Plant Omics Journal POJ

POJ 7(3):178-187 (2014) ISSN:1836-3644

Deep sequencing-based transcriptome analysis of the oil-bearing plant Physic Nut (*Jatropha curcas* L.) under cold stress

Haibo Wang[1,2], Zhurong Zou[1], Shasha Wang[1], Ming Gong[1,*]

[1]School of Life Sciences, Engineering Research Center of Sustainable Development and Utilization of Biomass Energy, Ministry of Education, Key Laboratory of Biomass Energy and Environmental Biotechnology of Yunnan Province, Yunnan Normal University, Kunming 650500, Yunnan, P. R. China
[2]College of Biological Resources and Environmental Science, Qujing Normal University, Qujing 655011, Yunnan, P. R. China

***Correspondence author: gongming6307@163.com**

Abstract

Nowadays *Jatropha curcas* L. has gained an increased attention in scientific and commercial fields as an important renewable bioenergy plant, aiming to prevent the possible energy crisis of fossil fuels. However, the studies on cold resistance of this biofuel shrub are still scarce, giving limited information for its genetic improvement and other biotechnological applications. In this work, the newly developed Illumina Hiseq™ 2000 RNA-seq, which is a deep high-throughput sequencing approach, preferentially was used for cold-resistance related transcriptome analysis of *J. curcas*. From the sequencing results, the total length of non-redundant sequences obtained was 4,960,092,780 bp, consisting of 106,749 contigs and 45,251 unigenes assembled by clean data. A total of 35,791 unigenes (79.09%) can be annotated to numerous databases (Nr, Swiss-Prot, GO, COG, KEGG) for functional classification. The 33,361 and 912 complete or partial CDSs are deduced by database alignment and ESTscan prediction, respectively. Among these unigenes, 27,293 can be categorized into 61 functional groups of GO, 11,887 of COG-annotated putative proteins were classified functionally into at least 25 molecular families, 18,787 were possibly involved in approximately 128 known metabolic or signaling pathways in KEGG. This study provided a comprehensive cold-resistance transcriptome analysis of *J. curcas* with remarkably more number of EST sequences than all previous relevant deposits in public databases. The results allowed us to decipher the key genes related to those coding for cold tolerance in this sequence library.

Keywords: *Jatropha curcas* L.; Transcriptome; High-throughput sequencing; Cold resistance.
Abbreviations: BLAST_Basic Local Alignment Search Tool; GO_Gene Ontology; COG_Cluster of Orthologous Groups of proteins; KEGG_Kyoto Encyclopedia of Gene and Genomes; CDS_Coding Region Sequence; SFR6_ Sensitive to Freezing 6; LOV1_Long Vegetative Phase 1; HOS9_High expression of Osmotic Stress regulated gene 9; SCOF1_Soybean Cold-inducible zinc Finger protein 1.

Introduction

Currently, inevitable increase in energy consumption consequently leads to continuous decrease in fossil fuel reserves and serious problem of global environmental worsening. To reconcile this conflict, more attention has been shifted to renewable energy resources such as biofuels obtained from either carbohydrate- or oil- based feedstock and biomass.

Among so many non-edible oleaginous plants, *Jatropha curcas* (*J.curcas*) is a multipurpose shrub belonging to the family of Euphorbiaceae. It is capable of growing in arid or semi-arid marginal lands with minimum cultivation inputs and also does not compete with food supply (Yang et al., 2012). This plant additionally confers considerable resistance to drought, growing well at rainfall levels as low as 200 mm per annum (Achten et al., 2010; Maes et al., 2009). Undoubtedly, such peculiarities make *J. curcas* endued with great significance in respects of agriculture, environment and forest. Most importantly, the eminent seed traits of high oil content and oil quality suitable for biodiesel production manifests this plant as a well-known bioenergy resource (Johnson et al., 2011).

In the past few years, a number of studies about *J. curcas* have been published with respect to regeneration (Deore and Johnson 2008; Khurana-Kaul et al., 2010; Kumar and Reddy, 2010; Kumar et al., 2010), genetic transformation (Li et al., 2008; Pan et al., 2010), genomics (Sato et al., 2011; Asif et al., 2010), transcriptomics (Costa et al., 2010; Natarajan and Parani, 2011), proteomics (Yang et al., 2009), and gene cloning particularly relevant to abiotic stress resistance (Tang et al., 2007; Wang et al., 2011; Zhang et al., 2007; Zhang et al., 2008). Nevertheless, the derived information is sporadic or limited, and incapable of providing plentiful knowledge, especially devoid of a global transcriptome profile, for genetic improvements on the important agronomic and economic traits such as the inherent cold-susceptibility of this energy plant.

In this study, a powerful deep high-throughput sequencing approach, Illumina Hiseq™ 2000 RNA-seq (Wang et al., 2009), was exploited for transcriptome sequencing analysis of *J. curcas* from a mixed sampling of the untreated and cold-treated seedlings. To this sequenced transcriptome of *J. curcas*, we obtained 45,251 assembled unigenes and completed their functional annotations and classifications into the reference sequences of different databases, using bioinformatics analyses. Among them, unigenes with relatively highest expression were revealed, and those correlated to the cold-tolerance were selectively elucidated. In all, the abundant sequences obtained

in this work represent a most global transcriptome of *J. curcas* till now, with a considerable increase in the total number of ESTs deposited in GenBank, and will be certainly beneficial to cloning the full genes of interest and further genetic improvements on this oil-bearing plant.

Results and Discussion

Statistical analysis of the sequenced transcriptome of J. curcas

Approximately 59,275,070 raw sequence reads were obtained for the transcriptome arising from a mixture sample of *J. curcas* seedlings under normal condition and three cold treatments (Table 1), using Illumina Hiseq™ 2000 RNA-seq approach. Prior to assembly of sequence reads for non-redundant unigenes, clean data was generated by discarding the dirty reads (described in Materials and Methods).

Trinity software was used to combine the clean reads with certain length of overlap into contigs. The reads were mapped back to contigs, in order to detect contigs for the same transcript and their interval distances. Trinity then connected these contigs, and got sequences unable to be extended on both ends. Such assembled sequences are defined as unigenes (Table 2). The total length of the final transcriptome sequences of *J. curcas* obtained was covered into 55,112,142 clean reads (4,960,092,780 bp total), consisting of 106,749 contigs (36,062,865 bpt total) and 45,251 unigenes (35,713,278 bp total). The average G+C content of the clean data was 42.16%, and the average length of contigs and unigenes was 338 bp and 789 bp, respectively. The length distributions of contigs and unigenes were further calculated and demonstrated in Fig 1.

