes accumulate higher levels of flavonoids, phenolic acids, and other secondary metabolites, coordinated with upregulation of phenylpropanoid, flavonoid, isoflavonoid, and TCA-cycle genes (Zhao et al., 2021). These datasets have identified pathway structural genes and TFs whose expression correlates strongly with drought-responsive metabolites, thereby nominating them as key levers for engineering metabolic drought defenses. In the drought-tolerant landrace LX, transcriptomic and metabolomic profiling uncovered constitutively higher expression of secondary metabolism genes and corresponding flavonoid accumulation, suggesting a pre-armed biochemical state that can be introgressed into elite backgrounds. Multi-omics dissection of melatonin-treated soybean further demonstrated how exogenous regulators reshape transcript and metabolite profiles in secondary metabolism pathways to enhance drought tolerance, offering additional agronomic levers (Cao et al., 2020).

Beyond pairwise integrations, more complex multi-omics frameworks are emerging for dissecting stress-specific and multi-stress architectures. Integrative systems-biology analysis of multiple drought transcriptome datasets has already defined co-expression modules, hub genes, and cis-regulatory elements underpinning drought tolerance, while feature-engineering pipelines incorporate co-functional networks (e.g., SoyNet) and non-omics data to prioritize robust DT genes (Kao et al., 2025). Multi-omics network approaches that combine transcriptome and metabolome data under simultaneous drought and pathogen stress have revealed largely distinct gene–metabolite modules for each stress, emphasizing the need for breeding strategies that account for stress-specific molecular architectures (Husein et al., 2025). In wild soybean, integration of transcriptomics, proteomics, and alternative splicing predictions has uncovered isoform-level regulation as an additional layer of drought adaptation and yielded co-expressed transcript–protein gene sets as high-confidence breeding targets (Kim et al., 2024). Looking forward, coupling these integrative omics platforms with high-throughput phenotyping and machine-learning-based genomic prediction is expected to deliver multi-trait, multi-stress-resilient soybean varieties, with transcriptomics serving as the central scaffold that links genetic variation, molecular networks, and field performance.

8 Summary and Outlook

Transcriptomic studies have greatly expanded understanding of how soybean senses, transduces, and mitigates drought stress from germination through reproductive stages. Genome-wide RNA-seq and microarray analyses across roots, leaves, and whole plants have identified thousands of drought-responsive differentially expressed genes (DEGs) involved in hormone signaling, osmotic adjustment, antioxidant defense, photosynthesis, and cell wall remodeling. Comparative analyses of contrasting genotypes at seedling and germination stages uncovered clear molecular distinctions between tolerant and sensitive materials, including broader or more precisely targeted DEG repertoires in tolerant cultivars such as Jindou 21, Heinong 44, and drought-tolerant wild accessions. Weighted gene co-expression network analyses have further organized these DEGs into key modules and pinpointed hub genes and transcription factors, including WRKY, NAC, bZIP, ERF, and NF-Y families, that coordinate downstream responses and represent high-value breeding targets.

Another notable achievement is the extension of transcriptomic analysis to critical yield-determining stages and complex stress scenarios. Studies at flowering and seed-filling stages have linked extensive transcriptional reprogramming in ABA biosynthesis, compatible solute metabolism, and ROS scavenging to sharp reductions in photosynthesis and yield under prolonged drought, while also identifying genes and pathways associated with partial recovery after rewatering. Work on combined drought and heat or flooding has delineated overlapping and stress-specific transcriptomic signatures, revealing an energy-saving core program and distinct hormonal and metabolic adjustments for each stress combination. Root-focused transcriptomes and alternative splicing landscapes have highlighted hormone (auxin/ethylene) signaling, carbohydrate and cell wall metabolism, and extensive drought-responsive splicing as key determinants of root architecture and water acquisition. Collectively, these achievements have created a rich catalog of candidate genes, pathways, and regulatory networks that now underpin molecular breeding strategies for drought-resilient soybean.

