RNA-seq Reveals Complicated Transcriptomic Responses to Drought Stress in a Nonmodel Tropic Plant, Bombax ceiba L.

Zhihui Zhou1,2, Huancheng Ma1, Kevin Lin3, Youjie Zhao1, Yuan Chen3,4, Zhi Xiong1, Liuyang Wang3 and Bin Tian1,2

1Key Laboratory of Biodiversity Conservation in Southwest China, State Forestry Administration, Southwest Forestry University, Kunming, China. 2Key Laboratory for Plant Diversity and Biogeography of East Asia, Chinese Academy of Sciences, Kunming, China. 3Department of Molecular Genetics and Microbiology. 4Division of Infectious Diseases, Department of Medicine, Duke University Medical Center, Durham, NC, USA.

Supplementary Issue: RNA: An Expanding View of Function and Evolution

ABSTRACT: High-throughput transcriptome provides an unbiased approach for understanding the genetic basis and gene functions in response to different conditions. Here we sequenced RNA-seq libraries derived from a Bombax ceiba L. system under a controlled experiment. As a known medicinal and ornamental plant, B. ceiba grows mainly in hot-dry monsoon rainforests in Southeast Asia and Australia. Due to the specific growth environment, it has evolved a unique system that enables a physiologic response to drought stress. To date, few studies have characterized the genome-wide features of drought endurance in B. ceiba. In this study, we first attempted to characterize and identify the most differentially expressed genes and associated functional pathways under drought treatment and normal condition. Using RNA-seq technology, we generated the first transcriptome of B. ceiba and identified 59 differentially expressed genes with greater than 1,000-fold changes under two conditions. The set of upregulated genes implicates interplay among various pathways: plants growth, ubiquitin-mediated proteolysis, polysaccharides hydrolyzation, oxidative phosphorylation and photosynthesis, etc. In contrast, genes associated with stem growth, cell division, fruit ripening senescence, disease resistance, and proline synthesis are repressed. Notably, key genes of high RPKM levels in drought are genes associated with stem growth, cell division, fruit ripening senescence, disease resistance, and proline synthesis are repressed. Notably, key genes of high RPKM levels in drought are AUX1, JAZ, and psbS, which are known to regulate the growth of plants, the resistance against abiotic stress, and the photosynthesis process. Furthermore, 16,656 microsatellite markers and 3,071 single-nucleotide polymorphisms (SNPs) were predicted by in silico methods. The identification and functional annotation of differentially expressed genes, microsatellites, and SNPs represent a major step forward and would serve as a valuable resource for understanding the complexity underlying drought endurance and adaptation in B. ceiba.

KEYWORDS: transcriptome, Bombax ceiba, RNA-seq, drought endurance

Introduction

Drought is a severe environment stressor that affects both the productivity and growth of plants.1 Drought tolerance is a complex trait, and the inability to appropriately respond to drought stress generally leads to loss of water in plant cells, decreasing turgor pressure, and a significant rise in the osmolarity of leaf sap in affected plants.2 Plants have evolved to adapt to drought stress using multiple strategies including changes of various morphological and physiological traits: leaf structure, root morphology, extension growth, etc.3,4 Previous studies reported that drought tolerance is associated with hundreds of genes and pathways, including regulation of the hormone abscisic acid (ABA), signal transduction, protein metabolism, and carbohydrate metabolism.4,5 Dehydration and water stress usually stimulate the production of ABA, which further induces various genes and pathways.6 Genes involved in drought stress have been identified through transcriptomic, proteomic, metabolomic, and epigenetic levels.7 However, the molecular basis of drought stress, especially regarding signal transduction and stress adaptation, remains unclear.

By taking advantage of next-generation sequencing, studies have broadened their scope from a few genes to the whole genome and from model species to nonmodel species. Analysis of the whole transcriptome enables us to better understand expression-level changes in response to environmental perturbation. Transcriptomes of nonmodel organisms via de novo transcriptome assembly has been reported for numerous plants.8-12 Similarly, progress has been made in the study of drought resistance and drought endurance by transcriptomic analysis. In Reaumuria soongorica, Shi et al.13
identified 123 candidate genes potentially associated with drought stress. Male *Populus yunnanensis* plants have greater remodeling of the leaf transcriptome in response to drought than their female counterparts.\textsuperscript{14} In *Macrotyloma uniflorum*, the metabolic pathways of valine degradation, gluconogenesis and purine nucleotide degradation are highly affected by drought stress, and transcription factors from NAC (NAM (no apical meristem), ATAF, and CUC (cup-shaped cotyledon) transcription factors) were significantly implicated.\textsuperscript{15} However, the extent to which drought-related genes change their expression level in response to drought stress and the molecular mechanisms that underlie drought resistance are yet to be discovered, especially in nonmodel plants.

