



# Combined GWAS and Transcriptome Analyses Provide New Insights Into the Response Mechanisms of Sunflower Against Drought Stress

Yang Wu<sup>1</sup>, Huimin Shi<sup>1</sup>, Haifeng Yu<sup>2</sup>, Yu Ma<sup>2</sup>, Haibo Hu<sup>1</sup>, Zhigang Han<sup>2</sup>, Yonghu Zhang<sup>2</sup>, Zilong Zhen<sup>1</sup>, Liuxi Yi<sup>1\*</sup> and Jianhua Hou<sup>1\*</sup>

<sup>1</sup> College of Agricultural, Inner Mongolia Agricultural University, Hohhot, China, <sup>2</sup> Institute of Crop Breeding and Cultivation, Inner Mongolia Academy of Agricultural and Husbandry Sciences, Hohhot, China

## OPEN ACCESS

### Edited by:

Dirk Walther,  
Max Planck Institute of Molecular  
Plant Physiology, Germany

### Reviewed by:

Samar Gamal Thabet,  
Fayoum University, Egypt  
Frank Maulana,  
Louisiana State University Agricultural  
Center, United States  
Zohreh Hajjbarat,  
Shahid Beheshti University, Iran

### \*Correspondence:

Jianhua Hou  
houjh@imau.edu.cn  
Liuxi Yi  
yiliuxivip@163.com

### Specialty section:

This article was submitted to  
Plant Bioinformatics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 02 January 2022

**Accepted:** 31 March 2022

**Published:** 03 May 2022

### Citation:

Wu Y, Shi H, Yu H, Ma Y, Hu H, Han Z,  
Zhang Y, Zhen Z, Yi L and Hou J  
(2022) Combined GWAS and  
Transcriptome Analyses Provide New  
Insights Into the Response  
Mechanisms of Sunflower Against  
Drought Stress.  
*Front. Plant Sci.* 13:847435.  
doi: 10.3389/fpls.2022.847435

Sunflower is one of the most important oil crops in the world, and drought stress can severely limit its production and quality. To understand the underlying mechanism of drought tolerance, and identify candidate genes for drought tolerance breeding, we conducted a combined genome-wide association studies (GWAS) and RNA-seq analysis. A total of 226 sunflower inbred lines were collected from different regions of China and other countries. Eight phenotypic traits were evaluated under control and drought stress conditions. Genotyping was performed using a Specific-Locus Amplified Fragment Sequencing (SLAF-seq) approach. A total of 934.08 M paired-end reads were generated, with an average Q30 of 91.97%. Based on the 243,291 polymorphic SLAF tags, a total of 94,162 high-quality SNPs were identified. Subsequent analysis of linkage disequilibrium (LD) and population structure in the 226 accessions was carried out based on the 94,162 high-quality SNPs. The average LD decay across the genome was 20 kb. Admixture analysis indicated that the entire population most likely originated from 11 ancestors. GWAS was performed using three methods (MLM, FarmCPU, and BLINK) simultaneously. A total of 80 SNPs showed significant associations with the 8 traits ( $p < 1.062 \times 10^{-6}$ ). Next, a total of 118 candidate genes were found. To obtain more reliable candidate genes, RNA-seq analysis was subsequently performed. An inbred line with the highest drought tolerance was selected according to phenotypic traits. RNA was extracted from leaves at 0, 7, and 14 days of drought treatment. A total of 18,922 differentially expressed genes were obtained. Gene ontology and Kyoto Encyclopedia of Genes and Genomes analysis showed up-regulated genes were mainly enriched in the branched-chain amino acid catabolic process, while the down-regulated genes were mainly enriched in the photosynthesis-related process. Six DEGs were randomly selected from all DEGs for validation; these genes showed similar patterns in RNA-seq and RT-qPCR analysis, with a correlation coefficient of 0.8167. Through the integration of the genome-wide association study and the RNA-sequencing, 14 candidate genes were identified. Four of them (LOC110885273, LOC110872899, LOC110891369, LOC110920644) were abscisic acid related protein kinases and transcription factors.

These genes may play an important role in sunflower drought response and will be used for further study. Our findings provide new insights into the response mechanisms of sunflowers against drought stress and contribute to further genetic breeding.

**Keywords:** sunflower, drought stress, genome-wide association studies (GWAS), RNA-seq, single-nucleotide polymorphisms (SNPs), specific-locus amplified fragment sequencing (SLAF-seq)

## INTRODUCTION

Sunflower (*Helianthus annuus*, L) belongs to the Compositae family (Schilling and Heiser, 1981), and is native to North America (Schilling and Heiser, 1981). As one of the major oilseed crops in the world, sunflower is considered an important source of high-quality oil and dietary fiber for human health (Khan et al., 2015). The world harvested area of sunflower seed has increased by 20% (from 23.07 million hectares to 27.87 million hectares), and the production has increased by more than 50% (from 31.45 million tons to 50.23 million tons) from 2010 to 2020 (FAO, 2021). China is the sixth-largest sunflower-producing country in the world. The main production areas of sunflowers in China are in the northwest region, such as Inner Mongolia Autonomous Region and Xinjiang Uygur autonomous region. The sunflower is an important economic source for local farmers, and the status of sunflower production directly affects farmers' living standards.

The global average temperature has risen by about 0.85°C from the year 1880 to 2012 (Adopted, 2014), resulting in a series of extreme weather events, such as heavy rains, flooding, drought, and desertification. Among them, drought is the most serious abiotic stress limiting global agricultural production (Wilhite and Buchanan-Smith, 2005). A persistent drought can cause a large number of deaths and force large-scale migration, while severe droughts can even impact human civilization (Ault, 2020). With the continued climate change and population growth, drought may pose a serious threat to global and regional food security in the coming decades (Riddell et al., 2018). Due to the strong root system, the sunflower was considered to be relatively tolerant to water stress. They are often seeded on beds and ridges with poor moisture conditions where many other crops are unable to survive (Hussain et al., 2018). As a result, it is more susceptible to drought stress leading to yield reduction (Pasda and Diepenbrock, 1990; Adeleke and Babalola, 2020; Grasso et al., 2020). Studies have shown that drought stress in sunflower seedlings can lead to severe yield loss (Mwale et al., 2003; Rauf and Ahmad Sadaqat, 2008).

The sunflower drought stress response behavior involves a series of changes in morphological, physiological, and molecular levels. The drought stress negatively influenced seed germination and seedling emergence at the germination stage (Kaya et al., 2006). Drought stress at the vegetative stage reduces plant height (PH), leaf surface area (LSA), and biomass production while causing pollen sterility at the reproductive stage (Turhan and Baser, 2004; Hussain et al., 2008). From a physiological perspective, drought affects the uptake of water and nutrition, leads to a reduction of relative water content (RWC), and the turgor of cells (Hussain et al., 2008, 2016; Ibrahim et al.,

2016). Plants respond to drought stress by reducing water evaporation through stomatal closure. As a result, it also reduces the photosynthetic rate (Flexas et al., 2004). The decreased photosynthesis rate leads to a decrease in CO<sub>2</sub> fixation, which affects the regeneration of the final acceptor of the electron transport chain (NADP<sup>+</sup>). The leaked electrons flow to O<sub>2</sub> to produce reactive oxygen species (ROS) (Flexas et al., 2004). ROS cause oxidation of membrane lipids, resulting in decreased cell membrane stability. The decrease in cell membrane permeability results in the accumulation of the relative electrical conductivity (REC) and malondialdehyde (MDA) (Gunes et al., 2008). From the molecular level, plants involve a series of pathways for signal perception, transduction, gene expression, and other stress metabolites to accommodate drought. Drought-induced genes can mainly be classified into two groups. The first group constitutes genes whose products directly function in tolerance to stress, such as LEA proteins, osmolytes, proline (Pro), CAT, POD. Another group includes genes playing a role in signal transduction as well as the regulation of gene expression including various transcription factors (TF), protein kinases (PK), and transcriptional regulators (TR) (Lata et al., 2015).

Some agronomic measures can mitigate the damage of drought impact on plants, such as exogenous applications of plant hormones, osmotic regulators, and mineral nutrients (Salami and Saadat, 2013; Rabert et al., 2014). However, these changes are not heritable, and need additional labor, capital, and technology investment. Coping with drought through the breeding approach is usually the most effective and economical strategy. The genetic modification within the plant is heritable. Once a gene is introduced into a breeding material, it will be a permanent source of drought tolerance (Rauf, 2008). Drought tolerance in plants is a complex quantitative trait involving many micro-effective genes (Blum, 2011). Molecular-based plant drought resistance breeding is a hot spot in recent years (Wang and Qin, 2017). Previous studies on the molecular mechanism of sunflower drought resistance were mostly based on linkage analysis (Kiani et al., 2007; Poormohammad Kiani et al., 2009; Haddadi et al., 2011). However, the linkage analysis population was on two parents with significantly different phenotypes and the recombinant inbred lines (RILs). Only genes in RILs that show a significant difference between parental lines could be detected.

Genome wide association study (GWAS) is an observational study to detect associations between genetic variants and traits in individuals (Togninalli et al., 2018). Compared to linkage analysis, GWAS uses a natural population, which eliminates the need to construct a population. Therefore, the time consumption is greatly reduced. The use of natural populations

allows GWAS to simultaneously detect many natural allelic variations (Ma et al., 2018). In addition, the natural population contains all the historical recombination information and thus provide relatively higher detection accuracy than bi-parental populations (Kofsky et al., 2020). GWAS has been widely used in plant drought research, such as wheat (*Triticum aestivum* L.), cotton (*Gossypium herbaceum* L.), rice (*Oryza sativa* L.), and potato (*Solanum tuberosum* L.) (Ma et al., 2016; Mwadzingeni et al., 2017; Hou et al., 2018; Tagliotti et al., 2021). RNA-sequencing (RNA-Seq) is another attractive omics tool to identify differentially expressed genes (DEGs) under different conditions. Further analysis can provide insight into the changes in the DEGs expression level, important biological processes, and pathways (Zhang et al., 2017). Combining GWAS with RNA-seq can decrease the higher false-positive rate (FDR) inherent in GWAS analysis, and improve the accuracy of gene selection (Xie et al., 2019; Wang et al., 2022). However, to our knowledge, there are no relevant studies on sunflowers.

Molecular marker-based genotyping is an important step in GWAS analysis. Most traditional molecular markers were based on sequence length polymorphism. However, it could not be used for large-scale genotyping due to low throughput (Sun et al., 2013b). Whole gene sequencing technology is restricted in its use for non-model organisms due to population size and price (Muir et al., 2016). One strategy to reduce the sequencing cost was to reduce representation libraries (RRL). Specific length amplified fragment sequencing (SLAF) is one of the representative techniques, which uses specific enzymes to digest the genomes, and select a given size range of restriction fragments based on personalized research purposes (Sun et al., 2013b). This approach maintains the marker density while reducing the volume of sequencing, lowering the cost.

In this study, we performed a GWAS analysis of 226 sunflower varieties based on SLAF-seq. Then, a drought-tolerant accession was selected for RNA-seq analysis. Several important candidate genes were obtained using a combined analysis. Our research objectives were to (1) investigate the phenotypic variations among accessions under different water conditions; (2) develop new drought-related SNPs and identify genetic variants; (3) understand gene expression patterns under different drought stress time points, and reveal important biological processes and pathways; (4) obtain important genes associated with drought tolerance.

## MATERIALS AND METHODS

### Plant Materials and Growth Condition

A total of 226 sunflower inbred lines were collected from different countries (Australia, U.S.A., and France) and different provinces in China (Inner Mongolia, Ningxia, Xinjiang, Liaoning, Jilin). Seventy-three of them were provided by the Inner Mongolia Academy of Agriculture and Animal Husbandry, and 153 were kept in our laboratory. The experiment was conducted in the summer of 2019 at the Inner Mongolia Agricultural University, China (111.71, 40.82, 1,000 m above sea level). Seeds with fully mature, healthy, and uniform sizes were sorted for drought-stress experiments. After sterilization with 0.2% (w/v) mercuric

chloride ( $\text{HgCl}_2$ ), all seeds were rinsed several times with distilled water and soaked in deionized water for 24 h. Then the seeds were sown in plastic flowerpots ( $25 \times 19 \times 16$  cm) filled with 3 kg soil (sandy soil and organic humus in a ratio of 2:1). Each pot was planted with 10 seeds and each accession had 6 pots. To avoid interference from natural rainfall and other factors, all pots were placed in a greenhouse (light/dark cycles: 14 h/10 h; 28/22°C;  $45 \pm 5\%$  relative humidity) without water and nutritional limitation.

### Experimental Design and Drought Treatments

When seedlings grew to the stage of three leaves, six pots of each accession were randomly and equally divided into two groups. Each group contained three pots as three biological replicates. The different watering regime was imposed on these two groups. One group continued to irrigate with sufficient water, and maintain the soil moisture content of  $30 \pm 2\%$  as a control group (WW). Another group kept the soil moisture content to  $10 \pm 2\%$  as a treatment group (DS). The soil moisture content of each pot was determined at 9 a.m. every day using the weight method described by Soni and Abdin (2017) and supplemented with water according to the target soil moisture content.

### Phenotypic Evaluation and Statistical Analysis

The experiment lasted for 15 days, then 5 plants were randomly selected from each pot for phenotypic evaluation. Plant height (PH) was measured directly with a ruler. Leaf surface area (LSA) was calculated by the leaf area co-efficient method (Alza and Fernandez-Martinez, 1997). Root shoot ratio (RSR) was measured by the gravimetric method. Total root length (RL), root volume (RV), and root surface area (RSA) were measured with an LA-S root scanner (Wanshen Testing Technology Co., Ltd., Hangzhou, China). The relative water content (RWC) was detected using the saturate water method by Galmes et al. (2011). The chlorophyll concentration was assessed using a SPAD chlorophyll meter (TYS-A, TOP Instrument Co., Ltd., Hangzhou, China).

Data were analyzed using SPSS software (SPSS for Windows, V20.0.0, SPSS, Chicago, Illinois). Normality distribution was preliminarily assessed by a one-sample Kolmogorov-Smirnov's goodness-to-fit test (K-S test). For statistical differences between WW and DS growth condition, the Student *t*-test (normal distribution) and Wilcoxon signed-rank test (non-normal distribution) was used. Spearman non-parametric correlations were used to determine the correlation coefficient and statistical significance. Corrplot and Pheatmap R package were used to visualize the correlation.

