ISSN 1021-4437, Russian Journal of Plant Physiology, 2017, Vol. 64, No. 5, pp. 728–737. © Pleiades Publishing, Ltd., 2017.

RESEARCH PAPERS

Transcriptomic Analysis of *Camellia oleifera* in Response to Drought Stress Using High Throughput RNA-seq[1]

H. Yang[a, d, 2], H. Y. Zhou[a, d, 2], X. N. Yang[a, d], J. J. Zhan[a, d], H. Zhou[a, d, e], C. Wang[a, d], Y. Yu[a, d], X. Y. Lu[a, d, e], Y. Z. Chen [b, c, *], and Y. Tian[a, b, d, e, **]

[a] *Hunan Province University Key Laboratory for Agricultural Biochemistry and Biotransformation, Changsha 410128, China*

[b] *National Engineering Research Center for Oil-tea Camellia, Changsha 410004, China*

[c] *Hunan Academy of Forestry, Changsha 410004, China*

[d] *College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha 410128, China*

[e] *Hunan Co-Innovation Center for Ultilization of Botanical Functional Ingredients, Changsha 410128, China*

**e-mail: chenyongzhong06@163.com*

***e-mail: tianyun@hunau.edu.cn*

Received April 18, 2016

Abstract—*Camellia oleifera* Abel, a woody oil plant of major economic value, has strong ability for stress resistance. However, insufficient genetic and genomic information hinders the research into the mechanisms of its stress response. In this work, Illumina Genome sequencing platform was used for de novo assembling the transcriptomes of leaves from *C. oleifera* seedlings grown under optimal (control) and drought conditions. A total of 66570 unigenes with a mean length of 659.78 bp were assembled, amongst which 35259 unigenes could be annotated using the NCBI nr database, Swiss-Prot protein database, Cluster of Orthologous Groups of protein (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database. In addition, 10869 simple sequence repeats (SSRs) were mined in the leaf transcriptome of *C. oleifera*. In a comparative transcriptome analysis, when large numbers of differentially expressed unigenes (DEUs) were detected at different stages of drought stress, most unigenes were downregulated under the stress. In the KEGG pathway enrichment analysis, some important KEGG metabolic pathways of *C. oleifera* were discovered, such as circadian rhythm, flavone and flavonol biosynthesis, and ribosomal structure. Our studies provide a comprehensive map of physiological and molecular responses of *C. oleifera* to drought stress.

Keywords: Camellia oleifera, drought resistance response, transcriptome analysis, RNA-seq

DOI: 10.1134/S1021443717050168

INTRODUCTION

Drought is one of the most acute abiotic stresses, which negatively affects the global agricultural production and food security [1]. Plants cannot escape from the habitat because of the sessile nature; therefore they tend to cope with drought stress by some particular strategies such as drought avoidance and/or tolerance. Both ways have been shown to involve the induction of specific gene expression and the accumulation of particular proteins, such as dehydrins, key enzymes for osmolyte biosynthesis, and detoxification enzymes. Moreover, a number of the regulatory genes for drought response, such as transcription factors and protein kinase encoding genes, were analyzed in details [2]. Therefore, it is important to clarify the molecular mechanisms of the drought response in plants and develop the drought-tolerant crops by molecular techniques.

Camellia oleifera Abel, an evergreen shrub, is widely grown in the low mountain area of southern China, with its planting area amounting to approximately four million hectares. Since its seed oil has become popular for a healthy lifestyle, many studies focused on its breeding, high-yield cultivation, pest and disease control, characteristic constituents and

[1] The article is published in the original.
Supplementary materials are available for this article at DOI 10.1134/S1021443717050168 and are accessible for authorized users.

[2] These authors contributed equally to this work.

Abbreviations: CAT—catalase; CHS—chalcone synthase; COGs—cluster of orthologous groups of proteins; CRY—cryptochrome; CTAB—hexadecyltrimethyl ammonium bromide; DEUs—differentially expressed unigenes; FDR—false discovery rate; GI—GIGANTEA; KEGG—Kyoto Encyclopedia of Genes and Genomes; MDA—malondialdehyde; NCBI—National Center for Biotechnology Information; PHYA—phytochrome A; PHYB—phytochrome B; PIF3—phytochrome-interacting factor 3; POD—peroxidase; PRR7—pseudo-response regulator 7; RNA-seq—RNA sequencing; RPKM—reads per kb per million reads; SSRs—simple sequence repeats; STEM—short time-series expression miner software.

developmental biology [3–5]. The plant is strongly resistant to drought, yet there are few reports about its physiological and biochemical response to drought stress. Thus, our transcriptomic analysis may shed some light on the molecular mechanism of *C. oleifera* response to drought.

For genetic analysis of non-model organisms, next-generation sequencing technologies offer a new approach with the advantage of high efficiency and low cost. On the other hand, when genome information of most plants is unavailable, RNA sequencing (RNA-seq) is an efficient method to study stress adaptation and other biological features [6]. In this report, to uncover the transcriptome-wide response to drought in *C. oleifera*, the global genetic expression profiles of *C. oleifera* were analyzed under drought conditions using the IlluminaHiSeq 2000 and Genome Analyzer IIx platforms. To the best of our knowledge, there were few previous reports about transcriptome of *C. oleifera* [3, 4], and our data will provide new insights into the molecular mechanisms of drought response in the wooden plants.

MATERIALS AND METHODS

Plant material. Population of *Camellia oleifera* Abel cultivar Xianglin 4 was grown in the greenhouse at 25°C with a 14-h light/10-h dark photoperiod. Two-year-old plants were about 30–35 cm high. The control plants were watered to field capacity, whereas for drought treatment, the seedlings were subjected to mild and severe drought stress followed by recovery. The mature and fully expanded leaves (third to fifth leaves from the apex) were sampled after 6 hours of continuous illumination and mixed equally for each of the three replicate groups.

