

SCIENTIFIC REPORTS



OPEN

Transcriptome sequence analysis and mining of SSRs in Jhar Ber (*Ziziphus nummularia* (Burm.f.) Wight & Arn) under drought stress

Radha Yadav¹, Showkat Ahmad Lone^{1,2}, Kishor Gaikwad¹, Nagendra Kumar Singh¹ & Jasdeep Chatrath Padaria¹

Ziziphus nummularia (Burm.f.) Wight & Arn., a perennial shrub that thrives in the arid regions, is naturally tolerant to drought. However, there are limited studies on the genomics of drought tolerance in *Ziziphus* sp. In this study, RNA-sequencing of one month old seedlings treated with PEG 6000 was performed using Roche GS-FLX454 Titanium pyrosequencing. A total of 367,176 raw sequence reads were generated, and upon adapter trimming and quality filtration 351,872 reads were assembled *de novo* into 32,739 unigenes. Further characterization of the unigenes indicated that 73.25% had significant hits in the protein database. Kyoto encyclopedia of genes and genomes database (KEGG) identified 113 metabolic pathways from the obtained unigenes. A large number of drought-responsive genes were obtained and among them differential gene expression of 16 highly induced genes was validated by qRT-PCR analysis. To develop genic-markers, 3,425 simple sequence repeats (SSRs) were identified in 2,813 unigene sequences. The data generated shall serve as an important reservoir for the identification and characterization of drought stress responsive genes for development of drought tolerant crops.

Incidence of drought is one of the important causes responsible for reduction in crop yields that arises due to inadequate and unevenly distributed rainfall. Plants respond to drought stress by activating an intricate network of protein-coding genes involving various biochemical and molecular mechanisms^{1,2}. The stress inducible genes expressed during drought either encode functional proteins which directly protect plants against environmental stresses or they encode proteins that help in regulation of stress responsive genes^{3,4}. Studies on response to drought stress have been performed in many arid region plants, including *Ziziphus nummularia*⁵⁻⁷ but the detailed study about the transcriptome involved in drought stress in *Z. nummularia* is scarce.

Z. nummularia (Burm.f.) Wight & Arn., commonly known as jhar ber or jhar beri belonging to the family Rhamnaceae, is a perennial herb native to tropical and dry regions, especially in the Thar Desert of western India, south-eastern Pakistan and southern Iran⁸. *Z. nummularia* is being utilized in a number of ways for the economic benefits. The leaves of this species are used as fodder^{9,10}, the fruit is used as food¹¹⁻¹³, particularly during food scarcity in many provinces of India. Since the shrub yields very hard wood, it is used to make tool handles, walking sticks and also for combustion purposes¹⁴. Over the years, the health benefits of Jhar ber have been extensively reported, including its antidiabetic, anti-inflammatory, antimicrobial, antioxidant, antitumor and immune system stimulant properties¹⁵⁻¹⁷. The anatomical characteristics like presence of high frequency of papilla on both surfaces, epidermis with thick outer wall and thick cuticle, and presence of stomata in crypts, place *Z. nummularia* in the xerophyte category of plants. As a xerophyte, *Z. nummularia* is bestowed with great tolerance to various abiotic stresses, such as drought, salinity and high temperature, which makes it an untapped non-model plant resource to uncover the mechanism underlying stress tolerance and to explore the candidate genes involved in stress tolerance⁷.

¹Biotechnology and Climate Change Group, National Research Centre on Plant Biotechnology (ICAR-NRCPB), New Delhi, 110012, India. ²Present address: Department of Agricultural Microbiology, Faculty of Agricultural Sciences, Aligarh Muslim University, Aligarh, 202002, India. Radha Yadav and Showkat Ahmad Lone contributed equally to this work. Correspondence and requests for materials should be addressed to J.C.P. (email: jasdeep_kaur64@yahoo.co.in)

Only a few studies have demonstrated the molecular mechanism and gene expression patterns involved in the drought tolerance of Indian jujube⁷. The problem of studying the genetic basis of drought tolerance in *Z. nummularia* is further complicated by lack of sequence data, as evidenced by the availability of only 1,243 ESTs (till May 30th 2017) in the GenBank database (<https://www.ncbi.nlm.nih.gov/nucest/?term=ziziphus+nummularia>). Therefore, it is imperative to carry out studies exploring the unique molecular adaptive strategies exhibited by this non-model plant species in response to drought stress.

Overall, the knowledge of *Ziziphus* genomics is in its infancy with limited genome-wide studies being available in the public domain. *Z. jujuba* (Chinese ber) represents the first and the only genome to be sequenced from the Rhamnaceae family¹⁸. The limited genomic information has further hampered the molecular breeding and SSR marker development. Molecular markers are robust tools to study the plant genomes and the genetic variations associated with particular phenotypes¹⁹. A number of molecular markers have been developed over the years and among them simple sequence repeats (SSRs) or microsatellite loci are preferred because they are simpler, reproducible, abundant, hyper variable, effective and co-dominant²⁰. Currently, no SSRs are seemingly reported in *Z. nummularia*, though a few have been reported in other species like *Z. mauritiana*²¹, *Z. jujuba*^{22,23}, and *Z. jujuba* var. *spinosa*²⁴. The absence of high-throughput marker collection is a major bottleneck in conducting studies on population genetics and high density mapping in *Z. nummularia*. Therefore, the identification of robust SSR markers is of paramount importance for conducting such studies in *Z. nummularia*.

Of late, the Next-Generation Sequencing (NGS) of RNA has emerged as an excellent approach for isolation of novel genes and development of molecular markers like SSRs^{25,26}, particularly, in the previously unannotated genomes^{27,28}. In the present study, we have performed transcriptome sequencing of *Z. nummularia* under drought stress using 454 GS-FLX sequencing platform. The *de novo* assembly and annotation led to the identification of candidate genes involved in drought tolerance. The differential expression patterns of the drought-responsive genes have been validated by qRT-PCR analysis. This is the first report on transcriptome sequencing in *Z. nummularia*, the large-scale sequencing data generated will help in the discovery of novel genes associated with drought stress. The SSR markers identified shall assist in breeding studies on *Z. nummularia* and the genes identified in this study can be characterized and further deployed for developing drought-stress tolerant transgenic crops.

Material and Methods

Plant materials and stress conditions. The seeds of *Z. nummularia* were procured from the Central Institute for Arid Horticulture (ICAR-CIAH), Bikaner, India. After removing the hard coat, the seeds of *Z. nummularia* were sown in 15 cm diameter pots containing soilrite and allowed to grow in growth chambers at 30 °C with 16 h light/8 h dark photoperiod at the National Phytotron Facility, Indian Agricultural Research Institute, New Delhi, India. Four-week-old seedlings were used to initiate drought stress. Fifteen whole seedlings (3 for each hour of treatment) were treated with 30% Polyethylene glycol M.W. 6000 (PEG 6000) in 0.5x Murashige and Skoog media (w/v) (Himedia, India) supplemented with 1 mM 2-(N-morpholino) ethanesulfonic acid (MES) (w/v) (Himedia, India) to introduce drought stress at 0.1 Mpa (osmotic potential). The seedlings (whole plant including roots, stem and leaves) were harvested after 6 h, 12 h, 24 h, 48 h and 72 h of treatment and three untreated (control) seedlings were also harvested simultaneously for further downstream processing. All the harvested seedlings were immediately frozen in liquid nitrogen and stored in a deep freezer (−80 °C) until use.

