**University of São Paulo
“Luiz de Queiroz” College of Agriculture**

Characterization of genes involved in lignin biosynthesis in *Tectona grandis*

Esteban Galeano Gómez

Thesis presented to obtain the degree of Doctor in
Science. Program: International Plant Cell and
Molecular Biology

**Piracicaba
2015**

Esteban Galeano Gómez
Forestry Engineer

Characterization of genes involved in lignin biosynthesis in *Tectona grandis*
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof^ª. Dr^ª. **HELAINÉ CARRER**

Thesis presented to obtain the degree of Doctor in
Science. Program: International Plant Cell and
Molecular Biology

Piracicaba
2015

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Galeano Gómez, Esteban

Characterization of genes involved in lignin biosynthesis in *Tectona grandis* / Esteban Galeano Gómez. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.
178 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Bioinformática 2. Fatores de transcrição 3. PCR em tempo real 4. RNAseq 5. Via fenilpropanóide 6. Xilema secundário I. Título

CDD 634.97338
G151c

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Angels exist, and they are really close. Those whom God has given me are my father, Gustavo Adolfo, and my mother, Alina del Carmen, who taught me to be a kind and moral person and to be strong beyond the difficulties. They showed me how to be amazed with the motivated beauty of life. Taking the time to help, smile, care and respect others are lessons which came from my angels.

And to make my life more interesting, God gave me the gift of a third little angel called Andrés López Rubio, who has given me his best moments, and which have made my life so much better.

I DEDICATE

ACKNOWLEDGMENTS

God, thank you for giving me everything I have needed throughout my life.

I feel such huge gratitude for my brother Pablo and my sister Paulina for the laughs, company and happiness throughout our lives, and for my grandmothers, Amanda and Elena, for their unconditional love. I am also grateful for having the best uncles and aunts: Sandra, Victor, Jader, Claudia, Omaira and Doris.

I want to thank Professor Helaine Carrer for believing in me and giving me all of her support, advice, friendship and lessons. Also, I thank Valentina de Fátima de Martin, Tarcísio Sales Vasconcelos, André Luiz Barbosa, Tânia Batista, Guilherme Hosaka and Ana Preczenhak for their help and friendship throughout the project.

Thanks to my lab partners, Daniel Alves Ramiro, Evandro Tambarussi, Geraldo Silva, Enio Tiago Oliveira, Keini Dressano, Fausto Andrés Ortiz, Flávio dos Santos, Nicole Labruto, Paulo Ceciliato, Frederico Almeida, Tabata Bergonci, Juan Carlos Guerrero and João Gabriel Vezza for all their help and suggestions during the research.

Thanks to Marcelo Brandão, Luiz Lehmann Coutinho and Professor Erich Grotewold for guiding and counseling me at various points in my research, and to all of the staff at the Center for Applied Plant Sciences, Ohio State University.

Thanks to my all lifelong friends for their true friendship and love: Alvaro Ramírez, Melina Alvarez, Miguel Angel Cano, Sebastian Salazar, Juan Manuel Zuluaga, Susana Montoya, Oscar Santiago Ariztizábal and Sebastián López.

Thanks to my friends Erick Espinoza, Edjane Freitas, Pedro Mansilla, Marco Arizapana, Berenice Alcantara, Javier Pulido, Eleonora Zambrano, Danilo Ignacio de Urzedo, Mariana Ferraz, Ivan Mozol, Diana Castillo, Jessica Johnson, Nelson Casas, Maryeimi Varón, Manuella Nóbrega, Rodrigo Keller and Diana Vasquez for all the conversations and exchanges of ideas during the last four years.

I am grateful to have received scholarships from CAPES (PEC-PG 5827108) and FAPESP (2013/06299-8).

Finally, I give a special thanks to all the professors, staff and friends from the University of São Paulo and Ohio State University.

SUMMARY

RESUMO	11
ABSTRACT	13
1 INTRODUCTION	15
References	21
2 IDENTIFICATION AND VALIDATION OF QUANTITATIVE REAL-TIME REVERSE TRANSCRIPTION PCR REFERENCE GENES FOR GENE EXPRESSION ANALYSIS IN TEAK (<i>Tectona grandis</i> L.f.).....	25
Abstract.....	25
2.1 Introduction	25
2.2 Materials and Methods	27
2.2.1 Plant material	27
2.2.2 Total RNA extraction, purification and quality controls	28
2.2.3 cDNA synthesis.....	29
2.2.4 Multiple sequence alignments, PCR and qRT-PCR primer design	29
2.2.5 Primer specificity, qRT-PCR Efficiency and R^2	29
2.2.6 Quantitative real-time reverse transcription PCR	31
2.2.7 Analysis of gene expression stability	31
2.2.8 Validation of reference genes.....	31
2.3 Results	32
2.3.1 Identification and cloning of references genes in teak.....	32
2.3.2 Primer specificity and PCR efficiency	33
2.3.3 Expression stability of the nine candidate reference genes.....	33
2.3.4 geNorm.....	35
2.3.5 NormFinder	37
2.3.6 BestKeeper	38
2.3.7 Delta Ct	39

2.3.8 Validation of <i>TgUBQ</i> and <i>TgEF-1a</i> as internal controls to assess expression of the teak <i>cinnamyl alcohol dehydrogenase</i> gene in lignified tissues	39
2.4 Discussion	40
2.5 Conclusions	44
Additional Files	45
References	71
3 CHARACTERIZATION OF <i>CINNAMYL ALCOHOL DEHYDROGENASE</i> GENE FAMILY IN LIGNIFYING TISSUES OF <i>Tectona grandis</i> L.f.	77
Abstract	77
3.1 Introduction	77
3.2 Materials and Methods	80
3.2.1 Plant material	80
3.2.2 RNA extraction and cDNA synthesis	81
3.2.3 Amplification of <i>CAD</i> family in <i>Tectona grandis</i>	82
3.2.3.1 Amplification of <i>TgCAD1</i>	82
3.2.3.2 Amplification of <i>TgCAD2</i> , <i>TgCAD3</i> , <i>TgCAD4</i>	83
3.2.4 Characterization and modeling of the <i>TgCAD1</i> protein	83
3.2.5 Phylogeny and characterization of <i>CAD</i> family in <i>Tectona grandis</i>	84
3.2.6 Gene expression of the <i>CAD</i> family in <i>Tectona grandis</i> by qRT-PCR	84
3.3 Results	85
3.3.1 Amplification of <i>CAD</i> family in <i>Tectona grandis</i>	85
3.3.1.1 Amplification of <i>TgCAD1</i>	85
3.3.1.2 Amplification of <i>TgCAD2</i> , <i>TgCAD3</i> , <i>TgCAD4</i>	85
3.3.2 Characterization and modeling of the <i>TgCAD1</i> protein	86
3.3.3 Phylogeny and characterization of <i>CAD</i> family in <i>Tectona grandis</i>	86
3.3.4 Gene expression of the <i>CAD</i> family in <i>Tectona grandis</i> by qRT-PCR	91
3.3.5 Gene expression in sapwood from mature and young <i>T. grandis</i> trees	91
3.4 Discussion	93

3.4.1 Characterization of <i>CAD</i> gene family	93
3.4.2 Differential expression of <i>TgCAD</i> gene family	95
3.5 Conclusions and Perspectives	96
Additional Files	98
References	110
4 RNA-SEQ REVEALS TRANSCRIPT PROFILING AND <i>MYB</i> TRANSCRIPTION FACTORS OF LIGNIFIED TISSUES IN <i>Tectona grandis</i>	115
Abstract.....	115
4.1 Introduction	115
4.2 Materials and Methods	118
4.2.1 Plant material	118
4.2.2 Total RNA extraction and Illumina sequencing.....	119
4.2.3 Mapping data against closely-related genomes.....	119
4.2.4 Cleaning and <i>de novo</i> assembly	120
4.2.5 Detection and annotations of differentially expressed unigenes between twelve- and sixty-year-old trees.....	120
4.2.6 Phylogeny of <i>MYB</i> transcription factors differentially expressed in teak.....	121
4.2.7 Gene expression of <i>MYBs</i> along the lignified teak tissues by qRT-PCR.....	121
4.3 Results	122
4.3.1 Quality of the RNA and the reads	122
4.3.2 Read mapping against the tomato, <i>Populus</i> and <i>Eucalyptus</i> genomes	122
4.3.3 <i>De novo</i> assembly	123
4.3.4 Unigenes differentially expressed in lignified tissues between 12- and 60-year-old trees	124
4.3.5 Functional annotations of unigenes differentially expressed in lignified tissues.....	124
4.3.6 Metabolic pathways of unigenes	127
4.3.7 Phylogenetic analysis of the teak R2R3- <i>MYB</i> gene family	130
4.3.8 Gene expression of <i>MYB</i> transcription factors in teak	132