### Genomic DNA Extraction and Restriction Enzyme Selection

Total genomic DNA was extracted from 100 mg of fresh leaves by the CTAB method with a plant genomic DNA kit DP305 (Tiangen Biotech, China). To ensure it met the requirements for SLAF-seq (concentration  $\geq 20$  ng/ $\mu\text{l}$ ; volume  $\geq 30/\mu\text{l}$ ), the concentration and quality of DNA were determined

using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

The SLAF-seq technique requires breaking the genome into small fragments using restriction enzymes. Then selecting restriction fragments of a specific length range (defined as SLAF-seq) for sequencing. To evaluate the number of target fragments produced *via* different combinations of restriction enzymes, a *in silico* pre-experiment for enzyme selection was conducted. The criteria for enzyme selection were as follows: (1) the proportion of restriction fragments located in repetitive sequences is as low as possible; (2) The restriction fragments are distributed evenly on the genome as far as possible; (3) Consistency between the length of restriction fragments and the specific experimental system (Davey et al., 2013); (4) The number of restriction fragments with lengths 364–464 pb (SLAF tags in sunflower) should exceed 300,000.

## SLAF Library Construction and High Throughput Sequencing

The SLAF library construction and high-throughput sequencing were performed as described by Sun et al. (2013b). After a series of polymerase chain reactions (PCR), adapter ligation reactions, and agarose gel purification, the SLAF-tags were isolated and subjected to PCR amplification following the guide of Illumine sample preparation. The paired-end sequencing was performed on an Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA) at Beijing Biomarker Technologies Corporation (Beijing, China). Sequencing quality was estimated by measuring the guanine-cytosine (GC) content and Q30 ratio. A Q value of 30 represents a 0.1% error probability and 99.9% confidence level. Reads with >90% identity were clustered into a single SLAF-tag using BLAT software, and SLAF-tags with a sequence that varied across samples were defined as polymorphic SLAF tags (Zhang et al., 2018). To test the accuracy of the restriction enzyme digestion protocol, we used the genome of *Oryza sativa ssp. japonica* as a control (374.30 Mb, <http://rapdb.dna.affrc.go.jp/>).

## SNP Genotyping and Linkage Disequilibrium Analysis

All reads were processed for quality control and filtered using Seqtk (<https://github.com/lh3/seqtk>) software. High-quality paired-end reads were aligned to the reference genome ([https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation\\_releases/4232/100/GCF\\_002127325.1\\_HanXRQr1.0/](https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/4232/100/GCF_002127325.1_HanXRQr1.0/)) using Burrows-Wheeler Aligner (BWA) software (Li and Durbin, 2009). SNP calling was conducted using the HaplotypeCaller function of Genome Analysis Toolkit (GATK) (McKenna et al., 2010). The VCF files obtained by GATK were converted to PLINK files using VCFtools (v0.1.16) (Danecek et al., 2011). SNPs with an integrity ratio of <0.8 and MAF <0.05 were filtered out *via* PLINK software (v1.90b6.21) (Purcell et al., 2007). Linkage disequilibrium (LD) was estimated by measuring the squared allele frequency correlations ( $r^2$ ) (VanLiere and Rosenberg, 2008) between pairs of SNPs *via* PLINK software, with  $r^2 = 1$  indicating complete LD, and  $r^2 = 0$  indicating absent LD. LD decay extent was defined as

the physical genomic distance at which the  $r^2$  decreased to half of its maximum value. PopLDdecay software (Zhang et al., 2019) was used to visualize the mean  $r^2$  of all chromosomes within the 100 kb region.

## Population Structure Analysis

Based on the filtered SNPs, population analysis, phylogeny analysis, and principal component analysis (PCA) were performed in turns. Admixture software v1.3.0 (Alexander et al., 2009) was used to analyze the population structure. The number of underlying population groups  $K$  was predefined as 1–13 using the maximum likelihood estimation approach. The cross-validation errors (CV) for each  $K$  value were calculated. The  $K$  value with the lowest CV error was selected as the optimal number of populations. The Pophelper R package was used to make multiline plots (Francis, 2017). The genetic distances were calculated using VCF2Dis-1.45 (<https://github.com/BGI-shenzhen/VCF2Dis>). The FastME (v 2.0) software (Lefort et al., 2015) was used to convert the mat file obtained in the previous step into a distance matrix file (\*.nwk). The phylogenetic trees were constructed using the neighbor-joining method in the iTOL server (<https://itol.embl.de/>) (Letunic and Bork, 2021). PCA was performed using PLINK software by the `-pca` function. The first three components were used to plot the PCA *via* the `rgl` (v. 0.107.14) R package (Adler et al., 2003).

## Genomic-Wide Association Study

The GWAS analysis was conducted using three methods: mixed linear model (MLM), Fixed and random model Circulating Probability Unificatin (FarmCPU), and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) in GAPIT R package (Lipka et al., 2012). The phenotypic data of each accession was represented using two indices: stress tolerance index (STI) (Fernandez, 1992), and stress susceptibility index (SSI) (Fischer and Maurer, 1978). These were calculated as follows:

$$STI = \frac{Y_{si} \times Y_{pi}}{\bar{Y}_{pi}^2}$$

$$SSI = \frac{1 - \frac{Y_{si}}{\bar{Y}_{pi}}}{1 - \frac{\bar{Y}_{si}}{\bar{Y}_{pi}}}$$

where  $Y_{si}$  = performance of a genotype under stress;  $Y_{pi}$  = performance of the same genotype under control conditions;  $\bar{Y}_{si}$  = mean  $Y_{si}$  of all genotypes,  $\bar{Y}_{pi}$  = mean  $Y_{pi}$  of all genotypes.

The first three principal components were used as covariates. The GAPIT uses genotype data to automatically generate kinship matrix and calculate population structure according to the needs of different methods. For the identification of true marker-trait association, the significant  $p$ -value was set as  $p < 1.062 \times 10^{-6}$  ( $p = 0.1/n$ ;  $n$  = total markers used, which is roughly a Bonferroni correction, corresponding to  $-\log_{10}(p) = 5.97$ , blue line in the Manhattan plots) (Zhou et al., 2017). The Manhattan plot was used to show the correlation between SNP and phenotypic traits.

The Quantile-quantile (Q-Q) plot was used to display the level of difference between observed and predicted values. Both the Manhattan plots and Q-Q plots were constructed using CMplot R package (Yin, 2018).

## GWAS Candidate Gene Search and Combined Analysis

The region of GWAS candidate genes was defined by the average LD decay distance. Genes located within 20 kb flanking regions on either side of the significantly associated SNPs were considered as candidate genes. Function annotations were conducted using the EggNog (Huerta-Cepas et al., 2019) and Pfam (Bateman et al., 2004) software. The blast software was used to search for *Arabidopsis thaliana* genes homologous to candidate genes in the TAIR database (<https://www.arabidopsis.org>). Transcription factors (TF), protein kinase (PK), and transcriptional regulators (TR) were identified using iTAK software (Zheng et al., 2016).

## Material Screening and RNA-Sequencing

To reveal important biological processes and significant pathways involved in sunflower drought-response, and narrow down the candidate genes, RNA-seq was conducted. We screened the 226 GWAS accessions based on phenotypic evaluation results. A comprehensive drought tolerance coefficient value ( $D$ -value) was used to evaluate the drought tolerance of all accessions (Li et al., 2015). The  $D$ -value integrated the results of multi-traits measured under two watering regimes and can represent the comprehensive drought tolerance of an accession. Finally, an inbred line with the highest  $D$ -value was selected and named “K58” (Zilong et al., 2021).

The drought stress experiment was the same as GWAS. Young leaves were sampled at 0, 7, and 14 days after drought treatment. Total mRNA was isolated using the RNA prep pure plant kit DP411 (Tiangen Biotech, China) according to the instruction manual. A total of 1  $\mu$ g RNA per sample was used for cDNA library construction. Sequencing libraries were generated using NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. The quality of libraries was assessed through the Agilent Bioanalyzer 2100 system. After the quality test, all samples were sequenced in the Illumina Novaseq 6000 system, and 150-bp paired-end sequences were obtained (raw reads). Clean reads were obtained by eliminating reads containing ploy-N, reads containing adapter and low-quality reads from raw reads. The Q30, GC content of clean reads were calculated simultaneously.

## Analysis of Differentially Expressed Genes

Differentially expressed genes analysis was conducted using the HISAT2-Stringtie(merge)-DESeq2 pipeline. High-quality clean reads were aligned to the reference genome using the Hisat2 software (version 2.2.1) (Kim et al., 2015) with default parameters. In the gene count step, we used a “Transcript merge mode” via StringTie software (Pertea et al., 2015). Briefly,

the alignment files (\*.BAM) of each sample was converted to GTF file using StringTie software. Then all the GTF files were merged into one single file containing a non-redundant set of transcripts. This file was then used as a reference to recalculate the count for each gene. With this model, novel genes/transcripts can be identified that differ from the reference genome.

A python script [prepDE.py (<https://ccb.jhu.edu/software/stringtie/dl/prepDE.py>)] was used to generate a gene count matrix from the GTF file of each sample. Normalization and differential expression analysis were performed using DESeq2 R packages (Love et al., 2014). By default, DESeq2 computes a Benjamini-Hochberg adjusted  $p$ -value ( $P_{adj}$ ) to control the false discovery rate (FDR) (Anders and Huber, 2012). The “Fold Changes” of a gene is the FPKM ratio at day 7 (or 14) to that at day 0. For comparison purposes, we take the logarithm of the fold change and calculate the absolute value ( $|\log_2(\text{Fold Changes})|$ ). The  $|\log_2(\text{Fold Changes})|$  of a gene equal to 1 means that the expression level of this gene has doubled or halved. Genes with  $P_{adj} \leq 0.01$  and  $|\log_2(\text{Fold Changes})| \geq 1$  was considered as DEG.

## Enrichment Analyses of Gene Ontology and KEGG Pathways

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed to reveal the biological functions and pathways of DEGs. The sequence file of each gene was input into EggNog software (version 2.0.1) to obtain gene annotation (Huerta-Cepas et al., 2019). GO and KEGG analysis was conducted using the ClusterProfiler (version 4.0.0) R package (Yu et al., 2012). Only GO-terms or KEGG pathways with  $p$ -value  $< 0.05$  were screened for subsequent analysis. The REVIGO program (<http://revigo.irb.hr/>) was used to remove redundant GO-terms (Supek et al., 2011).

## RT-qPCR Validation

To validate RNA-seq results, reverse transcription quantitative PCR (RT-qPCR) was conducted on 6 randomly selected DEGs with three technical replicates. Experimental samples are the same as for RNA-seq. Reverse transcription was conducted using Biomarker Script II 1st Strand cDNA Synthesis Kit (Biomarker Technologies, Beijing, China) with Oligo d(T)<sub>23</sub> VN as a primer, and qPCR reactions were performed with Biomarker 2X SYBR Green Fast qPCR Mix (Biomarker Technologies, Beijing, China) on the FTC-3000 qPCR system (Funglyn Biotech Inc., Toronto, ON, Canada). Gene expression levels were calculated using the method of  $2^{-\Delta\Delta C_t}$  according to Livak and Schmittgen (Livak and Schmittgen, 2001), and standard deviation was calculated among three biological replicates. The 18S rRNA gene was used as the endogenous control (Ebrahimi Khaksefidi et al., 2015).

## Combined Analysis of GWAS and RNA-Seq

To reduce the number of candidate genes, we conducted a combined analysis. The two gene sets obtained by GWAS and RNA-seq were subjected to the

intersection operation. Genes within the intersection were considered to be important genes and were investigated in depth.

## RESULTS

### Phenotypic Variation Among Accessions

Drought stress led to different degrees of changes in all phenotypic traits (**Figure 1; Table 1**). Drought stress inhibited plant height (PH). Mean PH was 31.37 cm (ranged from 15.07 to 56.10 cm) at WW condition, whereas it was 22.23 cm (ranged from 6.4 to 38.55 cm) under DS conditions. Over 90% of the accessions (208/226) had a decrease in PH under drought stress.

Mean leaf surface area (LSA) was 46.34 cm<sup>2</sup> (ranged from 4.62 to 143.62 cm<sup>2</sup>) for the WW condition compared with 24.21 cm<sup>2</sup> (ranged from 3.73 to 65.36 cm<sup>2</sup>) for the DS condition. Over 88% (200/226) of the accessions had a decrease in LSA under drought stress.

The root-shoot ratio (RSR) increased slightly under the DS condition compared with in WW condition. Mean RSR was 0.16 (ranged from 0.05 to 0.79) under DS condition, whereas it was 0.12 (ranged from 0.02 to 0.62) under WW condition, with 71.7% (162/226) of the accessions showing an increased RSR under DS conditions. Notably, drought stress significantly increased three root-related traits, the average root length (RL), root volume (RV), and root surface area (RSA) increased by 44.1, 131, and 76.4% under DS condition compared with plants under WW condition. Among the 226 accessions, 77.4% (175/226), 83.2% (188/226), 83.2% (188/226) of them showed an increased RL, RV, and RSA under drought conditions, respectively. Drought stress has relatively little effect on the relative water content (RWC) of sunflower leaves, and the mean value was reduced from 0.74% under WW condition to 0.69% under the DS condition, with a reduction rate of 7.3%. Among 226 sunflower plants, 83.6% (189/226) had lower RWC under the DS condition. Similarly, the SPAD value was also decreased slightly in DS compared to WW, with a reduction rate of 5.7%. Mean values were 31.08 (ranged from 22.1 to 39.77) and 29.31 (ranged from 18.6 to 38.67) under WW and DS, respectively, and 72.6% (164/226) accessions showed a decreased SPAD value under DS condition.

The coefficient of variation (CV) was used to describe the variance within accessions. In this study, the CV of some traits was very high, the average CV among all traits were 40.36%, varying from 11.94 to 71.86%. It shows that our experiment materials have strong heterogeneity. RSR had the highest CV values (61.42–65.49%) while the SPAD value showed the lowest CV values (11.94–14.09%) (**Supplementary Table 1**).

The correlation between the same indicator under different conditions is shown in **Supplementary Figure 1**. The correlation coefficients of LSA and SPAD were higher than 0.6 in the WW vs. DS, while the correlation coefficients of RSA, RL, and RSR were all lower than 0.1. The correlation between different indicators under the same condition is shown in **Figure 2**. The three root-related indexes (RL, RV, and RSA) showed positive correlation under both WW and DS growth conditions. Under DS conditions, RV was positively correlated with RSA (spearman Cor. = 0.776), whereas negatively correlated with

PH (spearman Cor. = -0.59). Under WW conditions, LSA is positively correlated with SPAD with a spearman correlation coefficient of 0.61.

