scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A comprehensive transcriptomic dataset of *Sorghum bicolor* seedlings under abiotic stress conditions

Won Cheol Yim[1,5], Yu Kyong Hu[2,5], Jonathan S. Lomas[2], Sang Hun Kim[2], Dong Yeol Song[2], Nguyen Viet Phong[3], Seo Young Yang[3,4] ✉ & Sung Don Lim[2] ✉

Sorghum [*Sorghum bicolor* (L.) Moench], a C_4 cereal crop, is the fifth most widely cultivated cereal globally and is cultivated in over 100 countries for food, fodder, and biofuel production. Its exceptional resilience to abiotic stresses such as salinity, drought, heat, and cold makes it a valuable model for crop adaptability research under climate change. Here, we present a comprehensive, time-resolved RNA-seq dataset from sorghum seedlings (cultivar Hwanggeumchal) subjected to salt, PEG-induced drought, heat, cold, and abscisic acid (ABA) treatments. Leaf tissue was sampled at 6, 12, and 24 hours post-treatment across three biological replicates, yielding 54 strand-specific libraries sequenced on the Illumina NovaSeq. 6000 platform. The dataset includes raw sequencing reads deposited in the NCBI SRA, together with processed data comprising differential gene expression profiles, Gene Ontology enrichment analyses, and a gene co-expression network, all hosted on Figshare. This resource is intended to facilitate comparative transcriptomic studies and functional gene discovery related to abiotic stress responses in sorghum and other cereal crops.

Background & Summary

Environmental stresses, such as soil salinity, drought, heat, and cold are significant challenges reducing crop yield and have an impact on the quality of agricultural products. As climate change continues and global population growth persists, abiotic stressors are projected to increase in severity, posing a significant threat to food security worldwide[1,2]. Abiotic stress leads to overproduction of reactive oxygen species (ROS) in plants, causing oxidative stress[3]. Oxidative stress damages essential cellular components such as membrane lipids, proteins, and nucleic acids, leading to impaired metabolic functions and plant growth[4,5]. In response to these abiotic stresses, plants produce the phytohormone abscisic acid (ABA), which in turn elicits many adaptive responses, including the expression of stress-responsive genes and modulation of the antioxidant defense system to scavenge the oxidative damage caused by ROS[6,7]. Understanding how plants integrate ABA signaling with abiotic stress response is critical for elucidating the gene regulatory networks that drive adaptive mechanisms and long-term stress resilience.

Sorghum [*Sorghum bicolor* (L.) Moench], the fifth most widely cultivated cereal crop globally, produced approximately 58.4 million metric tons in 2023[8–10]. As a C_4 photosynthetic plant, sorghum exhibits high photosynthetic efficiency and biomass productivity, enabling its widespread cultivation for food, livestock feed, and renewable energy[11–13]. Beyond its agricultural versatility, sorghum is distinguished by its robust tolerance to multiple abiotic stresses, including salinity, drought, heat, and to a lesser extent, cold[14–18]. Its exceptional adaptability allows it to thrive in marginal environments where other major cereals often underperform[18,19]. Sorghum exhibits inherent physiological and genetic traits, such as deep rooting systems, efficient stomatal regulation,

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV, 89557-0330, USA. [2]Molecular Plant Physiology Laboratory, Department of Smart-fam Life Sciences, Sangji University, Wonju, 26339, Republic of Korea. [3]Department of Biology Education, Teachers College and Institute for Phylogenomics and Evolution, Kyungpook National University, Daegu, 41566, Republic of Korea. [4]Department of Advanced Bioconvergence, Kyungpook National University, Daegu, 41566, Republic of Korea. [5]These authors contributed equally: Won Cheol Yim, Yu Kyong Hu. ✉e-mail: syy@knu.ac.kr; sdlim@sangji.ac.kr

osmotic adjustment, and protective antioxidant responses, which collectively confer tolerance to abiotic stresses and establish sorghum as a remarkable model for studying abiotic stress responses[20,21]. Therefore, identifying key genes associated with abiotic stress responses in sorghum provides insights that can be translated to improve stress resilience across other major crops.

Abiotic stresses encountered at the seedling stage can cause irreversible damage that limits crop establishment and yield potential. Although most crop species in the Poaceae family are susceptible to abiotic stresses during the early seedling stages, sorghum seedlings are notably more tolerant compared to the later growth stages[22–24]. This resilience in seedlings is underpinned by robust physiological adaptations and enhanced antioxidant systems[25,26]. Sorghum also relies on a complex regulatory landscape involving multiple transcription factor (TF) families, including NAC, MYB, WRKY, bZIP, AP2/ERF (DREB/ERF), and heat-shock factors (HSFs), implicated in abiotic stress responses. Many of these TFs are involved in ABA signaling, orchestrating transcriptional changes under salinity, drought, heat, and cold stress[27–31]. For example, the NAC transcription factor *SbNAC9* enhances expression of ABA-biosynthesis gene (*SbNCED3*) and a peroxidase gene, thus amplifying ABA accumulation and ROS detoxification to improve drought tolerance in seedlings[32]. While a number of physiological and transcriptomic responses have been characterized under individual stress conditions[21,33–35], the interplay between abiotic stress and ABA-mediated gene regulation remains incompletely understood.

Complicating transcriptomic analysis in grasses, mature monocot leaves exhibit spatially dynamic gene expression along their longitudinal axis due to their basipetal developmental pattern, with pronounced gradients from the base to the tip[36–38]. By contrast, whole-leaf sampling at the seedling stage reduces positional bias and captures more uniform transcriptional responses, thereby enabling reliable identification of stress-induced expression profiles. Given that early developmental stages are particularly critical for crop establishment and yield potential, elucidating these molecular responses in seedlings is vital for improving abiotic stress tolerance and overall crop productivity.

Advancements in high-throughput sequencing technologies have facilitated the comprehensive analyses of gene expression profiles under stress conditions[39]. RNA sequencing (RNA-seq) enables the identification of differentially expressed genes (DEGs) and provides insights into the molecular basis of abiotic stress tolerance[40]. Short-read RNA-seq has emerged as a powerful and cost-effective approach for transcriptome-wide analysis in plants, enabling high-resolution detection of DEGs across diverse experimental conditions[41–43]. Its affordability, scalability, and compatibility with multiplexed experimental designs make it especially suitable for high-throughput studies involving biological replicates, multiple treatments, and time-course analyses[44]. Moreover, parallel advances in sorghum genomics have provided important insights into biomass accumulation, stress tolerance, and bioenergy-related traits, while integrative transcriptomic and functional studies continue to expand our understanding of stress resilience mechanisms in sorghum[45]. Together, these developments underscore both the relevance of sorghum as a model for stress biology and the utility of RNA-seq in dissecting the molecular basis of abiotic stress responses.