4.4 Discussion.....	134
4.4.1 <i>T. grandis</i> transcriptome.....	134
4.4.2 RNAseq provided several useful unigenes differentially expressed in lignified tissues of <i>T. grandis</i>	136
4.4.3 Activation of stimulus response genes and heat-shock proteins with local environmental changes	137
4.4.4 <i>MYB</i> transcription factors revealed phylogenetic grouping and distinct expression during maturity	138
4.5 Implications and Perspectives of this study	140
4.6 Conclusions.....	141
Additional Files.....	142
References.....	163
5 GENERAL CONCLUSIONS	173
6 IMPLICATIONS AND PERSPECTIVES OF THIS STUDY	175
6.1 Heat-Shock proteins.....	175
6.2 Regulation in teak	175
6.3 It is essential to understand genes involved in teak wood	175
6.4 The transcriptomes lead to discover gene families, understand lignification processes and establish transcriptional regulatory networks	176
6.5 The next-generation sequencing in breeding programs	176
6.6 Wood and heartwood quality as characteristics in breeding programs.....	177
6.7 Three questions to be answered in future research projects	178

RESUMO

Caracterização de genes envolvidos na biossíntese de lignina em *Tectona grandis*

A árvore de teca (*Tectona grandis* L.f.) tem alto valor no comércio de madeira para a fabricação de produtos lenhosos, devido às suas qualidades extraordinárias de cor, densidade e durabilidade. Apesar da importância desta espécie, são poucos os estudos genéticos e moleculares disponíveis. Também, a falta de informação molecular sobre xilema secundário e maturação da árvore tem dificultado a exploração genética de teca. Assim, estudos de expressão gênica e perfis transcricionais são relevantes para explorar a formação da madeira e a biossíntese de lignina durante o desenvolvimento e envelhecimento das plantas vasculares. Visando os estudos de expressão gênica, foi essencial identificar e clonar genes de referência para a teca. Foram testados oito genes comumente usados em qRT-PCR, *TgRP60S*, *TgCAC*, *TgACT*, *TgHIS3*, *TgSAND*, *TgTUB*, *TgUBQ* e *TgEF1a*. Perfis de expressão destes genes foram avaliados por qRT-PCR em seis amostras de tecidos e órgãos (folhas, flores, plântulas, raiz, xilema secundário de caule e ramo). A validação da estabilidade pelos programas NormFinder, BestKeeper, geNorm e Delta CT mostrou que *TgUBQ* e *TgEF1a* são os genes mais estáveis para usar como genes de referência em teca nas condições testadas. Em virtude da disponibilidade de árvores de teca de diferentes idades, entre 12 e 60 anos, foi realizado o RNAseq de diferentes órgãos (plântulas, folhas, flores, raiz, ramos e caules de árvores de 12 e 60 anos). Obteve-se um total de 462.260 transcritos pela montagem com o software “Trinity”. Foram identificados 1.502 e 931 genes diferencialmente expressos para xilema secundário de caule e ramo, respectivamente, utilizando o programa DESeq e fatores de transcrição *MYB*, que foram posteriormente caracterizados. A sequência de aminoácidos do TgMYB1 exibiu um motivo “coiled-coil” (CC), enquanto TgMYB2, TgMYB3 e TgMYB4 mostraram domínio R2R3-MYB. Todos eles foram filogeneticamente agrupados com várias gimnospermas e angiospermas. Observou-se alta expressão do *TgMYB1* e *TgMYB4* em tecidos lignificados de árvores de 60 anos de idade. Neste trabalho também foi estudada a família gênica *Cinamil álcool desidrogenase* (*CAD*). Foi caracterizado um membro completo (*TgCAD1*) e três parciais (*TgCAD2* a *TgCAD4*). As quatro enzimas apresentaram resíduos de ação catalítica e estrutural de zinco, de ligação ao NADPH e de especificidade de substrato, em conformidade com o mecanismo conservado de álcool desidrogenases. *TgCAD3* e *TgCAD4* foram altamente expressos no alburno jovem e maduro e parecem estar duplicados e relacionados com a biossíntese de lignina. O melhoramento genético de árvores, a seleção assistida utilizando marcadores moleculares e a transformação de plantas parecem ser linhas promissoras de pesquisa, a partir dos dados obtidos nesta pesquisa. Este é o primeiro estudo sobre caracterização e expressão gênica, filogenia e perfis transcricionais em teca.

Palavras-chave: Bioinformática; Fatores de transcrição; PCR em tempo real; RNAseq; Via fenilpropanóide; Xilema secundário

ABSTRACT

Characterization of genes involved in lignin biosynthesis in *Tectona grandis*

Teak tree (*Tectona grandis* L.f.) has a high value in the timber trade for fabrication of woody products due to its extraordinary qualities of color, density and durability. Despite the importance of this species, genetic and molecular studies available are limited. Also, the lack of molecular information about secondary xylem and tree maturation has hindered genetic exploration of teak. Therefore, gene expression studies and transcriptomic profiling are essential to explore wood formation and lignin biosynthesis through the development and aging of vascular plants. Aiming the gene expression studies, it was essential to identify and clone reference genes for teak. Eight genes were tested, commonly used in qRT-PCR, including *TgRP60S*, *TgCAC*, *TgACT*, *TgHIS3*, *TgSAND*, *TgTUB*, *TgUBQ* and *TgEF1a*. Expression profiles of these genes were evaluated by qRT-PCR in six tissue and organ samples (leaf, flower, seedling, root, stem and branch secondary xylem). Stability validation by NormFinder, BestKeeper, geNorm and Delta Ct programs showed that *TgUBQ* and *TgEF1a* are the most stable genes to use as qRT-PCR reference genes in teak in the conditions tested. Due to the availability of 12- and 60-year-old teak trees, RNA-seq was performed in different organs (seedlings, leaves, flowers, root, stem and branch secondary xylem). A total of 462,260 transcripts were obtained by assembling with “Trinity” software. Also, 1,502 and 931 genes differentially expressed were identified for stem and branch secondary xylem, respectively, using DESeq program, and *MYB* transcription factors, which were characterized. *TgMYB1* amino acid sequence displayed a predicted coiled-coil (CC) motif while *TgMYB2*, *TgMYB3* and *TgMYB4* showed R2R3-MYB domain. All of them were phylogenetically grouped with several gymnosperms and flowering plants. High expression of *TgMYB1* and *TgMYB4* in lignified tissues of 60-year-old trees was observed. In this work, the *Cinnamyl Alcohol Dehydrogenase* (*CAD*) gene family was also studied. One complete (*TgCAD1*) and three partial (*TgCAD2* to *TgCAD4*) members were characterized. The four enzymes presented residues for catalytic and structural zinc action, NADPH binding and substrate specificity, consistent with the mechanism of alcohol dehydrogenases. *TgCAD3* and *TgCAD4* were highly expressed in young and mature sapwood and seem to be duplicated and highly related with lignin biosynthesis. Tree genetic improvement, marker-assisted selection and plant transformation seem to be promising lines of research for the data obtained from this research. This is the first study addressing gene characterization and expression, phylogeny and transcriptomic profiling in teak.

Keywords: Bioinformatics; Phenylpropanoid pathway; Quantitative real-time PCR; RNAseq; Secondary xylem; Transcription factors

1 INTRODUCTION

Wood, usually considered the secondary xylem, is a natural renewable resource which provides timber (construction, furniture), fibres (paper), energy (firewood) and biofuels, and represents a substantial advance in plant evolution during the Cretaceous because allowed trees to conquer land, to form stratified communities, to reach heights, to conduct water from roots to the crown and to support mass inspite wind, snow, slope and light (DÉJARDIN et al., 2010). Wood is essential to the world economy, human confort and terrestrial ecosystem carbon-cycling, but its exponential harvesting and constant growth demand are leading to natural forest degradation and a big concern about the future supplies.

Lignin, one of the wood components, is a phenolic polymer composted by *p*-coumaryl (H), coniferyl (G) and sinapyl (S) alcohols with essential roles in structural rigidity, pathogen defense, conduction of water and, after cell division, expansion and death, lignin constitutes the wood (BONAWITZ; CHAPPLE, 2010). Over the last decade, lignin biosynthesis has been a target for several studies due to its agricultural and economic importance (XU et al., 2013), which configures the final product of the phenylpropanoid pathway when the monolignols are joined together.

