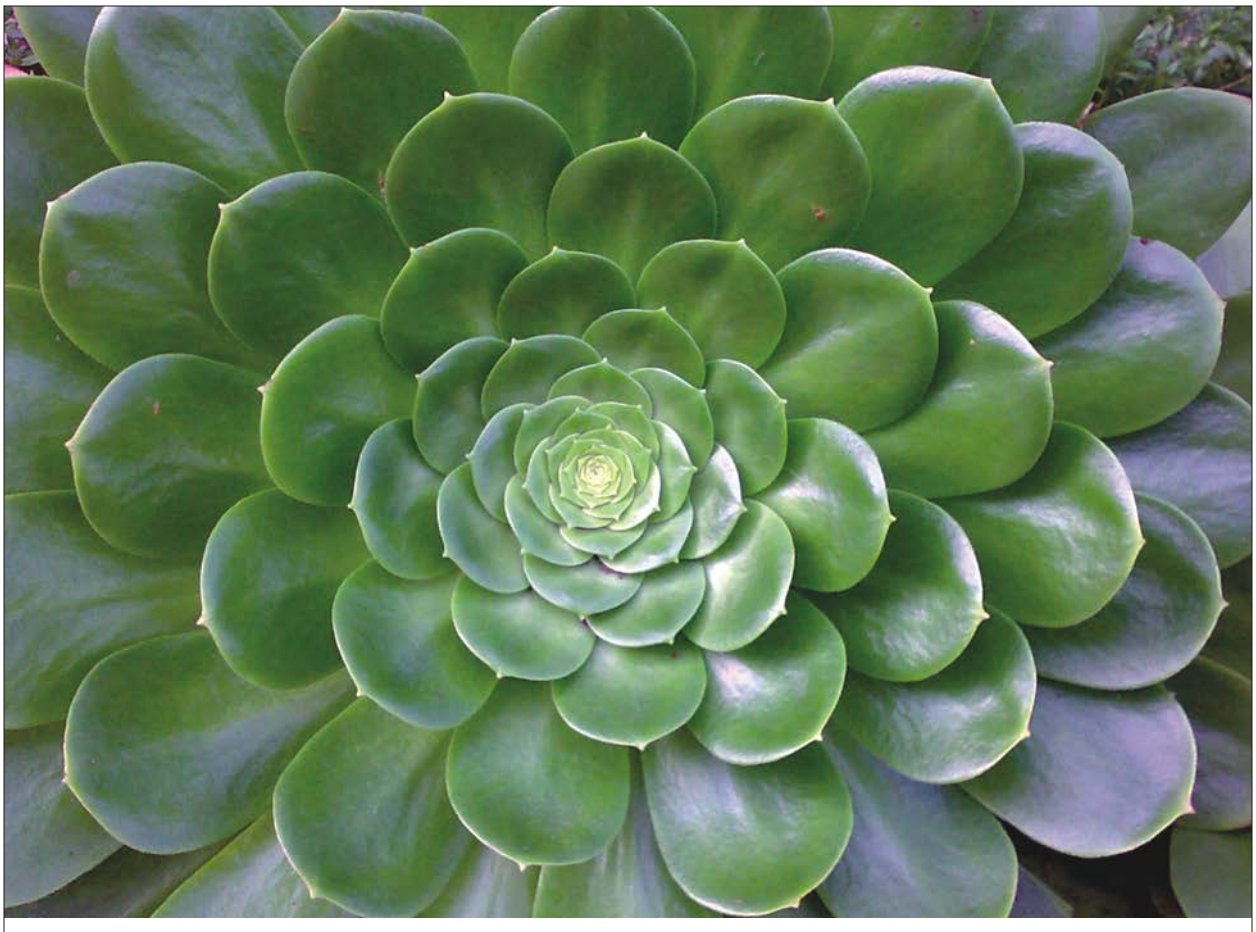

Journal of
Plant Sciences

ISSN 1816-4951

Academic
Journals Inc.

www.academicjournals.com

Journal of Plant Sciences 10 (1): 1-14, 2015
ISSN 1816-4951 / DOI: 10.3923/jps.2015.1.14

Transcriptome of Teak (*Tectona grandis*, L.f) in Vegetative to Generative Stages Development

[1,3]D.S. Diningrat, [1]S.M. Widiyanto, [1]A. Pancoro, [1]Iriawati, [2]D. Shim, [2]B. Panchangam, [2]N. Zembower and [2]J.E. Carlson

[1]School of Life Sciences and Technology, Institute Teknologi Bandung, Bandung, 40132, Indonesia
[2]Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA, 16802, USA
[3]Department of Biology, Mathematic and Natural Science Faculty, Medan State University, 20221, Medan

Corresponding Author: S.M. Widiyanto, School of Life Sciences and Technology, Institut Teknologi Bandung, Bandung, 40132, Indonesia

ABSTRACT

Teak is one of the highly famous woody plant species for its premier quality of wood. Teak has problem on productivity because of long reproductive cycle. The problem is basically related to mechanism of flower development. The aim of this study was preliminary development of expressed gene database to characterize the floral transcriptome in teak. Two subtracted cDNA libraries were constructed from vegetative and generative bud tissues. Libraries were sequenced using Illumina MiSeq technology which generated paired-end read sequences 3,778,316 for vegetative and 3,701,878 for generative. The sequences assembled *de novo* into 87,365 transcript contigs consisting of 42,435,728 bases with N50 of 498 bp using CLC-Genomics Workbench. 76,169 (87.18%) of the 87,365 assembled contigs exhibited significant similarity BLASTN to *Solanum lycopersicum* database (www.phytozome.com). The assembled contigs were annotated through high stringency BLASTX analysis to proteome of *S. lycopersicum*. Distribution of contigs abundance between vegetative and generative stages analyzed using the DEGseq approach. The numbers of contigs distribution are 24,730 in vegetative, 28,912 in generative and 33,723 in both stages. The functionally protein datasets characterized by Gene Ontology (GO) annotation and KEGG metabolic pathways assignments for the result of DEG analysis. These contigs, 18,756 (75.84%) from vegetative, 22,089 (76.40%) from generative and 22,917 (67.96%) from both stages were assigned to GO classes. A total of 1455 (13.77%) were mapped to 30 pathways from vegetative, 1,638 (13.70%) were mapped to 27 pathways from generative and 1,652 (12.20%) were mapped to 30 pathways from both by BLAST comparison against the KEGG database. The biological processes of flowering developments were identified in the biological process dataset and the numbers of contigs were discovered different between stages. This transcriptome dataset information will act as a valuable resource for further molecular genetic studies teak, as well as for isolation and characterization of functional genes involved in flowering development pathways.

