

RESEARCH ARTICLE

De novo transcriptomic analysis of Doum Palm (*Hyphaene compressa*) revealed an insight into its potential drought tolerance

Allen Johnny Borlay^{1,2*}, Cecilia Mbithe Mweu³, Steven Ger Nyanjom⁴, Kevin Mbogo Omolo⁴, Labode Hospice Stevenson Naitchede²

1 Department of Biological Sciences, University of Liberia, Monrovia, Liberia, **2** Department of Molecular Biology and Biotechnology, Pan African University Institute for Basic Sciences, Technology and Innovation, Nairobi, Kenya, **3** Institute for Biotechnology Research, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, **4** Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

* allenborlay@gmail.com



Abstract

OPEN ACCESS

Citation: Borlay AJ, Mweu CM, Nyanjom SG, Omolo KM, Naitchede LHS (2024) *De novo* transcriptomic analysis of Doum Palm (*Hyphaene compressa*) revealed an insight into its potential drought tolerance. PLoS ONE 19(3): e0292543. <https://doi.org/10.1371/journal.pone.0292543>

Editor: Ramegowda Venkatesh, University of Agricultural Sciences, INDIA

Received: October 27, 2022

Accepted: September 24, 2023

Published: March 12, 2024

Copyright: © 2024 Borlay et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: RNA-Seq data were deposited in the NCBI Sequence Read Archive (SRA) database with the accession number PRJNA861903.

Funding: This research was funded by the University of Liberia and the AU Commission through the Pan African University Institute for Basic Sciences, Technology, and Innovation. The funders had no role in the study design, data collection, analysis, the decision to publish, or the preparation of the manuscript.

Background

Doum palms (*Hyphaene compressa*) perform a crucial starring role in the lives of Kenya's arid and semi-arid people for empowerment and sustenance. Despite the crop's potential for economic gain, there is a lack of genetic resources and detailed information about its domestication at the molecular level. Given the doum palm's vast potential as a widely distributed plant in semi-arid and arid climates and a source of many applications, coupled with the current changing climate scenario, it is essential to understand the molecular processes that provide drought resistance to this plant.

Results

Assembly of the first transcriptome of doum palms subjected to water stress generated about 39.97 Gb of RNA-Seq data. The assembled transcriptome revealed 193,167 uni-genes with an average length of 1655 bp, with 128,708 (66.63%) successfully annotated in seven public databases. Unigenes exhibited significant differentially expressed genes (DEGs) in well-watered and stressed-treated plants, with 45071 and 42457 accounting for up-regulated and down-regulated DEGs, respectively. GO term, KEGG, and KOG analysis showed that DEGs were functionally enriched cellular processes, metabolic processes, cellular and catalytic activity, metabolism, genetic information processing, signal transduction mechanisms, and posttranslational modification pathways. Transcription factors (TF), such as the MYB, WRKY, NAC family, FAR1, B3, bHLH, and bZIP, were the prominent TF families identified as doum palm DEGs encoding drought stress tolerance.

Conclusions

This study provides a complete understanding of DEGs involved in drought stress at the transcriptome level in doum palms. This research is, therefore, the foundation for the

Competing interests: The authors have declared that no competing interests exist.

characterization of potential genes, leading to a clear understanding of its drought stress responses and providing resources for improved genetic modification.

1 Introduction

Stress from drought is among the most severe forms of environmental limitation, as it has a chilling effect on plant development and productivity. Worldwide, agriculture is becoming increasingly challenging to grow food in areas experiencing more extended periods of water stress due to climate change. Drought may induce organic process disproportion in plant cells and alter plant leaves' optical energy assimilation, damaging plants' photosynthetic organs [1]. Plants frequently encounter uncomplimentary environmental circumstances throughout their lifetime due to their sessile nature, and these circumstances, which include drought, restrict plant development and output [2]. Plant researchers and breeders are keenly interested in the plant's entire cellular workings that govern its reaction to drought stress to sustain and improve crop yields. Understanding the mechanisms that boost plants during drought is crucial for designing new drought-resistant strategies and maintaining global food security [3].

Plants have evolved various self-defense and defensive systems to deal with drought stress throughout their evolutionary process [4]. As indicated by [5], many genes are expressed in plants in retort to water deficit. Genes for protein structural maintenance, cell-water binding capacity, heat shock proteins, proteins prevalent in late embryogenesis, and genes that keep membrane proteins stable are some of the genes that help plants retort to drought. Plants are protected from the harmful effects of water shortages by proteins that change how genes are expressed and control metabolic processes. The synthesis of these proteins is a defense mechanism against the first effects of water deficit. Furthermore, transcription factors (TFs) normalize the workings of numerous genes to thwart these negative consequences [6]. Determining how crops can tolerate water deficits and creating new methods of plant cultivation to resist climate change would require understanding the genes translated and expressed in reaction to water stress.

Doum palms (*Hyphaene compressa*) perform a crucial starring role in the lives of Kenya's arid and semi-arid people (Nomadic pastoralists) for economic empowerment and sustenance [7]. During the lingering summertime, when there is widespread food insecurity across Africa, especially in Djibouti, Sudan, Kenya, Nigeria, and Namibia, the fruits of *Hyphaene* are considered an essential source of subsistence [8, 9]. Doum palm fruit aqueous extract possesses anti-inflammatory, antibacterial, and anticancer phytoconstituents [10]. In Kenya, the leaves of Doum Palm are used for making thatch, baskets, mats (tablemats, floor mats, sleeping mats), brooms, carpets, ropes, caps, seats, or door shutters.

Due to their unique adaptation to water scarcity and dry climates, doum palms can be found in Kenya's arid and semi-arid regions, particularly in the north river courses, lakes, and coasts [11]. *Hyphaene* species can be found in various habitats, including sandy lowlands, secondary woods open to the sky, and coastal and inland savannahs [8]. Coastal savannas in East Africa, particularly in Kenya and Tanzania, are dominated by their distinctive shapes.

Recent advances in molecular biology techniques have helped find plants' possible environmental stress response mechanisms. RNA Sequencing (RNA-Seq), a method for profiling expressed transcripts, is used to determine essential genes/factors related to abiotic/biotic stresses or diseases in plants and other species [12]. There are apparent advantages to using RNA-Seq, including the ability to carry out large-scale functional gene assignment, qualitative

and quantitative gene expression analysis, and improved understanding and precision in characterizing eukaryotic transcriptomes, particularly those from non-model species [13]. RNA-Seq has changed plant transcriptomics research since it is efficient, reliable, and effective. Many plant researchers and molecular breeder have successfully used RNA-Seq to classify abiotic condition retort mechanisms. This powerful technology is widely used to sequence transcriptomes, allowing for the generation of functional genomic data and the discovery of DEGs across cultivars, organs, and treatment conditions [14–16]. RNA-seq data, such as annotated genes, DEGs, and molecular markers, may offer insight into discovering novel genes [17]. Here, high automated sequencing technique was utilized to sequence the transcriptomes of drought-stressed leaf samples from doum palms to obtain functional genomic data that may shed light on the molecular processes underpinning drought responses in doum palms.

The transcriptomes of doum palms responding to drought stress have been systematically characterized in this work. These doum palm transcriptome data can potentially be exploited in crop enhancement strategies in the molecular breeding of doum palms for abiotic stress tolerance by identifying genes and markers involved in water deficit tolerance.

2 Materials and methods

2.1 Doum palm samples and stress imposition for *de novo* assay

Doum palm seeds were collected from Tharaka Nithi, Kenya, in May of 2021 and brought to the Institute for Biotechnology Research (IBR) greenhouse at the Jomo Kenyatta University of Agriculture and Technology (JKUAT) for further use. Doum palms are wild plants hence no permission is needed for seed collections as per the Kenya laws. Additionally, appropriate greenhouse guidelines and procedures were straightly followed as provided by the JKUAT charter. Before sowing, the seeds were unshelled using a knife to break seed dormancy (physical application by using a knife and applying force to remove the outer covering of the doum palms seeds) and surface sterilized by successive washings using 3% redomil and distilled water. Seedlings were soaked in distilled water for 72 hours to induce rooting until germination. After seven weeks, the germinated seedlings were transplanted and grown in pots with regular watering for ten months before being subjected to drought stress for 65 days with fully irrigated samples as a control. The control plants received constant watering (1.5 liters) every two days, while the drought-stressed plants received no water for 65 days with a greenhouse temperature of 26°C.

2.2 Total RNA isolation and cDNA synthesis

RNA was extracted using a modified ISOLATE II Plant Kit (Meridian Bioscience, London, UK).

Trizol (600 µL) was used to lyse up to 100 mg of tissue at a time. 250 µL of membrane desalting buffer (MEM) was use to desalt the silica membrane. RNA extracts were treated with DNase enzyme for 15 minutes at 20°C to eliminate contaminating DNA. RNase-free water (50 µL) was used to elute the RNA. Jenway Genova Nanodrop Spectrophotometer (Bibby Scientific Ltd Beacon, United Kingdom) and the MyGel Mini compact electrophoresis system (Benchmark Scientific Inc, Sayreville, NJ 08872 USA) were used to examine the isolated RNA for concentration, purity, and quality. Using a Tetro™ cDNA Synthesis Kit, the isolated RNA was converted into cDNA by reverse transcription (Meridian Bioscience, London, UK).

2.3 Library construction, sequencing, reads filtering, and *De novo* assembly

The Beijing Genomic Institution performed (BGI) the library construction and sequencing on a BGISEQ-500 platform. *De novo* assembly with clean readings was accomplished using the

Trinity package with Kmer = 25 as the standard parameter [18]. Using Tgicl, the assembled transcripts were grouped hierarchically [19]. Raw sequence reads were subjected to quality filtering using FasqQC (version 0.11.5). Raw reads are a characteristic of low-quality reads, adapter contamination, and a large concentration of unknown base (N) reads, which should be filtered out before further analysis. To get more accurate results and clean readings, we discarded short reads containing adapters, reads with values below 5, and reads with unknown bases (N) over 10% of the total bases. Clean data were calculated simultaneously, including GC content, Q20, Q30, and sequence repetition levels. The *de novo* transcriptome short reads and unigene refer to the sequences generated by Trinity's method.

Contigs were formed from the overlap regions of the short reads, and for the final set of Unigenes, gene family clustering was performed via Tgicl. Following hierarchical clustering, each cluster's lengthiest sequence (unigene) was utilized for further data on length circulation, gene identification, and DEG detection. The unigenes were then categorized into gene families, following which clusters and singletons were discovered as two distinct categories.

2.4 Functional annotation, open reading frames (ORFs), and Pfam domain detection

For gene annotation, the BLAST (E-value $< 1.0 \times 10^{-5}$) programs [20] and Diamond [21] against NT (Nucleotide database), NR (Protein database), KOG, KEGG, and SwissProt databases were utilized. The [22] method was used to annotate the InterPro database, and Blast2GO [23] was used to annotate the NR and GO terms. Sequence-identical proteins to the supplied unigenes were retrieved with annotations on their putative functions. The degree of association between biological repeats was calculated using the Pearson's correlation analysis on fragments per kilobase per million fragments (FPKM) values from reads. The possible coding region was subsequently determined using the TransDecoder application. Predicted coding sequences (CDSs) were obtained from Pfam homology sequences by blasting the most extended open reading frame (ORF) against the SwissProt and Hmmscan Coding DNA sequence databases. Furthermore, query sequences derived from unigenes were utilized to search the Pfam (Protein family) database.