The phenylpropanoid pathway includes several intermediates and enzymes (Figure 1) (VANHOLME et al., 2010). The first step is a deamination of phenylalanine by the phenylalanine ammonia-lyase (PAL) producing cinnamic acid, which consequently is hydroxylated by the cinnamate-4-hydroxylase (C4H) producing *p*-coumaric acid, followed by a generation of *p*-coumaroyl-CoA by the 4-coumarate: CoA ligase (4CL), and this substrate is processed by the cinnamoyl-CoA reductase (CCR) to coniferaldehyde followed by a conversion to coniferyl alcohol by the cinnamyl alcohol dehydrogenase (CAD); *p*-coumaroyl-CoA can be transformed to *p*-coumaroyl-CoA shikimate by the hydroxycinnamoyl transferase (HCT) (Figure 1) (BARAKAT et al., 2009). The enzymes *p*-coumarate 3-hydroxylase (C3H), HCT, caffeoyl-CoA O-methyltransferase (CCOMT) and CCR transform *p*-coumaroyl-CoA shikimate into caffeoyl shikimate, caffeoyl-CoA, feruloyl CoA and coniferaldehyde, respectively; CAD, ferulate 5-hydrolase (F5H) and caffeic/5-hydroxyferulic acid O-methyltransferase (COMT) transform coniferaldehyde into coniferyl alcohol, 5-Hydroxy-coniferaldehyde and sinapyl aldehyde, respectively (Figure 1) (BARAKAT et al., 2009). The same authors explained that CAD and F5H/COMT can produce sinapyl alcohol from sinapyl aldehyde and coniferyl alcohol, respectively. Therefore, CAD is a key enzyme that catalyzes

the conversion of cinnamyl aldehydes to cinnamyl alcohols and can synthesize both coniferyl and sinapyl alcohols (Figure 1). Several *CAD* genes have been characterized finding complete families and phylogeny, and they have also been used for genetic transformation (ANTEROLA; LEWIS, 2002; GUO; RAN; WANG, 2010).

Figure 1 – The phenylpropanoid pathway (VANHOLME et al., 2010). PAL, PHENYLALANINE AMMONIOLYASE; C4H, CINNAMATE 4-HYDROXYLASE; 4CL, 4-COUMARATE:CoA LIGASE; C3H, *p*-COUMARATE 3-HYDROXYLASE; HCT, *p*-HYDROXYCINAMOYL-CoA:QUINATE/SHIKIMATE *p*-HYDROXYCINAMOYLTRANSFERASE; CCoAOMT, CAFFEYOYL-CoA O-METHYLTRANSFERASE; CCR, CINAMOYL-CoA REDUCTASE; F5H, FERULATE 5-HYDROXYLASE; COMT, CAFFEIC ACID O-METHYLTRANSFERASE; CAD, CINNAMYL ALCOHOL DEHYDROGENASE.

In addition, the regulation of the phenylpropanoid pathway by *MYB* transcription factors and their influence in the lignin biosynthesis have been widely described (RAHANTAMALALA et al., 2010). The *MYB* transcription factors in plants (mainly the R2R3-type) have been extensively characterized with essential roles in vascular organization (ROGERS; CAMPBELL, 2004; BEDON; GRIMA-PETTENATI; MACKAY, 2007).

There is lack on the knowledge of how wood formation is genetically regulated and how all these mechanism could be used to ensure tree resources availability in the future. In recent years, genomics and transcriptomics have emerged as promising areas related to the molecular biology, aiming in the understanding of complex biological processes, such as lignin deposition and wood formation, particularly in important tropical trees such as eucalyptus and teak (*Tectona grandis* Linn. F.).

Among the molecular biology techniques and methods currently available for gene expression assessment and transcriptome analysis, the quantitative real-time reverse transcription RT-PCR (qRT-PCR) is a reliable and efficient technology available for quantifying levels of transcripts, along with the RNA sequencing (next-generation sequencing technology), which is a powerful tool for providing genetic information of organisms that do not yet have the complete genome available.

Teak plant is a deciduous tree from Lamiaceae family, which can reach up to 45-50 meters of height, composed by a rounded and open crown (Figure 2A), with a robust and stellate stem which can reach up to 150-250 cm of diameter at breast height, usually with large, petiolate, oppositely arranged leaves (ocasionally three at a node or rarely alternate in seedlings), dark green in the upper surface and silvery in the under surface, commonly elliptic or slightly obovate, with acute or acuminate apex with sizes of 20-55 cm x 15-37 cm of leaf length and 5-6 cm of leaf petiole (Figure 2B) (REDDY 2003). The inflorescence is mainly terminal with bracteate flowers arranged in cymose panicles, actinomorphic (Figure 2C) and the flowers are small, white, bisexual (hermaphroditic) and irregular, with white corolla, tube broadly cylindrical, filaments glabrous, yellow anthers (Figure 2D), ovary densely pubescent with one ovule, and when pollinated (usually by wind, bumble bees, bees and wasps, mainly *Ceratina* spp.) can produce fruits as a subglobose drupe (Figure 2E) enclosed by an expanded inflated calyx, with a pericarp densely felted-tomentose with irregularly branched pale brown hairs, containing from two to four seeds (Figure 2F) (REDDY 2003).

As ecological factors, teak trees grow well in fairly moist, warm tropical climate, with a mean annual temperature of 25-38°C, between 1,250 - 2,500 mm/year of rainfall, presenting the best yields under 600 meters above sea level and produce better wood quality with long dry periods, from 3 to 5 month long (KOLLERT; CHERUBINI, 2012; REDDY 2003). Teak tree is a strong light demander, intolerant of shade and requires complete overhead light, and plantations prefer deep, porous, well-drained soils (ABRAF 2013; REDDY 2003).

Teak is native to countries of southeast Asia such as Myanmar, Thailand, India, Laos and Java Islands, and is the most valuable commercial timber in the tropics due to its beauty, light weight, high durability, dimensional stability and resistance to external environmental factors (LUKMANDARU; TAKAHASHI, 2008). It is used for furniture, buildings, finishes, cabinets, sleepers, decorative veneers, lamination, house walls, flooring, joinery, carpentry, vehicles, mining and shipbuilding (BAILLÈRES; DURAND, 2000; BHAT; PRIYA; RUGMINI, 2001).

Figure 2 – Teak morphology. A= tree, B= leaves, C=inflorescence, D=flowers, E=fruit, F=seeds.

Teakwood is the only valuable hardwood that constitutes a globally emerging forest resource with a planted area of 4,346 million ha (0,5 million m³ of wood) and natural forest of 29,035 million ha (2 million m³ of wood) around the world, including America, Africa and Asia (Figure 3) (KOLLERT; CHERUBINI, 2012).

Several countries have introduced teak plantations: Bangladesh, Cambodia, Nepal, Pakistan, Japan, Sri Lanka, Taiwan, Vietman in South East Asia; Australia, Fiji Islands, U.S. Pacific Islands in the Pacific; Kenya, Malawi, Somalia, Sudan, Tanzania, Uganda, Zimbabwe in the East Africa; Benin, Ghana, Guinea, Ivory Coast, Nigeria, Senegal, Togo in the West Africa; South Africa; Cuba, Honduras, Jamaica, Nicaragua, Panama, Puerto Rico, West Indies in the Carribbean; Argentina, Brazil, Colombia, Surinam, Venezuela, Ecuador in South America; Belize, Costa Rica, El Salvador in Central America (KOLLERT; CHERUBINI, 2012).

Brazil presents the largest teak reforestation in South America, and plantations are concentrated in the Center-West and North of the country, in the states of Roraima, Amazonas, Amapá, Pará, Acre, Rondônia, Mato Grosso, Tocantins and Minas Gerais (Figure 4) (KOLLERT; CHERUBINI, 2012; ABRAF 2013), mainly in the municipalities of Cáceres and Alta Floresta (State of Mato Grosso) and Dom Eliseu (State of Pará) (CASTRO, 2011).

Teak is not a fast growing species but can produce a timber of optimum strength in relatively short rotations of 21 years (BHAT; INDIRA, 1997) depending on the sapwood-heartwood percentages. The timber quality produced will be the overriding commercial factor for the near future (GOH et al., 2007), and usually relates to the amount, color and durability of the heartwood.

Currently, the wood market has a great interest in teak extractives such as naphthoquinones and anthraquinones, which have shown remarkable properties against fungus and termites (GUERRERO-VÁSQUEZ et al., 2013).

Additionally, teak populations have significant environmental roles, as they can be used in agroforestry systems and forest recovery. These characteristics make teak one of the most widely grown and economically profitable trees around the world.