Physiological measurements. The malondialdehyde (MDA) concentration was measured using the thiobarbituric acid method [7]. Soluble sugar content was measured using the anthrone colorimetry method [8]. Proline content was measured using the acid ninhydrin method [9]. Peroxidase (POD) activity was determined by measuring the oxidation of guaiacol. Catalase (CAT) activity was determined as H_2O_2 consumption [10]. The measurements were taken on six replicates of the leaves from the same group.

Library preparation and sequencing. Total RNA was isolated with the modified Hexadecyltrimethyl Ammonium Bromide (CTAB) method [11] and extracted to construct libraries for gene expression profiling analyses. The reference transcriptome library was constructed by mixing equal amounts of RNA from each group sample. The sequencing was performed by Illumina HiSeq 2000 sequencer (GeneDenovo, China).

***De novo* assembly and annotation.** Prior to assembling libraries, raw reads were filtered to remove the reads with adaptors, reads with unknown nucleotide content exceeding 10% and low-quality reads containing over 50% bases with q-value ≤ 5. The contigs were de novo assembled by the short reads assembling program Trinity [12]. After sequence assembly, the resultant contigs were joined into scaffolds using the read mate pairs, and the scaffolds were clustered using TGI Clustering tools [13]. To annotate the unigenes, the BLASTx program (http://www.ncbi.nlm.nih.gov/BLAST/) was used with an E-value threshold of $1e^{-5}$ according to National Center for Biotechnology Information (NCBI) nr database (http://www.ncbi.nlm.nih.gov), the Swiss-Prot protein database (http://www.expasy.ch/sprot), the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg), and the COG database (http://www.ncbi.nlm.nih.gov/COG). The sequence direction of the unigenes was determined according to the best alignment results. When a unigene could not be aligned, sequence direction would be confirmed using ESTScan program [14]. KEGG pathway annotation was realized using the Blastall software against the KEGG database [15]. The GO annotation was performed by Blast2GO software (http://www.blast2go.com/b2ghome) [16]. The functional classification of the unigenes was performed using the WEGO software [17].

Gene expression annotation. Clean reads were mapped into the transcriptome reference database with the SOAPaligner/soap2 software [18]. Not more than 2 mismatch bases were permitted, and unique mapped reads were obtained. The number of unique-match reads was calculated and normalized to RPKM (Reads Per Kb per Million Reads) for gene expression analysis [19]. Unigene expression values from different treatments were compared as described by Audic and Claverie [20]. The p-value corresponds to differential gene expression test, and False Discovery Rate (FDR) is the method to determine the threshold of p-value. In this experiment, the differentially expressed genes from each group were restricted with FDR ≤ 0.001 and the absolute value of $\log_2$ Ratio ≥ 1. Statistical analysis of the differential gene expression was performed using the R package edgeR [21]. Next the differentially expressed genes were subjected to GO classifications using WEGO [17], and KEGG pathway annotation was performed by the Blastall software against the KEGG database [15].

RESULTS

The Physiological Responses of Camellia oleifera to Drought

The phenotypic and physiological responses of *C. oleifera* to drought were assessed using two-year-old seedlings exposed to progressive water loss. The leaves of seedlings started to droop after mild drought and then became paler and more wilted after severe stress. After recovery, the plants showed less wilting