2.5 Unigene Gene Ontology (GO) Classification and Enrichment Analysis for DEGs

The GO annotation of the unigenes was obtained using the NR with the application of the Blast2GO program. GO functional evaluation was carryout for the entire transcripts with the aid of the WEGO program to get a global view of the distribution of genes' roles in doum palms [24]. A gene list and gene counts for each of the DEGs annotated with GO terms were generated by first mapping all of the DEGs in the GO catalog. The hypergeometric test was utilized to discover highly improved GO terms among the DEGs using the formula below to calculate the p-value:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

2.6 Kyoto encyclopedia of genes and genomes (KEGG) pathway evaluation for DEGs

After considering the complete transcriptome, we find that enrichment analysis significantly enhanced specific pathways regarding DEGs. We utilized the p-value technique that mirrors the one applied in the GO term analysis. An analog of the p-value known as Q represents the FDR (false discovery rate) equivalent of a value of 0.01. Multiple test adjustments allow us to identify highly enriched DEG pathways with Q-values of ≤ 0.05 .

2.7 Transcription factor (TF) analysis of DEGs and simple sequence repeats (SSR) detection

We used the getorf program [25] to determine the open reading frame (ORF) of each Unigene, and then we used hmmsearch [26] to align the ORF to TF realms from the PlntfDB database, thereby identifying each TF following the PlantfDB standards. SSRs in the unigenes of doum palm were detected using the MICOSSatellite program with the default setting.

2.8 Gene expression quantification and DEGs screening

To get the number of reads count, we used the Bowtie aligner and the expectation-maximization method (RSEM) from the Trinity protocol to run the Perl script align and estimate its richness. Through assembly and clustering, FPKM scores representing relative quantification levels of individual unigene transcripts were determined using the normalization procedure [27] approach. We used DESeq2, an R package developed by [28], to analyze the DEGs. DEGs fold-change (FC) values were calculated based on the following cutoff criteria: $P < 0.05$ and $\log |FC| \geq 1.5$. Significant positive or negative changes in logarithmic values indicated the DEGs' up-regulated or down-regulated status.

2.9 Quantitative RT-PCR authentication of sequencing results

We utilized a CFX Connect qPCR detection instrument to confirm the sequencing findings for gene expression profiling. The elongation factor 1-alpha gene (accession number: AB061263) served as a reference gene, while four drought tolerance-related genes were randomly chosen (Table 1) for validating the RNA-Seq results. The NCBI Primer-BLAST program, a resource for generating gene-specific primers, was utilized to brand primers for evaluating relative gene expression. After that, validation analysis was performed in 96 well-plates (Applied Biosystems, Foster, California) with SYBR Green master mix (Luna[®] Universal qPCR Master Mix, NEB) in qTower (Applied Biosystems) at 95°C for 60 seconds, proceeded by 40 cycles of 95°C for 15 seconds, and 58°C for 30 seconds. The gene expression level was performed by the comparative $2^{-\Delta\Delta CT}$ methods [29] using all samples in triplicate for normalization.

Table 1. Gene used for RNA-Seq validation.

Primer ID	Gene accession	Forward sequence	Reverse sequence	Size
Efl α	AB061263	ATTGGAACGGATATGCTCCA	TCCTTACCTGAACGCCTGTCA	188
CL5334.Contig3	XP_008794197.1	GATCTTACCAAGCGCCAGCA	CCTGGGAATGCACAAGCAAG	140
CL5518.Contig6	XP_010920097.1	CATCCTCGTCTACTGCACCG	ATGATCCCTTGACGATGCC	165
CL1905.Contig16	XP_010926195.1	GACCCCTCGTCTGAAGTTG	CTTTGGACCTGCTTGCGAAC	254
CL8906.Contig36	XP_010916565.1	CCGGACTCGTTTGGGAACT	TCCCTCCTTTCCAGTCAGA	209

<https://doi.org/10.1371/journal.pone.0292543.t001>

2.9.1 Sample size, *De novo*, and qPCR statistical analysis. The experiment used a completely randomized design (CRD) setup with three biological and technical replicates for each treatment. The setup had two treatments: drought/stress and control/well-watered. There were 18 individual data points (3 biological replicates x 3 technical replicates x 2 treatments) used for quality control and assessing variability representing the total samples or dataset (a representative of the treatments and replicates). We used SAS version 9.2 and GraphPad Prism 8 for the statistical analysis. A one-way analysis of variance (ANOVA) with Tukey's post hoc test was employed to identify any significant differences between groups and treatments. All experimental findings were presented as means with standard deviations (SDs) with a significance threshold for statistical analysis set at 0.05.

3 Results

3.1 Validation of extracted total RNA and cDNA synthesis

This work validated extracted total RNA using qualitative and quantitative nucleic acid detection techniques. High brightness and clarity in the 28S and 18S stripes and the absence of a trailing degradation characteristic indicate intact RNA. The Nanodrop spectrophotometer results from the ISOLATE II Plant extraction Kit showed total RNA isolated from doum palm leaves, A260/A230 between 1.92 and 2.13, and concentrations ranging from 579.61 to 1601.4 ng/μL

3.2 Sequence data summary and *de novo* assembly

The raw reads from the sequenced doum palm leaves were uploaded to the Genbank portal at NCBI with the accession number PRJNA861903. The raw data generated was approximately 39.97 Gigabytes (Gb). The quality control (QC) results based on an average sequencing error rate of 0.01% showed that the sequence data had a median of 96.41%, 88.76%, and 42.83% for Q20, Q30, and GC content, respectively (Table 2).

De novo assembly with Trinity yielded 193,167 unigenes with an average length of 1655 bp and an N50 of 3165 bp (Table 3 and S2 Fig). The results of the NR database function annotation allowed us to determine the species proportion on Unigene annotation and created a distribution map, which showed that 51.8% of the annotations were for *Phoenix dactylifera*, 39.19% were for *Elaeis guineensis*, and 1.08% were for *Ananas comosus*, with the remaining 7.85% being for other species (S1 Fig).

3.3 Doum palm annotation matrices with public databases

We aligned our sequence data in various databases, including the protein sequence (NR), Nucleotide Sequence Database (NT), KEGG, Swiss-Prot, Interpro GO, and eukaryotic

Table 2. Clean reads quality data.

Sample	C	D	J	K	M	N	Average
Total Raw Reads (M)	50.83	50.83	50.83	50.83	52.59	50.83	51.12
Total Clean Reads (M)	44.11	44.18	44.06	44.34	45.27	44.47	44.41
Total Clean Bases (Gb)	6.62	6.63	6.61	6.65	6.79	6.67	6.66
Clean Reads Q20 (%)	96.47	96.52	96.32	96.25	96.46	96.41	96.41
Clean ReadsQ30 (%)	88.94	89.08	88.52	88.36	88.9	88.73	88.76
Clean Reads Ratio (%)	86.77	86.91	86.68	87.23	86.08	87.48	86.86
GC (%)	42.91	42.92	42.94	42.22	42.96	43	42.83

D, J, and M are control samples; C, K, and N are stress samples

<https://doi.org/10.1371/journal.pone.0292543.t002>

Table 3. *De novo* assembly of clean reads.

Sample	Total Number	Total Length	Mean Length	N50	N70	N90
C	241,065	202,188,614	838	1,571	829	306
D	262,476	215,137,283	819	1,590	803	293
J	249,710	206,523,523	827	1,594	816	297
K	288,674	244,491,248	846	1,647	847	302
M	274,264	218,897,862	798	1,575	775	281
N	244,877	197,684,025	807	1,577	792	287
All Unigene	193,167	319,715,115	1,655	3,167	1,997	759

N50: Is a weighted median statistic that 50% of the total length is contained in Unigenes that are equal to or larger than this value

N90: The length for which the collection of all contigs of that length or longer contains at least 90% of the sum of the lengths of all contigs

<https://doi.org/10.1371/journal.pone.0292543.t003>

orthologous groups (KOG) for homologs search of the assembled unigenes (Table 5 and S3 Fig). There were 128,708 unigenes effectively annotated with at least one database, including 115,388 (73%), 122,767 (63.55%), 75,426 (39.05%), 91,401 (47.32%), 89,403 (46.28%), 92,118 (47.69%) and 23,074 (11.95%) that were significantly matched with the NR, NT, Swissprot, KEGG, KOG, Interpro and GO databases (Table 4 and S3 Fig).

3.3.1 Functional KEGG annotation in doum palm. The KEGG documented five of the seven critical metabolic pathway components (KEGG) in the doum palm transcriptome analysis. Functional categories of the Unigenes studied here included cellular activities, environmental information processing, genetic information processing, metabolism, and organismal systems (Fig 1). Analysis of the doum palm transcriptome via the KEGG showed that "metabolism" and "genetic information processing" accounted for the most significant categories. Our initial impression is that the sequencing and assembly methods employed here are of the highest quality. In the metabolism category, "global and overview maps" accounted for 18180 genes, while carbohydrate metabolism was followed by coding for 7079 genes. The genetic information processing category coded 7189 and 5677 genes representing translation and "folding, sorting and degradation," respectively (Fig 1).

3.3.2 Eukaryotic Orthologous Groups (KOG) functional classification in doum palm. Fig 2 shows the distribution of annotated Unigenes in the KOG database according to the 25

Table 4. Function annotation summary.

Values	Total	NR	NT	Swissprot	KEGG	KOG	Interpro	GO	Intersection	Overall
Number	193,167	115,388	122,767	75,426	91,401	89,403	92,118	23,074	12,760	128,708
Percent	100%	59.73%	63.55%	39.05%	47.32%	46.28%	47.69%	11.95%	6.61%	66.63%

Intersection: The total amount of Unigenes annotated

Overall: Quantity of annotated Unigenes

<https://doi.org/10.1371/journal.pone.0292543.t004>

Table 5. Doum palm predicted CDS quality.