*Bombax ceiba*, a diploid deciduous tree ($2n = 72$),\textsuperscript{16} is widely distributed in temperate and tropical regions, such as Southeast Asia, Africa, and Australia.\textsuperscript{17} It has been cultivated as a medicinal and ornamental resource. In addition, *B. ceiba* has been recognized as a new source in the textile industry imbued with slenderer and softer fiber. The ability to grow in arid regions such as dry-hot valleys makes *B. ceiba* an important industrial crop. Although previous studies of *B. ceiba* mainly focused on phytochemistry,\textsuperscript{18} its inherent resistance to drought makes it an ideal system for understanding transcriptome-level adaptations and changes under drought stress. In this study, we used Illumina HiSeq 2500 to sequence two libraries of *B. ceiba* using a controlled approach. In particular, we compared transcriptomic changes of *B. ceiba* between drought and normal conditions. We aim to identify a number of differentially expressed genes, which may be associated with drought resistance, provide a primary analysis of their function and propose drought-resistant mechanisms in *B. ceiba*. To the best of our knowledge, this study represents the first genome-level characterization of drought-related genes in *B. ceiba*. Our data set is an important first step and will serve as a publicly available information platform for future gene expression, genomic, and functional studies.

**Materials and Methods**

Sample preparation and sequencing. The seeds of *B. ceiba* were collected from the dry-hot valley in Honghe Prefecture (23°21′33″ N, 113°12′40″ E). Seeds were grown in the greenhouse in Kunming, Yunnan province, China. To simulate drought stress, plants were cultured under dehydration conditions until they reached constant weight; control plants were frequently watered, so they were not subjected to dehydration stress. The fresh leaf tissues were immediately frozen in liquid nitrogen and stored at −80°C for RNA separation.

For Illumina sequencing, total RNA was extracted according to Cetyltrimethyl Ammonium Bromide procedure.\textsuperscript{19} The RNA quality and quantity were determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Fragmentation buffer (Ambion, Austin, TX, USA) was added to produce short mRNA fragments. After fragmentation, cDNA was synthesized. The cDNA library was constructed based on poly-A–enriched RNA method, and 200–300 bp fragments were selected for non-directional, 100-nucleotide, paired-end sequencing according to Illumina protocols (San Diego, CA, USA). With random hexamer primers, the first strand was synthesized and then the second strand. The short fragments were purified with the QIAquick PCR Purification kit (Qiagen, Valencia, CA, USA) for both end repair and poly-A addition reaction. Finally, the purified DNA libraries were amplified by polymerase chain reaction and sequenced by Illumina HiSeq\textsuperscript{20} 2500 (Encode Genomics Bio-Technology Company, China).

De novo assembly and clustering. In-house perl scripts were written for quality control (script is available upon request). Our filtering and quality control processes include four steps: (1) removing reads with adaptor, (2) removing reads that contained uncertain bases (N) more than 10%, (3) removing low-quality reads, and (4) removing duplicated reads. To obtain maximum differentially expressed genes, we pooled all filtered reads for both control and drought treatment samples for generating the reference sequences. *De novo* transcriptome assembly was performed using the Trinity program,\textsuperscript{20} proceeding with three independent modules: Inchworm, Chrysalis, and Butterfly. Trinity generated a reference genome with abundant similar transcripts and alternatively spliced isoforms. We then used Cluster Database at High Identity with Tolerance (CD-HIT) program\textsuperscript{21} (http://www.bioinformatics.org/cd-hit/) to cluster the similar contigs/isoforms to obtain a high-quality reference genome with nonredundant unigenes.

Homology search and functional annotation. All unigenes were annotated using BLASTx by searching against three commonly used databases: the National centre for biotechnology information (NCBI) nonredundant (NR) database, the Swiss-Prot database, and the KOG (Eukaryotic ortholog groups) database. The cut-off $E$-values were set as $10^{-5}$. Gene ontology (GO) annotation was carried out using Blast2go program,\textsuperscript{22} and visualized on the WEGO (Web Gene Ontology Annotation Plot)\textsuperscript{23} Web site (http://wego.genomics.org.cn/cgi-bin/wego/index.pl).

