International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 9 No. 3 Nov. 2014, pp. 1416-1427

http://www.ijias.issr-journals.org/

Transcriptomes of Teak (*Tectona grandis*, L.f) in Vegetative to Generative Transition Stage Development

Diky S. Diningrat[1], Srinanan M. Widiyanto[1], Adi Pancoro[1], Iriawati[1], Donghwan Shim[2], Bhargavi Panchangam[2], Nicole Zembower[2], and John E. Carlson[2]

[1]School of Life Sciences and Technology, Institut Teknologi Bandung, Bandung-40132, Indonesia

[2]Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA 16802, Indonesia

ABSTRACT: Teak (*Tectona grandis* L.f) is highly famous woody plant species for the quality and durability. Teak has two main problems on long reproductive cycle and produces low seeds. Both problems are basically related to mechanism of flower development. Hence, the determination of the genetic pathways and specific genes involved in teak flowering development could be beneficial for teak productivity improvement. 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 teak bud tissues. Libraries were sequenced using Illumina MiSeq technology which generated 3,778,316 in vegetative and 3,701,878 in generative paired-end reads sequences. The sequences were combined QC tested, trimmed, and de novo assembled conducted using CLCGenomics Workbench. The sequence reads assembled de novo into 87,365 transcript contigs consisting of 42,435,728 bases with N50 of 498bp. 64,961 (74.36%) of assembled contigs exhibited similarity BLASTN to *Arabidopsis thaliana* database. The assembled contigs were annotated through high stringency BLASTX analysis to proteome of *A. thaliana*. 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 transition stage. The functionally protein datasets characterized by gene ontology (GO) annotation and KEGG metabolic pathways assignments for the result of DEG analysis. This study allowed us compare the transcriptomes of vegetative and generative tissues of teak in flowering developmental stage, and identify potential biological processes involved in teak flowering developmental stage.

KEYWORDS: de novo assembly, teak flowering, NGS, DEG, Gene Ontology, EST, non model plant.

1 INTRODUCTION

Teak (*Tectona grandis*, L. f) is a tropical tree species distributed naturally in countries including India, Myanmar, Thailand, Myanmar and Indonesia ([18],[17]). 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 ([18]). 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, ([18],[19], [20]). 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 ([7]). Then, CLC bio bioinformatics technology tool was used to

Corresponding Author: Diky S. Diningrat

perform a de novo assembly and annotation without prior genome information ([1]). 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 ([15]). 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.

2 METHODS

2.1 TEAK TISSUES MATERIALS AND RNA ISOLATION

Vegetative and generative stage shoot tips of teak were collected from a 12 year old teak D plant in Institute Technology Bandung, Indonesia for RNA isolation.The following D-VS tissues were sampled from vegetative apical shoots. D-LB2 tissues were sampled from lateral (nodal) floral-Buds 2nd of generative stage shoots. Both of teak tissue samples were frozen in liquid nitrogen immediately upon collection and putted in Dry Shipper for shipping from ITB-Indonesia to Pennsylvania State University (PSU)-USA. Samples were immediately frozen at -80 upon arrival at PSU until use. Total RNA was obtained using the 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 ([3]). 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 ng/ µl and 206 ng/ µl for DVS and DLB2 sample, respectively. The integrity of RNA was assessed with the Agilent 6000 RNA Nano Chip Kit on 2100 Bioanalyzer (AgilentTechnologies).

2.2 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 (Illumina, USA) ([1], [7], [14]). The resulting library was sequenced at Penn State University using Illumina MiSeq™ 2000 (Illumina Inc., USA).

2.3 TRANSCRIPT ASSEMBLY

Two sequence data in FASTQ files computed with CLCbio for transcript assembly strategy ([1]). 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.

2.4 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 *Arahidopsis thaliana* and *Populus tricocharpa* (www.phytozome.com) using CLCbio workbench. Top hit species results use for homology based annotations of Teak ([2], [3]).

2.5 DEGSEQ ANALYSIS

Comparison of digitally gene expression (DEGseq) between vegetative and generative tissues was done using RNAseq analysis software test developed by CLCbio genomic work bench. DEGseq analysis was used to identify flowering development genes in transcript abundance because it integrates several statistical methods ([2], [3]). 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.

2.6 GO ANALYSIS

Further assessment of the quality of the de novo assembly was carried out as follows.We compared the dept hand the length of contig coverage with reference to orthologous genes in *A. thaliana*, by plotting the ratio of contig length to *A. thaliana* orthologue coding region length against coverage depth. Orthologous genes were retrieved performing a local BLASTX alignment (e-value< 10^{-6}) usingCLCbio workbench with the TAIR9 *A. thaliana* database 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/ ([10], [11], [12]). 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 ([12]), 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* 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.