In *Sorghum bicolor*, a stress-resilient C_4 cereal, RNA-seq studies have uncovered key TFs, stress-inducible genes, and hormone signaling pathways under specific abiotic stress conditions[33,34,46]. However, the existing datasets are often limited in scope, typically focusing on single stress types, generated using variable genotypes, tissue types, and sampling time points. This lack of integration limits direct comparison across stress conditions and hinders the identification of core stress-responsive genes and shared regulatory modules, especially those modulated by ABA signaling. To address this knowledge gap, the present study delivers a time-resolved, short-read RNA-seq dataset encompassing sorghum seedling responses to multiple abiotic stresses and ABA treatment under standardized conditions. Samples were collected at 6, 12, and 24 h post treatment: to capture temporal changes in transcript abundance. The dataset, accompanying processed tables, and analysis code are provided as a reusable resource to support comparative transcriptomic analyses of abiotic stress and ABA treatments in sorghum and related crops. All raw sequencing reads and processed outputs (DEG tables, GO enrichment results, co-expression network, and quality-control summaries) described in this Data Descriptor are newly generated for this study and have not been previously published or used in any other article.

Methods

Plant materials. Sorghum [*Sorghum bicolor* (L.) Moench] (cultivar Hwanggeumchal) seeds were used in this study. Hwanggeumchal is a Korean sorghum variety developed through pure-line selection from local landrace GWG 140 (Gangwon 140) by the Gangwon State Agricultural Research and Extension Service (https://www.ares.gangwon.kr/gwares). The GWG 140 was introduced in 2000, and yield performance trials were conducted from 2000 to 2001. Regional adaptability tests were carried out for three years from 2002 to 2004, and the cultivar was officially registered under as Hwanggeumchal (cultivar registration number 06-0013-8) with the Korea Seed & Variety Service (https://www.seed.go.kr/seed/index.do) on March 16, 2005[47].

The Hwanggeumchal sorghum variety is an early-maturing, high-yielding glutinous sorghum cultivar characterized by a high content of unsaturated fatty acids. Additionally, Hwanggeumchal contains a higher concentration of polyphenolic compounds, including gallic acid, compared to other sorghum varieties, resulting in strong antioxidant and anti-inflammatory activities[47–49]. Due to these desirable traits, Hwanggeumchal is the most widely cultivated sorghum variety in South Korea and holds significant agricultural importance. Hwanggeumchal seeds are preserved and managed by the Seed Bank of the National Institute of Agricultural Sciences, Rural Development Administration (RDA), South Korea. Seeds can be obtained by submitting an online distribution request (https://genebank.rda.go.kr/simpleSearch.do) with accession number (IT231310) or by contacting the designated personnel via email at genebank@korea.kr.

Comparison with widely used sorghum reference genotypes. The *Sorghum bicolor* v3.1.1 reference genome was assembled from BTx623[50], a non-waxy genotype widely used for genetic and transcriptomic studies. Hwanggeumchal differs in three aspects. First, it is a waxy cultivar with amylopectin-dominated starch, unlike

BTx623[51], which may affect carbohydrate metabolism under stress. Second, it shows elevated polyphenol levels and altered fatty acid profiles relative to typical sorghum lines[52], potentially influencing antioxidant and membrane-stability responses. Third, it originates from a Korean temperate landrace, unlike commonly studied genotypes such as Tx7000, IS3614-3, and M81-E[46,53], and may therefore show distinct cold-response signatures. Implications for data interpretation. Mapping to the BTx623 reference may reduce efficiency at divergent loci due to sequence variation[54], although overall rates remain high (71.0–91.9%). Observed stress-response patterns reflect the Hwanggeumchal background and should be compared with comparable datasets from other cultivars before generalization. This temperate, waxy, and polyphenol-rich genotype complements existing datasets and expands sorghum transcriptomic diversity.

Plant growth conditions. The sorghum seeds were surface-sterilized by immersion in 70% ethanol for 1 min, followed by immersion in 2% (v/v) sodium hypochlorite for 5 min, and subsequently rinsed thoroughly five times with sterile distilled water. Sterilized seeds were immediately sown in a TS2 soil mixture (Klasmann-Deilmann, Geeste, Germany) in 89-mm square plastic pots (Kord, Inc., Toronto, CA, USA) and germinated in a growth chamber. Growth conditions were maintained at a 16-h/8-h photoperiod (160–180 μmol m^{-2} s^{-1} light intensity) with day/night temperatures of 25 °C/23 °C. Once the coleoptiles emerged from the soil, the seedlings were irrigated twice weekly with half-strength Kimura B nutrient solution[55] and maintained under identical conditions for an additional two weeks.

Abiotic stress and ABA treatments. At the three-leaf stage (15 days after emergence), seedlings were subjected to five treatments: salinity, polyethylene glycol (PEG)-induced drought, heat, cold, and abscisic acid (ABA). For salinity, PEG, and ABA treatments, pots containing seedlings were placed in trays with half-strength Kimura B nutrient solution supplemented with 150 mM NaCl (S9888, Sigma-Aldrich), 20% PEG 8000 (P2139, Sigma-Aldrich), or 100 µM ABA (90769, Sigma-Aldrich), respectively. After 20 min of uptake, pots were transferred to empty trays to avoid waterlogging. For heat and cold treatments, seedlings were first equilibrated in nutrient solution for 20 min and then transferred to growth chambers set at 40 °C or 4 °C. Control plants were maintained at 25 °C under identical light conditions.

For RNA extraction, the entire third fully expanded leaf (from ligule to tip) was harvested at 6, 12, and 24 h post-treatment under both control and stress conditions. The decision to use the whole leaf, rather than segmenting along the longitudinal axis, was made to reduce positional bias in transcriptional profiling, as monocot leaves are known to exhibit spatial gene expression gradients from the basal meristematic zone to the distal tip[36]. Samples were immediately flash-frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

RNA extraction, library preparation, and sequencing. Total RNA was extracted from 100–200 mg of frozen tissue using the Apure™ Plant RNA kit (APBIO, Namyangju-si, Gyeonggi-do, South Korea). The total RNA concentration was evaluated using Quant-IT RiboGreen (Invitrogen, Carlsbad, CA, USA). The RNA-seq libraries were generated using the TruSeq Stranded mRNA Sample Prep Kit (Illumina, San Diego, USA) for all 54 samples. The fragmented RNA was reverse transcribed into first-strand cDNA using SuperScript II (Invitrogen) and random hexamer primers, followed by second-strand cDNA synthesis using DNA Polymerase I, RNase H, and dUTP. The cDNA fragments were then subjected to end repair, A-tailing, adapter ligation, and purification. Following adapter ligation and purification, the cDNA fragments were amplified by polymerase chain reaction (PCR) to generate the final sequencing libraries. Library quantification was performed using KAPA Library Quantification kits (KAPA Biosystems), following their qPCR guidelines. Paired-end sequencing (2 × 100 bp) was performed using the Illumina NovaSeq. 6000 platform (Macrogen Inc, https://dna.macrogen.com/). Approximately 31 million reads per sample were obtained (±2.28 million) (Fig. 1; Data Table 7[56]).