Despite these advances, several conceptual and technical challenges limit the full translation of transcriptomic knowledge into robust drought-tolerant cultivars. Many studies rely on a small number of genotypes, often single cultivars or a few contrasting pairs, which constrains the ability to distinguish genotype-specific responses from broadly conserved tolerance mechanisms and complicates the extrapolation of findings to diverse germplasm. Experimental conditions frequently involve PEG-induced or acute drought treatments in controlled environments, whereas field drought is typically moderate, intermittent, and accompanied by other stresses; this discrepancy raises questions about how representative some expression signatures are of real-world condition. Moreover, most datasets capture static “snapshots” at a few time points, providing limited insight into the dynamics of regulatory cascades and the temporal coordination of stress perception, acclimation, and recovery.

Another major challenge is the gap between transcript-level patterns and functional validation or breeding deployment. Only a small fraction of the many candidate DEGs and hub genes identified through GWAS–transcriptome integration, co-expression networks, or multi-omics analysis have been functionally characterized in soybean or heterologous systems . Yield-relevant phenotypes under multi-environment field trials remain even more rarely assessed, particularly for combinations of drought with heat, flooding, or biotic stresses. Additionally, drought tolerance is strongly stage- and tissue-dependent, yet many studies focus on a single stage, tissue, or stress level, making it difficult to build unified models that connect early-stage responses with reproductive performance and final yield. Finally, integrating large, heterogeneous datasets (different platforms, annotations, and analysis pipelines) into coherent, breeder-oriented resources is still technically challenging, and the development of user-friendly tools that link omics findings with markers, QTLs, and genomic selection pipelines is in its infancy.

Future transcriptomic research on soybean drought stress will benefit from more integrative, field-relevant, and functionally anchored approaches. Large, diverse panels of cultivated and wild accessions evaluated under realistic, multi-environment drought scenarios should be combined with time-series RNA-seq in key tissues (roots, leaves, reproductive organs) to capture both the breadth and dynamics of drought responses. Coupling GWAS or RTM-GWAS with tissue-specific transcriptomics and co-expression/network analysis can refine drought-related QTLs and prioritize causal genes for breeding and editing, building on current successes in identifying candidate genes within major QTLs at germination and seedling stages. There is also a need to systematically validate transcription factor hubs (e.g., NAC, ERF, NF-Y, WRKY) and structural genes in ABA, phenylpropanoid, flavonoid, and ROS pathways using overexpression, CRISPR/Cas-mediated editing, and allele-swapping in elite backgrounds, with emphasis on whole-plant performance and yield stability under field drought.

At the same time, multi-omics and post-transcriptional regulation should be more deeply integrated into drought research pipelines. Transcriptome–metabolome and transcriptome–proteome studies have already highlighted central roles for the TCA cycle, isoflavonoid and flavonoid biosynthesis, and tyrosine and linoleic acid metabolism in drought responses, while revealing co-expressed transcript–protein modules that likely represent robust tolerance nodes. Expanding these efforts across developmental stages and stress combinations (drought with heat, flooding, or pathogens) will clarify which pathways can be simultaneously optimized for multi-stress resilience. Genome-wide analyses of alternative splicing and splicing factor regulation indicate that isoform-level control is an important but underexploited layer of drought adaptation and should be incorporated into candidate-gene selection and functional assays. Ultimately, integrating transcriptomic, genomic, and multi-omics data into machine-learning-based genomic selection models, and embedding these models in breeder-friendly platforms linked to markers and decision tools, will be essential to convert molecular insights into climate-resilient soybean cultivars at scale.

Acknowledgments

Thanks to the reviewers for providing detailed comments and guidance on the manuscript of this study. The reviewers' keen insights into the issues and attention to detail have greatly benefited the authors.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdullah M., Waraich E., Ahmad M., Hussain S., Asghar H., Haider A., Zulfiqar U., Ahmad Z., Soufan W., Prasad P., and Djalović I., 2025, Improving soybean drought tolerance via silicon-induced changes in growth, physiological, biochemical, and root characteristics, Plant Signaling & Behavior, 20: 2465232.
https://doi.org/10.1080/15592324.2025.2465232

Aleem M., Raza M., Haider M., Atif R., Ali Z., Bhat J., and Zhao T., 2020, Comprehensive RNA-seq analysis revealed molecular pathways and genes associated with drought tolerance in wild soybean (Glycine soja Sieb. & Zucc.), Physiologia Plantarum, 3: 12-42.
https://doi.org/10.1111/ppl.13219