Total #	Total length	N50	N90	Max Length	Min Length	Sequence (GC%)
94741	108,239,283	1,476	567	15,321	279	46.41%

Total number: the number of CDS

<https://doi.org/10.1371/journal.pone.0292543.t005>

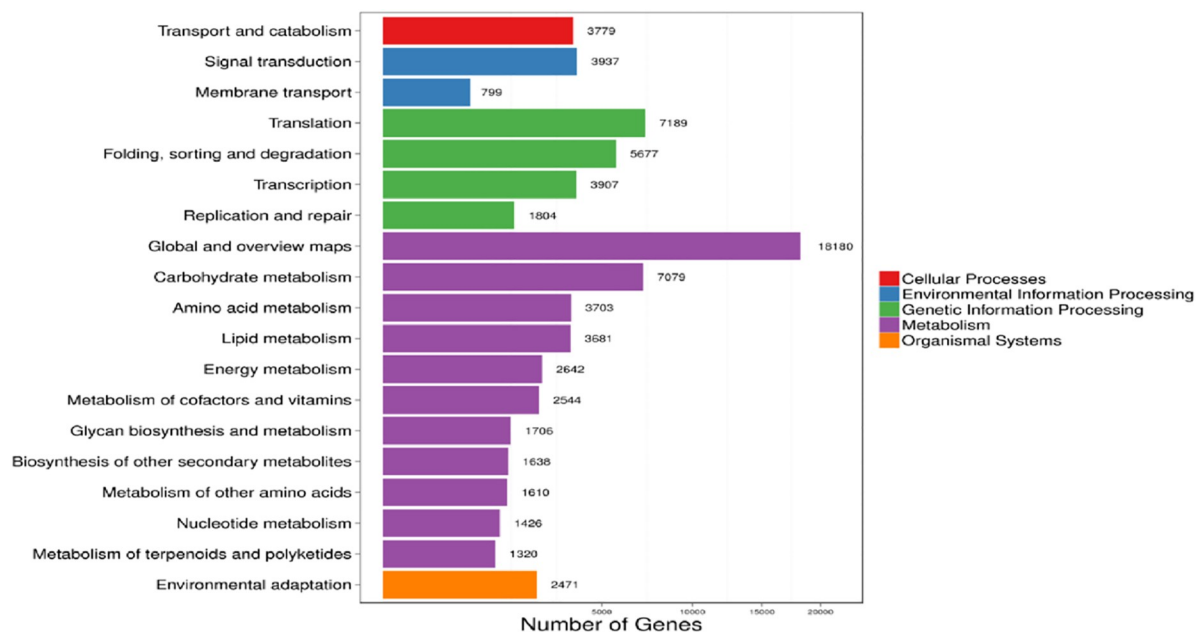


Fig 1. KEGG annotation functional distribution. Horizontal and vertical bars indicate the sum of unigenes and KEGG functional grouping, respectively. Different color represents five of the seven categories of the KEGG metabolic pathway.

<https://doi.org/10.1371/journal.pone.0292543.g001>

functional categories. The functional distribution result shows that "general function prediction (23425 genes)", "signal transduction mechanisms (16168 genes)", "posttranslational modification, protein turnover, chaperones (9804 genes)", "unknown function (9438 genes)", and "transcriptional groups (9313 genes)" were abundantly annotated among the 25 functional groups in the KOG database. Nuclear structure and cell motility accounted for the least annotated functional group, with 992 and 219 genes, respectively.

3.3.3 Doum palm gene ontology (GO) functional classification and predicted CDS quality. Following these quality checks, the unigenes were annotated with biological roles. Regarding biological function, the bulk of the annotated unigenes was found to be linked with the cellular process (9025 genes), metabolic process (8391 genes), biological regulation (2327 genes), and regulation of biological processes (2173 genes). After cells and cell parts coded 9889 and 9693 genes, respectively, the membrane ranking third coded for 8038 genes while the membrane part ranked fourth coded for 7360 genes ranked among cellular components. Most unigenes exhibited binding (10,215 genes) when tested for molecular function and catalytic activity (10097 genes). Additionally, 1292 and 925 genes were involved in transporter and structural molecule activity, respectively. Fig 3 presents the GO term functional classification.

Most DEGs were down-regulated, as evidenced by the enrichment of GO terms in the D control sample set against drought samples (Fig 4). The most down-regulated biological categories were binding and catalytic activity, cellular component and metabolic processes, and cells and their parts.

In the biological category, the most up-regulated expression were those connected to cells and their constituent parts; in the cellular component category, the utmost up-regulated processes were those linked to cells and their constituents; and in the cellular activity and cellular component categories, respectively, the most up-regulated processes were those related to catalytic activity and binding (Fig 4). The CDSs result of the unigene is presented in Table 5.

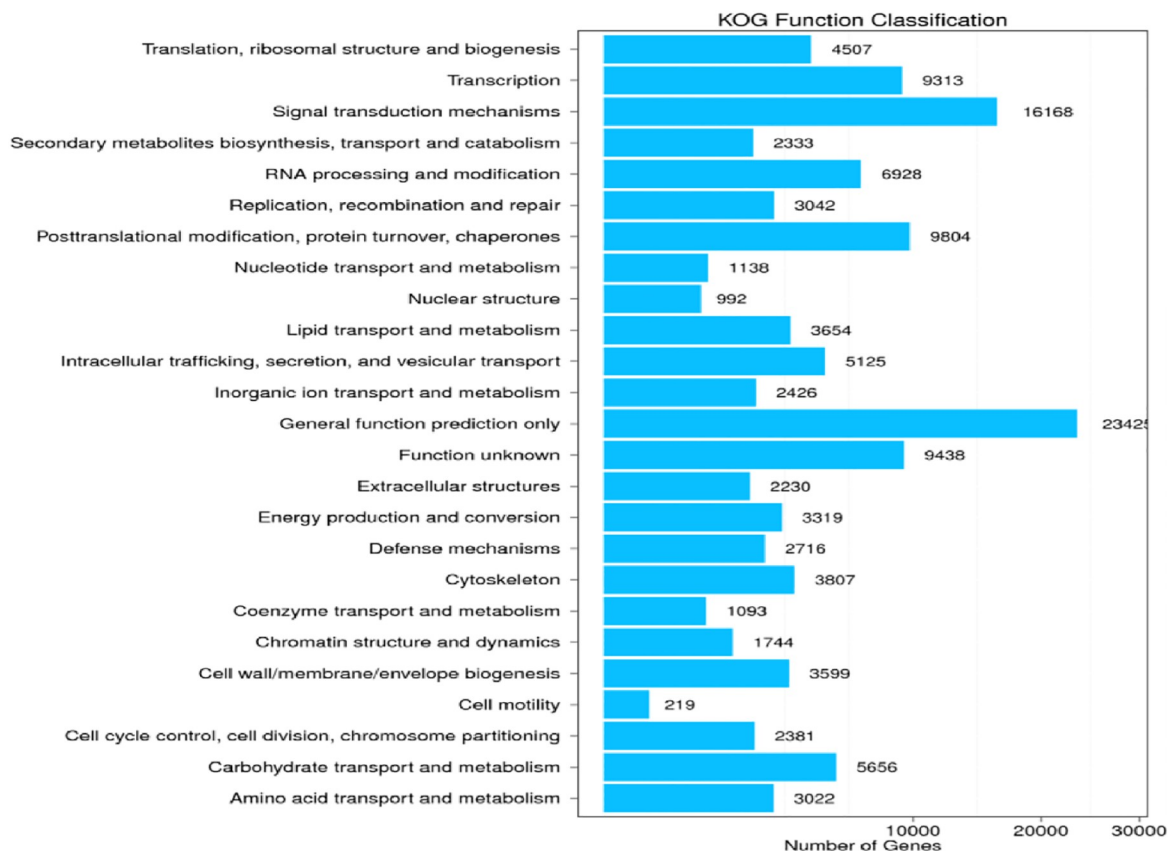


Fig 2. KOG functional classification. The abscissa and vertical coordinate depict the amount of Unigenes and KOG functional categories, respectively.

<https://doi.org/10.1371/journal.pone.0292543.g002>

3.4 Classification of TFs in doum palm and identified SSR

3.4.1 TFs in doum palm. The DEGs analysis of doum palms under well water and drought stress circumstances revealed the presence of fifty-eight (58) TFs families. The leading group of TFs was the MYB and MYB-related family, accounting for 683 and 502 genes, respectively, followed by FAR1 (378), C3H (315), bHLH (256), and WRKY (242), whereas other TFs families identified were those belonging to the AP2-EREBP (230), NAC (202), G2-like (176), and ARF (171) (S5 Fig). The MYB and MYB-related family accounted for the highest TF in doum palm, while HRT, GRF, and NOZZLE families were identified as the least. S5 Fig depict this study's identified TFs in doum palms.

3.4.2 Identified SSRs in doum palm. The popularity of SSR markers can be attributed to their ease, speed, low cost, and reproducibility. Based on the sequences of the functional genes of interest, SSR markers can be used in molecularly-assisted selection breeding, allowing for the investigation of genetic variation without identifying biases that arise in non-gene areas [30]. The majority of SSRs detected were dinucleotides (41.28%) and mononucleotides (34.1%), followed by trinucleotide and pentanucleotide repeats accounting for 20.64% and 1.79%, respectively (S6 Fig). The AG/CT repeat was the most prevalent SSR among mononucleotide repeats, followed by the AT/AT repeat. AGG/CCT and AAG/CTT accounted for the most common repeats in the dinucleotide type.

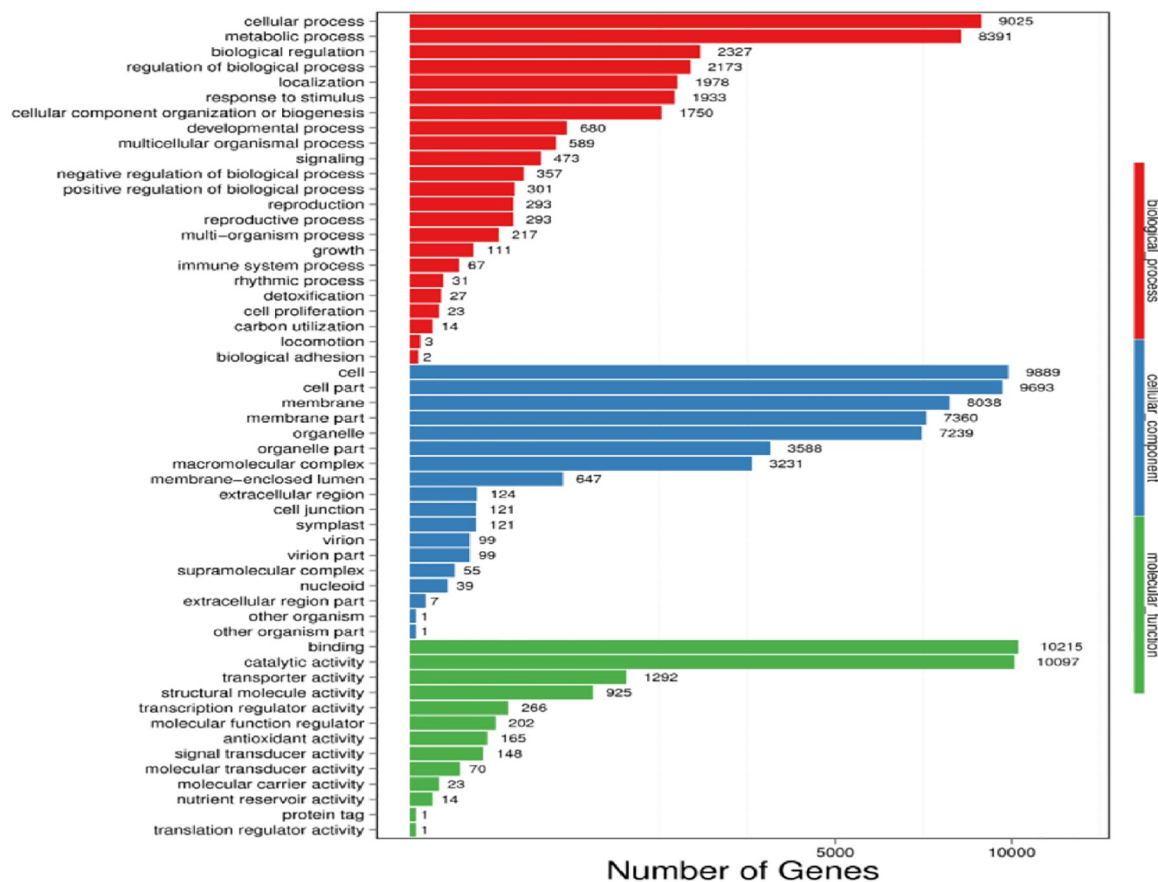


Fig 3. Functional classification of GO term. The abscissa and vertical coordinate depict the amount of Unigenes and GO term functional categories, respectively. Different color represents categories of GO term function.