Read processing and alignment. Raw reads were processed with Trim Galore (v0.6.6) to remove adapters and bases with Phred quality scores < 20. Reads <50 bp after trimming were discarded. Quality metrics were evaluated with FastQC (v0.11.9) (Data Table 7[51]). A STAR genome index was built from the Sorghum bicolor v3.1.1 primary assembly together with the corresponding Phytozome GTF annotation using–runMode genomeGenerate with–sjdbOverhang 99 to match the 100-bp paired-end read length. Trimmed reads were aligned to this index with STAR (v2.7.10a) under default parameters except for the following:–runMode alignReads,–readFilesCommand zcat,–outSAMtype BAM SortedByCoordinate,–quantMode TranscriptomeSAM (to emit a transcriptome-coordinate BAM for RSEM), and–outFilterType BySJout. All other filtering thresholds (–outFilterMultimapNmax 20,–outFilterMismatchNoverReadLmax 0.04,–alignIntronMin 20,–alignIntronMax 1000000) were left at STAR defaults. Gene- and transcript-level abundances were estimated by running rsem-calculate-expression (RSEM v1.3.3) on the transcriptome BAM with–paired-end–strandedness reverse,–no-bam-output, and–estimate-rspd. The RSEM reference was prepared from the same GTF used for the STAR index. Multi-mapped reads were assigned probabilistically across isoforms/genes using RSEM's expectation-maximization (EM) algorithm. Expression outputs included raw expected read counts per gene (Data Table 8[56]) and normalized values in transcripts per million (TPM) (Data Table 9[56]).

Differential expression analysis. Differential expression analysis was performed in R (v4.0.3) using DESeq. 2 (v1.30.1)[57]. RSEM gene-level expected counts for all 54 samples were rounded to integers, and a global expression filter was applied to remove genes with fewer than 25 reads summed across the full set of 54 samples. For each of the five treatment-versus-control comparisons (salinity, PEG-induced drought, heat, cold, and ABA), the corresponding subset of 18 samples (treated samples + control samples × 3 time points × 3 biological replicates) was then extracted from the globally filtered count matrix and imported into a separate DESeq 2 dataset via DESeqDataSetFromMatrix. Sample metadata for each comparison included two factors: Condition (treatment vs. control) and Time (6, 12, 24 h), both encoded as factors with explicit reference levels (Condition: control;

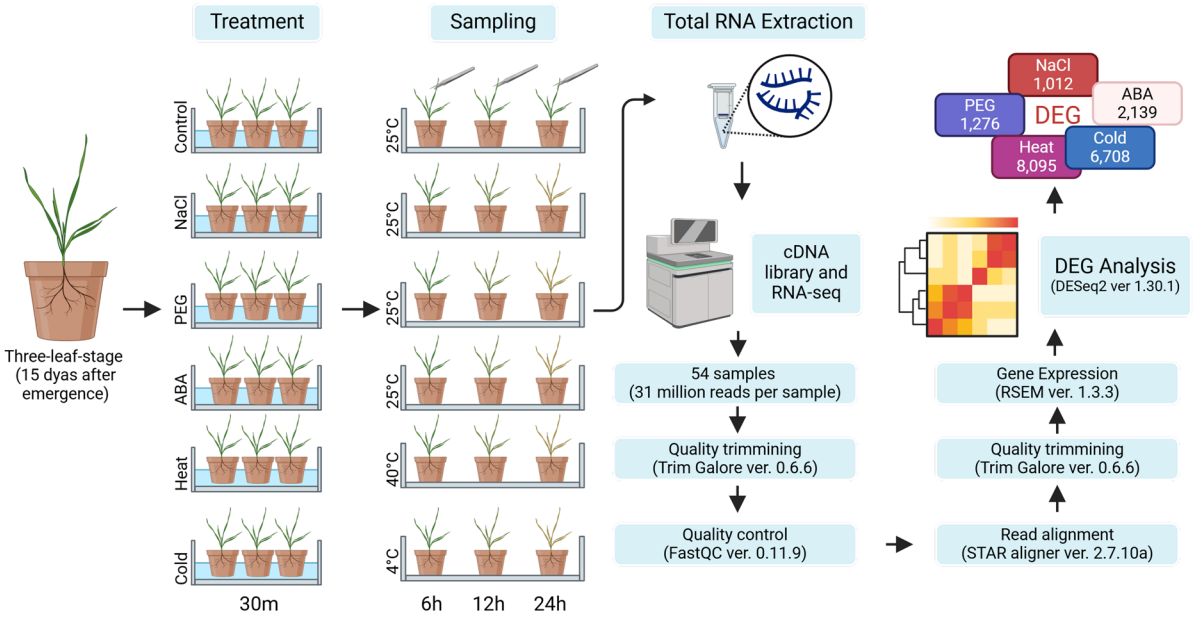

Fig. 1 Schematic representation of experimental process. Leaf tissue samples were treated and collected with three biological replicates. The workflow for this study includes the analysis procedures, programs, and parameters used. Detailed methodologies are described in the Methods sections.

Time: 6, 12, 24 h). To identify genes responsive to the given treatment while accounting for temporal effects shared between treated and control samples, the full model ~ Condition + Time + Condition:Time was fitted and a likelihood ratio test (LRT) was performed against the reduced model ~ Time. This LRT identifies genes whose expression depends on Condition at any time point or whose response trajectory differs between conditions. Size factors were estimated by the median-of-ratios method, dispersions by the default empirical Bayes shrinkage, and Cook's-distance outlier handling and independent filtering were left at DESeq. 2 defaults. Multiple testing correction used the Benjamini-Hochberg procedure. Per-time-point $\log_2$ fold-changes (treated vs. control at each of 6, 12, and 24 h) were obtained from the corresponding model coefficients of each per-treatment fit, and genes meeting both LRT adjusted p-value < 0.05 and $|\log_2 \text{fold-change}| \geq 1$ in at least one time point were retained as DEGs and reported in Data Tables 1–5.

Functional Annotation and Enrichment

Network analysis visualization. A gene co-expression network was constructed from the union of non-redundant DEGs identified across the five stress treatments. The input expression matrix consisted of TPM values for these DEGs (extracted from Data Table 9) across all 54 samples, transformed as log2(TPM + 1) to stabilize variance and reduce the impact of high-magnitude outliers. Genes for which the log2(TPM + 1) value was zero in more than 50% of samples were removed prior to correlation analysis to avoid spurious correlations driven by sparse expression. Pairwise gene-gene Pearson correlation coefficients (PCC) were computed across the 54-sample expression vector for every pair of remaining DEGs. To retain only high-confidence co-expression relationships, edges were filtered at PCC ≥ 0.99; this stringent threshold was chosen empirically to yield a sparsely connected, interpretable network rather than a near-complete graph. Self-loops and duplicate undirected edges were removed. The resulting edge list (Data Table 6) was imported into Cytoscape (v3.10)[58], and topology metrics including node degree, betweenness centrality, and edge betweenness were computed using the NetworkAnalyzer plug-in.