Annotation of non-redundant unigenes of J. curcas

All assembled 45,251 *J. curcas* unigenes were aligned to different databases such as Nr, Nt, Swiss-Prot, GO, COG, KEGG, Rc, Jc_EST, and Jc_CDS, with the annotated results shown in Table 3. Among them, 35,791 (79.09%) non-redundant unigenes have known protein annotations. The best alignment results were used to decide sequence direction from 5' end to 3' end of unigenes. If the aligned results from different databases conflicted with each other, a priority order of Nr, Swiss-Prot, KEGG, and COG was followed when deciding sequence direction of unigenes. When a unigene was aligned to none of the above databases, a program named ESTScan was used to decide its sequence direction.

By Blastx alignment (e-value<1e-5), 33,848 (74.80%) unigenes of *J. curcas* were matched against the non-redundant (Nr) protein database of GenBank, and 22,366 (66.10%) against a sole predicted proteome of *Ricinus communis* (Rc), which is another species within the same family of *Euphorbiaceae*. All unigenes of *J. curcas* were further evaluated by a most relevant alignment against its EST database (Jc_EST) at GenBank and predicted CDS sequences based on published draft genomic database (Jc_CDS, http://www.kazusa.or.jp/jatropha/) with an e-value cut-off of 1e-5. We found 20,485 (45.27%) and 36,321 (80.27 %) matches, respectively (Table 3). Additionally, the aligned results by Nr database showed that the total percentage of *J. curcas* unigenes with a similarity larger than 60% is 86.4%, in which the percentages with similarity range of 80-95% and 95-100% is 43.1% and 6.4%, respectively. The alignment of *J. curcas* unigenes against other individual plant species such as *Populus trichocarpa* was also carried out. Conclusively, the percentage maps of e-value distribution, similarity distribution and species distribution from such alignment against Nr database were showed in Fig. 2.

Furthermore, by Blastx and ESTscan analyses, 34,274 of total 45,251 *J. curcas* unigenes were speculated to have reliable coding sequences (CDSs) (33,362 unigenes by Blastx alignment and 912 by ESTscan prediction). Among CDS-containing *J. curcas* unigenes from database alignment, the length distribution ranged from 200 bp to 3,000 bp, and 198 of which are even of the length more than 3,000 bp. In contrast, those predicted by ESTScan have the length spectrum from 200 bp to 1,100 bp, with the main focus on a narrow range of 300~600 bp. Accordingly, the deduced protein sequences of unigenes analyzed by Blastx are of a length range from 200 aa to 2,100 aa, while those by ESTScan have a length distribution ranging from 200 aa to 700 aa (Fig. 3).

Functional classification of unigenes of J. curcas

According to the annotated information by different databases such as GO, COG, KEGG (Table 3), it is able to achieve a comprehensive functional classification for all 35,791 annotated *J. curcas* unigenes. GO annotation information of *J. curcas* unigenes was firstly obtained through Blast2GO program and then GO functional classification was completed through WEGO software as indicated in Fig. 4.

Based on sequence homology, 27,293 unigenes can be categorized into 61 functional groups within three GO ontologies (biological process, 29; cellular component, 16; and molecular function, 16) that each has a dominant percentage of unigenes involved in ‘metabolic process’(GO:0008152) and ‘cellular process’ (GO:0009987), ‘cell’ (GO:0005623) and ‘cell part’ (GO:0044464), and ‘binding’ (GO:0005488) and ‘catalytic activity’ (GO:0003824), respectively. Additionally, 11,887 unigenes encoding COG-annotated putative proteins were classified into at least 25 functional families as shown in Fig. 5, of which a remarkably high number participate the biological activities such as ‘General function’, ‘Transcription’, ‘Replication, recombination and repair’, ‘Signal transduction mechanisms’, and ‘Posttranslational modification, protein turnover, chaperones’. KEGG is a database that links genes to metabolism, signal transduction and other related cellular processes. To identify the biological pathways that are active in *J. curcas*, the 35,791 annotated sequences from 45,251 unigenes were evaluated using the reference canonical pathways in KEGG. In total, 18,787 unigenes can be grouped into 128 known metabolic or signaling pathways, including molecular metabolism, apoptosis, material migration and signaling transduction. The pathways with most number of engaged unigenes are ‘Metabolic pathways’ (3,930, 20.92%), ‘Biosynthesis of secondary metabolites’ (1,830, 9.74%) and ‘Plant-pathogen interaction’ (1,023, 5.45%). Among all metabolic pathways, the number of unigenes involved in ‘Starch and sucrose metabolism’ (430, 2.29%), ‘Glycerophospholipid metabolism’ (374, 1.99%), ‘Purine metabolism’ (334, 1.78%), ‘Pyrimidine metabolism’ (288, 1.53%), and ‘Ether lipid metabolism’ (235, 1.25%) are in the range of top 5.

Most highly expressed unigenes of J. curcas

For each unigene, the number of raw sequencing rea ds can be used to estimate its transcript abundance or expression level. Accordingly, 20 unigenes or the related contigs of *J. curcas*

Table 1. Brief statistics of sequencing data,

Total raw reads	Total clean reads	Total clean nucleotides (nt)	Q20 percentage	N percentage	GC percentage
59,275,070	55,112,142	4,960,092,780	98.04%	0.00%	42.16%

Q20: proportion of nucleotides with quality value larger than 20; N percentage: Unknown nucleotide percentage.