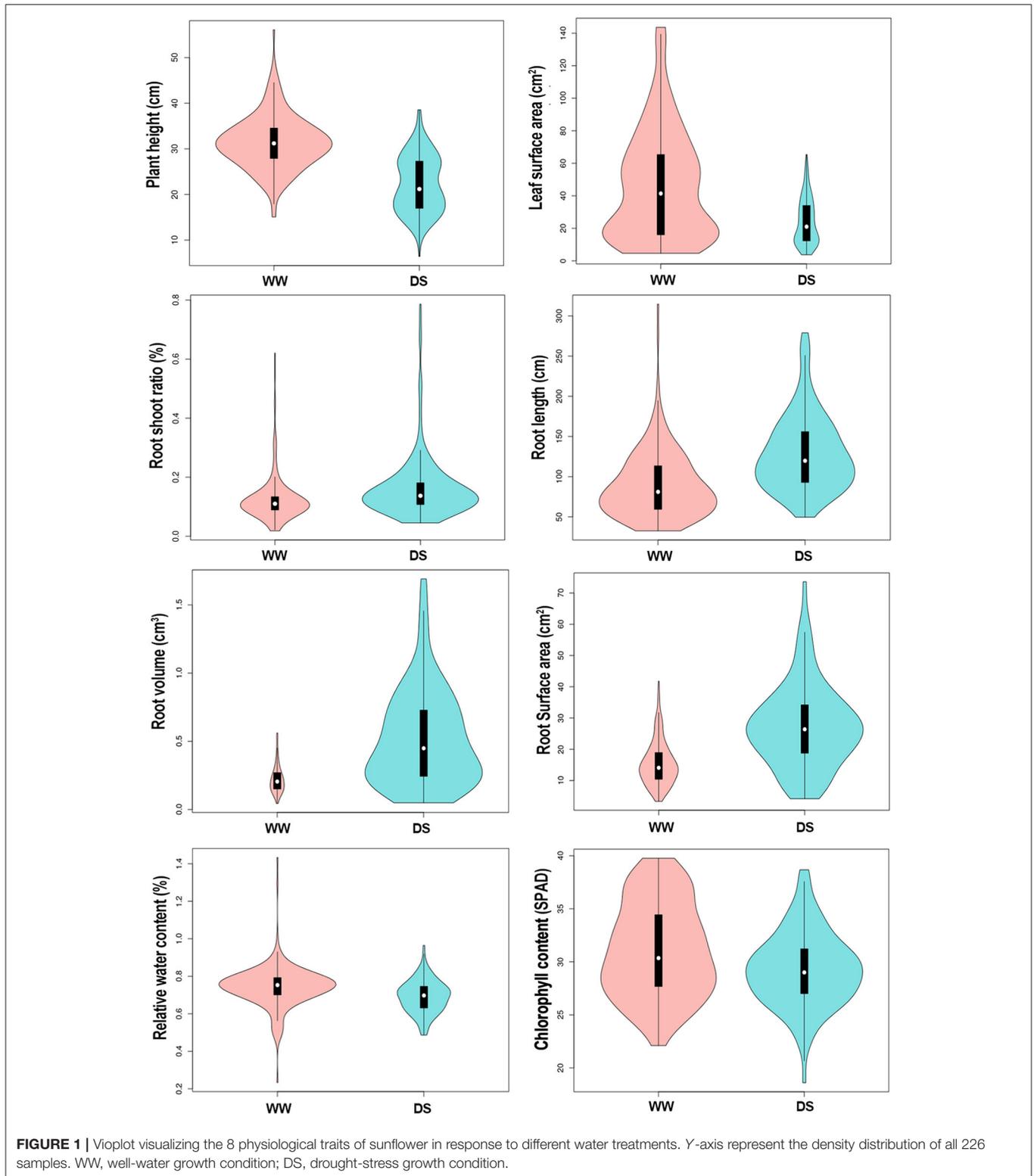
### SLAF-Sequencing, Genotyping, and Linkage Disequilibrium

Enzyme digestion efficiency is an important indicator of SLAF-seq quality. According to the results of the pre-experiment, Hae III was selected to digest the genomic DNA. The enzyme digestion efficiency of control genome *Oryza sativa* ssp. *japonica* was 94.12%, indicating the enzyme digestion reaction was normal. A total of 934.08 MB paired-end reads were obtained, with an average Q30 of 91.97% (89.04–93.44%) and a GC content of 43.67% (42.13–45.56%) (**Supplementary Table 2**). The mapping rate and the proper mapped rate were 98.20 and 90.96%, respectively (**Supplementary Table 3**).

A total of 565,668 SLAF tags were obtained, 243,291 of them were polymorphic SLAF tags. These SLAF-tags were evenly distributed on 17 chromosomes (**Figure 3; Supplementary Table 4**). SLAF tags on chromosome 13 had the highest polymorphic rate (48.25%), while chromosome 12 had the lowest polymorphic rate (38.85%). A total of 2,124,143 population SNP markers were developed *via* GATK software (**Supplementary Table 5; Figure 4**). After quality control, 94,162 high-quality SNPs were obtained for subsequent analysis (**Supplementary Table 6; Figure 5**). Chromosome 10 harbored the highest proportion of SNPs (8.68%, 8,173 of 94,162), while the shortest chromosome 6 contained the lowest proportion of SNPs (3.08%, 2,898 of 94,162). There were 31.37 SNP per 1 MB on average across 17 chromosomes. Chromosome 10 had the highest SNPs/Mb ratio (47.68 SNPs per Mb), while chromosome 6 had the lowest SNPs/Mb ratio (19.56 SNPs per Mb) (**Supplementary Table 6**). LD was estimated as the  $r^2$  value,  $r^2$  ranged from 0.135 on chromosome 6 to 0.218 on chromosome 10, with an average of 0.174, revealing differences in the level of LD among chromosomes (**Supplementary Table 7**). The average distance of LD decay was about 20 kb (**Figure 6**).

### Genetic Diversity and Population Structure

Divergence of the 226 accessions during evolution was the major factor leading to high rates of false positive errors in GWAS analysis (Yu and Buckler, 2006). The admixture software was used to analyze the population structure, and the CV for  $K = 1-13$  was examined. The results showed that when  $K = 11$ , the CV dropped to the lowest value (0.659), suggesting the entire population most likely originated from 11 ancestors (**Figures 7, 8A**). The phylogenetic tree has divided the accessions into 7 main clusters with identical tree topologies (**Figure 8B**). PCA analysis revealed that all the 11 principal components had eigenvalues of over 1, and the first 8 principal components can explain 85.73% of the total variance. The first three principal components PC1 (with variance explain 15.71%), PC2 (with variance explain 13.55%), and PC3 (with variance explain 11.77%) were displayed in **Figure 8C**. All these results showed that our experimental materials are highly heterogeneous and is ideal for GWAS analysis.



## Genome-Wide Association Analysis

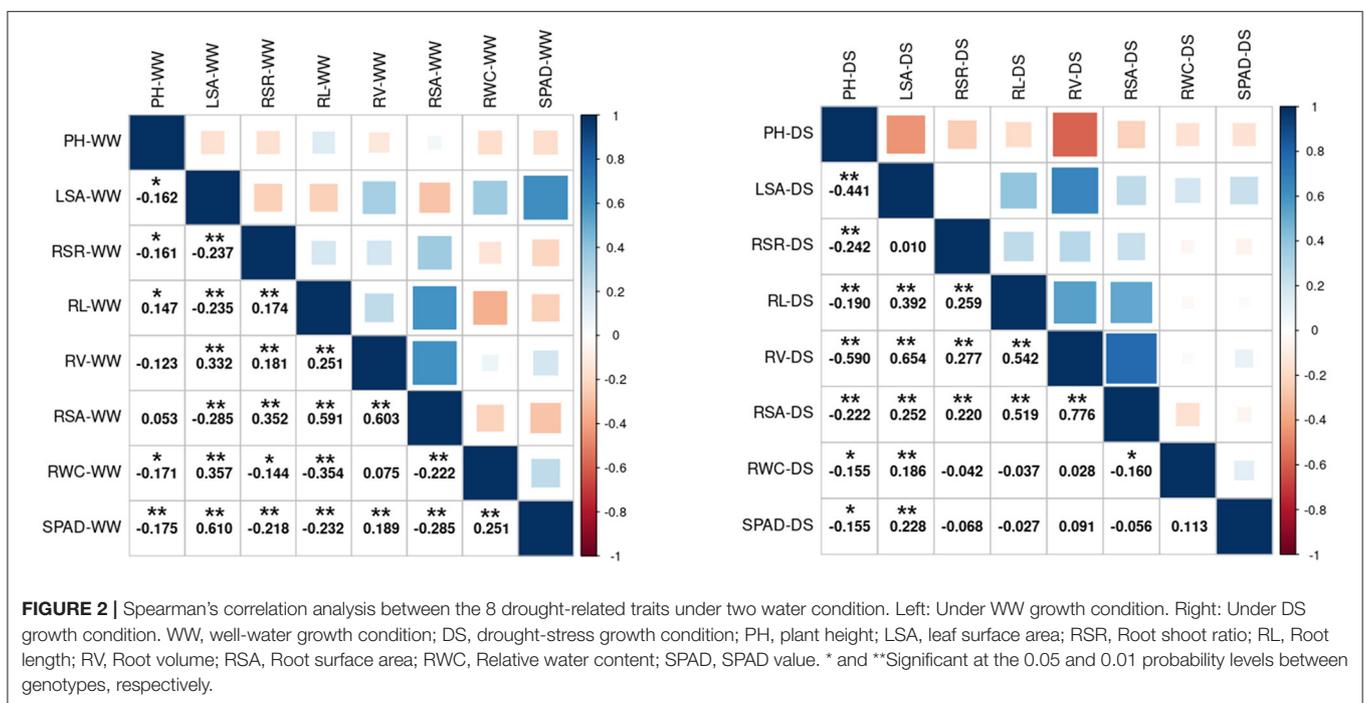
The GWAS was performed on 8 traits using 3 methods (MLM, FarmCPU, BLINK). A total of 80 SNPs were detected under

the significance threshold of  $p < 1.062 \times 10^{-6}$ . Among them, 59 were obtained by STI, and 22 were obtained by SSI, and there was only one common SNP between the two indicators

**TABLE 1** | Descriptive statistics values for traits of 226 sunflowers under drought stress.

Traits	Trt.	Min.	Max.	Mean	SD.	CV. (%)	Skewness	Kurtosis
Plant height	WW	15.07	56.10	31.37	5.80	18.48	0.50	1.56
	DS	6.40	38.55	22.23	6.16	27.72	0.25	-0.65
Leaf surface area	WW	4.62	143.62	46.34	33.30	71.85	0.91	0.30
	DS	3.73	65.36	24.21	14.03	57.93	0.71	-0.31
Root shoot ratio	WW	0.02	0.62	0.12	0.08	61.42	3.24	14.85
	DS	0.05	0.79	0.16	0.11	65.49	3.21	13.05
Root length	WW	32.56	314.68	89.47	40.83	45.63	1.63	5.11
	DS	49.66	279.06	128.90	46.76	36.28	0.96	0.86
Root volume	WW	0.04	0.56	0.22	0.11	46.99	1.21	1.53
	DS	0.05	1.69	0.52	0.35	67.17	1.01	0.88
Root surface area	WW	3.33	41.79	15.50	7.04	45.43	1.00	1.03
	DS	4.22	73.59	27.34	13.19	48.23	0.74	0.88
Relative water content	WW	0.23	1.43	0.74	0.11	14.59	0.84	11.26
	DS	0.49	0.96	0.69	0.09	12.54	0.10	0.19
SPAD	WW	22.10	39.77	31.08	4.38	14.09	0.19	-0.97
	DS	18.60	38.67	29.31	3.50	11.93	0.18	0.42

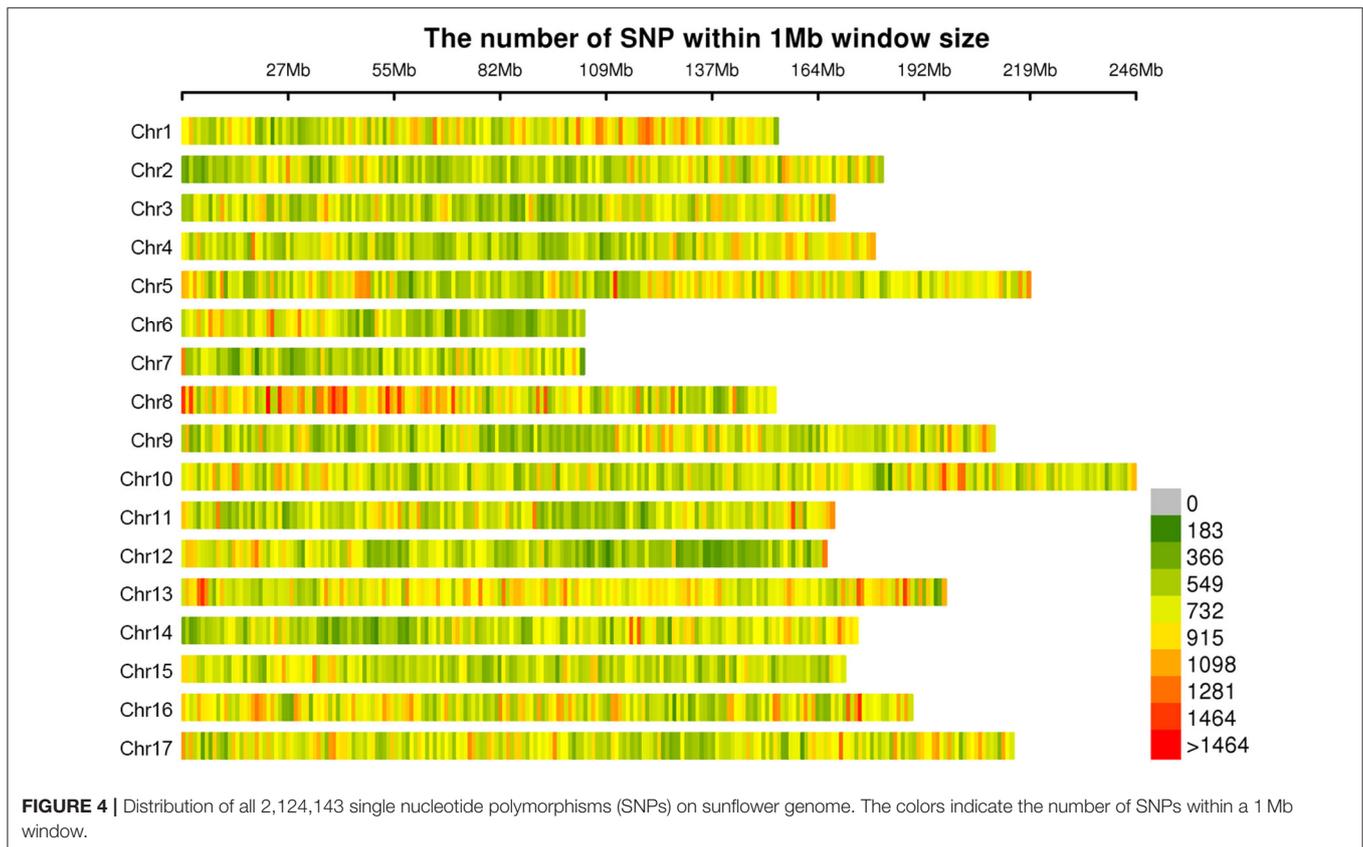
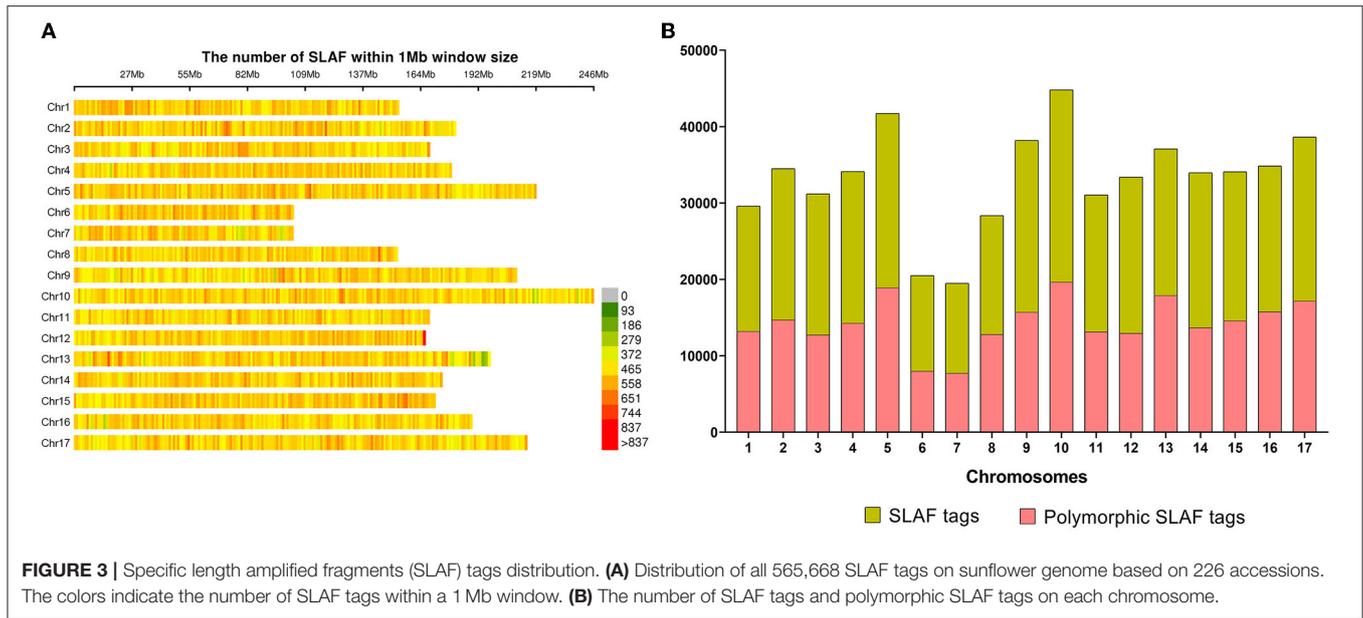
Trt., Treatment; Min., Minimum; Max., Maximum; SD, Standard deviation; CV, Coefficient of variance; CK, Well water condition; DS, Drought stress condition.