Differential gene expression analysis. All raw reads were mapped to *de novo* assembled reference sequence using BWA software\textsuperscript{24} and then proceeded by RSEM (RNA-Seq by Expectation-Maximization) methods\textsuperscript{25} to normalize the raw read counts. The RPKM value (reads per kilobase per million reads), Fold Change value, and GFOld value for each transcript were measured according to GFOld algorithm,\textsuperscript{26} which are still robust when there is no available replicates. In general, a gene with zero GFOld value should never be considered as differentially expressed while upregulated genes usually have positive GFOld values, and negative values represent downregulated genes. On this basis, we use GFOld value as an initial step to exclude the genes that are not differentially expressed under the two conditions. Then, in order to reduce the likelihood of false positives, differential expressed genes (DEGs) were identified with Fold Change $\geq 1$ between
two conditions and false discovery rate (FDR) ≤0.001. GO enrichment analysis and the functional associated pathways for DEGs were further analyzed by WEGO and KEGG (the Kyoto Encyclopedia of Genes and Genomes).27 KAAS (KEGG Automatic Annotation Server) was used to characterize associated pathways. To find the most highly significant differential expression genes, the criterion of the absolute value of RPKM ratio >1,000 was used. Blast searches were conducted against NCBI and The Arabidopsis Information Resource (TAIR) to seek the homologs with *Gossypium hirsutum* and *Arabidopsis thaliana*.

**Single-sequence repeat mining and single-nucleotide polymorphism identification.** We predicted microsatellites using MISA (MlcroSATellite identification tool; available at http://pgrc.ipk-gatersleben.de/misa/). The criteria used for the single-sequence repeat (SSR) detection is unigenes containing 2–6 bp motifs and the minimum number of repeats are as follows: seven for dinucleotide motifs and five for trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide motifs.

We used the SOAP Program28 to align the raw reads with the assembled reference sequence and used the SOAPsnp program to identify potential single-nucleotide polymorphisms (SNPs). To reduce the possibility of false-positive hits, we filtered out the reads with quality values <20 or read depth below 10×.

**Results and Discussion**

**Reads generation and de novo assembly.** With Illumina HiSeq™ 2500 sequencing technology, we obtained a total of 136,000,000 short reads. Most reads have quality scores >99%, and the quality distribution is shown in Supplementary Figure S1. The filtering steps are shown in Supplementary Figure S2. After trimming, 56,782,314 and 46,561,748 reads, covering 78.86% and 72.75% for drought-treated and control samples, respectively, were retained. The clean reads were assembled using the Trinity program to obtain the de novo reference sequence. We then used the CD-HIT program to cluster redundant and similar isoforms. Finally, we obtained a high-quality reference sequence with a total of 160,601 contigs and an average length of 1,093 bp. The combined length of the reference genome reached 175,523,499 bp (Table 1). The N50 for reference genome is 418 bp, and the length of contigs ranged from 200 to 14,533 bp, of which 67.03% fall into the 200–600 bp range (Fig. 1A). The GC content is about 39.51%. The nonredundant contigs were defined as unigenes and subjected to further analysis.

**Characterization of unigenes.** All unigenes were annotated against the NCBI NR database, Eukaryotic ortholog groups (KOG) database, and Swiss-Prot protein database. With a cutoff of *E*-value of $10^{-5}$, 95,586 (59.52%) unigenes had hits in the three public databases (Fig. 1B). In particular, we obtained 94,904 hits (59.09%) from the NR database, of which 64,655 (40.26%) hits had *E*-value ≤1e-30 indicating a high similarity. We also searched all the unigenes against KOG database and identified 24,774 (34.7%) homolog genes. Those genes were classified into 25 functional categories (Supplementary Fig. S3). The largest category was “General function prediction only” (8,957 of 24,774 unigenes, about 15.97%), followed by “Posttranslational modification, protein turnover, chaperones” (2,042, 8.24%), “Translation, ribosomal structure and biogenesis” (1765, 7.1%), “Carbohydrate transport and metabolism” (1388, 5.60%), “Amino acid transport and metabolism” (1025, 4.14%), and the “Cell motility”, which accounted for 0.1% unigenes. Compared with the NCBI NR and KOG database, fewer hits (65,288) were found in Swiss-Prot database, and the majority of hits (63.83%) had very low *E*-values (1e-150 to 1e-5). In total, 95,586 (59.52%) unigenes were annotated in the three main public databases.

**Functional classification.** All unigenes were assigned with the GO classification by Blast2go, and then visualized using WEGO. A total of 9,349 unigenes were categorized into 38 functional subgroups (Supplementary Fig. S4), including Molecular Function (86,763), Biological Process (206,944), and Cellular Components (94,459). In general, the groups of “cell” (87.2%) and “cell part” (87.2%) were the most common. For the Cellular Components category, the cell and cell part were the most highly represented categories. For the Molecular Function category, genes associated with binding (45.5%) and catalytic proteins (42.3%) have the highest fractions. Under the Biological Process category, proteins related to cellular processes (66.4%) and metabolic processes (57.8%) were most frequent.