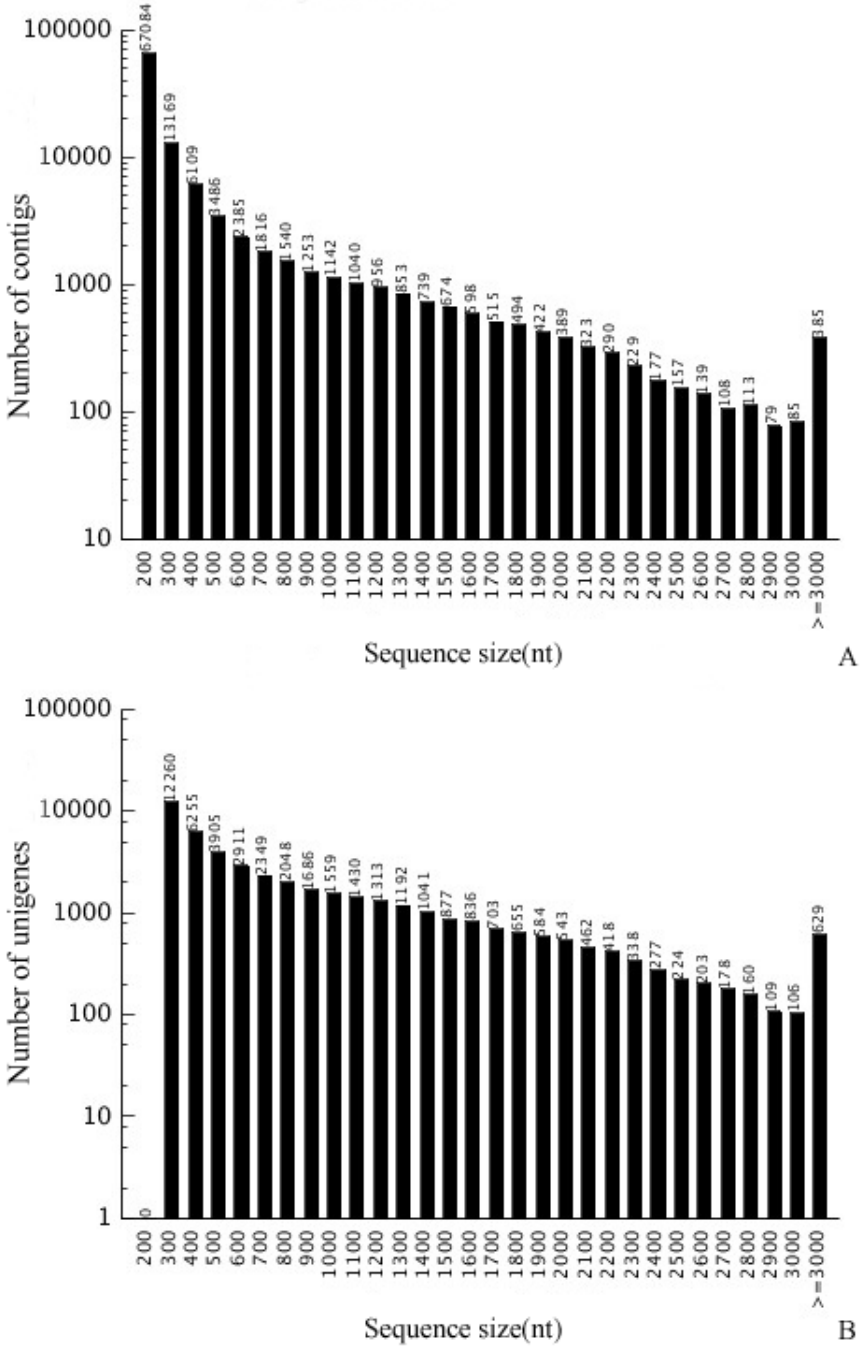

Fig 1. Sequence length distributions of assembled contigs (A) and unigenes (B).

Table 2. Brief statistics of sequence assembly.

	Total number	Total length (nt)	Mean length (nt)	N50
Contig	106,749	36,062,865	338	715
Unigene	45,251	35,713,278	789	1,225

N50: median length of all non-redundant sequences.

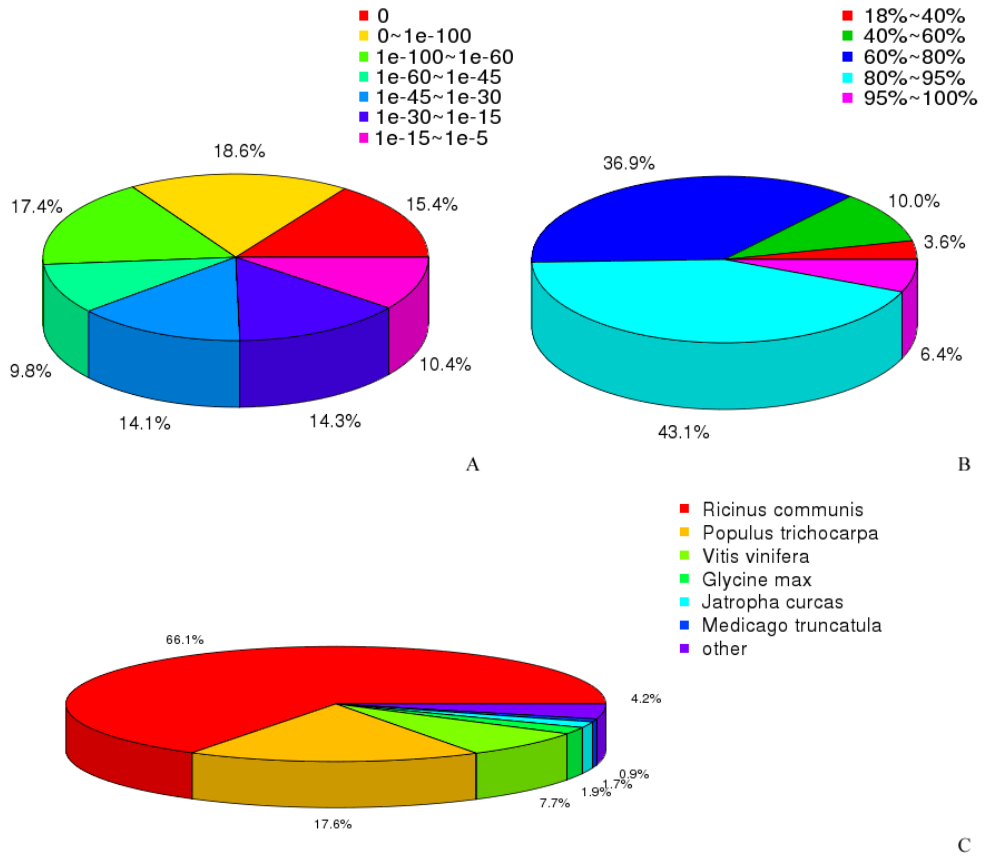

Fig 2. Percentage maps of e-value distribution (A), similarity distribution (B) and species distribution (C) from the alignments of *J. curcas* unigenes against Nr database.

with relatively highest expression were uncovered (Table 4). The most highly expressed unigene possibly encodes the cell wall-associated hydrolase (Unigene503_ JC-CK_1A), followed by ones encoding the ribulose 1,5-bisphosphate carboxylase small subunit (Unigene501_JC-CK_1A) and atpA (Unigene499_JC-CK_1A), in turn. Moreover, three of them can also be annotated into the known *J. curcas* ESTs (Unigene501 _JC-CK_1A, Unigene499_JC-CK_1A and CL2774.Contig1_ JC-CK_1A).