The osmotic potential was calculated as follows:

$$\text{Osmotic potential } (\psi\pi) = -CRT$$

where, C denotes the molar concentration of solute in molL^{−1}, R is the universal gas constant (R = 8.314 kPa L/mol K) and T is the absolute temperature (i-e, experimental temperature 30 °C + absolute temperature).

RNA extraction and transcriptome pyrosequencing. The total RNA from both treated and control seedlings was extracted using Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO 63103, USA) as described by the manufacturer and DNA contamination was removed by on column DNase I Digestion set (Sigma-Aldrich). The integrity and concentration of RNA were checked by electrophoresis on 1.2% agarose/EtBr gel in 1X TAE buffer, UV spectrophotometry (Thermo scientific, Waltham, MA, USA) and a Bioanalyzer Chip RNA7500 series II (Agilent Technologies, USA). The mRNA was purified from 3 μg total RNA using Oligotex mRNA mini kit (Qiagen, Germany) as per manufacturer's instructions. Equal concentrations of purified mRNA (200 ng) from both control and treated samples were used as template for random hexamer-primed synthesis of first strand cDNA, followed by double strand cDNA synthesis using cDNA Synthesis System Kit (Roche, Switzerland). The double strand cDNA was purified by PCR purification kit (Qiagen) and concentration of the purified cDNA was determined using Bioanalyzer 2100 (Agilent Technologies). The purified double strand cDNAs were sheared via nebulization into small fragments, followed by ligation to adapters using GS FLX Titanium Rapid Library MID Adaptors kit (Roche, Switzerland). The libraries were pooled before sequencing and loaded on to Titanium PicoTiter Plate followed by pyrosequencing on Roche GS-FLX454 Titanium platform.

De novo assembly, annotation and functional characterization. The raw sequences generated from pyrosequencing were subjected to quality control analysis for removal of short-sized and poor quality reads (like adaptor sequences, reads containing poly N and sequences <50 bp) using FastQC tools v0.11.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The *de novo* assembly of high quality reads was performed using iAssembler v1.0 (beta) (<http://bioinfo.bti.cornell.edu/tool/-iAssembler/>). The assembled unigenes were annotated against non-redundant (NR) protein database through BLASTx with E-value cut-off of 10^{−6} (<http://www.ncbi.nlm.nih.gov/>). The functional categorizations were searched against TAIR (<http://www.arabidopsis.org/>) through BLASTx using threshold E-value value of 10^{−6}. Gene ontology analysis of annotated transcripts was performed using Blast2GO version 2.8 (<https://www.blast2go.com/>)²⁹ and the transcripts were divided into three groups,

Gene	Forward primer sequence (5' to 3')	Reverse primer sequence (5' to 3')
Cell wall-associated hydrolase	GTTTCATCTTCGGCGCAA	CTCAGTGATAAAGGAGGTAGGG
Secretory carrier membrane protein	TTGGGAGATCATGCTTTGG	GCTGCTCGGAAATACAT
Glycine-rich RNA-binding protein	GTCGTGAGGGAGGATACAA	CCACCATCACCGTATCCA
ATP synthase	CCCTACCCGTATTGAAAC	CCGCGTTAATAGCAGGTC
Calcium-dependent protein kinase	GAACTCTGTGCTGGTTTGG	CACGATGCATCACACCCATA
ABC transporter C family	CCTTAAGACATCTGCTTCTTC	TGCTTCAACTCCAGCTCTA
Hydroxyproline-rich glycoprotein family protein	GGAATGTGGGTGATGGTGTAG	AGCGAGTCACCAGAAGAAAC
Late embryogenesis abundant protein	TCATGAGCCAAGAACAGC	TCGGCGCAATTGGTTTAG
Serine/threonine-protein phosphatase,	CCTGCAAACAAGGTGGTACT	CATTCTTCTCAGGCTCTTCC
Asparagine synthetase	GCTAGGAGTGAAGATGGTTATC	CATACATATGCAGTGCCTTAATC
Auxin efflux carrier	CGTATTCCAGCTCATTGG	GAAGAGCCATAAAGAGACCTAAA
Catalase	CCCAATTCCTCTGCCATCT	AATCGCTCTGCCTGTCTG
Heat shock protein 20	CTAAGGCCGAGAAGAAGGAAG	GAGGCACAGGAACAACTACA
<i>Wrky</i> transcription factor	TTCTCTGAGTCTCCAATCT	CCATCTGGCAATACCTTTCT
Heat shock protein 70	GCAAAATGGTCTTGGTGAGT	CAACCGATTGAGGCTTCAT
Chaperone protein <i>ClpB3</i>	CAACAATCAGAGAGGGAGAAGC	TCCGATAGACCTGCTCTTGA
<i>Ziziphus</i> elongation factor1	GCTGACTGTGCTGTTCTCATC	GACACCAAGAGTGAAGCGAG

Table 1. Description of the putative drought responsive genes and their respective primer sequences used for qRT-PCR analysis.

viz. biological process, molecular function and cellular component. The biochemical pathways of annotated transcripts were carried out by KEGG mapping³⁰ using Blast2GO program.

Mining of simple sequence repeats (SSRs). The Perl script of MISA–MicroSatellite identification tool (<http://pgrc.ipk-gatersleben.de/misa/>) was used for the identification of mono-nucleotide, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide and hexa-nucleotide simple sequence repeats (SSRs) in the unigenes generated by sequencing. Ten contiguous repeats for mono-nucleotides, six contiguous repeats for di-nucleotides and five contiguous repeats each for tri-nucleotide, tetra-nucleotide, penta-nucleotide and hexa-nucleotides were considered as search criteria for SSR identification.

Differential expression analysis of drought-responsive genes. In order to validate the robustness of the obtained transcriptome profile read analysis, 16 putative drought-responsive genes were selected and their expression patterns were studied at different intervals of PEG6000-induced drought stress by qRT-PCR analysis. Primers were designed by using Integrated DNA technology (IDT) software. The description of the genes selected and their corresponding primer sequences are given in Table 1; the related BLAST annotations of these genes is provided in supplementary Table S1. The total RNA was isolated from control and stressed samples, as described above and cDNA was synthesized by SuperScript[®] III first strand synthesis system (Invitrogen, USA) using oligo (dT) primer. *Ziziphus* elongation factor1 (*Zjef1*) [GenBank: EU916202] gene was used as internal control³¹. The reaction mixtures of qRT-PCR were set up in a total volume of 10 μ l containing 100 ng template cDNA, 200 nM of each primer and 1 μ l of 10X KAPA SYBR FAST qPCR master mix buffer (KAPA Biosystems, USA). The qRT-PCR was carried out in LightCycler[®] 480 II (Roche) with thermal cycling conditions as: initial denaturation at 94 °C for 3 min, followed by 40 cycles at 94 °C for 30 sec, 60 °C for 30 sec and 72 °C for 30 sec. The Ct (threshold cycles) value were calculated by $2^{-\Delta\Delta CT}$ method as described by Padaria *et al.*⁷, where Ct values from three replicates were averaged and normalized with the Ct values of internal control *Zjef1* and specificity of the reactions was confirmed by melting curve. The statistical significance of the difference in relative expression of treated and control plants was calculated by Tukey's multiple comparison test.