Key words: *Tectona grandis*, transcriptome analysis, vegetative stage, generative stage, transitional development

INTRODUCTION

Teak (*Tectona grandis*, L. f) is a tropical tree species distributed naturally in countries including India, Myanmar, Thailand, Myanmar and Indonesia (Orwa *et al*., 2009; Palupi *et al*., 2010; Lyngdoh *et al*., 2010). Teak is one of the world's premier hardwood tree species, highly

famous for its quality, profile and durability of timber. In Indonesia, teak flowering usually appears every year at the beginning of the rainy season (October-November) and only few flowers (about 1%) develop into fruits. Fruits fall gradually during the dry season (Orwa *et al*., 2009). According to the fact, the main limitations of teak improvement are it has a long reproductive cycle and produces low seeds. Both problems are basically related to mechanism flower development (De Gyves *et al*., 2007; Rosli *et al*., 2009; Widiyanto *et al*., 2009; Palupi *et al*., 2010). Hence, the determination of the genetic pathways and identifying specific genes involved in teak flowering and flower development could be beneficial for teak productivity improvement. We are interested in studying more about the roles of genes that control development of flowers in teak especially during the transition period between shifts of the vegetative to reproductive phase. This study was preliminary of teak floral transcriptome characterization, before isolation and characterization of functional genes involved in flowering development pathways.

In this study, we sequenced the transcriptome of *T. grandis* using the next generation of high throughput paired-end RNA sequencing (RNA-seq) technology, Illumina MiSeq^{TM} 2000 (Fan *et al*., 2013). Then, CLC bio bioinformatics technology tool was used to perform a *de novo* assembly and annotation without prior genome information (Agusti *et al*., 2011; Angeloni *et al*., 2011). This transcriptome database helped to reveal much about the functional genomics of *T. grandis* and was then used to predict the functional classification of many unigenes using GO and KEGG pathway analysis (Chang *et al*., 2013; Rosero *et al*., 2011). These results lay the foundation for understanding the relation between gene expression patterns and plant development, physiology and structure and will be helpful for the molecular approach to improve of *T. grandis*. Furthermore, we focused on the sequences that are related to flowering developmental biological process in the aim of exploring the relationship between genes in transition development vegetative to generative stage.

MATERIALS AND METHODS

Teak tissues materials and RNA isolation: Vegetative and generative stage shoot tips of teak were collected from a 12 years old teak plant in Institute Technology Bandung, Indonesia for RNA isolation. The following VS tissues were sampled from vegetative apical shoots. LB2 tissues were sampled from lateral (nodal) floral-buds 2nd of generative stage shoots (Fig. 1). Both of teak tissue samples were frozen in liquid nitrogen immediately upon collection and put in Dry Shipper for shipping from ITB-Indonesia to Pennsylvania State University (PSU)-USA. Samples were immediately frozen at -80°C upon arrival at PSU until use. Total RNA was obtained using the

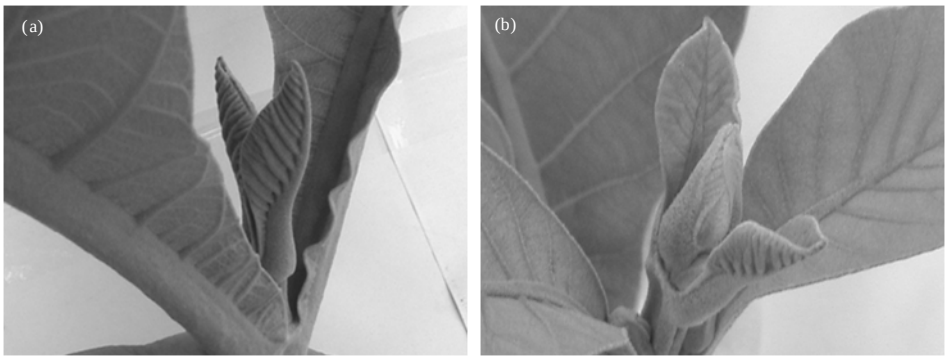

Fig. 1(a-b): Vegetative and generative stage shoot tips of teak, (a) Vegetative stage and (b) Generative stage

method for RNA isolation protocol that developed by Dr. Carlson's team at Schatz Center Laboratory, PSU-USA. Frozen tissue were ground to a fine powder under liquid nitrogen and dispersed in CTAB buffer. Following 2 chloroform extractions, RNA was precipitated with $LiCl_2$,again extracted with chloroform and precipitated with ethanol. The resulting RNA pellet was resuspended in 20-100 μL of DEPC-treated water (Barakat *et al*., 2012). RNA concentration analysis on a Qubit™ fluorometer (www.invitrogen.com/qubit) to showed a total yield of RNA sample. The concentration of RNA are 555 and 206 ng μL^{-1} for VS and LB2 sample, respectively. The integrity of RNA was assessed with the Agilent 6000 RNA Nano Chip Kit on 2100 Bioanalyzer (Agilent Technologies).

Paired-end cDNA library preparation and MiSeq Illumina sequencing: Total RNA of teak was extracted from the two tissues using the protocol described previously. The double-stranded cDNA was synthesized using the cDNA Synthesis System using random hexamer primers (illumina) according to manufacturer's instructions. The paired-end library was developed according to the protocol of the Paired-End sample Preparation kit (lllumina, USA) (Angeloni *et al*., 2011; Li *et al*., 2012; Fan *et al*., 2013). The resulting library was sequenced at Penn State University using Illumina MiSeq™ 2000 (Illumina Inc., USA).

Transcript assembly: Two sequence data in FASTQ files computed with CLCbio for transcript assembly strategy (Angeloni *et al*., 2011). Paired-End reads were trimmed for quality score and the presence of repeated sequences >50 bp using the modified Mott-trimming algorithm present (default parameters) in CLCbio. We assembled *de novo* the Illumina-trimmed paired-end reads into transcript contigs using the software CLC Genomics Workbench by setting minimum 95% identity, minimum 40% overlap and 200 bp as minimum contig length.

Contig annotation: The quality of the *de novo* assembly was assessed with a local BLASTn (e-value<10^{-6}) alignment of all the contigs against *A. thaliana, P. tricocharpa, M. guttatus*, *S. tuberosum* and *S. lycopersicum* (www.phytozome.com) using CLCbio Workbench. Top hit species results use for homology based annotations of teak (Rosero *et al*., 2011; Barakat *et al*., 2009, 2012).