3 RESULT

3.1 ILLUMINA SEQUENCING OUTPUT STATISTICS AND READS ASSEMBLY

T. grandis vegetative and generative cDNA library was constructed from a pool of RNA isolated from Vegetative and generative bud tissues of D 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 (Table1). The average length of the reads was 151 nucleotides (Figure1). 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 (Table2). Those contigs having an average length of 486 nt, 225nt for minimum length and 4,361 nt for maximum length (Figure2).

Table 1. Summary statisitcs of cDNA library.

Library		
Vegetative Shoot (D-VS)	:	3,701,878 sequences in pairs
Generative Shoot (D-LB2)	:	3,778,316 sequences in pairs

Table2. Summary statisitcs of sequencing and de novo assembly results

	Value
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

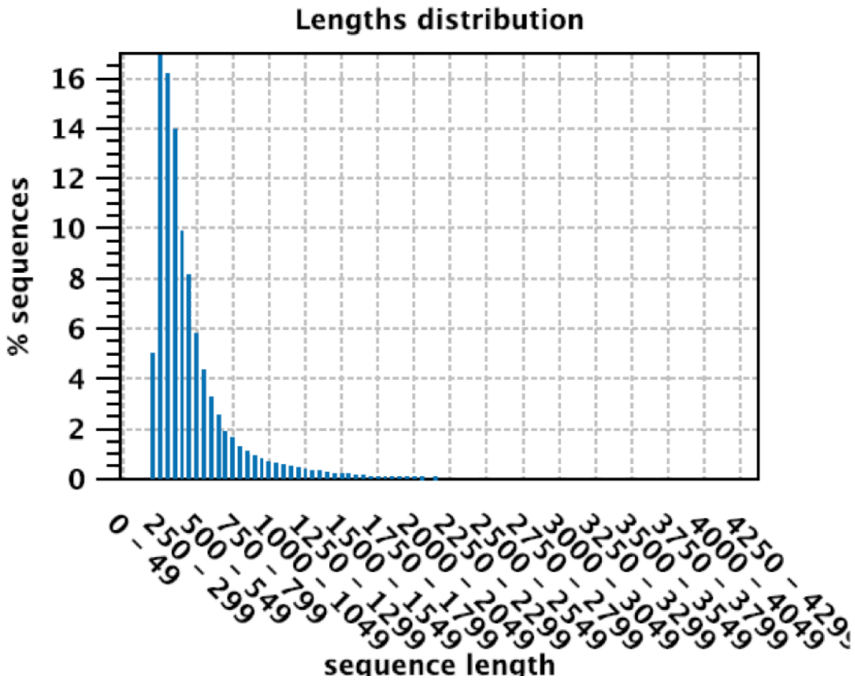


Fig. 1. Paired reads distance distribution

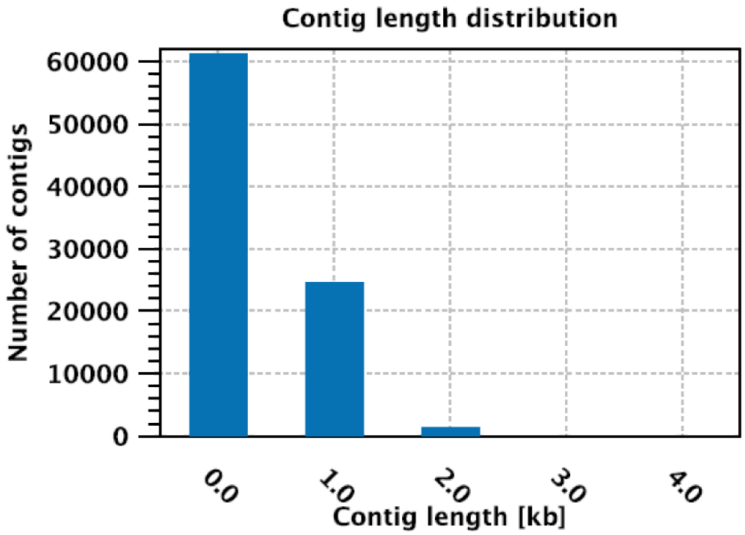


Fig. 2. Histogram of the frequency of different contigs sizes in transcriptome assemblies of Teak samples.

3.2 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 to the model species (Table 3) were: *Arabidopsis thaliana* (74.36%) and *Populus trichocarpa* (8.56 %). *A. thaliana* was the highest blast hits model species for teak. This is a remarkable result when considering the current state of functional annotation of teak to the *A. thaliana* proteome database (www.phytozome.com).

Table 3. Number of Blast Hit Top Species for Homology Based Annotations of Teak Contigs

Species	Number of Hits	%
Arabidopsis thaliana	64,961	74.36%
Populus trichocarpa	7,479	8.56 %

3.3 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 (Figure 3).