Ali S., Mir R., Haque A., Almalki M., Alfredan M., Khalifa A., Mahmoudi H., Shahid M., Tyagi A., Mir Z., Manghwar H., Aslam M., and Khan M., 2025, Exploring physiological and molecular dynamics of drought stress responses in plants: challenges and future directions, Frontiers in Plant Science, 16: 1565635.
https://doi.org/10.3389/fpls.2025.1565635

Amin N., Lu L., Rehman F., Imran M., Gai Y., Wang P., and Jian W., 2025, Transcriptional regulation of GmNAC3-mediated drought stress tolerance in soybean, GM Crops and Food, 16: 435-449.
https://doi.org/10.1080/21645698.2025.2516295

Aslam M., Waseem M., Jakada B., Okal E., Lei Z., Saqib H., Yuan W., Xu W., and Zhang Q., 2022, Mechanisms of abscisic acid-mediated drought stress responses in plants, International Journal of Molecular Sciences, 23: 1084.
https://doi.org/10.3390/ijms23031084

Aziez A., Prasetyo A., and P., 2022, The effect of drought stress on the growth and yield of soybean (Glycine max L.), Applied Ecology and Environmental Research, 20(4): 3569-3580.
https://doi.org/10.15666/aeer/2004_35693580

Bolc P., Puchta-Jasińska M., Motor A., Maździarz M., and Boczkowska M., 2025, Regulatory landscapes of non-coding RNAs during drought stress in plants, International Journal of Molecular Sciences, 26: 9892.
https://doi.org/10.3390/ijms26209892

Cao L., Jin X., Zhang Y., Zhang M., and Wang Y., 2020, Transcriptomic and metabolomic profiling of melatonin treated soybean under drought stress during grain filling period, PLoS ONE, 15: e0239701.
https://doi.org/10.1371/journal.pone.0239701

Chen J., Shrestha L., Green G., Leier A., and Marquez-Lago T., 2023, The hitchhikers' guide to RNA sequencing and functional analysis, Briefings in Bioinformatics, 5: 24.
https://doi.org/10.1093/bib/bbac529

Chen W., Yao Q., Patil G., Agarwal G., Deshmukh R., Lin L., Wang B., Wang Y., Prince S., Song L., Xu D., An Y., Valliyodan B., Varshney R., and Nguyen H., 2016, Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-seq, Frontiers in Plant Science, 7: 1044.
https://doi.org/10.3389/fpls.2016.01044

Conesa A., Madrigal P., Tarazona S., Gómez-Cabrero D., Cervera A., McPherson A., Szcześniak M., Gaffney D., Elo L., Zhang X., and Mortazavi A., 2016, A survey of best practices for RNA-seq data analysis, Genome Biology, 17: 13.
https://doi.org/10.1186/s13059-016-0881-8

Cui X., Tang M., Li L., Chang J., Yang X., Chang H., Zhou J., Liu M., Wang Y., Zhou Y., Sun F., and Chen Z., 2024, Expression patterns and molecular mechanisms regulating drought tolerance of soybean conferred by transcription factor gene GmNAC19, International Journal of Molecular Sciences, 25: 2396.
https://doi.org/10.3390/ijms25042396

Demirtas C., Yazgan S., Candoğan B., Sincik M., Büyükcangaz H., and Göksoy A., 2010, Quality and yield response of soybean to drought stress in sub-humid environment, African Journal of Biotechnology, 9(41): 6873-6881.

Dong S., Jiang Y., Dong Y., Wang L., Wang W., Yan C., and Liu L., 2019, A study on soybean responses to drought stress and rehydration, Saudi Journal of Biological Sciences, 26: 2006-2017.
https://doi.org/10.1016/j.sjbs.2019.08.005

Falcioni R., De Oliveira C., Vedana N., Mendonça W., Gonçalves J., De Fatima Da Silva Haubert D., De Matos D., Reis A., Antunes W., Crusiol L., Sibaldelli R., Nepomuceno A., Neumaier N., Farias J., Furlanetto R., Demattê J., and Nanni M., 2025, Progressive water deficit impairs soybean growth, alters metabolic profiles, and decreases photosynthetic efficiency, Plants, 14: 2615.