<https://doi.org/10.1371/journal.pone.0292543.g003>

3.5 Gene expression and DEGs analysis

Using the FPKM method, we calculated the relative expression of protein-coding genes [31]. Also, the expression levels were used to rank the six samples into three groups: those with high, medium, and low levels (FPKM ≥ 10 , FPKM 10, and FPKM 10). The control sample (M) (8515 genes) and stress sample (N) (8486 genes) had the most highly expressed genes, according to an analysis of the genes with FPKM ≥ 10 (S7 Fig).

Differential expression of genes between well-watered and stressed-treated doum palms was analyzed to help us find novel candidate genes involved in water scarcity stress resistance. There were 87528 DEGs found between the control and tree treatment groups out of a total of 193,167 unigenes (Fig 5). Up-regulated DEGs accounted for 45071 unigenes, and 42457 unigenes accounted for down-regulated DEGs.

3.6 qPCR results validating the RNA-Seq data

Four DEGs were arbitrarily chosen for relative quantitative analysis to validate the RNA sequencing data. These selected DEGs exhibited considerable up-regulation of all selected DEGs in the RNA-Seq study. Drought-treated plants exhibited strong upregulation of the four unigenes after the qPCR assay. Figs 6A, 6B, 7A, and 7B display that qPCR and RNA sequencing data showed consistent patterns (up-regulation) of unigene expression.

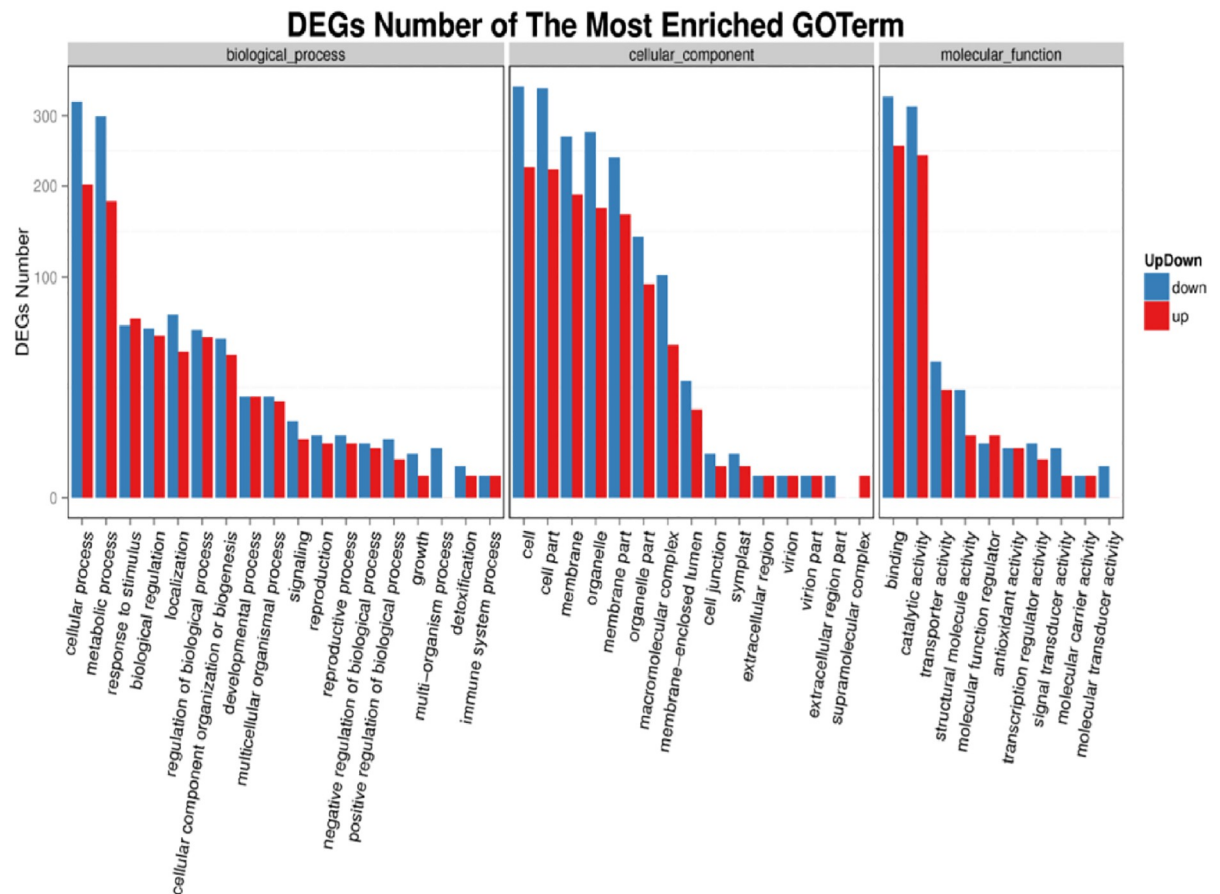


Fig 4. Doum palm GO terms categorized up-and down-regulated genes. The abscissa and vertical coordinate depict the GO term and how the genes were regulated, respectively.

<https://doi.org/10.1371/journal.pone.0292543.g004>

4 Discussion

4.1 Perspectives on the doum palm de novo transcriptome assembly and annotation

Several genes in a plant regulate and modulate the complicated biological process known as drought resistance. The cellular response processes to water deficit stress were investigated in the doum palm, a plant without a reference genome, utilizing an all-encompassing approach. Genomic and transcriptomic data for doum palms are still missing to provide a complete picture of the plant's genetic history or functional genes. As a result, there are many unannotated unigenes; in the current analysis, just 66.63 percent of all unigenes had annotations in any database. Annotation results from filtered reads used to sequence reported species' transcriptomes revealed that 83.48% of unigenes were present in public databases. This includes transcripts from common vetch [32]. Compared to data on common buckwheat, which had 39.33% of unigenes annotated in the GO catalog, the fraction of annotated genes in the doum palm genome is significantly more significant [33]; likewise, the *Taxus* species had 8426 and 12,642 annotated unigenes in the KEGG and GO term databases, respectively [32]. These findings suggest that a substantial amount of transcriptome data still needs further studies for organisms that do not contain a reference genome, such as the doum palm. The unmatched

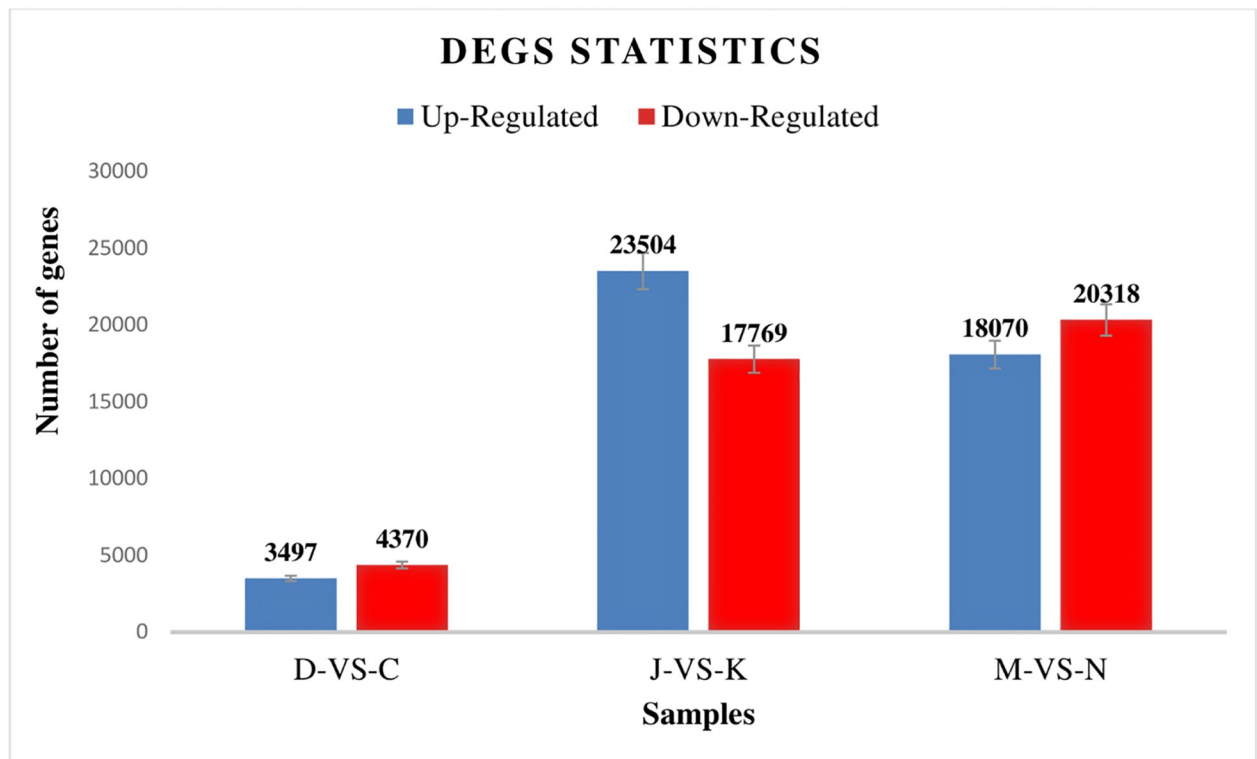


Fig 5. Genes differentially expressed in doum palm. NB: D, J, and M are well water samples; C, K, and N are stress-treated samples.

