# A field-focused systems approach reveals mRNA covalent modifications linked to sorghum growth and development under drought stress

Li'ang Yu[1], Giovanni Melandri[2], Anna C. Nelson Dittrich[1], Sebastian Calleja[2], Kyle Palos[1], Aparna Srinivasan[1], Emily K. Brewer[1], Riley Henderson[1], Ciara Denise Garcia[2], Xiaodan Zhang[1], Bruno Rozzi[2], Andrea Eveland[3], Susan J. Schroeder[4], David Stern[1], Aleksandra Skirycz[5], Eric Lyons[2], A. Elizabeth Arnold[2,6], Brian D. Gregory[7]*, Andrew D. L. Nelson[1]*, Duke Pauli[2,6]*

[1]Boyce Thompson Institute, Cornell University, Ithaca, NY 14850, USA
[2]School of Plant Sciences, University of Arizona, Tucson, AZ, 85721, USA
[3]Donald Danforth Plant Science Center, Olivette, MO, 63131, USA
[4]Department of Chemistry and Biochemistry, The University of Oklahoma, Norman, OK, 73019
[5]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48824
[6]Center for Agroecosystem Research in the Desert (ARID), University of Arizona, Tucson, AZ 85721
[7]Department of Biology, University of Pennsylvania, Philadelphia, PA, US

***Address correspondence to:**
Dr. Duke Pauli
The University of Arizona
School of Plant Sciences
1140 E. South Campus Drive, Tucson, AZ, 85721
dukepauli@arizona.edu

or

Dr. Andrew D. L. Nelson
The Boyce Thompson Institute
Cornell University
533 Tower Road, Ithaca, NY, 14850.
an425@cornell.edu

or

Dr. Brian D. Gregory
University of Pennsylvania
Department of Biology
103H Lynch Laboratory
433 South University Avenue, Philadelphia, PA, 19104
bdgregor@sas.upenn.edu

**Current Email addresses:**
LY: ly76@cornell.edu, GM: gmelandri@arizona.edu, ACND: an426@cornell.edu, SC: scalleja@arizona.edu, KP: krp75@cornell.edu, ApS: as2967@cornell.edu, EKB: ekb78@cornell.edu, RH: rnh53@cornell.edu, CDG: ciarag@arizona.edu, XZ: xz828@cornell.edu, BR: XXXXXX, AE: AE: veland@danforthcenter.org, SJS: susan.schroeder@ou.edu, DS: ds28@cornell.edu, AlS: skirycza@msu.edu, EL: ericlyons@arizona.edu, AEA: barnoldaz@gmail.com,

**Running title:** The RNA modification Dihydrouridine positively impacts plant stress responses


## Abstract

Covalent RNA modifications (RCMs) are post-transcriptional changes to the chemical composition of RNA. RCMs influence mRNA stability, regulate transcription and translation efficiency, and play critical roles in the growth and development of eukaryotes. However, their role in plants remains poorly understood, particularly as they relate to performance in the field under stress conditions. In this study, we grew a panel of six diverse sorghum (*S. bicolor*) accessions in the field during the summer in central Arizona and examined their physiological and molecular responses to drought and heat stress over time. We then explored the molecular features that contributed to plant performance under stress. To do so, we combined genomic, transcriptomic, epitranscriptomic, physiological, and metabolomic data in a systems-level approach. Co-expression network analyses uncovered two modules of interest, one controlled primarily by a single stress-responsive transcription factor, SbCDF3, and the other by an RCM, dihydrouridine. While the CDF3 module largely contained a set of stress response and photosynthesis-associated genes that were positively correlated with plant performance, the dihydrouridine-associated module was largely comprised of photosynthesis and metabolism genes, including SbPPDK1, which is integral to C4 photosynthesis in the grasses. In addition, the transcript encoding the enzyme responsible for this RCM, dihydrouridine synthase (*SbDUS2*), was also present in this module, and its abundance was positively correlated with photosynthetic traits and *SbPPDK1* abundance. Given that this highly conserved RCM has never been characterized in plants, we examined loss of function mutants for the DUS2 enzyme in Arabidopsis, demonstrating decreases in plant growth and performance under heat stress in this background. Our work highlights both a key transcription factor, CDF3, for breeding in the *Poaceae*. In addition, for the first time in plants, we reveal a role for the RCM dihydrouridine in modifying conserved, core metabolic and photosynthesis-associated transcripts in plants.



**Keywords:**

Drought response, RNA covalent modifications, Gene regulatory network, Systems biology

## Introduction

*Sorghum bicolor* (sorghum) is the fifth most important crop worldwide and is grown for grain, brewing, and cellulosic biomass production. As a member of the *Poaceae* and a close relative of maize, sorghum is believed to have been domesticated around 4,000–6,000 years ago in Africa (Winchell et al., 2017), followed by the formation of a variety of landraces with diversified morphological variation due to selection across broad geographical locations[1]. Most sorghum cultivars were historically, and are currently, grown in arid and semi-arid areas due to their adaptability to limited watering during growth. However, prolonged water scarcity reduces germination rates and leads to declined grain yield and quality, as well as reductions in other agronomic traits [2,3]. Yield, like many other agronomic traits, is a complex trait influenced by a number of developmental and environmental cues. Thus, a better understanding of the genetic mechanisms that interpret and respond to these environmental cues is essential for maintaining yield in sorghum under stress conditions [4]

Drought is a form of abiotic stress that leads to numerous unfavorable phenotypic and physiological changes such as reduced yield and altered plant growth, and can lead to increased susceptibility to biotic stressors. Plant phenotypic responses to reduced water availability are often paired with physiological and molecular responses, with more tolerant plants increasing their water use efficiency (WUEi), more tightly controlling transpiration rates, but also reducing their $CO_2$ net assimilation. Metabolite profiles and enzyme activity levels also change in more tolerant accessions in response to drought, as seen by the elevation of certain sugars and amino acids that act as osmoprotectants. [5–7] Combined, these phenotypic and physiological traits can serve as key indicators to interpret the levels of drought tolerance among different accessions within a species.

These drought responses are regulated at the cellular level through a complex set of regulatory control points guided by the activation and repression of transcription factors that in turn regulate the expression of their target genes [8–10]. These regulatory networks receive and interpret cues from the environment and then govern the transcriptional cascades that ultimately link stress signaling to downstream phenotypes such as yield [11,12]. Systems-level approaches have been quite successful at identifying the elements that influence stress and yield in a number of agronomic species [10,13,14] . While some elite sorghum accessions have shown strong drought tolerance phenotypes during breeding and selection, the crosstalk between many upstream regulatory genes and downstream trait-associated genes remains unclear, particularly on field-derived data and even more so across the diversity within a species. Thus, we are still limited in the genetic and molecular targets at our disposal to develop new cultivars that show improved stress response without the off-target effects that come from expressing a transcriptional regulator out of context.

Often the act of transcription is seen as a direct proxy for translation, and therefore insights are derived on protein function based on RNA abundance. However, RNAs encounter several co- and post-transcriptional events that influence their stability, translation, and even cellular localization, which are not immediately apparent in RNA-sequencing data. One such event is the addition of RNA covalent modifications (RCMs). RCMs were initially discovered in the 1970s on tRNAs and rRNAs, but are now known to impact all classes of RNAs in the epitranscriptome (e.g., mRNAs, lncRNAs, and miRNAs) and influence their stability, translation efficiency (mRNAs), structure, and splicing [15–17]. To date, more than 100 RCMs have been discovered, predominantly on housekeeping RNAs in mammalian systems. However, RCMs, and in particular N6-methyladenosine ($m^6A$), are known to influence plant development and stress responses [15,18,19]. Several other RCMs, such as those that impact structure by disrupting Watson-Crick-Franklin base pairings, like dihydrouridine (D), pseudouridine (U), N1-methyladenosine (m1A), and 5-methylcytidine (m5C), are less studied but also believed to be critical for RNA turnover and translation. However, the functional roles of these latter RCMs are poorly understood in plants, predominantly due to difficulty in their detection.

Detection of RCMs generally relies on antibody-based ($m^6A$) or modification-specific approaches using chemical probing or conversions (m5C and U, respectively) [20–22]. However, RCMs that occur along the Waston-Crick-Franklin base pairing edges can be detected in Illumina RNA-seq data due to nucleotide misincorporation during the reverse transcription step of library preparation. This RT-induced nucleotide misincorporation is non-random, and several statistical algorithms have been developed to infer modification class from the pattern of non-biallelic SNPs present in conventional transcriptome sequencing [23,24]. These algorithms provide us with the ability to monitor both transcript abundance and RCM abundance genome-wide and associate these molecular phenotypes with changes at the whole plant level in response to environmental perturbations.

To identify the species-wide genetic basis by which sorghum effectively responds to drought stress, we leveraged a systems biology approach to study six phylogenetically distinct sorghum accessions with different levels of drought tolerance. These accessions were field evaluated in the hot, arid environment of the Arizona low desert and exposed to two different watering regimes throughout their development. We collected paired transcriptomic, epitranscriptomic, metabolomic, physiological, as well as leaf-level and whole plant phenotypic data at multiple points throughout the season to monitor responses to soil water conditions. Gene co-expression networks and transcription factor (TF) binding data were integrated with all of the other measured traits to identify the sets of TFs conferring the phenotypic responses to drought that we observed. Epitranscriptomic data, inferred from Illumina RNA-seq, were also integrated into our analyses, uncovering associations between RCM abundance, phenotype, and plant performance within these accessions. The levels of one such RCM,

Dihydrouridine (DHU), and the enzyme responsible for its deposition, Dihydrouridine Synthase (DUS) were found to be positively associated with plant performance traits (e.g., $CO_2$ net assimilation, Fv'/Fm') during drought treatment. Despite its deep evolutionary conservation in living organisms, DUS is minimally characterized in plants. Thus, we examined loss of function DUS2 mutants in Arabidopsis, where we discovered reduced germination, growth rates, and $CO_2$ net assimilation under favorable conditions that were exacerbated under heat stress. A transcriptomic time course comparison between Arabidopsis wild type and *dus2* mutants under control and heat conditions revealed a decrease in (DHU) levels and a concomitant reduction in photosynthesis associated transcript abundance, similar to what we observed in sorghum in the field. In sum, our integrative systems approach uncovered protein and RNA-level control points that impact responses to drought stress in sorghum and are strong candidates for targeted breeding approaches.

# Methods and Materials

## Sorghum growth and phenotypic measurement

Six selected sorghum accessions (**PI 533871: M1, PI 533961: DL/59/1530, PI 656041: 80M, PI 656053: N290-B, PI 656076: SC1271, and PI 656096: SC391**) were grown underneath the Gantry at the University of Arizona Maricopa Agricultural Center (MAC). All plants were watered at the same rate (~24% soil volume metric water content, approximately field capacity) until flag leaf appearance in the whorl (~47 days after planting), after which irrigation was reduced to a subset of plots to mimic drought treatment (Water limited, WL; 16% SVWC).

The destructive plant phenotyping was conducted at the end of the season. Traits that were measured included above-ground biomass, root biomass, and three-leaf weights (five replicates per plant). For other time-series molecular phenotypes, samples were collected weekly (on Thursdays) for seven weeks (Week 1-Week 7, from August 13th to September 24th, 2020) on Range 28 of the gantry field at MAC (**See supplementary note for details**). For oxidative stress enzyme activity, samples consisting of six leaf disks from five different plants per accession and treatment per time point were collected in a 1.5 mL tube and immediately snap-frozen in liquid nitrogen. Two leaf disks were used for the oxidative stress analyses based on the sample fresh weight (FW). Licor-6800 derived photosynthesis-related traits (***N = 29,*** https://www.licor.com/) were measured with the following settings: Flow=600; Fan=12,000 rpm; $CO_2$= 420 ppm; Light= 2,000. The top fully expanded leaf on the main stem of each plant was measured and data were logged as soon as assimilation and stomatal conductance curves (monitored in real-time on the instrument display) when reaching a plateau (on

average 30 seconds after a leaf was clamped). The 42 selected stress-responsive metabolites were quantified using the untargeted liquid chromatography-mass spectrometry (LC-MS). Best Linear Unbiased Estimators (BLUEs) of each phenotypic trait and metabolite value were generated for each sorghum sample.

## ***Arabidopsis*** **growth and phenotypic measurement**

Arabidopsis Col-0 and *dus2* mutant lines (**Line ID:** SALK_079129C, CS857388, and SALK_023006, https://abrc.osu.edu/) were grown at the growth chamber at the Boyce Thompson Institute (BTI), Ithaca NY. The growth conditions were set as follows: temperature (22°C: control condition, 37°C: heat treatment), humidity: 60%, and lighting period: 8:00 - 21:00). All seeds were stratified at 4°C condition for 48~72 hours after planting. For image-based phenotyping, plants were placed under Raspi-derived phenotypic tools (Pheno-Rigs [25]) to collect time-series plant images (**imaging-interval: 30 mins, length: 12 days**) and processed by PlantCV pipelines to extract rosette area represented by RGB images, followed by statistical analyses and plots described in previous tools [25]. Germination assays were performed on ½ MS medium with Col-0 and *dus2* mutant plant seeds placed side-by-side under 22°C conditions [26], and the number of germinated seeds was recorded during each 3-hour interval for 27 hours after seed stratification. The photosynthetic efficiency of *dus2* and Col-0 lines was evaluated using a gas exchange approach with the Licor-6800 (https://www.licor.com/). Briefly, plants were maintained in the growth chamber under short-day (SD) growth conditions (photo-period: **8:00-17:00,** temp: **22°C**) to ensure a sufficient rosette area to cover the $2cm^2$ small chamber of the Licor-6800 device. The 4-week seedlings under 22°C were used as the control group and plants were placed under 37°C for 24 hours for heat treatment before the Licor-6800 measurement. Parameters were set based on previous leaf gas exchange experiments[27] (flow rate: 400 μmol $s^{-1}$, relative humidity of air: 55–60 %, reference $CO_2$: 400 μmol $^{-1}$, mixing fan speed: 10,000 rpm, heat exchanger temperature: 22°C or 37°C, chamber pressure: 0.2 kPa, light source composition 90 % red 10 % blue light, measurement time: between ZT4 (12:00) and ZT6 (14:00)).

## **RNA-seq libraries construction**

RNA extraction was performed using methods from previous studies for both Arabidopsis and sorghum tissue (**Pang et al., 2011; Wu et al., 2002**). Briefly, hot borate buffer was prepared containing 0.2 M sodium borate pH 9.0, 30 mM EGTA, 1% SDS, and 1% sodium deoxycholate., The PVP-40, NP-40, and DTT were added to the hot borate buffer to a final concentration of 2%, 1%, and 10 mM, respectively before the ise. During the extraction, a total of approximately 0.75 ml hot borate buffer per

50 mg tissue (~5 sorghum leaf disks) was used. Each sample was initially ground in 250 ul of hot buffer (80 °C) supplemented 15 µl of 20 mg/ml Proteinase K (**Roche #46950800/Sigma #3115887001**). After tissue homogenization, additional hot buffer was added to a final volumen of 750 ul. Lysate was added to a Qiashredder column (**Qiagen #7965**6) and centrifuged at 13,000 × for 1 minute. Flow through was added to 0.5 volumes of 100% ethanol. The mixture was used as input for RNA cleanup using the RNeasy kit (**Qiagen #74104**). The final RNA-seq libraries were then constructed using the Amaryllis Nucleics Full Transcript library prep kit (YourSeq Duet, https://amaryllisnucleics.com/kits/duet-rnaseq-library-kit ).