(Supplementary Figures 2, 3; Supplementary Table 8). A total of 19, 44, and 33 SNPs were discovered by MLM, FarmCPU, and BLINK methods, respectively. For 8 phenotypic traits, LSA detected the most associated SNPs (27), followed by RWC detected 13, SPAD, RSR, and PH detected 12, 11, and 11, respectively. RL, RV, and RSA were detected 2, 4, and 3 SNPs, respectively. A total of 118 genes were found within the 20 kb of 80 significant SNPs, 85 of them were protein-coding genes (Supplementary Table 9).

### RNA-Sequencing and Expression Analysis

A total of 70 Gb clean data were obtained after filtering and quality control. The Q30 of each library ranged from 93.57 to 94.97%, and the GC content ranged from 44.86 to 45.68% (Supplementary Table 10). A total of 18,922 DEGs were obtained (Supplementary Table 11), 6,698 of them were newly discovered. In general, there were more DEGs under 14 days of drought stress compared with the 7 days, and down-regulated DEGs were more than

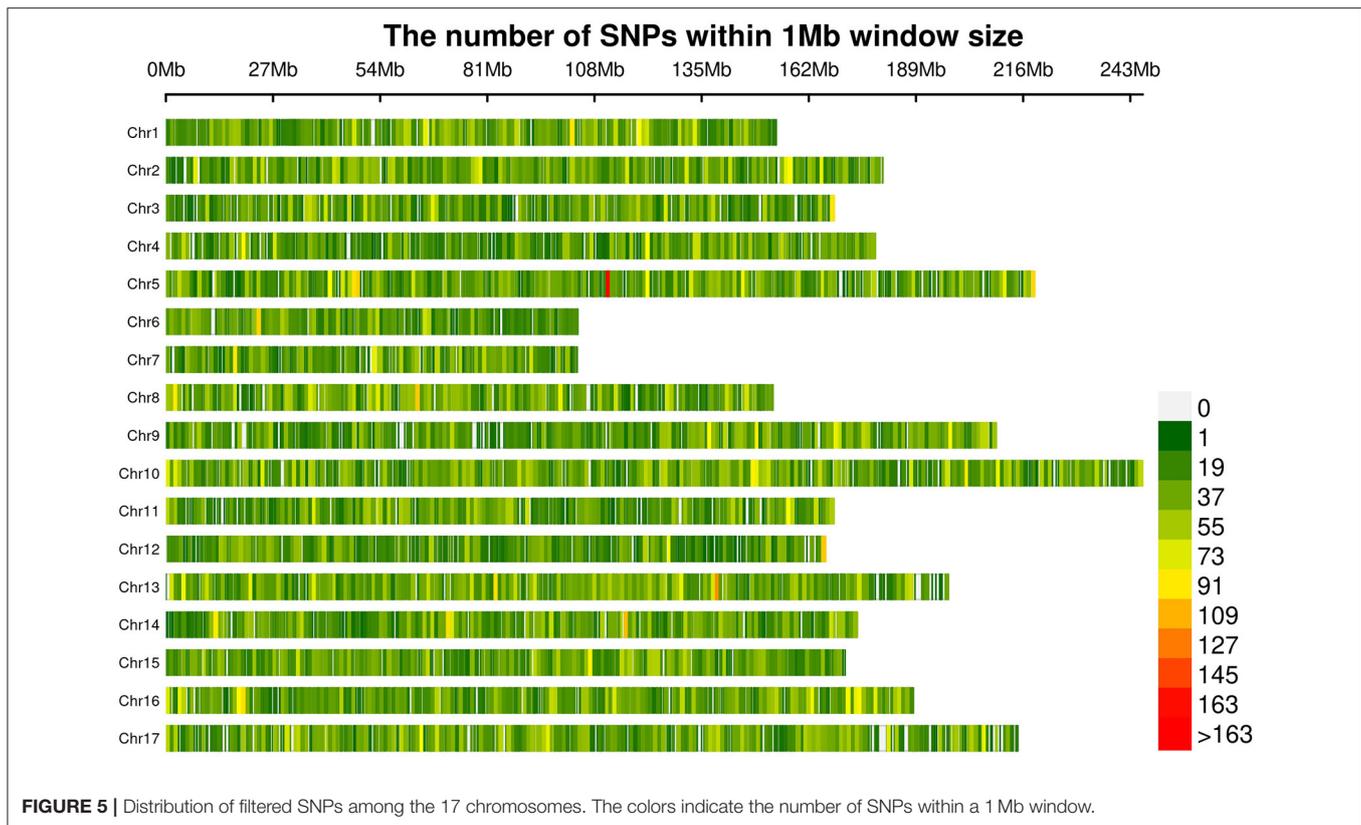


up-regulated DEGs (Figure 9). From day-7 to day-14, the up-regulated DEGs were increasing from 3,848 to 7,174, whereas the down-regulated DEGs were increasing from 5,201 to 8,521, respectively.

## Enrichment Analysis

### GO Analysis

The up-regulated genes were enriched in 46, 90 GO-terms at 7, 14 days. On day-7, the most significant GO-terms were



cellular amino acid catabolic process (GO:0009063), branched-chain amino acid catabolic process (GO:0009083), and seed maturation (GO:0010431). On day-14, the most significant GO-terms were leaf senescence (GO:0010150), aging (GO:0007568), and carboxylic acid catabolic process (GO:0046395). For down-regulated genes, there were 127, and 199 GO-terms enriched at 7, 14 days. On day-7, the most significant GO-terms were cellular polysaccharide metabolic process (GO:0044264), cell wall biogenesis (GO:0042546), and photosynthesis, light reaction (GO:0019684); At day-14, the most significant GO-terms were photosynthesis (GO:0015979), photosynthesis, light reaction (GO:0019684), and plastid organization (GO:0009657) (**Supplementary Figure 4**).

### KEGG Analysis

Up-regulated genes were enriched in 13 and 48 significant KEGG pathways at 7 and 14 days. On day-7, the most significant pathways were Valine, leucine and isoleucine degradation, MAPK signaling pathway—plant, and FoxO signaling pathway; On day-14, the most significant pathways were valine, leucine, and isoleucine degradation, MAPK signaling pathway—plant, and longevity regulating pathway. For down-regulated genes, there were 36, 48 significant KEGG pathways enriched at 7, 14 days. On day-7 and day-14, the most significant KEGG pathways were both related to photosynthesis, such as photosynthesis proteins (BR:ko00194), photosynthesis-antenna proteins, and photosynthesis (**Supplementary Figure 5**).

### RT-qPCR Validation

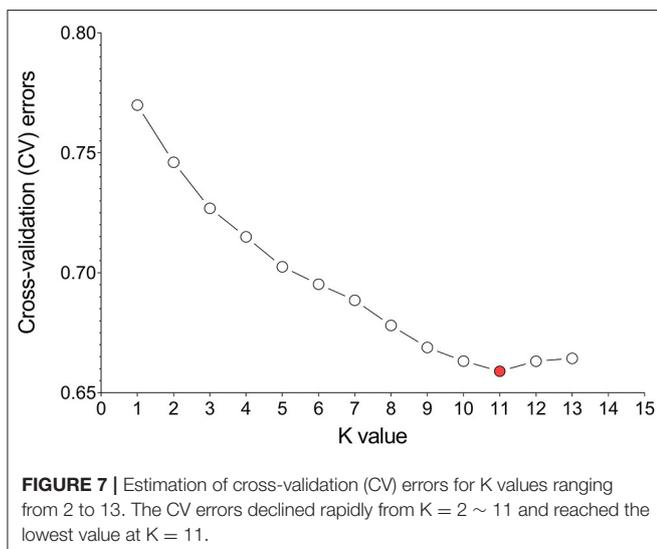
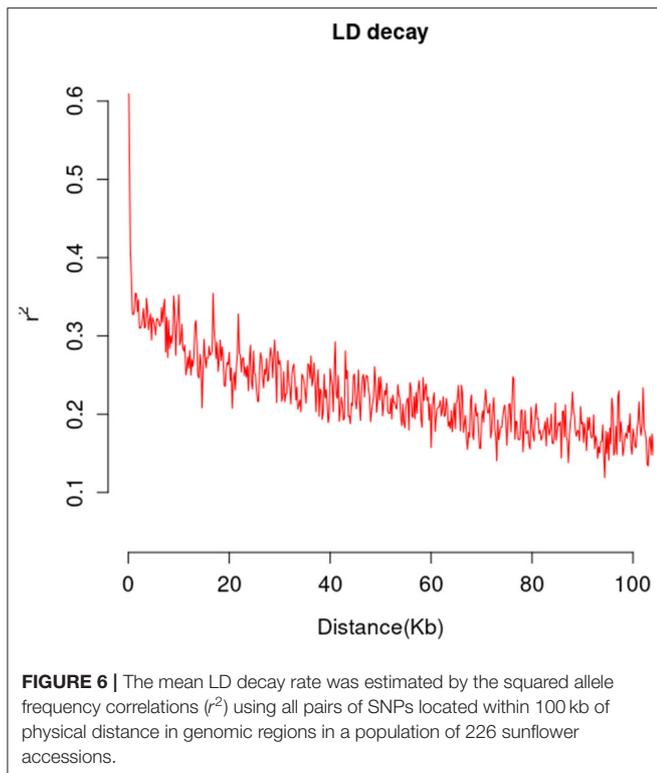
To validate the accuracy of RNA-seq, RT-qPCR was performed. Six genes were randomly selected from all DEGs. The primer sequence was shown in **Supplementary Table 12**. Correlation analysis showed that RNA-seq was closely related to RT-qPCR results. The correlation coefficient ( $R^2$ ) was 0.8167, endorsing our RNA-seq data were reliable (**Figure 10; Supplementary Figure 6**).

### Candidate Genes Identification

By integrating the results of GWAS and RNA-seq analysis, a total of 18 common genes were obtained, 14 of them were protein-coding genes (**Table 2; Figure 11**). These genes are distributed on chromosomes 4, 5, 8, 9, 10, 11, 12, 13, 16, and 17. Two genes are associated with both LSA and PH. One gene is associated with both LSA and SPAD. Their details are as follows.

### Candidate Genes Associated With Plant Height

There were 2 candidate genes that were screened using combined analysis. Both of them were located on chromosome 13. The LOC110899235 gene encoding “inosine-uridine preferred nuclear hydrate” is homologous to the AT5G18860.2 gene in *Arabidopsis thaliana*. Another LOC110899238 gene encoding “ABC transporter c family member 3-like” is homologous to the AT3G13080.1 gene in *Arabidopsis thaliana*. Both two genes were down-regulated with the extension of drought stress time in K58.



### Candidate Genes Associated With Leaf Surface Area

There were 8 common candidate genes associated with LSA, 2 of which were also associated with PH. The function of the gene LOC10936334 located on chromosome 4 was annotated as “Jacalin-like lectin domain”, which is homologous to the AT1G73040.1 gene in *Arabidopsis thaliana*, and its expression level continues to decrease under drought stress in K58. Gene LOC110941963 located on chromosome 5 was annotated as “microtubule-associated protein RP EB family member”, which

was homologous to the AT3G47690.1 gene in *Arabidopsis thaliana*. This gene was down-regulated after 14 days of drought stress in K58. At 19.52 kb upstream of an SNP (S10\_123892851) on chromosome 10, a gene (LOC110885273) encoding “Serine threonine-protein kinase” was identified. It is worth noting that the gene was also associated with SPAD. This gene belongs to the protein kinase family of RLK-Pelle\_SD-2b, and is homologous to the *Arabidopsis* AT4G32300.1 gene. RNA-seq showed it was down-regulated with the extension of drought stress in K58. Gene LOC110894816 encoding “Equilibrative nucleotide transporter” were down-regulated at 7, 14 days in K58, which is homologous to AT1G70330.1 in *Arabidopsis thaliana*. Gene LOC110920644 belongs to the PLATZ transcription factor family. It was up-regulated at 7 days and down-regulated at 14 days of drought stress in K58. Gene LOC110891369 encoding “receptor-like protein kinase” was sharply up-regulated at 14 days. This protein kinase belongs to the RLK-Pelle\_SD-2b RLK-Pelle\_CrRLK1L-1 protein kinase family.