Data availability. The high-quality sequencing reads obtained in this study were deposited to Short Read Archive (SRA) database of NCBI (<http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=studies>) and were assigned the accession number SRX1118834.

Results

***De novo* assembly, annotation and Functional characterization.** To gain insights into the transcriptome profile of *Z. nummularia* in response to drought stress, a pooled cDNA library at different intervals of PEG 6000 induced drought stress was constructed and sequenced. A total of 367,176 raw reads were generated and after removing the low-quality reads, 351,872 (95.83% of all the reads) high quality (HQ) reads were used for subsequent analysis. Among these HQ reads, control (C) and stress (S) library contained 1,14,012 and 2,37,860 sequences, respectively (Table 2). The *de novo* assembly was carried out using iAssembler v1.0 (beta) tool with default or optimized parameters. A total of 351,872 high quality reads were assembled into 32,739 unigenes (Supplementary Table S1). The total size of transcripts including gaps was 1,34,76,223 bp. The length of unigenes ranged from 200 bp to 5,167 bp, with an average length of 412 bp, the mean transcript length and N50 values were 412 bp and 396 bp, respectively (Table 2). The nucleotide length of 442 transcripts (1.35%) was upto

Statistics	Results
Raw reads from control library	1,18,386
Raw reads from stress library	2,48,790
Reads after trimming from C library	1,14,012
Reads after trimming from S library	2,37,860
High quality sequence used for assembly	3,51,872
Total number of unigenes	32,739
Size of transcripts including gaps	1,34,76,223 bp
Longest transcript	5,167 bp
Shortest transcript	200 bp
Mean transcript size	412 bp
Median transcript size	337 bp
N50 transcript length	396 bp
Transcript %GC	42.93

Table 2. Summary of transcriptome sequencing and assembly in *Ziziphus nummularia*.

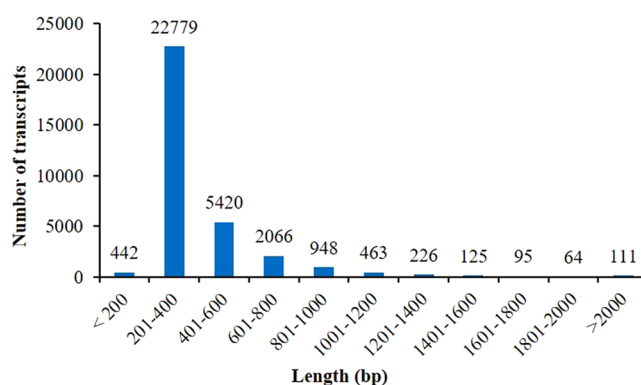


Figure 1. Sequence length distribution for assembled sequence reads.

200 bp, of 22,779 transcripts (69.57%) was 201–400 bp, of 5,420 transcripts (16.55%) was 401–600 bp, of 2,066 transcripts (6.31%) was 601–800 bp, of 948 transcripts (2.89%) was 801–1,000 bp, of 463 transcripts (1.41%) was 1,001–1,200 bp, of 226 transcripts (0.69%) was 1,201–1,400 bp, of 125 transcripts (0.38%) was 1,401–1,600 bp, of 95 transcripts (0.29%) was 1,601–1,800 bp, of 64 transcripts (0.19%) was 1,801–2,000 bp and of 111 transcripts (0.33%) was >2,001 bp (Fig. 1) with an average GC content of 42.93%.

For annotation and functional characterization, non-redundant (NR) databases of NCBI and TAIR were used. All the unigenes were aligned with unigenes of other plant species using BLASTx program with an E-value cut-off of 10^{-06} . Out of 32,739 unigenes annotated, 20,291 (61.97%) showed significant BLAST matches and it was observed that many unigenes (two or more) were annotated against the same protein. A total of 9,487 (28.97%) unigenes did not show any hits with the known protein, so they may be considered as putative novel unigenes. The number of annotated sequence showing similarity to *Ziziphus jujuba* was the largest (74.36%), followed by *Morus notabilis* (2%), *Glycine max* (1.58%), *Vitis vinifera* (1%) and *Citrus sinensis* (0.89%) (Supplementary Figure S1, Supplementary Table S2).

Gene ontology assignments, functional annotation and categorization of assembled unigenes of *Z. nummularia* were carried out by using TAIR database. BLASTx results indicated that 20,345 of the 32,739 unigenes matched with a particular accession number or gene. Out of 20,345 unigenes, 9,906 unigenes were annotated to GO terms one or more times and they were categorized into 47 functional groups. All these functional groups were further categorized into three main categories of gene ontology classification *viz.* biological process, molecular function and cellular component. The biological process group was further classified into 23 sub-groups with the majority of unigenes being explicated in cellular (6,576, 66.4%) and metabolic process (6,124, 61.8%) categories (Fig. 2). The molecular function category was divided into 14 sub-groups with majority of the unigenes being delineated in binding activity (4,773, 48.2%) and catalytic activity (4,211, 42.5%) (Fig. 2). The cellular component category was classified into 10 sub-functional groups and unigenes under cell group were the most dominant (8,938, 90.2%) followed by organelle (7,011, 70.8%), organelle part (2,113, 21.3%), macromolecular complex (1093, 11.0%), extracellular region (746, 7.5%), envelope (652, 6.6%), membrane-enclosed lumen (349, 3.5%) and extracellular region (16, 0.2%) sub-groups (Fig. 2).

KEGG pathway mapping. The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis revealed that out of 2,571 annotated unigenes, 1,140 had KEGG orthologs which were assigned to 114 pathways

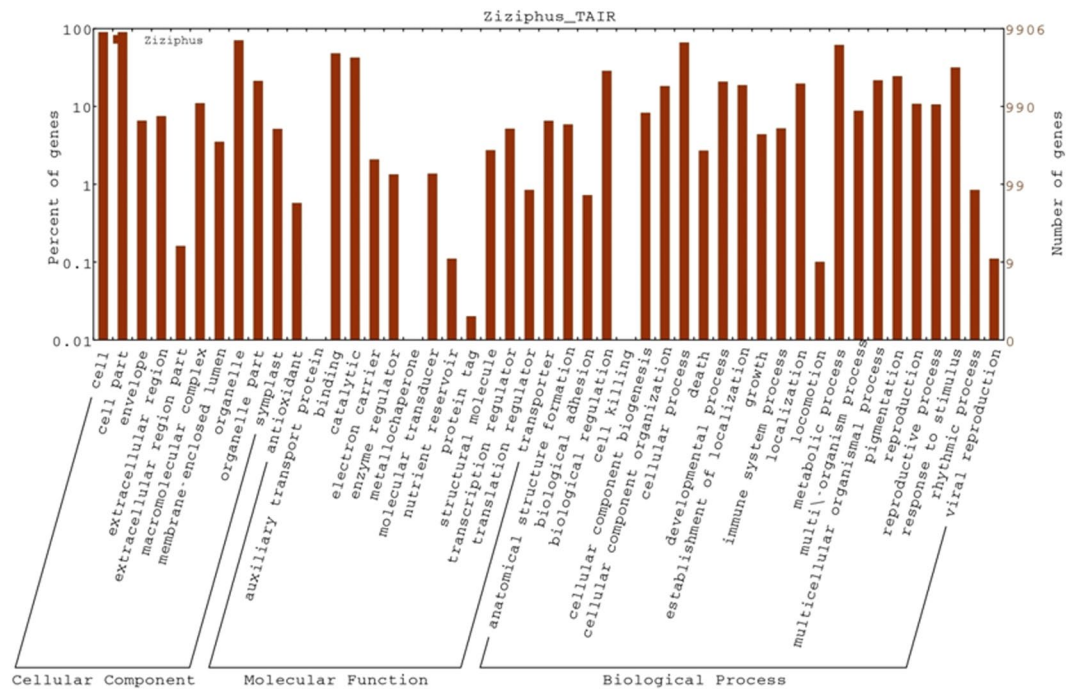


Figure 2. Histogram representing gene ontology (GO) classification of the assembled unigenes. The results are summarized in three main GO categories: cellular component, molecular function and biological process. The x-axis depicts the sub-categories of the unigenes, and the left and right y-axes represent the percentage and the number of unigenes, respectively.