Fan X., Zhang Y., Gu P., and Naz M., 2025, Epitranscriptomic control of drought tolerance in rice: the role of RNA methylation, Plants, 14: 2002.
https://doi.org/10.3390/plants14132002

Fracasso A., Trindade L., and Amaducci S., 2016, Drought stress tolerance strategies revealed by RNA-seq in two sorghum genotypes, BMC Plant Biology, 16: 115.
https://doi.org/10.1186/s12870-016-0800-x

Gebre M., Rajcan I., and Earl H., 2022, Genetic variation for effects of drought stress on yield formation traits among soybean cultivars, Frontiers in Plant Science, 13: 1020944.
https://doi.org/10.3389/fpls.2022.1020944

Gelaw T., and Sanan-Mishra N., 2021, Non-coding RNAs in response to drought stress, International Journal of Molecular Sciences, 22: 12519.
https://doi.org/10.3390/ijms222212519

Geng A., Lian W., Wang Y., Liu M., Zhang Y., Wang X., and Chen G., 2024, Molecular mechanisms and regulatory pathways underlying drought stress response in rice, International Journal of Molecular Sciences, 25: 1185.
https://doi.org/10.3390/ijms25021185

Haghpanah M., Hashemipetroudi S., Arzani A., and Araniti F., 2024, Drought tolerance in plants: physiological and molecular responses, Plants, 13: 2962.
https://doi.org/10.3390/plants13212962

Husein G., Castro-Moretti F., Prado M., Amorim L., Mazzafera P., Canales J., and Monteiro-Vitorello C., 2025, Integrative multi-omics analysis reveals stress-specific molecular architectures in soybean under drought and rust infection, bioRxiv, 07: 663534
https://doi.org/10.1101/2025.07.07.663534

Jia Q., Zhou M., Xiong Y., Wang J., Xu D., Zhang H., Liu X., Zhang W., Wang Q., Sun X., and Chen H., 2024, Development of KASP markers assisted with soybean drought tolerance, Frontiers in Plant Science, 15: 1352379.
https://doi.org/10.3389/fpls.2024.1352379

Jones S., and Vodkin L., 2013, Using RNA-seq to profile soybean seed development from fertilization to maturity, PLoS ONE, 8: e59270.
https://doi.org/10.1371/journal.pone.0059270

Kao P., Baiya S., Lee C., Tseng C., Chen S., Huang Y., and Kao C., 2025, Identification of key drought-tolerant genes in soybean using an integrative pipeline, Journal of Big Data, 12: 78.
https://doi.org/10.1186/s40537-025-01078-w

Khatamov D., Kadirova Z., Usmanov A., Nazarov K., and Aberkulov M., 2025, Cloning and bioinformatic analysis of drought-related genes in soybean, BIO Web of Conferences, 181: 01004.
https://doi.org/10.1051/bioconf/202518101004

Kim T., Hwang H., Bang G., Ha J., Park Y., and Kim J., 2024, Molecular mechanisms of drought tolerance in wild soybean via multi-omics, Environmental and Experimental Botany, 220: 105872.
https://doi.org/10.1016/j.envexpbot.2024.105872

Kong K., Xu M., Wu L., Zhou H., Wang R., Zhao T., Wang C., and Song Y., 2025, GWAS and transcriptome analysis reveal drought-related genes in soybean, Frontiers in Plant Science, 16: 1621869.
https://doi.org/10.3389/fpls.2025.1621869

Li B., Liu Y., Cui X., Fu J., Zhou Y., Zheng W., Lan J., Jin L., Chen M., Xu Z., and Min D., 2019, Genome-wide characterization of soybean TGA transcription factors, Frontiers in Plant Science, 10: 549.
https://doi.org/10.3389/fpls.2019.00549

Li H., Zhang Q., Xu P., Wang X., Dai S., Liu Z., Xu M., Cao X., and Cui X., 2024, GmTRAB1 positively regulates drought tolerance in soybean, Plants, 13: 3104.
https://doi.org/10.3390/plants13213104

Li M., Li H., Sun A., Wang L., Ren C., Liu J., and Gao X., 2022, Transcriptome analysis reveals drought-responsive genes in soybean, Frontiers in Genetics, 13: 1060529.
https://doi.org/10.3389/fgene.2022.1060529

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.