## Identification and annotation of RNA covalent modifications (RCMs)

Identification of RNA covalent modifications (RCMs) was performed by intersecting results derived from high-throughput annotation of modified ribonucleotides[23,28] (**HAMR v1.4**) and ModTect[24] (**v1.1.5.1**). Initially, the paired-end stranded RNA reads were trimmed by "trimmomatic" tools [29] (**v0.32**, parameter: LEADING:3 TRAILING:3 MINLEN:36, TruSeq3-PE.fa:2:30:10:2) to remove adaptors and low-quality reads. For HAMR, the forward and reverse reads were mapped to the sorghum reference genome respectively[30] (**v3.1.1**) using Tophat2 (v2.1.1, --read-mismatches 12 --read-edit-dist 12, --max-multihits 10, --transcriptome-max-hits 10, --library-type fr-second strand, and --b2-very-sensitive)[31], followed by mapped reads merge, multiple-mapped reads removal, and splicing alignment resolvement using Picard (v3.1.1, http://broadinstitute.github.io/picard) and GATK[32] (**v3.3.5**). The clean alignment files were processed by HAMR with a stringent parameter setting, as studies been performed previously: min_read_qual 30, min_read_coverage: 50, seq_error_rate0.01, hypothesis $H_4$, max_FDR 0.05, which filtered out potential modified sites with a < 50 mapping depth with high-confidence (FDR < 0.05) to call modified sites [28,33]. For ModTect prediction, trimmed reads were mapped to the sorghum genome and associated annotation files as above using STAR[34] (**v 2.7.10b**, with the default arguments were used for mapping: --outFilterIntronMotifs RemoveNoncanonical, --alignIntronMax 10000, --out SAMstrandField intronMotif. Raw BAM outputs were then used as input for ModTect with the following arguments: –minBaseQual 30, --readlength 150, and --regionFile genes. To harvest the high-confidence PTMs, only the overlapped sites derived from two replicates and two methods were retained as the modified sites for each sorghum sample.

To simplify the annotation of RCMs relative to transcripts, only primary transcripts (the longest transcripts per gene) were selected to group the number of modified sites that fell into respective exon regions. The output file of HAMR was merged with the gene annotation (primary transcripts only) to assign sites to transcripts. Here, three criteria were applied to characterize the RCMs, including the presence and absence of RCMs per transcript (**PAM**), the number of modified sites per transcript

(**modification density**), and the ratio of modified reads over total mapped per transcript (**modification frequency**).

## Characterization and comparative analyses of RCMs

The distribution pattern of RCMs was studied by classifying modified sites across the genome by comparing the abundance of single-nucleotide modified sites over the 5' UTR, coding sequence regions, and 3' UTR of respective primary transcripts using the metaplotR package and in-house R scripts package (https://github.com/olarerin/metaPlotR ). Further, the density and portions of different types of RCMs over each sample were calculated using in-house R scripts (https://github.com/Leon-Yu0320/Sorghum_Range28). For those sites modified with particular PTMs, we harvested the 6 bp extended flanking regions of each modified site under each type of PTM to explore their consensus motifs using the MEME tool[35] (parameter: E-value: 0.05, minimum width: 5bp, maximum width: 10bp). These motifs which shared statistical consensus were mapped to their respective genes for functional enrichment.

## Genomic analyses of sorghum genome and population

Orthologs of important sorghum candidate genes in other species were identified using CoGe blast (https://genomevolution.org/coge/CoGeBlast.pl ) with the following parameter (E-value: 1e-5, word_size: 3, matrix: BLOSUM62) to identify the best hits from *Arabidopsis*[30] (TAIR 11 and **Sbicolor_454.v3**). The genomic variations of the six sorghum accessions were identified and annotated using the published sorghum association panel resequencing data[36]. Briefly, cleaned DNA reads were mapped to the sorghum reference genome (**v3.1.1**) using BWA-mem with default parameters[37] (**v0.7.17-r1188**). Duplicated read-pairs and low-quality mapping in each sample were cleaned using "sambamba"[38] (**v1.28.1,** parameter: mapping_quality >= 1 and not (unmapped or secondary_alignment) and not ([XA] != null or [SA] != null) followed by reads sorting using "samtools"[39]. Variant calling was conducted using the GATK4 "HaplotypeCaller" function[32] (**v4.2.5.0**, Parameter: --max-alternate-allele:6, --Ploidy:2). The single nucleotide polymorphisms (SNPs) were filtered using GATK hard-filtering protocol 'QD < 2.0||FS > 60.0|| MQ< 40.0||MQRankSum < 12.5||ReadPosRankSum < 8.0', and insertions and deletions (INDELs) with parameters 'QD < 2.0||FS > 200.0 || ReadPosRankSum < 20.0'. Those high-quality SNPs and INDELs were further annotated by variants effect predictor (**VEP**) using the *S.bicolor* gene annotation[30,40]**.** For phylogenetic analysis, we retained the bi-allelic SNPs at the four-fold degenerate (4DTV) loci filtered by plink[41] (v1.90b4**,** --types SNPs -m 2 -M 2) that derived from 365 out of 407 sorghum accessions with complete geographical

information[36]. Using the SNP filtering with minor allele frequency (MAF > 0,05, missing data < 0.1) and linkage disequilibrium (LD) prune[41] (**plink v1.90b4** parameter: --indep-pairwise 1000 500 0.2), a total of 7,520 SNPs were kept to reconstruct the phylogenetic tree using IQtree [42](**v1.6.12,** Parameter: -nt 100 -m MFP).

## Transcriptome data processing

Transcript abundance was measured using the same sets of RNA reads alignment files used for RCM identification. The FeatureCounts tool was used for quantification using the mRNA (transcripts only) as the meta-level to count reads[43] (**v2.0.1,** parameter: -s 1 -p -t mRNA -g ID -O). Read counts of 34,130 primary transcripts were normalized using DESeq2 after the removal of transcripts with less than an average of 1 count across 96 samples[44] .We incorporated the time point (Week 1, Week 3, Week 5, and Week 7), accession IDs (A, B, C, D, E, F), and watering conditions (control and drought) as the experimental design factors. Then condition-wise differentially abundant transcripts (DATs) over each time point per accession were identified using the cutoff of adjusted P-value < 0.05. A $\log_2$-transformed fold change (FC) greater than 1 was used to filter up-regulated transcripts and a less than -1 $\log_2$FC filtered the down-regulated transcripts.

## Construction of co-expression network

The co-expression network was constructed using the WGCNA package with variance stabilizing transformed (VST) normalized read counts with 34,130 transcripts across 96 samples[45]. Initially, a soft threshold (Beta score) was applied to fit the scale-free topology model assumption which generates a minimum number to make the index curve flatten out reaching a high value (> 0.85)[13]. Further, the adjacency matrix represented by gene expression was transformed into a topological overlap matrix (TOM) to classify genes using a dynamic tree-cut approach. Finally, genes that shared high-confidence correlation (abs(coefficient) > 0.8) were merged into the same module (parameter: deepsplit = 4, tree cut height = 0.2). Further, the pairwise Pearson correlation between each gene from modules and the module-wise eigengene values was tested to characterize each gene's module membership (MM) under the same module. The weight score, a parameter to denote the robustness of connections of each edge, was filtered by taking genes associated with the top 10% high-confidence weight ranking connections for downstream functional analysis. Lastly, the metabolite profiles, oxidative stress-related enzyme activities, and Licor-6800-derived traits were integrated into the network using Pearson correlation (PC) between gene expression and these molecular phenotypes. Trait-correlated modules were selected based on their correlation level between MM scores and phenotypes (cutoff: r > 0.5, or r

< -0.5, *P-value* < 0.05). The correlation between gene expression and phenotypes was estimated as gene significance (GS) to interpret the importance of genes related to each molecular phenotype.

## Functional and regulatory network analysis

To study the transcription factors (TFs) and their downstream target genes (protein-DNA interactions), TFs and protein kinases from each co-expression module were predicted using the iTAK tool based on conserved protein-coding domains[46] To identify potential upstream regulators of genes in each module, we conducted the enrichment of consensus binding motifs within 2kb upstream promoter region of genes in the same module by selecting *Arabidopsis* DAPseq consensus motif database ( adjusted *P-value:* 0.05) to perform analysis of motif enrichment (AME parameter: --scoring avg --method fisher --hit-lo-fraction 0.25)[35,47]. The protein-protein interactions (PPI), PPIs among genes present in the co-expression network were acquired from the STRING database as the third layer of evidence to infer the functions or pathways of modules (parameter: confidence level: > 0.7, source: experiments and database source)[48]. As a result, co-expression patterns, PPIs, and TF-binding interactions were merged into an integrated network.

## Functional enrichment analysis

Functional enrichment analysis of pathways (Kyoto Encyclopedia of Genes and Genomes (KEGG) database) and Gene ontology (GO) was performed to infer the biological function of genes associated with certain RCMs, transcripts assigned in co-expression modules, and differentially abundant transcripts. Here, we used a customized database of GO and KO information derived from sorghum functional annotation (https://phytozome-next.jgi.doe.gov) by topGO and clusterprofile R packages (**Alexa et al., 2016, Yu et al., 2012, cut-off:** P-adj or FDR < 0.05). For each KEGG term significantly enriched (adjusted *P-value* < 0.05).

## Data visualization and usage of databases

All statistical visualizations were generated in R (v4.3.2)[49] with multiple packages (https://rpubs.com/LeonYu). Briefly, plots include the line, boxplot, PCA, enrichment graphs generated by the “ggplot2” package[50], and heatmaps using the “pheatmap” package[51]. The integrated gene networks containing TF-binding, protein-protein interaction (PPI), and co-expression derived from WGCNA were generated using Cytoscape[52]

# Results

## Genomic diversity contributes to phenotypic variation among sorghum accessions

Phenotypic diversity under water-limited conditions is ultimately traced to variation in genomic sequence and rewiring of gene expression. Therefore, to capture the genetic and phenotypic diversity present within sorghum, we initially generated an unrooted phylogeny from publicly available whole genome skim sequencing data of 356 out of 400 accessions with demographic information found within the Sorghum Association Panel (SAP)[30,36]. We developed a small but diverse panel of sorghum accessions by combining this phylogeny with performance data from 356 sorghum accessions grown in the field in Arizona in 2019 under standard field conditions (**Figure 1A**). From these 356 accessions, six were selected based on seed viability, genetic distinctiveness, and phenotypic variation (shoot and root biomass, plant height, and yield). Among the selected lines, four accessions (A, B, E, and F) are African accessions obtained from four countries across the continent (**Figure 1A**). The other two accessions (**C** and **D**) are U.S. breeding accessions. Variant calling and annotation for SNPs and INDELs were performed to more closely examine the genomic variation between these accessions. In total, 8,725,449 variants were identified among the six accessions relative to the published reference genome and more than 50% of those variants were located in intergenic regions, followed by gene regulatory regions (21.00% in the upstream region and 19.5% in the downstream region, **Figure S1, left panel**). Different proportions of unique SNPs and INDELs were uncovered among the six accessions, with each genotype contributing around 10-20% of genomic variation to the panel (**Figure S1 right panel**). In sum, these data suggest that we were successful in our goal of developing a genetically diverse panel for identifying drought and heat tolerance-associated mechanisms in sorghum.

Based on our initial field experiments to select these six accessions, we observed phenotypic variation under standard growth conditions, although all six accessions successfully reached reproductive maturity. To assess performance under water-limiting conditions, we established all six genotypes in replicated plots with field trials at the University of Arizona Maricopa Agricultural Center (MAC). All plants were watered at the same rate (~24% soil volume metric water content) until flag leaf appearance in the whorl (~ 47 days after planting), after which irrigation was reduced to a subset of plots to mimic drought treatment (Water limited, WL; 16% SVWC). Phenotypic and molecular sampling occurred during the 1st, 3rd, 5th, and 7th week (Week 1-7) after the implementation of drought treatment with destructive measurements (biomass, etc) occurring at the end of the season. Treatment effects were readily apparent at the end of the season (**Figures 1B-C**), with the M1 (“A”) accession showing the most pronounced leaf necrosis and alterations to root structure. The “A” and “D” (N290-B)

accessions had the lowest total biomass under both well-watered (control) and water-limited (drought) conditions, potentially indicating poor adaptation to heat stress (**Figure 1D**). In contrast, the "B" (DL/59/1530) and "C" (80M) accessions displayed a minimal reduction in biomass or leaf weight in response to drought stress. Interestingly, the "A" accession showed an increase in root height and width under drought relative to control conditions, indicative of a water foraging mechanism at play (**Figure 1E**). Leaf level physiological traits such as water use efficiency (WUE), Fv'/Fm', and internal $CO_2$ concentrations exhibited similar results, with the "A" accession displaying hallmarks of water stress, particularly at the later time points (Week 5 and Week 7, **Figure 1F**), whereas other accessions, such as "B" and "C", showed a robust response. Thus, in addition to genotypic diversity, these data suggest that this panel of six accessions displayed robust phenotypic diversity to drought stress and may provide information on the molecular mechanisms by which sorghum responds to water limiting conditions.

## Sorghum accessions displayed molecular response variation in their drought response

To characterize the molecular responses to drought stress in our panel, we next assessed the change in RNA abundance over the two months of the treatment (Week 1-7). Initial quality control revealed high correlation between replicates (Pearson Correlation; r > 0.9, **Supplementary Table 1**), along with high mapping rates to primary transcripts among samples (**Supplementary Table 2**). Using hierarchical clustering and Pearson Correlation analysis of global transcript abundance (***N = 22,314***, mean TPM > 1), 96 samples (**2 rep * 4-time points * 2 conditions * 6 accessions**) were clustered into subgroups that fit their genotypic classification (**Figure 2A**). Separation due to treatment effect was observed for accession "B", whereas accession "A" W3-7 timepoints for control and treatment were separated from W1, suggesting the combined stresses of heat and drought were having a distinct impact on this accession (**Figure 2A**). This global view of the transcriptome suggests that these accessions respond to stress in unique ways that are largely driven by genotype, not time or treatment.

To unpack the transcriptomic responses we observed in response to treatment in these six accessions, we examined the transcripts that were most highly variable among all 96 datasets as well as those that were differentially abundant between treatment conditions (**Supplementary Table 3)**. The top 10% of most variable transcripts were selected based on the standard deviation of the global transcript abundance (***N = 22,31***, mean TPM > 1) for principle component analysis (PCA). This approach uncovered strong genotypic and treatment effects represented by both PC1 and PC2 (**PC1 = 19.35%, PC2 = 14.88%**, **Figure 2B**). In agreement with the hierarchical clustering, samples from the six accessions clustered separately, with accession "C" exhibiting the least, and accession "A" the most, separation due to stress or time **(Figure 2B**). Pairwise comparisons of differentially abundant

transcripts (DATs; *Padj* < 0.05) between control and treatment for each timepoint and genotype uncovered different transcriptional responses to treatment. Accession "C" showed the lowest response to treatment, whereas accessions "E" and "A" showed the most DATs across the four timepoints (**Figure 2C**). In addition to vastly different numbers of DATs among accessions, these DATs were often specific to a genotype at any given time point during the season (**Figure S3**). Gene ontology (GO) term enrichment of these DATs (up and down-regulated, Padj < 0.05) reinforced prior observations that each of these genotypes was responding to the treatment by modulating unique pools of transcripts (**Figure 2D**). Interestingly, the "A" and "F" accessions did not appear to be inducing a molecular stress response, as "stress response" GO terms were absent for both of these accessions (**Figure 2D**). Thus, one explanation for the decreased biomass and survival rates for accessions A and F could in part be explained by a lack of initiating an adequate stress response.