**Identification of differentially expressed genes.** We computed the RPKM values for each unigene using the GFOOLD algorithm and detected 20,666 differentially expressed unigenes between control and drought conditions after excluding the unigenes with zero GFOOLD values (Supplementary Table S1). Then, DEGs were identified with Fold Change ≥1 between two conditions and FDR ≤0.001. Altogether 1416 DEGs unigenes were identified, with 871 of them upregulated and 545 downregulated. GO enrichment was used to identify the major function groups of DEGs affected by drought stress, and 38 overrepresented groups in the GO classification were found (Fig. 2). Among

**Table 1. Characteristics of clustered contigs.**

| ITEM                               | NUMBER     |
|------------------------------------|------------|
| Number of contigs after clustering | 160,601    |
| Total length of contigs (bp)       | 175,523,499|
| Average length (bp)                | 1,093      |
| Length of minimum contigs (bp)     | 200        |
| Length of maximum contigs (bp)     | 14,533     |
| N50 (bp)                           | 1987       |
| N90 (bp)                           | 418        |
| GC (%)                             | 40         |

Complicated transcriptomic responses to drought stress of *Bombax ceiba* L.
those groups, we are particularly interested in the “response to stimulus” (40.3%) category under “biological process”, because the most drought-related genes may be contained in this group. Genes associated with stimulus comprised the “response to stress” and “response to abiotic stimulus” with fractions of 28.9% and 22.0%, respectively. Interestingly, we also found genes associated with “response to water deprivation” that were expressed differentially between the control and drought conditions (Fig. 3A). Those genes contained the putative conserved domains of DNA-binding domains, Annexin, NAD(P)+-dependent aldehyde dehydrogenase, lysozyme-like domain and WAX2 C-terminal.

For the KEGG analysis, 462 differentially expressed genes obtained hits in KEGG database, and those genes were associated with 140 enzyme classes and 341 KEGG pathways. These enzymes were further categorized into oxidoreductases,

![Figure 1](image1.png)

**Figure 1.** Length distributions of all-unigenes (A) and the characterization of unigenes against known public databases (B). The majority of contig length fall in the range of 201–600 bp. Altogether 95,586 unigenes (59.52%) had hits in the three public databases.

![Figure 2](image2.png)

**Figure 2.** GO classifications of differentially expressed unigenes. The differential unigenes are summarized into three main categories: biological process, cellular components, and molecular function. In total, 614 unigenes were assigned to gene ontologies.
transfères, hydrolases, lyases, isomerases, and ligases, of which oxidoreductases (28.57%) and transferases (28.05%) were found to dominate (Supplementary Fig. S5). In addition, the top 12 KEGG pathways were shown in Figure 3B. Of the top 12 KEGG pathways, genes associated with metabolic pathways (33.24%) were most representative, followed by biosynthesis of secondary metabolites (19.2%). Those pathways include plant hormone signal transduction (ko04075), glycolysis/gluconeogenesis (ko00010), glyoxylate and dicarboxylate metabolism (ko00630), photosynthesis (ko00195), oxidative phosphorylation (ko00190), and DNA replication (ko03030).

Using the criterion of the absolute value of RPKM ratio >1,000, 59 differentially expressed unigenes were retained. Among those genes, 54 showed overexpression and 5 decreased expression (Fig. 4A). Blast searches were conducted against NCBI and TAIR (Supplementary Table S2, Fig. 4B). We obtained 15 homologs from G. hirsutum and 55 homologs from A. thaliana. Several processes, including the metabolisms of nucleic acid, lipid metabolism, carbohydrate, and heat

![Figure 3](image-url)  
*Figure 3.* The heatmap of genes in related to response to water deprivation (A) and the top 12 pathways assignment based on KEGG (B).
resistance, were significantly changed, indicating the importance and requirement of these processes in plant growth and survival under drought condition. Unigenes associated with the processes involved should be paid attention and followed up in future.