<https://doi.org/10.1371/journal.pone.0292543.g005>

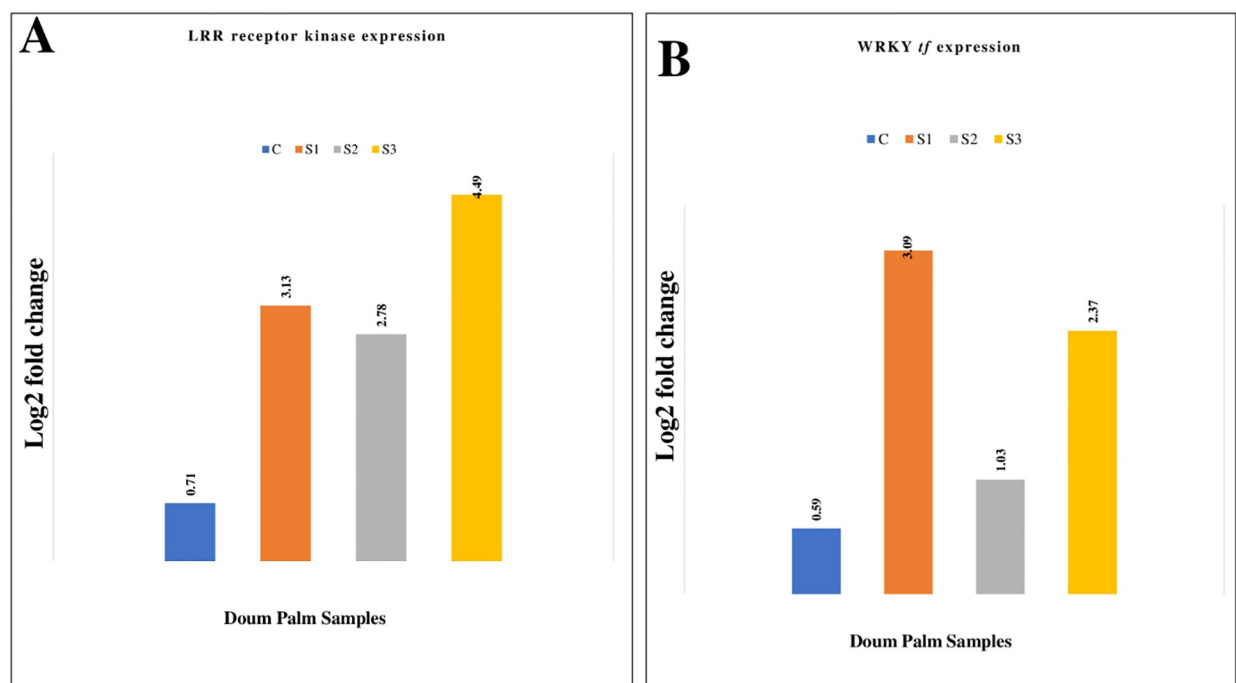


Fig 6. (A) LRR receptor-like gene and (B) WRKY tf gene expression in doum palm. Relative quantifications of leucine-rich repeats (LRR) receptor-like serine/threonine-protein kinase and WRKY transcription factor genes in 65 days stressed doum palm samples (Fold change in log 2 ratio). NB: C = control sample, S1 = stressed sample1, S2 = stressed sample 2, and S3 = stressed sample 3.

<https://doi.org/10.1371/journal.pone.0292543.g006>

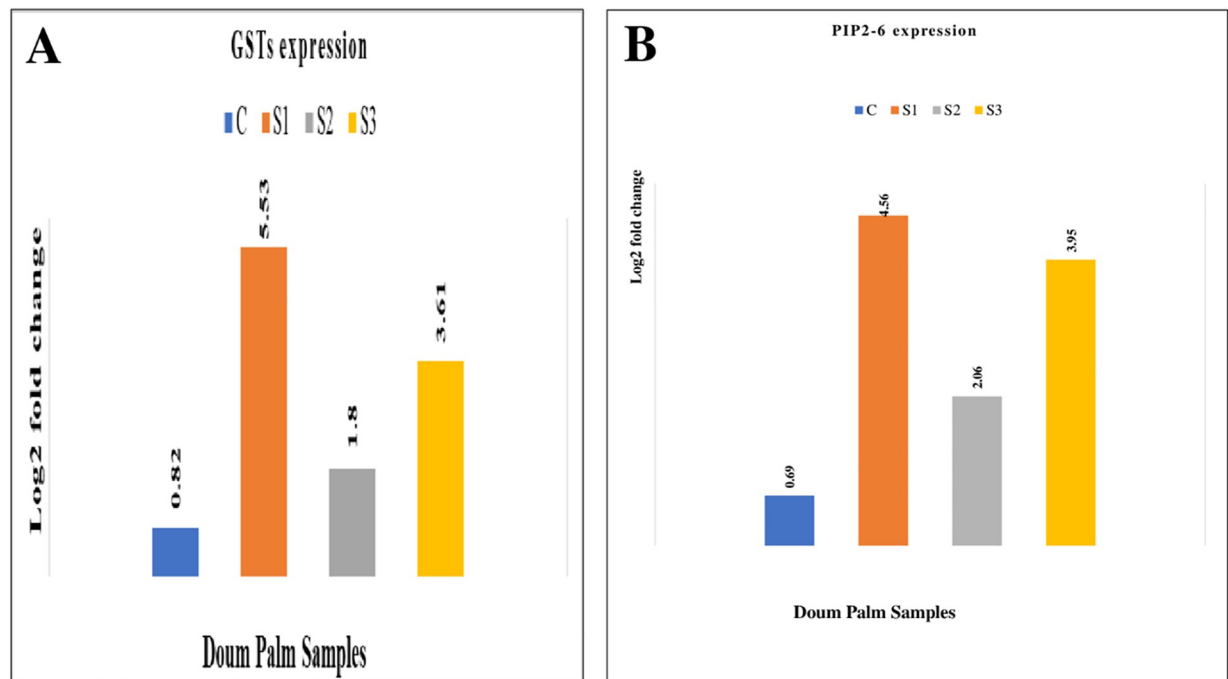


Fig 7. (A) GSTs gene and (B) PIP2-6 gene expression in doum palm. Relative quantifications of glutathione S-transferases T1 (GSTs) and aquaporins PIP2-6 genes in 65 days stressed doum palm samples (Fold change in log 2 ratio). NB: C = control sample, S1 = stressed sample1, S2 = stressed sample 2, and S3 = stressed sample 3.

<https://doi.org/10.1371/journal.pone.0292543.g007>

unigenes probably represent doum palm-specific genes. Transcriptome outcomes from this work were similar to those from other previous transcriptomic investigations of crop species, including maize [34], mulberry (*Morus* L.), [35], and *Ammopiptanthus mongolicus* [36].

The species with the best hits in the annotated sequences were the date palm (*Phoenix dactylifera*), the oil palm (*Elaeis guineensis*), and the *ananas comosus*. Doum palm is the nearest species to date, and oil palm with sequenced genomes belongs to the *Arecaceae* family. Therefore, these findings are consistent. The genomic sequences for date palms and oil palms have been publicly available since 2013 [37, 38]. It is anticipated that when the doum palm genome, a member of the family *Arecaceae*, is accessible, a comparison analysis would further illuminate the differences in their features and the evolutionary links.

4.2 KOG functional classification analysis

The most significant fraction of genes in the KOG analysis were classified as having "general function prediction," "signal transduction mechanisms," "posttranslational modification," "protein turnover," and "chaperones" categories. These data support the findings of [39], who tackled the difficulties associated with *de novo* sequence analysis of the sweet potato genome. Stress from drought triggers a cascade of molecular activities in plants, including signal transduction, the cellular process that responds to a physical or chemical stimulus [40, 41]. When a trigger is detected, the plant responds by changing the enzyme activity of proteins, striking an equilibrium between protein production and breakdown [42]. Chaperone proteins, which facilitate protein folding as a chemical process due to extreme environmental stress, also help prevent or repair damage caused by misfolded proteins [43]. Recent research has divided drought-inducible genes into two groups: functional proteins and regulatory proteins. Proteins

such as those found in the late embryonic abundant (LEA) fraction, chaperones, dehydrin, detoxifying enzymes, and osmolytes mediate abiotic stress tolerance; on the other hand, a network of regulatory proteins controls genes involved in drought resistance [43].

4.3 GO term functional category reveal in doum palm under stress

In agreement with several prior studies [44], transcripts associated with "binding" and "catalytic activity" were over-represented under the Molecular function category of Gene Ontology, as revealed in the current study. Gene interaction is a common and crucial phenomenon that controls many biological processes. In response to drought stress, plants may express downstream target genes better or accumulate microRNAs if specific genes are activated. Therefore, gene binding activity regulates plant stress responses [45].

Generally, there was more than one unigene present in a given category. The two main categories in the cellular component group were "Cell" and "Cell part." This finding conforms to that of [46], in which Cell and Cell components accounted for the most significant functional category of unigenes annotated in the GO database. Similar GO terms in retort to external environmental factors have also been observed in chickpea transcriptome investigations [45, 47, 48].

4.4 Doum palm KEGG pathway analysis for DEGs

KEGG enrichment analyses on DEGs revealed that several differentially expressed genes were most strongly enriched, including cellular activities, environmental information processing, genetic information processing, metabolism, and organismal systems, indicating various functional activities/processes in leaves of doum palm under stress. These annotations are an excellent resource for learning about plants, mainly when cultivating climate-smart crops [35]. The regulation of multiple critical metabolic processes is intrinsically linked to the KEGG pathways enrichment analysis of doum palm, as similar findings were observed in rice [49, 50]. Our results also corroborated the findings of prior studies showing that environmental adaption was enhanced after water stress treatment [51, 52].

4.5 Drought-induced increases TFs activities in doum palm

In response to stressful conditions, plants use transcription regulation to temporarily and spatially regulate the expression of their target transcription genes [53, 54]. According to recent research [55], TFs significantly contribute to plant drought resistance by controlling the transcription of downstream genes to increase plant stress tolerance. This investigation will help researchers pinpoint and categorize TFs, providing a deeper insight into the molecular processes behind doum palms' drought tolerance. The leading TF families of doum palm, which include the myeloblastosis (MYB) and MYB-related, WRKY, NAC family (NAM, ATAF, and CUC), FAR-RED-IMPAIRED RESPONSE1 (FAR1), B3, and bZIP, were represented by DEGs encoding TFs. Many of these identified TFs families have been linked to improved plant stress resistance [56]. Accurate modulation of gene activities by TFs is essential for many plant functions, such as responding to environmental stresses [57, 58]. Multiple transcription factor genes, including ethylene-responsive element binding protein, bZIP, the myelocytomatosis tfs, NAC, and WRKY, have been associated with plant reactions to stress [59].

The current investigations are comparable to those of [60], who conducted a DEGs analysis on drought-stressed Giant Juncao and discovered that 56 TFs families were represented. In retort to a drought signal, the plant initiates a series of signaling cascades that ultimately leads to the activation of a TF. Furthermore, in response to a water deficit, an active TF attaches to a cis-acting site in the interested gene, which then triggers transcription of a specific response

gene further down the gene's regulatory pathway [56]. This research uncovered various transcription factors with varying expression levels crucial for withstanding drought stress.

The MYB-related class of TF is among the numerous plant TFs and is essential for coping with various adversities such as drought, light signaling, and stress. Increased activity of MYB family genes in transgenic rice, as seen with OsMYB6, improves the plant's resilience to water stress [61]. According to the research of [62], the MYB and MYB-related families can regulate drought tolerance in plants.

Numerous studies have linked WRKY TF to drought tolerance [63]. At least 17 rice WRKY genes were triggered by drought using massively parallel signature sequencing (MPSS) technology [64]. For example, [65] discovered that transgenic tobacco plants expressing TaWRKY44, a member of the WRKY TFs, were extra tolerant to water deficit and salt than non-transgenic tobacco crops. These outcomes indicate that TaWRKY44 may have a beneficial regulatory function in retorts to drought, salt, and osmotic conditions by expediting the effective removal of ROS by promoting the production of genes involved in dealing with stress. Similarly, the expression of CsWRKY2, like that of other WRKY TFs, was shown to be up-regulated in response to exogenous ABA, indicating that this gene contributes to tea plants' ability to withstand water stress [66].

Identification and investigation of TF types in plants and other organisms benefit significantly from de novo analysis. Most bHLH family members have been linked with water deficit conditions [67]. For instance, the rice, maize, and wheat genomes have 183, 23,1, and 571 members of the bHLH family, respectively [54], while *Arabidopsis* has 162, Chinese cabbage has 230, *Brachypodium distachyon* has 146, and apple has 175 [67, 68]. And therefore, breeding crops to thrive in water-poor environments would benefit significantly from a better understanding of which TFs are involved in this process and which genes they regulate.