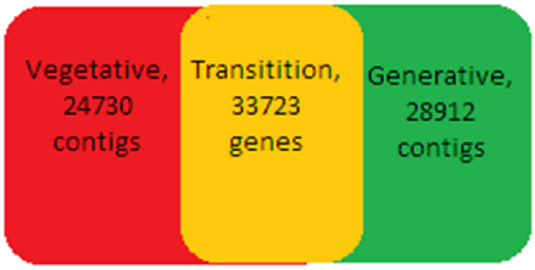


Fig. 3. Contigs Disribution Result of DEG Analysis

Detailed comparison of the gene ontology (GO) transcriptomes in vegetative stage, generative stage and transi showed the different percentage of biological processes, cellular component and molecular function (Figure 4). Figure 5 showed top 25 of biological process, cellular component and molecular function that occurred in the tissue samples. In the category of biological process, regulation of transcription (vegetative stage 0%; generative 0%; transition stage 8%), phosphorus metabolic process (vegetative stage 6.4%; generative 6.2%; transition stage 5.9 %), phosphate metabolic process (vegetative stage 6.4%; generative 6.2%; transition stage 5.9 %), phosphorylation (vegetative stage 5.9%; generative 5.8%; transition stage 5.5%), protein amino acid phosphorylation (vegetative stage 5.4%; generative 5.4%; transition stage 5.3%), response to abiotic stimulus (vegetative stage 6.1%; generative 5.3%; transition stage 5.2%), transcription (vegetative stage 0%; generative 0%; transition stage 5.2%), proteolysis (vegetative stage 4.6%; generative 0%; transition stage 4.1%) of the total, respectively. According to flowering development biological processes, there are post-embryonic development (vegetative stage 4.5%; generative 3.8%; transition stage 3.9%), reproductive developmental process (vegetative stage 4.2%; generative 3.6%; transition stage 3.6%), reproductive structure development (vegetative stage 3.8%; generative 3.3%; transition stage 3.3%) comprised part of the top ten largest proportion.

In the category of cellular components, plastid comprised (vegetative stage 15.3%; generative 13.6%; transition stage 13.9%), chloroplast was (vegetative stage 15.3%; generative 13.6%; transition stage 13.9%), and intrinsic to membrane (vegetative stage 11.6%; generative 10.8%; transition stage 11%) these three subgroups were dominant over the others. In the category molecular function, sequences with the functions of nucleotide binding, ion binding and cation binding comprised (18.4% in vegetative stage, 16.7% in generative stage and 17.1% in transition stage), (0% in vegetative stage, 14.9% in generative stage and 15.3% in transition stage) and (0% in vegetative stage, 14.8% in generative stage and 15.2% in transition stage) of the total (Table 4).

On the other hand we also identified the other biological processes of flower development. We compare 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 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 of alkaloids derived from terpenoid and polyketide, alkaloids derived from histidine and purine, alkaloids derived from shikimate pathway, terpenoids and steroids, and plant hormones biosynthesis, indicative that those pathways, seemingly paired in response to reproductive developmental process (Table 6).