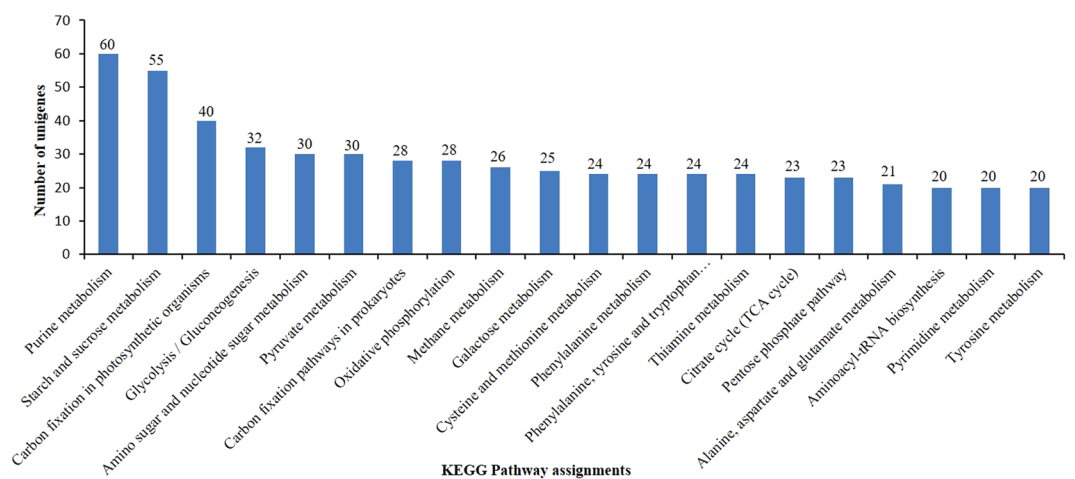


Figure 3. Histogram representations for the top twenty pathways of unigenes assigned based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). The columns signify the number of unigenes in each subcategory.

(Supplementary Table S3). The purine metabolism, starch and sucrose metabolism, carbon fixation in photosynthetic organisms, glycolysis/gluconeogenesis and amino sugar and nucleotide sugar metabolism pathways were the most representative with 60, 55, 40, 32 and 30 unique sequences, respectively. The top 20 KEGG pathway groups are represented in Fig. 3.

Distribution and frequency of SSRs in the *Z. nummularia* transcriptome. Using MISA tool, a total of 32,739 sequences containing 1,34,76,223 bp were examined and 3,425 SSRs distributed in 2,813 sequences were identified. The EST-SSR frequency in the transcriptome was observed at 8.59%. As per the classification criteria proposed by Weber³², the most abundant repeat motif was mono-nucleotide (1,453, 42.4%), followed by di-nucleotide (982, 28.67%), tri-nucleotide (874, 25.5%), tetra-nucleotide (80, 2.33%), hexa-nucleotide (21, 0.60%) and penta-nucleotide (15, 0.45%) (Table 3). 414 out of 3,425 SSRs were present in compound formation

Motif length (nucleotides)	Repeat numbers							Total motifs	Percentage in total (%)
	5	6	7	8	9	10	>10		
Mono-						458	995	1,453	42.42
Di-		224	145	115	129	103	266	982	28.67
Tri-	412	198	120	54	33	28	29	874	25.51
Tetra-	50	21	2	1	0	1	5	80	2.33
Penta-	8	2	4	1		0	0	15	0.43
Hexa-	14	3	3	0	1	0		21	0.61
Total								3,425	

Table 3. Frequency of different SSRs observed in *Ziziphus nummularia*.

and the remaining 3,011 were observed to be the perfect SSRs. Among 3,011 perfect SSRs, 333 were found to be compound SSRs and the summarized details are in supplementary Table S4. The number of sequences having more than one SSR was 491. The frequencies of SSRs with different number of tandem repeats were calculated and it was observed that SSRs with five tandem repeats (24.54%) were the most abundant, followed by six tandem repeats (22.71%), >10 tandem repeats (21.9%), seven tandem repeats (13.89%), eight tandem repeats (8.67%), nine tandem repeats (8.26%) and 10 tandem repeats (6.69%) (Supplementary Table S5). Among the di-nucleotides repeats identified, AT/AT motif was the most predominant, constituting 25.30% of all the di-nucleotides characterized, which was followed by AG/CT (18.96%), AC/GT (5.22%) and CG/CG (0.30%) motifs. Among the tri-nucleotide repeats, AAG/CTT motif was the most predominant amounting to 14.09% of all the tri-nucleotides characterized, followed by AAT/ATT (9.22%), ACC/GGT (5.52%), ATC/ATG (4.56%), AGC/CTG (2.89%), AGG/CCT (2.78%), ACG/CGT (0.96%), CCG/CGG (0.76%) and ACT/AGT (0.50%) motifs. A total of 80 tetra-nucleotide repeat motifs were identified, among them AAAT/ATTT (2.43%) was found to be the most abundant. The number of penta-nucleotide and hexa-nucleotide motifs identified was 15 and 21, respectively, with AAAAT/ATTTT (0.45%) as the most predominant penta-nucleotide motif and AAAAAT/ATTTTT (0.20%) as the most frequent hexa-nucleotide motif. The frequencies of all the identified repeat motifs are summarized in supplementary Table S5.

Differential expression analysis of genes involved in drought tolerance. Based on the qRT-PCR data analysis, the selected genes displayed different expression profiles, with chaperone protein ClpB3 and secretory carrier membrane protein encoding genes showing significant down regulation compared to the control, and genes encoding calcium-dependent protein kinase and heat shock protein 20 showing expression at par with the control (Fig. 4). All other genes displayed significant up-regulation with time compared to the control, with *LEA* gene showing the maximum up-regulation (56.2-fold) at 48 h of PEG 6000 treatment (Fig. 4). The qRT-PCR expression patterns of all the candidate genes were in agreement with the transcriptome data, and therefore validated the results of the first *de novo* transcriptome data.