4.6 Generic SSR makers in doum palm under stress

Research on the transcriptome can reveal genetic markers, making them an excellent choice for molecular cultivation. Due to their rapid amplification levy and good transferability between organisms, markers derived from the transcriptome have proven superior to those derived from non-transcribed areas [69]. Although single-nucleotide polymorphism (SNP) markers are best for understanding trait architecture [70], simple sequence repeats (SSRs) markers and other genes/length polymorphism-based markers are preferable for breeding applications. Comparable to what was described in previous plant research [71, 72], the preponderance of SSRs identified in doum palms were dinucleotides (41.28%), followed by mononucleotides (34.1%). Our findings are consistent with [73] *de novo* transcriptome assembly and gene quantification investigation in several moth bean tissues (*Vigna aconitifolia* Jacq.). Previous research has also reported on the function of SSRs in counterpoise to various abiotic conditions [74, 75].

4.7 Doum palm gene expression via transcriptomic profiling

Due to variations in the number of expressed genes and the level of gene activities in samples, the sample expression value of FPKM can be divided into various intervals. Statistics on gene expression levels are calculated and presented as stacked histograms of data from multiple time points. When [76] analyzed the groundnut transcriptome globally, they found that transcription factors, biological activities, and related pathways were similarly distributed across the genome, in line with our findings.

4.7.1 Analysis of selected doum palm transcripts based on their response to drought.

The selected transcripts in the qPCR analysis include leucine-rich repeats (LRR) receptor-like

serine/threonine-protein kinase, WRKY transcription factor, glutathione S-transferases T1 (GSTs), and aquaporins PIP2-6, were found to be significantly up-regulated in doum palm based on the expression analysis data. These genes/gene products or regulatory sequences have been linked to improved drought tolerance using a combination of in-depth literature reviews and extensive searches of publicly available databases.

LRR-RLKs are involved in numerous processes vital to plant life, including growth and development and detecting signals relevant to retorts to abiotic conditions [40]. For example, LRR-RLKs are crucial in controlling how plants react to water deficit and salinity conditions [77, 78]. In another example, drought, osmotic stress, auxin, and abscisic acid stimulated the *Arabidopsis* LRR-II type RLK genes. These genes affect the crop's response to external environmental influences [79].

The expression of rice LRR-RLKs (OsSIK1, OsGIRL1, OsLP2) was regulated in response to a wide range of stresses, including salinity, water deficit, abscisic acid, salicylic acid, and hydrogen peroxide, suggesting that these genes may function in various signaling networks that drive development and stress responses [15, 80]. Multiple studies have revealed that LRR-RLKs respond to drought by increasing their expression levels. In response to water deficit and salt treatment, for example, the expression of PnLRR-RLK27 in the Antarctic moss *Pohlia nutans* significantly up-regulated transcript levels of essential transcription factors and stress-related genes [81]. The Pkinases GhMKK1 and GhMKK3 have also been linked to cotton's resilience to salinity and water deficit [82].

The WRKY protein family is among plants' most numerous transcription factors (TFs). A highly conserved WRKY domain made up of the DNA-binding heptapeptide WRKYGQK, and the zinc finger binding motif spanning roughly 60 amino acids in length is a defining feature of WRKY TF [83]. WRKY is a member of the transcription factor superfamily and is involved in numerous biological activity processes during the plant's development and growth. Still, it is crucial in the defense reaction to harsh environmental conditions by modulating the transcription of genes involved in responding to those stresses through phytohormones. Moreover, extensive research shows that WRKY enhances the plant's retort to water shortage conditions [84–86].

The expression of WRKY TFs varies in response to environmental factors like salinity, water deficit, pathogen inoculation, the application of growth hormones, and oxidative stress, according to several scientific investigations utilizing transcriptome analysis [63]. Consistent with our findings, [87] demonstrated increased expression of 15 EgWRKYs TFs gene candidates in oil palms after drought, salinity, and heat exposure. Water-stress-induced upregulation of numerous transcription factors (TFs) may result in higher expression of genes of interest. [88] demonstrated that co-expression of AtWRKY28, a member of the WRKY TFs family, boosted the activities of numerous genes of interest in *Arabidopsis* in response to abiotic conditions such as drought, salt, and oxidative stress. To a large extent, WRKY TFs appear to regulate water deficit conditions via altering the osmotic equilibrium, ROS sifting, and the expression of several drought-related genes [89]. WRKY TF's potential led to its selection as one of the candidate genes for sequencing data.

Plants utilize many genes in water status maintenance in response to various environmental stresses. Many different cellular processes in plants also depend on glutathione S-transferases (GSTs), an enzyme family with many different roles, including shielding cells from oxidative stress damage and scavenging ROS due to water deficit duress [90–92]. ThGSTZ1 transient increased expression in transgenic *T. hispida*, for instance, elevated GST and GPX activities and boosted ROS scavenging abilities in response to NaCl and mannitol treatments [91]. These findings, consistent with the current study's expression pattern, indicate that GSTs can increase plant resistance to drought and salinity by boosting ROS scavenging.

Many plant species have undergone genome-wide functional characterization of the glutathione S-transferases (GSTs) gene family. A recent study published by [93] found that cold, heat, drought, salt, and osmotic stress led to significant upregulation in pepper GST transcript expression and overall pepper GST activity, further supporting the current finding. As a result of its demonstrated association with several pathways involved in the control of plant stress, doum palm RNA-Seq data validation with GST transcripts was of significant interest.

As a subfamily of the MIP superfamily, aquaporins (AQPs) create channels in membranes that are permeable to water and other tiny neutral substances. It has been established that abiotic stress causes plants to up-regulate AQPs [94–96]. The functional characterization of AQPs in many crop species revealed that this gene had improved resistance to abiotic stress [97, 98]. It is well established that the AQP family, of which the PIP proteins are a subset, plays crucial roles in water transport across the plasma membranes of many different crop species [99, 100].

In line with the findings of this study, the PIP2-6 genes showed significant expression under increased water duress in a prior investigation [101]. Similar results were found in earlier publications on chickpeas [102], which examined variations in water transport networks in the root apoplast and found that PIP2-6 gene expression was similarly significantly elevated. Considering the possible role of the aquaporin (AQP) genes family in harsh environmental condition resistance, specifically in its active involvement in water transport and desiccation resistance, the aquaporins PIP2-6 were chosen for use in confirming the RNA-Seq data.

5 Conclusion

Here, we present the first transcriptome data from doum palms that allowed us to characterize the genes involved in their response to drought. *De novo* assembly yielded 93,907 unigenes, and seven functional databases provided annotations for 57,941 genes. Deciphering the response mechanism of doum palms under drought stress can be aided by studying the expression of DEGs and finding transcription factors/genes related to drought. This study sheds light on the genetics of drought tolerance adaptation in doum palms and identifies prospective genetic resources for abiotic stress resistance in plant development. From a drought tolerance perspective, this study shows that doum palm resources can be exploited as a breeding crop that can thrive well during drought stress. This study has made the first comprehensive molecular/transcriptome resources available to the scientific community to facilitate increasing studies on doum palm species.

Supporting information

S1 Fig. NR annotation indicating the range of species.

(DOCX)

S2 Fig. Total length of doum palm assembled unigene.

(DOCX)

S3 Fig. Annotation matrices for NR, KOG, KEGG, Swissprot, and Interpro.

(DOCX)

S4 Fig. The range of doum palm CDS lengths.

(DOCX)

S5 Fig. Doum palm transcription factor (TF) family classification.

(DOCX)

S6 Fig. Detected SSRs in doum palm.
(DOCX)

S7 Fig. Gene expression distribution per sample.
(DOCX)

Acknowledgments

The authors are grateful to Dr. Cecilia Mbithe Mweu and Ms. Agnes Omire for assisting in collecting doum palm seeds. For this, we are grateful.

Author Contributions

Conceptualization: Allen Johnny Borlay, Cecilia Mbithe Mweu, Kevin Mbogo Omolo.

Data curation: Allen Johnny Borlay, Labode Hospice Stevenson Naitchede.

Formal analysis: Allen Johnny Borlay, Labode Hospice Stevenson Naitchede.

Funding acquisition: Allen Johnny Borlay.

Investigation: Allen Johnny Borlay.

Methodology: Allen Johnny Borlay, Kevin Mbogo Omolo.

Project administration: Allen Johnny Borlay.

Resources: Allen Johnny Borlay, Kevin Mbogo Omolo, Labode Hospice Stevenson Naitchede.

Software: Allen Johnny Borlay, Steven Ger Nyanjom, Labode Hospice Stevenson Naitchede.

Supervision: Cecilia Mbithe Mweu, Steven Ger Nyanjom.

Validation: Allen Johnny Borlay.

Visualization: Allen Johnny Borlay, Steven Ger Nyanjom.

Writing – original draft: Allen Johnny Borlay.

Writing – review & editing: Allen Johnny Borlay, Cecilia Mbithe Mweu, Steven Ger Nyanjom, Kevin Mbogo Omolo, Labode Hospice Stevenson Naitchede.

References

1. Hein JA, Sherrard ME, Manfredi KP, Abebe T. The fifth leaf and spike organs of barley (*Hordeum vulgare* L.) display different physiological and metabolic responses to drought stress. *BMC Plant Biol.* 2016; 16: 1–12.
2. Jones C, de Vega J, Worthington M, Thomas A, Gasior D, Harper J, et al. A comparison of differential gene expression in response to the onset of water stress between three hybrid *Brachiaria* genotypes. *Front Plant Sci.* 2021; 12: 393.
3. Chaves MM, Oliveira MM. Mechanisms underlying plant resilience to water deficits: prospects for water-saving agriculture. *J Exp Bot.* 2004; 55: 2365–2384. <https://doi.org/10.1093/jxb/erh269> PMID: 15475377
4. Farooq M, Hussain M, Siddique KHM. Drought stress in wheat during flowering and grain-filling periods. *CRC Crit Rev Plant Sci.* 2014; 33: 331–349.
5. Gong Z, Xiong L, Shi H, Yang S, Herrera-Estrella LR, Xu G, et al. Plant abiotic stress response and nutrient use efficiency. *Sci China Life Sci.* 2020; 63: 635–674. <https://doi.org/10.1007/s11427-020-1683-x> PMID: 32246404
6. Maruyama K, Urano K, Yoshiwara K, Morishita Y, Sakurai N, Suzuki H, et al. Integrated analysis of the effects of cold and dehydration on rice metabolites, phytohormones, and gene transcripts. *Plant Physiol.* 2014; 164: 1759–1771. <https://doi.org/10.1104/pp.113.231720> PMID: 24515831