**Regulation of plant hormone pathways genes.** Plant hormones are known to play a vital role for plant growth and development including drought resistance. They are able to regulate many important processes, such as cell enlargement, shoot initiation, stem growth, stomata closure, and cell division (Fig. 5). Auxin, a class of plant growth hormones, promotes plant growth against a wide range of stresses including drought.\(^{29,30}\) In *A. thaliana*, Shi et al.\(^{31}\) found that plants with high IAA (indoleacetic acid) level exhibited significantly lower water loss rate than IAA-mutant plants. In this study, pathway enrichment analysis suggested that a number of the modulated unigenes contribute to plant growth and development, ie, auxin signal pathways. We found five unigenes, encoding the SAUR family protein (SAUR), auxin influx carrier protein (AUX1), and auxin-responsive protein (IAA), which were expressed differentially under drought treatment. Two unigenes encoding AUX1 and IAA were found to be upregulated under drought condition. Three unigenes encoding SAUR showed different expression patterns; two of them were upregulated and one of them was downregulated. Our results suggest that several overexpressed genes are positively regulated, while other genes such as SAUR act as negative regulators to adjust auxin synthesis,\(^{32}\) indicating that auxins may play an important role in response to drought stress in *B. ceiba*.

We also identified differentially expressed genes involved in the brassinosteroid and jasmonic acid signal transduction pathways. Some of them are upregulated under drought condition, including CYCD3 and two unigenes encoding jasmonate ZIM domain-containing protein (JAZ), suggesting the importance of brassinosteroid and jasmonic acid under stress.\(^{33,34}\) This assertion is consistent with previous finding. With EBR (one kind of brassinosteroid) treatment, *A. thaliana* can reduce leaf wilting with high survival rate under drought.\(^{35}\)

Notably, plants have different modulation of hormones in response for plant abiotic stress. Gibberellin, cytokinin, ethylene, and salicylic acid are involved in regulating several aspects of plant growth and development.\(^{36-38}\) In the gibberellin signal transduction pathway, the GID1 and PIF4 gene, encoding the gibberellin receptor and phytochrome-interacting factor 4 respectively, decreased expression and negatively regulated stem growth. Similarly, in the cytokinin, ethylene and salicylic acid signal transduction pathway, unigenes encoding histidine-containing phosphotransfer protein (AHP), regulatory protein NPR1 (NPR1), ethylene-insensitive protein 3 (EIN3), and ethylene-responsive transcription factor I (ERF1) showed decreased expression. Differential expression profiles indicated that gibberellin, cytokinin, ethylene, and salicylic acid catabolism and their repressive signaling are involved in drought condition.\(^{39-46}\)

ABA, a plant growth regulator, is involved in the regulation of the final stage of plant development and accelerates the yellowing of both attached and detached leaves.\(^{47}\) Its function can be enhanced by drought and could reduce stomatal closure.\(^{48}\) In the early phase of drought, stomata close in response to water deficiency. But as drought conditions persist, the stomata reopens passively.\(^{49}\) In this study, we found that ABA receptor family PYL is upregulated. Phosphatase 2C (PP2C), serine/threonine-protein kinase SRK2 (SNRK2), and ABA-responsive element binding factors (ABF) are downregulated. We hypothesize that those genes may control the level of stomatal closure to ensure air exchange with external environment.\(^{50}\)

In summary, we found a number of genes associated with various hormone pathways, including AUX1, SAUR, GID1, PIF4, ERF1, PYL, SNRK2, ABA, etc (Supplementary Table S3), indicating that complicated regulation and interaction among those genes contributes to drought tolerance in *B. ceiba*. Plant hormones enhance resistance against water stress by regulating physiological processes. Auxin may participate in the positive regulation of drought stress resistance, through regulation of root architecture, ABA-responsive gene expression, Reactive oxygen species metabolism, and metabolic homeostasis.\(^{31}\) Brassinosteroid not only increases the maximum quantum yield of photosystem II, the water potential, and the content of soluble sugars and proline but also decreases the malondialdehyde content and electrical conductivity of leaves under drought stress.\(^{31}\) ABA, coordinating with other signals such as ethylene, regulates stomatal closure in response to drought.\(^{3,50}\)

**Metabolism associated drought resistance.** Under drought conditions, plant cells could accumulate a variety of small organic metabolites including sugars, amino acids, and related compounds.\(^{51,52}\) Soluble sugars are known to regulate osmotic pressure of plants, such as starch and sucrose.\(^{53,54}\) Drought exposure induces the hydrolyzation of polysaccharides and accumulation of soluble sugars and causes a decrease in leaf osmotic potential.\(^{4,55}\) We identified six genes associated with metabolism of starch and sucrose (Supplementary Fig. S6) in this study. Regarding sucrose degradation, two genes, encoding the β-fructofuranosidase sacA and catalyzing the conversions of sucrose→6P into α-d-Glucose→6P, increased their expression and facilitated the conversions of sucrose into β-d-Fructose. Regarding starch degradation, one unigene encoded beta-glucosidase bgX, which converts cellulose into cellobiose, and another gene encoded endoglucanase, may be involved in the conversion of cellobiose into β-d-Glucose. Both of them are upregulated at the transcriptome level. In addition, genes which encoded sucrose synthase could also promote the conversion of Uridine Diphosphate-glucose into sucrose. Those results suggest that polysaccharide and disaccharide metabolism tend toward degradative processes under
Complicated transcriptomic responses to drought stress of *Bombax ceiba* L.