To gather additional clues as to how these accessions are responding to drought treatment at the molecular level, we next examined the stress-induced changes of metabolites, oxidative stress markers and antioxidant enzyme activities in the sorghum leaves[53,54]. Metabolites were sampled at the same time points as RNA, with gas-chromatography mass spectrometry (GC-MS)-based quantification and annotation of primary metabolites and spectrophotometric-based quantification of oxidative stress traits. A PCA of global changes in metabolite levels across our panel revealed that, as with the transcriptome, genotype explained most of the variation in the metabolome (**Figure 2E**, left panel; **PC1: 73.12 %**). Interestingly, samples associated with accession "A", weeks 3-7 (both treatments), separated from all other samples, and this separation was largely driven by sugars (sucrose, glucose, and fructose) and chlorogenic acid (**Figure 2E, right panel**).

In contrast to transcriptomic and metabolomic profiles, markers for oxidative stress did not exhibit genotypic effects. We thus analyzed the data by hierarchical clustering-based, pairwise correlation (**Figure 2F**). This clustering approach uncovered several subgroups of enzymes and enzymes within each subgroup that were highly correlated (> 0.8), indicating these enzymes were involved in the same biological pathways. For instance, both chlorophyll *a* and chlorophyll *b* are the product of chlorophyll biosynthetic pathways (**Figure 2F**, bottom right). We used the four largest clusters from this analysis to assign representative markers and then plot their changes over time for each accession. Interestingly, each of these markers displayed a similar profile between genotypes over time in response to treatment (**Figure 2G**). As many of these molecules are indicative of perceived stress, their similar profiles suggest that all accessions are perceiving stress at the metabolic level.

Overall, we identified strong genotypic effects on both visible and molecular phenotypes. As expected for this species, most accessions displayed a robust resilience to combined heat and drought stress, whereas others (e.g., accession "A") performed poorly in this environment. Thus, this panel

presents an ideal comparative framework to identify the components responsible for drought and heat tolerance in sorghum under field conditions.

## Various covalent RNA modification (RCM) types are associated with similar effects on sorghum transcripts across time points and conditions

As changes in RNA covalent modification (RCM) levels have been associated with plant stress responses (**CITES**, including Palos et al., 2024), we next examined changes in the epitranscriptome of our panel. A subset of RCMs interfere with RNA-seq library preparation (specifically the reverse transcription step), leaving patterns of misincorporated nucleotides that can be assigned to specific modification classes bioinformatically (CITES). We processed our RNA-seq data through two independent RCM predictors (HAMR and Modtect) and took the intersection of called sites, resulting in 20,096 high-confidence RCMs associated with 3,715 transcripts (**Supplementary Table 4 and 5**). These RCMs are mainly composed of six modification classes based on the resolution of the RCM caller, including 3-methylcytidine ($m^3C$, *N = 1,688* transcripts), N1-Methyladenosine / 1-methylinosine / 2-methylthio-N6-isopentenyladenosine ($m^1A$ / $m^1I$/$ms^2i^6A$, *N = 2,324* transcripts), N1-methylguanosine ($m^1G$. *N = 1,250* transcripts), Pseudouridine (Y, *N = 1,435* transcripts), D (Dihydrouridine, *N* = 1,101 transcripts), and N6-isopentenyl adenosine or N6-Threonylcarbamoyladenosine ($i^6A$/$t^6A$, *N* = 1,236 transcripts).

In order to associate epitranscriptomic changes with other molecular and physiological traits, we developed three metrics to explain variation in the modification status for each transcript across our data: **1**) Presence or absence of a modification (PAM) within a transcript, scored as 1 (present/modified) or 0 (absent/not modified), which simply denotes if a transcript was modified or not in a particular context, **2**) RCM density, which describes the actual number of modifications within a transcript within a dataset, and **3**) RCM frequency, which describes the ratio of total modified reads for a transcript relative to the total mapped reads for that transcript (**Figure S4A, Supplementary Tables 6 and 7; see Methods**). As with the other molecular markers, genotype, not treatment, explained much of the variation in RCM PAM and density in our samples regardless of RCM class (**PC1: 30.78%; Figure S5A**). When the density of each RCM class was examined individually, separation based on treatment was more apparent, but only for the “A” accession (**Figures S5A and S5B**). By examining all RCMs per transcript, regardless of class (**Supplementary Tables 6 and 7**), a stronger genotypic separation was observed for PAM, density, and frequency, again predominantly driven by the “A” accession (**Figure S6A**). An examination of average density and frequency changes over time revealed that accessions “A” and “F” displayed the most variation, and predominantly under drought conditions (**Figure S6B**). To further characterize the RCM variation driven by genotype and treatment, we

performed pairwise comparisons of RCM features under control and drought using the numbers of modified transcripts, mean modification density, and mean modification frequency for each accession (**Figures S7A-D**). While accession "A" had a higher number of modified transcripts when combining treatments (**Figure S7A**), both accessions "A" and "F" displayed a reduction in PAM (**Figure S7B**), frequency (**Figure S7C**), and density (**Figure S7D**) under drought conditions relative to control. As accessions "A" and "F" are more closely related to each other than to the other accessions (**Figure 1A**), these data may point to a shared epitranscriptomic response to drought stress that is in close concordance with transcriptomic and metabolomic responses.

To better understand the biological implications of a transcript being modified, we selected a core set of 382 transcripts that had a non-zero PAM across all time points, treatments, and genotypes. A hierarchical clustering split these 382 transcripts into three groups based on RNA abundance changes (Z-score normalized, **Figure 3A, upper panel, Supplementary Table 8**). Gene Ontology (GO) enrichment analysis revealed an overrepresentation of stress-response-related terms in the Group I genes, translation and metabolic processes for Group II, and photosynthetic processes for Group III (**Figure 3B,** top 20 FDR GO terms shown). Within these groups, we highlighted time points and accessions where these modified transcripts were most or least abundant (**C1-C6**, **Figure 3A**). Interestingly, samples associated with accession "A" showed the highest abundance for stress response transcripts (**C2; Figure 3A**) and the least abundance of transcripts associated with photosynthesis (**C6; Figure 3A**). An examination of the relationship between RNA and RCM frequency for the transcripts found within each of these groups revealed an inverse relationship for most clusters (**Figure 3C**). In contrast, RCM density is largely correlated with RNA abundance, with the exception of photosynthesis-related transcripts within Cluster 6 associated with accession "A" (**Figure 3D**). Directly comparing the scaled transcript abundance vs RCM frequency and density values revealed that transcripts with high RCM density and low RCM frequency are associated with high transcript abundance (**Figure 3E**). This result is consistent with the global transcript-wise correlation between transcript abundance and RCM status (**Figure S4C and S4D**). We also observed this pattern in transcripts that are only modified in a single genotype (genotype-specific RCMs; **Figure S6E**). Despite these genotypic-specific modified transcripts being transcribed across all genotypes, their RNA abundance was lower in genotypes that were not modified compared with genotypes in which the transcripts were modified (**Figure S6F**), indicating RCM density is positively associated with the transcript abundance. In sum, these data suggest that RCMs are enriched in a transcript and genotype-specific manner that may influence how each accession responds to its environment.

## Integrated co-expression network reveals two modules with strong RCM and trait associations

With each of these molecular and physiological traits examined, we next took a comparative systems-level approach to connect these traits to their molecular pathways and putative regulators. To do so, we constructed a co-expression network starting with the top 10% most variable transcripts, as measured by median absolute deviation (MAD) score. We then recursively explored parameter space, increasing the number of included transcripts until we achieved an optimized network with the fewest number of unassigned transcripts and the highest scale-free topology model (**Supplemental Table 9**). The resulting network contained 22,314 transcripts classified into 44 modules with a low portion of unclassified transcripts (***N = 2,646***) and high scale-free model correlation ($R^2$ = 0.91, slope = -1.93, **Supplemental Table 10**). Using this framework, we incorporated phenotype data into modules using trait-module correlations. Briefly, the first principal component value (eigenvalue) per module was calculated using their transcript abundance PCA. Leaf level physiological measurements (LiCor-6800), oxidative stress enzyme activities, and metabolite abundances were then correlated with the eigenvalues of each module (**Supplemental Tables 11, 12, and 13**), leading to 13 modules with significant trait-module correlations ($P < 0.05$, **Figure 4A, left panel**). Several modules (e.g., “lightyellow”, “salmon”, “darkturquoise”, and “darkmagenta”) were strongly correlated with multiple traits. In addition to correlating traits to the eigenvalue of a module, pairwise correlations were performed between traits and each gene within a module (referred to as gene significance, **GS**), to uncover transcripts with a strong association with traits ($P < 0.05$).

To gain insights into the biological processes associated with each of these modules, GO and KEGG pathway enrichment were performed (modules with ***FDR* < 0.05** are displayed, **Figure 4A, middle panel**). This approach revealed modules with biological processes of interest, such as stress response, translation, metabolic pathways, and RNA modification terms. In addition, we performed a motif enrichment of transcription factors (TFs) for the 13 modules using the Arabidopsis DAP-seq database (**O’Malley et al., 2016,** *P-adj* < 0.05). We then determined which of the enriched motifs corresponded to TFs found within the corresponding module, suggesting those TFs may target co-expressed genes in our network (**Figure 4A, right panel**). Lastly, we overlaid RCM features onto each module by calculating the portion of modified transcripts with high GS value, and proportions of modified transcripts/unmodified transcripts associated with the enriched GO/KEGG terms in a module (**Figure 4B**). These data uncovered several modules with strong associations between each of our molecular and physiological datasets that are associated with stress responses or photosynthesis. Of these, the dark magenta module stands out for its strong positive association with leaf-level physiological parameters and enrichment of stress-response GO terms, whereas the salmon module

shows a negative correlation with leaf-level physiology and RCM status but a positive correlation with photosynthetic terms.

Based on the information above, we further characterized co-expression modules to determine if any exhibit distinct patterns that can be attributed to genotype, condition, or time point. To do so, we plotted the eigenvalue of each accession under each treatment per module to select modules with strong genotypic and treatment effects (or time, treatment). Using this approach, we identified two modules with strong genotypic effects (**Figure 4C,** "darkmagenta" and "salmon" modules, **Figure SX**) where accession "A" performed quite differently at weeks 3-7 relative to the other accessions. Interestingly, of the 382 constantly modified transcripts (**Figure 3A**), a significant number of the photosynthesis associated transcripts (Group III) appear in the Salmon and Darkmagenta modules (**Figure 4D**), further reinforcing a connection between RCM status and leaf-level responses to the environment. In sum, this network-based approach overlaid multiple lines of evidence to uncover a subset of transcripts whose abundance is strongly associated with physiological and metabolic traits. In addition, this approach uncovered putative transcriptional regulators, both TFs and RCMs, that might lead to the accession-specific stress resilience we observe in this panel.

## Transcripts associated with drought resilience and phenotypic variation are regulated by SbCDF3 and RCMs

To further explore the sets of transcripts that have the strongest trait-associations, we initially focused on identifying TFs within the "Darkmagenta" and "Salmon" modules that likely exert the largest effect on transcription within their respective modules (hub TFs). Using the Analysis of Motif Enrichment (AME; **CITE**), we found the binding motifs of 10 TFs that were significantly enriched in the 2 Kbp upstream promoter regions of downstream genes in the "Darkmagenta" module (**Supplemental Table 14**). Among these putative hub TFs, one in particular, CDF3 (Arabidopsis DAP-seq database ID: C2C2dof_tnt.CDF3, TAIR ID: AT3G47500, sorghum ID: Sobic.002G421900), stood out based on AME ranking, numbers of edges, and number of connected target genes within the module. Importantly, CDF3 has previously been reported as a transcriptional activator that enhances drought and oxidative stress tolerance in Arabidopsis and tomato[55,56]. CDF3 is single copy in sorghum, and importantly, appears to be connected to similar classes of stress response genes based on its location within the "Darkmagenta" module. Out of the 1093 genes within the "Darkmagenta" module, 681 were predicted to be bound by CDF3 (**Figure 5A**). Further, we carefully examined the binding motifs of those 681 putative CDF3-bound and co-expressed genes. The motif contains the core binding motif (**5'-T/AAAAG-3'**; **Figure 5B**) based on previous molecular studies [57,58], suggesting conservation in both

CDF3 as a regulator of environmental responses, and in its preferred binding motif. Interestingly, an examination of CDF3's transcript abundance pattern across the 96 sorghum samples revealed a dramatic decrease in the "A" accession in weeks 3-7 under both control and drought conditions (**Figure 4A**). Further, we examined the correlation between CDF3 and each CDF3 target (***N = 410,*** selected by random extraction) in the "Darkmagenta" module, a similar number of non-CDF3 targeted transcripts within the module, as well as randomly selected transcripts from outside of the "Darkmagenta" module (***N = 410***). A significantly higher correlation was observed between CDF3 and within-module targets compared with the other two groups (***P*** = 0.0018 and ***P*** < 2.2e-16, **Figure 5C**). Lastly, we compared the $\log_2$(fold-change) score (Control/Drought) of CDF3 targets and non-CDF3 targets across all accessions and time points. This comparison revealed a significantly higher $\log_2$FC (***P*** = 0.001) of CDF3-targets (**Figure 5D)**, suggesting these target genes are induced under drought conditions in a CDF3-dependent manner.

The above CDF3-target data, along with the traits correlated with the "Darkmagenta" module, suggest that CDF3 is acting as a potential central regulator connecting the drought response in these accessions to downstream photosynthetic traits. To further investigate this hypothesis, we more closely examined the relationship between CDF3 target genes and the physiological and molecular traits (e.g., RCMs and metabolites) associated with the "Darkmagenta" module. To do so, we selected the transcripts (***N* = XXX**) with the highest trait-RNA abundance correlation (cut-off: |r| > 0.7) and then integrated GO terms, TF-binding, and protein-protein interaction (**Figure 5E; Supplementary Table 15**). We further filtered these CDF-interacting transcripts based on RCM status (boxes = modified, circles = unmodified; **Figure 5E**), resulting in a sub-network of 230 out of the 1093 transcripts in the original "Darkmagenta" module. Of the 230 modified transcripts found within this sub-network, XX displayed a strong correlation with CDF3 and was predicted to be a CDF3 target based on AME (blue dashed lines, **Figure 5E**). Of these, ten genes were strongly correlated with CO2 net assimilation (Blue boxes, **Figure 5E**), eleven with flavonoid content (grey boxes, **Figure 5E**), four with leaf temperature (orange boxes, **Figure 5E**), and six with total antioxidant capacity (peach boxes, **Figure 5E**). The remaining ~200 genes were associated with GO terms for abiotic stress response, RNA modification (predominantly spliceosomal RNAs), and metabolic processes. In particular, some genes with strong trait-transcript correlation coincide with several candidate genes identified through QTL mapping from previous studies (**marked with red pentagrams, Figure 5E, Supplementary Table 15 in the "QTL" column**), these genes could be ideal candidates for downstream functional validation.

Interestingly, ten of these CDF3 targets were not only strongly trait-correlated but were also part of the previously identified constantly modified transcripts (**Figures 3A** and **5F**). In addition to being strongly correlated with CDF3 transcript abundance (R > 0.6, P < 0.05; **Figure 5F**) and RCM status

(**Figure 5G,** P < 0.05 with PCC marked), these ten transcripts were also positively correlated with $CO_2$ net assimilation, and negatively associated with flavonoid content and total antioxidant activity (**Figure 5E**). Indeed, the poorly performing accession "A" had reduced transcript abundance for each of these genes (and CDF3) and showed reduced CO2 net assimilation in the later time points (Weeks 3-7; **Figure 5E**). Some of these trait and CDF3-associated genes have been functionally characterized in previous studies, including the orthologs of *SbPPC2, SbCP5,* and *SbPSAL* in Arabidopsis [59–61]. In agreement with the data we describe here, these genes are integral to photosynthesis-related processes and confer stress tolerance. In sum, we uncovered a suite of photosynthesis and stress-associated genes that are regulated at the transcriptional level by CDF3 and at the post-transcriptional level by a suite of RCMs in sorghum.