DEGseq analysis: Comparison of Digitally Gene Expression (DEGseq) between vegetative and generative tissues was done using RNAseq analysis software test developed by CLCbio-Genomic Workbench. DEGseq analysis was used to identify flowering development genes in transcript abundance because it integrates several statistical methods (Barakat *et al*., 2009, 2012). The number of reads per contig for each gene was compared between vegetative stage as control and generative tissues in teak separately. Similar analyses were performed for gene orthologs from both tissues. Orthologs were identified using a reciprocal best hit approach. RNAseq employs a random sampling model based on the read count in vegetative and generative tissues libraries and performs a hypothesis test based on that model. Genes expression in vegetative, generative and both of them are identified and go to GO enrichment.

GO analysis: Further assessment of the quality of the *de novo* assembly was carried out as follows. We compared the depth and the length of contig coverage with reference to orthologous genes in *S. lycopersicum* and *A. thaliana*, by plotting the ratio of contig length to *S. lycopersicum* and *A. thaliana* orthologue coding region length against coverage depth. Orthologous genes were

retrieved performing a local BLASTX alignment (e-value<10^{-6}) using CLCbio Workbench with the TAIR9 *A. thaliana* database and *S. lycopersicum* predicted proteins (Unipro/Swissprot database). To further assess the coverage and the quality of the assembly, we used BLASTX to align the contigs to the manually curated protein database Uniprot/Swissprot using DAVID Bioinformatics Resources at http://david.abcc.ncifcrf.gov/ (Huang *et al*., 2007; Jiao *et al*., 2012). DAVID Bioinformatics is an automated tool for the assignment of Gene Ontology (GO) terms to BLAST hits and it has been designed for use with novel sequence data (Jiao *et al*., 2012), Assignment of GO terms to contigs with significant BLASTX match with Swissprot (http://www.expasy.ch/sprot/) and the KEGG pathway (http://www.genome.jp/kegg/) were also performed using DAVID Bioinformatics. In addition, we generated GO assignments for *A. thaliana* and *S. lycopersicum* annotated proteins to compare the distribution of functional annotation in teak to those plants species with a well-characterized transcriptome, we did the GO analysis for the result of DEG analysis from vegetative stage and generative stage of teak samples.

RESULTS

Illumina sequencing output statistics and reads assembly: *T. grandis* vegetative and generative cDNA libraries were constructed from a pool of RNA isolated from vegetative and generative bud tissues teak tree using the Illumina MiSeq^{TM} 2000 system at Penn State University. A total of 3,778,316 and 3,701,878 reads were generated from vegetative and generative teak transcriptomes, respectively (Table 1). The average length of the reads was 151 nucleotides (Fig. 2). *De novo* contig construction of the Illumina reads using the CLCbio assembly software led to the construction of 87,365 contigs from combined vegetative and generative teak (Table 2). Those contigs were having an average length of 486 nt, 225 nt for minimum length and 4,361 nt for maximum length (Fig. 3).

Table 1: Summary statistics of cDNA library

Library	Sequences
Vegetative Shoot (VS)	3, 701, 878 sequences in pair
Generative shoot (LB2)	3, 778, 316 sequences in pair

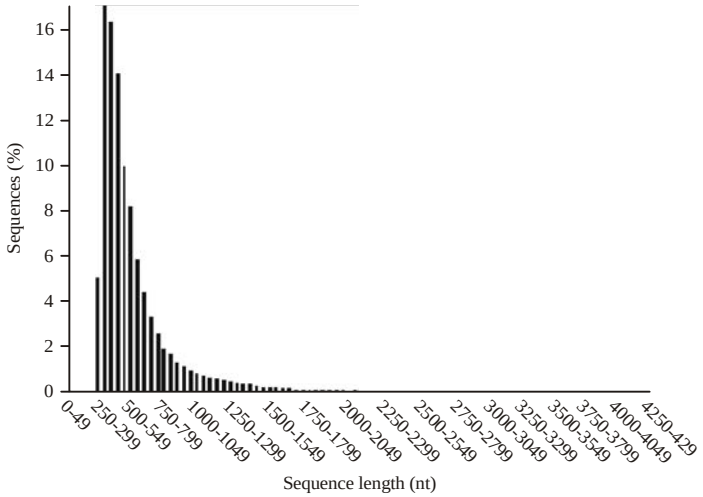

Fig. 2: Paired reads distance distribution

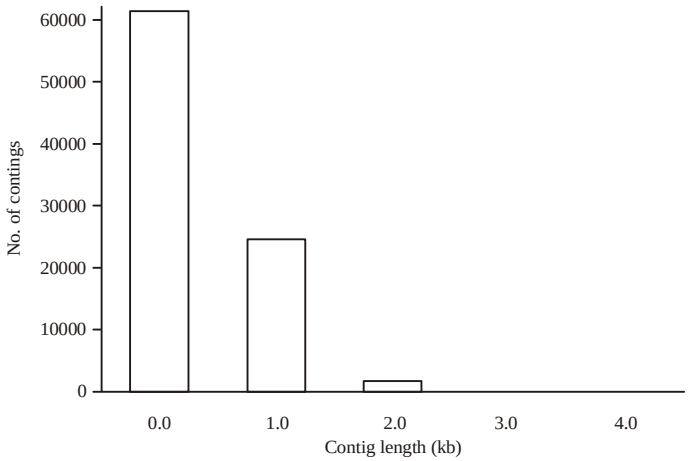

Fig. 3: Histogram of the frequency of different contigs sizes in transcriptome assemblies of teak samples

Table 2: Summary statistics of sequencing and *de novo* assembly results

Category	Values
Input sequence	3,701, 878 and 3, 778, 316
Total bases	42, 435, 728
Contigs	87, 365
Minimum length of contigs	225
Maximum length of contigs	4, 361
Average length of contigs	486
N75	359
N50	498
N25	805

Table 3: No. of BLAST hit top species for homology based annotations of teak contigs

Species	No. of hits	Percentage
A. thaliana	64,961	74.36
P. trichocarpa	7,479	8.56
M. guttatus	16,525	18.91
S. tuberosum	16,525	18.91
S. lycopersicum	76,169	87.18

Contigs annotations: Collection of 87,365 contigs, enriched in vegetative to generative transition stage related transcripts, was obtained from both vegetative and generative bud subtracted libraries. Top hit species for homology based annotations of teak contigs (Table 3) were: *A. thaliana* (74.36%), *P. trichocarpa* (8.56%), *M. guttatus* (18.91%), *S. tuberosum* (18.91%) and *S. lycopersicum* (87.18%). *S. lycopersicum* was the highest blast hits species for teak. This is a remarkable result when considering the current state of functional annotation of teak to the *S. lycopersicum* proteome database (www.phytozome.com).