Discussion

Z. nummularia, an inherently drought-tolerant non-model plant, is an important untapped genetic resource to study the genetic and metabolic control of adaptation to drought stress⁷. Next-generation sequencing technology (NGS) has been extensively used to study the model plant transcriptomes and due to accuracy, rapidity and low cost, its use has also been extended to unearth the unique genetic characteristics in the non-model plant systems³³. The sequences generated were *de novo* assembled using iAssembler software. The iAssembler generates initial assemblies by MIRA and CAP3 assemblers and makes correction in two common types of transcriptome assembly errors: i) ESTs from different transcripts (mainly alternatively spliced transcripts or paralogs) are incorrectly assembled into same contigs; and ii) ESTs from same transcripts fail to be assembled together to deliver highly accurate *de novo* assemblies of EST sequences³⁴. In spite of the development of several bioinformatics tools for data assembly and analysis, *de novo* assembly of short reads without a reference genome, is very challenging task^{35,36}. It is very important to choose a suitable assembler and parameters to assemble large-scale ESTs into consensus sequences with significantly high accuracy.

In the present study, the average GC content of *Z. nummularia* transcripts was observed to be 42.93% which is in agreement with the previously reported value in other dicots *viz.*, 44.6% for *Picrorhiza kurroa*³⁷, 33.41% for genome of *Z. jujuba*¹⁸, and 41.0% for *Nicotiana benthamiana*³⁸. The GC content represents the stability of DNA as well as of the genes and genomic composition including evolution, gene structure and regulation³⁹, and the changing GC content may reveal the adaptability to different climatic conditions.

The functional annotation results showed that a total of 23,984 (73.25%) unigenes have significant BLAST matches, which is higher than the previously reported data, 54.9% in bamboo⁴⁰, and 54.31% in *Ziziphus jujuba*⁵. This higher value indicates the extensive coverage of the *Z. nummularia* transcriptome in this study. A total of 8,112 (24.77%) unigenes did not match with any of the known proteins, this is a common feature associated with *de novo* sequencing and varying percentages like 45.73%, 40.70%, 62.48%, 42.78% by Dang *et al.*⁴¹, Long *et al.*⁴², Wu *et al.*⁴³ and Ma *et al.*⁶, respectively of such unigenes has been reported previously. The presence of these un-identified unigenes can be partly ascribed to the dearth of genome and EST information of *Z. nummularia* in the public domain and also to the short read length commonly associated with the next-generation sequencing technology⁶. The gene ontology results revealed that a large number of unigenes were involved in cellular process, metabolic process and response to stimulus (Fig. 2). The functional GO results indicated the dominance of gene

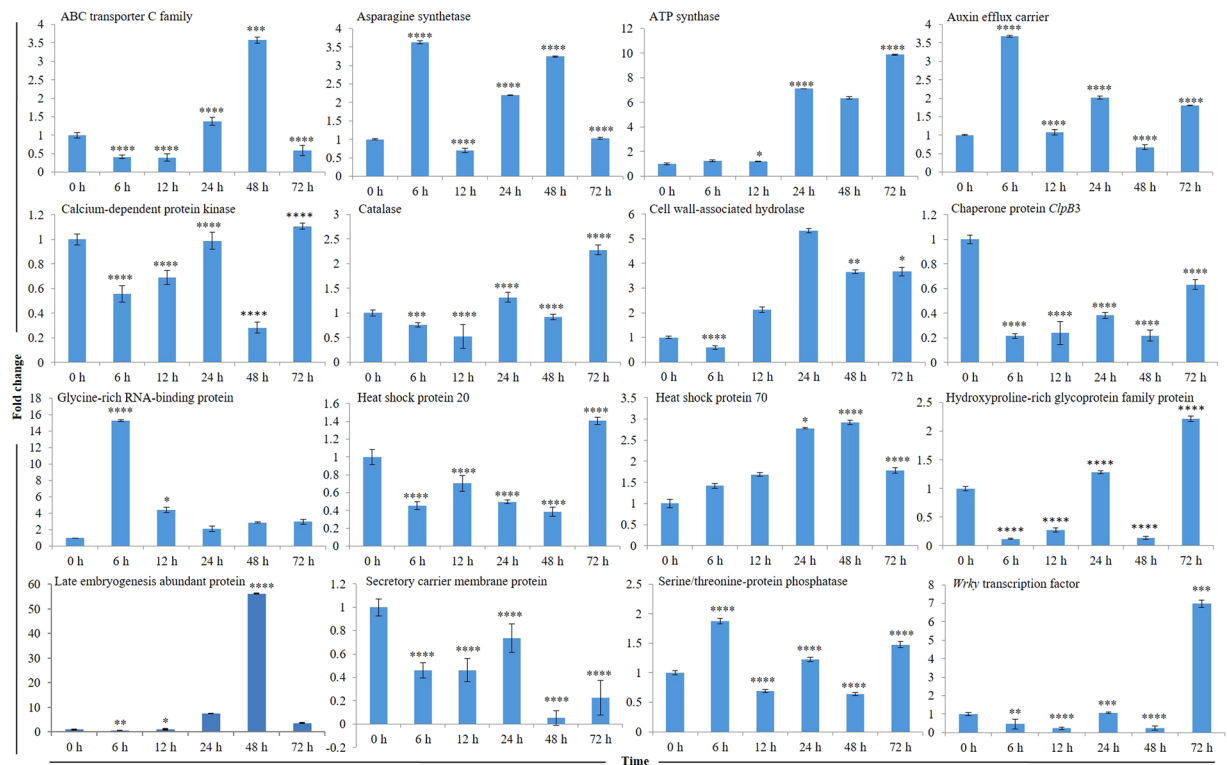


Figure 4. qRT-PCR expression profile of the putative drought responsive genes in *Z. nummularia*. Duration of PEG-induced drought stress (in hours) and Fold change in gene expression are shown on X and Y axes, respectively. *Ziziphus elongation factor1 (Zjef1)* was used as an internal control. The data represents mean values \pm SE ($n = 3$) and vertical lines indicate calculated standard error. The asterisks designate statistical significance between the control and treated seedlings (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ and **** $P < 0.0001$).

regulation, signal transduction and enzymatically active processes in transcriptome, as maximum unigenes were found to be belonging to the binding activity, catalytic activity, transport activity and transcription regulator activity categories (Fig. 2). The various categories identified by the GO analysis classified the unigenes as, those involved in response to abiotic and biotic stimulus. Among these, the unigenes were further found to be involved in heat acclimation, cellular response to hydrogen peroxide, osmotic stress, oxidative stress, unfolded protein, water deprivation, drought recovery, endoplasmic reticulum unfolded protein response, hydrogen peroxide catabolic process, starvation, osmosensory signaling pathway, priming of cellular response to stress and cation stress. All these categories are reported to participate in drought tolerance mechanisms^{44,45}. Drought tolerance mechanism acts in combination with drought avoidance and tolerance periods; it also involves many genes that play a major role in osmotic and redox homeostasis processes and participate in the readjustment process. Some of the unigene categories identified were found to be involved in response to salt, cold and heat stresses, hyperosmotic response, and hyperosmotic salinity response, indicating that various abiotic stresses follow multiple pathways for stress perception and signaling, which cross-talk at various points⁴⁶. These results suggested that transcriptome sequencing was successful in collecting the information about cytosolic response to drought stress in *Z. nummularia* and it would contribute to our understanding about the drought tolerance mechanism in *Z. nummularia*.