Among the top 20 members listed in Table 4, two different unigenes, encode for carbon assimilation-related enzymes, the small subunit and the activase of ribulose 1,5-bisphosphate carboxylase (Unigene501_ JC-CK_1A and Unigene539_ JC-CK_1A), two for chlorophyll a/b binding proteins (Unigene498_JC-CK_1A and Unigene508_JC-CK_1A), and three for the subunits of ATP synthase (Unigene499_ JC-CK_1A, Unigene500_JC-CK_1A and Unigene502_JC-CK_1A). All of them are known to be absolutely involved in plant photosynthesis. In additional, four different unigenes (Unigene503_JC-CK_1A, Unigene530_JC-CK_1A, Unigene 531_JC-CK_1A and Unigene521_JC-CK_1A), individually encoding the cell wall-associated hydrolase, plastidic aldolase, polygalacturonase and carbonic anhydrase, are known to be involved in the metabolisms of carbohydates and lipids. It is worth noting another abundantly expressed unigene (Unigene516_JC-CK_1A) encodes the non-specific lipid-transfer protein 3 which is through to be involved in oil transferring. This finding is consistent with the high level of oil in *J. curcas*.

Unigenes of J. curcas involved in cold tolerance

Under abiotic stress, plants generally undergo a series of physiological, biochemical and molecular changes for acclimation. For example, plants cold tolerance can be ascribed to several cellular behaviors such as changing the membrane fluidity, increasing the entocyte concentration to lower the osmotic potential, improving the activity of antioxidation and altering the expression of correlated genes (Ruelland et al., 2009). In order to decipher those additional genes induced by cold, a mixture sample for trancriptome analysis were taken from *J. curcas* seedlings under the normal condition as well as three chilling treatments at 12 ºC (see Materials and Methods).

Cold exposure usually induces notable changes in plant membrane composition and fluidity (Uemura and Steponkus, 1997; Ivanov et al., 2006), in which the content of polyunsaturated fatty acids is important for membrane integrity maintenance. The $\triangle^9$-stearoyl-ACP desaturase (SAD), catalyzing the first step reaction from saturated fatty acids to unsaturated fatty acids is often up-regulated for its expression at low temperature. Trienoic fatty acid is also a type of unsaturated fatty acid significant in *J. curcas* and might be functional in cold resistance. FAD3, an endoplasmic reticulum omega-3 desaturase, is the key enzyme to catalyze the formation of trienoic fatty acids. For these two desaturases, 1 unigene for SAD and 7 unigenes for FAD3 were found in our transcriptome sequencing library.

Table 3. Annotation analysis of non-redundant unigenes of *J. curcas.*

Database	Total assembled 45,251 unigenes	
	Number of annotated unigenes	Percentage of annotated unigenes
Nr	33,848	74.80%
Swiss-Prot	19,850	43.87%
GO	27,293	60.31%
COG	11,887	26.27%
KEGG	18,787	41.52%
Rc	22,366	66.10% *
Nt	33,697	74.47%
Jc_EST	20,485	45.27%
Jc_CDS	36,321	80.27%
ALL	35,791	79.09%

Rc: predicted proteome of *Ricinus communi*; *: Percentage of Rc matches versus the entirty annotation in Nr database; Jc_EST: EST sequences of *J.curcas* at NCBI (Release 130101, 46,944 items); Jc_CDS: predicted CDS sequences (57,437 items) based on published draft genomic sequences of *J.curcas* (http://www.kazusa.or.jp/jatropha/, Release 4.5, 39,277 items with 297,661,187bp).

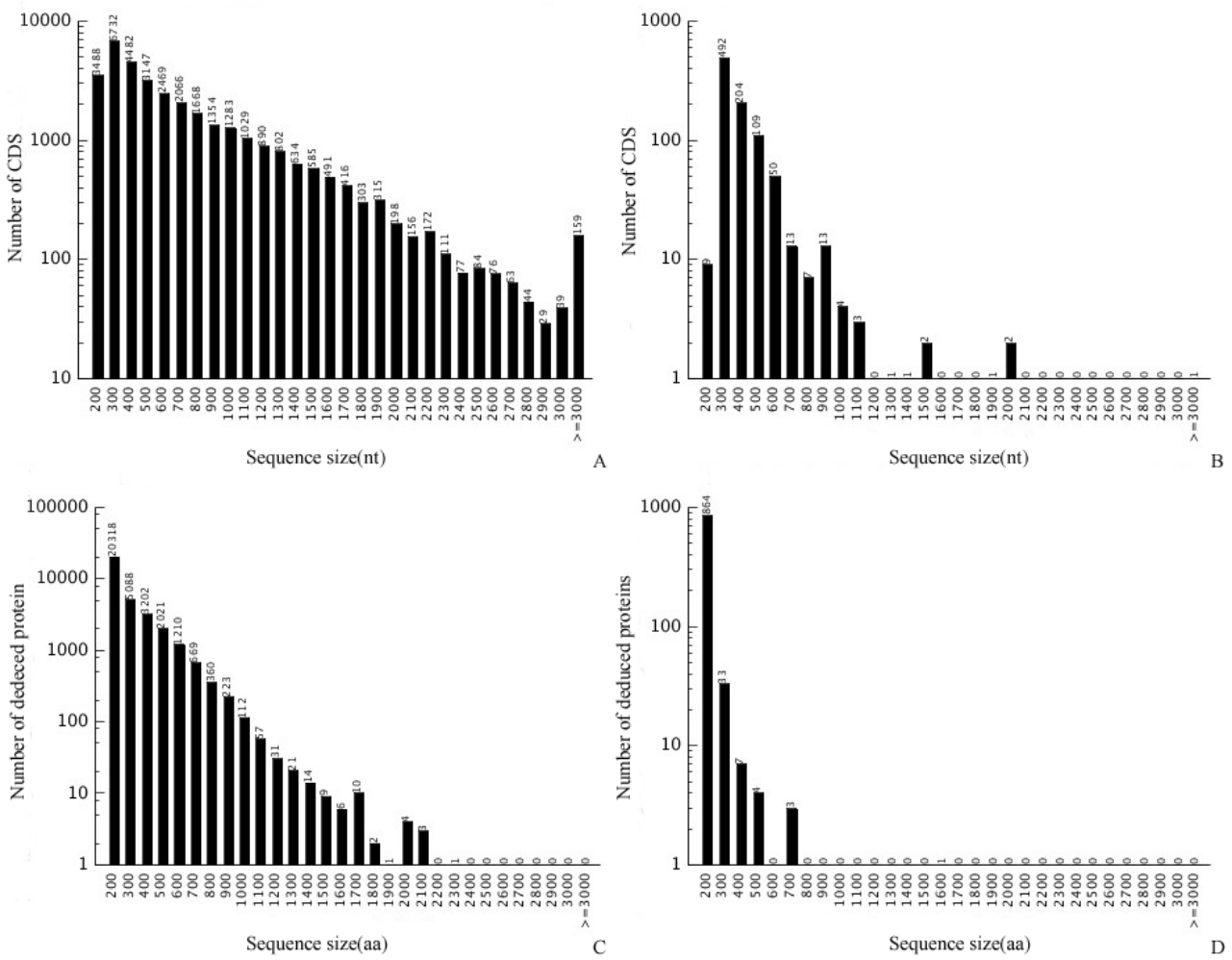

Fig 3. Length distribution of CDSs (A, B) and deduced protein sequences (C, D) of *J. curcas* unigenes analyzed by blastx (A, C) and ESTscan (B, D).