Transcriptome comparison between vegetative and generative tissues: We compared the transcriptomes from teak vegetative tissues and generative tissues to gain insight into the differences in the gene activity of the transition vegetative to generative stages in teak development. This comparison showed that the distribution of contigs in vegetative stage, generative stage and both using DEG analysis software (Fig. 4).

Detailed comparison of the Gene Ontology (GO) transcriptomes in vegetative stage, generative stage and both showed the different percentage of biological processes, cellular component and molecular function (Fig. 5). Figure 6 showed top 25 of biological process, cellular component and

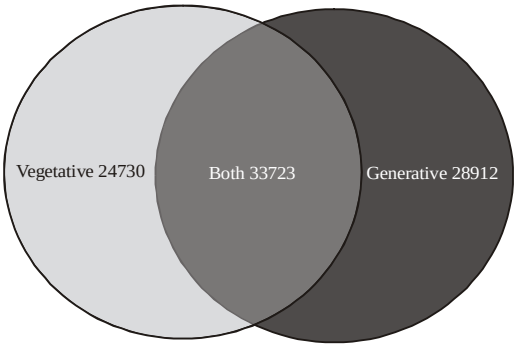

Fig. 4: Contigs distribution result of DEG analysis

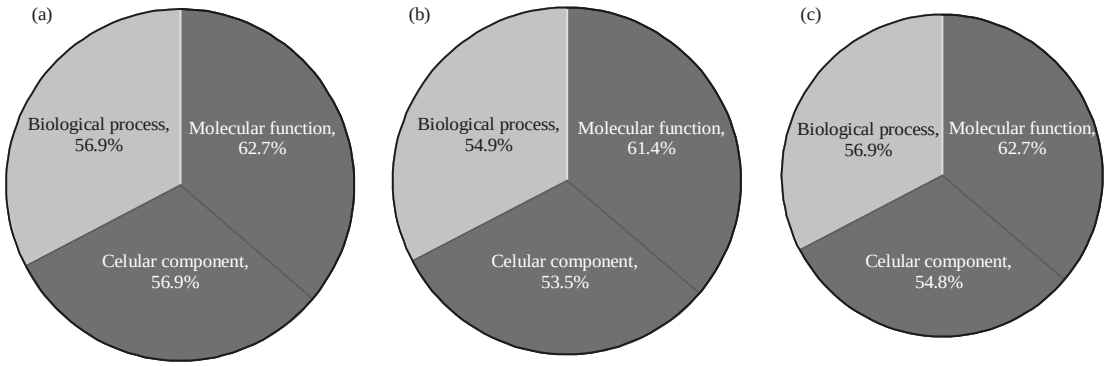

Fig. 5(a-c): Chart of GO categories of teak, (a) Vegetative, (b) Both and (c) Generative

molecular function that occurred in the tissue samples. In the category of biological process, response to abiotic stimulus, phosphorus metabolic process and phosphate metabolic process comprised the largest proportion of sequences, accounting for (6.4% in vegetative stage, 5.53% in generative stage and 5.3% in both), (6% in vegetative stage, 6.3% in generative stage and 6% in both) and (6% in vegetative stage, 6.3% in generative stage and 6% in both) of the total, respectively. According to flowering development biological processes, there are post-embryonic development (4.4% in vegetative stage, 4.1% in generative stage and 4% in both) and reproductive developmental process (4.1% in vegetative stage, 4% in generative stage and 3.8% in both), reproductive structure development (3.8% in vegetative stage, 3.6% in generative stage and 3.4% in both) comprised part of the top ten largest proportion.

In the category of cellular components, plastid comprised (16.20% in vegetative stage, 13.7% in generative stage and 13.8% in both), chloroplast was (15.8% in vegetative stage, 13.3% in generative stage and 13.5% in both) and intrinsic to membrane (11.3% in vegetative stage, 11.6% in generative stage and 11.3% in both) these three subgroups were dominant over the others. In the category molecular function, sequences with the functions of nucleotide binding, metal ion binding and purine nucleotide binding comprised (17.4% in vegetative stage, 18.4% in generative stage and 17.3% in both), (14.7% in vegetative stage, 14.9% in generative stage and 14.5% in both) of the total (Table 4).

On the other hand we also identified the other biological processes of flower development. We compared the biological processes of flower development between vegetative stage, generative stage

and both. In the flowering developmental biological processes, positive regulation of developmental process, positive regulation of flower development and regulation of meristem development comprised only in vegetative stage. The pollen tube development, tube development and negative