The KEGG pathway mapping resulted in the identification of 114 biological pathways from 1,140 annotated unigenes (Fig. 3, Supplementary Table S3). Li *et al.*⁵ reported 16,693 *Z. jujuba* annotated unigenes, consisting of 6,485 isotigs and 10,117 singletons, which were assigned to 319 KEGG pathways. Some of the pathways, identified in our study, that are involved in biosynthesis of genes responsive to drought stress are purine metabolism⁴⁷, starch and sucrose metabolism⁴⁸, fructose and mannose metabolism⁴⁹, arginine and proline metabolism⁵⁰, aminobenzoate degradation⁴³, thiamine metabolism⁵¹, and phenylalanine metabolism⁵². These analyses showed that the diversity and variation in *Z. nummularia* genome provide precious information for investigating specific processes, functions and pathways and thereby making possible the discovery of novel genes which are responsive to drought stress in *Z. nummularia*.

In transcriptome of *Z. nummularia*, a total of 3,425 SSRs were identified with a frequency of 8.59%, which is lower than that observed in the fruit transcriptome of the related species *Z. jujuba* (10.1%)⁵, and floral transcriptome of *Z. celata* (17%)⁵³. A total of 3,425 SSRs were recognised in 2,813 sequences, with one SSR observed for every 3.93 kb of the analysed sequences. Previously, one SSR locus for every 0.94 kb (kb/SSR) was detected in *Triticum aestivum*⁵⁴, 3.87 kb/SSR in *Z. jujuba*⁵, 3.55 kb/SSR in *Camellia sinensis*⁵⁵, 6.22 kb/SSR in *Arachis hypogaea*⁵⁶, 6.69 kb/SSR in *Epimedium sagittatum*⁵⁷, and 17.56 kb/SSR for *Cymbidium ensifolium*⁵⁸ transcriptomes. The incidence of SSRs in plant genomes is governed by the repeat length and the standard used for searching SSRs in databases⁵⁹. Moreover, the depth of sequencing coverage also influences the detection rate and since

454-pyrosequencing gives less depth compared to other contemporary sequencing platforms, the detection could be low in the present case. Among all the nucleotide repeat motifs, AT/AT, AG/CT and AAG/CTT were the most abundant SSR motifs and CG/CG was the least frequent. Our results are consistent with those observed in *Sesamum indicum*⁶⁰, *Epimedium sagittatum*⁵⁷, and other plant species^{61,62}. These observations suggest that the transcriptome sequencing data is an ideal resource for mining the SSRs in *Z. nummularia* and its cross transferability will help in studying different species of this family.

Drought tolerance is a complex trait and various adaptive mechanisms are involved at multiple levels like cellular, molecular and morphological to cope with water-deficit conditions. In our study, a large number of drought-responsive genes were obtained in the transcriptome of *Z. nummularia* and their expression patterns were studied by qRT-PCR analysis. Among them, the cell wall associated hydrolase gene has been reported to play an important role in accumulation of compatible extra-cellular solutes through altering cell wall extensibility and other cell wall modifications under drought and salt stress. High expression levels of the cell wall associated hydrolase gene was reported in *Ammopiptanthus mongolicus* and may help the specific structural adaptations to cope with severe drought in deserts⁶³. Glycine-rich RNA-binding protein harboring RNA chaperon activity has importance in RNA metabolism to prevent the adverse effect of environmental stresses on crop yield. Transgenic rice over-expressing *Arabidopsis thaliana* glycine-rich RBP (*AtGRP* or *AtGRP7*) has shown much higher stress-recovery rates and grain yields in comparison to the wild-type plants under drought stress conditions and has confirmed the importance of post-transcriptional regulation of RNA metabolism in plant under abiotic stress conditions⁶⁴. ABC transporter gene is a membrane-intrinsic gene that has shown high expression level and plays an important role in physiological processes to adapt the plant in changing environments and cope with biotic and abiotic stresses⁶⁵. Based on the qRT PCR analysis, the maximum upregulation was observed for *Lea* gene (Fig. 4), many studies have reported that over-expression of LEA proteins improve abiotic stress tolerance in transgenic plants. For example, the expression of barley LEA protein (HVA1) conferred drought tolerance in transgenic wheat and rice^{66,67}, and wheat LEA proteins PMA80 and PMA1959 enhanced dehydration tolerance in transgenic rice⁶⁸. The involvement of WRKY proteins in various abiotic stresses, such as salinity, drought, and cold has been reported previously^{69,70}. *Wrky* gene from *T. aestivum* and *Oryza sativa* have been identified to play an important role in drought tolerance mechanism^{71,72}. The enzyme catalase is a main component of anti-oxidative machinery⁷³ and its expression and activities are always triggered by the environmental stresses^{74–76}. The upregulation of catalase gene in *Z. nummularia* at 72 h has suggested that abiotic stresses enhance the transcription of catalase and subsequently control homeostasis in plant cell as described earlier⁷⁷. The qRT-PCR expression patterns of all the candidate genes were found in agreement with the transcriptome data, therefore, the high-throughput RNA sequencing data generated could be considered fairly conclusive.

Conclusion. *Z. nummularia* is categorized as a xerophyte plant species with inherent tolerance to drought stress and hence is a rich genetic resource to investigate drought-responsive genes and mechanism of drought tolerance. In this study, we have used a combination of *de novo* transcriptome sequencing, differential gene expression profiling and SSR marker identification to perform comprehensive analysis of the mechanism underlying drought stress and also for identification of drought-responsive genes in *Z. nummularia*. Computational analysis has confirmed that the maximum portion of *Z. nummularia* transcriptome is sequenced, and numerous genes and pathways are involved in tolerance to drought stress. The maximum number of ESTs showing similarity to the known databases and a substantial number of ESTs remaining un-identified do endorse the discovery of putative novel genes from the transcriptome data. Various SSR molecular markers have been characterized, which shall be useful in genotyping and breeding studies in *Ziziphus* and related species. The data generated can serve as a significant resource for bio-prospecting drought-responsive candidate genes. Our findings will complement the available genomic resources and shall further provide the momentum to the ongoing efforts to ameliorate drought tolerance in crop plants.

References

- Chaves, M. M., Maroco, J. P. & Pereira, J. S. Understanding plant responses to drought- from genes to the whole plant. *Funct. Plant Biol.* **30**(3), 239–264 (2003).
- Jenks, M. A., Hasegawa, P. M. & Jain, S. M. *Advances in molecular breeding toward drought and salt tolerant crops.* (Springer, Netherlands, 2007).
- Shinozaki, K., Yamaguchi-Shinozaki, K. & Seki, M. Regulatory network of gene expression in the drought and cold stress responses. *Curr. Opin. Plant Biol.* **6**(5), 410–417 (2003).
- Valliyodan, B. & Nguyen, H. T. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr. Opin. Plant Biol.* **9**(2), 189–195 (2006).
- Li, Y. *et al.* *De Novo* Assembly and Characterization of the Fruit Transcriptome of Chinese Jujube (*Ziziphus jujuba* Mill.) Using 454 Pyrosequencing and the Development of Novel Tri-Nucleotide SSR Marker. *PLoS One* **9**(9), e106438 (2014).
- Ma, X., Wang, P. & Zhou, S. *et al.* *De novo* transcriptome sequencing and comprehensive analysis of the drought-responsive genes in the desert plant *Cynanchum komarovii*. *BMC genomics* **16**(1), 1 (2015).
- Padaria, J. C. *et al.* Molecular cloning and characterization of drought stress responsive abscisic acid-stress-ripening (*Asr 1*) gene from wild jujube, *Ziziphus nummularia* (Burm.f.) Wight & Arn. *Mol. Biol. Rep.* **43**(8), 849–59 (2016).
- Singh, N. P., Vohra, J. N., Hazra, P. K. & Singh, D. K. *Ziziphus*. In: Flora of India, Vol. 5. Botanical Survey of India, Calcutta, pp. 221–245 (2000).
- Shankar, V. Silviculture research: a review. *Forage Res.* **6**, 107–122 (1980).
- Shankar, V. & Kumar, S. Distribution pattern of jharber (*Ziziphus nummularia*) and its leaf fodder and bushwood production on different habitats and land use types. *Forage Res.* **7**(2), 119–126 (1981).
- Bhandari, M. M. Famine foods in the Rajasthan Desert. *Econ Bot.* **28**, 73–81 (1974).
- Shankar, V. Life support species in the Indian Thar deserts. In: Paroda R. S., Kapoor, P., Arora, R. K., Bhagmal, (eds) Life support plant species: diversity and conservation. National Bureau of Plant Genetic Resources, New Delhi, p 38 (1988).
- Rathore, M. Nutrient content of important fruit trees from arid zone of Rajasthan. *J. Horticult. For.* **1**(7), 103–108 (2009).
- Pandey, A., Singh, R., Radhamani, J. & Bhandari, D. C. Exploring the potential of *Ziziphus nummularia* (Burm. f.) Wight et Arn. from drier regions of India. *Genet. Resour. Crop Ev.* **57**(6), 929–36 (2010).