Figure 3 –Surface of teak plantations in the world (KOLLERT; CHERUBINI, 2012). In green, countries with teak plantations

Figure 4 –States with teak plantations in Brazil (KOLLERT; CHERUBINI, 2012; ABRAF 2013). RR, Roraima; AM, Amazonas; AP, Amapá; PA, Pará; AC, Acre; RO, Rondônia; MT, Mato Grosso; TO, Tocantins; MG, Minas Gerais

Studies characterizing genes related to secondary xylem, vessel formation, sapwood and heartwood differentiation, volume growth and abiotic stress have been documented in *Populus tremula* (SCHRADER et al., 2004), *Populus trichocarpa* (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010), eucalyptus (MIZRACHI et al., 2010), conifers (BEDON; GRIMA-PETTENATI; MACKAY, 2007; PAVY et al., 2008) and *Fraxinus spp.* (BAI et al., 2011). Despite the great economic, biological and ecological importance of teak, there was no gene characterization and expression studies available before this project.

The purpose of this research was to discover and characterize genes related to lignin biosynthesis and wood formation in *Tectona grandis* through cloning, gene expression, phylogenetic analyzes, transcriptomic profiling and bioinformatics tools. With this investigation, the characterization of control genes for gene expression studies, the discovery of members from *CAD* and *MYB* gene family and the RNA sequencing covering the transcripts profile of *T. grandis* during young to mature transition and in several tissues were obtained.

References

- ABRAF. Anuário estatístico da Associação Brasileira dos produtores das florestas plantadas: Ano Base 2012. ABRAF: Brasília, 2013. 142 p.
- ANTEROLA, A.M.; LEWIS, N.G. Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. **Phytochemistry**, New York, v. 61, n. 3, p. 221–294, 2002.
- BAI, X.; RIVERA-VEGA, L.; MAMIDALA, P.; BONELLO, P.; HERMS, D.A.; MITTAPALLI, O. Transcriptomic signatures of ash (*Fraxinus spp.*) phloem. **PloS One**, San Francisco, v. 6, n. 1, p. e16368, 2011.
- BAILLÈRES, H.; DURAND, P.Y. Non-destructive techniques for wood quality assessment of plantation grown teak. **Bois et Forêts des Tropiques**, Nogent-sur-Marne, v. 263, n. 1, p. 17–29, 2000.
- BARAKAT, A.; BAGNIEWSKA-ZADWORNA, A.; CHOI, A.; PLAKKAT, U.; DILORETO, D.S.; YELLANKI, P.; CARLSON, J.E. The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. **BMC Plant Biology**, London, v. 9, p. 26, 2009.

BEDON, F.; GRIMA-PETTENATI, J.; MACKAY, J. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). **BMC Plant Biology**, London, v. 7, p. 17, 2007.

BHAT, K.M.; INDIRA, E.P. **Effect of faster growth on timber quality of teak**. Thrissur: Kerala Forest Research Institute, 1997. 60 p.

BHAT, K.M.; PRIYA, P.B.; RUGMINI, P. Characterisation of juvenile wood in teak. **Wood Science and Technology**, New York, v. 34, n. 6, p. 517–532, 2001.

BONAWITZ, N.D.; CHAPPLE, C. The genetics of lignin biosynthesis: connecting genotype to phenotype. **Annual Review of Genetics**, Palo Alto, v. 44, p. 337–363, 2010.

CASTRO, V.R. de. **Aplicação de métodos não destrutivos na avaliação das propriedades físicas do lenho de árvores de *Pinus caribaea* var. *hondurensis* Barr. et Golf. e *Tectona grandis* (L.f.)**. 2001. 104 p. Dissertação (Mestrado em Recursos Florestais) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.

DÉJARDIN, A.; LAURANS, F.; ARNAUD, D.; BRETON, C.; PILATE, G.; LEPLÉ, J. Wood formation in Angiosperms. **Comptes Rendus Biologies**, Paris, v. 333, p. 325–334, 2010.

DHARMAWARDHANA, P.; BRUNNER, A.M.; STRAUSS, S.H. Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. **BMC Genomics**, London, v. 11, p. 150, 2010.

GOH, D.K.S.; CHAIX, G.; BAILLÈRES, H.; MONTEUUIS, O. Mass production and quality control of teak clones for tropical plantations : the Yayasan Sabah Group and CIRAD Joint Project as a case study. **Bois et Forêts des Tropiques**, Nogent-sur-Marne, v. 293, n. 3, p. 65–77, 2007.

GUERRERO-VÁSQUEZ, G.A.; ANDRADE, C.K.Z.; MOLINILLO, J.M.G.; MACÍAS, F.A. Practical first total synthesis of the potent phytotoxic (±)-naphthotectone, isolated from *Tectona grandis*. **European Journal of Organic Chemistry**, Weinheim, v. 2013, n. 27, p. 6175–6180, 2013.

GUO, D.-M.; RAN, J.-H.; WANG, X.-Q. Evolution of the Cinnamyl/Sinapyl Alcohol Dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of Bona Fide CAD. **Journal of Molecular Evolution**, New York, v. 71, n. 3, p. 202–218, 2010.

KOLLERT, W.; CHERUBINI, L. **Teak resources and market assessment 2010 (*Tectona grandis* Linn. F.)**. Rome: FAO, 2012. 42 p.

LUKMANDARU, G.; TAKAHASHI, K. Variation in the natural termite resistance of teak (*Tectona grandis* Linn. fil.) wood as a function of tree age. **Annals of Forest Science**, Les Ulis, v. 65, p. 708, 2008.

MIZRACHI, E.; HEFER, C.A.; RANIK, M.; JOUBERT, F.; MYBURG, A.A. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. **BMC Genomics**, London, v. 11, n. 1, p. 681, 2010.

PAVY, N.; BOYLE, B.; NELSON, C.; PAULE, C.; GIGUÈRE, I.; CARON, S.; PARSONS, L. S.; DALLAIRE, N.; BEDON, F.; BÉRUBÉ, H.; COOKE, J.; MACKAY, J. Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. **The New phytologist**, Cambridge, v. 180, n. 4, p. 766–786, 2008.

RAHANTAMALALA, A.; RECH, P.; MARTINEZ, Y.; CHAUBET-GIGOT, N.; GRIMA-, J.; PACQUIT, V. Coordinated transcriptional regulation of two key genes in the lignin branch pathway - CAD and CCR - is mediated through MYB- binding sites. **BMC Plant Biology**, London, v. 10, p. 130, 2010.

REDDY, S.M. Gymnosperms, Plant Anatomy, Genetics, Ecology. New Age International Publishers: New Delhi, 2003. 560 p.

ROGERS, L.A.; CAMPBELL, M.M. The genetic control of lignin deposition during plant growth and development. **New Phytologist**, Cambridge, v. 164, n. 1, p. 17–30, 2004.

SCHRADER, J.; NILSSON, J.; MELLEROWICZ, E.; BERGLUND, A.; NILSSON, P.; HERTZBERG, M. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. **The Plant Cell**, Rockville, v. 16, p. 2278–2292, Sept. 2004.

VANHOLME, R.; DEMEDTS, B.; MORREEL, K.; RALPH, J.; BOERJAN, W. Lignin biosynthesis and structure. **Plant Physiology**, Bethesda, v. 153, n. 3, p. 895–905, 2010.

XU, Y.; THAMMANNAGOWDA, S.; THOMAS, T.P.; AZADI, P.; SCHLARBAUM, S.E.; LIANG, H. LtuCAD1 is a cinnamyl alcohol dehydrogenase ortholog involved in lignin biosynthesis in *Liriodendron tulipifera* L., a basal angiosperm timber species. **Plant Molecular Biology Reporter**, Athens, v. 31, n. 5, p. 1089–1099, 2013.

2 IDENTIFICATION AND VALIDATION OF QUANTITATIVE REAL-TIME REVERSE TRANSCRIPTION PCR REFERENCE GENES FOR GENE EXPRESSION ANALYSIS IN TEAK (*Tectona grandis* L.f.).