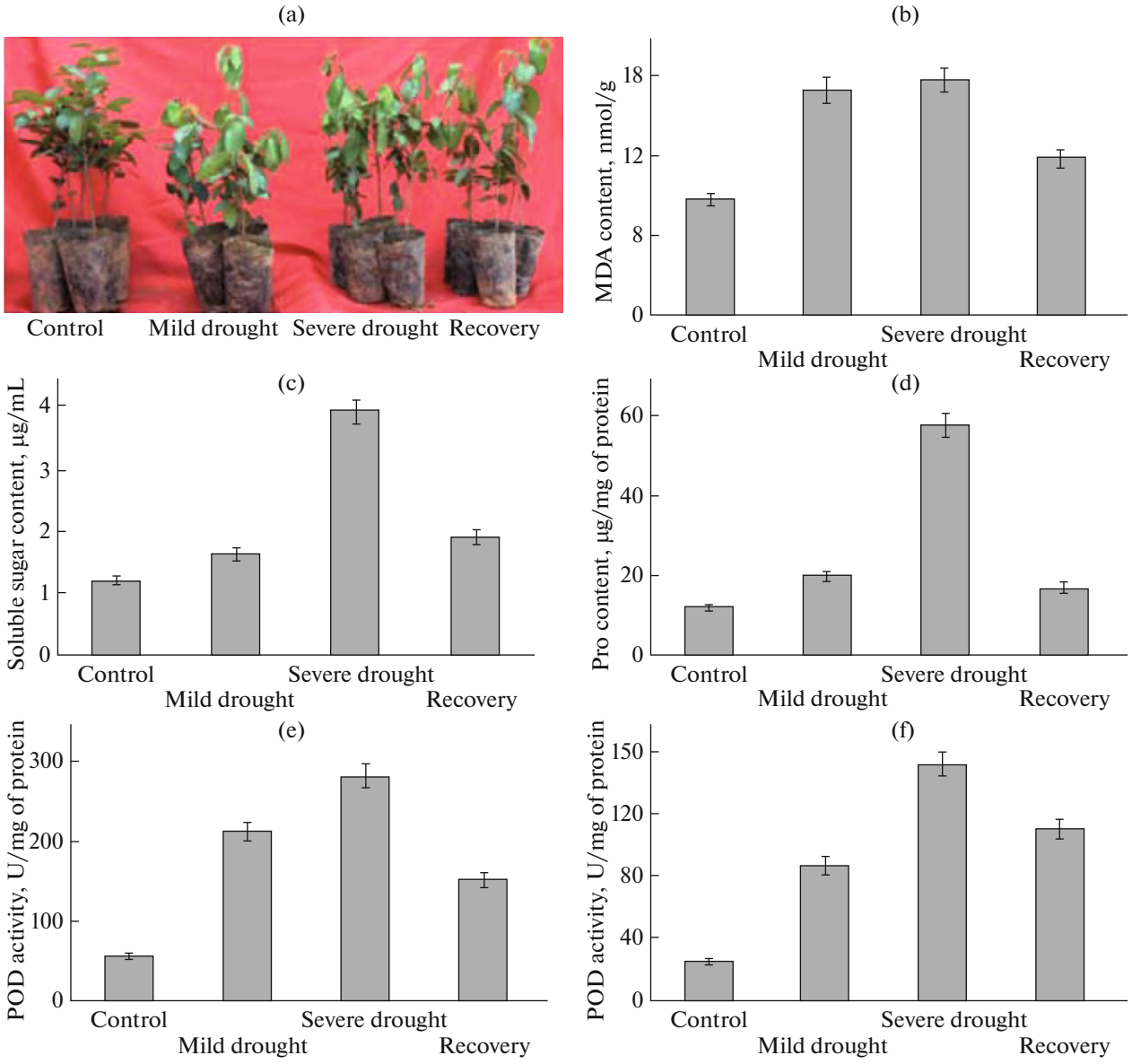

Fig. 1. Physiological manifestations of *Camellia oleifera* leaves in response to drought: (a) phenotype; (b) malondialdehyde (MDA) content; (c) soluble sugar content; (d) proline content; (e) peroxidase (POD) activities; (f) catalase (CAT) activities. Values represent the means and SD from six replicates.

(Fig. 1a). Such response to drought significantly promoted the accumulation of malondialdehyde (MDA), soluble sugar, proline, as well as peroxidase (POD) and catalase (CAT) activities. As compared to the control plants, the leaves of *C. oleifera* under mild and severe drought conditions showed substantial increases in the content of MDA, soluble sugar and proline, and then these indices decreased dramatically after renewed cultivation (Fig. 1b–1f). These results indicate drought injury in *C. oleifera* leaves.

Illumina Sequencing and de novo Assembly

Our transcriptome sequencing project produced a total of 91744458 raw reads with 11468057250 nt (GC content 46.02%). Of note, the Q20 data percentage (an error probability <1%) of the final sequence generated in this study (97.11%) indicated that sequencing throughput and quality were acceptable for further analysis. *De novo* assembly of all clean reads produced 92.036 contigs with a mean contig size of 733.25 bp. Then the assembled scaffolds were further clustered into 66570 unigenes (Table 1). The size distributions of these unigenes, CDS through blast and ESTscan respectively, are shown in Table S1. 52.42% (34897/66570) of unigenes were annotated by BLASTx (E-value<$1e^{-5}$) using the NCBI nr database, while 24315 were annotated using the Swiss-Prot protein database. In addition, 10943 and 9801 unigenes were annotated

Table 1. Summary of the *Camellia oleifera* tests for transcriptome assembly

Name	Number of sequences	
Total clean reads	91744458	
Total nucleotides (nt)	11468057250	
Total number of Contigs	92036	Mean length 733.25
Total number of Unigenes	66570	Mean length 659.78 N50 1066
Unigene annotation	Number of annotations	
tBLAST×known transcripts	35259	
Unknown or predicted transcripts	31311	
NR	34897	
Swiss-Prot	24315	
KEGG	9801	
COG	10943	
GO	15815	

according to the Cluster of Orthologous Groups of protein (COG) and KEGG database respectively. About 8.61% (5735/66570) unigenes were assigned to a homolog in all four databases (Fig. 2a). Based on the NCBI nr database, 21.32% of the unigenes showed homology ($1e^{-20} < E\text{-value} < 1e^{-5}$), 45.76% of them showed strong homology ($1e^{-100} < E\text{-value} < 1e^{-20}$) and the remaining 32.92%, high homology ($E\text{-value} < 1e^{-100}$) to available plant sequences (Fig. 2b). As shown in Fig. 2c, 19 and 112 unigenes were annotated to 3 top-hit species, including *Vitis vinifera*, *Theobroma cacao* and *Solanum lycopersicum*. Hence, these transcriptome data might expand the genetic resources available for *C. oleifera* in the public database and could become the basis of further research on the molecular genetics and functional genomics of *C. oleifera*.

GO, COGs and KEGG Ontology (KO) Classification

A total of 66570 unigenes from *C. oleifera* were assigned for GO analysis. Based on the sequence similarity, 15815 sequences (23.76%) were further classified into three main categories (biological processes, cellular components and molecular functions) and 71 subcategories. Majority of the unigenes are involved in various biological processes and other are related to molecular functions and cellular components. Analysis of their further distribution to various subcategories revealed that the majority of the unigenes were involved in metabolic and cellular processes (55.01 and 52.25%, respectively). About 46.06% and 44.78% of unigenes under the cellular component category were involved in cell and cell part activities, respectively. As to the molecular functions, the metabolic processes (57.15%), binding (50.28%) and catalytic activities (49.81%) accounted for most of the unigenes. The smallest groups were organelles of biological processes. As for metallochaperone activity and translation regulator activity of molecular functions, only one unigene was predicted from each group (Fig. S1A).