## Uncovering dihydrouridine links to photosynthesis by network analysis

Using a similar approach, we examined transcripts within the salmon module, which was also associated with photosynthesis and metabolite-related genes. We first prioritized hub TFs based on AME motif enrichment, identifying two potential upstream TF regulators, ARID-HMG DNA-binding protein 15 (**HMGDB15**) and basic pentacysteine 6 (**BPC6**, **Supplementary Table 14**). Both of these TFs were reported to be involved in abiotic stress response in Arabidopsis (AtHMGDB15: AT5G21430; AtBPC6: AT5G42520, **Antosch et al., 2012**; **Zhang et al., 2023; Mallik et al., 2020**)[62–64]. When examining genes with strong GS and module membership, we identified a transcript that encodes dihydrouridine synthase (***SbDUS2*: Sobic.006G158100.1**), the enzyme responsible for synthesizing the RCM dihydrouridine. We confirmed this annotation in sorghum using reciprocal best BLASTn and subsequent protein alignment to the single copy Arabidopsis DUS2 as well as searching against the conserved domain database (CDD; **ID: AT3G49640, CDD number: cd02801-**FMN-binding domain). DUS2 is an ancient enzyme shared between eukaryotic, prokaryotic, and some archaeal lineages, and the RCM dihydrouridine has been predominantly reported on housekeeping RNAs (e.g., tRNAs and rRNAs). Intrigued by its inclusion in the salmon module, we examined DUS2 transcript abundance over time and across treatments in each of the accessions (**Figure 6A**). Accessions "A" and "F" showed decreased DUS2 transcript abundance over time, with "A" displaying the lowest values (**Figure 6A**). In addition, the presence of the dihydrouridine RCM itself showed a strong genotype effect, with the later time points (weeks 3-7) separating from all other accessions and time points in a PCA (**Figure 6B**). Interestingly, the "A" and "F" accessions showed an elevated number (*N* = 119) of homozygous non-reference alleles within the *SbDus2* coding sequencing relative to the other four accessions (**Figure 6C**), in agreement with their phylogenetic relationship (**Figure 1A**). Of these, eight SNPs were

predicted to impact splicing and potentially affect transcript abundance (**VEP annotation results, see methods**). The true impact of these SNPs is unclear, but altogether, the Dus2 expression and pattern of D deposition suggest that dihydrouridine is used differentially across these six accessions.

To further explore the relationship between dihydrouridine and the traits associated with the salmon module, we developed a sub-network on phenotypes and their potential interaction with the downstream transcript. We built up the sub-network (***N* = 77**) of transcripts with high GS and module membership and then overlaid RCM status, TF-binding, trait association, and GO term data (**Figure 6D**). Within this module, there was a set of nine transcripts that consistently showed a non-zero dihydrouridine status (blue boxes, **Figure 6D**), with one of them, *SbPpdk1*, regulated by the TF HMGD15. Among the nine dihydrouridine-targeted transcripts, their transcript abundance was positively correlated with the *Dus2* transcription (PCC > 0.38, **Figure 6E left indicators**). We also noticed that two out of nine DUS2-co-expressed genes are highly correlated with phenotypic traits (**Figure 6F,** Pyruvate phosphate dikinase *(PPDK)**:*** **Sobic.009G132900 and** Chloro-respiratory reduction L (CRRL)***:*** **Sobic.008G102000**), and these two genes are found within QTL intervals that were significantly associated with leaf carbon assimilation rate derived from a GWAS of sorghum [65,66](p = 2.9e-4 and p = 4.5e-5). The Arabidopsis orthologs of these two genes (**ID: *AtPPDK:* AT4G15530 and *AtCRRL:* AT5G21430**) have been functionally validated. *AtPPDK* is known to enhance photosynthesis while improving stress tolerance and is also important for seed development in Arabidopsis [67,68]. *AtCRRL* is essential for NAD(P)H dehydrogenase activity (NDH) that serves as the machinery of stress resistance[69,70]. In maize, the C4 grass ortholog of sorghum PPDK1 was identified using Blastp (***Zm00001eb287770.1p:*** *ZmPPDK-C4***,** identities: 92.87% and coverage: 95%). Maize *ZmPPDK-C4* loss-of-function mutants have also been characterized as displaying reduced $CO_2$ net assimilation capacity and down-regulation of other C4-related genes in previous study[71].

To determine if the relationship between SbPPDK1 transcript abundance and dihydrouridine RCM status held true in a broader setting, we reprocessed a publicly available transcriptomic tissue atlas for sorghum (**McCormick et al., 2017**). In this tissue atlas, which comprises 38 tissue types and developmental stages (148 samples), *SbPPDK* was modified by dihydrouridine in 18 tissues, including seedlings, lower, middle, and upper leaves, as well as head regions (**Figure 6G**). Notably, *SbPPDK* exhibited higher transcript abundance in upper leaves, recapitulating what we observed in the field. Furthermore, we analyzed the modification status of dihydrouridine in *SbPPDK* transcripts relative to their abundance (**Figure 6H**). The results revealed a strong correlation between modification density (r = 0.94) and frequency (r = -0.58). These findings suggest that PPDK1 RNAs are targeted for modification by dihydrouridine, and the presence of this modification is associated with increased transcript abundance and CO2 net assimilation levels.

## ***DUS2*** **recruits dihydrouridine affecting** ***PPDK*** **transcript abundance and $CO_2$ net assimilation in** ***Arabidopsis***

The data presented so far suggests a correlation between dihydrouridine levels, PPDK1, and photosynthesis in response to heat stress. To determine if DUS2 has a direct role in influencing plant performance under stress, we focused on DUS2 loss-of-function mutants in the genetic model system, *Arabidopsis thaliana*. The uncharacterized Arabidopsis DUS2 ortholog (**AT3G49640**) is 79.81% identical to the SbDUS2 protein and also contains the dihydrouridine synthase domain (**Figure 7A**). We obtained three T-DNA insertion lines for DUS2 (SALK_079129C, CS857388, and SALK_023006, https://abrc.osu.edu/), and based on reduced dihydrouridine levels measured by mass-spec (**Supplemental Figure X**), we selected one homozygous T-DNA mutant line (**CS857388**) harboring an insertion within the coding region of DUS2 for further phenotypic analysis (Exon 3; **Figure 7A**).

*AtDUS2* transcript abundance peaks in germinating seeds, leaves, and flowers (**Supplemental FIGURE X**). In agreement with its expression, we observed reduced germination (**Figure 7B**; $P < 0.05$, $N > 50$) and growth rates (**Figure 7C**), as well as shorter stature and increased time to flower, in dus2 mutants (**Figure 7D**) relative to wild type controls (Col-0). We also examined the ability of the dus2 mutant to withstand heat stress by growing ~2-week seedlings at 37°C (heat) and 22°C (control) for 12 days. While both wild-type Arabidopsis and dus2 mutants exhibited reduced growth rates, the survival rate at the end of the 12 days of treatment was significantly reduced (**Figure 7C**, right and inset, $\boldsymbol{P < 0.05}$). As SbDUS2 was positively associated with $CO_2$ net assimilation under heat conditions in our field data, we examined the leaf-level responses to heat stress in the Arabidopsis dus2 mutants. We observed slightly reduced (**$p = 0.12$** and **$p = 0.33$, $N = 10$; Figure 7E**) $CO_2$ net assimilation and stomatal conductance in *dus2* mutant lines under control conditions (**22°C**). However, heat-induced (**time: 24h temperature: 37°C**) changes in photosynthetic parameters were significantly reduced in the Atdus2 background relative to Col-0 controls (**$p = 0.00016$** and **$p = 0.0024$, $N = 10$, Figure 7E**), suggesting that DUS2 is critical for a proper heat stress response in Arabidopsis. Interestingly, although Arabidopsis does not utilize *PPDK1* to perform C4 photosynthesis, we still observe a positive correlation between DUS2 and *PPDK1* across the Arabidopsis tissue atlas (**Supplemental Figure X**), highlighting a deeply conserved potentially regulatory connection between these two genes across flowering plants. In sum, we uncovered, for the first time, a role for the RCM, dihydrouridine, in modulating plant performance under abiotic stress conditions in Arabidopsis and sorghum.

## Discussion

Combined heat and drought stress typically induce complex phenotypic and molecular responses that often lead to reduced biomass, seed set, and photosynthesis. Sorghum is generally well adapted to hot and dry environments, although there is variation in how sorghum responds to these stressors. Given the complex nature of plant responses to water deprivation, many stress-responsive genes have been identified in sorghum and other systems[8–10] and these cohorts of genes often change depending on the developmental stage[72]. Thus, we still lack an understanding of which of these regulatory mechanisms and their targets contribute to improved agricultural traits in a field setting, and therefore could serve as breeding markers or targets for manipulation.

We used a representative subset of the genetic and phenotypic diversity found within sorghum to identify transcriptional and post-transcriptional regulatory mechanisms associated with improved responses to hot and dry environments. Our panel of six distantly related sorghum accessions revealed distinct physiological and molecular responses to hot and arid conditions, with some accessions exhibiting minimal reductions in biomass and photosynthetic efficiency, whereas others (e.g., “A”) performed quite poorly. Leaf-level physiological traits tracked closely with metabolomic, transcriptomic, and epitranscriptomic features across the six accessions, allowing us to make inferences as to which molecular mechanisms were contributing to phenotypic success. Using a comparative systems-level approach, we identified cohorts of transcripts whose expression patterns were most closely linked to molecular and physiological responses to drought stress (**Figure 4**). The benefit of the approach applied here is that systemically induced stress-response genes are ignored in favor of those that track with phenotypic and genetic variation. Two co-expression modules in particular exhibited variation due to both treatment and genotype, allowing us to identify regulators that were associated with stress tolerance. At the transcriptional level, we identified three stress-induced TFs that served as key positive regulators in their respective modules (*SbCDF3;* **Figure 5***, SbHMGDB15*, and *SbBPC6;* **Figure 6**). As expected based on their expression in leaf tissue, these TFs, and their AME-predicted targets, such as *SbPPDK, SbCRRL, SbPPC2, SbCP5,* and *SbPSAL,* were highly correlated with photosynthetic traits. Several of these genes, such as CDF3 and PPDK1, have been characterized in other systems [55,56,62–64], and have also been shown to be part of a large suite of genes induced under drought stress across development in sorghum [72]. In addition, association studies in sorghum under abiotic stresses uncovered a strong association between $CO_2$ assimilation and QTLs harboring these genes[65]. As these genes in particular show strong trait-linked transcriptional variation between sorghum accessions and are conserved across flowering plants (*Arabidopsis* to sorghum), they are ideal candidates for manipulation or selective breeding in sorghum or other grasses.

While RCMs have been shown to influence or be associated with plant stress responses **(**[19,33,73–77]. we were still surprised to see that an enzyme associated with the reduction of uridine to 5'-6'-dihydrouridine, DUS2, was clustered into one of the two photosynthesis and stress-response modules. The dihydrouridine RCM is most commonly associated with tRNAs, and indeed the DUS protein family is conserved across the tree of life, including eukaryotes, prokaryotes, and some archaea. Due to advances in detection capabilities, dihydrouridine has recently been described on mRNAs in humans and yeast, where it has been reported to influence splicing, translation, and fundamental biological processes such as meiosis. However, no reports exist on dihydrouridine in plants, and thus the findings presented here offer insights into a deeply conserved post-transcriptional regulatory mechanism.

Interestingly, not only was DUS2 strongly co-expressed with a group of transcripts related to photosynthesis, the presence of the RCM itself was also linked to these transcripts. One gene in particular, PPDK1, is integral to both photosynthesis and gluconeogenesis in plants. Both PPDK1's transcript and dihydrouridine abundance were positively associated with photosynthetic parameters under drought stress. As we did not observe any differences in splicing for PPDK1 in accessions with high dihydrouridine levels, we hypothesize that this RCM may be positively influencing translation over splicing. In sum, our transcriptomic and epitranscriptomic approach uncovered both the transcriptional and post-transcriptional regulators of key enzymes associated with central metabolism and photosynthesis.

Given that this was the first report of dihydrouridine in plants - and in particular plant stress responses, we sought to develop a clearer picture of what this RCM might be doing by examining Arabidopsis lines with insertions in the orthologous DUS2 locus. These *dus2* mutant lines exhibited both decreased germination, slower growth, and reduced biomass at time of flowering. Each of these traits could be linked to overall reduced translation rates due to inadequate dihydrouridine on tRNAs. However, we also observed a heat-stress-dependent reduction in photosynthetic traits and reduced survival after treatment in Arabidopsis *dus2* mutants, suggesting a conserved role for dihydrouridine in mounting an effective stress response in plants. In line with this hypothesis, a time-series transcriptomic analysis of *dus2* mutants during heat stress revealed a number of differentially expressed transcripts, with photosynthetic and stress response GO terms being particularly enriched. Interestingly, despite the fact that Arabidopsis performs C3-photosynthesis, we still observe a positive correlation between DUS2 and PPDK1 in Arabidopsis, both in our data and in the *Arabidopsis* tissue atlas[78], suggesting these two transcripts are coregulated across the plant lineage. Indeed, DUS2 and the RCM dihydrouridine are positively associated with many nuclear-encoded genes associated with photosynthesis and metabolism in Arabidopsis and sorghum. While the mechanism by which DUS2 targets specific

transcripts is unclear, the deep conservation of many of these genes, across plants and eukaryotes in general, suggests that this modification is an important addition to the repertoire of critical mRNA RCMs.

In summary, our comparative systems-level approach has allowed us to better understand the genetic basis underlying heat and drought tolerance in sorghum. Using a small panel that represents the diversity within sorghum, alongside intensive phenotypic and molecular phenotypic, we were able to identify regulatory factors, such as CDF3, associated with improved biomass and photosynthetic parameters. Among the many differentially abundant transcripts, we were able to identify a subset that are strong candidates as breeding targets, similar to our past efforts in other agronomic species (Yu et al., 2024). Lastly, our integrative approach uncovered an unexpected role for the RCM dihydrouridine in plant stress responses. Indeed, this deeply conserved RCM influences heat-stress responses in both Arabidopsis and sorghum by acting on many of the same photosynthetic and metabolic mRNA transcripts. Thus, these data highlight the complex transcriptional and post-transcriptional regulatory landscape influencing plant stress responses, but offer promising insights as to how they might be used to select for more stress-resilient plants in the future.

## Data availability

Transcriptome sequencing reads have been deposited in the National Center for Biotechnology Information (NCBI) BioProject database under the accession number: **PRJNA1119650.** Computational pipelines used for data analysis were deposited under the Github page (https://github.com/Leon-Yu0320/Sorghum_Range28).

## Acknowledgments

We would like to thank the members of the epitranscriptomic working group within the Nelson, Pauli, Schroeder, and Gregory labs for their insightful comments. This work was supported by NSF IOS PGRP 2023310 (to ADLN, EHL, DP, and BDG), NSF IOS PGRP 1849708 (to BDG), and NSF MCB 2120131 (to ADLN and DP).

## Author contributions

LY, GM, AEA, EHL, BDG, ADLN, and DP developed the experiment. GM, SC, BR, and DP designed the field experiment and collected field-related data. ACND and EKB extracted RNA and performed other sorghum-related molecular experiments. AE provided accessions from the SAP for this study. LY, APS, XZ, and EKB worked with the Arabidopsis DUS2 mutants. CDG and AEA worked on the microbiome-related experiments. GM worked on the metabolomics-related experiments. LY, RH, and DS helped and provided expertise in acquiring physiological measurements for the Arabidopsis DUS2 mutants. AlS performed mass-spec experiments on the DUS2 mutants. LY analyzed all RNA-seq and developed co-expression networks and data integration. LY, KP, and SJS analyzed epitranscriptomic data. LY, ADLN, and DP wrote the manuscript.