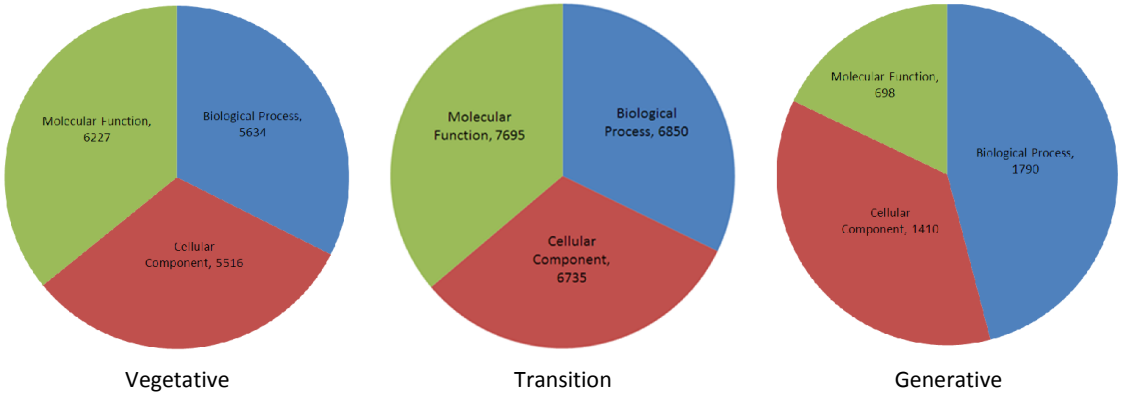
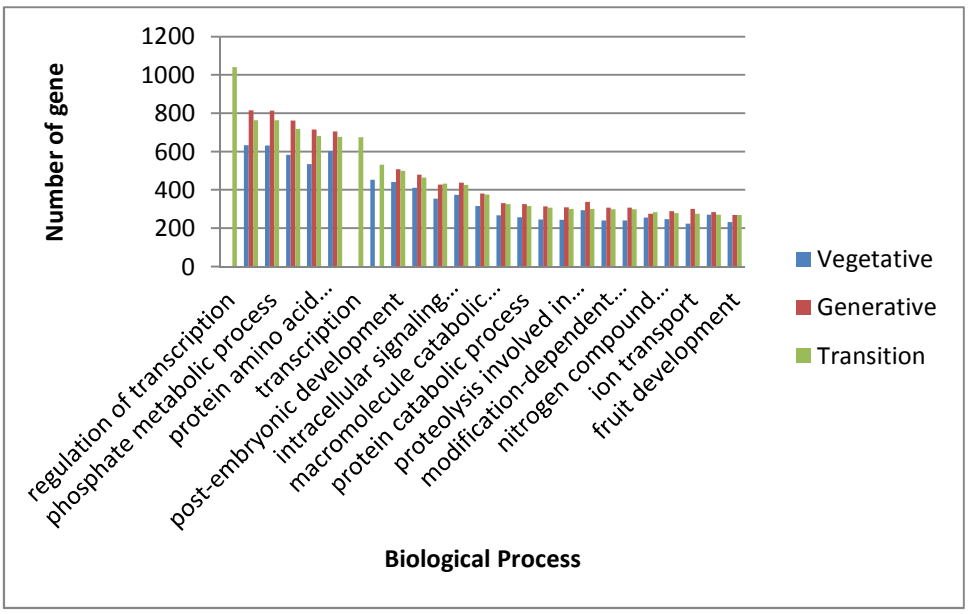
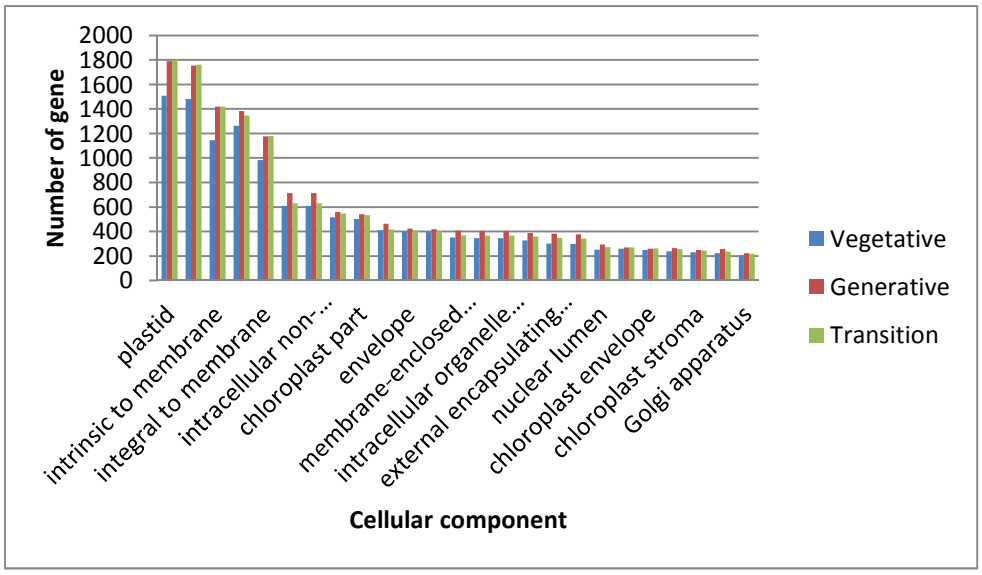