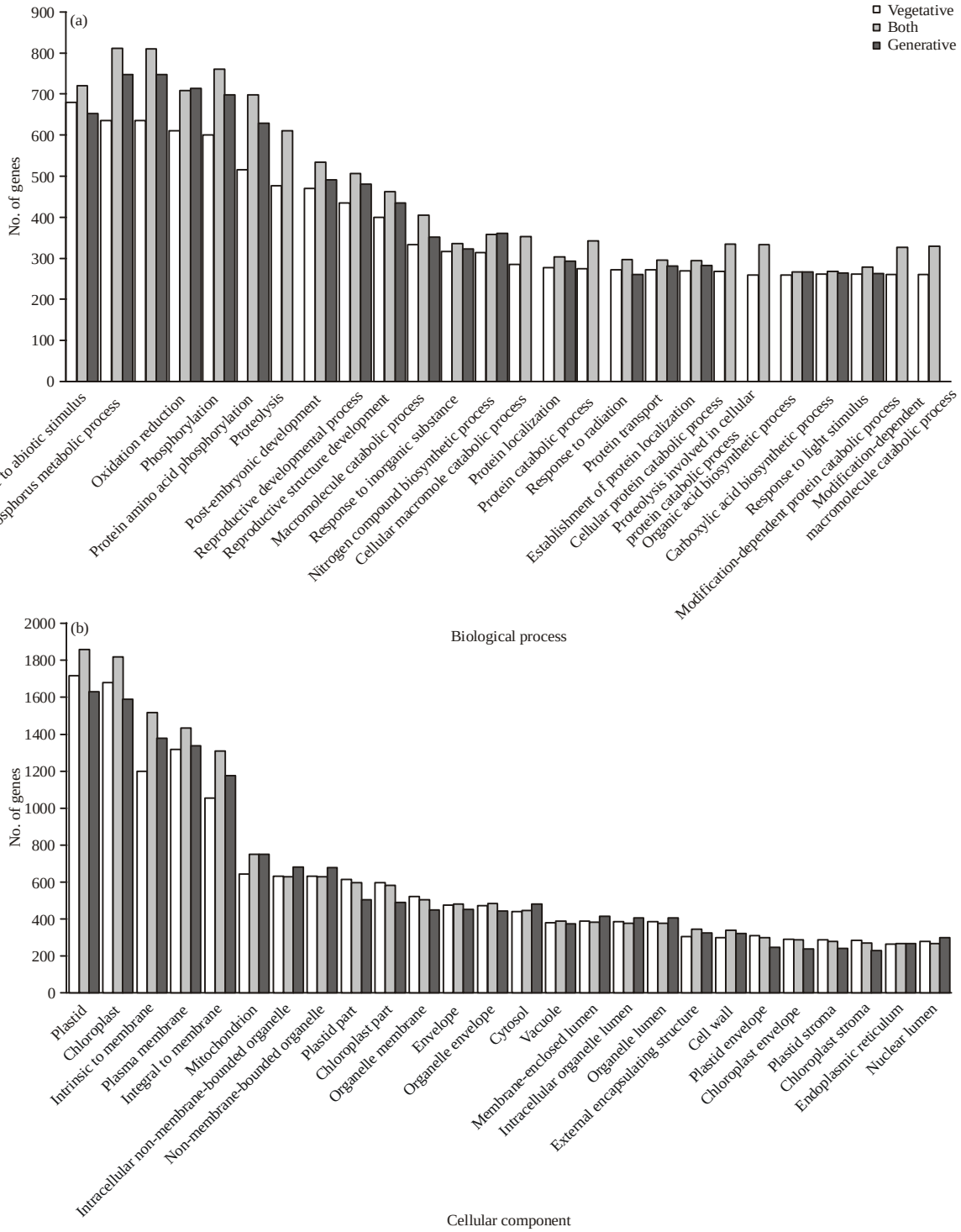

Fig. 6(a-c): Continue

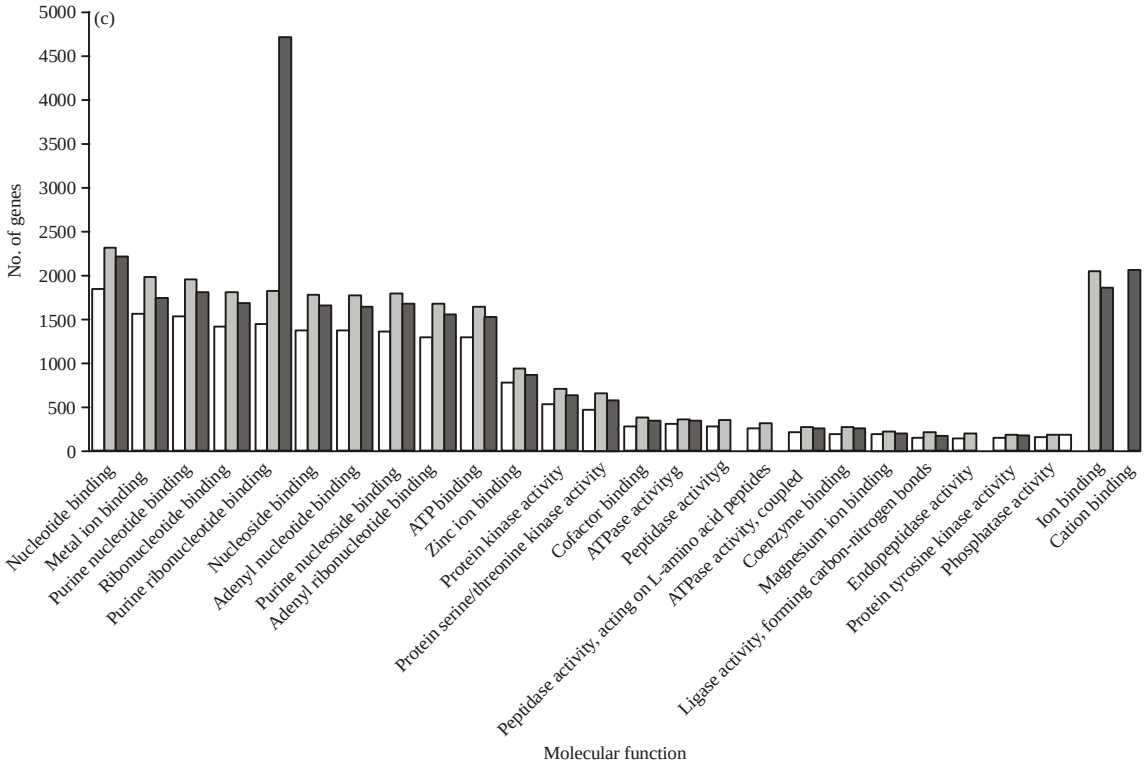

Fig. 6(a-c): Histogram of GO classification of teak, (a) Biological process, (b) Cellular component and (c) Molecular function

Table 4: GO classification of teak

	Vegetative		Both		Generative	
Categories	Value	%	Value	%	Value	%
Biological process categories						
Response to abiotic stimulus	680	6.4	719	5.3	654	5.5
Phosphorus metabolic process	636	6.0	813	6.0	747	6.3
Phosphate metabolic process	635	6.0	812	6.0	747	6.3
Oxidation reduction	611	5.8	711	5.3	713	6.0
Phosphorylation	580	5.5	758	5.6	698	5.9
Protein amino acid phosphorylation	518	4.9	699	5.2	630	5.3
Proteolysis	479	4.5	613	4.5	0	0.0
Post-embryonic development	470	4.4	535	4.0	493	4.1
Reproductive developmental process	436	4.1	508	3.8	480	4.0
Reproductive structure development	400	3.8	463	3.4	435	3.6
Macromolecule catabolic process	333	3.1	407	3.0	353	3.0
Response to inorganic substance	318	3.0	338	2.5	327	2.7
Nitrogen compound biosynthetic process	312	2.9	358	2.7	362	3.0
Cellular macromolecule catabolic process	285	2.7	354	2.6	0	0.0
Protein localization	279	2.6	303	2.2	295	2.5
Protein catabolic process	277	2.6	347	2.6	0	0.0
Response to radiation	272	2.6	298	2.2	260	2.2
Protein transport	271	2.6	294	2.2	284	2.4
Establishment of protein localization	271	2.6	294	2.2	284	2.4
Cellular protein catabolic process	267	2.5	337	2.5	0	0.0
Proteolysis involved in cellular protein catabolic process	264	2.5	334	2.5	0	0.0
Organic acid biosynthetic process	261	2.5	269	2.0	267	2.2
Carboxylic acid biosynthetic process	261	2.5	269	2.0	267	2.2
Response to light stimulus	261	2.5	282	2.1	250	2.1
Modification-dependent protein catabolic process	261	2.5	331	2.5	0	0.0
Modification-dependent macromolecule catabolic process	261	2.5	331	2.5	0	0.0