## References


1. Morris, G. P. *et al.* Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 453–458 (2013).
2. Abreha, K. B. *et al.* Sorghum in dryland: morphological, physiological, and molecular responses of sorghum under drought stress. *Planta* **255**, 20 (2021).
3. Sehgal, A. *et al.* Drought or/and heat-stress effects on seed filling in food crops: Impacts on functional biochemistry, seed yields, and nutritional quality. *Front. Plant Sci.* **9**, 1705 (2018).
4. Alizadeh, M. R. *et al.* A century of observations reveals increasing likelihood of continental-scale compound dry-hot extremes. *Sci. Adv.* **6**, eaaz4571 (2020).
5. Hasanuzzaman, M., Nahar, K., Gill, S. S. & Fujita, M. Drought stress responses in plants, oxidative stress, and antioxidant defense. *Climate change and plant abiotic stress tolerance* 209–250 (2013).
6. Zargar, S. M. *et al.* Impact of drought on photosynthesis: Molecular perspective. *Plant Gene* **11**, 154–159 (2017).
7. Fàbregas, N. & Fernie, A. R. The metabolic response to drought. *J. Exp. Bot.* **70**, 1077–1085 (2019).
8. Zhu, J.-K. Abiotic stress signaling and responses in plants. *Cell* **167**, 313–324 (2016).
9. Kim, J.-S., Kidokoro, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. Regulatory networks in plant responses to drought and cold stress. *Plant Physiol.* **195**, 170–189 (2024).
10. Mahmood, T. *et al.* Insights into drought stress signaling in plants and the molecular genetic basis of cotton drought tolerance. *Cells* **9**, 105 (2019).
11. Sato, H., Mizoi, J., Shinozaki, K. & Yamaguchi-Shinozaki, K. Complex plant responses to drought and heat stress under climate change. *Plant J.* **117**, 1873–1892 (2024).
12. Joshi, R. *et al.* Transcription factors and plants response to drought stress: Current understanding and future directions. *Front. Plant Sci.* **7**, 1029 (2016).
13. Yu, L. *et al.* Regulation of a single inositol 1-phosphate synthase homeologue by HSFA6B contributes to fibre yield maintenance under drought conditions in upland cotton. *Plant Biotechnol. J.* **22**, 2756–2772 (2024).

14. Bang, S. W. *et al.* Overexpression of OsTF1L, a rice HD-Zip transcription factor, promotes lignin biosynthesis and stomatal closure that improves drought tolerance. *Plant Biotechnol. J.* **17**, 118–131 (2019).
15. Prall, W., Ganguly, D. R. & Gregory, B. D. The covalent nucleotide modifications within plant mRNAs: What we know, how we find them, and what should be done in the future. *Plant Cell* **35**, 1801–1816 (2023).
16. Gilbert, W. V. & Nachtergaele, S. MRNA regulation by RNA modifications. *Annu. Rev. Biochem.* **92**, 175–198 (2023).
17. Lewis, C. J. T., Pan, T. & Kalsotra, A. RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* **18**, 202–210 (2017).
18. Arribas-Hernández, L. *et al.* An m6A-YTH module controls developmental timing and morphogenesis in Arabidopsis. *Plant Cell* **30**, 952–967 (2018).
19. Tang, Y. *et al.* OsNSUN2-mediated 5-methylcytosine mRNA modification enhances rice adaptation to high temperature. *Dev. Cell* **53**, 272-286.e7 (2020).
20. Sun, H., Li, K., Liu, C. & Yi, C. Regulation and functions of non-m6A mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **24**, 714–731 (2023).
21. Zhang, Y., Lu, L. & Li, X. Detection technologies for RNA modifications. *Exp. Mol. Med.* **54**, 1601–1616 (2022).
22. Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.* **37**, e12 (2009).
23. Vandivier, L. E. *et al.* Chemical modifications Mark alternatively spliced and uncapped messenger RNAs in Arabidopsis. *Plant Cell* **27**, 3024–3037 (2015).
24. Tan, K.-T., Ding, L.-W., Wu, C.-S., Tenen, D. G. & Yang, H. Repurposing RNA sequencing for discovery of RNA modifications in clinical cohorts. *Sci. Adv.* **7**, eabd2605 (2021).
25. Yu, L. *et al.* Development of a mobile, high-throughput, and low-cost image-based plant growth phenotyping system. *Plant Physiol.* (2024) doi:10.1093/plphys/kiae237.
26. Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497 (1962).
27. Brazel, A. J., Manoj, N. S., Turck, F. & Ó'Maoiléidigh, D. S. Measuring CO2 assimilation of Arabidopsis thaliana whole plants and seedlings. *Plant Sci.* **350**, 112295 (2025).
28. Ryvkin, P. *et al.* HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684–1692 (2013).
29. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
30. McCormick, R. F. *et al.* The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* **93**, 338–354 (2018).
31. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
32. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
33. Palos, K., Nelson Dittrich, A. C., Lyons, E. H., Gregory, B. D. & Nelson, A. D. L. Comparative analyses suggest a link between mRNA splicing, stability, and RNA covalent modifications in flowering plants. *BMC Plant Biol.* **24**, 768 (2024).
34. Dobin, A. & Gingeras, T. R. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11.14.1-11.14.19 (2015).
35. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39-49 (2015).
36. Boatwright, J. L. *et al.* Sorghum Association Panel whole-genome sequencing establishes cornerstone resource for dissecting genomic diversity. *Plant J.* **111**, 888–904 (2022).

37. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).
38. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
39. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *BIOINFORMATICS APPLICATIONS NOTE* **25**, (2009).
40. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
42. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
43. Liao, Y., Smyth, G. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).
44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
45. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
46. Zheng, Y. *et al.* ITAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670 (2016).
47. O'Malley, R. C. *et al.* Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016).
48. von Mering, C. *et al.* STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).
49. Team, R. C. RA language and environment for statistical computing, R Foundation for Statistical. *Computing* (2020).
50. Wickham, H. Ggplot2: Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).
51. Kolde, R. & Kolde, M. R. *R package* (2015).
52. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
53. Obata, T. *et al.* Metabolite profiles of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiol.* **169**, 2665–2683 (2015).
54. Das, K. & Roychoudhury, A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* **2**, (2014).
55. Corrales, A.-R. *et al.* Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in Arabidopsis. *Plant Cell Environ.* **40**, 748–764 (2017).
56. Corrales, A.-R. *et al.* Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. *J. Exp. Bot.* **65**, 995–1012 (2014).
57. Yanagisawa, S. & Schmidt, R. J. Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J.* **17**, 209–214 (1999).
58. Noguero, M., Atif, R. M., Ochatt, S. & Thompson, R. D. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci.* **209**, 32–45 (2013).
59. Feria, A. B. *et al.* Phosphoenolpyruvate carboxylase (PEPC) and PEPC-kinase (PEPC-k) isoenzymes in Arabidopsis thaliana: role in control and abiotic stress conditions. *Planta* **244**, 901–913 (2016).
60. Leonetti, P. *et al.* Endogenous activated small interfering RNAs in virus-infected Brassicaceae crops show a common host gene-silencing pattern affecting photosynthesis and stress response. *New Phytol.* **229**, 1650–1664 (2021).

61. Ascencio-Ibáñez, J. T. *et al.* Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol.* **148**, 436–454 (2008).
62. Antosch, M., Mortensen, S. A. & Grasser, K. D. Plant proteins containing high mobility group box DNA-binding domains modulate different nuclear processes. *Plant Physiol.* **159**, 875–883 (2012).
63. Zhang, Z., Zhang, T. & Ma, L. Analysis of basic pentacysteine6 transcription factor involved in abiotic stress response in Arabidopsis thaliana. *Front. Genet.* **14**, 1097381 (2023).
64. Mallik, R. *et al.* Identification of genome-wide targets and DNA recognition sequence of the Arabidopsis HMG-box protein AtHMGB15 during cold stress response. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194644 (2020).
65. Mace, E. *et al.* The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Züchter Genet. Breed. Res.* **132**, 751–766 (2019).
66. Ortiz, D., Hu, J. & Salas Fernandez, M. G. Genetic architecture of photosynthesis in Sorghum bicolor under non-stress and cold stress conditions. *J. Exp. Bot.* **68**, 4545–4557 (2017).
67. Eastmond, P. J. *et al.* Arabidopsis uses two gluconeogenic gateways for organic acids to fuel seedling establishment. *Nat. Commun.* **6**, 6659 (2015).
68. Yadav, S. & Mishra, A. Ectopic expression of C4 photosynthetic pathway genes improves carbon assimilation and alleviate stress tolerance for future climate change. *Physiol. Mol. Biol. Plants* **26**, 195–209 (2020).
69. Yamamoto, H., Peng, L., Fukao, Y. & Shikanai, T. An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in Arabidopsis. *Plant Cell* **23**, 1480–1493 (2011).
70. Wang, P. *et al.* Chloroplastic NAD(P)H dehydrogenase in tobacco leaves functions in alleviation of oxidative damage caused by temperature stress. *Plant Physiol.* **141**, 465–474 (2006).
71. Zhang, Y. *et al.* Characterization of maize leaf pyruvate orthophosphate dikinase using high throughput sequencing. *J. Integr. Plant Biol.* **60**, 670–690 (2018).
72. Varoquaux, N. *et al.* Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 27124–27132 (2019).
73. Anderson, S. J. *et al.* N6-methyladenosine inhibits local ribonucleolytic cleavage to stabilize mRNAs in Arabidopsis. *Cell Rep.* **25**, 1146-1157.e3 (2018).
74. Cheng, Q. *et al.* Coordination of m6A mRNA methylation and gene transcriptome in rice response to cadmium stress. *Rice (N. Y.)* **14**, 62 (2021).
75. Yang, W. *et al.* The N1-methyladenosine methylome of petunia mRNA. *Plant Physiol.* **183**, 1710–1724 (2020).
76. Dominissini, D. *et al.* The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
77. Luo, J.-H. *et al.* Natural variation in RNA m6A methylation and its relationship with translational status. *Plant Physiol.* **182**, 332–344 (2020).
78. Mergner, J. *et al.* Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature* **579**, 409–414 (2020).

# Figure Legends

## Main Figures

### Figure 1. Genomic and phenotypic response diversity of selected sorghum genotypes

**(A) Phylogenetic tree of 356 domesticated sorghum accessions:** Reconstruction of an unrooted phylogenetic tree was performed using 7,520 bi-allelic resulting SNPs after low-quality loci removal and LD-pruning. The six selected sorghum accessions were marked along with the geographical locations colored to represent African (green), Asian(purple), and North American (blue) areas. The "pentagram" logo highlights the six accessions in our studies. **(B, C) Morphological responses of sorghum accessions under drought:** Leaf and root tissues collected at the terminal stage of sampling in the field. **(D, E) Measurement of leaf and root characteristics:** The biomass, leaf weight, and root length(width) were measured under two watering conditions. **(F) Photosynthesis-related status of sorghum under drought:** Line graphs highlight the highest stressed condition in accession "A" represented by $CO_2$ concentration, Fv'/Fm', and WUEi under drought conditions over the morning and afternoon measurement sessions.

### Figure 2. Molecular response variation to drought among sorghum accessions

**(A) Transcriptome features of global transcripts:** A hierarchical clustering of 34,130 transcripts normalized from DESeq2 has been conducted to group transcriptomic features among samples derived from six accessions under two watering conditions over four-time points. The level of similarities of samples was displayed based on the heatmap. **(B) Signatures of the top 10% of most variables transcripts:** Two PCA plots highlight the strong genotypic effects among six sorghum accessions relative to time points and treatment effects. **(C) Numbers of DATs (Drought vs Control) among six sorghum accessions:** The line graph represents changes in DATs across developmental stages where the accession "C" with the most mild stress response. **(D) GO enrichment of DATs (Drought vs Control) of sorghum:** The enrichment levels of shared GO terms been over-represented in each gene set ($-\log_{10}(Padj) < 0.05$) among six sorghum genotypes were displayed by a heatmap to uncover the variation of up-regulated and down-regulated DATs among genotypes. **(E) Profiles of metabolites under two watering conditions:** The left PCA panel uncovered the unique metabolic features of genotype "A" and the right PCA loading plot highlights the chlorogenic acid and sucrose as the major contributors to PCA patterns. **(F) Clustering of oxidative enzyme activity entries:** Pearson correlation-based clustering was performed to group enzymes with similarities with four major subgroups with similar patterns ($|r| > 0.6$). (**G) Profiles of representative enzymes under drought:** One enzyme per group was selected to display the activities across time points among six genotypes.

### Figure 3. RCMs landscape of a core set transcripts

**(A) Transcript abundance of constantly modified transcripts:** A heatmap with clustering highlights the level of transcription among three groups of transcripts (Group I, Group II, and Group III) and six (C1-C6) subgroup samples. Most samples of accession A (except Week 1) were grouped. **(B) GO enrichment of modified transcripts per group:** Numbers of transcripts per term (Counts) and level of the enrichment ($-\log_{10}(Padj)$) were plotted to represent different sets of transcripts being modified among three groups. **(C, D) Relationship between transcript abundance and modification frequency:** The scaled (Z-score) inputs of transcription and modification were compared to show the situation where the higher transcript abundance was associated with lower modification frequency. **(E,**

**F) Relationship between transcript abundance and modification density:** The scaled (Z-score) inputs of transcription and modification were compared to show the situation where the higher transcript abundance was associated with higher modification density.

## Figure 4. Selection of core co-expression modules contributing to drought tolerance

**(A) Transcriptome features of modules:** The number of transcripts per module, correlation between traits and module (**blue and red scale**), levels of KEGG and GO enrichment levels (**purple and brown scale**), and numbers of TFs with enriched motif within modules were displayed to select modules with biological meanings related to drought response. **(B) RCM features of modules:** The ratio of modified transcripts with "module-trait" correlation (**green and yellow scale**) and the ratio of modified traits involved in enriched KEGG and GO terms **(dark blue and yellow scale)** were represented to show the status of modules containing transcripts been modified. **(C) Eigene value of selected modules:** The first principle score of each module was plotted to characterize the features of modules in terms of genotypes, conditions, and time points. **(D) Ratio of constantly modified transcripts within modules:** Portions of constantly modified transcripts over 382 transcripts (Figure 3C) highlight the "salmon" and "Darkmagenta" modules as major inputs of constantly modified transcripts that are involved in photosynthesis and metabolic pathways.

## Figure 5. Regulatory gene network involved in photosynthesis-related traits

**(A) Transcript abundance of CDF3:** Relative abundance of CDF3 among genotypes, time points, and treatments suggests the lower transcript abundance in accessions "A" relative to other genotypes. **(B)** Numbers of genes with enriched CDF binding motif in "darkmagenta" module and the consensus motif of CDF3 in Arabidopsis. **(C) Pairwise correlation of CDF3 and each module gene:** The correlation between CDF3 and each module was compared among three groups, including CDF3 targets, CDF3 non-targets, and randomly selected genes outside of the module (*N = 392*). These three sets of correlations were compared (t-test, $P < 0.05$). **(D) $log_2$FC score of CDF3 targeted (non-target) downstream genes:** The $log_2$FC score (Drought/Control) was calculated to compare the levels of response between CDF3 targets and non-CDF3 targets (t-test $P < 0.05$). **(E) Integration of regulatory network of selected genes in the "darkmagenta" module:** 230 genes involved in different categories were visualized using the Cytoscpae network integration, including genes associated with enriched GO terms (abiotic stress response, RNA modification, and metabolic process), as well as genes with high correlation ($|r| > 0.65$) with respective traits (listed above with legends). The solid 'grey' lines connected the co-expressed genes within the module, the dashed 'red' lines indicated protein-protein interactions (PPI) derived from the STRING database, and the 'blue' dash-arrowed lines highlighted targets bound by CDF3, as supported by Arabidopsis DAP-seq data. The red "pentagram" marked the genes overlapped with key QTLs from previous sorghum GWAS (**Ortiz et al 2017**) **(F) Trait-associated genes targeted by CDF3:** Transcript abundance of 10 selected genes targeted by CDF3 was displayed in a clustering heatmap, where the respective correlation with CDF3 was listed on the left, and the gene name and correlated traits on the right side (Upper panel). **(G) Correlation between transcript abundance and RCM frequency (density):** The sample-wise correlation between transcript abundance and RCM frequency (left side), transcript abundance and RCM density (right side) was plotted, indicating the close relationship between RCM status and transcript level of genes associated with traits.