Fig. 4. Chart of GO Categories of Teak





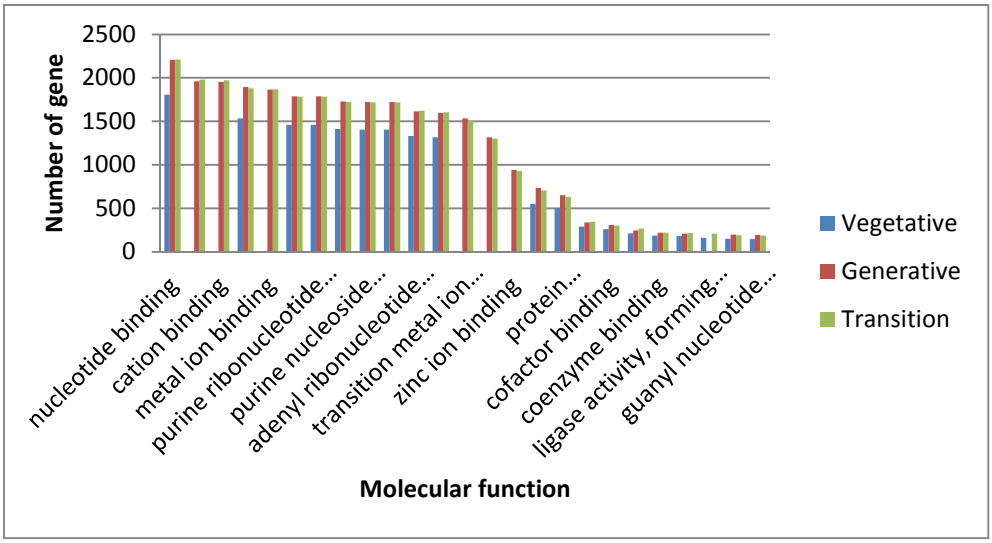


Fig. 5. Histogram of GO Classification of Teak

Table 4. GO classification of Teak

Biological Process Categories	Vegetative		Generative		Transition	
	Value	%	Value	%	Value	%
regulation of transcription	0	0	0	0	1040	8
phosphorus metabolic process	633	6.4	815	6.2	764	5.9
phosphate metabolic process	632	6.4	814	6.2	763	5.9
phosphorylation	583	5.9	761	5.8	718	5.5
protein amino acid phosphorylation	535	5.4	715	5.4	681	5.3
response to abiotic stimulus	602	6.1	704	5.3	677	5.2
transcription	0	0	0	0	674	5.2
proteolysis	452	4.6	0	0	531	4.1
post-embryonic development	440	4.5	507	3.8	499	3.9
reproductive developmental process	411	4.2	480	3.6	465	3.6
intracellular signaling cascade	354	3.6	427	3.2	433	3.3
reproductive structure development	374	3.8	437	3.3	425	3.3
macromolecule catabolic process	316	3.2	381	2.9	375	2.9
cellular macromolecule catabolic process	267	2.7	331	2.5	325	2.5
protein catabolic process	257	2.6	325	2.5	316	2.4
cellular protein catabolic process	246	2.5	313	2.4	307	2.4
proteolysis involved in cellular protein catabolic process	243	2.5	309	2.3	301	2.3
response to inorganic substance	293	3	337	2.6	300	2.3
modification-dependent protein catabolic process	240	2.4	307	2.3	298	2.3
modification-dependent macromolecule catabolic process	240	2.4	307	2.3	298	2.3
nitrogen compound biosynthetic process	256	2.6	276	2.1	284	2.2
response to radiation	247	2.5	289	2.2	279	2.2
ion transport	224	2.3	300	2.3	275	2.1
protein localization	271	2.8	283	2.1	271	2.1
fruit development	232	2.4	268	2	268	2.1
regulation of transcription	0	0	0	0	1040	8

Cellular Component Categories	Vegetative		Generative		Transition	
	Value	%	Value	%	Value	%
plastid	1508	15.3	1789	13.6	1797	13.9
chloroplast	1482	15.1	1753	13.3	1761	13.6
intrinsic to membrane	1146	11.6	1420	10.8	1420	11
plasma membrane	1263	12.8	1383	10.5	1345	10.4
integral to membrane	984	10	1176	8.9	1180	9.1
non-membrane-bounded organelle	609	6.2	713	5.4	630	4.9
intracellular non-membrane-bounded organelle	609	6.2	713	5.4	630	4.9
plastid part	516	5.2	559	4.2	548	4.2
chloroplast part	503	5.1	542	4.1	535	4.1
cytosol	411	4.2	464	3.5	415	3.2
envelope	398	4	424	3.2	408	3.2
organelle envelope	396	4	420	3.2	404	3.1
membrane-enclosed lumen	351	3.6	408	3.1	368	2.8
organelle lumen	347	3.5	406	3.1	367	2.8
intracellular organelle lumen	347	3.5	406	3.1	367	2.8
vacuole	328	3.3	387	2.9	359	2.8
external encapsulating structure	301	3.1	382	2.9	348	2.7
cell wall	298	3	378	2.9	342	2.6
nuclear lumen	253	2.6	293	2.2	272	2.1
plastid envelope	259	2.6	270	2	269	2.1
chloroplast envelope	248	2.5	259	2	261	2
plastid stroma	239	2.4	264	2	257	2
chloroplast stroma	231	2.3	250	1.9	245	1.9
endoplasmic reticulum	224	2.3	257	1.9	237	1.8
Golgi apparatus	205	2.1	224	1.7	219	1.7
Molecular Function Categories	**Vegetative**		**Generative**		**Transition**	
	Value	**%**	**Value**	**%**	**Value**	**%**
nucleotide binding	1806	18.4	2207	16.7	2210	17.1
ion binding	0	0	1961	14.9	1978	15.3
cation binding	0	0	1954	14.8	1971	15.2
purine nucleotide binding	1535	15.6	1895	14.4	1881	14.5
metal ion binding	0	0	1865	14.1	1870	14.5
ribonucleotide binding	1459	14.8	1787	13.5	1783	13.8
purine ribonucleotide binding	1459	14.8	1787	13.5	1783	13.8
nucleoside binding	1411	14.3	1728	13.1	1722	13.3
purine nucleoside binding	1406	14.3	1721	13	1716	13.3
adenyl nucleotide binding	1406	14.3	1721	13	1716	13.3
adenyl ribonucleotide binding	1333	13.5	1616	12.2	1621	12.5
ATP binding	1315	13.4	1598	12.1	1603	12.4
transition metal ion binding	0	0	1533	11.6	1503	11.6
DNA binding	0	0	1318	10	1303	10.1
zinc ion binding	0	0	940	7.1	931	7.2
protein kinase activity	552	5.6	735	5.6	704	5.4
protein serine/threonine kinase activity	499	5.1	651	4.9	632	4.9
ATPase activity	290	2.9	337	2.6	345	2.7
cofactor binding	261	2.7	308	2.3	301	2.3
ATPase activity, coupled	213	2.2	247	1.9	268	2.1