15. Dwivedi, S. P. D., Pandey, V. B., Shah, A. H. & Eckhardt, G. Cyclopeptide alkaloids from *Zizyphus nummularia*. *J. Nat. Prod.* **50**(2), 235–237 (1987).
16. Singh, A. K., Raghubanshi, A. S. & Singh, J. S. Medical ethnobotany of the tribals of Sonaghati of Sonbhadra district, Uttar Pradesh, India. *J. Ethnopharmacol.* **81**(1), 31–41 (2002).
17. Bachaya, H. A., Iqbal, Z., Khan, M. N. & Jabbar, A. Anthelmintic activity of *Zizyphus nummularia* (bark) and *Acacia nilotica* (fruit) against Trichostrongylid nematodes of sheep. *J. Ethnopharmacol.* **123**(2), 325–329 (2009).
18. Liu, M. J. *et al.* The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.* **5** (2014).
19. Appleby, N., Edwards, D. & Batley, J. New technologies for ultra-high throughput genotyping in plants. In: D. A., Somers, P. Langridge and J. P., Gustafson (eds.), *Plant genomics: methods and protocols*, Springer, pp. 19–39 (2009).
20. Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids res.* **17**(16), 6463–6471 (1989).
21. Singh, A. K., Devanshi & Sharma, P. *et al.* Assessment of Genetic Diversity in *Zizyphus mauritiana* Using Inter-Simple Sequence Repeat Markers. *J. Plant Biochem. Biotechnol.* **16**(1), 35–40 (2007).
22. Ma, Q., Wang, G. & Liang, L. Development and characterization of SSR markers in Chinese jujube (*Zizyphus jujuba* Mill.) and its related species. *Sci. Hort.* **129**(4), 597–602 (2011).
23. Wang, S., Liu, Y. & Ma, L. Isolation and Characterization of Microsatellite Markers and Analysis of Genetic Diversity in Chinese Jujube (*Zizyphus jujuba* Mill.). *PLoS One* **9**(6), e99842 (2014).
24. Zhang, C., Yin, X. & Li, X. *et al.* Development and characterization of microsatellite markers for sour jujube (*Zizyphus jujuba* var. *spinosa*). *Indian J. Genet. Plant Breed.* **73**(3), 338–341 (2013).
25. Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **24**(3), 133–141 (2008).
26. Morozova, O. & Marra, M. A. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**(5), 255–264 (2008).
27. Wang, X. W. & Luan, J. B. Li *et al.* *De novo* characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics* **11**(1), 400 (2010).
28. Shi, C. Y., Yang, H. & Wei, C. L. *et al.* Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC genomics* **12**(1), 131 (2011).
29. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
30. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
31. Sun, H. F. *et al.* Selection of housekeeping genes for gene expression studies on the development of fruit bearing shoots in Chinese jujube (*Zizyphus jujuba* Mill.). *Mol. Biol. Rep.* **36**, 2183–2190 (2009).
32. Weber, J. L. Informativeness of human (dC-dA)n. (dG-dT)n polymorphisms. *Genomics* **7**, 524–530 (1990).
33. Unamba, C. I. N., Nag, A. & Sharma, R. K. Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants. *Front. Plant Sci.* **6**, 1074 (2015).
34. Zheng, Y., Zhao, L., Gao, J. & Fe, Z. iAssembler: a package for *de novo* assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* **12**, 453 (2011).
35. Wang, S. M. Understanding SAGE data. *Trends Genet.* **23**, 42–50 (2007).
36. Zerbin, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
37. Gahlan, P., Singh, H. R. & Shankar, R. *et al.* *De novo* sequencing and characterization of *Picrorhiza kurroa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics* **13**, 126 (2012).
38. Nakasugi, K. *et al.* *De Novo* Transcriptome Sequence Assembly and Analysis of RNA Silencing Genes of *Nicotiana benthamiana*. *PLoS One* **8**(3), e59534 (2013).
39. Carels, N., Hatey, P., Jabbari, K. & Bernardi, G. Compositional properties of homologous coding sequences from plants. *J. Mol. Evol.* **46**, 45–53 (1998).
40. Zhang, X. M., Zhao, L., Larson-Rabin, Z., Li, D. Z. & Guo, Z. H. *De novo* sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae). *PLoS One* **7**(8), e42082, (2012).
41. Dang, Z. H. Transcriptomic profiling of the salt-stress response in the wild recretohalophyte *Reaumuria trigyna*. *BMC Genomics* **14**(3), 397–402 (2013).
42. Long, Y., Zhang, J. & Tian, X. *et al.* *De novo* assembly of the desert tree *Haloxylon ammodendron* (C. A. Mey.) based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genomics* **15**(1), 1–11 (2014).
43. Wu, Y., Wei, W. & Pang, X. *et al.* Comparative transcriptome profiling of a desert evergreen shrub, *Ammopiptanthus mongolicus*, in response to drought and cold stresses. *BMC Genomics* **15**(1), 671 (2014).
44. Farooq, M., Wahid, A., Kobayashi, N., Fujita, D. & Basra, S. M. A. Plant drought stress: effects, mechanisms and management. *Agron. Sustain. Dev.* **29**(1), 185–212 (2009).
45. Fang, Y. & Xiong, L. General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell. Mol. Life Sci.* **72**(4), 673–89 (2015).
46. Huang, G. T., Ma, S. L. & Bai, L. P. *et al.* Signal transduction during cold, salt, and drought stresses in plants. *Mol. Biol. Rep.* **39**, 969–987 (2012).
47. Watanabe, S., Kounosu, Y., Shimada, H. & Sakamoto, A. Arabidopsis *xanthine dehydrogenase* mutants defective in purine degradation show a compromised protective response to drought and oxidative stress. *Plant Biotechnol.* **31**(2), 173–8 (2014).
48. Moradi, F. Comparison of the Drought Stress Responses of Tolerant and Sensitive Wheat Cultivars during Grain Filling: Impact of Invertase Activity on Carbon Metabolism during Kernel Development. *J. Agr. Sci.* **3**(2), 32–44 (2011).
49. Moumeni, A., Satoh, K. & Kondoh, H. *et al.* Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. *BMC Plant Biol.* **11**, 174 (2011).
50. Dong, Y., Fan, G., Zhao, Z. & Deng, M. Transcriptome expression profiling in response to drought stress in *Paulownia australis*. *Int. J. Mol. Sci.* **15**, 4583–4607 (2014).
51. Rapala-Kozik, M., Kowalska, E. & Ostrowska, K. Modulation of thiamine metabolism in *Zea mays* seedlings under conditions of abiotic stress. *J. Exp. Bot.* **59**(15), 4133–4143 (2008).
52. Yao, L. M., Wang, B., Cheng, L. J. & Wu, T. L. Identification of Key Drought Stress-Related Genes in the hyacinth bean. *PLoS One* **8**(3), e58108 (2013).
53. Edwards, C. E., Parchman, T. L. & Weekley, C. W. Assembly, Gene Annotation and Marker development Using 454 Floral Transcriptome Sequences in *Zizyphus Celata* (Rhamnaceae), a Highly Endangered, Florida Endemic Plant. *DNA Res.* **19**, 1–9 (2012).
54. Goyal, E. *et al.* Transcriptome profiling of the salt-stress response in *Triticum aestivum* cv. Kharchia Local. *Sci. rep.* **6** (2016).
55. Sharma, R. K., Bhardwaj, P., Negi, R., Mohapatra, T. & Ahuja, P. S. Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biol.* **9**, 53 (2009).
56. Zhang, J., Liang, S. & Duan, J. *et al.* *De novo* assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC genomics* **13**, 90 (2012).
57. Zeng, S., Xiao, G. & Guo, J. *et al.* Development of an EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC genomics* **11**, 94 (2010).
58. Li, X., Luo, J. & Yan, T. *et al.* Deep Sequencing-Based Analysis of the *Cymbidium ensifolium* Floral Transcriptome. *PLoS one* **8**, e85480 (2013).