### Candidate Genes Associated With Root-Shoot Ratio

There were 2 candidate genes obtained by combined analysis. One gene LOC110937937 encoding “Component of the peroxisomal and mitochondrial division machineries” was up-regulated at 14 days post drought stress, another gene LOC110915715 encoding “Protein of unknown function (DUF1666)” were continuously down-regulated with the drought stress.

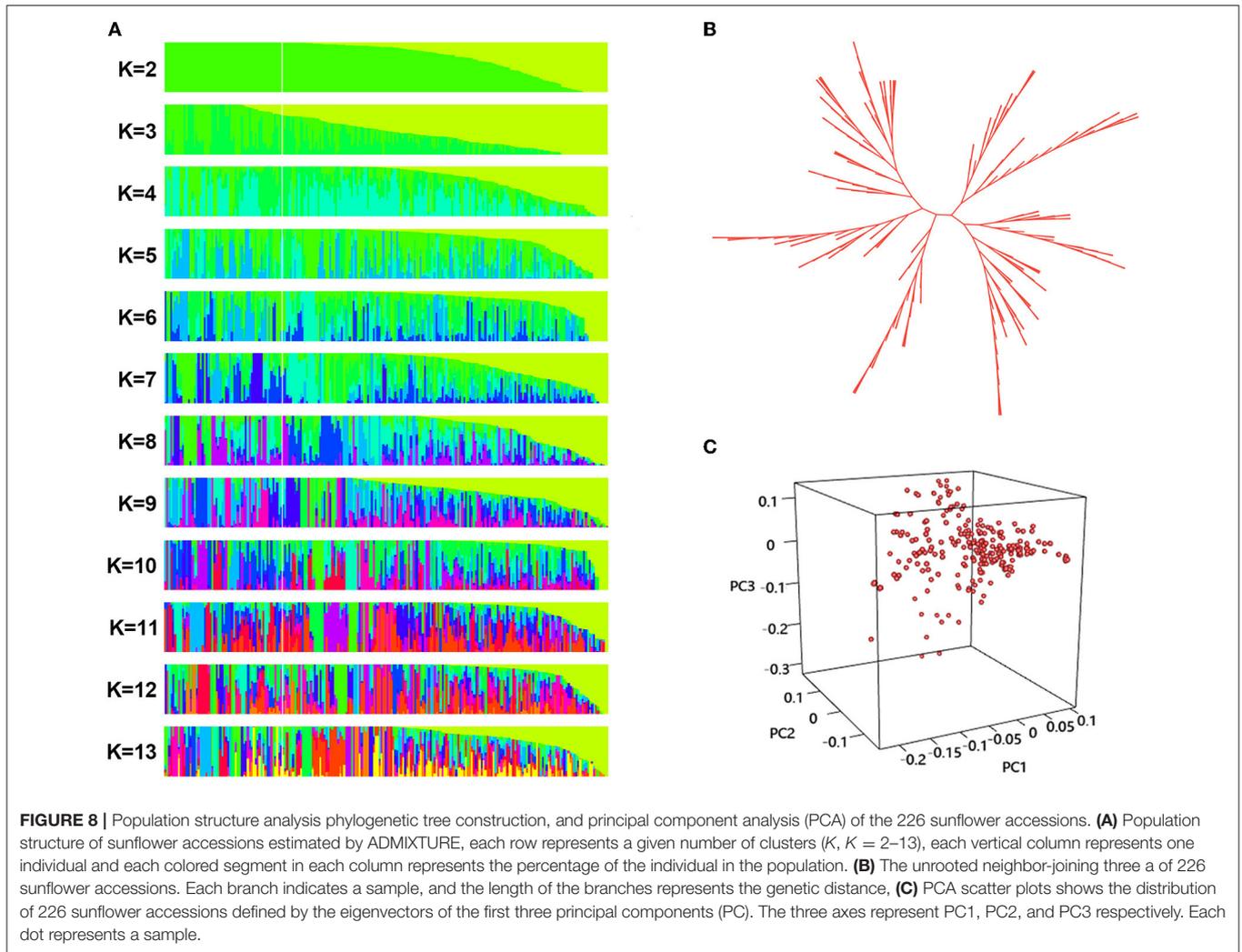
### Candidate Genes Associated With Three Root Related Traits

Notably, there are relatively fewer SNPs related to three root traits (RL, RV, and RSA). No genes were found within the 20 kb region of RL associated SNPs. The combined analysis identified 2 genes associated with RV and 1 gene associated with RSA. For RV, gene LOC110877324 on chromosome 9 was annotated as “Belongs to the UDP-glycosyl transferase family”, which was down-regulated in K58 after 14 days of drought stress. Another gene (LOC110917707) located on chromosome 16 was annotated as “domain presence in VPS-27, Hrs and Stam”, which was up-regulated in K58 after 14 days of drought stress. These two genes are homologous to the AT2G18570.1 gene and AT2G38410.1 gene in *Arabidopsis thaliana*, respectively.

For RSA, gene LOC110872899 was located on chromosome 8, and annotated as “Inactive leucine-rich repeat receptor-like serine threonine-protein kinase”. This gene is homologous to the *Arabidopsis* AT1G10850.1 gene. It was slightly up-regulated in K58 at 7 days and then sharply down-regulated at 14 days of drought stress.

### Candidate Genes Associated With Relative Water Content

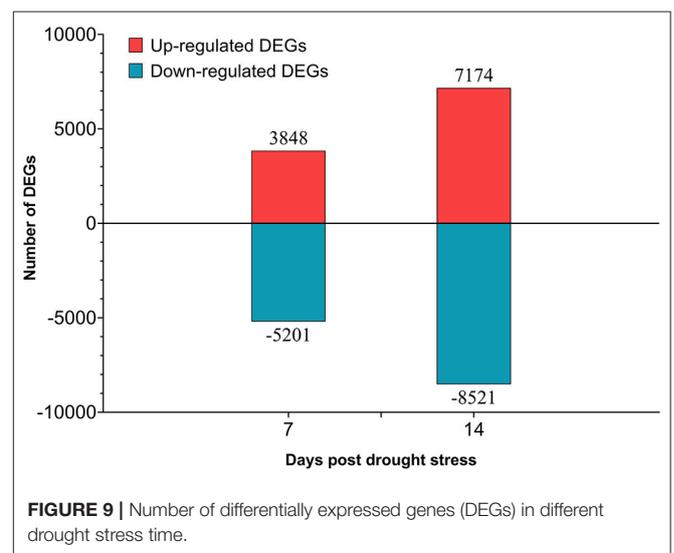
LOC110941862 is the unique gene screened by the combined analysis. This gene encodes the “Topless-related protein”, which is homologous to the AT1G15750.3 gene in *Arabidopsis thaliana*. RNA-seq results showed that this gene was continuously down-regulated in K58 under drought stress.



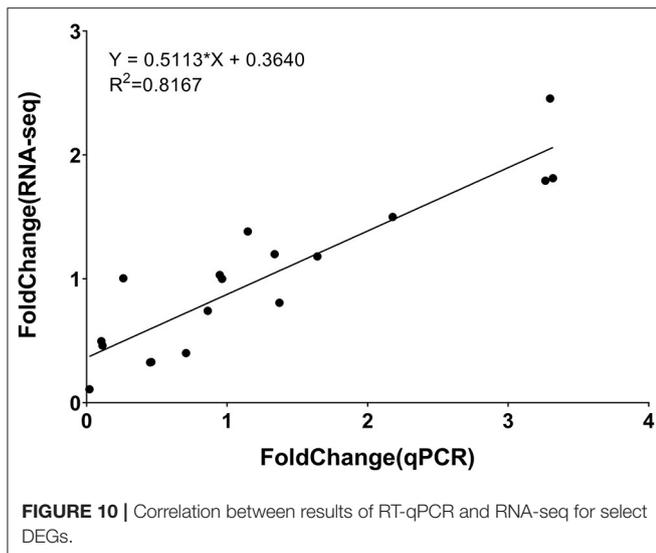
## DISCUSSION

Global climate change threatens crop production worldwide. Plants adopt diverse strategies to combat drought stress such as reducing the stomatal conductance, decreased photosynthetic rate, accumulation of different osmoprotectants, activation of stress-responsive genes and transcription factors, etc. (Farooq et al., 2009; Kaur and Asthir, 2017). Drought resistance is a complex quantitative trait. One difficulty in drought-tolerant genetic breeding is the unequivocal evaluation of plant response to soil-water deficits (Pereyra-Irujo et al., 2007). Based on the previous research, we evaluated 8 phenotypic traits among 226 accessions under WW and DS conditions. Compared to the WW condition, the average PH, LSA, RWC, and SPAD value were decreased, while RSR and three root related traits (RL, RV, RSA) were increased under the DS condition.

It has long been known that drought stress at the vegetative stage impedes phenotypic traits like PH, LSA, whereas an increase in RL at the expense of above-ground dry matter occurs resulting in higher RSR (Petcu et al., 2001; Hussain et al., 2010; Javaid et al., 2015). In our results, the change trends of mean PH, RL, RSR, and LSA were consistent with previous



studies. However, the mean RV increased under drought, which was not consistent with a previous study. Geetha et al. (2012) found that the RV decreased by 40.2% under drought stress



among 29 sunflower varieties, while we found 83% of accessions have an increase in RV. This may be due to differences in the genotypes of the study materials. Different genotypes of plants have different adaptability to drought stress (Petcu et al., 2001). Even in the most consistent trend of variation in PH (92% decreased under drought stress), there were still 16 accessions increased under drought stress. These specific materials may include important drought-tolerance genes and will be good sources for our drought tolerance molecular breeding. In some previous studies, the relationship between SPAD and chlorophyll content per unit leaf area is fitted as linear regression. SPAD value is often used to represent chlorophyll content (Costa et al., 2001; Martínez and Guiamet, 2004). Our results show that under WW growth conditions, SPAD value is positively correlated with LSA. It demonstrates that a larger LSA has more chlorophyll, which increases the photosynthetic rate (Espina et al., 2018). The correlation coefficients of LSA and SPAD in WW vs. DS conditions were higher than 0.6, indicating that drought affects these two traits more by environment than by genotype. The correlation coefficients of RSA, RL and RL were very low, indicating that they were more influenced by genotype.

Studies have shown that the genetic relatedness of the mapping population can increase the false positive risk of GWAS results (Ali et al., 2020). A population with enough genotype and trait diversity is considered to be the expected GWAS population (Flint-Garcia et al., 2005). In this study, the population panel consisting of 226 accessions were collected from different ecological regions. Three population structure analysis methods (admixture, phylogenetic, and PCA) were conducted. Results showed that 226 sunflower materials had large genetic differences and were an ideal GWAS population. Linkage disequilibrium (LD) is the basis of GWAS (Ali et al., 2020). When LD declines rapidly with distance, LD mapping is potentially very precise (Gaut and Long, 2003). Since our materials have high genetic variability, the LD-decay distance is about 20 kb. Overall patterns of LD decay show chromosome specificity. Chr10 showed the highest LD value, followed by Chr7, Chr5,

Chr13, and Chr17. This result is consistent with a previous study conducted by Filippi et al. (2020). They have reported different patterns of LD across chromosomes, with Chr10, Chr17, Chr5, and Chr2 showing the highest LD. The extended LD in Chr10 and Chr5 were also reported by other researchers (Cadic et al., 2013; Mandel et al., 2013). Owens et al. showed that the extended LD on Chr10 could be the result of the wild introgression in the fertility restoring male lines (Owens et al., 2019).

GWAS methods have evolved over years. Several new methods are being developed to improve the statistical power and reduce the computational time. FarmCPU uses a set of markers associated with a casual gene as a co-factor instead of kinship to avoid overfitting and eliminate confounding between kinship and testing markers iteratively (Liu et al., 2016). More recently, along with improvements in statistical power and reduction in computing time compared to FarmCPU, the new method called BLINK is set to eliminate FarmCPU requirement that quantitative trait nucleotides (QTNs) are evenly distributed in the genome (Huang et al., 2019). In the present study, we used 3 methods simultaneously to conduct GWAS. The FarmCPU method detected 44 SNPs, the BLINK method detected 33 SNPs, and the MLM method detected the lowest of 19 SNPs, respectively. There were 12 SNPs found simultaneously by FarmCPU and BLINK method, and only 3 common SNPs were found by 3 methods. Most SNPs were only found in one method. Therefore, it may be prudent to use multiple methods to conduct a GWAS survey (Nida et al., 2021).

STI and SSI are two commonly used evaluation indexes in the study of plant abiotic stress. According to the research of Mehdi GHAFARI, STI is more efficient for identifying drought-resistant lines, and SSI is more efficient for identifying drought-sensitive lines (Ghaffari et al., 2012). Applying both indicators simultaneously could provide a complete and accurate assessment of drought tolerance. Strangely, the calculation methods of STI in different articles are inconsistent (Sukumaran et al., 2018; Khanzada et al., 2020; Chaurasia et al., 2021). In the present study, we carefully chose a scientific STI calculation method for GWAS analysis. A total of 80 significant SNP markers associated with 8 phenotypic traits were detected, 22 of them were detected using SSI, and 59 of them were detected using STI, only one common SNP was detected by both of the two indexes.

To further understand the biological processes, pathways, and gene expression patterns in sunflowers under drought stress, we conducted an RNA-seq analysis. Based on the phenotypic traits, a drought-tolerant plant was selected from the GWAS population. We sampled the leaves at 0, 7, and 14 days after drought stress. A total of 18,922 differentially expressed genes were obtained.