To further evaluate the completeness of the transcriptomic library and effectiveness of the annotation process, we collated the annotated sequences with the COG classifications and KEGG metabolic pathways. Based on sequence homology, 10943 (16.44%) unigenes were categorized into 24 COG groups (Fig. S1B). The highest represented biological processes included 'general function prediction only' (2993; 27.35%), followed by 'posttranslational modification, protein turnover, chaperones' (1523; 13.92%), 'replication, recombination and repair' (1444; 13.20%), 'transcription' (1404; 12.83%), 'signal transduction mechanisms' (1276; 11.66%) and 'translation, ribosomal structure and biogenesis' (1010; 9.23%). The cell motility and nuclear structure group contained only 8 and 2 unigenes, respectively. The unigene metabolic pathways were investigated using the KEGG annotation system. In the 34897 unigene nr hits, a total of 9801 unigenes were associated directly with 124 predicted KEGG metabolic pathways, and the number of unigenes in different pathways ranged from 1 to 2591. Among them, metabolic pathway group ko01100, which included 2591 unigenes (26.44%), was found to be significantly larger than other pathways; it was followed by biosynthesis of secondary metabolites (ko01110, 1288 unigenes, 13.14%), ribosomal structure (ko03010, 495 unigenes, 5.05%), protein processing in endoplasmic reticulum (ko04141, 368 unigenes, 3.75%) and oxidative phosphorylation (ko00190, 328 unigenes, 3.35%) (Fig. S1C). The smallest group was indole alkaloid biosynthesis and anthocyanin biosynthesis, both contained only one unigene (Table S2). All these analyses indicated that most of the unigenes did not match any known proteins, although they

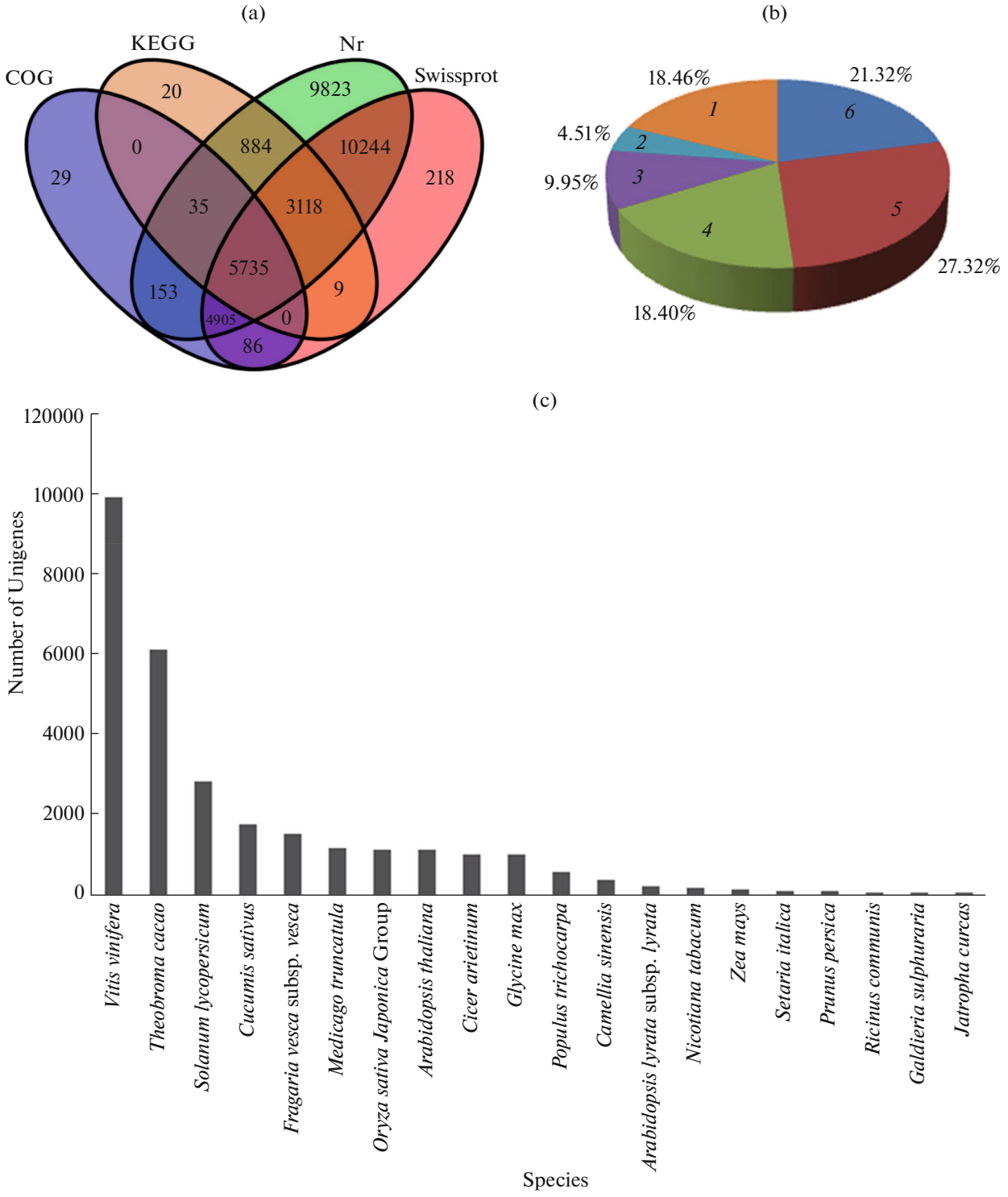

Fig. 2. Homology search for *Camellia oleifera* unigenes: (a) venn diagram of number of unigenes annotated by BLASTx with a cut-off E-value $1e^{-05}$ against protein databases (numbers in the circles indicate the number of unigenes annotated by single or multiple databases); (b) *E*-value distribution of the top BLASTx hits against the Nr database. *1*—0; *2*—0 < *E*-value < $1e^{-150}$; *3*—$1e^{-150}$ < *E*-value < $1e^{-100}$; *4*—$1e^{-100}$ < *E*-value < $1e^{-50}$; *5*—$1e^{-50}$ < *E*-value < $1e^{-20}$; *6*—$1e^{-20}$ < *E*-value < $1e^{-5}$; c—number of unigenes matching the 20 top species using BLASTx in the Nr database.

Table 2. The statistical data of SSRs in the *Camellia oleifera*

Statistical Item	Numbers
Total number of sequences examined	66570
Total size of examined sequences, bp	43921855
Total number of identified SSRs	10869
Number of SSR containing sequences	9049
Number of sequences containing more than 1 SSR	1518
Number of SSRs present in compound formation	697
Dinucleotide	6452
Trinucleotide	3097
Tetranucleotide	660
Pentanucleotide	335
Hexanucleotide	325

probably somehow contributed to drought response in *C. oleifera*.