59. Poncet, V., Rondeau, M. & Tranchant, C. *et al.* SSR mining in coffee tree EST databases: potential use of EST–SSRs as markers for the *Coffea* genus. *Mol. Genet. Genomics* **276**(5), 436–449 (2006).
60. Dure, L. & Galau, G. A. Developmental Biochemistry of Cottonseed Embryogenesis and Germination: XIII. Regulation of biosynthesis of principal storage proteins. *Plant Physiol.* **68**, 187–94 (1981).
61. Kantety, R. V., La Rota, M., Matthews, D. E. & Sorrells, M. E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* **48**, 501–510 (2002).
62. Kumpatla, S. P. & Mukhopadhyay, S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **48**, 985–998 (2005).
63. Pang, T., Guo, L. & Shim, D. *et al.* Characterization of the Transcriptome of the Xerophyte *Ammopiptanthus mongolicus* Leaves under Drought Stress by 454 Pyrosequencing. *PLoS One*, **10**(8), e0136495 (2015).
64. Yang, D. H. *et al.* Expression of Arabidopsis glycine-rich *AtGRP2* or *AtGRP7* improves grain yield of rice (*Oryza sativa*) under drought stress conditions. *Plant Sci.* **214**, 106–12 (2014).
65. Kretzschmar, T., Burla, B., Lee, Y., Martinoia, E. & Nagy, R. Functions of ABC transporters in plants. *Essays Biochem.* **50**, 145–160 (2011).
66. Xu, D. *et al.* Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* **110**, 249–257 (1996).
67. Sivamani, E., Bahieldin, A. & Wraith, J. M. *et al.* Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. *Plant Sci.* **155**, 1–9 (2000).
68. Cheng, Z., Targolli, J., Huang, X. & Wu, R. Wheat LEA genes, PMA80 and PMA 1959, enhance dehydration tolerance of transgenic rice (*Oryza sativa* L.). *Mol. Breed.* **10**, 71–82 (2002).
69. Zhou, Q. Y., Tian, A. G. & Zou, H. F. *et al.* Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21*, and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol. J.* **6**(5), 486–503 (2008).
70. Ren, X. Z., Chen, Z. Z. & Liu, Y. *et al.* *ABO3*, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in Arabidopsis. *Plant J.* **63**(3), 417–29 (2010).
71. Okay, S., Derelli, E. & Unver, T. Transcriptome-wide identification of bread wheat WRKY transcription factors in response to drought stress. *Mol. Genet. Genomics* **289**(5), 765–781 (2014).
72. Chen, H. *et al.* Roles of Arabidopsis *WRKY18*, *WRKY40* and *WRKY60* transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biol.* **10**, 281 (2010).
73. Shao, H. B., Liang, Z. S. & Shao, M. A. Changes of anti-oxidative enzymes and MDA content under soil water deficits among 10 wheat (*Triticum aestivum* L.) genotypes at maturation stage. *Colloids Surf. B Biointerfaces* **45**, 7–13 (2005).
74. Luna, C. M. *et al.* Drought controls on H₂O₂ accumulation, catalase (CAT) activity and CAT gene expression in wheat. *J. Exp. Bot.* **56**, 417–423 (2005).
75. Verslues, P. E. *et al.* Interaction of SOS2 with Nucleoside Diphosphate Kinase 2 and Catalases Reveals a Point of Connection between Salt Stress and H₂O₂ Signaling in *Arabidopsis thaliana*. *Mol. Cell. Biol.* **27**, 7771–7780 (2007).
76. Xing, Y., Jia, W. & Zhang, J. AtMEK1 mediates stress-induced gene expression of CAT1 catalase by triggering H₂O₂ production in *Arabidopsis*. *J. Exp. Bot.* **58**(404), 2969–2981 (2007).
77. Moloudi, F., Navabpour, S., Soltanloo, H., Ramazanpour, S. S. & Sadeghipour, H. Catalase and Metallothionein genes expression analysis in wheat cultivars under drought stress condition. *J. Plant Mol. Breed.* **1**(2), 54–68 (2013).

Acknowledgements

This work was supported by the grants, C30033 of the National Agricultural Innovation Project (NAIP) and National Innovations in Climate Resilient Agriculture (NICRA) of the Indian Council of Agricultural Research (ICAR), New Delhi, India. We express sincere thanks to the Project Director, ICAR-National Research Centre on Plant Biotechnology for providing the 454 GS-FLX sequencing platform and other lab facilities to carry out this research work. We are also highly thankful to the Director, ICAR-Indian Agricultural Research Institute, New Delhi, India, for providing the facilities for seed germination and plant stress experiments at the National Phytotron Facility. We sincerely appreciate suggestions and encouragement by Dr. T. Mohapatra, Secretary, Department of Agricultural Research and Education & Director General, Indian Council of Agricultural Research, New Delhi, during the course of study.

Author Contributions

J.C.P. conceived and designed the experiments and coordinated the study, R.Y. and S.A.L. carried out the experiments and wrote the manuscript, R.Y., S.A.L., K.G. and J.C.P. analyzed the data, J.C.P. and K.G. Contributed reagents/materials/analysis tools, J.C.P. and N.K.S. checked and performed the proofing of the final manuscript. All authors have read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-20548-1>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018