coenzyme binding	186	1.9	221	1.7	220	1.7
magnesium ion binding	182	1.8	207	1.6	214	1.7
ligase activity, forming carbon-nitrogen bonds	160	1.6	0	0	207	1.6
protein tyrosine kinase activity	148	1.5	199	1.5	190	1.5
guanyl nucleotide binding	147	1.5	192	1.5	185	1.4
guanyl ribonucleotide binding	144	1.5	189	1.4	183	1.4

Tabel 5. Biological processes of flower development

Biological processes of flower development categories	Vegetative		Generative		Transition	
	Value	%	Value	%	Value	%
post-embryonic development	440	4.5	507	3.8	499	3.9
reproductive developmental process	411	4.2	480	3.6	465	3.6
reproductive structure development	374	3.8	437	3.3	425	3.3
fruit development	232	2.4	268	2	268	2.1
seed development	221	2.2	257	1.9	257	2
embryonic development ending in seed dormancy	201	2	222	1.7	232	1.8
shoot system development	140	1.4	166	1.3	164	1.3
shoot development	139	1.4	164	1.2	162	1.3
root development	108	1.1	124	0.9	129	1
flower development	117	1.2	138	1	126	1
phyllome development	107	1.1	128	1	120	0.9
gametophyte development	99	1	0	0	115	0.9
leaf development	97	1	115	0.9	110	0.9
shoot morphogenesis	82	0.8	98	0.7	103	0.8
developmental growth	87	0.9	113	0.9	103	0.8
developmental growth involved in morphogenesis	0	0	98	0.7	88	0.7
regulation of post-embryonic development	80	0.8	91	0.7	85	0.7
ectoderm development	69	0.7	78	0.6	79	0.6
pollination	64	0.7	83	0.6	76	0.6
pollen development	73	0.7	0	0	74	0.6
embryo sac development	0	0	0	0	53	0.4
meristem development	55	0.6	62	0.5	53	0.4
post-embryonic morphogenesis	42	0.4	0	0	49	0.4
developmental cell growth	38	0.4	47	0.4	45	0.3
hair cell differentiation	34	0.3	39	0.3	38	0.3
developmental maturation	32	0.3	0	0	37	0.3
root epidermal cell differentiation	29	0.3	30	0.2	36	0.3
negative regulation of post-embryonic development	0	0	0	0	35	0.3
photomorphogenesis	29	0.3	0	0	29	0.2
pollen-pistil interaction	0	0	0	0	27	0.2
organ formation	20	0.2	24	0.2	24	0.2
meristem structural organization	25	0.3	0	0	24	0.2
pollen germination	23	0.2	23	0.2	21	0.2
meristem initiation	14	0.1	0	0	16	0.1
regulation of cell morphogenesis	14	0.1	16	0.1	16	0.1
embryonic meristem development	15	0.2	0	0	15	0.1
negative regulation of photomorphogenesis	0	0	0	0	8	0.1