**Figure 5.** The pathway map of plant hormone signal transduction. The identified genes involved in plant hormone signal transduction are marked with green. Reproduced with permission from Kanehisa Laboratories.\(^{(56,57)}\)
drought conditions to increase substance solubility and weaken osmotic pressure.

Proline can be functional as a mediator of osmotic adjustment, a stabilizer of macromolecules, a compatible solute to protect enzymes, and as a store for carbon and nitrogen for use during water-deficit and other stress regimes. In *A. thaliana*, proline concentrations slightly increased under mild drought stress and reached a remarkable level under strong stress. In *B. ceiba*, the genes associated with proline synthesis are downregulated during drought treatment. Two genes encoding proline iminopeptidase (pip) and betaine-aldehyde dehydrogenase (betB, gbsA) (Supplementary Fig. S7) in the proline pathway demonstrated decreased expression. This suggests that the drought tolerance mechanism is complicated and species-specific in *B. ceiba*.

**The oxidative phosphorylation and photosynthesis.** Detoxification pathways could also enhance plant drought tolerance. Variation of drought tolerance under similar conditions has been reported in different plants and cultivars of crops, emphasizing the significance in drought adaptation. We found 20 genes associated with oxidative phosphorylation and photosynthesis that showed obvious changes in expression level (Fig. 6 and Supplementary Fig. S8). Seven genes associated with oxidative phosphorylation increased expression under drought treatment. Eleven photosynthesis genes demonstrated increase in expression, and two genes, encoding photosystem II oxygen-evolving enhancer protein 2 (psbP) and photosystem II PsbW protein (psbW), decreased expression. The predicted explanation was that unless simultaneous and proportionate reductions in growth and carbon consumption take place, a negative carbon balance can occur as a result of diminished photosynthetic capacity during drought. Similar to *B. ceiba*, RNA-seq study of *P. euphratica* revealed that increasing expression of genes in photosynthesis and a high photosynthetic rate occurred under drought stress. On the other hand, foliar photosynthetic rate in some plants decrease with lower relative water content.

Under drought conditions, the fraction of carbohydrate lost through respiration determines the overall metabolic efficiency of the plant. Photosynthesis fuels plant metabolism by converting the light energy into carbohydrate molecules. We hypothesized that the increased respiration and photosynthesis provide sufficient energy for growth during drought adaption and tolerance. When lacking water, *B. ceiba* actively reprograms gene expression of metabolism.

---

**Figure 6.** The pathway map of oxidative phosphorylation. The identified genes involved in oxidative phosphorylation are marked with green. Reproduced with permission from Kanehisa Laboratories.
and growth pathways. Regulatory changes of hormone like ABA, wax composition, amino acids like proline, oxidative phosphorylation, and photosynthesis should prevent the plant cell damage from drought stress and enable the maintenance of normal function, an adaptation, which is similar to that seen in *A. thaliana*. *B. ceiba* has evolved specific mechanisms in response to drought stress by adapting to drought conditions over a long period. Broadly, drought resistance mechanisms include avoidance and tolerance to prevent drought through water conservation while preserving function under extremes of both internal and external stress. ABA and cytokinin as well as soluble sugars affect photosynthetic efficiency of plants. Initially, the concentration of ABA increased, driving stomata closure and permitting maintenance of the wax composition. To reduce the energy consumption, synthesis of proline was increased to accumulate solution and oxidative phosphorylation and photosynthesis reduced. Following appropriate adaption to the environment, ABA decreased to reopen stoma, the wax composition changed to have a higher average chain length, the amount of proline declined, and the activity of oxidative phosphorylation and photosynthesis recovered. All these adaptations enable the plant to maintain its normal functions and gradually adapt to the arid environment. Furthermore, drought has been shown to affect many physiological functions in plants. Taken together, drought resistance in *B. ceiba* is the result of many interacting genes and pathways, tying together multiple physiological processes. As a drought-endurable plant, an efficient response to the environment is particularly necessary. It is crucial for *B. ceiba* that coordinates and executes stress responses in terms of metabolic and developmental adjustments.