## Figure 6. Potential regulations between Dihydrouridine and *SbPPDK*

**(A) Transcript abundance of DUS2:** Relative abundance of *SbDUS2* among genotypes, time points, and treatments suggests the lower transcript abundance in accessions "A" and "F" relative to other genotypes. **(B) Presence and absence of modification (PAM) status of dihydrouridine-modified transcripts:** The dihydrouridine-modified transcripts (PAM) were characterized using PCA (*N = 1101 transcripts*) with samples of accession "A" and "F" clustered away from the rest samples (PC1: 59.91%). **(C) Genomic variation within *SbDUS2* among six genotypes:** 119 SNPs were identified over the genomic region of SbDUS2 where 8 SNPs potentially affect mRNA splicing. Both "A" and "F" genotypes distinguish from the four other accessions that have homozygous reference alleles across most loci. **(D) Integration of regulatory network of selected genes in the "salmon" module:** 77 selected genes involved in different categories were displayed using Cytoscape, including genes associated with several traits (marked with different colors on the right side), and nine genes been constantly modified by dihydrouridine (dark blue). The solid 'grey' lines connected the co-expressed genes within the module, the dashed 'red' lines indicated PPI derived from the STRING database, and the 'blue' dash-arrowed lines highlighted targets bound by HMGDB15 and BPC6, as supported by Arabidopsis DAP-seq data. The red "pentagram" marked the genes overlapped with key QTLs from previous sorghum GWAS (**Ortiz et al 2017**) **(E) Heatmap of Dihydrouridine-modified transcripts:** Transcript abundance of 9 selected genes been constantly modified was displayed in a clustering heatmap, where the respective correlation with DUS2 was listed on the left ($P < 0.05$). **(F) Correlation between $CO_2$ net assimilation and transcript abundance of *SbPPDK* and *SbDnaJ*:** Two transcripts that were constantly modified by dihydrouridine (D) were correlated with the level of $CO_2$ net assimilation. **(G) Transcript and D modification status of *SbPPDK* in sorghum tissue atlas dataset:** Public RNA-seq dataset (**McCormick *et al.,* 2018**) uncovered *SbPPDK* were modified by D among19 out of 48 tissues and *SbPPDK* highly expressed in young leaf tissues. **(H) Correlation between modification and transcript status of *SbPPDK:*** *SbPPDK* transcript abundance was positively (negatively) correlated with modification density (frequency) of dihydrouridine.

## Figure 7. Functional characterization of *Atdus2* mutants

**(A) Comparison of *DUS2* in sorghum and Arabidopsis:** *DUS2* encodes the conserved DUS-FMN-like domain in both species, sharing 79.81% sequence identities (94% coverage) at the amino acid levels. The arrow above indicates the genomic position (3rd exon from the 5' end) of the T-DNA insertion line (**CS857388**) used for subsequent phenotypic analyses. **(B) Germination essay of *AtDUS2* mutants and Col-0 plants:** Germination was performed on ½ MS medium plate along with seed counts per 3 hours since 24 hours after stratification. A total of 9 plates comprised of 133 seeds were included for analysis ($p < 0.05$). **(C) Vegetative growth monitoring of Col-0 and dus2 mutants under control and heat conditions:** Image-based phenotyping over 12 days of vegetative growth with 30-minute intervals was conducted to uncover the rosette area per individual plant. The right panel uncovered the rate of surviving plants under heat treatment for both genotypes (**Top panel**). Selected plant images on the 10th day of image collection (**Bottom panel**). **(D) Photosynthesis efficiency of Col-0 and dus2 Arabidopsis mutants:** Photosynthesis efficiency was measured using Licor-6800 system derived from 4-week seedlings with large rosette leaves, represented by $CO_2$ net assimilation rate and stomatal conductance under two conditions. **(E) Reproductive growth features of Col-0 and DUS2 mutants:** Features of plant architecture and buds initiation rate between Col-0 and dus2 mutants during the transition from vegetative growth into reproductive stages.

## Supplementary Figures

**Figure S1. Genomic composition of variants of six sorghum accessions:** 8,725,449 variants (SNPs and INDELs) derived from variants calling (GATK4, reference genome: **McCormick et al., 2017)** were annotated based on genomic classifications using VEP tool, the unique variants of each accession were calculated across upstream, downstream, intron, and splicing regions.

**Figure S2. Phenotypic variation of six sorghum accessions**
**(A, B) Mebolites contents:** The content of sucrose and chlorogenic acids measured across six genotypes, four-time points, and two conditions highlights the distinguished response of accession "A".
**(C, D) Features of microbiome derivatives:** The PCA and PCA loading of bacteria (*N = 515*) and fungi (*N = 46*) uncovered the lower level of response of accession "A" under drought conditions.

**Figure S3. Upset plots of DATs between two watering conditions**
**(A, B)** The up-regulated (**left**) and down-regulated (**right**) transcripts among samples per genotype across each time point were identified in a pair-wise manner. Numbers of DATs were displayed using the upset plots. The horizontal bar charts denote the shared (or unique) DATs corresponding to the inclusion of samples below, and vertical bars denote the total numbers of DATs derived from each pair-wise comparison.

**Figure S4. Global characteristics of RNA covalent modifications (RCMs)**
**(A) Modification density and modification frequency:** The RCM density is characterized by the number of modified sites per transcript **(numbers of red triangles)**, and the RCM frequency is estimated by the ratio between the total reads containing modifications **(red lines)** and the total mapped reads per transcript. **(B) PCC of modification frequency between replicates among samples:** The correlation levels of modification frequency between replicates per sample derived from HAMR and Modtech were compared using a boxplot where the HAMR has a more stable estimation of modification frequency for each sample. **(C) Relationship between RCM density (frequency) and transcript abundance:** The transcripts-wise correlation between RCM frequency and transcript abundance was estimated using the $\log_2$(TPM) of each transcript and the $\log_2$(modification frequency). The correlation between RCM density and transcript abundance was measured using the $\log_2$(TPM) score and the $\log_2$(transcript length normalized density).

**Figure S5. Effects of modification type, genotype, time points, and treatment on global RCMs**
**(A) RCMs landscape of modified transcripts:** The effects of six modification types, six genotypes, four-time points, and two watering conditions were estimated using PCA based on the presence and absence of particular modification types (denoted by a matrix comprised of "0" and "1" to represent the status of each RCM type). **(B) RCMs density of modified transcripts for each modification type:** The effects of six genotypes, four-time points, and two watering conditions were estimated using PCA based on the modification density of particular modification types (denoted by a matrix comprised of modification density score of each RCM type).

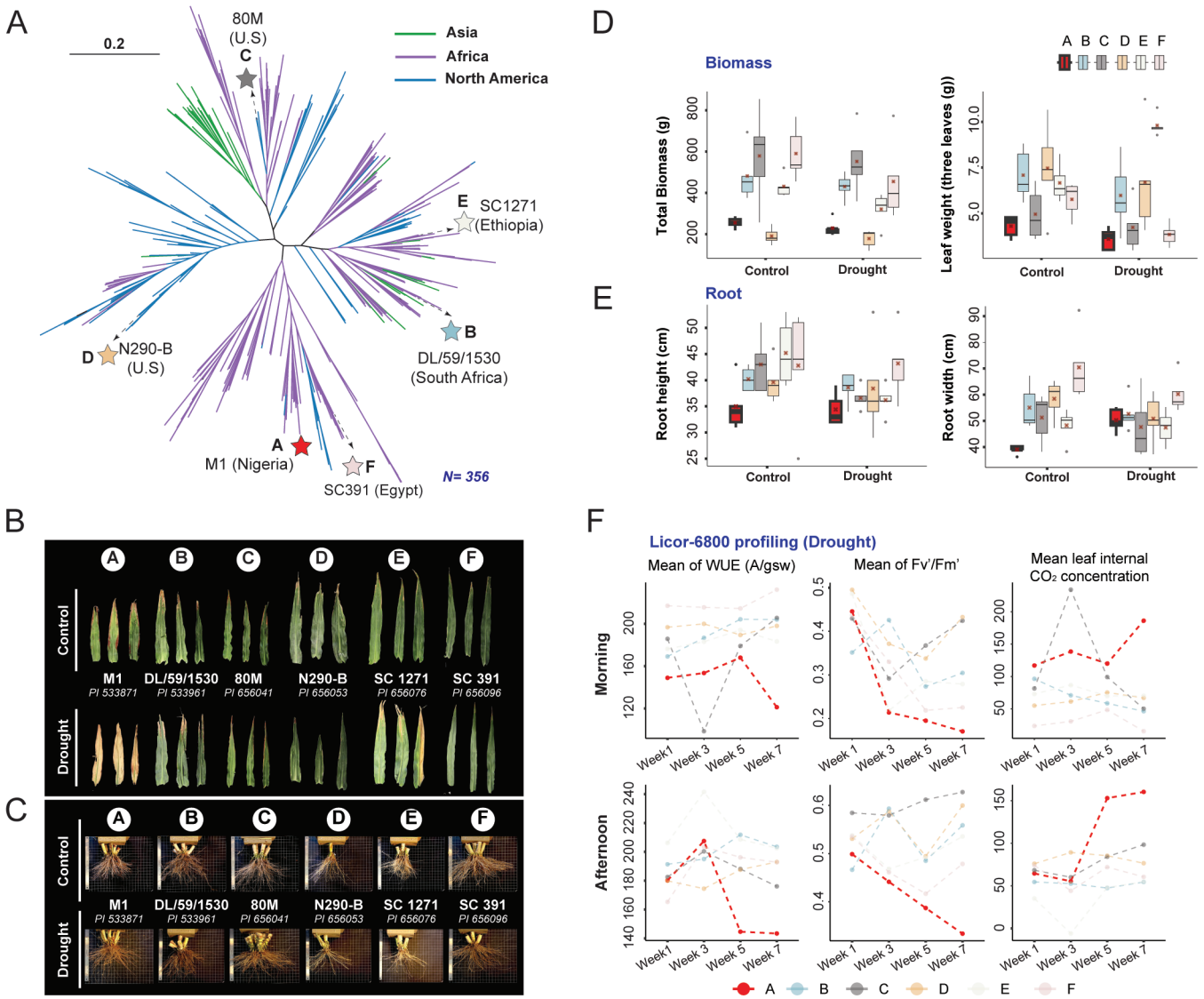
**Figure S6. Characterization of global RCM status**

**(A) RCMs landscape of modified transcripts:** Four sets of inputs were used to uncover the status of modified transcripts, including the presence and absence of RMCs (PAV), modification density, modification frequency, and transcript abundance. **(B) Characteristics of modification density (frequency) relative to treatment, time, and genotypes:** The mean of density and frequency score uncovered the accession "A" with lower density under drought and higher frequency under control relative to other accessions.

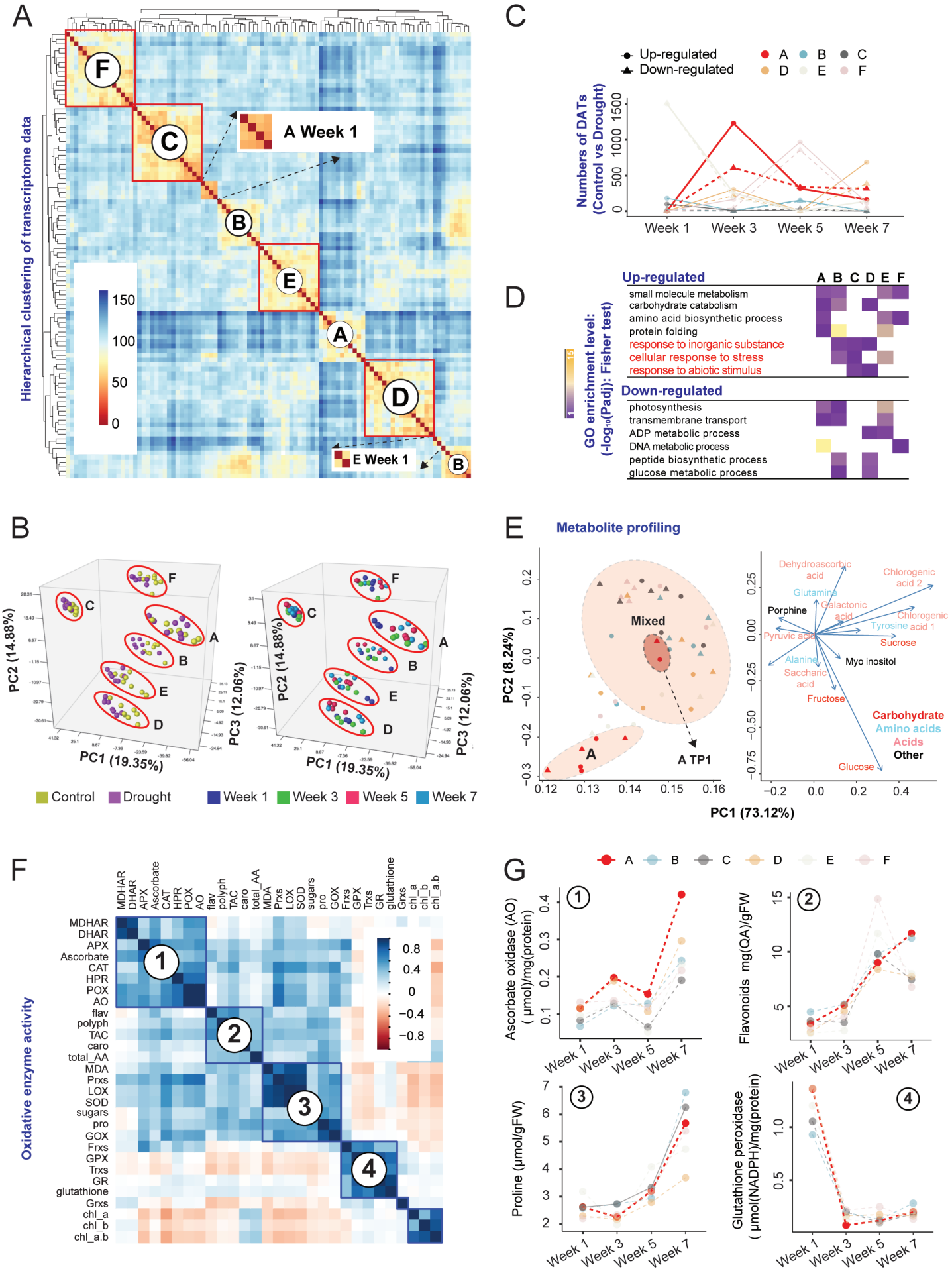
**Figure S7. Comparison of RCMs status among genotypes**

**(A, B) Number of modified transcripts:** Numbers of modified transcripts per genotype were compared using a boxplot (t-test, $P < 0.05$). The accessions "A" and "F" revealed a reduced number of modified transcripts under drought conditions while other accessions exhibited a reverse pattern. **(C) Mean modification frequency under two conditions per genotype:** The mean modification frequency scores were compared for each genotype under drought and control conditions. **(D) Mean modification density under two conditions per genotype:** The mean modification density scores were compared for each genotype under drought and control conditions. **(E) Status of transcription of modified transcripts only observed in each genotype:** Transcript abundance of genotypic-specific modified transcripts were displayed using a heatmap where the result suggest the higher levels of transcription of genotypic-specific modified transcripts. **(F) Comparison of transcript abundance of modified transcripts only observed in each genotype relative to background:** The same sets of transcripts relative to the genotypic-specific modified transcript for each genotype were selected as the background for comparison. The boxplot highlights the higher transcription level in certain genotypes compared with the background (t-test, $P < 0.05$).