Network topology parameters were calculated with the NetworkAnalyzer plug-in from Cytoscape. Node size was scaled by degree, and node color was mapped to betweenness centrality. Edge thickness was proportional to PCC, and edge color reflected edge betweenness. These visualization settings are reported to document how the network figure was generated and to enable readers to reproduce or adapt the visualization for their own analyses.

Data Records

All data has been deposited in Figshare[56].

Raw sequencing data. Raw FASTQ files for all 54 libraries have been deposited at the NCBI Sequence Read Archive under BioProject accession SRP53196755[59]. Each sample is represented by a pair of gzipped FASTQ files corresponding to read 1 and read 2 of the 2 × 100 bp paired-end run. The mapping between SRR run accessions and our internal sample identifiers is provided in Data Table 7.

Processed data on figshare. All processed data tables and the GO enrichment outputs are deposited together on Figshare[56]. The deposit also contains a README.md file summarizing contents and column

definitions reproduced below. Each table is provided in two equivalent formats, Microsoft Excel (.xlsx) and UTF-8 comma-separated text (.csv), and the two largest matrices (Data Tables 8, 9) are additionally provided as gzipped tab-separated text (.tsv.gz) for use with command-line, R, and Python workflows.

Sample naming convention. Across all data tables, each of the 54 samples is identified by a code of the form Condition_Time_Replicate, where Condition ∈ {ABA, Cold, Heat, PEG, Salt, control}, Time ∈ {6 h, 12 h, 24 h}, and Replicate ∈ {Rep1, Rep2, Rep3}. Thus, for example, Salt_12h_Rep2 refers to the second biological replicate of the salinity treatment harvested 12 h post-treatment. The keyword "control" (lowercase) denotes matched untreated samples grown at 25 °C under standard conditions and harvested at the same three time points to provide a within-experiment baseline for each stress.

Data Tables 1–5. Differential expression results per treatment. Filenames: DataTable 1_DEGs_Salinity, DataTable 2_DEGs_PEG, DataTable 3_DEGs_Heat, DataTable 4_DEGs_Cold, DataTable 5_DEGs_ABA (each available as.xlsx and.csv). Each file contains the differentially expressed gene set for the indicated treatment versus matched controls, identified using the global DESeq. 2 LRT described in the Methods (adjusted p-value < 0.05 and $|\log_2 FC| \geq 1$ for the corresponding treatment-specific contrast). Across the five treatments, 19,230 non-redundant DEGs were identified. Columns common to all five files are: gene_id (gene identifier from the S. bicolor v3.1.1 annotation), Description (functional description from Phytozome / InterProScan), baseMean (DESeq. 2 mean of normalized counts across all 54 samples), log2FoldChange (treatment-vs-control $\log_2$ fold-change for this treatment), lfcSE (standard error of the $\log_2$ fold-change), stat (DESeq. 2 LRT statistic; shared across treatments), pvalue (raw LRT p-value; shared across treatments), padj (Benjamini-Hochberg-adjusted LRT p-value; shared across treatments), and the 54 sample-level RSEM expected count columns named per the convention above. Note that for Data Table 3 (Heat), the Description column is included with empty values to maintain a consistent column structure across the five tables, as the original annotation was not available for that table.

Data Table 6. Co-expression network (nodes). Filename: DataTable 6_CoExpressionNetwork_Nodes (.xlsx and.csv). The file contains node-level topology metrics for the 414 genes that participate in at least one PCC ≥ 0.99 co-expression edge. Columns are: gene_id (*S. bicolor* v3.1.1 identifier), Description (functional description), AverageShortestPathLength, BetweennessCentrality, ClosenessCentrality, ClusteringCoefficient, Degree, Eccentricity, NeighborhoodConnectivity, NumberOfDirectedEdges, NumberOfUndirectedEdges, PartnerOfMultiEdgedNodePairs, Radiality, SelfLoops, Stress, TopologicalCoefficient (all from Cytoscape NetworkAnalyzer), and cl1.Status (cluster-membership label assigned by the clusterMaker / ClusterONE plug-in). These topology metrics are provided as a reusable resource that users can apply to prioritize candidate genes or modules in their own downstream analyses; we do not assign biological interpretations to specific nodes here. The underlying edge list (PCC ≥ 0.99) is not redistributed in this deposit but can be regenerated from Data Table 9 (TPM) by users with their preferred preprocessing.

Data Table 7. Sequencing and alignment quality. Filename: DataTable 7_QC_Alignment (.xlsx and.csv). One row per sample (n = 54). Columns: Sample_ID (using the naming convention above), Million_Reads (total sequencing reads in millions), Adapter_Pct (fraction of bases identified as adapter by Trim Galore), Million_Aligned (reads aligned to *S. bicolor* v3.1.1, in millions), Aligned_Pct (mapping rate as a fraction; uniquely + multi-mapped, STAR), Duplication_Pct (estimated duplication rate), and GC_Content (mean GC content of the library).

Data Table 8. Raw expected count matrix. Filename: DataTable 8_RSEM_ExpectedCounts (.xlsx,.csv,.tsv.gz). Rows are the 29,027 sorghum gene models (gene_id in the first column) and the remaining 54 columns are samples named per the convention above. Cell values are RSEM expected counts (non-integer; round to integer before importing into DESeq. 2, as described in Methods).

Data Table 9. TPM matrix. Filename: DataTable 9_RSEM_TPM (.xlsx,.csv,.tsv.gz). Identical structure to Data Table 8 but with cell values reported as transcripts per million (TPM).

GO enrichment results. Ten files of the form GO_Enrichment_<Treatment>_<UP|DOWN>.tsv (one per treatment per direction, comprising Salinity, PEG, Heat, Cold, and ABA × UP and DOWN). Each file is a tab-separated GOATools output with columns: GO (GO term identifier), NS (namespace; BP / MF / CC), enrichment (e for enriched, p for purified/depleted), name (GO term name), ratio_in_study (foreground term hits / foreground size), ratio_in_pop (background term hits / background size), p_uncorrected (raw Fisher's exact test p-value), depth (depth in the GO ontology), study_count (foreground hits in term), p_bonferroni, p_sidak, p_holm, p_fdr_bh (multiple-testing-adjusted p-values), and study_items (comma-separated foreground gene_ids contributing to the term).

Technical Validation

Sequencing quality was assessed using FastQC (v0.11.9), with per-read quality score distributions and read count summaries shown in Fig. 2. All RNA extractions produced high-quality material suitable for RNA-seq, and each library was generated using strand-specific protocols to minimize bias in transcript orientation.