Generally, small compatible organic solutes such as amino acids (proline, glycine, alanine, serine), betaines and polyamines are argued as important osmotic protectants or stabilizing elements for membranes and macromolecules under almost all abiotic stresses (Ruelland et al., 2009). In our sequencing results, 2 unigenes encoding the key enzyme P5CS for proline biosynthesis were found. Glycine betaine, a quaternary ammonium compound, is synthesized in chloroplasts through two step-oxidations of choline catalyzed by choline monooxygenase (3 unigenes found herein) and betaine aldehyde dehydrogenase (1 unigene), respectively.

Cold stimuli often elicit an oxidative stress which results from an excess of reactive oxygen species (ROS) over the antioxidants (AOX) leading to abnormal redox status (Ruelland et al., 2009). Ascorbate and glutathione are two major abundant antioxidants in plants and function directly as the scavengers of ROS (singlet oxygen, superoxide and hydroxyl radicals). Both of them as well as their oxidative forms constitute the ascorbate-glutathione cycle, in which the involved enzymes such as ascorbate peroxidases, glutathione-S-transferase can be induced in response to cold (Yan et al., 2006; Cui et al., 2005). In our sequenced transcriptome, unigenes were found for all the enzymes related to this cycle, including glutathione peroxidase (10 unigenes), glutathione reductase (5), glutathione S-transferase (54), ascorbate peroxidase-1 (2), monodehydro-ascorbate reductase (10) and dehydroascorbate reductase (12). In addition, unigenes encoding the detoxifying enzymes such as superoxide dismutase (17 unigenes), catalase (22), peroxidase (106) with the ability to quench ROS were also found and likely up-regulated for their expression to adapt the cold environment.

Transcription factors are important for cold signaling and tolerance by modulating the expression of correlated functional genes (Stockinger et al., 1997; Gilmour et al., 1998; Chinnusamy et al., 2003). At the present time, the chilling transduction pathways can be grouped into two manners of ABA-dependent and ABA-independent. Especially, in the CBF pathway, belong to the part of ABA-independent pathways, ICE and CBF are the main transcription factors being with a considerable number of studies up to now, but its positive regulators such as SFR6 (Knight et al., 1999), LOV1 (Yoo et al., 2007), HOS9 (Zhu et al., 2004), SCOF1 (Kim et al., 2001) are scarce. Their encoding unigenes in *J. curcas* were definitely found in our sequenced transcriptome (2 for ICE, and 9 for CBF). Taken together, numerous unigenes of *J. curcas* involved in cold tolerance are listed in Table 5.

Materials and Methods

Seed germination and cold treatment of J. curcas

Seeds of *J. curcas* were surface-sterilized in 1.5% $CuSO_4$ for 30 min and rinsed thoroughly with sterile distilled water according to our previous methods (Li and Gong, 2011), and then soaked in distilled water for 24 h. The imbibed seeds were sown on six layers of wetted filter papers in trays and germinated in climate chamber at 26 ºC in the dark for 5 days. Then the geminated seeds were transferred into the pots containing sterilized soil with perlite, peat and sand (1:2:1) in climate chamber with the parameters of 26/20 ºC (day/night), 75% RH (Relative Humidity) and 16 h photoperiod, and sequentially grown for 14d.

For cold treatment, 2-week-old *J. curcas* seedlings were subjected to chilling at 12 ºC for 12h, 24h and 48h, respectively, as described in our early study (Ao et al., 2013). The leaves from each treatment as well as the control seedlings (continually under normal growth conditions) were harvested and froze in liquid nitrogen and stored at -80 ºC until RNA extraction.

RNA isolation and transcriptome sequencing

Tissue samples from three cold treatments and the control seedlings of *J. curcas* were equally mixed for RNA preparation. Total RNA was extracted using a TRIzol reagent (Invitrogen) following the manufacture's protocol. The concentration of the RNA was determined with Qubit Fluorometer, and the quality was checked by Agilent 2100 Bioanalyzer and 1% (w/v) agarose gel electrophoresis. Only RNAs with a 260nm/280nm ratio between 1.8 and 2.2, 260nm/230nm ratio≥2.0, 28S/18S ratio >1.0 and RIN (RNA integrated number) ≥7.0 were processed further.

By using oligo(dT)-attached magnetic beads, poly(A)-containing mRNAs were enriched from the total RNA extracts. Fragmentation buffer was added for interrupting mRNA to short segments, which were taken as the templates to synthesize cDNAs by random hexamer-primers. Then, the short DNA products were purified with QiaQuick PCR extraction kit and ligated with sequencing adapters. By agarose gel electrophoresis, suitable fragments with 200-700bp were selected for PCR amplification and sequencing templates by the Illumina Hiseq™ 2000 RNA-seq system.

Data cleaning and de novo sequence assembly into unigenes

The raw reads produced from sequencing machines contain dirty ones, including several types as follows: (1) Reads with adaptors (2) Reads with unknown nucleotides of a percentage larger than 5% (3) Low quality reads (more than 20% nucleotides with sequencing quality value Q≤10. Denote E as sequencing error rate, and Q as sequencing quality value calculated by the equation of Q = -10lgE). These dirty raw reads must be discarded to generate the clean ones for subsequent sequence assembly.

De novo sequence assembly was carried out with short reads assembling program of Trinity (Grabherr et al., 2011). Trinity firstly combines the reads with certain length of overlap to form longer fragments as contigs, and then the reads are mapped back to contigs, with paired-end reads. It is able to detect contigs from the same transcript as well as the distances between these contigs. Finally, contigs are connected constantly, and get sequences that cannot be extended on either end which are defined as unigenes of *J. curcas*.

Annotation of unigenes

Unigene sequences of *J. curcas* were firstly aligned to protein databases Nr, Swiss-Prot and Rc (the predicted proteome of *Ricinus communis*) using Blastx, and then by Blastn to nucleotide database Nt as well as the particular EST database (Jc_EST) at NCBI and predicted CDS sequences (Jc_CDS) based on published draft genomic database of *J. curcas* (e-value<1e-5 with identity of at least 80% over 100bp to detect pairwise similarities) (Costa et al., 2010), retrieving the proteins with highest sequence similarity to the query unigenes along with their protein functional annotations. Likewise, the annotation information of GO, COG, KEGG and expression level for each unigene was achieved.