There was a noticeable consistency between the results of GO and KEGG analysis. For example, up-regulated genes were enriched in GO-terms such as cellular amino acid catabolic process (GO:0009063), branched-chain amino acid catabolic process (GO:0009083), while KEGG analysis showed “Valine, leucine and isoleucine degradation” was the most significant pathway. Down-regulated genes were enriched in photosynthesis (GO:0015979), photosynthesis, light reaction (GO:0019684) according to GO analysis, while KEGG analysis showed down-regulated genes enriched in pathways such as

**TABLE 2** | Detail information of 14 genes obtained by combine-analysis of GWAS and RNA-seq.

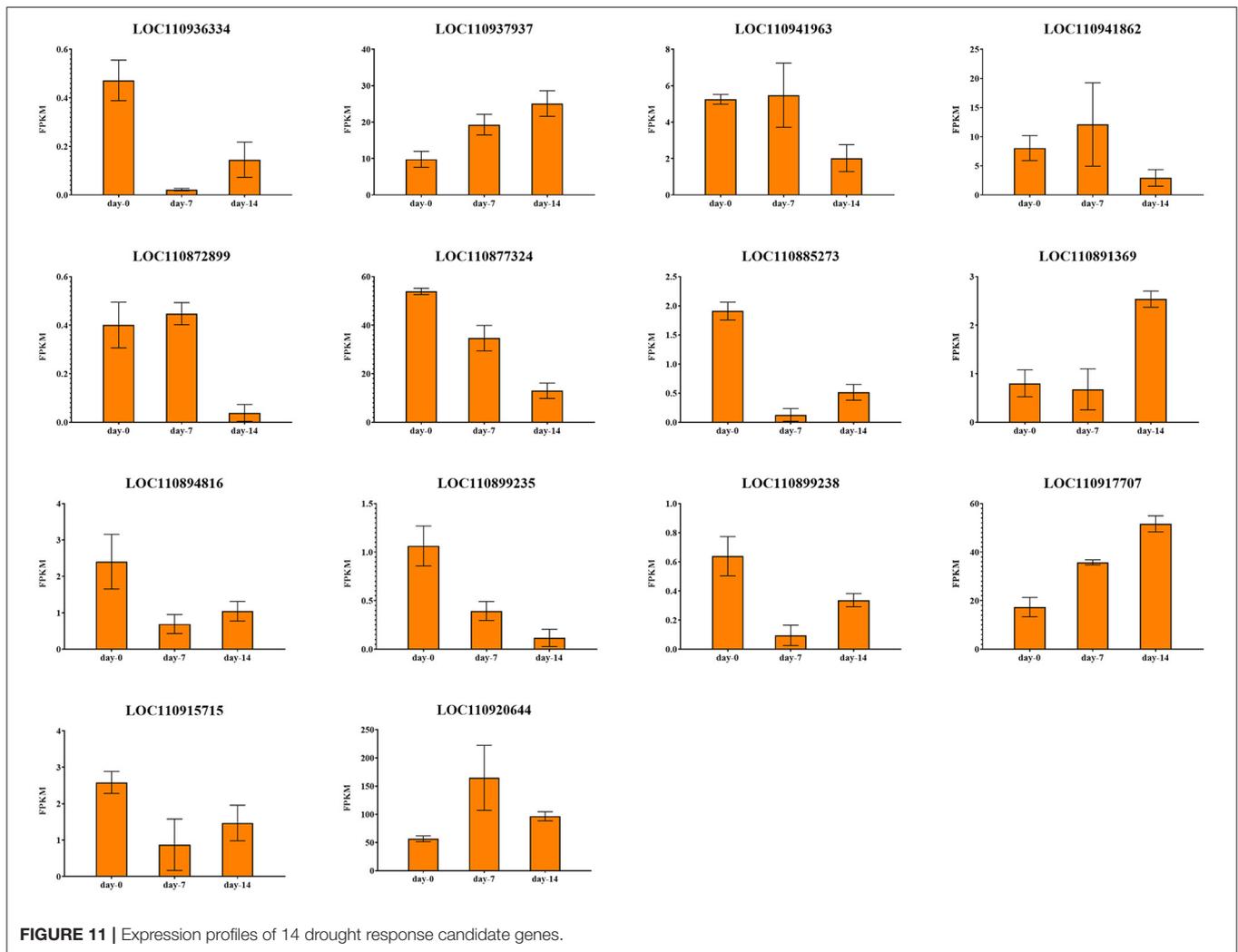
Traits	Gene name	Chromosome	Gene_start	Gene_end	Description	iTak	Families
(PH/LSA)-STI	LOC110899235	13	138759411	138769673	Inosine-uridine preferring nucleoside hydrolase		
	LOC110899238	13	138795923	138801286	ABC transporter C family member 3-like		
(LSA/SPAD)-(SSI/STI)	LOC110885273	10	123870286	123873341	Serine threonine-protein kinase	PK	RLK-Pelle_SD-2b
LSA-SSI	LOC110894816	12	55570908	55573165	Equilibrative nucleotide transporter		
	LOC110936334	4	60470023	60472767	Jacalin-like lectin domain		
	LOC110941963	5	200316650	200319863	Microtubule-associated protein RP EB family member		
	LOC110891369	11	160106964	160111888	Receptor-like protein kinase	PK	RLK-Pelle_CrRLK1L-1
	LOC110920644	17	8881157	8883452	PLATZ transcription factor	TF	PLATZ
RSR-SSI	LOC110937937	4	169924932	169927583	Component of the peroxisomal and mitochondrial division machineries. Plays a role in promoting the fission of mitochondria and peroxisomes		
	LOC110915715	16	39633096	39638580	Protein of unknown function (DUF1666)		
RV-STI	LOC110877324	9	29925998	29927713	Belongs to the UDP-glycosyltransferase family		
	LOC110917707	16	74795170	74799439	Domain present in VPS-27, Hrs and STAM		
RSA-STI	LOC110872899	8	68366663	68376805	Inactive leucine-rich repeat receptor-like serine threonine-protein kinase	PK	RLK-Pelle_LRR-III
RWC-SSI	LOC110941862	5	195719872	195730082	Topless-related protein		

The content in brackets indicates simultaneous, for example, (PH/LSA)-STI, indicating that this gene is recognized by both PH-STI and LSA-STI.

Photosynthesis proteins (BR:ko00194), Photosynthesis—antenna proteins, Photosynthesis. The branched-chain amino acids (BCAAs), including isoleucine, leucine, and valine, are essential for plants (Binder et al., 2007). Pires et al. (2016) results highlight that catabolism of BCAA appears to play an important role in the mechanism of tolerance to short-term drought, most likely by delaying the onset of stress. Our results also proved that the degradation of BCAA may be an important mechanism of sunflower drought resistance. Abiotic stress damage the thylakoid membrane, disturb its functions, and ultimately decrease photosynthesis. Down-regulated expression of photosynthesis-related genes under drought stress has been reported in several plants, such as *Arabidopsis* (Bechtold et al., 2016; Bouzid et al., 2019), wheat (Derakhshani et al., 2020), and grapevines (Franck et al., 2020). In a previous study, Escalante et al. found a down-regulation of photosynthesis-related genes in the aerial part of sunflowers (Moschen et al., 2017). However, another study revealed that the expression levels of photosynthesis-related genes were increased under drought stress in sunflowers (Escalante et al., 2020). This difference may be caused by differences in drought intensity and genotype, and our results were identical with the former.

With the development of high-throughput technologies, omics research is also undergoing a shift from a single-omics to a large-scale multi-omics approach (Liu et al., 2020). Through the multi-omics approach, researchers can obtain a deeper understanding of the fundamental biological processes, a more accurate prediction of the response variable, and gain further insight into mechanistic aspects of the system (Cavill et al., 2015). By integrating the transcriptome and metabolome, Sebastián Moschen et al. (2017) gained a deeper insight into the sunflower drought-response mechanism. The integration of genomic and transcriptomic analysis has also been reported in many recent studies. This approach can be used as an effective way to identify candidate genes. For example, eight salt stress-related candidate genes were identified by a combination of GWAS analysis and transcriptome analysis in Alfalfa (*Medicago sativa* L.) (He et al., 2021). Seven candidate genes for seminal root length in maize (*Zea mays* L.) were identified by integrating the results of the GWAS, the common DEGs, and the co-expression network analysis (Guo et al., 2020). Using a combined analysis, we identified 18 common genes.

The total genes in the sunflower reference genome were 81,496, and we found 18,922 DEGs via RNA-seq. According to this proportion, we should find at least 29 DEGs among the 118



genes of GWAS. However, the number of common genes that we have found was relatively small (18). This is because among the 18,922 DEGs, only 12,124 of them exist in the reference genome and the rest are novel genes. A subsequent chi-square test using this number found no significant difference between the two proportions ( $P = 0.908$ ). Nonetheless, the proportion of overlapped genes was still lower than we expected. The reason we speculate is that GWAS candidate genes are mainly regulatory genes that act in all accessions. A slight regulation of expression level under drought stress, which did not reach the threshold of significant difference, can affect the physiological processes in plants, whereas the DEGs of RNA-seq are mainly a series of drought-responsive functional genes that are regulated in K58 under drought stress. The difference in the class and function of the genes from these two gene sets results in a low percentage of overlapping genes. Of course, this needs further confirmation.

Among these 18 genes, 14 are protein-coding genes, of which 3 are encoding PK and 1 encodes TF. These genes may play an important role in drought response in sunflowers.

The LOC110885273 gene encodes G-type lectin S-receptor-like serine/threonine-protein kinase (LecRLKs). The protein kinase is involved in plant responses to biotic and abiotic stresses (Bonaventure, 2011; Singh et al., 2012; Zhao et al., 2016). Overexpression of G-type LecRLKs enhances the drought tolerance of *Arabidopsis thaliana* (Sun et al., 2013a), which may be achieved by controlling stomata size through interaction with abscisic acid (ABA) (Arnaud et al., 2012). Pan et al. (2020) identified a LecRLKs gene OsESG1 in rice and found it could be induced by treating with PEG, NaCl, and ABA. However, we found the LOC110885273 gene was down-regulated under drought stress, which may lead to the decrease of SPAD value under drought stress.

The receptor like kinase (RLKs) family has been defined as the most abundant gene family in *Arabidopsis*. Leucine rich repeat-RLKs (LRRRLKs) are the largest group of receptor-kinases in *Arabidopsis*, which is widely involved in responses to various biotic and abiotic stresses (Diévar and Clark, 2003; Lehti-Shiu et al., 2009). Osakabe et al. (2005) found that an LRRRLKs gene (RPK1) is involved in the early steps in the

ABA signaling pathway through a gene knock-out experiment. The overexpression of receptor-like kinase rich in the Leucine Repetition gene improves the *Arabidopsis thaliana* drought resistance (Xing et al., 2011). Receptor-like cytoplasmic kinase GUDK and OsSIK1 were shown to enhance drought tolerance in rice (Ouyang et al., 2010; Harb et al., 2020). In the present study, a down-regulated LRRRLKs gene LOC110872899 was identified, which is located at chromosome 8, and associated with RSA, maybe the mechanism of this gene in sunflower drought tolerance response is different. Another receptor-like protein kinase gene LOC110891369 was up-regulated at 14-days of drought stress in K58, which belongs to the family of RLK-Pelle\_CrRLK1L-1, and is associated with LSA.

PLATZ transcription factors play important roles in plant growth, development, and biotic and abiotic stress responses. Liu et al. (2021) reveal that PLATZ4 interacts with AITR6 to increase ABA sensitivity and drought tolerance in *Arabidopsis* by regulating the expression of different genes. Zenda et al. (2019) identified a PLATZ gene (Zm00001d051511) in maize. It was up-regulated in tolerant line YE8112, whilst down-regulated in drought-sensitive line after drought stress. This result indicated the TF genes could be the key contributors to drought stress tolerance in the drought-tolerant maize inbred line. This different expression pattern was also proved in Ray's research on rice (Ray et al., 2011), PLATZ (LOC\_Os10g42410) gene was down-regulated in panicle, while up-regulated in vegetative tissues under drought stress. Even in the same tissue at the same time, it was found that the expression levels of two PLATZ genes were up-regulated and down-regulated, respectively, which indicated the complexity of drought stress regulation. In this study, a PLATZ gene LOC110920644, which is related with LSA, was up-regulated at the early stage in K58 under drought stress.

ABA is an important hormone for plant drought response (Zotova et al., 2018). The cell ABA level increases under drought stress, leading to stomatal closure and active several stress-responsive genes (Cutler et al., 2010). Drought stress increased ABA levels in sunflowers have been reported (Robertson et al., 1985). In this study, the functions of the four TF/PK genes are all related to ABA, indicating the important role of the ABA-dependent process in the drought response of sunflowers.

## CONCLUSION

Sunflower is one of the most important oil crops in the world, which is often grown as a rainfed crop. Water limitation at the seedling stage can severely reduce stand establishment and negatively impact yields. However, the molecular mechanism underlying drought resistance is still not fully understood. In this study, we used SLAF-seq to perform GWAS for 8 important phenotypic traits in 226 sunflower inbred lines. Using three methods (i.e., MLM, FarmCPU, and BLINK) for sunflower grown in two conditions (i.e., well-water and drought stress), we identified a total of 80 SNP displaying a significant association ( $p < 1.062 \times 10^{-6}$ ). Candidate genes were searched in the 20 kb up/down-stream of each SNP. There were 85 protein-coding candidate genes possibly related to the 8 important

phenotypic traits. Next, we conducted an RNA-seq based on a drought-tolerance inbred line (K58). A total of 18,922 DEGs were identified on 7 and 14 days after drought treatment. Up-regulated genes were mainly enriched in BCAA catabolic process, while down-regulated genes were mainly enriched in the photosynthesis process. Using a combined analysis, we found 14 common genes between GWAS and RNA-seq, three of them were PK genes, and one of them was TF gene. LOC110885273 was associated with LSA and SPAD, belongs to the RLK-Pelle\_SD-2b protein kinase family. LOC110872899 belongs to the RLK-Pelle\_LRR-III protein kinase family and is associated with RSA. LOC110891369 belongs to the RLK-Pelle\_CrRLK1L-1 protein kinase family and is associated with LSA. The PLATZ gene LOC110920644 is related to LSA, and belongs to PLATZ TF family. Through functional analysis, there are 4 genes involving the ABA-dependent drought response pathway of plants.

The integrative analysis of omics data is a promising approach to identify candidate genes for complex traits. This study is the first attempt to combine GWAS and RNA-seq to explore the genetic mechanism of sunflower drought tolerance to our knowledge. We will further validate the functions of these genes, possibly by overexpression or by CRISPER/Cas genome editing. Our research reveals the phenotypic and molecular mechanisms of drought response in sunflowers. The results will be useful for the genetic enhancement of drought-resistant sunflowers.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: <https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA797473>.