Tabel 6. KEGG Pathway

KEGG Pathway Categories	Vegetative		Generative		Transition	
	Value	%	Value	%	Value	%
Histidine metabolism	0	0	0	0	12	0.1
Glycerolipid metabolism	0	0	0	0	15	0.1
Homologous recombination	0	0	0	0	20	0.2
Glycerophospholipid metabolism	0	0	0	0	23	0.2
Inositol phosphate metabolism	0	0	0	0	26	0.2
Aminoacyl-tRNA biosynthesis	37	0.4	35	0.3	33	0.3
Nucleotide excision repair	0	0	0	0	37	0.3
Pentose phosphate pathway	0	0	0	0	37	0.3
Citrate cycle (TCA cycle)	45	0.5	43	0.3	43	0.3
Endocytosis	0	0	0	0	43	0.3
Pyruvate metabolism	50	0.5	51	0.4	46	0.4
Glycolysis / Gluconeogenesis	60	0.6	67	0.5	55	0.4
Carbon fixation in photosynthetic organisms	50	0.5	52	0.4	56	0.4
Spliceosome	69	0.7	0	0	69	0.5
Biosynthesis of alkaloids derived from terpenoid and polyketide	105	1.1	110	0.8	95	0.7
Biosynthesis of alkaloids derived from histidine and purine	109	1.1	113	0.9	108	0.8
Biosynthesis of alkaloids derived from shikimate pathway	112	1.1	119	0.9	110	0.9
Biosynthesis of terpenoids and steroids	123	1.2	131	1	119	0.9
Biosynthesis of plant hormones	177	1.8	197	1.5	191	1.5

4 DISCUSSION

4.1 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 ([7], [8], [9]). 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 ([5], [14], [15]). 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 ([16]). 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* ([4]) and have applied it to the non-model tree species *Castanea dentata* and *C. mollissima* ([2], [3], [5])

Tectona grandis, also known as teak is a tropical deciduous tree native from moist tropical forests of Asia ([18], [13]). *T. grandis* is lamiaceae family is known for the quality of its wood ([17]). Despite its ecological and increasing economic importance, very little is known about the biology of this species at the genetic, molecular and biochemical levels ([4]). 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 ([5], [6], [10]). 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 *Arabidopsis* proteome than *Poplar* (Table2). Most of the genes in teak hits to the *Arabidopsis* proteome encoded proteins annotated. Those genes could be

homology to *Arabidopsis* using the Blast algorithm. Over 80% of the teak reads could be annotated using the *Arabidopsis* proteome. By taking into consideration the sequences that have homologies in the *Arabidopsis* proteome, assuming that the two samples of teak have a similar gene number as *Arabidopsis*. 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.

4.2 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 illumine 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, 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, 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.

5 CONCLUSION

In conclusion, this study allowed us to (i) Obtain 87,365 contigs from vegetative and generative tissue of teak, (ii) Transciptomes of teak could be annotated using the *Arabidopsis* proteome according to Blast result, (iii) Compare the transcriptomes of vegetative and generative tissues of teak in flowering developmental stage, and (iv) Identify potential biological processes involved in teak flowering developmental stage.

ACKNOWLEDGMENT

The Schatz Center for Tree Molecular Genetics, Department of Ecosystem Science and Management, Pennsylvania State University University Park, PA 16802, USA;

The School of Life Sciences and Technology, Institut Teknologi Bandung, Bandung-40132, Indonesia.

REFERENCES

[1] Angeloni, F., C. A. M. Wagemaker, M. S. M. Jetten, H. J. M. O. D. Camp, E. M. J. Megens, K. J. Francoijs, H. G. Stunnenberg, & N. J . Ouborg. 2011. De Novo Transcriptome Characterization and Development of Genomic Tools for Scabiosa columbaria L. Using Next-Generation Sequencing Techniques. Molecular Ecology Resources.11:662-674.

[2] Barakat, A., D. S. DiLoreto, Y. Zhang, C. Smith, K. Baier, W. A. Powell, N. Wheeler, R. Sederoff & J. E. Carlson. 2009. Comparison of TheTranscriptomes of American Chestnut (*Castanea dentata*) and Chinese Chestnut (*Castanea mollissima*) in Response to The Chestnut Blight Infection. BMC Plant Biology. 9(51): 1471-2229 http://www.biomedcentral.com/1471-2229/9/51

[3] Barakat, A., M. Staton, C. H. Cheng, J. Park, N. B.M. Yassin, S. Ficklin,C. C. Yeh, F. Hebard, K. Baier, W. Powell, S. C. Schuster, N. Wheeler, AI. Abbott, J. E. Carlson, & R. Sederoff. 2012. Chestnut Resistance to The Blight Disease: Insights from Transcriptome Analysis. BMC Plant Biology. 12(38):1471-2229. http://www.biomedcentral.com/1471-2229/12/38

[4] Borges, F., G. Gomes, R. Gardner, N. Moreno, S. MeCormick, J. A. Feijo, & J. D. Becker. 2008. Comparative Transcriptomics of Arabidopsis Sperm Cells. Plant Physiology. 148:1168-1181. www.plantphysiol.org