For CDS (coding region sequence) analysis, unigenes of *J. curcas* were aligned by Blastx to protein databases Nr, Swiss-Prot (e-value<1e-5). The retrieved proteins with highest ranks in blast results were taken as the references to determine the CDSs of unigenes with the deduced amino acid sequences according to standard codon table. Those unigenes not aligned

Table 4. The 20 most highly expressed contigs or unigenes of *J. curcas*.

No.	Unigenes ID	No. of raw reads	BLAST annotation	COG functional classification
1	Unigene503_JC-CK_1A	869,154	Cell wall-associated hydrolase, partial [*Medicago truncatula*]	M
2	Unigene501_JC-CK_1A	775,863	Ribulose 1,5-bisphosphate carboxylase small subunit [*Jatropha curcas*]	C
3	Unigene499_JC-CK_1A	604,265	atpA [*Jatropha curcas*]	C
4	Unigene500_JC-CK_1A	513,881	ATP synthase CF0 B subunit [*Hevea brasiliensis*]	C
5	Unigene498_JC-CK_1A	498,815	Chlorophyll A/B binding protein, putative [*Ricinus communis*]	C
6	Unigene502_JC-CK_1A	428,818	ATP synthase CF0 C chain [*Welwitschia mirabilis*]	C
7	CL1387.Contig2_JC-CK_1A	274,650	Hypothetical protein MTR_5g051130 [*Medicago truncatula*]	S
8	CL2774.Contig1_JC-CK_1A	267,239	Pathogenesis-related protein 10a [*Jatropha curcas*]	V
9	Unigene508_JC-CK_1A	236,435	Chlorophyll A/B binding protein, putative [*Ricinus communis*]	C
10	Unigene505_JC-CK_1A	221,742	Predicted protein [*Populus trichocarpa*]	S
11	CL1260.Contig1_JC-CK_1A	128,976	Photosystem II CP43 chlorophyll apoprotein [*Medicago truncatul*a]	C
12	Unigene504_JC-CK_1A	127,594	Unknown	S
13	Unigene516_JC-CK_1A	118,790	Non-specific lipid-transfer protein 3 [*Prunus dulcis*]	I
14	CL2597.Contig1_JC-CK_1A	110,464	Hypothetical protein MTR_5g050970 [*Medicago truncatula*]	S
15	Unigene521_JC-CK_1A	98,658	Carbonic anhydrase, putative [*Ricinus communis*]	G
16	CL2429.Contig1_JC-CK_1A	94,087	Elongation factor 1-alpha, putative [*Ricinus communis*]	J
17	Unigene539_JC-CK_1A	88,401	Ribulose bisphosphate carboxylase/oxygenase activase 1[*Ricinus communis*]	C
18	Unigene530_JC-CK_1A	86,346	Latex plastidic aldolase-like protein [*Hevea brasiliensis*]	G
19	Unigene515_JC-CK_1A	85,487	Predicted protein [*Arabidopsis lyrata subsp. lyrata*]	S
20	Unigene531_JC-CK_1A	80,437	Polygalacturonase, putative [*Ricinus communis*]	G

C: Energy production and conversion; G: Carbohydrate transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; M: Cell wall/membrane/envelope biogenesis; S: Function known; V: Defense mechanisms.

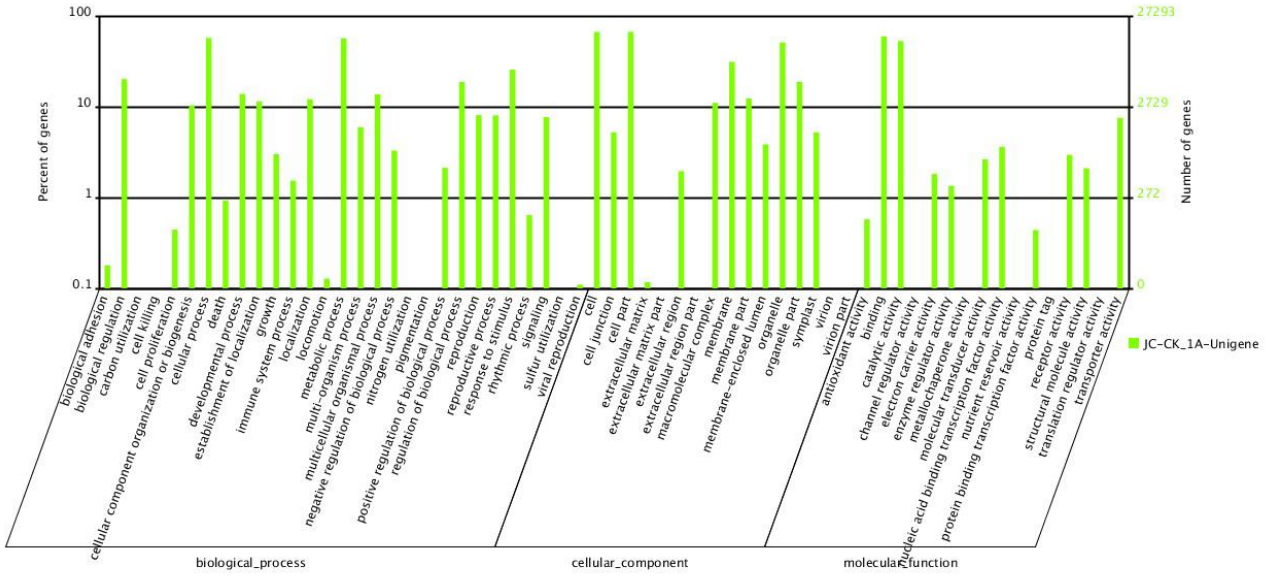

Fig 4. GO annotations of non-redundant unigenes of *J. curcas.*

Table 5. Unigenes of *J. curcas* involved in cold tolerance.