**SSR mining and SNP detecting.** We used the MISA program to identify possible microsatellite (SSR) markers for *B. ceiba*. We found 16,656 microsatellites from 15,589 unigenes (Supplementary Table S4), of which 1,067 unigenes contained more than one SSR. According to the repetitive motifs, all microsatellites were further grouped into dinucleotide motifs (7,998), trinucleotide motifs (7,624), tetranucleotide motifs (657), pentanucleotide motifs (61), and hexanucleotide motifs (45) and present in compound formation (271) (Table 2). Among all microsatellites, the most common one was dinucleotide motifs (48%) and the most abundant repeat type was (AT/TA), followed by (TA/AT), (AG/TC), (GA/CT), (TC/AG), and (GAA/CTT).

We also identified putative SNPs by aligning raw reads against the de novo–assembled reference genome. In total, 3,071 high-confidence SNPs were detected, of which 377 and 558 unigenes had more than two SNPs for control and drought conditions, respectively. On average, one SNP exists per 57 bp. The putative SNPs comprised transitions and transversions, two classes and six variations in detail (Supplementary Table S5). The transitions between C and T were most frequent than any other variation. In addition, the frequency of transitions was much higher than transversions in both samples (Table 3).

**Conclusions**

To the best of our knowledge, this study represents the first attempt to identify possible genes associated with drought using transcriptome sequencing of *B. ceiba*. We obtained a total of 136,000,000-bp de novo reference transcriptome. Based on this reference, we identified 20,666 DEGs genes by comparing the transcriptional profiles between control and drought treatment. Moreover, 16,656 microsatellite markers and 3,071 SNPs were also identified. This data set and its corresponding annotations allow us to explore the

---

**Table 2. Single sequence repeats (SSRs) identification.**

| UNIT SIZE        | NUMBER OF SSRs | MOST DISTRIBUTED IN EACH UNIT | NUMBERS |
|------------------|----------------|-------------------------------|---------|
| Dinucleotide     | 7,998          | (AT/TA)                       | 2079 (26%) |
|                  |                | (TA/AT)                       | 1639 (20%) |
| Trinucleotide    | 7,624          | (GAA/CTT)                     | 633 (8%)  |
|                  |                | (TTC/AAG)                     | 479 (6%)  |
| Tetranucleotide  | 657            | (AAAT/TTTA)                   | 65 (10%)  |
|                  |                | (TTTA/AAAT)                   | 43 (7%)   |
| Pentanucleotide  | 61             | (ATTTG/TAACC)                 | 9 (15%)   |
|                  |                | (CCCAA/GGGT)                  | 8 (13%)   |
| Hexanucleotide   | 45             | (GTAGT/CAATCA)                | 6 (13%)   |
|                  |                | (GGGAGG/GGCTCC)               | 5 (11%)   |
| Number of SSRs present in compound formation | 271 |
| Number of sequences containing more than one SSR | 1,067 |
| Total number of identified SSRs | 16,656 |
| Total number of sequences examined | 15,589 |

*Note:* Details of SSRs found in *B. ceiba.*
molecular mechanism of drought response for B. ceiba. Those DEGs genes associated with various pathways, including the hormone metabolisms and photosynthesis, indicate that drought response in B. ceiba is the consequence of interactions between multiple genes and pathways at multiple physiologic levels. Our results provide valuable resources for understanding drought mechanism in B. ceiba.

Acknowledgments
We thank Zhao Gaojuan, Chen Jiafu, and Tang Junrong for partial data processing and field sampling.

Author Contributions
Conceived and designed the experiments: BT, HM. Analyzed the data: ZZ, YZ, YC. Contributed to the writing of the manuscript: ZZ, KL, ZX, BT, LW. Agreed with manuscript results and conclusions: BT. All authors reviewed and approved the final manuscript.

Supplementary Materials
Supplementary Table S1. Gene expression level measured by GFOLD.
Supplementary Table S2. Information of 59 significant differential expression genes.
Supplementary Table S3. Regulation of plant hormone pathways genes.
Supplementary Table S4. SSRs results.
Supplementary Table S5. SNP results.
Supplementary Figure S1. The quality distribution of two samples.
Supplementary Figure S2. Reads of filtering steps. After trimming and quality control, 56,782,314 and 46,561,748 reads were remained for each sample.
Supplementary Figure S3. Eukaryotic ortholog group (KOG) classification. In total, 24,774 sequences were grouped into 24 COG classifications.

Supplementary Figure S4. GO classifications of assembled unigenes. The unigenes are summarized into three main categories: biological process, cellular location, and molecular function. In total, 9,349 unigenes were assigned with GO number.

Supplementary Figure S5. Enzyme assignment based on KEGG pathways. Oxidoreductases and transferases were dominated.