Abstract

Background: Teak (*Tectona grandis* L.f.) is currently the preferred choice of the timber trade for fabrication of woody products due to its extraordinary qualities and is widely grown around the world. Gene expression studies are essential to explore wood formation of vascular plants, and quantitative real-time reverse transcription PCR (qRT-PCR) is a sensitive technique employed for quantifying gene expression levels. One or more appropriate reference genes are crucial to accurately compare mRNA transcripts through different tissues/organs and experimental conditions. Despite being the focus of some genetic studies, a lack of molecular information has hindered genetic exploration of teak. To date, qRT-PCR reference genes have not been identified and validated for teak. Results: Identification and cloning of nine commonly used qRT-PCR reference genes from teak, including *ribosomal protein 60s* (*RP60S*), *clathrin adaptor complexes medium subunit family* (*CAC*), *actin* (*ACT*), *histone 3* (*HIS3*), *sand family* (*SAND*), β -*Tubulin* (*B-TUB*), *ubiquitin* (*UBQ*), *elongation factor 1- α* (*EF-1 α*), and *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*). Expression profiles of these genes were evaluated by qRT-PCR in six tissue and organ samples (leaf, flower, seedling, root, stem and branch secondary xylem) of teak. Appropriate gene cloning and sequencing, primer specificity and amplification efficiency was verified for each gene. Their stability as reference genes was validated by NormFinder, BestKeeper, geNorm and Delta Ct programs. Results obtained from all programs showed that *TgUBQ* and *TgEF-1 α* are the most stable genes to use as qRT-PCR reference genes and *TgACT* is the most unstable gene in teak. The relative expression of the teak *cinnamyl alcohol dehydrogenase* (*TgCAD*) gene in lignified tissues at different ages was assessed by qRT-PCR, using *TgUBQ* and *TgEF-1 α* as internal controls. These analyses exposed a consistent expression pattern with both reference genes. Conclusion: This study proposes a first broad collection of teak tissue and organ mRNA expression data for nine selected candidate qRT-PCR reference genes. NormFinder, Bestkeeper, geNorm and Delta Ct analyses suggested that *TgUBQ* and *TgEF-1 α* have the highest expression stability and provided similar results when evaluating *TgCAD* gene expression, while the commonly used *ACT* should be avoided.

doi:10.1186/1756-0500-7-464

Keywords: Relative expression; Trees; Transcript stability; Lignin

2.1 Introduction

The flux of information from DNA to protein is connected by mRNA, and the level of mRNA transcription is one of the factors determining the degree of gene expression (HAUPTMAN; GLAVAC, 2013). Changes in gene expression are critical for cell development (CONTE; BANFI; BOVOLENTA, 2013), integration of metabolism (OHDAN et al., 2005) and resistance to biotic and abiotic stresses (MACKAY; ARDEN; NITSCHKE,

2002; ZHU et al., 2013), and as such are a research area of great interest to the fields of medicine, pharmacy, life sciences and agronomy.

Methods currently available for gene expression assessment include microarray analysis, Northern blotting, *in situ* hybridization, RNase protection assay, RNA sequencing (RNA-seq), qualitative RT-PCR, competitive RT-PCR, and quantitative real-time reverse transcription RT-PCR (qRT-PCR). The qRT-PCR is considered an efficient, safe (free of radioactive reagents), fast, affordable, reproducible, reliable and specific for quantifying levels of transcripts (KUBISTA et al., 2006). However, some variables such as the integrity, amount and purity of the RNA used as well as enzyme efficiency during cDNA synthesis and PCR amplification make an additional step to normalize the data necessary (BUSTIN et al., 2009). Normalization requires the use of one or more reference genes (also called internal control genes) for which expression is constant and stable at different developmental stages, nutritional conditions or experimental conditions (DHEDA et al., 2005). Unfortunately, a gene has not been found for which expression is absolutely stable under all circumstances or across species that can be used indiscriminately for qRT-PCR analysis.

Bioinformatics tools have been developed to assess and identify the most suitable reference genes for qRT-PCR data normalization. geNorm shows expression stability throughout a set of housekeeping candidates (VANDESOMPELE et al., 2002), the Normfinder algorithm chooses the best candidate reference genes according to its calculations (ANDERSEN; JENSEN; ØRNTTOFT, 2004), while the Excel-based tool called BestKeeper determines the best candidate of pair-wise correlations (PFAFFL et al., 2004). Other statistical approaches used include Delta Ct (SILVER et al., 2006) and “Stability index” methods (BRUNNER; YAKOVLEV; STRAUSS, 2004).

Reference genes commonly used that present sufficiently stable expression are those related to cell maintenance such as *actin*, *tubulin*, *glyceraldehyde-3-phosphate dehydrogenase*, *elongation factor 1- α* and *18S ribosomal RNA* (RADONIĆ et al., 2004; DHEDA et al., 2005). New genes have been studied as internal controls in model or commercial plants, such as *Arabidopsis* (CZECHOWSKI et al., 2005), *Populus* (XU et al., 2011) and *Brachypodium* (CHAMBERS et al., 2012). Tests for the selection of reference genes for qRT-PCR in teak have not been published yet.

Teak is a deciduous tree, native to countries of southeast Asia such as Myanmar, Thailand, India, Laos and Java Islands (VERHAEGEN et al., 2010). Its wood is known internationally for its beauty, weightlessness, durability and weather resistance and it is used in the building of ships, furniture, house floors and walls, and general carpentry

(LUKMANDARU; TAKAHASHI, 2008; MIRANDA; SOUSA; PEREIRA, 2011). Currently, the wood market has a great interest in teak extractives such as naphthoquinones and anthraquinones, which have shown remarkable antifungal and antitermitic effect (HEALEY; GARA, 2003; GUERRERO-VÁSQUEZ et al., 2013). Additionally, teak populations serve significant environmental roles, as they can be used in agroforestry systems and forest recovery (HEALEY; GARA, 2003). These characteristics make teak one of the most widely grown and economically profitable trees around the world (HALLETT et al., 2011). Despite the great economic importance of teak, there are no studies of gene expression, the genome sequence is not available and sequenced genes are limited.

To select suitable qRT-PCR internal control genes for teak, this study analyzed the expression levels of candidate reference genes in different tissues and organs such as leaves, flowers, seedlings, roots, stem and branch secondary xylem of trees. Eight candidate reference genes were identified by their orthologous genes in model plants. These candidates were cloned, sequenced and tested. The selected genes are involved in different biological functions such as the formation of cellular cytoskeleton (*Actin and β -Tubulin*), elongation phase of translation (*Elongation factor 1- α*), DNA packaging (*Histone 3*) protein modification (*Ubiquitin*), intracellular transport (*Clathrin adaptor complexes medium subunit family*), vesicular transport (*SAND family*), protein biosynthesis (*Ribosomal protein 60s*) and carbohydrate metabolism (*Glyceraldehyde-3-phosphate dehydrogenase*).

Finally, in order to validate our results, the most stable reference genes were used to assess the *TgCAD* gene expression levels in different tissues and organs.

2.2 Materials and Methods

2.2.1 Plant material

Roots, seedlings and leaves were obtained from fifteen four month-old greenhouse grown teak. Flowers and branch and stem secondary xylem (Figure 1) were collected from fifteen twelve year-old teak trees located in Piracicaba, São Paulo State, Brazil. All the harvested tissues were immediately frozen by immersion in liquid nitrogen and stored at -80°C.

2.2.2 Total RNA extraction, purification and quality controls

Frozen tissue samples of 1.0 g were weighed and ground to fine powder in liquid nitrogen using a sterilized mortar and pestle. The fifteen samples from each tissue or organ were divided into three different RNA extractions (five samples for each extraction).

Figure 1 - Teak tissue and organ sample set. A= leaf, B= flower, C=seedling, D=root, E=stem secondary xylem, F=branch secondary xylem

Total RNA was extracted following a protocol developed for lignified tissues by Salzman et al. (1999). RNA quality assessment included purity (absence of protein and DNA) and integrity (absence of RNA degradation). 1 μ l of each extraction was analyzed spectrophotometrically using a Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., USA) and only RNA samples with 260/280 ratio between 1.9 and 2.1 and 260/230 ratio greater than 2.0 were used for subsequent analyses. The concentration of each sample was approximately 2 μ g/ μ l, so they were diluted to a final concentration of 1 μ g/ μ l and 4 μ g of total RNA from each sample was treated with DNase I (Promega). Then, 0.5 μ l of each treated sample was analyzed in agarose gels, all displaying clear bands corresponding to rRNA, absence of DNA and no degradation. In addition, PCR control reactions to examine for genomic DNA contamination were performed using total RNA without reverse transcription as template, and negative results (absence of bands) were assessed by electrophoresis on a 1% (w/v) agarose gel with ethidium bromide staining.

2.2.3 cDNA synthesis

Two cDNA samples were synthesized from the three extractions of each tissue or organ from 1.0 µg of the treated RNA using the SuperScriptTM III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer's instructions. Each cDNA sample concentration was determined using the Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., USA) to be approximately 2000 ng/µl. A concentration of 100 ng/µl (1:20 dilution) and 25 ng/µl (1:80 dilution) was used for PCR amplification and qRT-PCR expression experiments, respectively.