Aspects related to cold-tolerance	Symbol	Enzyme	No. of unigenes
Fluidity of membrane			
	SAD	Δ^9-stearoyl-ACP desaturase	1
	FAD3	Fatty acid desaturase 3	7
Entocyte			
	P5CS	Δ^1-pyrroline-5-carboxylate synthase	2
	CMO	Choline monooxygenase	3
	BADH	Betaine aldehyde dehydrogenase	1
	NR	Nitrate reductase	10
	ENO3	Enolase 3	2
	DHN	Dehydrin	4
	LEA-5	Late embryogenesis protein-5	2
Balance of AOX and ROX			
	SOD	Superoxide dismutase	17
	CAT	Catalase	22
	POD	Peroxidase	106
	GPX	Glutathione peroxidase	10
	GSTs	Glutathione S-transferase	54
	MDAR	Monodehydroascorbate reductase	10
	DHAR	Dehydroascorbate reductase	12
	GR	Glutathione reductase	5
	APX-1	Ascorbate peroxidase-1	10
Transcription factor			
	ICE1	Inducer of CBF Expression 1	2
	CBF	CRT/BRE binding factor (C-repeat binding factor)	9

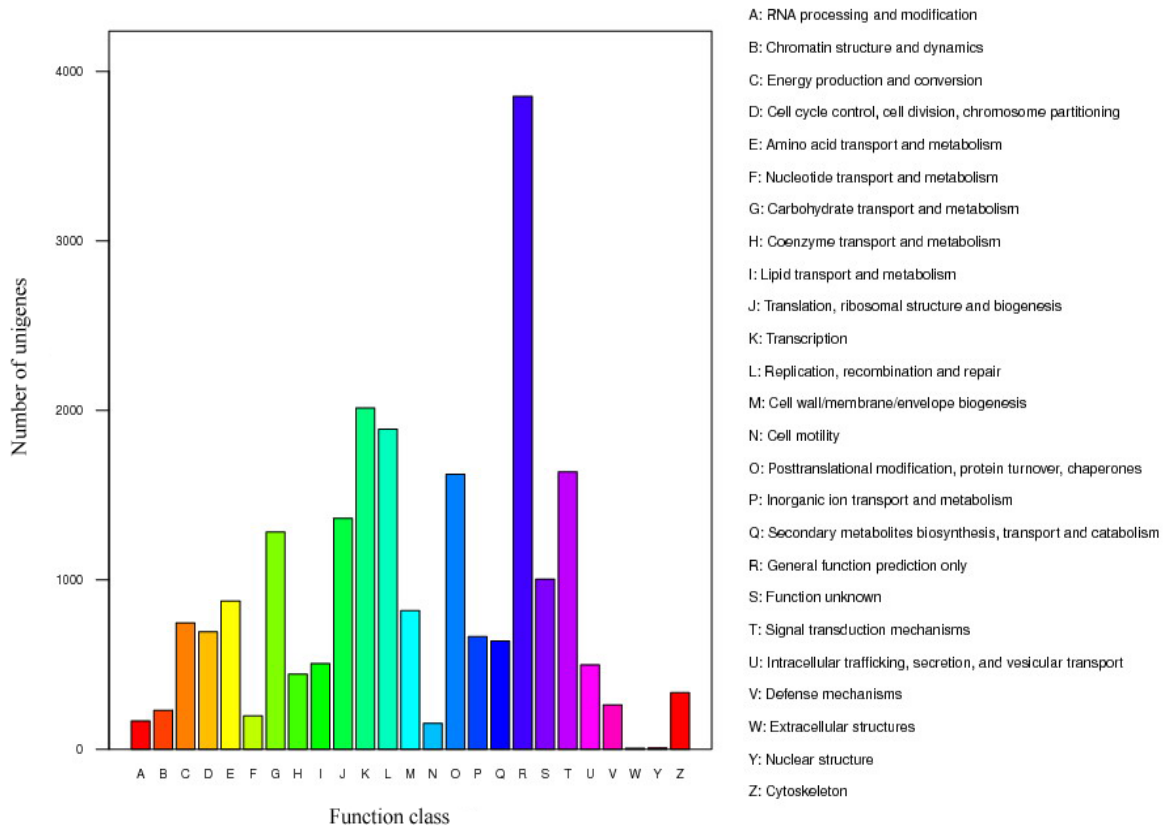

Fig 5. COG annotations of putative proteins deduced from unigenes of *J. curcas.*

to any databases were predicted by ESTScan program (Iseli et al., 1999), giving the sequence information of nucleotide and amino acid of their putative coding regions.

Functional classification of unigenes

With the aid of protein Nr annotation, GO functional annotation of *J. curcas* unigenes was performed. As we know, GO has three ontologies: molecular function, cellular component and biological process. Firstly, Blast2GO program (Conesa et al., 2005) was used to get GO annotation of all unigenes, and then WEGO software (Ye et al., 2006) was exploited for GO functional classification and systematically understanding the gene function and distribution in this species. COG is a database which was built on the basis of coding genes from a complete genome and their evolutional relationships among prokaryotes, archea, and eukaryotic organisms. With its help, orthologous genes can be classified. All unigenes of *J. curcas* were aligned to COG database for function prediction and classification in this work.

With the help of KEGG database, the cellular pathways and biological complexity for unigenes can be illustrated. According to KEGG annotated results, *J. curcas* unigenes were classified into different metabolic pathways, those of which involved in encoding cold-tolerance related components such as enzymes for synthesis of unsaturated fatty acids and osmolytes (proline, betaine), enzymes for antioxidant system, and transcription factors were retrieved from this sequenced transcriptome.

Conclusion

J. curcas is known as an elite bioenergy plant due to its high oil content and diesel-like quality of seed oil. Although an increased number of studies on this species are available in many respects such as regeneration, genetic transformation, physiologic and biochemical adaptations to abiotic stresses, transcriptomics (mainly focus on oil accumulation related genes) and proteomics of particular organs or developmental stages, a global cold-resistance trasncriptome analysis was still in waiting list until our work presented here. Through the deep high-throughput sequencing approach Illumina Hiseq™ 2000 RNA-seq, we obtained a comparatively largest cold-resistance transcriptome of *J. curcas*, from which 45,251 unigenes were derived and 35,791 could be annotated to known public databases. Functional classification for such a huge library of *J. curcas* unigenes was delicately accomplished and most highly expressed individuals were outlined. Furthermore, unigenes of *J. curcas* were elucidated particularly in cold tolerance. These data can give useful insights in understanding gene expression patterns correlated to the growth, development, cellular metabolism and environment adaptation of *J. curcas*, but also provide crucial gene targets for genetic improvements on the traits of most commercial concerns to develop new varieties with higher seed yield and oil content, better oil quality, low seed toxicity, effectual cold tolerance and so on.

Acknowledgements

We thank Ming Gong for suggestions on design and data analysis. We also thank Zhurong Zou for helpful discussions and improving the English. This work was supported by several grants from the National Foundations of Natural Sciences, China (No. 31260064 to MG, No.31060160, 31160169 to ZZ) and the Education Bureau of Yunnan Province (No. ZD2010004 to MG, No.2010Z087 to ZZ).

Data archiving statement

All the clean reads have been submitted to the sequence read archive (SRA) at NCBI with accession SRR653198. This Transcriptome Shotgun Assembly (TSA) project has been deposited at GenBank/EMBL/DDBJ under the accession GAHK00000000. The version described in this paper is the first version, GAHK01000000 (GAHK01000001-GAHK01045171).