Simple Sequence Repeat Mining

Transcriptome data are widely used as a source for mining molecular markers such as the simple sequence repeats (SSRs). Therefore, the transcripts generated in this study were mined using the MISAPerl script (http://pgrc.ipk-gatersleben.de/misa/) so as to identify SSRs (dinucleotide to hexanucleotide repeats) in transcripts of *C. oleifera* leaves. Total of 10869 SSRs were identified in 9049 transcripts of *C. oleifera* (Table 2). The average frequency of SSRs was found to be one SSR per 4.04 kb of the transcriptome sequence. Moreover, 1518 sequences contained more than one SSR, and 697 SSRs were presented in compound formation. Unlike cereal species such as barley, maize, rice, wheat and sorghum, with trinucleotide repeats as the main type of SSR, the largest portion of SSRs identified in *C. oleifera* were dinucleotides (59.36%) followed by trinucleotides (28.49%) (Table 2). Among the nucleotide repeats, AG/CT predominated (46.9%) followed by AT/AT which accounted for 7.6% of all nucleotide repeats (Fig. S2). To our knowledge, there are only a few reports about SSRs in *C. oleifera* [4], and our report on SSRs will help future genetic characterization and germplasm utilization of *C. oleifera*.

Transcriptional Profiling of Camellia oleifera Leaves under Drought Stress

In order to clarify the transcriptional changes in *C. oleifera* leaves under drought stress, we performed four pairwise comparisons of the transcriptomes as follows: control versus mild drought (C vs. M), control versus severe drought (C vs. S), control versus recovery (C vs. R), and mild drought versus severe drought (M vs. S). Large numbers of differentially expressed unigenes (DEUs) were detected, and the results of the pairwise comparisons of the DEUs are shown in Fig. 3a. It was observed that majority of DEUs was downregulated during the drought stress. A comparison of control and mild drought group showed that 4463 and 8761 DEUs were upregulated and downregulated, respectively. By contrast, 305 DEUs were downregulated and 159 DEUs were upregulated in mild drought group compared with severe drought group. To further represent differentially expressed genes during the drought stress response, we created a heat map through hierarchical clustering across the four time points. Red and blue colors indicate up- and downregulated transcripts, respectively, from both control and drought treated leaves. FDR ≤ 0.001 and the maximum value of $|\log_2 (\text{ratio of stress/control})| \geq 1$ were used as cut-off to evaluate significant differences in expression (Fig. 3b).These results indicated that the changes of gene expression mainly occurred at the early stage of drought stress response in *C. oleifera*.

Furthermore, the DEUs were sorted into clusters using Short Time-series Expression Miner software (STEM) to determine the optimal number of clusters and in this way to classify gene expression patterns under drought stress. Among 20 representative clusters, we selected five major clusters with significant enrichment (p-value ≤ 0.05) that displayed distinct expression profiles during the drought response in *C. oleifera* (Fig. 4a–4e). The unigenes (9072) in the clusters A to C showed a decrease expression, while those (5181) in the clusters D to E indicated enhanced expression under the drought stress. Next, DEUs of the significantly enriched cluster groups was subjected to GO-term analysis. They were also classified into three main categories: cellular components, biological processes and molecular functions. Under the biological process category, many DEUs were classified into metabolic process, cellular process and single-organism process categories. Most of cellular component categories belonged to cell, cell part and organelle and