Trimmed reads were aligned to the *S. bicolor* BTx623 reference genome (v3.1.1) using STAR (v2.7.10a), and alignment quality metrics for all samples are shown in Fig. 3. Mapping rates ranged from 71.0% to 91.9% across all 54 samples, with unmapped reads primarily classified as too short (3.7–16.1%). Duplication rates ranged from 16.2% to 27.9%, within acceptable ranges for RNA-seq libraries. No abnormalities were observed in read

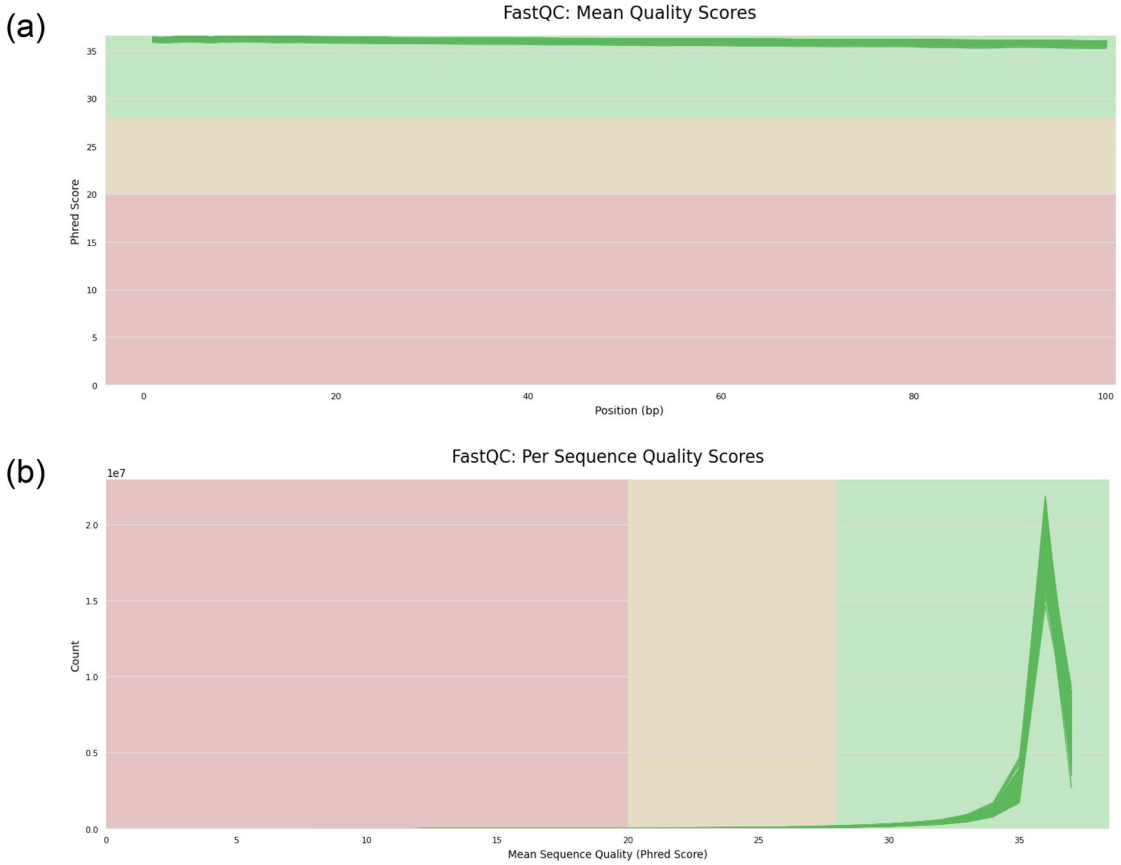

Fig. 2 Evaluation of sequence quality in raw FASTQ data. (**a**) The distribution of read counts for mean sequence quality is indicated. (**b**) The distribution of per-read mean quality scores for each sequencing run is indicated.

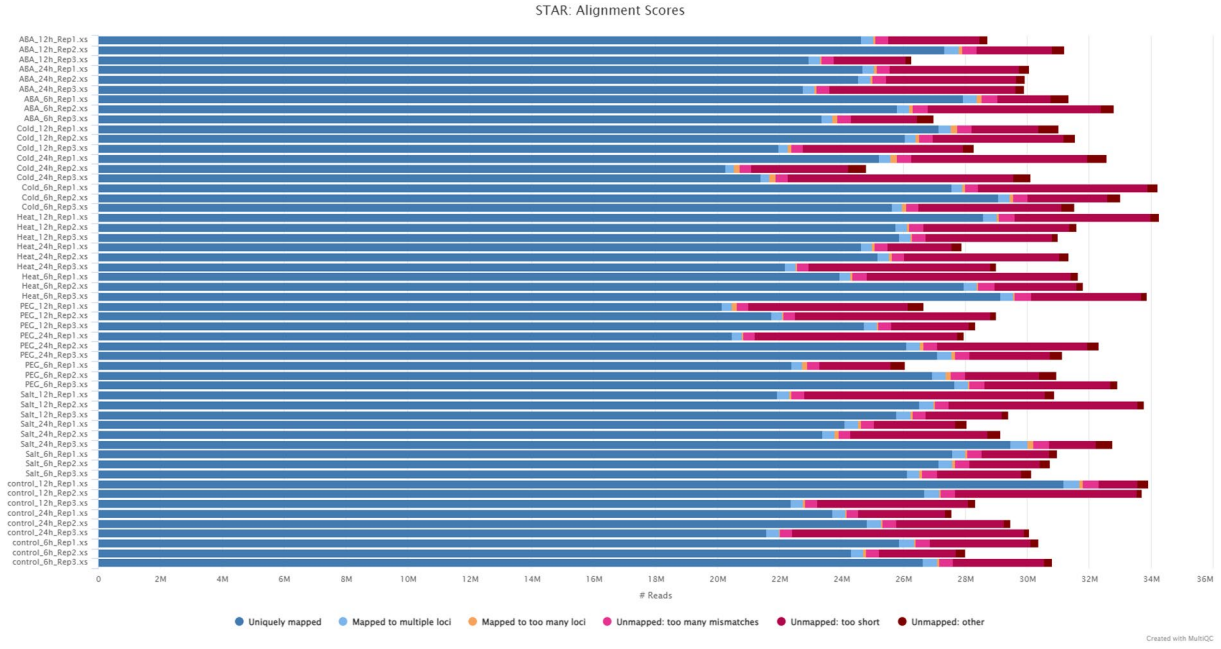

Fig. 3 Evaluation of read alignment quality. Trimmed reads were aligned to the sorghum reference genome using STAR aligner (ver. 2.7.10a). Each horizontal bar represents a sample and displays the distribution of reads according to their quality.