References

Achten MJ, Nielsen LR, Aerts R, Lengkeek AG, Kjar ED, Trabucco A, Hansen JK, Maes H, Graudal L, Akinnifesl FK, Muys R (2010) Towards domestication of *Jatropha curcas*. Adv Biochem Eng Biot. 1:91-107.

Ao PX, Li ZG, Fan DM and Gong M (2013) Involvement of antioxidant defense system in chill hardening-induced chilling tolerance in *Jatropha Curcas* seedlings. Acta Physiol Plant. 35:153-160.

Asif MH, Mantri SS, Sharma A, Srivastava A, Trivedi I, Gupta P, Mohanty CS, Sawant SV, Tuli R (2010) Complete sequence and organisation of the *Jatropha curcas* (*Euphorbiaceae*) chloroplast genome. Tree Genet Genomes. 6:941-952.

Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK (2003) ICE1: A regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev. 17:1043-1054.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21: 3674-3676.

Costa GGL, Cardoso KC, Del Bem LEV, Lima AC, Cunha MAS, de Campos-Leite L, Vicentini R, Papes F, Moreira RC, Yunes JA, Campos FAP, Silva MJ (2010) Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L.. BMC Genomics. 11:462.

Cui SX, Huang F, Wang J, Ma X, Cheng YS, Liu JY (2005) A proteomic analysis of cold stress responses in rice seedlings. Proteomics. 5:3162-3172.

Deore AC, Johnson TS (2008) High-frequency plant regeneration from leaf-disc cultures of *Jatropha curcas* L.: an important biodiesel plant. Plant Biotechnol Rep. 2:7-11.

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in coldinduced COR gene expression. Plant J. 16:433-442.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompason DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma FD, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29:644-652.

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol. 138-148.

Johnson TS, Eswaran N, Sujatha M (2011) Molecular approaches to improvement of *Jatropha curcas* Linn. As a sustainable energy crop. Plant Cell Rep. 30:1573-1591.

Khurana-Kaul V, Kachhwaha S, Kothari SL (2010) Direct shoot regeneration from leaf explants of *Jatropha curcas* in response to thidiazuron and high copper contents in the

medium. Biol Plantarum. 54:369-372.
Kim JC, Lee SH, Cheong YH, Yoo CM, Lee SI, Chun HJ, Yun DJ, Hong JC, Lee SY, Lim CO, Cho MJ (2001) A novel cold inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants. Plant J. 25:247-259.
Knight H, Veale EL, Warren GJ, Knight MR (1999) The sfr6 mutation in *Arabidopsis* suppresses low-temperature induction of genes dependent on the CRT/DRE sequence motif. Plant Cell. 11:875-886.
Kumar N, Anand KG, Reddy MP (2010) Shoot regeneration from cotyledonary leaf explants of *Jatropha curcas*: a biodiesel plant. Acta Physiol Plant. 32:917-924.
Kumar N, Reddy MP (2010) Plant regeneration through the direct induction of shoot buds from petiole explants of *Jatropha curcas*: a biofuel plant. Ann Appl Biol. 156:367-375.
Li MR, Li HQ, Jiang HW, Pan XP, Wu GJ (2008) Establishment of an *Agrobacterium* mediated cotyledon disc transformation method for *J. curcas*. Plant Cell Tiss Org. 92:173-181.
Li ZG and Gong M (2011) Effects of different chemical disinfectant on seed germination and seedling growth of *Jatropha curcas* L.. Seed. 30:4-7,12.
Natarajan P, Parani M (2011) De novo assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing. BMC Genomics. 12:191.
Pan J, Fu Q, Xu ZF (2010) *Agrobacterium tumefaciens*-mediated transformation of biofuel plant *Jatropha curcas* using kanamycin selection. Afr J Biotechnol. 9:6477-6481.
Ruelland E, Vaultier MN, Zachowski A, Hurry V (2009) Cold signalling and cold acclimation in plants. Adv Bot Res. 49:35-150.
Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-i T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S, Fukui K (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L.. DNA Res. 18:65-76.
Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana* CBF1 encodes an AP2 domain‐containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA. 94:1035-1040.
Tang M, Sun J, Liu Y, Chen F, Shen S (2007) Isolation and functional characterization of the JcERF gene, a putative AP1/EREBP domain-containing transcription factor, in the woody oil plant *Jatropha curcas*. Plant Mol Biol. 63:419-428.
Uemura M, Steponkus PL (1997) Effect of cold acclimation on the lipid composition of the inner and outer membrane of the chloroplast envelope isolated from rye leaves. Plant Physiol. 114:1493-1500.
Wang Y, Huang J, Gou CB, Dai X, Chen F, Wei W (2011) Cloning and characterization of a differentially expressed cDNA encoding myo-inositol-1-phosphate synthase involved in response to abiotic stress in *Jatropha curcas*. Plant Cell Tiss Organ Cult. 106: 269-277.
Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 10:57-63.
Yan SP, Zhang QY, Tang ZC, Su WA, Sun WN (2006) Comparative proteomic analysis provides new insights into chilling stress responses in rice. Mol Cell Proteomics. 5:484-496.
Yang CY, Fang Z, Li B, Long YF (2012) Review and prospects of *Jatropha* biodiesel industry in China. Renew Sust Energ Rev. 16:2178-2190.
Yang MF, Liu YJ, Liu Y, Chen H, Chen F, Shen SH (2009) Proteomic analysis of oil mobilization in seed germination and postgermination development of *Jatropha curcas*. J Proteome Res. 8:1441-1451.
Ye J, Fang L, Zheng HK, Zhang Y, Chen J, Zhang ZJ, Wang J, Li ST, Li RQ, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34:W293-297.
Yoo SY, Kim Y, Kim SY, Lee JS, Ahn JH (2007) Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. PLoS One. 2: e642.
Zhang Y, Wang Y, Jiang L, Xu Y, Wang Y, Lu D, Chen F (2007) Aquaporin JcPIP2 is involved in drought responses in *Jatropha curcas*. Acta Biochim Biophys Sin. 39:787-794.
Zhang FL, Niu B, Wang YC, Chen F, Wang SH, Xu Y, Jiang LD, Gao S, Wu J, Tang L, Jia YJ (2008) A novel betaine aldehyde dehydrogenase gene from *Jatropha curcas*, encoding an enzyme implicated in adaptation to environmental stress. Plant Sci. 174:510-518.
Zhu J, Shi H, Lee BH, Damsz B, Cheng S, Stirm V, Zhu JK, Hasegawa PM, Bressan RA (2004) An *Arabidopsis* homeodomain transcription factor gene, HOS9, mediates cold tolerance through a CBF independent pathway. Proc Natl Acad Sci USA. 101:9873-9878.