Table 4: Continue

Categories	Vegetative		Both		Generative	
	Value	%	Value	%	Value	%
Cellular component categories						
Plastid	1719	16.2	1862	13.8	1632	13.7
Chloroplast	1683	15.8	1820	13.5	1591	13.3
Intrinsic to membrane	1205	11.3	1524	11.3	1377	11.6
Plasma membrane	1323	12.5	1435	10.6	1341	11.3
Integral to membrane	1057	9.9	1313	9.7	1181	9.9
Mitochondrion	646	6.1	753	5.6	752	6.3
Intracellular non-membrane-bounded organelle	633	6.0	631	4.7	682	5.7
Non-membrane-bounded organelle	633	6.0	631	4.7	682	5.7
Plastid part	618	5.8	599	4.4	506	4.2
Chloroplast part	602	5.7	586	4.3	491	4.1
Organelle membrane	524	4.9	502	3.7	453	3.8
Envelope	476	4.5	485	3.6	450	3.8
Organelle envelope	473	4.5	483	3.6	446	3.7
Cytosol	440	4.1	447	3.3	486	4.1
Vacuole	377	3.5	393	2.9	380	3.2
Membrane-enclosed lumen	394	3.7	385	2.9	416	3.5
Intracellular organelle lumen	389	3.7	381	2.8	410	3.4
Organelle lumen	389	3.7	381	28.0	410	3.4
External encapsulating structure	309	2.9	348	2.6	331	2.8
Cell wall	304	2.9	346	2.6	326	2.7
Plastid envelope	313	2.9	304	2.3	252	2.1
Chloroplast envelope	297	2.8	292	2.2	242	2.0
Plastid stroma	293	2.8	281	2.1	243	2.0
Chloroplast stroma	284	2.7	271	2.0	231	1.9
Endoplasmic reticulum	264	2.5	270	2.0	267	2.2
Nuclear lumen	282	2.7	270	2.0	299	2.5
Molecular function categories						
Nucleotide binding	1853	17.4	2329	17.3	2190	18.4
Metal ion binding	1566	14.7	1961	14.5	1772	14.9
Purine nucleotide binding	1523	14.3	1956	14.5	1824	15.3
Ribonucleotide binding	1442	13.6	1833	13.6	1713	14.4
Purine ribonucleotide binding	1442	13.6	1833	13.6	4713	14.4
Nucleoside binding	1401	13.2	1802	13.4	1680	14.1
Adenyl nucleotide binding	1394	13.1	1795	13.3	1674	14.0
Purine nucleoside binding	1394	13.1	1795	13.3	1674	14.0
Adenyl ribonucleotide binding	1315	12.4	1675	12.4	1565	13.1
ATP binding	1296	12.2	1656	12.3	1543	12.9
Zinc ion binding	777	7.3	963	7.1	862	7.2
Protein kinase activity	536	5	728	5.4	651	5.5
Protein serine/threonine kinase activity	477	4.5	650	4.8	571	4.8
Cofactor binding	298	2.8	386	2.9	363	3.0
ATPase activity	313	2.9	365	2.7	349	2.9
Peptidase activity	285	2.7	354	2.6	0	0.0
Peptidase activity, acting on L-amino acid peptides	271	2.6	332	2.5	0	0.0
ATPase activity, coupled	233	2.2	283	2.1	272	2.3
Coenzyme binding	218	2.1	281	2.1	265	2.2
Magnesium ion binding	204	1.9	229	1.7	215	1.8
Ligase activity, forming carbon-nitrogen bonds	180	1.7	220	1.6	194	1.6
Endopeptidase activity	161	1.5	199	1.5	0	0.0
Protein tyrosine kinase activity	154	1.4	195	1.4	181	1.5
Phosphatase activity	162	1.5	188	1.4	192	1.6
Ion binding	0	0	2067	15.3	1870	15.7
Cation binding	0	0	2061	15.3	1864	15.6

regulation of flower development accounted only in generative stage (Table 5). Analysis of KEGG metabolic pathway assignments revealed that our contig catalog covers all major plant metabolic pathways with a certain dominance of plant hormone biosynthesis and many alkaloid biosynthesis, indicative that those pathways, seemingly paired in response to reproductive developmental process (Table 6).

Table 5: Biological processes of flower development

	Vegetative		Both		Generative	
Biological processes of flower development categories	Value	%	Value	%	Value	%
Positive regulation of developmental process	26	0.2	0	0.0	0	0.0
Regulation of meristem development	31	0.3	0	0.0	0	0.0
Meristem development	54	0.5	63	0.5	0	0.0
Shoot development	155	1.5	169	13.0	0	0.0
Shoot system development	156	1.5	170	13.0	0	0.0
Reproductive developmental process	436	4.1	508	3.8	480	4.0
Reproductive structure development	400	3.8	463	3.4	435	3.6
Regulation of flower development	58	0.5	0	0.0	57	0.5
Flower development	123	1.2	134	1.0	125	1.0
Positive regulation of flower development	19	0.2	0	0.0	0	0.0
Gametophyte development	107	1.0	129	1.0	117	1.0
Pollen development	75	0.7	89	0.7	82	0.7
Pollen tube development	0	0.0	0	0.0	49	0.4
Tube development	0	0.0	0	0.0	49	0.4
Ovule development	0	0.0	25	0.2	0	0.0
Negative regulation of flower development	0	0.0	27	0.2	24	0.2
Embryonic meristem development	13	0.1	15	01.0	0	0.0
Embryo sac development	42	0.4	50	0.4	0	0.0
Fruit development	254	2.4	299	2.2	276	2.3
Seed development	242	2.3	288	2.1	268	2.2