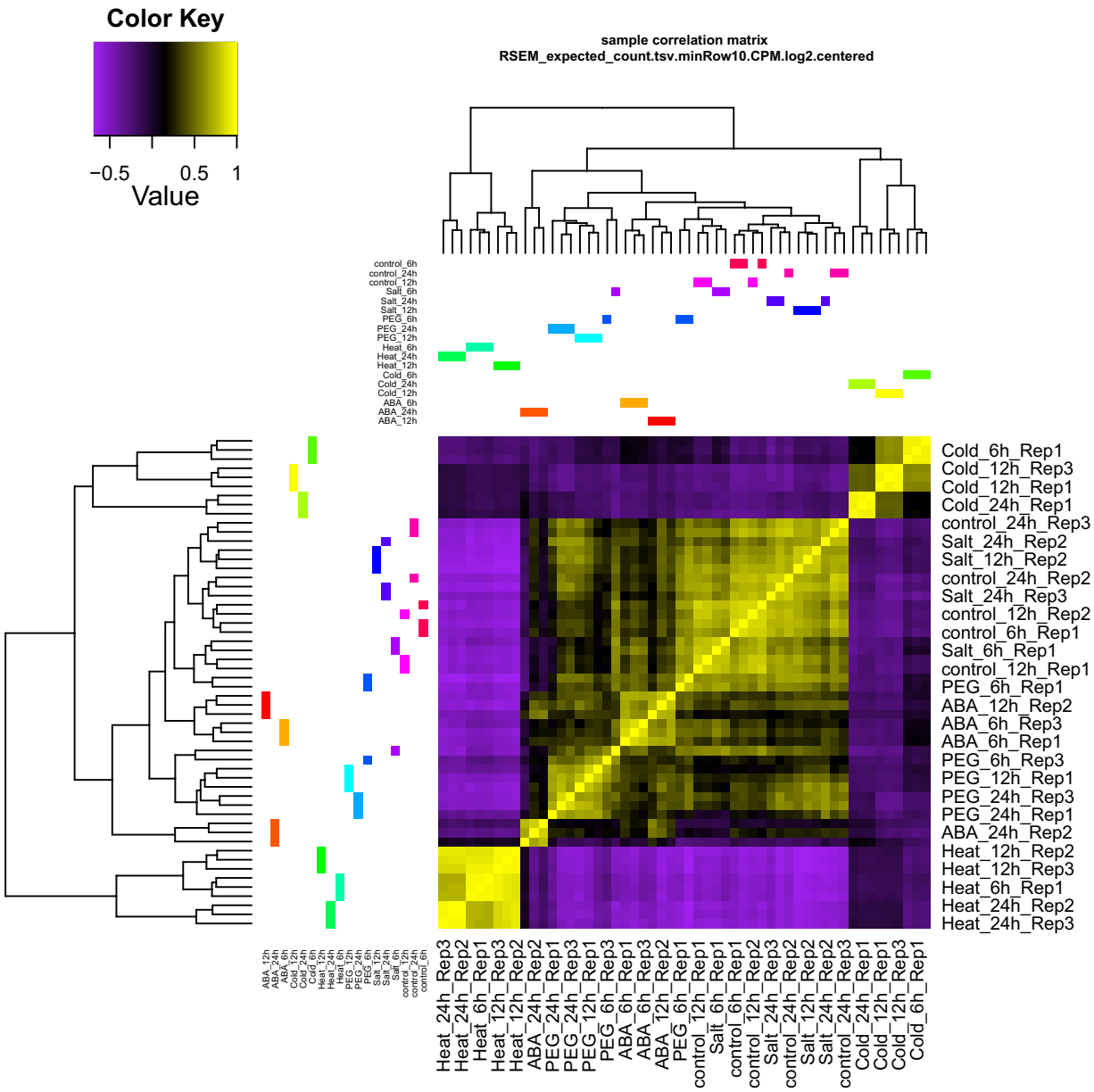

Fig. 4 Heatmap of a sample correlation matrix based on the clustering of transcriptomic data across different abiotic stress and ABA treatments and time points. The data matrix was derived from RSEM expected counts, normalized with a minimum threshold of 10 rows, centered by CPM, and clustered using Spearman correlation with the Ward method.

length distributions or base composition across samples. Sequencing depth and alignment statistics for all samples are summarized in Data Table 7[56].

Sample reproducibility across treatments and time points was evaluated by Spearman correlation of CPM-normalized RSEM expected counts, with hierarchical clustering using the Ward method[60–63] in Fig. 4. Biological replicates clustered tightly by condition and time point, confirming high reproducibility and minimal batch effects. Raw count matrices and normalized expression values are provided in Data Tables 8, 9[56]. Co-expression network data, including node attributes and edge lists with pairwise Pearson correlation coefficients, are deposited at Figshare[56].

As an additional quality check, GO enrichment analysis was performed on DEGs from each treatment to confirm that the dataset recovered transcriptional signatures already documented for these stresses. Enriched categories included protein folding-related terms under heat, translation-related terms under cold, and ion transport and redox-related terms under salinity and ABA treatments, broadly consistent with prior sorghum reports[46]. These observations are reported here as a sanity check on dataset quality rather than as new biological conclusions; full GO enrichment outputs are deposited on Figshare to support user-driven re-analysis.

Usage Notes

This dataset is intended to support comparative transcriptomic analyses of abiotic stress responses in *S. bicolor* at the seedling stage. Within a single genotype and standardized growth conditions, users can (i) identify treatment-specific or shared stress-responsive genes by combining or intersecting the per-treatment DEG tables, (ii) examine temporal dynamics across the 6, 12, and 24 h sampling points using the count or TPM matrices, and (iii) extend or rebuild the co-expression network with alternative correlation thresholds or input transformations starting from the TPM matrix. The full file index, sample naming convention, and column definitions are documented in Data Records and reproduced in the README accompanying the Figshare deposit[56]. Custom scripts implementing the read processing, STAR alignment, RSEM quantification, DESeq. 2 differential expression analysis, and GOATools enrichment workflow are available at https://github.com/wyim-pgl/Sorghum_RNA_SEQ; the documented pipeline uses the same parameter settings and software versions reported in the Methods, enabling reproduction on the deposited data or application to comparable RNA-seq experiments.

Data availability

Raw sequencing data are deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP531967[59]. All processed data (Data Tables 1–9) and GO enrichment results, are publicly available at Figshare[56].

Code availability

All custom scripts for data preprocessing, quality control, read alignment, quantification, and downstream analyses are publicly available at https://github.com/wyim-pgl/Sorghum_RNA_SEQ.