## AUTHOR CONTRIBUTIONS

JH and LY: conceptualization. LY, YM, and YW: methodology. YW: software, writing—original draft preparation, and visualization. JH: validation, project administration, and funding acquisition. YW, ZZ, and HS: formal analysis. YM, ZZ, HH, and ZH: investigation. HY: resources. YM and ZZ: data curation. LY, YZ, and HY: writing—review and editing. LY: supervision. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This work was financially supported by grants from the National Natural Science Foundation of China (Nos. 32160450 and 31760396).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.847435/full#supplementary-material>

## REFERENCES

- Adeleke, B. S., and Babalola, O. O. (2020). Oilseed crop sunflower (*Helianthus annuus*) as a source of food: nutritional and health benefits. *Food Sci. Nutr.* 8, 4666–4684. doi: 10.1002/fsn3.1783
- Adler, D., Nenadic, O., and Zucchini, W. (2003). “Rgl: a r-library for 3d visualization with OpenGL,” in *Proceedings of the 35th Symposium of the Interface: Computing Science and Statistics*. Salt Lake City, 1–11.
- Adopted, I. (2014). *Climate Change 2014 Synthesis Report*. Geneva: IPCC.
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Ali, N., Li, D., Eltahawy, M. S., Abdulmajid, D., Bux, L., Liu, E., et al. (2020). Mining of favorable alleles for seed reserve utilization efficiency in *Oryza sativa* by means of association mapping. *BMC Genetics* 21, 1–15. doi: 10.1186/s12863-020-0811-3
- Alza, J., and Fernandez-Martinez, J. (1997). Genetic analysis of yield and related traits in sunflower (*Helianthus annuus* L.) in dryland and irrigated environments. *Euphytica* 95, 243–251. doi: 10.1023/A:1003056500991
- Anders, S., and Huber, W. (2012). *Differential Expression of RNA-Seq Data at the Gene Level—the DESeq Package*. Heidelberg: European Molecular Biology Laboratory (EMBL), 10.
- Arnau, D., Desclos-Theveniau, M., and Zimmerli, L. (2012). Disease resistance to *Pectobacterium carotovorum* is negatively modulated by the Arabidopsis Lectin Receptor Kinase LecRK-V. 5. *Plant Signal. Behav.* 7, 1070–1072. doi: 10.4161/psb.21013
- Ault, T. R. (2020). On the essentials of drought in a changing climate. *Science* 368, 256–260. doi: 10.1126/science.aaz5492
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al. (2004). The Pfam protein families database. *Nucleic Acids Res.* 32(Suppl. 1), D138–D141. doi: 10.1093/nar/gkh121
- Bechtold, U., Penfold, C. A., Jenkins, D. J., Legaie, R., Moore, J. D., Lawson, T., et al. (2016). Time-series transcriptomics reveals that AGAMOUS-LIKE22 affects primary metabolism and developmental processes in drought-stressed Arabidopsis. *Plant Cell* 28, 345–366. doi: 10.1105/tpc.15.00910
- Binder, S., Knill, T., and Schuster, J. (2007). Branched-chain amino acid metabolism in higher plants. *Physiol. Plantarum* 129, 68–78. doi: 10.1111/j.1399-3054.2006.00800.x
- Blum, A. (2011). Drought resistance—is it really a complex trait? *Func. Plant Biol.* 38, 753–757. doi: 10.1071/FP11101
- Bonaventure, G. (2011). The *Nicotiana attenuata* LECTIN RECEPTOR KINASE 1 is involved in the perception of insect feeding. *Plant Signal. Behav.* 6, 2060–2063. doi: 10.4161/psb.6.12.18324
- Bouzi, M., He, F., Schmitz, G., Häusler, R., Weber, A., Mettler-Altman, T., et al. (2019). Arabidopsis species deploy distinct strategies to cope with drought stress. *Ann. Botany* 124, 27–40. doi: 10.1093/aob/mcy237
- Cadic, E., Coque, M., Vear, F., Grezes-Basset, B., Pauquet, J., Piquemal, J., et al. (2013). Combined linkage and association mapping of flowering time in Sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 126, 1337–1356. doi: 10.1007/s00122-013-2056-2
- Cavill, R., Jennen, D., Kleinjans, J., and Briedé, J. J. (2015). Transcriptomic and metabolomic data integration. *Briefings Bioinform.* 17, 891–901. doi: 10.1093/bib/bbv090
- Chaurasia, S., Singh, A. K., Kumar, A., Songachan, L., Yadav, M. C., Kumar, S., et al. (2021). Genome-wide association mapping reveals key genomic regions for physiological and yield-related traits under salinity stress in wheat (*Triticum aestivum* L.). *Genomics* 113, 3198–3215. doi: 10.1016/j.ygeno.2021.07.014
- Costa, C., Dwyer, L. M., Dutilleul, P., Stewart, D. W., Ma, B. L., and Smith, D. L. (2001). Inter-relationships of applied nitrogen, SPAD, and yield of leafy and non-leafy maize genotypes. *J. Plant Nutr.* 24, 1173–1194. doi: 10.1081/PLN-100106974
- Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., and Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *Ann. Rev. Plant Biol.* 61, 651–679. doi: 10.1146/annurev-arplant-042809-112122
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., and Blaxter, M. L. (2013). Special features of RAD Sequencing data: implications for genotyping. *Mol. Ecol.* 22, 3151–3164. doi: 10.1111/mec.12084
- Derakhshani, B., Ayalew, H., Mishina, K., Tanaka, T., Kawahara, Y., Jafary, H., et al. (2020). Comparative analysis of root transcriptome reveals candidate genes and expression divergence of homoeologous genes in response to water stress in wheat. *Plants* 9, 596. doi: 10.3390/plants9050596
- Diévar, A., and Clark, S. E. (2003). Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. *Curr. Opin. Plant Biol.* 6, 507–516. doi: 10.1016/S1369-5266(03)00089-X
- Ebrahimi Khaksefidi, R., Mirlolhi, S., Khalaji, F., Fakhari, Z., Shiran, B., Fallahi, H., et al. (2015). Differential expression of seven conserved microRNAs in response to abiotic stress and their regulatory network in *Helianthus annuus*. *Front. Plant Sci.* 6, 741. doi: 10.3389/fpls.2015.00741
- Escalante, M., Vigliocco, A., Moschen, S., Fernández, P., Heinz, R., Garcia-Garcia, F., et al. (2020). Transcriptomic analysis reveals a differential gene expression profile between two sunflower inbred lines with different ability to tolerate water stress. *Plant Mol. Biol. Rep.* 38, 1–16. doi: 10.1007/s11105-020-01192-4
- Espina, M. J., Ahmed, C. M. S., Bernardini, A., Adeleke, E., Yadegari, Z., Arelli, P., et al. (2018). Development and phenotypic screening of an ethyl methane sulfonate mutant population in soybean. *Front. Plant Sci.* 9, 394. doi: 10.3389/fpls.2018.00394
- FAO (2021). *FAO Statistical Yearbook – World Food and Agriculture*. FAO.
- Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., and Basra, S. (2009). Plant drought stress: effects, mechanisms and management. *Sustain. Agric.* 29, 153–188. doi: 10.1007/978-90-481-2666-8\_12
- Fernandez, G. C. (1992). “Effective selection criteria for assessing plant stress tolerance,” in *Proceeding of the International Symposium on Adaptation of Vegetables and other Food Crops in Temperature and Water Stress*. Shanhuah (1992), 257–270.
- Filippi, C. V., Merino, G. A., Montecchia, J. F., Aguirre, N. C., Rivarola, M., Naamati, G., et al. (2020). Genetic diversity, population structure and linkage disequilibrium assessment among international sunflower breeding collections. *Genes* 11, 283. doi: 10.3390/genes11030283
- Fischer, R., and Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses. *Austr. J. Agric. Res.* 29, 897–912. doi: 10.1071/AR9780897
- Flexas, J., Bota, J., Loreto, F., Cornic, G., and Sharkey, T. (2004). Diffusive and metabolic limitations to photosynthesis under drought and salinity in C3 plants. *Plant Biol.* 6, 269–279. doi: 10.1055/s-2004-820867
- Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44, 1054–1064. doi: 10.1111/j.1365-313X.2005.02591.x
- Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Res.* 17, 27–32. doi: 10.1111/1755-0998.12509
- Franck, N., Zamorano, D., Wallberg, B., Hardy, C., Ahumada, M., Rivera, N., et al. (2020). Contrasting grapevines grafted into naturalized rootstock suggest scion-driven transcriptomic changes in response to water deficit. *Sci. Horticul.* 262, 109031. doi: 10.1016/j.scienta.2019.10.9031
- Galmes, J., Conesa, M. A., Ochogavía, J. M., Perdomo, J. A., Francis, D. M., Ribas-Carbó, M., et al. (2011). Physiological and morphological adaptations in relation to water use efficiency in Mediterranean accessions of *Solanum lycopersicum*. *Plant Cell. Environ.* 34, 245–260. doi: 10.1111/j.1365-3040.2010.02239.x
- Gaut, B. S., and Long, A. D. (2003). The slowdown on linkage disequilibrium. *Plant Cell.* 15, 1502–1506. doi: 10.1105/tpc.150730
- Geetha, A., Suresh, J., and Saidaiah, P. (2012). Study on response of sunflower (*Helianthus annuus* L.) genotypes for root and yield characters under water stress. *Curr. Biot.* 6, 32–41. doi: 10.1111/j.1365-313X.2003.01987.x
- Ghaffari, M., Toorchi, M., Valizadeh, M., and Shakiba, M. R. (2012). Morpho-physiological screening of sunflower inbred lines under drought stress condition. *Turkish J. Field Crop.* 17, 185–190. doi: 10.2298/GENSR1203701Z
- Grasso, S., Pintado, T., Pérez-Jiménez, J., Ruiz-Capillas, C., and Herrero, A. M. (2020). Potential of a sunflower seed by-product as animal fat replacer in healthier frankfurters. *Foods* 9, 445. doi: 10.3390/foods9040445

- Gunes, A., Pilbeam, D. J., Inal, A., and Coban, S. (2008). Influence of silicon on sunflower cultivars under drought stress. I: growth, antioxidant mechanisms, and lipid peroxidation. *Commun. Soil Sci. Plant Anal.* 39, 1885–1903. doi: 10.1080/00103620802134651
- Guo, J., Li, C., Zhang, X., Li, Y., Zhang, D., Shi, Y., et al. (2020). Transcriptome and GWAS analyses reveal candidate gene for seminal root length of maize seedlings under drought stress. *Plant Sci.* 292, 110380. doi: 10.1016/j.plantsci.2019.110380
- Haddadi, P., Yazdi-Samadi, B., Naghavi, M. R., Kalantari, A., Maury, P., and Sarrafi, A. (2011). QTL analysis of agronomic traits in recombinant inbred lines of sunflower under partial irrigation. *Plant Biotechnol. Rep.* 5, 135–146. doi: 10.1007/s11816-011-0164-5
- Harb, A., Simpson, C., Guo, W., Govindan, G., Kakani, V. G., and Sunkar, R. (2020). The effect of drought on transcriptome and hormonal profiles in barley genotypes with contrasting drought tolerance. *Front. Plant Sci.* 11, 618491. doi: 10.3389/fpls.2020.618491
- He, F., Wei, C., Zhang, Y., Long, R., Li, M., Wang, Z., et al. (2021). Genome-wide association analysis coupled with transcriptome analysis reveals candidate genes related to salt stress in alfalfa (*Medicago sativa* L.). *Front. Plant Sci.* 12, 826584–826584. doi: 10.3389/fpls.2021.826584
- Hou, S., Zhu, G., Li, Y., Li, W., Fu, J., Niu, E., et al. (2018). Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum* L.). *Front. Plant Sci.* 9, 1276. doi: 10.3389/fpls.2018.01276
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience.* 8, giy154. doi: 10.1093/gigascience/giy154
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Hussain, M., Farooq, S., Hasan, W., Ul-Allah, S., Tanveer, M., Farooq, M., et al. (2018). Drought stress in sunflower: physiological effects and its management through breeding and agronomic alternatives. *Agric. Water Manag.* 201, 152–166. doi: 10.1016/j.agwat.2018.01.028
- Hussain, M., Malik, M., Farooq, M., Ashraf, M., and Cheema, M. (2008). Improving drought tolerance by exogenous application of glycinebetaine and salicylic acid in sunflower. *J. Agron. Crop. Sci.* 194, 193–199. doi: 10.1111/j.1439-037X.2008.00305.x
- Hussain, R. A., Ahmad, R., Nawaz, F., Ashraf, M. Y., and Waraich, E. A. (2016). Foliar NK application mitigates drought effects in sunflower (*Helianthus annuus* L.). *Acta Physiol. Plant.* 38, 1–14. doi: 10.1007/s11738-016-2104-z
- Hussain, S., Saleem, M., Ashraf, M., Cheema, M., and Haq, M. (2010). Abscisic acid, a stress hormone helps in improving water relations and yield of sunflower (*Helianthus annuus* L.) hybrids under drought. *Pakis. J. Bot.* 42, 2177–2189. doi: 10.3417/2008072
- Ibrahim, M., Faisal, A., and Shehata, S. (2016). Calcium chloride alleviates water stress in sunflower plants through modifying some physio-biochemical parameters. *American-Eurasian J. Agric. Environ. Sci.* 16, 677–693. doi: 10.5829/idosi.ajeaes.2016.16.4.12907
- Javaid, T., Bibi, A., Sadaqat, H., and Javed, S. (2015). Screening of sunflower (*Helianthus annuus* L.) hybrids for drought tolerance at seedling stage. *Int. J. Plant Sci. Ecol.* 1, 6–16.
- Kaur, G., and Asthir, B. (2017). Molecular responses to drought stress in plants. *Biol. Plantarum* 61, 201–209. doi: 10.1007/s10535-016-0700-9
- Kaya, M. D., Okçu, G., Atak, M., Cikili, Y., and Kolsarici, Ö. (2006). Seed treatments to overcome salt and drought stress during germination in sunflower (*Helianthus annuus* L.). *Eur. J. Agron.* 24, 291–295. doi: 10.1016/j.eja.2005.08.001
- Khan, S., Choudhary, S., Pandey, A., Khan, M. K., and Thomas, G. (2015). Sunflower oil: efficient oil source for human consumption. *Emerg. Life Sci. Res.* 1, 1–3. Available online at: [https://www.emergentresearch.org/uploads/38/1768\\_pdf.pdf](https://www.emergentresearch.org/uploads/38/1768_pdf.pdf)
- Khanzada, H., Wassan, G. M., He, H., Mason, A. S., Keerio, A. A., Khanzada, S., et al. (2020). Differentially evolved drought stress indices determine the genetic variation of *Brassica napus* at seedling traits by genome-wide association mapping. *J. Adv. Res.* 24, 447–461. doi: 10.1016/j.jare.2020.05.019
- Kiani, S. P., Talia, P., Maury, P., Grieu, P., Heinz, R., Perrault, A., et al. (2007). Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments. *Plant Sci.* 172, 773–787. doi: 10.1016/j.plantsci.2006.12.007
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Kofsky, J., Zhang, H., and Song, B.-H. (2020). Genetic architecture of early vigor traits in wild soybean. *Int. J. Mol. Sci.* 21, 3105. doi: 10.3390/ijms21093105
- Lata, C., Muthamilarasan, M., and Prasad, M. (2015). “Drought Stress Responses and Signal Transduction in Plants,” in *Elucidation of Abiotic Stress Signaling in Plants: Functional Genomics Perspectives, Vol. 2*, ed. G. K. Pandey (New York, NY: Springer New York), 195–225.
- Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800. doi: 10.1093/molbev/msv150
- Lehti-Shiu, M. D., Zou, C., Hanada, K., and Shiu, S.-H. (2009). Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol.* 150, 12–26. doi: 10.1104/pp.108.134353
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, R., Zeng, Y., Xu, J., Wang, Q., Wu, F., Cao, M., et al. (2015). Genetic variation for maize root architecture in response to drought stress at the seedling stage. *Breed. Sci.* 65, 298–307. doi: 10.1270/jsbbs.65.298
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, M., Wang, C., Ji, Z., Zhang, L., Li, C., Huang, J., et al. (2021). Regulation of drought tolerance in Arabidopsis involves PLATZ4-mediated transcriptional suppression of PIP2. *bioRxiv*. doi: 10.1101/2021.10.06.463369
- Liu, S.-H., Shen, P.-C., Chen, C.-Y., Hsu, A.-N., Cho, Y.-C., Lai, Y.-L., et al. (2020). DriverDBv3: a multi-omics database for cancer driver gene research. *Nucleic Acids Res.* 48, D863–D870. doi: 10.1093/nar/gkz964
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8
- Ma, J., Geng, Y., Pei, W., Wu, M., Li, X., Liu, G., et al. (2018). Genetic variation of dynamic fiber elongation and developmental quantitative trait locus mapping of fiber length in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics* 19, 882–882. doi: 10.1186/s12864-018-5309-2
- Ma, X., Feng, F., Wei, H., Mei, H., Xu, K., Chen, S., et al. (2016). Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. *Front. Plant Sci.* 7, 1801. doi: 10.3389/fpls.2016.01801
- Mandel, J. R., Nambeesan, S., Bowers, J. E., Marek, L. F., Ebert, D., Rieseberg, L. H., et al. (2013). Association mapping and the genomic consequences of selection in sunflower. *PLoS Genetics* 9, e1003378. doi: 10.1371/journal.pgen.1003378
- Martínez, D., and Guaiamet, J. (2004). Distortion of the SPAD 502 chlorophyll meter readings by changes in irradiance and leaf water status. *Agronomie* 24, 41–46. doi: 10.1051/agro:2003060
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Moschen, S., Di Rienzo, J. A., Higgins, J., Tohge, T., Watanabe, M., González, S., et al. (2017). Integration of transcriptomic and metabolic data reveals hub transcription factors involved in drought stress response in sunflower (*Helianthus annuus* L.). *Plant Mol. Biol.* 94, 549–564. doi: 10.1007/s11103-017-0625-5

- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., et al. (2016). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol.* 17, 53. doi: 10.1186/s13059-016-0917-0
- Mwadzingeni, L., Shimelis, H., Rees, D. J. G., and Tsilo, T. J. (2017). Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. *PLoS ONE* 12, e0171692. doi: 10.1371/journal.pone.0171692
- Mwale, S., Hamusimbi, C., and Mwansa, K. (2003). Germination, emergence and growth of sunflower (*Helianthus annuus* L.) in response to osmotic seed priming. *Seed Sci. Technol.* 31, 199–206. doi: 10.15258/sst.2003.31.1.21
- Nida, H., Girma, G., Mekonen, M., Tirfessa, A., Seyoum, A., Bejiga, T., et al. (2021). Genome-wide association analysis reveals seed protein loci as determinants of variations in grain mold resistance in sorghum. *Theor. Appl. Genetics* 134, 1167–1184. doi: 10.1007/s00122-020-03762-2
- Osakabe, Y., Maruyama, K., Seki, M., Satou, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2005). Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in arabidopsis. *Plant Cell* 17, 1105–1119. doi: 10.1105/tpc.104.027474
- Ouyang, S. Q., Liu, Y. F., Liu, P., Lei, G., He, S. J., Ma, B., et al. (2010). Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (*Oryza sativa*) plants. *Plant J.* 62, 316–329. doi: 10.1111/j.1365-3113X.2010.04146.x
- Owens, G. L., Baute, G. J., Hubner, S., and Rieseberg, L. H. (2019). Genomic sequence and copy number evolution during hybrid crop development in sunflowers. *Evolut. Appl.* 12, 54–65. doi: 10.1111/eva.12603
- Pan, J., Li, Z., Wang, Q., Yang, L., Yao, F., and Liu, W. (2020). An S-domain receptor-like kinase, OsESG1, regulates early crown root development and drought resistance in rice. *Plant Sci.* 290, 110318. doi: 10.1016/j.plantsci.2019.110318
- Pasda, G., and Diepenbrock, W. (1990). The physiological yield analysis of sunflower (*Helianthus annuus* L.). Part II. Climatic factors. *Fett Wissenschaft Technol.* 93, 155–168. doi: 10.1002/lipi.19910930501
- Pereyra-Irujo, G. A., Velázquez, L., Granier, C., and Aguirrezabal, L. A. (2007). A method for drought tolerance screening in sunflower. *Plant Breed.* 126, 445–448. doi: 10.1111/j.1439-0523.2007.01375.x
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Petcu, E., Arsintescu, A., and Stanciu, D. (2001). The effect of hydric stress on some characteristics of sunflower plants. *Romanian Agric. Res.* 16, 15–22.
- Pires, M. V., Pereira Júnior, A. A., Medeiros, D. B., Daloso, D. M., Pham, P. A., Barros, K. A., et al. (2016). The influence of alternative pathways of respiration that utilize branched-chain amino acids following water shortage in Arabidopsis. *Plant Cell Environ.* 39, 1304–1319. doi: 10.1111/pce.12682
- Poormohammad Kiani, S., Maury, P., Nouri, L., Ykhlef, N., Grieu, P., and Sarrafi, A. (2009). QTL analysis of yield-related traits in sunflower under different water treatments. *Plant Breed.* 128, 363–373. doi: 10.1111/j.1439-0523.2009.01628.x
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* 81, 559–575. doi: 10.1086/519795
- Rabert, G. A., Manivannan, P., Somasundaram, R., and Panneerselvam, R. (2014). Triazole compounds alters the antioxidant and osmoprotectant status in drought stressed *Helianthus annuus* L. plants. *Emir. J. Food Agric.* 26, 265–276. doi: 10.9755/efja.v26i3.17385
- Rauf, S. (2008). Breeding sunflower (*Helianthus annuus* L.) for drought tolerance. *Commun. Biomet. Crop Sci.* 3, 29–44. doi: 10.3923/pjbs.1999.846.848
- Rauf, S., and Ahmad Sadaqat, H. (2008). Effect of osmotic adjustment on root length and dry matter partitioning in sunflower (*Helianthus annuus* L.) under drought stress. *Acta Agric. Scand. Section B–Soil Plant Sci.* 58, 252–260. doi: 10.1080/09064710701628958
- Ray, S., Dansana, P. K., Giri, J., Devshwar, P., Arora, R., Agarwal, P., et al. (2011). Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice. *Funct. Integr. Genomics* 11, 157–178. doi: 10.1007/s10142-010-0187-y
- Riddell, E. A., Odom, J. P., Damm, J. D., and Sears, M. W. (2018). Plasticity reveals hidden resistance to extinction under climate change in the global hotspot of salamander diversity. *Sci. Adv.* 4, eaar5471. doi: 10.1126/sciadv.aar5471
- Robertson, J. M., Pharis, R. P., Huang, Y. Y., Reid, D. M., and Yeung, E. C. (1985). Drought-induced increases in abscisic acid levels in the root apex of sunflower 1. *Plant Physiol.* 79, 1086–1089. doi: 10.1104/pp.79.4.1086
- Salami, M., and Saadat, S. (2013). Study of potassium and nitrogen fertilizer levels on the yield of sugar beet in jolge cultivar. *J. Novel Appl. Sci.* 2, 94–100.
- Schilling, E. E., and Heiser, C. B. (1981). Infrageneric classification of Helianthus (Compositae). *Taxon* 30, 393–403. doi: 10.2307/1220139
- Singh, P., Kuo, Y.-C., Mishra, S., Tsai, C.-H., Chien, C.-C., Chen, C.-W., et al. (2012). The lectin receptor kinase-VI. 2 is required for priming and positively regulates Arabidopsis pattern-triggered immunity. *Plant Cell* 24, 1256–1270. doi: 10.1105/tpc.112.095778
- Soni, P., and Abidin, M. Z. (2017). Water deficit-induced oxidative stress affects artemisinin content and expression of proline metabolic genes in *Artemisia annua* L. *FEBS Open Bio.* 7, 367–381. doi: 10.1002/2211-5463.12184
- Sukumaran, S., Reynolds, M. P., and Sansaloni, C. (2018). Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. *Front. Plant Sci.* 9, 81. doi: 10.3389/fpls.2018.00081
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013b). SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* 8, e58700. doi: 10.1371/journal.pone.0058700
- Sun, X.-L., Yu, Q.-Y., Tang, L.-L., Ji, W., Bai, X., Cai, H., et al. (2013a). GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress. *J. Plant Physiol.* 170, 505–515. doi: 10.1016/j.jplph.2012.11.017
- Supek, F., and Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6, e21800. doi: 10.1371/journal.pone.0021800
- Tagliotti, M. E., Deperi, S. I., Bedogni, M. C., and Huarte, M. A. (2021). Genome-wide association analysis of agronomical and physiological traits linked to drought tolerance in a diverse potatoes (*Solanum tuberosum*) panel. *Plant Breed.* 140, 654–664. doi: 10.1111/pbr.12938
- Togninalli, M., Roqueiro, D., Investigators, C. O., and Borgwardt, K. M. (2018). Accurate and adaptive imputation of summary statistics in mixed-ethnicity cohorts. *Bioinformatics* 34, i687–i696. doi: 10.1093/bioinformatics/bty596
- Turhan, H., and Baser, I. (2004). *In vitro* and *in vivo* water stress in sunflower (*Helianthus Annuus* L.)/Estrés Hídrico En Girasol (*Helianthus annuus* L.) En Las Condiciones *in vitro* E *in vivo*/stress D'eau Du Tournesol (*Helianthus annuus* L.) Dans Les Conditions *in vitro* Et *in vivo*. *Helia* 27, 227–236. doi: 10.2298/HEL040227T
- VanLiere, J. M., and Rosenberg, N. A. (2008). Mathematical properties of the r2 measure of linkage disequilibrium. *Theor. Popul. Biol.* 74, 130–137. doi: 10.1016/j.tpb.2008.05.006
- Wang, H., and Qin, F. (2017). Genome-wide association study reveals natural variations contributing to drought resistance in crops. *Front. Plant Sci.* 8, 1110. doi: 10.3389/fpls.2017.01110
- Wang, L., Liu, Y., Gao, L., Yang, X., Zhang, X., Xie, S., et al. (2022). Identification of candidate forage yield genes in sorghum (*Sorghum bicolor* L.) using integrated genome-wide association studies and RNA-Seq. *Front. Plant Sci.* 12, 433. doi: 10.3389/fpls.2021.788433
- Wilhite, D. A., and Buchanan-Smith, M. (2005). Drought as hazard: understanding the natural and social context. *Drought Water Crises Sci. Technol. Manag. Issues* 3, 29. doi: 10.1201/9781420028386.pt1
- Xie, D., Dai, Z., Yang, Z., Tang, Q., Deng, C., Xu, Y., et al. (2019). Combined genome-wide association analysis and transcriptome sequencing to identify candidate genes for flax seed fatty acid metabolism. *Plant Sci.* 286, 98–107. doi: 10.1016/j.plantsci.2019.06.004
- Xing, H. T., Guo, P., Xia, X. L., and Yin, W. L. (2011). PdERECTA, a leucine-rich repeat receptor-like kinase of poplar, confers enhanced water use efficiency in Arabidopsis. *Planta* 234, 229–241. doi: 10.1007/s00425-011-1389-9
- Yin, L. (2018). *CMplot: Circle Manhattan Plot*. San Francisco, CA: GitHub, Inc.
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic J. Integ. Biol.* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003

- Zenda, T., Liu, S., Wang, X., Liu, G., Jin, H., Dong, A., et al. (2019). Key maize drought-responsive genes and pathways revealed by comparative transcriptome and physiological analyses of contrasting inbred lines. *Int. J. Mol. Sci.* 20, 1268. doi: 10.3390/ijms20061268
- Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., and Yang, T.-L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875
- Zhang, T., Zhang, X., Han, K., Zhang, G., Wang, J., Xie, K., et al. (2017). Analysis of long noncoding RNA and mRNA using RNA sequencing during the differentiation of intramuscular preadipocytes in chicken. *PLoS ONE* 12, e0172389. doi: 10.1371/journal.pone.0172389
- Zhang, X., Wang, G., Chen, B., Du, H., Zhang, F., Zhang, H., et al. (2018). Candidate genes for first flower node identified in pepper using combined SLAF-seq and BSA. *PLoS ONE* 13, e0194071. doi: 10.1371/journal.pone.0194071
- Zhao, W., Liu, Y.-W., Zhou, J.-M., Zhao, S.-P., Zhang, X.-H., and Min, D.-H. (2016). Genome-wide analysis of the lectin receptor-like kinase family in foxtail millet (*Setaria italica* L.). *Plant Cell Tissue Organ Cult.* 127, 335–346. doi: 10.1007/s11240-016-1053-y
- Zheng, Y., Jiao, C., Sun, H., Rosli, H., Pombo, M. A., Zhang, P., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014
- Zhou, Q., Zhou, C., Zheng, W., Mason, A. S., Fan, S., Wu, C., et al. (2017). Genome-wide SNP markers based on SLAF-seq uncover breeding traces in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* 8, 648. doi: 10.3389/fpls.2017.00648
- Zilong, Z., Jianhua, H., Liuxi, Y., HaiFeng, Y., Huimin, S., Jinglin, W., et al. (2021). Drought resistance identification and drought resistance index screening of sunflower germplasm resources at seedling stage. *Agric. Res. Arid Area.* 39, 228–238. doi: 10.7606/j.issn.1000-7601.2021.04.29
- Zotova, L., Kurishbayev, A., Jatayev, S., Khassanova, G., Zhubatkanov, A., Serikbay, D., et al. (2018). Genes encoding transcription factors TaDREB5 and TaNFYC-A7 are differentially expressed in leaves of bread wheat in response to drought, dehydration and ABA. *Front. Plant Sci.* 9, 1441. doi: 10.3389/fpls.2018.01441

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

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

Copyright © 2022 Wu, Shi, Yu, Ma, Hu, Han, Zhang, Zhen, Yi and Hou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.