Table 6: KEGG pathway

	Vegetative		Both		Generative	
KEGG pathway categories	Value	%	Value	%	Value	%
Biosynthesis of plant hormones	201	1.9	230	1.7	217	1.8
Biosynthesis of terpenoids and steroids	130	1.2	147	1.1	140	1.2
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	122	1.1	128	0.9	127	1.1
Biosynthesis of alkaloids derived from shikimate pathway	119	1.1	133	1.0	126	1.1
Biosynthesis of alkaloids derived from histidine and purine	116	1.1	131	1.0	127	1.1
Biosynthesis of alkaloids derived from terpenoid and polyketide	113	1.1	121	0.9	115	1.0
Spliceosome	80	0.3	86	0.6	84	0.7
Purine metabolism	73	0.7	0	0.0	82	0.7
Glycolysis/Gluconeogenesis	66	0.6	65	0.5	66	0.6
Carbon fixation in photosynthetic organisms	56	0.5	58	0.4	0	0.0
Pyruvate metabolism	52	0.5	51	0.4	50	0.4
Pyrimidine metabolism	52	0.5	0	0.0	0	0.0
Citrate cycle (TCA cycle)	51	0.5	52	0.4	50	0.4
Amino sugar and nucleotide sugar metabolism	50	0.5	0	0.0	0	0.0
Proteasome	44	0.4	46	0.3	49	0.4
Aminoacyl-tRNA biosynthesis	37	0.3	42	0.3	43	0.4
Arginine and proline metabolism	36	0.3	40	0.3	40	0.3
Glycine, serine and threonine metabolism	32	0.3	37	0.3	40	0.3
Alanine, aspartate and glutamate metabolism	31	0.3	36	0.3	39	0.3
DNA replication	29	0.3	33	0.2	32	0.3
Fatty acid metabolism	26	0.2	29	0.2	29	0.2
Ascorbate and aldarate metabolism	25	0.2	0	0.0	0	0.0
N-Glycan biosynthesis	24	0.2	0	0.0	27	0.2
Glyoxylate and dicarboxylate metabolism	24	0.2	0	0.0	0	0.0
Alpha-Linolenic acid metabolism	22	0.2	24	0.2	0	0.0
Biosynthesis of unsaturated fatty acids	21	0.2	23	0.2	0	0.0
Butanoate metabolism	19	0.2	0	0.0	0	0.0
Propanoate metabolism	16	0.2	17	0.1	17	0.1
One carbon pool by folate	15	0.1	0	0.0	0	0.0
Lysine degradation	13	0.1	15	0.1	0	0.0
Nucleotide excision repair	0	0.0	41	0.3	42	0.4
Nitrogen metabolism	0	0.0	0	0.0	34	0.3
Inositol phosphate metabolism	0	0.0	0	0.0	30	0.3
Mismatch repair	0	0.0	25	0.2	25	0.2
Selenoamino acid metabolism	0	0.0	0	0.0	24	0.2
Steroid biosynthesis	0	0.0	23	0.2	23	0.2

DISCUSSION

Vegetative and generative stage of teak transcriptome sequencing and annotation: Next Generation Sequencing (NGS) technology during the last decade have dramatically impacted genome sequencing and transcriptome analysis (Fan *et al*., 2013; Fu *et al*., 2013; Fox *et al*., 2013). This technique could be used for model plants with known genome sequences and also has been successfully used to analyze the transcriptome in non-model plants (Collins *et al*., 2008; Logacheva *et al*., 2011; Li *et al*., 2012). However, this technique requires cDNA cloning and individual RNA preparations for each sample stages, is time consuming and very costly. Pyrosequencing like 454 and illumina plat form introduced recently constitutes a better alternative for transcriptomics (Lulin *et al*., 2012). The high number of reads generated per run together with the low sequencing error rate in the contigs obtained makes it a good tool to deeply sequence the transcriptome of plants. This approach has been used successfully for analyzing the transcriptomes of maize and *Arabidopsis* and have applied it to the non-model tree species *Castanea dentata* and *C. mollissima* (Collins *et al*., 2008; Barakat *et al*., 2009, 2012).

Tectona grandis, also known as teak is a tropical deciduous tree native from moist tropical forests of Asia (Palupi *et al*., 2010; Khanduri, 2012). *Tectona grandis* is lamiaceae family is known for the quality of its wood (Lyngdoh *et al*., 2010). Despite its ecological and increasing economic importance, very little is known about the biology of this species at the genetic, molecular and biochemical levels (Borges *et al*., 2008). Genomic tools have recently increased the numbers and volume of genomic resources for several crop plants and trees and have contributed to enlarge our knowledge on basic aspects of plant biology; furthermore, they represent valuable sources of candidate genes and new molecular markers to assist improvement programs (Collins *et al*., 2008; Eveland *et al*., 2010; Huang *et al*., 2012). Biological sequences reported to date in public databases and belonging to teak do not exceed 20 entries: This very narrow availability of genetic information is the main problem to initiate improvement programs in *T. grandis*.

Our study generated 3,778,316 and 3,701,878 reads and 87,365 high quality contigs from vegetative and generative teak transcriptomes, respectively. A fraction of teak contigs could be annotated using the *S. lycopersicum* proteome than *Arabidopsis* or the others (Table 2). Most of the genes in teak hits to the *S. lycopersicum* proteome encoded proteins annotated. Those genes could be homology to *S. lycopersicum* using the Blast algorithm. Over 80% of the teak reads could be annotated using the *S. lycopersicum* proteome. By taking into consideration the sequences that have homologies in the *S. lycopersicum* proteome, assuming that the two samples of teak have a similar gene number as *S. lycopersicum*. cDNA sequences generated from both teak samples cover various biological processes and molecular functions indicating that the technique constitutes a powerful tool for sequencing the transcriptome of non-model species. These results confirm that pyrosequencing constitutes a powerful tool for transcriptome characterization and gene discovery.

Transcriptome comparison between vegetative and generative tissues from *Tectona grandis*: Gene Ontology (GO) annotation analyses showed that, overall vegetative and generative tissues from teak present a similar transcriptome. Gene function categories associated with response to abiotic stimuli and metabolic process are highly represented in both transcriptomes. The second most highly represented category includes genes involved in reproductive development. The category represented the most is composed of genes associated with various reproductive processes as previously described in other systems such as Gerbera, Fagopyrum and Prunus.

Detailed analysis of illumina sequences from both vegetative and generative tissue showed that the tagged genes included a large number associated with response to abiotic stimuli, metabolic process and reproductive development.

These include genes involved in regulation of development, meristem development and reproductive development genes. Comparison of flowering developmental genes highly expressed in the vegetative and generative tissues of teak showed that a fraction were either preferentially expressed in vegetative or in generative stage. Genes of positive regulation of developmental process, regulation of meristem development, meristem development, shoot development, shoot system development, positive regulation of flower development, embryonic meristem development and embryo sac development represented the functional category with the largest number of reads in vegetative stage. Genes of reproductive developmental process, reproductive structure development, regulation of flower development, flower development, gametophyte development, pollen development, pollen tube development, tube development, negative regulation of flower development, fruit development and seed development complied the largest number of reads in generative stage.

Positive regulation of developmental process, regulation of meristem development and positive regulation of flower development genes category expressed only in vegetative stage. Pollen tube development, tube development and negative regulation of flower development genes category founded only in generative stages.