7. Amwatta CJM. Diversity of use of Doum palm (*Hyphaene compressa*) leaves in Kenya. *Palms*. 2004; 48: 184–190.
8. Aboshora W, Lianfu Z, Dahir M, Gasmalla MAA, Musa A, Omer E, et al. Physicochemical, nutritional and functional properties of the epicarp, flesh and pitted sample of doum fruit (*Hyphaene Thebaica*). 2014.
9. Cheikhyyoussef A, Embashu W. Ethnobotanical knowledge on indigenous fruits in Ohangwena and Oshikoto regions in Northern Namibia. *J Ethnobiol Ethnomed*. 2013; 9: 1–13.
10. Lokuruka MNI. Amino acids and some minerals in the nut of the Turkana doum palm (*Hyphaene coriacea*). *African Journal of Food, Agriculture, Nutrition and Development*. 2007; 7: 1–14.
11. Maurya RR, Qureshimatva UM, Gamit SB, Solanki HA. New Record to the Flora of Gujarat, India: *Ceropegia vincaefolia* Hook.(Family Asclapadaceae). *Indian Forester*. 2018; 144: 101.
12. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009; 10: 57–63. <https://doi.org/10.1038/nrg2484> PMID: 19015660
13. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb)*. 2011; 107: 1–15. <https://doi.org/10.1038/hdy.2010.152> PMID: 21139633
14. Tao X, Gu Y-H, Wang H-Y, Zheng W, Li X, Zhao C-W, et al. Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam.]. *PLoS One*. 2012; 7: e36234. <https://doi.org/10.1371/journal.pone.0036234> PMID: 22558397
15. Liu Z, Wang J, Yang F, Yang L, Yue Y, Xiang J, et al. A novel membrane-bound E3 ubiquitin ligase enhances the thermal resistance in plants. *Plant Biotechnol J*. 2014; 12: 93–104. <https://doi.org/10.1111/pbi.12120> PMID: 24102799
16. Bhardwaj AR, Joshi G, Kukreja B, Malik V, Arora P, Pandey R, et al. Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop Brassica juncea. *BMC Plant Biol*. 2015; 15: 1–15.
17. Zhang H, He L, Cai L. Transcriptome sequencing: RNA-seq. *Computational systems biology*. Springer; 2018. pp. 15–27.
18. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 2011; 29: 644.
19. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 2003; 19: 651–652. <https://doi.org/10.1093/bioinformatics/btg034> PMID: 12651724
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
21. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015; 12: 59–60. <https://doi.org/10.1038/nmeth.3176> PMID: 25402007
22. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro, progress and status in 2005. *Nucleic Acids Res*. 2005; 33: D201–D205. <https://doi.org/10.1093/nar/gki106> PMID: 15608177
23. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21: 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610> PMID: 16081474
24. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006; 34: W293–W297. <https://doi.org/10.1093/nar/gkl031> PMID: 16845012
25. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends in genetics*. 2000; 16: 276–277. [https://doi.org/10.1016/S0168-9525\(00\)00204-2](https://doi.org/10.1016/S0168-9525(00)00204-2) PMID: 10827456
26. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 2013; 41: e121–e121. <https://doi.org/10.1093/nar/gkt263> PMID: 23598997
27. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28: 511–515. <https://doi.org/10.1038/nbt.1621> PMID: 20436464
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15: 1–21. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
29. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *methods*. 2001; 25: 402–408.
30. Gao Y, Lu Y, Li X, Li N, Zhang X, Su X, et al. Development and application of SSR markers related to genes involved in leaf adaxial-abaxial polarity establishment in chinese cabbage (*Brassica rapa* L.

- ssp. pekinensis). *Front Genet.* 2020; 11: 773. <https://doi.org/10.3389/fgene.2020.00773> PMID: 32793286
31. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* 2011; 12: 1–14. <https://doi.org/10.1186/gb-2011-12-3-r22> PMID: 21410973
 32. Zhu Y, Liu Q, Xu W, Zhang J, Wang X, Nie G, et al. De novo assembly and discovery of genes that involved in drought tolerance in the common vetch. *Int J Mol Sci.* 2019; 20: 328. <https://doi.org/10.3390/ijms20020328> PMID: 30650531
 33. Wu Q, Zhao G, Bai X, Zhao W, Xiang D, Wan Y, et al. Characterization of the transcriptional profiles in common buckwheat (*Fagopyrum esculentum*) under PEG-mediated drought stress. *Electronic Journal of Biotechnology.* 2019; 39: 42–51.
 34. Waititu JK, Zhang X, Chen T, Zhang C, Zhao Y, Wang H. Transcriptome analysis of tolerant and susceptible maize genotypes reveals novel insights about the molecular mechanisms underlying drought responses in leaves. *Int J Mol Sci.* 2021; 22: 6980. <https://doi.org/10.3390/ijms22136980> PMID: 34209553
 35. Wang H, Tong W, Feng L, Jiao Q, Long L, Fang R, et al. De novo transcriptome analysis of mulberry (*Morus L.*) under drought stress using RNA-Seq technology. *Russ J Bioorg Chem.* 2014; 40: 423–432.
 36. Gao F, Wang J, Wei S, Li Z, Wang N, Li H, et al. Transcriptomic analysis of drought stress responses in *Ammopiptanthus mongolicus* leaves using the RNA-Seq technique. *PLoS One.* 2015; 10: e0124382. <https://doi.org/10.1371/journal.pone.0124382> PMID: 25923822
 37. Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, et al. Genome sequence of the date palm *Phoenix dactylifera L.* *Nat Commun.* 2013; 4: 1–9. <https://doi.org/10.1038/ncomms3274> PMID: 23917264
 38. Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature.* 2013; 500: 335–339. <https://doi.org/10.1038/nature12309> PMID: 23883927
 39. Isobe S, Shirasawa K, Hirakawa H. Challenges to genome sequence dissection in sweetpotato. *Breed Sci.* 2017; 16186. <https://doi.org/10.1270/jsbbs.16186> PMID: 28465666
 40. Osakabe Y, Osakabe K, Shinozaki K, Tran L-SP. Response of plants to water stress. *Front Plant Sci.* 2014; 5: 86. <https://doi.org/10.3389/fpls.2014.00086> PMID: 24659993
 41. Osakabe Y, Yamaguchi-Shinozaki K, Shinozaki K, Tran L-SP. Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. *J Exp Bot.* 2013; 64: 445–458. <https://doi.org/10.1093/jxb/ers354> PMID: 23307915
 42. Xiong L, Schumaker KS, Zhu J-K. Cell signaling during cold, drought, and salt stress. *Plant Cell.* 2002; 14: S165–S183. <https://doi.org/10.1105/tpc.000596> PMID: 12045276
 43. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Annu Rev Biophys.* 2008; 37: 289. <https://doi.org/10.1146/annurev.biophys.37.092707.153558> PMID: 18573083
 44. Garg V, Khan AW, Kudapa H, Kale SM, Chitikineni A, Qiwei S, et al. Integrated transcriptome, small RNA and degradome sequencing approaches provide insights into *Ascochyta* blight resistance in chickpea. *Plant Biotechnol J.* 2019; 17: 914–931. <https://doi.org/10.1111/pbi.13026> PMID: 30328278
 45. Garg R, Bhattacharjee A, Jain M. Genome-scale transcriptomic insights into molecular aspects of abiotic stress responses in chickpea. *Plant Mol Biol Report.* 2015; 33: 388–400.
 46. Dong Y, Fan G, Zhao Z, Deng M. Transcriptome expression profiling in response to drought stress in *Paulownia australis*. *Int J Mol Sci.* 2014; 15: 4583–4607. <https://doi.org/10.3390/ijms15034583> PMID: 24642880
 47. Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, Tuteja R, et al. Large-scale transcriptome analysis in chickpea (*Cicer arietinum L.*), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol J.* 2011; 9: 922–931. <https://doi.org/10.1111/j.1467-7652.2011.00625.x> PMID: 21615673
 48. Pradhan S, Bandhiwal N, Shah N, Kant C, Gaur R, Bhatia S. Global transcriptome analysis of developing chickpea (*Cicer arietinum L.*) seeds. *Front Plant Sci.* 2014; 5: 698. <https://doi.org/10.3389/fpls.2014.00698> PMID: 25566273
 49. Ithal N, Reddy AR. Rice flavonoid pathway genes, OsDfr and OsAns, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, OsC1-MYB. *Plant Science.* 2004; 166: 1505–1513.
 50. Lenka SK, Katiyar A, Chinnusamy V, Bansal KC. Comparative analysis of drought-responsive transcriptome in *Indica* rice genotypes with contrasting drought tolerance. *Plant Biotechnol J.* 2011; 9: 315–327. <https://doi.org/10.1111/j.1467-7652.2010.00560.x> PMID: 20809928
 51. Xu Y, Gao S, Yang Y, Huang M, Cheng L, Wei Q, et al. Transcriptome sequencing and whole genome expression profiling of *chrysanthemum* under dehydration stress. *BMC Genomics.* 2013; 14: 1–15.

52. Patel M, Fatnani D, Parida AK. Silicon-induced mitigation of drought stress in peanut genotypes (*Arachis hypogaea* L.) through ion homeostasis, modulations of antioxidative defense system, and metabolic regulations. *Plant Physiology and Biochemistry*. 2021; 166: 290–313. <https://doi.org/10.1016/j.plaphy.2021.06.003> PMID: 34146784
53. Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res*. 2014; 42: D1182–D1187. <https://doi.org/10.1093/nar/gkt1016> PMID: 24174544
54. Wei K, Chen H. Comparative functional genomics analysis of bHLH gene family in rice, maize and wheat. *BMC Plant Biol*. 2018; 18: 1–21.
55. Gonçalves LP, Boscardioli Camargo RL, Takita MA, Machado MA, dos Soares Filho WS, Costa MGC. Rootstock-induced molecular responses associated with drought tolerance in sweet orange as revealed by RNA-Seq. *BMC Genomics*. 2019; 20: 1–14.
56. Zhu CB, Xu CF, Liu RM, Lin ZZ, Cheng LR. Expression change of transcription factors of rice under drought stress. *Chin Agric Sci Bull*. 2019; 35: 108–114.
57. Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, et al. Transcription factors and plants response to drought stress: current understanding and future directions. *Front Plant Sci*. 2016; 7: 1029. <https://doi.org/10.3389/fpls.2016.01029> PMID: 27471513
58. Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front Plant Sci*. 2014; 5: 170. <https://doi.org/10.3389/fpls.2014.00170> PMID: 24904597
59. Gahlaut V, Jaiswal V, Kumar A, Gupta PK. Transcription factors involved in drought tolerance and their possible role in developing drought tolerant cultivars with emphasis on wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*. 2016; 129: 2019–2042. <https://doi.org/10.1007/s00122-016-2794-z> PMID: 27738714
60. Zhou J, Chen S, Shi W, David-Schwartz R, Li S, Yang F, et al. Transcriptome profiling reveals the effects of drought tolerance in Giant Juncao. *BMC Plant Biol*. 2021; 21: 1–20.
61. Tang Y, Bao X, Zhi Y, Wu Q, Guo Y, Yin X, et al. Overexpression of a MYB family gene, OsMYB6, increases drought and salinity stress tolerance in transgenic rice. *Front Plant Sci*. 2019; 10: 168. <https://doi.org/10.3389/fpls.2019.00168> PMID: 30833955
62. Xiong H, Li J, Liu P, Duan J, Zhao Y, Guo X, et al. Overexpression of OsMYB48-1, a novel MYB-related transcription factor, enhances drought and salinity tolerance in rice. *PLoS One*. 2014; 9: e92913. <https://doi.org/10.1371/journal.pone.0092913> PMID: 24667379
63. Tripathi P, Rabara RC, Rushton PJ. A systems biology perspective on the role of WRKY transcription factors in drought responses in plants. *Planta*. 2014; 239: 255–266. <https://doi.org/10.1007/s00425-013-1985-y> PMID: 24146023
64. Chen L, Song Y, Li S, Zhang L, Zou C, Yu D. The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2012; 1819: 120–128. <https://doi.org/10.1016/j.bbagr.2011.09.002> PMID: 21964328
65. Wang X, Zeng J, Li Y, Rong X, Sun J, Sun T, et al. Expression of TaWRKY44, a wheat WRKY gene, in transgenic tobacco confers multiple abiotic stress tolerances. *Front Plant Sci*. 2015; 6: 615. <https://doi.org/10.3389/fpls.2015.00615> PMID: 26322057
66. Luo X, Bai X, Sun X, Zhu D, Liu B, Ji W, et al. Expression of wild soybean WRKY20 in Arabidopsis enhances drought tolerance and regulates ABA signalling. *J Exp Bot*. 2013; 64: 2155–2169. <https://doi.org/10.1093/jxb/ert073> PMID: 23606412
67. Cui X, Wang Y-X, Liu Z-W, Wang W-L, Li H, Zhuang J. Transcriptome-wide identification and expression profile analysis of the bHLH family genes in *Camellia sinensis*. *Funct Integr Genomics*. 2018; 18: 489–503. <https://doi.org/10.1007/s10142-018-0608-x> PMID: 29651641
68. Mao K, Dong Q, Li C, Liu C, Ma F. Genome wide identification and characterization of apple bHLH transcription factors and expression analysis in response to drought and salt stress. *Front Plant Sci*. 2017; 8: 480. <https://doi.org/10.3389/fpls.2017.00480> PMID: 28443104
69. Barbara T, PALMA-SILVA C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol Ecol*. 2007; 16: 3759–3767. <https://doi.org/10.1111/j.1365-294X.2007.03439.x> PMID: 17850543
70. Nepolean T, Singh I, Hossain F, Pandey N, Gupta HS. Molecular characterization and assessment of genetic diversity of inbred lines showing variability for drought tolerance in maize. *J Plant Biochem Biotechnol*. 2013; 22: 71–79.
71. Magandhi M, Kusumo YWE, Matra DD. Development and characterization of Simple Sequence Repeats (SSRs) markers in durian kura-kura (*Durio testudinarius* Becc.) using NGS data. *IOP Conference Series: Earth and Environmental Science*. IOP Publishing; 2021. p. 012082.