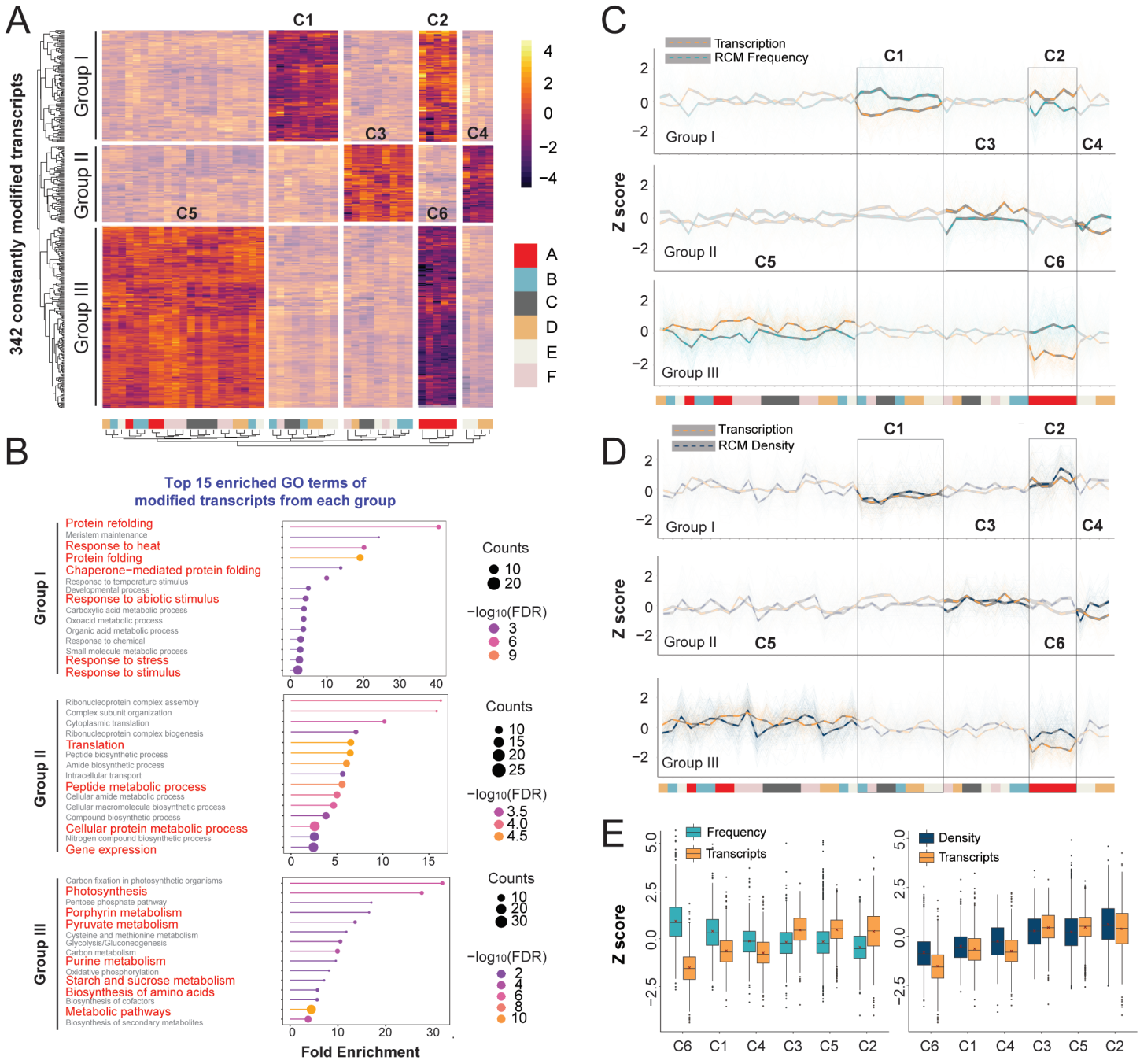
# Figure 1


A
0.2
80M
(U.S)
C
Asia
Africa
North America
E
SC1271
(Ethiopia)
B
DL/59/1530
(South Africa)
D
N290-B
(U.S)
A
M1 (Nigeria)
F
SC391 (Egypt)
N= 356
B
Control
Drought
M1
PI 533871
DL/59/1530
PI 533961
80M
PI 656041
N290-B
PI 656053
SC 1271
PI 656076
SC 391
PI 656096
C
D
Biomass
Total Biomass (g)
Leaf weight (three leaves (g))
Control
Drought
E
Root
Root height (cm)
Root width (cm)
F
Licor-6800 profiling (Drought)
Mean of WUE (A/gsw)
Mean of Fv'/Fm'
Mean leaf internal
CO₂ concentration
Morning
Afternoon
Week1
Week 3
Week 5
Week 7

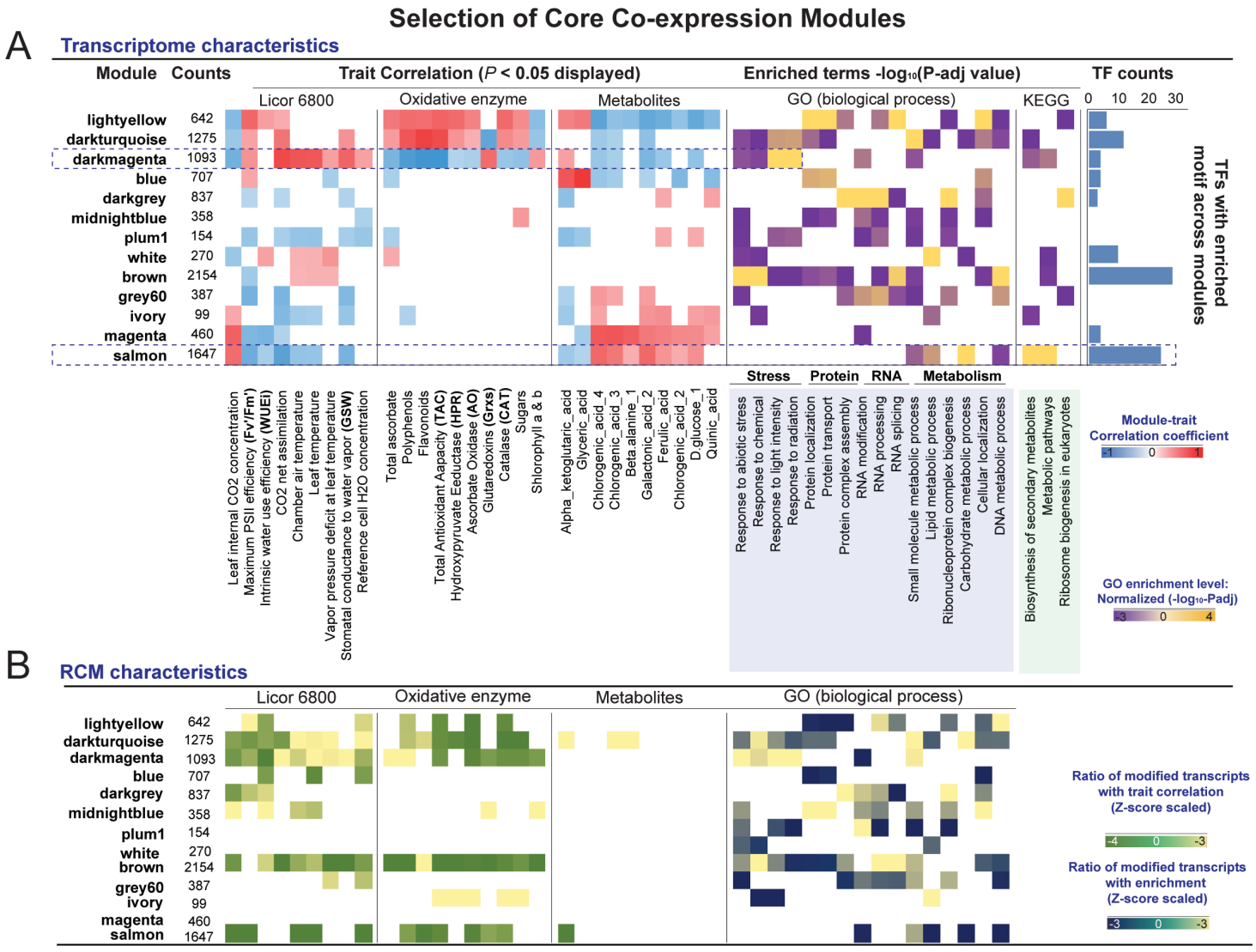
# Figure 2


A
Hierarchical clustering of transcriptome data
A Week 1
E Week 1
150
100
50
0
C
Up-regulated
Down-regulated
A B C
D E F
Numbers of DATs (Control vs Drought)
Week 1 Week 3 Week 5 Week 7
D
Up-regulated
small molecule metabolism
carbohydrate catabolism
amino acid biosynthetic process
protein folding
response to inorganic substance
cellular response to stress
response to abiotic stimulus
Down-regulated
photosynthesis
transmembrane transport
ADP metabolic process
DNA metabolic process
peptide biosynthetic process
glucose metabolic process
GO enrichment level: (-log10(Padj)): Fisher test
B
PC1 (19.35%)
PC2 (14.88%)
PC3 (12.06%)
Control Drought
Week 1 Week 3 Week 5 Week 7
E
Metabolite profiling
Mixed
A
A TP1
PC1 (73.12%)
PC2 (8.24%)
Dehydroascorbic acid
Chlorogenic acid 2
Glutamine
Galactonic acid
Porphine
Chlorogenic acid 1
Tyrosine
Pyruvic acid
Sucrose
Alanine
Myo inositol
Saccharic acid
Fructose
Carbohydrate
Amino acids
Acids
Other
Glucose
F
Oxidative enzyme activity
MDHAR DHAR APX Ascorbate CAT HPR POX AO flav polyph TAC caro total_AA MDA Prxs LOX SOD sugars pro GOX Frxs GPX Trxs GR glutathione Grxs chl_a chl_b chl_a.b
G
Ascorbate oxidase (AO) ( μmol)/mg(protein)
Flavonoids mg(QA)/gFW
Proline (μmol/gFW)
Glutathione peroxidase ( μmol(NADPH)/mg(protein)
Week 1 Week 3 Week 5 Week 7

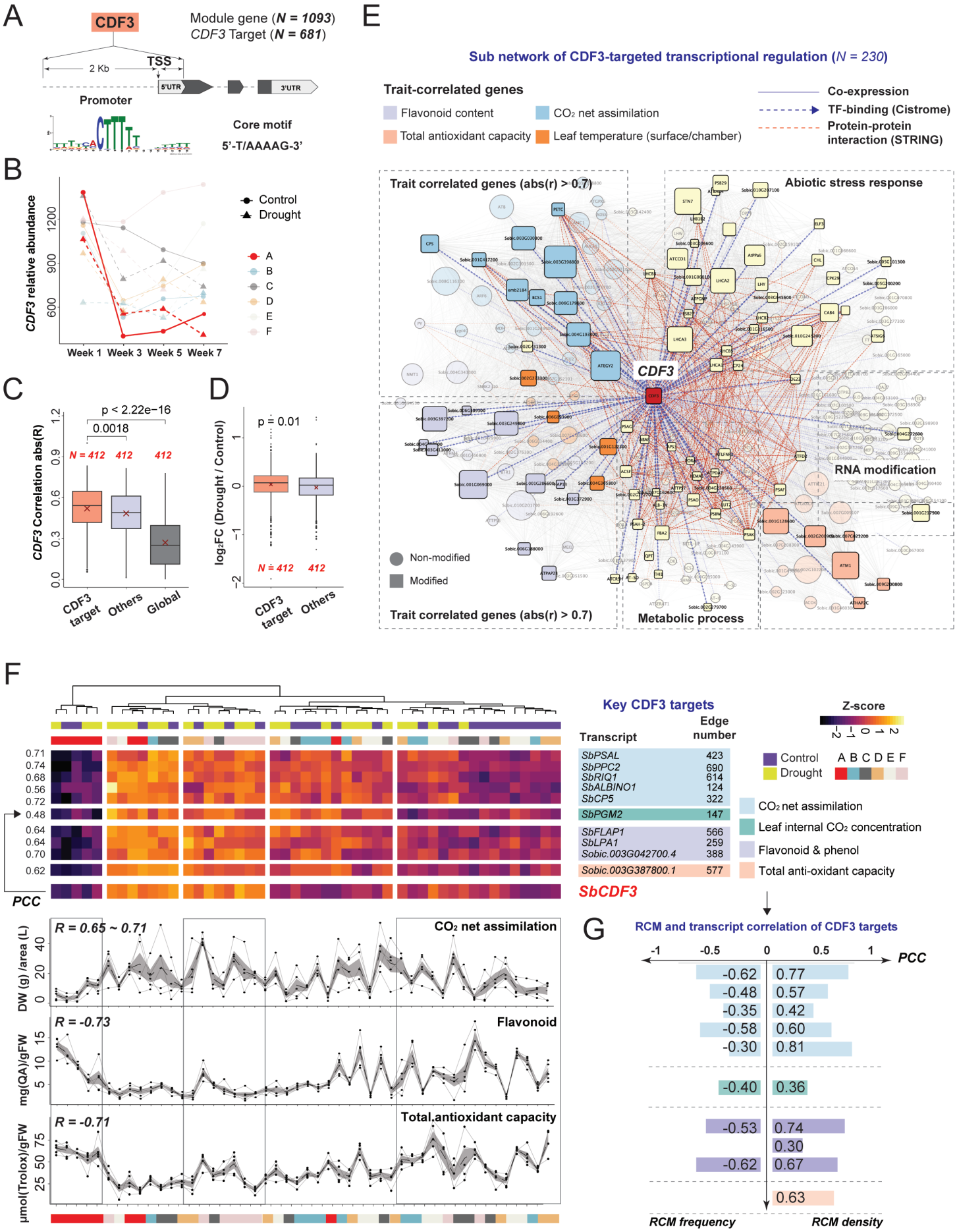
# Figure 3


A
342 constantly modified transcripts
Group I
Group II
Group III
C1
C2
C3
C4
C5
C6
4
2
0
−2
−4
A
B
C
D
E
F
B
Top 15 enriched GO terms of modified transcripts from each group
Group I
Protein refolding
Meristem maintenance
Response to heat
Protein folding
Chaperone−mediated protein folding
Response to temperature stimulus
Developmental process
Response to abiotic stimulus
Carboxylic acid metabolic process
Oxoacid metabolic process
Organic acid metabolic process
Response to chemical
Small molecule metabolic process
Response to stress
Response to stimulus
Counts
10
20
−log10(FDR)
3
6
9
Group II
Ribonucleoprotein complex assembly
Complex subunit organization
Cytoplasmic translation
Ribonucleoprotein complex biogenesis
Translation
Peptide biosynthetic process
Amide biosynthetic process
Intracellular transport
Peptide metabolic process
Cellular amide metabolic process
Cellular macromolecule biosynthetic process
Compound biosynthetic process
Cellular protein metabolic process
Nitrogen compound biosynthetic process
Gene expression
Counts
10
15
20
25
−log10(FDR)
3.5
4.0
4.5
Group III
Carbon fixation in photosynthetic organisms
Photosynthesis
Pentose phosphate pathway
Porphyrin metabolism
Pyruvate metabolism
Cysteine and methionine metabolism
Glycolysis/Gluconeogenesis
Carbon metabolism
Purine metabolism
Oxidative phosphorylation
Starch and sucrose metabolism
Biosynthesis of amino acids
Biosynthesis of cofactors
Metabolic pathways
Biosynthesis of secondary metabolites
Counts
10
20
30
−log10(FDR)
2
4
6
8
10
Fold Enrichment
C
Transcription
RCM Frequency
Group I
Group II
Group III
Z score
D
Transcription
RCM Density
E
Frequency
Transcripts
Density
Transcripts
Z score
C6
C1
C4
C3
C5
C2