[5] Collins, L. J., P. J. Biggs, C. Voelckel & S. Joly. 2008. An Approach To Transcriptome Analysis of Non- Model Organisms Using Short-Read Sequences. Genome Informatics. 21:3-14

[6] Eveland, A. L., N. S. Nagasawa, A. Gofdshmidt, S. Meyer, M. Beatty, H. Sakai, D. Ware & D. Jackson. 2010. Digital Gene Expression Signatures for Maize Development. Plant Physiology. 154:1024-1039. www.plantphysiol.org

[7] Fan H, Y. Xiao, Y. Yang, W. Xias, A. S. Mason, Z. Xia, F. Qiao, S. Zhao & H. Tang. 2013. RNA-Seq Analysis of *Cocos nucifera*: Transcriptome Sequencing and De Novo Assembly for Subsequent Functional Genomics Approaches. PLoS ONE. 8(3): e59997. doi:10.1371/journal.pone.0059997

[8] Fox, S. E., J. Preece, J.A. Kimbref, G. L. Marchini, A. Sage, K. Y. Clark, M. B. Cruzart & P. Jaiswal.2013. Sequencing and De Novo Transcriptome Assembly of *Brachypodium sylvaticum* (Poaceae). Applications in Plant Sciences. 1(3): 1200011

[9] Fu, N., 0. Wang & L. H. Shen. 2013. De Novo Assembly, Gene Annotation and Marker Development Using Illumina Paired-End Transcriptome Sequences in Celery (*Apium graveolens* L.). PLoS ONE. 8(2):e57686. doi:10.1371/journal.pone.0057686

[10] Huang, D. W., B. T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M.W. Baseler, H. C. Lane & R. A. Lempicki. 2007. DAVID Bioinformatics Resources: Expanded Annotation Database and Novel Algorithms to Better Extract Biology from Large Gene Lists. Nucleic Acids Research. 35:169-175. doi:10.l093/nar/gkm415

[11] Huang, H. H., L. L. Xu, Z. K. Tong, E. P. Lin, Q. P. Liu, L J. Cheng & 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. 13:1471-2164. http://www.biomedcentral.com/1471-2164/13/648

[12] Jiao, X., B. T. Sherman, D. W. Huang, R. Stephens, M. W. Baseler, H. C. Lane & R. A. Lempicki. 2012. DAVID-WS: A Stateful Web Service to Facilitate Gene/Protein List Analysis. Bioinformatics. 13(28):1805-1806. doi:10.1093/bioinformatics/bts251

[13] Khanduri, V. P. 2012. Annual Variation in Floral Phenology and Pollen Production in a 25-Year-Old Plantation of *Tectona grandis*. Nordic Journal of Botany. 30:82-89. doi: 10.1111/j.1756-1051.2011.01157.x

[14] Li, D., Z. Deng, B. Qin, X. Liu, & Z. Men. 2012. De Novo Assembly and Characterization of Bark Transcriptome Using Illumina Sequencing and Development of EST-SSRMarkers in Rubber Tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics. 13:192. http://www.biomedcentral.com/1471-2164/13/192

[15] Logacheva, M. D., A. S. Kasianov, D. V. Vinogradov, T. H. Samigullin, M. S. Gelfand, V. J. Makeev & A. A. Penin. 2011. De Novo Sequencing and Characterization of Floral Transcriptome in Two Species of Buckwheat (*Fagopyrum*). BMC Genomics. 12(30):1471-2164. http://www.biomedcentral.com/1471-2164/12/30

[16] Lulin H., Y. Xiao, S. Pei, T. Wen & H. Shangqin. 2012. The First Illumina-Based De Novo Transcriptome Sequencing and Analysis of Safflower Flowers. PLoS ONE. 7(6): e38653. doi:10.1371/journal.pone.0038653

[17] Lyngdoh, N., G. Joshi, G. Ravikanth, R. U. Shaanker & R. Vasudeva. 2010. Influence of Levels of Genetic Diversity on Fruit Quality in Teak (*Tectona grandis* L.f.) Current Science. 5(99):639-644

[18] Palupi, E. R., J. N. Owens, S. Sadjad, Sudarsono & 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

[19] Rosli , H. M., N. M. Saleh, M. I. M. Noor, M. Norwati, B. Norlia & A. Norwati. 2009. Analysis of Expressed Sequence Tags Derived from Inflorescence Shoot of *Tectona grandis* (Teak). African Journal of Biotechnology. 8(21):5633-5640.

[20] Widiyanto, S. N, A. Sukmawan, N. Haro & H. Rahmania. 2009. Transient Expression of β-glucuronidase Reporter Gene in Agrobacterium Inoculated Shoots of Various Teak Clones. African Journal of Biotechnology. 8(10):2143-2150.