Received: 2 January 2025; Accepted: 30 April 2026;
Published online: 22 May 2026

References

1. Bashir, K., Matsui, A., Rasheed, S. & Seki, M. Recent advances in the characterization of plant transcriptomes in response to drought, salinity, heat, and cold stress. *F1000Research* **8**, https://doi.org/10.12688/f1000research.18424.1 (2019).
2. IPCC. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. Intergovernmental Panel on Climate Change (IPCC), 35-115, https://doi.org/10.59327/IPCC/AR6-9789291691647 (2023).
3. Zhu, J.-K. Abiotic Stress Signaling and Responses in Plants. *Cell* **167**, 313–324, https://doi.org/10.1016/j.cell.2016.08.029 (2016).
4. Apel, K. & Hirt, H. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* **55**, 373–399, https://doi.org/10.1146/annurev.arplant.55.031903.141701 (2004).
5. Petrov, V., Hille, J., Mueller-Roeber, B. & Gechev, T. S. ROS-mediated abiotic stress-induced programmed cell death in plants. *Front. Plant Sci.* **6**, 69, https://doi.org/10.3389/fpls.2015.00069 (2015).
6. Chong, L., Guo, P. & Zhu, Y. Mediator Complex: A Pivotal Regulator of ABA Signaling Pathway and Abiotic Stress Response in Plants. *Int. J. Mol. Sci.* **21**, https://doi.org/10.3390/ijms21207755 (2020).
7. Zhu, J.-K. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53**, 247–273, https://doi.org/10.1146/annurev.arplant.53.091401.143329 (2002).
8. Hao, H. *et al.* Sorghum breeding in the genomic era: opportunities and challenges. *Theor. Appl. Genet.* **134**, 1899–1924, https://doi.org/10.1007/s00122-021-03789-z (2021).
9. Ndlovu, E., van Staden, J. & Maphosa, M. Morpho-physiological effects of moisture, heat and combined stresses on Sorghum bicolor [Moench (L.)] and its acclimation mechanisms. *Plant Stress* **2**, 100018, https://doi.org/10.1016/j.stress.2021.100018 (2021).
10. Assefa, Y., Holman, J. D., Obour, A. K., O'Brien, D. & Prasad, P. V. V. Historic Grain Sorghum Production, Value, Yield Gap, and Weather Relation Trends. *Agronomy* **14**, 2582 (2024).
11. Bushra, S. *et al.* in Genetic Variation (eds Maia Rafael Trindade & Campos Magnólia de Araújo) Ch. 7 (IntechOpen, 2021).
12. Mundia, C. W., Secchi, S., Akamani, K. & Wang, G. A Regional Comparison of Factors Affecting Global Sorghum Production: The Case of North America, Asia and Africa's Sahel. *Sustainability* **11**, 2135 (2019).
13. Yang, L. *et al.* Harnessing the Genetic Basis of Sorghum Biomass-Related Traits to Facilitate Bioenergy Applications. *International Journal of Molecular Sciences* **24**, 14549 (2023).
14. Bhattarai, B., Singh, S., West, C. P., Ritchie, G. L. & Trostle, C. L. Water Depletion Pattern and Water Use Efficiency of Forage Sorghum, Pearl millet, and Corn Under Water Limiting Condition. *Agric. Water Manag.* **238**, 106206, https://doi.org/10.1016/j.agwat.2020.106206 (2020).
15. Chakrabarty, S. *et al.* Genetic Architecture of Novel Sources for Reproductive Cold Tolerance in Sorghum. *Front. Plant Sci.* **12**, https://doi.org/10.3389/fpls.2021.772177 (2021).
16. Hatfield, J. L. & Dold, C. Water-Use Efficiency: Advances and Challenges in a Changing Climate. *Front. Plant Sci.* **10**, https://doi.org/10.3389/fpls.2019.00103 (2019).
17. Rajabi Dehnavi, A., Zahedi, M. & Piernik, A. Understanding salinity stress responses in sorghum: exploring genotype variability and salt tolerance mechanisms. *Front. Plant Sci.* **14**, https://doi.org/10.3389/fpls.2023.1296286 (2024).
18. Zheng, H., Dang, Y., Diao, X. & Sui, N. Molecular mechanisms of stress resistance in sorghum: Implications for crop improvement strategies. *J. Integr. Agric.* **23**, 741–768, https://doi.org/10.1016/j.jia.2023.12.023 (2024).
19. Khalifa, M. & Eltahir, E. A. B. Assessment of global sorghum production, tolerance, and climate risk. *Frontiers in Sustainable Food Systems* Volume 7 - 2023, https://doi.org/10.3389/fsufs.2023.1184373 (2023).
20. Prasad, V. B. R. *et al.* Drought and High Temperature Stress in Sorghum: Physiological, Genetic, and Molecular Insights and Breeding Approaches. *International Journal of Molecular Sciences* **22**, 9826 (2021).
21. Tu, M. *et al.* Current advances in the molecular regulation of abiotic stress tolerance in sorghum via transcriptomic, proteomic, and metabolomic approaches. *Front. Plant Sci.* **14**, 1147328, https://doi.org/10.3389/fpls.2023.1147328 (2023).
22. Nicholson, R. L., Jamil, F. F., Snyder, B. A., Lue, W. L. & Hipskind, J. Phytoalexin synthesis in the juvenile sorghum leaf. *Physiol. Mol. Plant Pathol.* **33**, 271–278, https://doi.org/10.1016/0885-5765(88)90027-6 (1988).
23. Rankenberg, T. *et al.* Age-Dependent Abiotic Stress Resilience in Plants. *Trends Plant Sci.* **26**, 692–705, https://doi.org/10.1016/j.tplants.2020.12.016 (2021).
24. Sarkar, S. *et al.* in Climate-Resilient Agriculture, Vol 2: Agro-Biotechnological Advancement for Crop Production (ed Mirza Hasanuzzaman) 319-362 (Springer International Publishing, 2023).
25. Ma, S. *et al.* Genome-Wide Analysis of Abscisic Acid Biosynthesis, Catabolism, and Signaling in Sorghum Bicolor under Saline-Alkali Stress. *Biomolecules* **9**, https://doi.org/10.3390/biom9120823 (2019).
26. Mansour, M. M. F., Emam, M. M., Salama, K. H. A. & Morsy, A. A. Sorghum under saline conditions: responses, tolerance mechanisms, and management strategies. *Planta* **254**, 24, https://doi.org/10.1007/s00425-021-03671-8 (2021).