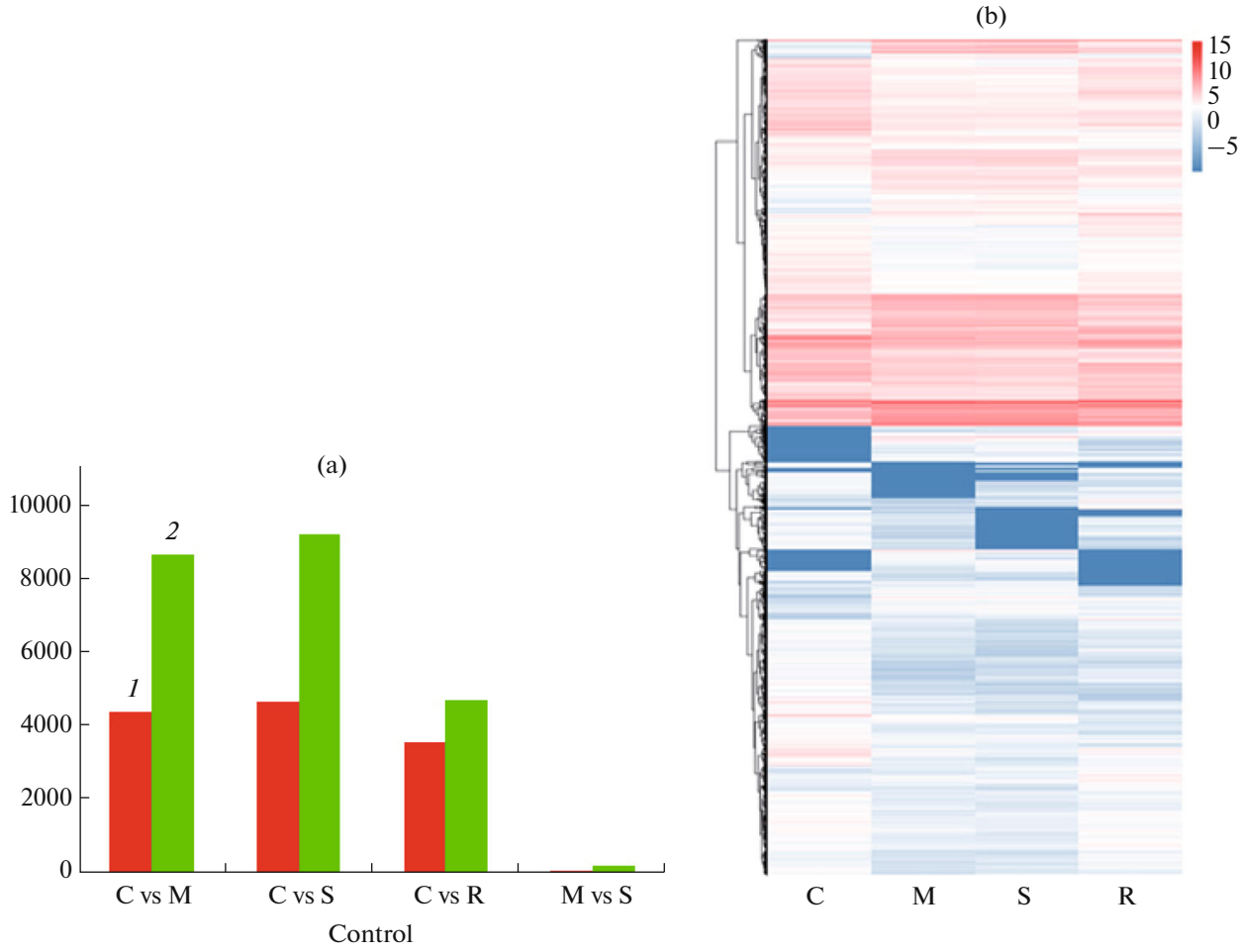

Fig. 3. Statistics of differentially expressed genes in each pairwise comparison (a) and heatmap of differentially expressed transcript isoforms across the four time points (b). *1*—Upregulated genes; *2*—downregulated ones. False Discovery Rate (FDR) ≤ 0.001 and the maximum value of $|\log_2$ (ratio of stress/control)$| \geq 1$ were used as cut-off to evaluate significant differences in expression. C—control group; M—mild drought group; S—severe drought group; R—recovery group.

to molecular function subcategories, with metabolic processes, catalytic activities and binding being the most abundant subcategories. The detailed lists with all downregulated unigenes (cluster A, B and C) and upregulated unigenes (clusters D and E) under drought stress, including unigene descriptions, changes in expression, KEGG metabolic pathways and their GO annotation are presented in Table S3 (A–E).

The Important KEGG Metabolic Pathways of C. oleifera under Drought Stress

The KEGG pathway enrichment analysis demonstrated that the major KEGG metabolic pathways of *C. oleifera* affected by drought stress included circadian rhythm, flavone and flavonol biosynthesis, and ribosomal structure (Table S4 and Fig. S3–5). The detailed lists with all unigenes include the changes in expression, unigenes descriptions (Nr annotation, Swissprot annotation, COG function description, ko definition), and their GO annotation.

Previous studies have demonstrated that drought stress could affect the expression levels of putative circadian rhythm genes in plants such as *Arabidopsis*, soybean [22–24]. GIGANTEA (GI) is also known to play some role in drought tolerance and circadian clock control [25]. In this experiment, 33 unigenes that belonged to the circadian rhythm pathway were downregulated during the drought stress in *C. oleifera*, and these unigenes were upregulated after the recovery. These unigenes might encode the Phytochrome B (PHYB) (Unigene0030838, Unigene0030839, Unigene0030840), Phytochrome A (PHYA) (Unigene0043190, Unigene0043191), Phytochrome-Interacting Factor 3 (PIF3) (Unigene0030334, Unigene0045142, Unigene0045143), Pseudo-response Regulator 7 (PRR7) (Unigene0022103, Unigene0025709, Unigene0031899, Unigene0035040, Unigene0045762), Cryptochrome