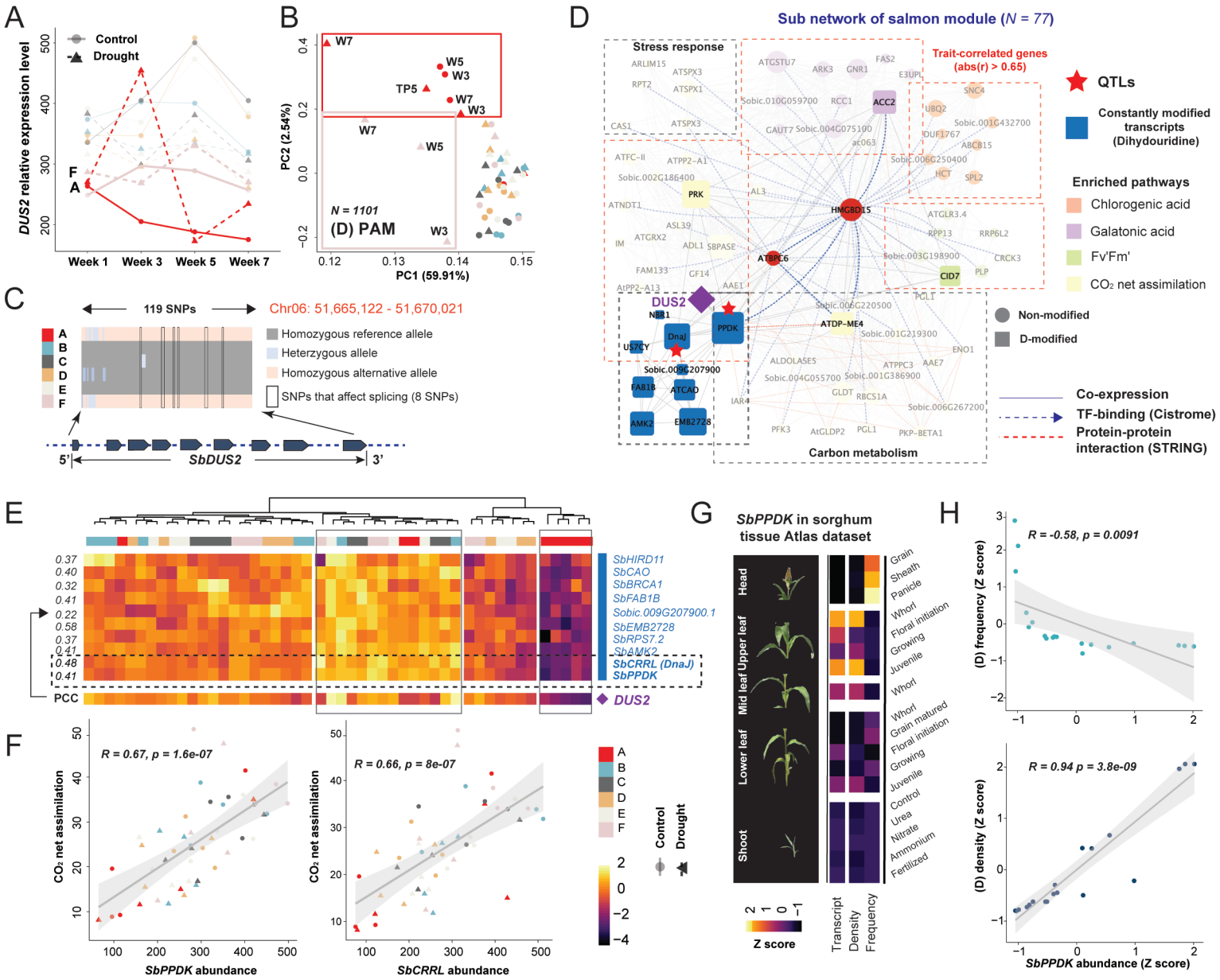
# Figure 4


Selection of Core Co-expression Modules
A
Transcriptome characteristics
Module
Counts
Trait Correlation (P < 0.05 displayed)
Enriched terms -log10(P-adj value)
TF counts
Licor 6800
Oxidative enzyme
Metabolites
GO (biological process)
KEGG
lightyellow 642
darkturquoise 1275
darkmagenta 1093
blue 707
darkgrey 837
midnightblue 358
plum1 154
white 270
brown 2154
grey60 387
ivory 99
magenta 460
salmon 1647
TFs with enriched motif across modules
Leaf internal CO2 concentration
Maximum PSII efficiency (Fv'/Fm')
Intrinsic water use efficiency (WUEi)
CO2 net assimilation
Chamber air temperature
Leaf temperature
Vapor pressure deficit at leaf temperature
Stomatal conductance to water vapor (GSW)
Reference cell H2O concentration
Total ascorbate
Polyphenols
Flavonoids
Total Antioxidant Aapacity (TAC)
Hydroxypyruvate Eeductase (HPR)
Ascorbate Oxidase (AO)
Glutaredoxins (Grxs)
Catalase (CAT)
Sugars
Shlorophyll a & b
Alpha_ketoglutaric_acid
Glyceric_acid
Chlorogenic_acid_4
Chlorogenic_acid_3
Beta.alanine_1
Galactonic_acid_2
Ferulic_acid
Chlorogenic_acid_2
D.glucose_1
Quinic_acid
Stress
Protein
RNA
Metabolism
Response to abiotic stress
Response to chemical
Response to light intensity
Response to radiation
Protein localization
Protein transport
Protein complex assembly
RNA modification
RNA processing
RNA splicing
Small molecule metabolic process
Lipid metabolic process
Ribonucleoprotein complex biogenesis
Carbohydrate metabolic process
Cellular localization
DNA metabolic process
Biosynthesis of secondary metabolites
Metabolic pathways
Ribosome biogenesis in eukaryotes
Module-trait Correlation coefficient
GO enrichment level: Normalized (-log10-Padj)
B
RCM characteristics
Licor 6800
Oxidative enzyme
Metabolites
GO (biological process)
Ratio of modified transcripts with trait correlation (Z-score scaled)
Ratio of modified transcripts with enrichment (Z-score scaled)
C
Eigenvalue of module
Genotype A B C D E F
Condition Control Drought
Salmon module
Darkmagenta module
Week 1
Week 3
Week 5
Week 7
D
Percent of constantly modified transcripts in module
Group I
99%
N = 103
Group II
11%
89%
N = 71
Group III
17%
25%
58%
N = 167
Other modules
darkmagenta
Salmon

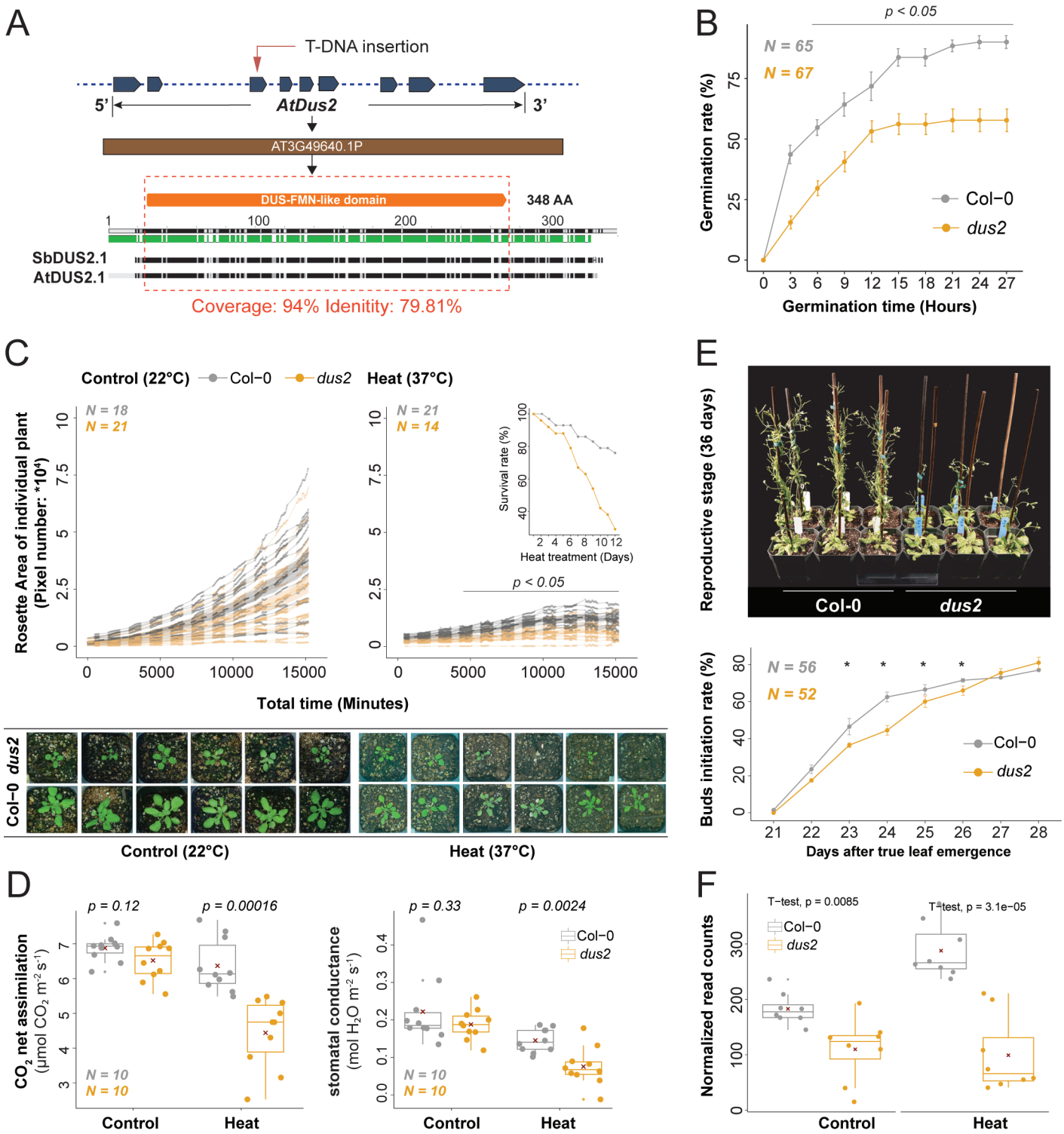
# Figure 5


A
CDF3
Module gene (N = 1093)
CDF3 Target (N = 681)
2 Kb
TSS
5'UTR
3'UTR
Promoter
Core motif
5'-T/AAAAG-3'
B
CDF3 relative abundance
Control
Drought
A
B
C
D
E
F
Week 1
Week 3
Week 5
Week 7
C
CDF3 Correlation abs(R)
p < 2.22e−16
0.0018
N = 412
412
412
CDF3 target
Others
Global
D
log2FC (Drought / Control)
p = 0.01
N = 412
412
CDF3 target
Others
E
Sub network of CDF3-targeted transcriptional regulation (N = 230)
Trait-correlated genes
Flavonoid content
CO2 net assimilation
Total antioxidant capacity
Leaf temperature (surface/chamber)
Co-expression
TF-binding (Cistrome)
Protein-protein interaction (STRING)
Trait correlated genes (abs(r) > 0.7)
Abiotic stress response
CDF3
RNA modification
Non-modified
Modified
Trait correlated genes (abs(r) > 0.7)
Metabolic process
F
PCC
Key CDF3 targets
Transcript
Edge number
SbPSAL 423
SbPPC2 690
SbRIQ1 614
SbALBINO1 124
SbCP5 322
SbPGM2 147
SbFLAP1 566
SbLPA1 259
Sobic.003G042700.4 388
Sobic.003G387800.1 577
SbCDF3
Z-score
Control
Drought
A B C D E F
CO2 net assimilation
Leaf internal CO2 concentration
Flavonoid & phenol
Total anti-oxidant capacity
R = 0.65 ~ 0.71
CO2 net assimilation
DW (g) /area (L)
R = -0.73
Flavonoid
mg(QA)/gFW
R = -0.71
Total antioxidant capacity
μmol(Trolox)/gFW
G
RCM and transcript correlation of CDF3 targets
PCC
-0.62 0.77
-0.48 0.57
-0.35 0.42
-0.58 0.60
-0.30 0.81
-0.40 0.36
-0.53 0.74
0.30
-0.62 0.67
0.63
RCM frequency
RCM density

# Figure 6


A
Control
Drought
DUS2 relative expression level
Week 1
Week 3
Week 5
Week 7
B
PC2 (2.54%)
PC1 (59.91%)
N = 1101
(D) PAM
C
119 SNPs
Chr06: 51,665,122 - 51,670,021
Homozygous reference allele
Heterzygous allele
Homozygous alternative allele
SNPs that affect splicing (8 SNPs)
SbDUS2
D
Sub network of salmon module (N = 77)
Stress response
Trait-correlated genes (abs(r) > 0.65)
QTLs
Constantly modified transcripts (Dihydouridine)
Enriched pathways
Chlorogenic acid
Galatonic acid
Fv'Fm'
CO2 net assimilation
Non-modified
D-modified
Co-expression
TF-binding (Cistrome)
Protein-protein interaction (STRING)
Carbon metabolism
E
PCC
DUS2
SbHIRD11
SbCAO
SbBRCA1
SbFAB1B
Sobic.009G207900.1
SbEMB2728
SbRPS7.2
SbAMK2
SbCRRL (DnaJ)
SbPPDK
F
R = 0.67, p = 1.6e-07
R = 0.66, p = 8e-07
CO2 net assimilation
SbPPDK abundance
SbCRRL abundance
Control
Drought
G
SbPPDK in sorghum tissue Atlas dataset
Head
Upper leaf
Mid leaf
Lower leaf
Shoot
Grain
Sheath
Panicle
Whorl
Floral initiation
Growing
Juvenile
Grain matured
Control
Urea
Nitrate
Ammonium
Fertilized
Z score
Transcript
Density
Frequency
H
R = -0.58, p = 0.0091
R = 0.94 p = 3.8e-09
(D) frequency (Z score)
(D) density (Z score)
SbPPDK abundance (Z score)

# Figure 7


A
T-DNA insertion
5'
AtDus2
3'
AT3G49640.1P
DUS-FMN-like domain
348 AA
1
100
200
300
SbDUS2.1
AtDUS2.1
Coverage: 94% Idenitity: 79.81%
B
p < 0.05
N = 65
N = 67
Germination rate (%)
Germination time (Hours)
Col−0
dus2
C
Control (22°C)
Col−0
dus2
Heat (37°C)
N = 18
N = 21
N = 21
N = 14
Rosette Area of individual plant (Pixel number: *10⁴)
Survival rate (%)
Heat treatment (Days)
p < 0.05
Total time (Minutes)
Col−0
dus2
Control (22°C)
Heat (37°C)
E
Reproductive stage (36 days)
Col-0
dus2
N = 56
N = 52
Buds initiation rate (%)
Days after true leaf emergence
Col−0
dus2
D
p = 0.12
p = 0.00016
p = 0.33
p = 0.0024
CO2 net assimilation (µmol CO2 m-2 s-1)
stomatal conductance (mol H2O m-2 s-1)
N = 10
N = 10
N = 10
N = 10
Control
Heat
Col−0
dus2
F
T−test, p = 0.0085
T−test, p = 3.1e−05
Normalized read counts
Control
Heat