These different suggest that these tissues may modulate the expression of flowering development genes in transition vegetative to generative in teak. The important thing after this step is select the candidate genes involved in regulation of teak vegetative to generative transition. Overall, this study allowed us to conclude that teak tree responds to abiotic stimuli before entering to flowering developmental stage. The different category of flowering developmental processes between vegetative and generative stage showed us the regulation of transitional vegetative to generative.

CONCLUSION

In conclusion, this study allowed us to (1) Obtain 87,365 contigs from vegetative and generative tissue of teak, (2) Transcriptomes of teak could be annotated using the *S. lycopersicum* proteome according to BLAST result, (3) Compare the transcriptomes of vegetative and generative tissues of teak in flowering developmental stage and (4) Identify potential biological processes involved in teak flowering developmental stage.

REFERENCES

Agusti, J., R. Lichtenberger, M. Schwarz, L. Nehlin and T. Greb, 2011. Characterization of transcriptome remodeling during cambium formation identifies *MOL1* and *RUL1* as opposing regulators of secondary growth. PLoS Gen., Vol. 7.

Angeloni, F., C.A.M. Wagemaker, M.S.M. Jetten, H.J.M. Op den Camp and E.M. Janssen Megens *et al*., 2011. *De novo* transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next generation sequencing techniques. Mol. Ecol. Resour., 11: 662-674.

Barakat, A., D.S. DiLoreto, Y. Zhang, C. Smith and K. Baier *et al*., 2009. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. BMC Plant Biol., Vol. 9. 10.1186/1471-2229-9-51

Barakat, A., M. Staton, C.H. Cheng, J. Park and N.B. Yassin *et al*., 2012. Chestnut resistance to the blight disease: Insights from transcriptome analysis. BMC Plant Biol., Vol. 12. 10.1186/1471-2229-12-38

Borges, F., G. Gomes, R. Gardner, N. Moreno, S. McCormick, J.A. Feijo and J.D. Becker, 2008. Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiol., 148: 1168-1181.

Chang, L., J. Chen, Y. Xiao and Y. Xia, 2013. De novo characterization of *Lycoris sprengeri* transcriptome using Illumina GA II. Afr. J. Biotechnol., 10: 12147-12155.

Collins, L.J., P.J. Biggs, C. Voelckel and S. Joly, 2008. An approach to transcriptome analysis of non-model organisms using short-read sequences. Genome Inform., 21: 3-14.

De Gyves, E.M., J.I. Royani and E. Rugini, 2007. Efficient method of micropropagation and in vitro rooting of teak (*Tectona grandis* L.) focusing on large-scale industrial plantations. Ann. For. Sci., 64: 73-78.

Eveland, A.L., N. Satoh-Nagasawa, A. Goldshmidt, S. Meyer and M. Beatty *et al*., 2010. Digital gene expression signatures for maize development. Plant Physiol., 154: 1024-1039.

Fan, H., Y. Xiao, Y. Yang, W. Xia and A.S. Mason *et al*., 2013. RNA-Seq analysis of *Cocos nucifera*: Transcriptome sequencing and *de novo* assembly for subsequent functional genomics approaches. PloS One, Vol. 8. 10.1371/journal.pone.0059997

Fox, S.E., J. Preece, J.A. Kimbrel, G.L. Marchini and A. Sage *et al*., 2013. Sequencing and *de novo* transcriptome assembly of *Brachypodium sylvaticum* (Poaceae). Applic. Plant Sci. 10.3732/apps.1200011

Fu, N., Q. Wang and H.L. Shen, 2013. *De novo* assembly, gene annotation and marker development using Illumina paired-end transcriptome sequences in celery (*Apium graveolens* L.). PloS One, Vol. 8. 10.1371/journal.pone.0057686

Huang, D.W., B.T. Sherman, Q. Tan, J. Kir and D. Liu *et al*., 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res., 35: W169-W175.

Huang, H.H., L.L. Xu, Z.K. Tong, E.P. Lin, Q.P. Liu, L.J. Cheng and M.Y. Zhu, 2012. *De novo* characterization of the Chinese Fir (*Cunninghamia lanceolate*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. BMC Genomics. 10.1186/1471-2164-13-648

Jiao, X., B.T. Sherman, D.W. Huang, R. Stephens, M.W. Baseler, H.C. Lane and R.A. Lempicki, 2012. DAVID-WS: A stateful web service to facilitate gene/protein list analysis. Bioinformatics, 28: 1805-1806.

Khanduri, V.P., 2012. Annual variation in floral phenology and pollen production in a 25-year-old plantation of *Tectona grandis*. Nordic J. Bot., 30: 82-89.

Li, D., Z. Deng, B. Qin, X. Liu and Z. Men, 2012. *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics, Vol. 13. 10.1186/1471-2164-13-192

Logacheva, M.D., A.S. Kasianov, D.V. Vinogradov, T.H. Samigullin, M.S. Gelfand, V.J. Makeev and A.A. Penin, 2011. *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). BMC Genomics, Vol. 12. 10.1186/1471-2164-12-30

Lulin, H., Y. Xiao, S. Pei, T. Wen and H. Shangqin, 2012. The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. PloS One, Vol. 7. 10.1371/journal.pone.0038653.g005

Lyngdoh, N., G. Joshi, G. Ravikanth, R.U. Shaanker and R. Vasudeva, 2010. Influence of levels of genetic diversity on fruit quality in teak (*Tectona grandis* Lf). Curr. Sci., 99: 639-644.

Orwa, C., A. Mutua, R. Kindt, R. Jamnadass and A. Simons, 2009. Agroforestree database: A tree reference and selection guide, version 4.0. http://www.worldagroforestry.org

Palupi, E.R., J.N. Owens, S. Sadjad, S. Sudarsono and D.D. Solihin, 2010. The importance of fruit set, fruit abortion and pollination success in fruit production of teak (*Tectona grandis*). Can. J. For. Res., 40: 2204-2214.

Rosero, C., X. Argout, M. Ruiz and W. Teran, 2011. A drought stress transcriptome profiling as the first genomic resource for white teak-Gamhar-(*Gmelina arborea* Roxb) and related species. BMC Proc., Vol. 5. 10.1186/1753-6561-5-S7-P178

Rosli, H.M., N.M. Saleh, M.M. Noor, M. Norwati, B. Norlia and A. Norwati, 2009. Analysis of expressed sequence tags derived from inflorescence shoot of *Tectona grandis* (teak). Afr. J. Biotechnol., 8: 5633-5640.

Widiyanto, S.N., A. Sukmawan, N. Haro and H. Rahmania, 2009. Transient expression of β-glucuronidase reporter gene in Agrobacterium-inoculated shoots of various teak clones. Afr. J. Biotechnol., 8: 2143-2150.