Supplementary Figure S6. The pathway map of starch and sucrose metabolism. The identified genes involved in starch and sucrose metabolism are marked with green.56,57

Supplementary Figure S7. The pathway map of proline metabolism. The identified genes involved in proline metabolism are marked with green.56,57

Supplementary Figure S8. The pathway map of photosynthesis. The identified genes involved in photosynthesis are marked with green.56,57

REFERENCES
1. Wang W, Vinocur B, Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Plant. 2003;218:1–14.
2. Egert M, Tewini M. Influence of drought on some physiological parameters of Allium schoenoprasum. Environ Exp Bot. 2002;48:43–9.
3. Schulze ED. Whole-plant responses to drought. Funct Plant Biol. 1986;13:127–41.
4. Chaves MM, Maroco JP, Pereira J. Understanding plant responses to drought: from genes to the whole plant. Funct Plant Biol. 2003;30:239–64.
5. Pei ZM, Murata Y, Benning G, et al. Calcium channels activated by hydrogen peroxide mediate abscisic acid signaling in guard cells. Nature. 2000;406:731–4.
6. Shinorizki K, Yamaguchi-Shinozaki K. Gene expression and signal transduction in water-stress response. Plant Physiol. 1997;115:327–34.
7. Hu HH, Xiong LZ. Genetic engineering and breeding of drought-resistant crops. Annu Rev Plant Biol. 2014;65:715–41.
8. Zhang JA, Liang S, Duan JL, et al. De novo assembly and characterization of the transcriptome during seed development and generation of genic-SSR markers in peanut (Arachis hypogaea L.). BMC Genomics. 2012;13:90–5.
9. Iorizzo M, Senalik DA, Grzebelus D, et al. De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. BMC Genomics. 2011;12:389–402.
10. Fu N, Wang Q, Shen HL. De novo assembly, gene annotation and marker development using Illumina paired-end transcriptome sequences in celery (Apium graveolens L.). PLoS One. 2013;8:e57686–97.
11. Liu SQ, Li WS, Wu YM, Chen CM, Lei JJ. De novo transcriptome assembly in chili pepper (Capsicum frutescens) to identify genes involved in the biosynthesis of capsaicinoids. PLoS One. 2013;8:e48156–63.
12. Huang HH, Xu LL, Tong ZK, et al. De novo characterization of the Chinese fir (Cunninghamia lanceolata) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. BMC Genomics. 2012;13:648–61.
13. Shi Y, Yan X, Zhao FS, et al. Transcriptomic analysis of a tertiary relict plant, extreme xerophytes Richea texanae to identify genes related to drought adaptation. PLoS One. 2013;8:e63993–4003.
14. Peng SM, Jiang H, Zhang S, et al. Transcriptional profiling reveals sexual differences of the leaf transcriptomes in response to drought stress in Populus simonii. Tree Physiol. 2012;32:1541–55.
15. Bhardwaj J, Chauhan R, Swarnkar MK, et al. Comprehensive transcriptomic study on horse gram (Macrotyloma uniflorum): de novo assembly, functional characterization and comparative analysis in relation to drought stress. BMC Genomics. 2013;14:646–73.
16. Baum DA, Ogimuna K. A review of chromosome numbers in Bombacaceae with new counts for Adenanthera. Taxon. 1994:43:11–20.
17. Wang GK, Lin BB, Ruo R, et al. A new lignan with anti-HBV activity from the roots of Bombax ceiba. Nat Prod Res. 2013;27:1348–52.
18. Patel SS, Patel NK, Rathore B, Nayak G, Singhai AK, Singh P. Cardioprotective effect of Bombax ceiba flowers against acute adriamycin-induced myocardial infarction in rats. Revista Brasileira de Farmacognosia Brazilian J. Pharmacogn. 2011;21:704–9.
34. Olszewski N, Sun TP, Gubler F. Gibberellin signaling: biosynthesis, catabolism, and physiological response to drought and short-day photoperiods in Arabidopsis thaliana. Plant Cell. 2014;164:1151–60.

35. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

36. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

37. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

38. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

39. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

40. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

41. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

42. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

43. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

44. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

45. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

46. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

47. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

48. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

49. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

50. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

51. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

52. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

53. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.

54. Olszewski N, Sun TP, Gubler F. Gibberellin metabolic genes in soybean under normal and drought conditions in response to 19A (a myb protein) expression. Mol Plant. 2014;7:942–55.

55. Dejardin A, Sokolov L, Kleczkowski L. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis thaliana. J Exp Bot. 2012;63:2185–97.