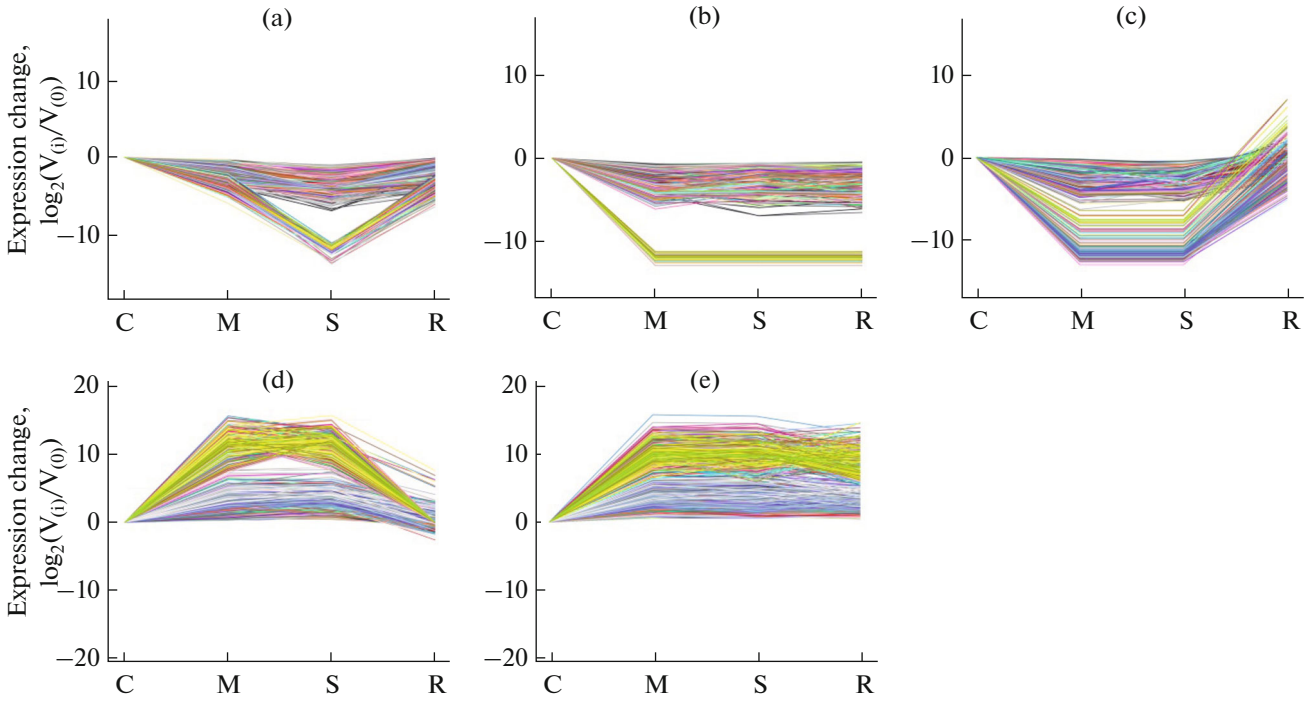

Fig. 4. The five major clusters with significant enrichment trend displayed distinct profiles of expression during the drought response in *Camellia oleifera*: (a) the expression profile of unigenes was 0.0, −1.0, −2.0 and −1.0 of $\log_2(V_{(i)}/V_{(0)})$; (b) the expression profile of unigenes was 0.0, −1.0, −1.0 and −1.0 of $\log_2(V_{(i)}/V_{(0)})$; (c) the expression profile of unigenes was 0.0, −1.0, −1.0 and 0.0 of $\log_2(V_{(i)}/V_{(0)})$; (d) the expression profile of unigenes was 0.0, 1.0, 1.0 and 0.0 of $\log_2(V_{(i)}/V_{(0)})$; (e) the expression profile of unigenes was 0.0, 1.0, 1.0 and 1.0 of $\log_2(V_{(i)}/V_{(0)})$. C—control group; M—mild drought group; S—severe drought group; R—recovery group.

(CRY) (Unigene0029679, Unigene0038643), GIGANTEA (GI) (Unigene0013010, Unigene0016976, Unigene0018657, Unigene0018658, Unigene0018659, Unigene0024567, Unigene0029354, Unigene0047505, Unigene0047506), and Chalcone Synthase (CHS) (Unigene0043143, Unigene0043144, Unigene0043145), respectively (Table S4A and Fig. S3). These results also indicate that the drought stress could significant affect circadian rhythm in *C. oleifera*.

Change in the secondary metabolism is one of the stress responses in plants. It has been reported that in *Arabidopsis,* a class of specialized metabolites, such as flavonoids and flavonols, with strong radical scavenging activity, could alleviate drought stress [26, 27]. In *C. oleifera*, six unigenes relating to the flavone and flavonol biosynthesis pathways were downregulated under drought, and they were upregulated after renewed growth; these genes include UDP-glycosyltransferase (Unigene0023780, Unigene0024708, Unigene0024710), flavonoid 3',5'-hydroxylase 2 (Unigene0036522, Unigene0036523), and anthocyanin 3-o-galactosyl transferase (Unigene0042525), respectively (Table S4B and Fig. S4). These results showed that the drought stress could partially damage the non-enzymatic antioxidant system; they also demonstrated that the secondary metabolites were involved in the stress response in *C. oleifera*.

To date, it is known that ribosomal proteins are involved in response to abiotic stresses; however, the mechanism of this involvement is still unclear [28–30]. In this research, 197 unigenes relating to the ribosomal structure were upregulated under drought stress, and during the recovery these unigenes were downregulated in *C. oleifera*. These unigenes encode 49 proteins of large ribosome subunit and 29 proteins of small subunit, respectively (Table S4C and Fig. S5). Therefore, the ribosomal proteins may play the vital roles in the response to drought stress in *C. oleifera*.

DISCUSSION

High-throughput transcriptome analysis provides an unbiased approach for understanding gene functions in response to different conditions, especially in non-model organisms with unsequenced genome. Transcriptomic changes in *Camellia oleifera* have been previously reported [4, 5]. In this study, we combined physiological analyses with RNA-Seq to investigate the physiological responses to drought stress with transcriptomic changes. As a result, we assembled leaf transcriptome of *C. oleifera* seedlings grown under control condition and drought stress which comprised 66570 unigenes of average length 659.78 bp. Among these unigenes, about 52.97% (35259) were annotated

using the NCBI nr database, Swiss-Prot protein database, COG and KEGG database. Furthermore, these unigenes were annotated into 21 GO subcategories, 24 COG groups, and 124 KEGG pathways.

Analysis of the whole transcriptome enabled us to better understand expression changes as related to drought tolerance. Among the DEUs, several categories associated with carbon metabolism were enriched. The results of our study provided new information pertaining to the differential regulation of genes in response to drought. These data suggest that 33 unigenes belonging to the circadian rhythm pathway were downregulated during the drought, and these unigenes were upregulated after the recovery. Many genes involved in various syntheses were differentially expressed in response to the drought stress. They are major targets for further functional and physiological studies to unravel the complex mechanisms associated with drought tolerance in *C. oleifera*.

DATA ARCHIVING STATEMENT

The sequence reads were submitted to the NCBI Sequence Read Archive (SRA) database, and the accession numbers of SRA was SRP068898. The supplementary materials for this project have been archived on Figshare (https://figshare.com/account/home#/collections/3501180).

ACKNOWLEDGMENTS

This research was supported by the Opening Project of National Engineering Research Center for Oil-tea Camellia, project no. 2014CY03, the Key Scientific Research Fund of Hunan Provincial Education Department (project no. 13A041), the Hunan Provincial Innovation Foundation for Postgraduate, project no. CX2015B240 and the Construct Program of the Key discipline in Hunan Province.