2.2.4 Multiple sequence alignments, PCR and qRT-PCR primer design

Primers (Table 1) were manually designed flanking the conserved domains of *RP60S*, *CAC*, *ACT*, *HIS3*, *SAND*, *B-TUB*, *UBQ*, and *EF-1α* after doing Clustal alignment (<http://www.ebi.ac.uk/Tools/msa/clustalw2>) of several orthologous plant sequences obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank>) to amplify by PCR those genes from teak leaf cDNA (Additional File 2). The eight amplified fragments gel electrophoresis were excised, purified with Fragment CleanUp® (Invisorb, USA) and inserted into the pJET1.2/Blunt vector from the CloneJetTM PCR Cloning Kit (Thermo Scientific, USA) following the manufacturer's recommendations. Plasmids were cloned in *DH5α*TM competent cells (Life Technologies, USA) and recombinant colonies were sequenced with the 3100 Genetic Analyzer (Applied Biosystems, USA) using pJET1.2/Blunt vector specific primers. Finally, sequences were blasted using blastx (<http://blast.ncbi.nlm.nih.gov>) to confirm their percentage amino acid similarity to the conserved domains and were translated (<http://web.expasy.org/translate/>) to amino acid sequences, which were submitted to PFAM search (<http://pfam.sanger.ac.uk/search>) (Sanger Institute, England) to confirm the presence of each gene's canonical protein domains. The primers for qRT-PCR were designed flanking the eight cloned teak sequences and *GAPDH* (Table 2) with OligoPerfectTM Designer (Life technologies, USA) with default parameters. Teak candidate reference genes, *TgCAD* target gene, NCBI accession numbers, qRT-PCR primer information and different parameters derived from qRT-PCR analysis are shown in Table 2.

2.2.5 Primer specificity, qRT-PCR Efficiency and R²

Confirmation of primer specificity was based on the dissociation curve at the end of each run. To determine the amplification efficiencies of the candidate genes, it was used

cDNA samples from the teak leaf with five dilutions to obtain the standard curve, and then the PCR efficiency for each gene was calculated according to the equation $(1+E)=10^{\text{slope}}$. The correlation coefficient (R^2) and slope values were obtained from the standard curve (Table 2).

Table 1 - Candidate reference genes, primers used to amplify in teak and their PCR parameters. Degenerate bases are indicated in bold

Gene symbol	Gene name	Primer sequences (5'-3') forward/reverse*	T _m (°C)	Amplicon (bp)
<i>RP60S</i>	<i>Ribosomal protein 60S</i>	ATGGTGAAGTTCTTGAAGCC/ TGGTTCTTTACCAAGCTC	55	399
<i>CAC</i>	<i>Clathrin adaptor complex</i>	AAGGATAACTTTGTCATTGT/ TGGGAAATACATGAAGGCG	58	794
<i>ACT</i>	<i>Actin</i>	GTTAGCAATTGGGATGATATGG/ ATCCAGACACTGTACTTCCT	57	797
<i>HIS3</i>	<i>Histone 3</i>	ACNGGTGGAGTGAAGAAGCC/ TCCTTGGGCATGATNGTNAC	61	275
<i>SAND</i>	<i>Sand family protein</i>	ATATATTCCAGATATGGAGATGA/ TAYATGAAATGCCAAAGTCCA	55	941
<i>B-TUB</i>	<i>B-Tubulin</i>	ACNCARCAAATGTGGGATGC/ TCCCCAGTGTACCARTGCAA	60	335
<i>UBQ</i>	<i>Ubiquitin</i>	TRACGGGNAAGACCATAAC/ ACCTTCTTNTTCTTGCTT	56	271
<i>EF-1α</i>	<i>Elongation factor 1- α</i>	CATCAACATTGTGGTCATTGG/ CCAGANCGCCTGTCAATCTTG	55	1095

*Degenerate nucleotides used in some primers: N=any base, Y=C or T, R=A or G.

Table 2 - Candidate reference genes, *TgCAD* target gene, specific qRT-PCR primers and different parameters derived from qRT-PCR analysis

Gene symbol	Accession Number	Primer sequences (5'-3') forward/reverse	T _m (°C)	Amplicon length (bp)	Primer efficiency	R ²
<i>TgRP60S</i>	JZ515972	AGAAGCAGGCGAAGAAATCA/ GTGGGCATGATGTGGTTGTA	75.9	70	91.4	0.998
<i>TgCAC</i>	JZ515973	ATCTTGTGGAAGAAATGGATGC/ TTCGCAAACAACAGAGTGAGAT	77.4	127	91.7	0.994
<i>TgACT</i>	JZ515974	TCCAGAAGAGCACCCAATTC/ CAGGGGCATTAAAGGTCTCA	77.9	100	91.6	0.995
<i>TgHIS3</i>	JZ515975	TGGCTTTGGAACCTCAAATC/ CCCTGGAACTGTTGCTCTTC	81.2	135	92.4	0.998
<i>TgSAND</i>	JZ515976	GCCCAAAAAGCATCTCTTCA/ TTGTGGTGAGCAAGATCAGG	77.1	187	108.5	0.987
<i>TgB-TUB</i>	JZ515977	CAAGATGAGCACGAAAGAAGTG/ CGGAACATCTCCTGTATCGAC	81.1	180	93.8	0.982
<i>TgUBQ</i>	JZ515978	CGGGTAAGACCATAACTCTGGA/ GTCGATTCCTTTGGATGTTGT	85.6	171	92.8	0.998
<i>TgEF-1α</i>	JZ515979	ACCACACCAAAATACTCCAAGG/ TGGACCTCTCAATCATGTTGTC	78.1	145	93.5	0.999
<i>TgGAPDH</i>	FN431983.1	GGCCACCTATGAGGAGATCA/ CCAAGATGCCCTTTAGCTTG	79.2	102	101.9	0.998
<i>TgCAD</i>	JZ515980	CGGCAAGGTCTACAAAGGAG/ GGCTGTTTATCGCTTGCTTC	78.8	200	98.4	0.993

2.2.6 Quantitative real-time reverse transcription PCR

The qRT-PCR mixture contained 5.0 µl of a 1:80 dilution of the six synthesized cDNAs from each tissue or organ, primers to a final concentration of 0.4 µM each, 12.5 µl of the SYBR Green PCR Master Mix (Applied Biosystems, USA) and PCR-grade water up to a total volume of 25 µl. Each gene reaction was performed in technical replicate. PCR reactions without template were also done as negative controls for each primer pair. The quantitative PCRs were performed employing the StepOnePlus™ System (Applied Biosystems, USA). All PCR reactions were performed under the following conditions: 2 min at 50°C, 2 min at 95°C, and 45 cycles of 15 s at 95°C and 1 min at 65°C in 96-well optical reaction plates (Applied Biosystems, USA). Leaf samples were used as calibrator to normalize the values between different plates.

2.2.7 Analysis of gene expression stability

Gene expression stability was evaluated by applying four different statistical approaches: geNorm (VANDESOMPELE et al., 2002), NormFinder (ANDERSEN; JENSEN; ØRNTØFT, 2004), Bestkeeper (PFAFFL et al., 2004) and Delta Ct (SILVER et al., 2006). qRT-PCR data was exported from the StepOnePlus™ System (Applied Biosystems, USA) into an Excel datasheet (Microsoft Excel 2003) as Raw Crossing Point data (Additional File 6) and those values were log transformed by the $2^{-\Delta Ct}$ method for further requirements. Each of these approaches generated a measure of reference gene stability, by which each gene was ranked. Venn diagrams were constructed with the Smartdraw® program.

2.2.8 Validation of reference genes

One gene of interest, putatively coding for a *cinammyl alcohol dehydrogenase* (CAD) (Table 2), an enzyme involved in lignin biosynthesis, one of the terminal steps of the phenylpropanoid pathway, was used to validate the best two reference genes. The relative expression level of the target gene was determined in leaves, and the lignified tissues of stem and branch secondary xylem of 60 year-old trees, stem of 1 year-old trees and branch secondary xylem of 12 year-old trees, expecting a higher expression level in younger tissues with a continuous secondary wall formation. The experimental procedure was the same used for the selection of the reference genes. Stem secondary xylem of 60 year-old tree samples were chosen as calibrator.

2.3 Results

Normalization of gene expression experiments, especially of qRT-PCR using a set of reference genes is currently a critical procedure when analyzing expression levels of target genes in different tissues or under different conditions. In the present study, nine potential reference genes for qRT-PCR of teak were assessed. A total of 36 cDNA samples including several organ types (leaf, root, and flower) and secondary xylem tissues from stems and branches of different ages were analyzed (Figure 1).