72. Ul-Haq S, Kumar P, Singh RK, Verma KS, Bhatt R, Sharma M, et al. Assessment of functional EST-SSR markers (Sugarcane) in cross-species transferability, genetic diversity among poaceae plants, and bulk segregation analysis. *Genet Res Int*. 2016; 2016. <https://doi.org/10.1155/2016/7052323> PMID: 27340568
73. Suranjika S, Pradhan S, Nayak SS, Parida A. De novo transcriptome assembly and analysis of gene expression in different tissues of moth bean (*Vigna aconitifolia*)(Jacq.) Marechal. *BMC Plant Biol*. 2022; 22: 1–15.
74. Jiang B, Xie D, Liu W, Peng Q, He X. De novo assembly and characterization of the transcriptome, and development of SSR markers in wax gourd (*Benicasa hispida*). *PLoS One*. 2013; 8: e71054. <https://doi.org/10.1371/journal.pone.0071054> PMID: 23951078
75. Jiang Q, Wang F, Tan H-W, Li M-Y, Xu Z-S, Tan G-F, et al. De novo transcriptome assembly, gene annotation, marker development, and miRNA potential target genes validation under abiotic stresses in *Oenanthе javanica*. *Molecular Genetics and Genomics*. 2015; 290: 671–683. <https://doi.org/10.1007/s00438-014-0953-y> PMID: 25416420
76. Soni P, Pandey AK, Nayak SN, Pandey MK, Tolani P, Pandey S, et al. Global transcriptome profiling identified transcription factors, biological process, and associated pathways for pre-harvest aflatoxin contamination in groundnut. *Journal of Fungi*. 2021; 7: 413. <https://doi.org/10.3390/jof7060413> PMID: 34073230
77. Hwang S-G, Kim DS, Jang CS. Comparative analysis of evolutionary dynamics of genes encoding leucine-rich repeat receptor-like kinase between rice and *Arabidopsis*. *Genetica*. 2011; 139: 1023–1032. <https://doi.org/10.1007/s10709-011-9604-y> PMID: 21879323
78. Wu F, Sheng P, Tan J, Chen X, Lu G, Ma W, et al. Plasma membrane receptor-like kinase leaf panicle 2 acts downstream of the DROUGHT AND SALT TOLERANCE transcription factor to regulate drought sensitivity in rice. *J Exp Bot*. 2015; 66: 271–281. <https://doi.org/10.1093/jxb/eru417> PMID: 25385766
79. Chae L, Sudat S, Dudoit S, Zhu T, Luan S. Diverse transcriptional programs associated with environmental stress and hormones in the *Arabidopsis* receptor-like kinase gene family. *Mol Plant*. 2009; 2: 84–107. <https://doi.org/10.1093/mp/ssn083> PMID: 19529822
80. Ouyang S, Liu Y, Liu P, Lei G, He S, Ma B, et al. Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (*Oryza sativa*) plants. *The Plant Journal*. 2010; 62: 316–329. <https://doi.org/10.1111/j.1365-3113X.2010.04146.x> PMID: 20128882
81. Zhang L, Li Y, Lu W, Meng F, Wu C, Guo X. Cotton GhMKK5 affects disease resistance, induces HR-like cell death, and reduces the tolerance to salt and drought stress in transgenic *Nicotiana benthamiana*. *J Exp Bot*. 2012; 63: 3935–3951. <https://doi.org/10.1093/jxb/ers086> PMID: 22442420
82. Lu W, Chu X, Li Y, Wang C, Guo X. Cotton GhMKK1 induces the tolerance of salt and drought stress, and mediates defence responses to pathogen infection in transgenic *Nicotiana benthamiana*. *PLoS One*. 2013; 8: e68503. <https://doi.org/10.1371/journal.pone.0068503> PMID: 23844212
83. Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Sci*. 2000; 5: 199–206. [https://doi.org/10.1016/s1360-1385\(00\)01600-9](https://doi.org/10.1016/s1360-1385(00)01600-9) PMID: 10785665
84. Chen X, Li C, Wang H, Guo Z. WRKY transcription factors: evolution, binding, and action. *Phytopathology Research*. 2019; 1: 1–15.
85. Jiang J, Ma S, Ye N, Jiang M, Cao J, Zhang J. WRKY transcription factors in plant responses to stresses. *J Integr Plant Biol*. 2017; 59: 86–101. <https://doi.org/10.1111/jipb.12513> PMID: 27995748
86. Phukan UJ, Jeena GS, Shukla RK. WRKY transcription factors: molecular regulation and stress responses in plants. *Front Plant Sci*. 2016; 7: 760. <https://doi.org/10.3389/fpls.2016.00760> PMID: 27375634
87. Lee FC, Yeap WC, Appleton DR, Ho C-L, Kulaveerasingam H. Identification of drought responsive *Elaeis guineensis* WRKY transcription factors with sensitivity to other abiotic stresses and hormone treatments. *BMC Genomics*. 2022; 23: 1–17.
88. Babitha KC, Ramu S v, Pruthvi V, Mahesh P, Nataraja KN, Udayakumar M. Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in *Arabidopsis*. *Transgenic Res*. 2013; 22: 327–341. <https://doi.org/10.1007/s11248-012-9645-8> PMID: 22948308
89. Agarwal PK, Gupta K, Lopato S, Agarwal P. Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance. *J Exp Bot*. 2017; 68: 2135–2148. <https://doi.org/10.1093/jxb/erx118> PMID: 28419345
90. Jha B, Sharma A, Mishra A. Expression of SbGSTU (tau class glutathione S-transferase) gene isolated from *Salicornia brachiata* in tobacco for salt tolerance. *Mol Biol Rep*. 2011; 38: 4823–4832. <https://doi.org/10.1007/s11033-010-0625-x> PMID: 21136169

91. Qi M, Sun T, Xue S, Yang W, Shao D, Martínez-López J. Competitive ability, stress tolerance and plant interactions along stress gradients. *Ecology*. 2018; 99: 848–857. <https://doi.org/10.1002/ecy.2147> PMID: 29345307
92. Rong W, Qi L, Wang A, Ye X, Du L, Liang H, et al. The ERF transcription factor Ta ERF 3 promotes tolerance to salt and drought stresses in wheat. *Plant Biotechnol J*. 2014; 12: 468–479. <https://doi.org/10.1111/pbi.12153> PMID: 24393105
93. Islam S, das Sajib S, Jui ZS, Arabia S, Islam T, Ghosh A. Genome-wide identification of glutathione S-transferase gene family in pepper, its classification, and expression profiling under different anatomical and environmental conditions. *Sci Rep*. 2019; 9: 1–15.
94. Chinnusamy V, Schumaker K, Zhu J. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J Exp Bot*. 2004; 55: 225–236. <https://doi.org/10.1093/jxb/erh005> PMID: 14673035
95. Quigley F, Rosenberg JM, Shachar-Hill Y, Bohnert HJ. From genome to function: the Arabidopsis aquaporins. *Genome Biol*. 2001; 3: 1–17. <https://doi.org/10.1186/gb-2001-3-1-research0001> PMID: 11806824
96. Ranganathan K, el Kayal W, Cooke JEK, Zwiazek JJ. Responses of hybrid aspen over-expressing a PIP2; 5 aquaporin to low root temperature. *J Plant Physiol*. 2016; 192: 98–104. <https://doi.org/10.1016/j.jplph.2016.02.001> PMID: 26895330
97. Sun H, Wang S, Lou Y, Zhu C, Zhao H, Li Y, et al. A bamboo leaf-specific aquaporin gene PePIP2; 7 is involved in abiotic stress response. *Plant Cell Rep*. 2021; 40: 1101–1114. <https://doi.org/10.1007/s00299-021-02673-w> PMID: 34100122
98. Zhang M, Liu R, Liu H, Yang H, Li X, Wang P, et al. Citrus NIP5; 1 aquaporin regulates cell membrane water permeability and alters PIPs plasma membrane localization. *Plant Mol Biol*. 2021; 106: 449–462. <https://doi.org/10.1007/s11103-021-01164-6> PMID: 34173150
99. Song J, Ye G, Qian Z, Ye Q. Virus-induced plasma membrane aquaporin PsPIP2; 1 silencing inhibits plant water transport of *Pisum sativum*. *Bot Stud*. 2016; 57: 1–10. <https://doi.org/10.1186/s40529-016-0135-9> PMID: 28597425
100. Zargar SM, Mahajan R, Bhat JA, Nazir M, Deshmukh R. Role of silicon in plant stress tolerance: opportunities to achieve a sustainable cropping system. *3 Biotech*. 2019; 9: 1–16.
101. Reddy PS, Dhaware MG, Sivasakthi K, Divya K, Nagaraju M, Cindhuri KS, et al. Pearl millet aquaporin gene PgPIP2; 6 improves abiotic stress tolerance in transgenic tobacco. *Front Plant Sci*. 2022; 13.
102. Sivasakthi K, Tharanya M, Kholová J, Wangari Muriuki R, Thirunalasundari T, Vadez V. Chickpea genotypes contrasting for vigor and canopy conductance also differ in their dependence on different water transport pathways. *Front Plant Sci*. 2017; 8: 1663. <https://doi.org/10.3389/fpls.2017.01663> PMID: 29085377