REFERENCES

1. Deikman, J., Petracek, M., and Heard, J.E., Drought tolerance through biotechnology: improving translation from the laboratory to farmers' fields, *Curr. Opin. Biotechnol.*, 2012, vol. 23, pp. 243–250.
2. Hu, H. and Xiong, L., Genetic engineering and breeding of drought-resistant crops, *Annu. Rev. Plant Biol.*, 2014, vol. 65, pp. 715–741.
3. Liao, T., Yuan, D.Y., Zou, F., Gao, C., Yang, Y., Zhang, L., and Tan, X.F., Self-sterility in *Camellia oleifera* may be due to the prezygotic late-acting self-incompatibility, *PLoS One*, 2014, vol. 9: e99639.
4. Xia, E.H., Jiang, J.J., Huang, H., Zhang, L.P., Zhang, H.B., and Gao, L.Z., Transcriptome analysis of the oil-rich tea plant, *Camellia oleifera*, reveals candidate genes related to lipid metabolism, *PLoS One*, 2014, vol. 9: e104150.
5. Tai, Y., Wei, C., Yang, H., Zhang, L., Chen, Q., Deng, W., Wei, S., Zhang, J., Fang, C., Ho, C., and Wan, X., Transcriptomic and phytochemical analysis of the biosynthesis of characteristic constituents in tea (*Camellia sinensis*) compared with oil tea (*Camellia oleifera*), *BMC Plant Biol.*, 2015, vol. 15, p. 190.
6. McGettigan, P.A., Transcriptomics in the RNA-seq era, *Curr. Opin. Chem. Biol.*, 2013, vol. 17, pp. 4–11.
7. Schmedes, A. and Hølmer, G., A new thiobarbituric acid (TBA) method for determining free malondialdehyde (MDA) and hydroperoxides selectively as a measure of lipid peroxidation, *J. Am. Oil Chem. Soc.*, 1989, vol. 66, pp. 813–817.
8. DuBois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., and Smith, F., Colorimetric method for determination of sugars and related substances, *Anal. Chem.,* 1956, vol. 28, pp. 350–356.
9. Claussen, W., Proline as a measure of stress in tomato plants, *Plant Sci.,* 2005, vol. 168, pp. 241–248.
10. Dionisio-Sese, M.L. and Tobita, S., Antioxidant responses of rice seedlings to salinity stress, *Plant Sci.*, 1998, vol. 135, pp. 1–9.
11. Wang, X., Xiao, H., Chen, G., Zhao, X., Huang, C., Chen, C., and Wang, F., Isolation of high-quality RNA from *Reaumuria soongorica*, a desert plant rich in secondary metabolites, *Mol. Biotechnol.*, 2011, vol. 48, pp. 165–172.
12. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.*, 2011, vol. 29, pp. 644–652.
13. Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., and Quackenbush, J., TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets, *Bioinformatics*, 2003, vol. 19, pp. 651–652.
14. Iseli, C., Jongeneel, C.V., and Bucher, P., ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences, *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 1999, pp. 138–148.
15. Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y., KEGG for linking genomes to life and the environment, *Nucleic Acids Res.*, 2008, vol. 36: D480–D484.
16. Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics,* 2005, vol. 21, pp. 3674–3676.
17. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., and Wang, J., WEGO: a web tool for plotting GO annotations, *Nucleic Acids Res.*, 2006, vol. 34: W293–W297.
18. Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J., SOAP2: an improved ultrafast tool for short read alignment, *Bioinformatics,* 2009, vol. 25, pp. 1966–1967.
19. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B., Mapping and quantifying mammalian

transcriptomes by RNA-Seq, *Nat. Methods*, 2008, vol. 5, pp. 621–628.

20. Audic, S. and Claverie, J.M., The significance of digital gene expression profiles, *Genome Res.*, 1997, vol. 7, pp. 986–995.
21. Robinson, M.D., McCarthy, D.J., and Smyth, G.K., edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, 2010, vol. 26, pp. 139–140.
22. Wilkins, O., Bräutigam, K., and Campbell, M.M., Time of day shapes Arabidopsis drought transcriptomes, *Plant J.*, 2010, vol. 63, pp. 715–727.
23. Marcolino-Gomes, J., Rodrigues, F.A., Fuganti-Pagliarini, R., Bendix, C., Nakayama, T.J., Celaya, B., Molinari, H.B., de Oliveira, M.C., Harmon, F.G., and Nepomuceno, A., Diurnal oscillations of soybean circadian clock and drought responsive genes, *PLoS One*, 2014, vol. 9, p. e86402.
24. Rodrigues, F.A., Fuganti-Pagliarini, R., Marcolino-Gomes, J., Nakayama, T.J., Molinari, H.B., Lobo, F.P., Harmon, F.G., and Nepomuceno, A.L., Daytime soybean transcriptome fluctuations during water deficit stress, *BMC Genomics*, 2015, vol. 16, p. 505.
25. Mishra, P. and Panigrahi, K.C., GIGANTEA-an emerging story, *Front. Plant Sci.*, 2015, vol. 6, p. 8.
26. Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., Matsuda, F., Kojima, M., Sakakibara, H., Shinozaki, K., Michael, A.J., Tohge, T., Yamazaki, M., and Saito, K., Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids, *Plant J.*, 2014, vol. 77, pp. 367–379.
27. Nakabayashi, R., Mori, T., and Saito, K., Alternation of flavonoid accumulation under drought stress *in Arabidopsis thaliana, Plant Signal. Behav.*, 2014, vol. 9, p. e29518.
28. Bae, M.S., Cho, E.J., Choi, E., and Park, O.K., Analysis of the Arabidopsis nuclear proteome and its response to cold stress, *Plant J.*, 2003, vol. 36, pp. 652–663.
29. Bai, B., Wu, J., Sheng, W.T., Zhou, B., Zhou, L.J., Zhuang, W., Yao, D.P., and Deng, Q.Y., Comparative analysis of anther transcriptome profiles of two different rice male sterile lines genotypes under cold stress, *Int. J. Mol. Sci.*, 2015, vol. 16, pp. 11 398–11 416.
30. Wang, J., Lan, P., Gao, H., Zheng, L., Li, W., and Schmidt, W., Expression changes of ribosomal proteins in phosphate- and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition, *BMC Genomics*, 2013, vol. 14, p. 783.