2.3.1 Identification and cloning of reference genes in teak

As teak does not have the relevant genetic sequence information available in databases, it was necessary to design degenerate primers to amplify, clone and sequence the reference genes according to the most common genes used for qRT-PCR analysis in trees such as *Platycladus orientalis* (CHANG et al., 2012), *Vernicia fordii* (HAN et al., 2012), *Quercus suber* (MARUM et al., 2012), *Populus euphratica* (WANG et al., 2014) and *Pyrus pyrifolia* (IMAI et al., 2014). *GAPDH* (FN431982.1) was the only teak sequence available in GenBank (<http://www.ncbi.nlm.nih.gov/genbank>). Therefore, we performed multiple nucleotide sequence alignment of the reference genes of different species for the remaining genes (Additional File 1). For each gene, at least four sequences were used in the alignments. Degenerate primers were designed to amplify the most conserved domains and at least 250 bp of the teak cDNA (Table 1, Additional File 2).

The degenerate primers were able to produce specific amplicons ranging from 271 to 1440 bp using cDNA of teak leaves as template (Table 1). After gel purification, PCR fragments were cloned into the pJET1.2/Blunt vector (Thermo Scientific, USA) and transformed into *DH5 α* TM competent cells (Life Technologies, USA). Recombinant colonies were selected to extract plasmid DNA for sequencing.

The teak nucleotide identities were checked by BLAST (ALTSCHUP et al., 1990) against NCBI non redundant sequences (<http://blast.ncbi.nlm.nih.gov>) and the results showed that all the clones contained the expected fragments. The most conserved genes were *TgACT* and *TgEF-1 α* with 92% of similarity, followed by *TgUBQ* and *Tg β -TUB* with 91% and 87%, respectively (data not shown). All genes showed at least 79% of similarity. Translated amino acid sequences were obtained by Expasy Translation Tool (<http://web.expasy.org/translate/>) and used to check for the presence of the expected domains in Pfam Database (<http://pfam.sanger.ac.uk>). Thereafter, teak amino acid sequences were compared against NCBI protein sequences with the algorithm tBLASTn (<http://blast.ncbi.nlm.nih.gov>). Results

of the *in silico* analysis showed that all teak putative protein sequences possess the predicted domains, presenting high similarity with the selected reference genes (Additional file 3). At protein level, the most conserved genes were *TgACT*, *TgEF-1 α* , *TgHIS3*, *Tg β -TUB* and *TgUBQ* presenting 99% of similarity (data not shown).

2.3.2 Primer specificity and PCR efficiency

Real-time PCR primers (Table 2, Additional File 4) were designed to amplify the teak sequences of the eight clones (*TgRP60S*, *TgCAC*, *TgACT*, *TgHIS3*, *TgSAND*, *Tg β -TUB*, *TgUBQ* and *TgEF-1 α*) and *TgGAPDH* (Table 2), and were used to detect transcript levels. Primer specificity was evaluated with a single peak in all ten melting curves (Figure 2) and as a single band in the agarose gel analysis (Additional File 5). qRT-PCR efficiency (E) varied from 91.4% for *TgRP60s* to 108.5% for *TgSAND* and correlation coefficients (R^2) oscillated from 98.2% for *TgB-TUB* to 99.9% for *TgEF-1 α* (Table 2). The acceptable range for PCR efficiencies calculated using standard curve serial dilution experiments is 90–110% (i.e. a slope between 3.1 and 3.58) (PFAFFL, 2004). The annealing temperature of 65°C was effective for all primers; nevertheless, its choice can impact on the efficiency of the reaction. Altogether, the results showed that the chosen primers accurately amplified the candidate reference genes.

To compare the differences in transcript levels between reference genes, the Cq range was determined and the coefficient of variance was calculated for each gene across all samples based on the interquartile range (25-75% percentiles). The average Cq values of the different genes ranged from 22 to 34 cycles (Figure 3, Table 3). *TgUBQ*, *Tg β -TUB* and *TgRP60S* showed the narrowest variance (lowest Cq dispersion), while *TgACT* and *TgCAC* exhibited widest variance (highest dispersion). The gene with the most abundant transcript level was *TgEF-1 α* while *TgCAC* was the least abundant, reaching mean threshold fluorescence with 23 and 31 amplification cycles, respectively.

2.3.3 Expression stability of the nine candidate reference genes.

To evaluate the reference genes' expression stability, four different methodologies were used: geNorm, NormFinder, BestKeeper and Delta Ct.

Figure 2 - Specificity of qRT-PCR amplification. Melting curves (dissociation curves) of the 10 amplicons (*RP60S*, *ACT*, *CAD*, *CAC*, *EF-1 α* , *GAPDH*, *HIS3*, *SAND*, *B-TUB*, *UBQ* genes) after the qRT-PCR reactions, all showing one peak

Figure 3 - Expression levels of candidate reference genes in different plant samples. Expression data displayed as Cq values for each reference gene in all *T. grandis* samples. The horizontal lines of the box indicate the 25th and 75th quartiles. The central horizontal line across the box is depicted as the median. The whisker caps represent the maximum and minimum values. Dots represent outliers. Genes are in order from the most (lower Cq, on the left) to the least abundantly expressed (higher Cq, on the right)

Table 3 - Descriptive statistics and expression level obtained by BestKeeper

Factor	<i>RP60S</i>	<i>SAND</i>	<i>ACT</i>	<i>CAC</i>	<i>GAPDH</i>	<i>B-TUB</i>	<i>HIS3</i>	<i>UBQ</i>	<i>EF-1α</i>
N	36	36	36	36	36	36	36	36	36
GM [CP]	28.59	29.92	26.21	31.05	25.63	29.59	28.24	26.77	23.45
AM [CP]	28.59	29.94	26.27	31.07	25.64	29.62	28.27	26.78	23.47
Min [CP]	27.52	28.08	24.15	28.71	24.21	28.00	26.31	25.99	22.50
Max [CP]	29.82	32.15	29.91	33.11	27.36	33.54	31.66	28.71	25.23
SD [$\pm$ CP]	0.51	0.90	1.42	0.99	0.64	1.21	1.02	0.60	0.73
CV [%CP]	1.78	3.01	5.40	3.20	2.49	4.10	3.62	2.23	3.12
Min [x-fold]	-1.99	-3.82	-3.84	-4.58	-2.72	-2.62	-3.53	-1.63	-1.79
Max [x-fold]	2.22	5.13	11.12	3.85	3.38	10.89	9.29	3.37	2.99
SD [$\pm$ x-fold]	1.39	1.79	2.50	1.90	1.51	2.19	1.94	1.47	1.61

Abbreviations: N: number of samples; CP: crossing point; GM [CP]: geometric CP mean; AM [CP]: arithmetic CP mean; Min [CP] and Max [CP]: CP threshold values; SD [$\pm$ CP]: CP standard deviation; CV [%CP]: variance coefficient expressed as percentage of CP level; Min [x-fold] and Max [x-fold]: threshold expression levels expressed as absolute x-fold over- or under-regulation coefficient; SD [$\pm$ x-fold]: standard deviation of absolute regulation coefficient

2.3.4 geNorm

geNorm was used to rank the reference genes by calculating the gene expression stability value M , which corresponds to the average pairwise variation (V) of a particular gene with all other control genes (VANDESOMPELE et al., 2002). The most stable reference gene has the lowest M value, while the least stable has the highest M value. To identify reference genes with stable expression, geNorm indicates genes with M values below the threshold of 1.5, however Vandesompele et al. (2002) suggests M values lower than 1.0 to ensure the selection of the most stable genes. When all 36 samples were analyzed together with geNorm (Figure 4), eight genes had $M < 1.0$, with *TgUBQ* and *TgEF-1 α* showing the highest expression stability ($M = 0.295$) in different tissues. *Act* was the only gene with $M > 1.0$, with the lowest expression stability of 1.035.

To obtain reliable results from qRT-PCR studies, two or more reference genes should be used for data normalization. The optimal number of reference genes can be determined by calculating the pairwise variation ($V_{n/n+1}$) using the geNorm algorithm (VANDESOMPELE et al., 2002). It is calculated between the two sequential normalization factors (NF), NF_n and NF_{n+1} , for all the samples analyzed. Slight variations mean addition of another gene has a low effect on the normalization. Vandesompele et al. (2002) proposed 0.15 as the cut-off value for V , below which the inclusion of an additional control gene is not required.