27. Alzahrani, Y., Abdulbaki, A. S. & Alsamadany, H. Genotypic variability in stress responses of Sorghum bicolor under drought and salinity conditions. *Frontiers in Genetics* Volume 15 - 2024, https://doi.org/10.3389/fgene.2024.1502900 (2025).
28. Baillo, E. H., Kimotho, R. N., Zhang, Z. & Xu, P. Transcription Factors Associated with Abiotic and Biotic Stress Tolerance and Their Potential for Crops Improvement. *Genes* **10**, 771 (2019).
29. Lee, S. *et al*. Expression Profile of Sorghum Genes and Cis-Regulatory Elements under Salt-Stress Conditions. *Plants* **11**, 869 (2022).
30. Luhua, Y. *et al*. Unlocking the Synergy: ABA Seed Priming Enhances Drought Tolerance in Seedlings of Sweet Sorghum Through ABA-IAA Crosstalk. Plant, Cell & Environment n/a, https://doi.org/10.1111/pce.15575.
31. Nagaraju, M. *et al*. Genome-wide Scanning and Characterization of Sorghum bicolor L. Heat Shock Transcription Factors. *Current genomics* **16**, 279–291, https://doi.org/10.2174/1389202916666150313230812 (2015).
32. Jin, X. *et al*. SbNAC9 Improves Drought Tolerance by Enhancing Scavenging Ability of Reactive Oxygen Species and Activating Stress-Responsive Genes of Sorghum. *Int J Mol Sci* **24**, https://doi.org/10.3390/ijms24032401 (2023).
33. Varoquaux, N. *et al*. Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proceedings of the National Academy of Sciences* **116**, 27124–27132, https://doi.org/10.1073/pnas.1907500116 (2019).
34. Zhang, D.-f. *et al*. Transcriptomic profiling of sorghum leaves and roots responsive to drought stress at the seedling stage. *Journal of Integrative Agriculture* **18**, 1980–1995, https://doi.org/10.1016/S2095-3119(18)62119-7 (2019).
35. Behera, P. P., Saharia, N., Borah, N., Devi, S. H. & Sarma, R. N. Sorghum Physiology and Adaptation to Abiotic Stresses. *International Journal of Environment and Climate Change* **12**, 1005–1022, https://doi.org/10.9734/ijecc/2022/v12i1030891 (2022).
36. Li, P. *et al*. The developmental dynamics of the maize leaf transcriptome. *Nature Genetics* **42**, 1060–1067, https://doi.org/10.1038/ng.703 (2010).
37. Hu, W. *et al*. New insights into the evolution and functional divergence of the SWEET family in Saccharum based on comparative genomics. *BMC Plant Biology* **18**, 270, https://doi.org/10.1186/s12870-018-1495-y (2018).
38. Li, N. *et al*. Nonuniform gene expression pattern detected along the longitudinal axis in the matured rice leaf. *Scientific Reports* **5**, 8015, https://doi.org/10.1038/srep08015 (2015).
39. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63, https://doi.org/10.1038/nrg2484 (2009).
40. Conesa, A. & Beck, S. Making multi-omics data accessible to researchers. *Scientific data* **6**, 251, https://doi.org/10.1038/s41597-019-0258-4 (2019).
41. Tu, M., Zeng, J., Zhang, J., Fan, G. & Song, G. Unleashing the power within short-read RNA-seq for plant research: Beyond differential expression analysis and toward regulomics. *Frontiers in Plant Science* **ume 13**, 2022, https://doi.org/10.3389/fpls.2022.1038109 (2022).
42. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11**, 1650–1667, https://doi.org/10.1038/nprot.2016.095 (2016).
43. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nature Reviews Genetics* **20**, 631–656, https://doi.org/10.1038/s41576-019-0150-2 (2019).
44. Zhao, S. & Zhang, B. A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics* **16**, 97, https://doi.org/10.1186/s12864-015-1308-8 (2015).
45. Sun, X. *et al*. MicroRNAs balance growth and salt stress responses in sweet sorghum. *The Plant Journal* **113**, 677–697 (2023).
46. Dugas, D. V. *et al*. Functional annotation of the transcriptome of Sorghum bicolor in response to osmotic stress and abscisic acid. *BMC Genomics* **12**, 514, https://doi.org/10.1186/1471-2164-12-514 (2011).
47. Young Ho Kim, I. T. O., Dong Kyu Lee, Do Yeon Jeon, Myung Chul Seo, Min Hee Nam, Byung Keun Oh. A composition comprising the extract of Sorghum bicolor L. Moench as an active ingredient for preventing and treating inflammatory disease. Korea patent KR101226824B1 (2013).
48. Do Youn, J., Hyun Joo, W., Jee Youn, K. & Young Ho, K. Anti-inflammatory Activity of Sorghum bicolor (L.) Moench var. Hwanggeumchal Grains in Lipopolysaccharide-stimulated RAW264.7 Murine Macrophage Cell Line. *Journal of Life Science* **32**, 929–937 (2022).
49. Ra, J.-E. *et al*. Antioxidative and Anti-inflammatory Activity of Extract from Milling By-products of Sorghum Cultivar, 'Hwanggeumchal'. The Korean Society. *of Crop Science* **59**, 463–469, https://doi.org/10.7740/KJCS.2014.59.4.463 (2014).
50. McCormick, R. F. *et al*. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Nat. Plants* **4**, 1003–1013, https://doi.org/10.1038/s41477-018-0286-7 (2018).
51. Huang, X. *et al*. Genome-wide association study of starch biosynthesis in sorghum. *Plant Genome* **8**, plantgenome2014.09.0046, https://doi.org/10.3835/plantgenome2014.09.0046 (2015).
52. Awika, J. M. & Rooney, L. W. Sorghum phytochemicals and their potential impact on human health. *Phytochemistry* **65**, 1199–1221, https://doi.org/10.1016/j.phytochem.2004.04.001 (2004).
53. Franks, C. D., Burow, G. B. & Burke, J. J. A. comparison of US and landrace sorghum genotypes for drought tolerance and recovery. *Crop Sci.* **46**, 1375–1382, https://doi.org/10.2135/cropsci2005.09-0337 (2006).
54. Mace, E. S. *et al*. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **4**, 2320, https://doi.org/10.1038/ncomms3320 (2013).
55. Lim, S. D. *et al*. Role of OsCZMT1 in Na+ and Mg2+ transport and salinity insensitivity. *Environ. Exp. Bot.* **194**, 104754, https://doi.org/10.1016/j.envexpbot.2021.104754 (2022).
56. Yim, W. C. *et al*. A comprehensive transcriptomic dataset of Sorghum bicolor seedlings under abiotic stress conditions. *Figshare* https://doi.org/10.6084/m9.figshare.32129923 (2026).
57. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol.* **15**, 550, https://doi.org/10.1186/s13059-014-0550-8 (2014).
58. Shannon, P. *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498–2504 (2003).
59. Yim, W. C. *et al*. *Sorghum bicolor* seedlings under abiotic stress conditions. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP531967 (2025).
60. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* **31**, 274–295, https://doi.org/10.1007/s00357-014-9161-z (2014).
61. Janská, A., Maršík, P., Zelenková, S. & Ovesná, J. Cold stress and acclimation-what is important for metabolic adjustment? *Plant Biology* **12**, 395–405 (2010).
62. Munns, R. & Tester, M. Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* **59**, 651–681 (2008).
63. Mittler, R. & Blumwald, E. The roles of ROS and ABA in systemic acquired acclimation. *The Plant Cell* **27**, 64–70 (2015).

Acknowledgements

This research was supported by the Regional Innovation System & Education (RISE) program through the Gangwon RISE Center, funded by the Ministry of Education (MOE) and the Gangwon State, Republic of Korea (2026-RISE-10-005); the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) (RS-2025-16066223); the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Agriculture and Food Convergence Technologies Program

for Research Manpower Development Project, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (RS-2024-00400922); and the U.S. National Science Foundation (NSF) under Award No. OAC-2427408.

Author contributions

W.C.Y., S.D.L., Y.K.H., S.Y.Y. - Designed the study; S-H.K., Y.K.H., D.Y.S., N.V.P. - Prepared abiotic stress treatments, sampling, total RNA isolation; W.C.Y., J.S.L., S.Y.Y. - Analyzed the data and in silico analysis; W.C.Y., J.S.L., S.D.L.,Y.K.H., S.Y.Y. - Wrote original draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-026-07425-7.

Correspondence and requests for materials should be addressed to S.Y.Y. or S.D.L.