Figure 4 - Gene expression stability of the candidate reference genes calculated by different statistical methods. Ranking of each candidate reference gene (*RP60S*, *ACT*, *CAC*, *EF-1α*, *GAPDH*, *HIS3*, *SAND*, *B-TUB*, *UBQ*) calculated by NormFinder, BestKeeper, geNorm and Delta Ct methods, for all tested cDNA samples (leaf, flower, seedling, root, stem secondary xylem, branch secondary xylem)

This means that if $V_{n/n+1} < 0.15$, it is not necessary to use $\geq n+1$ reference genes for normalization. In this study, the paired variable coefficients indicated that the inclusion of the third reference gene (i.e. *TgUBQ*, *TgEF-1α* and *TgGAPDH*) would be useful for normalization when considering total samples and only lignified samples, whereas two stable reference genes (*TgGAPDH* and *TgRP60S*) can be employed when analyzing non-lignified tissues. However, if the samples of stem secondary xylem (the most lignified tissue) are excluded from the analysis, two reference genes (*TgUBQ* and *TgEF-1α*) would be optimal for normalizing gene expression (Figure 5).

Figure 5 - Pairwise variations (V) calculated by geNorm to determine the optimal number of reference genes. The average pairwise variations $V_{n/n+1}$ was analyzed between the normalization factors NF_n and NF_{n+1} to indicate the optimal number of reference genes required for qRT-PCR data normalization in all the samples, lignified tissues (root, stem secondary xylem, branch secondary xylem), non-lignified tissues (leaf, flower, seedling) and in all samples minus stem secondary xylem

2.3.5 NormFinder

NormFinder is a Microsoft Excel-based Visual Basic application that allows estimation of stability values of single candidate reference genes. The algorithm is based on intra- and inter-group variations and combines both results into a stability value for each candidate reference gene (ANDERSEN; JENSEN; ØRNTTOFT, 2004). The results of the NormFinder analysis were somewhat similar to those of geNorm. Both methods ranked *TgUBQ*, *TgEF-1α* and *TgRP60S* as among the four most stable reference genes (Figure 6) and *TgACT* and *TgHIS3* as the least stable (Figure 4). However, *Tgβ-TUB* emerged as the third most stably expressed using NormFinder, whereas it was ranked seventh by geNorm. These discrepant results could be explained due to inter-tissue expression variations detected by NormFinder analysis, which is not take account for gene stability calculations in the geNorm algorithm. When considering only intra-tissue variations, *Tgβ-TUB* was the most stable gene in lignified tissues (i.e. roots, branches and stems) (Table 4). However, in non-lignified tissues (leaves, flowers and seedlings) *Tgβ-TUB* was ranked eighth of nine genes, corroborating results obtained by geNorm.

Figure 6 - Gene expression differences among the candidate reference genes analyzed by NormFinder. Black circles represent the log-transformed gene expression levels. Vertical bars give a confidence interval for the inter-tissue variation. Top and bottom lines from the graphic represent the maximum standard deviation of the candidate reference genes, with the difference log expression levels between 0.2 and -0.2

Table 4 - NormFinder intragroup expression stability for teak candidate reference genes. Between parenthesis: ranking of stability

Gene	non-lignified *	lignified **	Total
<i>TgSAND</i>	0.024 (5)	0.015 (4)	0.039 (3)
<i>TgACT</i>	0.043 (7)	0.023 (6)	0.066 (7)
<i>TgCAC</i>	0.012 (2)	0.037 (8)	0.059 (6)
<i>TgGAPDH</i>	0.017 (3)	0.023 (6)	0.050 (5)
<i>TgB-TUB</i>	0.069 (8)	0.011 (1)	0.080 (8)
<i>TgHIS3</i>	0.113 (9)	0.037 (8)	0.150 (9)
<i>TgUBQ</i>	0.009 (1)	0.012 (2)	0.021 (1)
<i>TgEF-1α⁺</i>	0.017 (3)	0.014 (3)	0.031 (2)

* Column 1: non-lignified tissues (flower, leaf, seedling)

** Column 2: lignified tissues (root, branch, stem)

2.3.6 BestKeeper

The Bestkeeper software was adopted for descriptive analysis. The program is an Excel-based software tool that estimates gene expression stability based on the coefficient of correlation (r) between each reference gene and an index, defined as the geometric mean of all candidate reference gene Ct (or CP) values (PFAFFL et al., 2004). The BestKeeper also calculates the CP standard deviation (SD) and the coefficient of variance (CV) of each candidate gene. Reference genes with SD values >1 are considered not stable and should be avoided.

Results of analysis are shown in Table 3. Similarly to geNorm and NormFinder, BestKeeper ranked *ACT* and *HIS3* among the three least stable reference genes with SD values >1.0 (1.42 and 1.02 respectively). In addition, *Tgβ-TUB* presented unstable expression

with SD value of 1.21, being one of the least stable genes as observed in the geNorm analysis (Figure 4). The best reference genes are those that have the lowest coefficient of variance and standard deviation. In this study, *TgRP60S* and *TgUBQ* had CV $\pm$ SD values of 1.78 $\pm$ 0.51 and 2.23 $\pm$ 0.60, respectively, displaying a stable expression in all samples. The results of the BestKeeper analysis showed a similar pattern of stability to those obtained from geNorm, which described *TgUBQ*, *TgRP60S*, *TgEF-1 α* and *TgGAPDH* as the four best reference genes for the normalization of qRT-PCR data in teak. In the NormFinder analysis, *TgGAPDH* was replaced by *Tg β -TUB* among the top four regarding stability (Figure 4).

2.3.7 Delta Ct

The Delta Ct method is based on the ‘pairs of genes’ comparison using a simple Δ Ct approach (SILVER et al., 2006). The formula used in this method is similar to the standard comparative Ct method ($\Delta\Delta$ Ct) (LIVAK; SCHMITTGEN, 2001) except that no endogenous reference gene is incorporated since the purpose is to define stably expressed genes to normalize. In this approach, all pairs of genes are compared to each other and the genes are ranked according to the Δ Ct values, from lowest to highest (SILVER et al., 2006). As observed in geNorm and Bestkeeper analysis, *TgEF-1 α* , *TgUBQ* and *TgGAPDH* were the best reference genes, as they had the highest expression stability (lowest Δ Ct values) (Figure 4). As was detected in all used programs, *TgHIS3* and *TgACT* were the least stable internal controls for gene expression normalization.

2.3.8 Validation of *TgUBQ* and *TgEF-1 α* as internal controls to assess expression of the teak *cinnamyl alcohol dehydrogenase* gene in lignified tissues

The use of different reference genes to evaluate relative expression data has an important impact on the final normalized results. As *TgUBQ* and *TgEF-1 α* showed the best stability values by geNorm, Delta Ct, NormFinder and BestKeeper analyses (Figure 7), they were used to evaluate the transcript level of a gene of interest, the teak *cinnamyl alcohol dehydrogenase* (*TgCAD*). *TgCAD* was identified and cloned using the same methodologies described for the reference genes. To validate the selected reference genes, the transcript levels were quantified in leaves from four month-old greenhouse grown teak, and in lignified tissues and organs such as stem secondary xylem from 60 year-old trees, stem from 1 year-old plants and branch secondary xylem from 60 and 12 year-old teak trees. Results showed that no matter which gene is used (*TgUBQ* or *TgEF-1 α*), *TgCAD* expression decreased in the

following order: leaf > stem from 1 year-old plants > branch secondary xylem from 12 year-old trees > branch secondary xylem from 60 year-old trees > stem secondary xylem from 60 year-old trees (Figure 8). On the other hand, between tissues, leaf was the tissue with highest expression and stem and branch secondary xylem from 60 year-old teak trees were the tissues with lowest expression of the *TgCAD* gene.

Figure 7 - Venn diagrams. (a) the most stable reference genes present in the first four positions and (b) the least stable genes present in the last five positions identified by the NormFinder, BestKeeper, geNorm and Delta Ct methods. Diagrams were performed with the Smartdraw® program

Figure 8 - Expression levels of the *TgCAD* gene. It was used different tissues and organ ages of teak tree, using the best validated reference genes (*TgUBQ* and *TgEF-1a*) for normalization and the results are represented as mean fold changes in relative expression compared to stem secondary xylem from 60 years-old trees. Bars are mean standard deviation calculated from the 3 biological replicates

2.4 Discussion

In plant molecular biological research, qRT-PCR has improved the detection and quantification of expression profiles of target genes due to its sensitivity, specificity and

accuracy. For correct qRT-PCR measurements, reference genes are used as endogenous controls for gene expression normalization when analyzing the expression of genes of interest (CHANG et al., 2012; ZHU et al., 2013). Therefore, a careful choice of reference genes is essential to obtain an accurate quantification of the target gene transcript levels (HAN et al., 2012).

Currently, teak is one of the most important trees worldwide due to its wood's properties. Despite the growing importance of teak wood and extractives to the world market, the number of studies adopting techniques of modern biology for teak improvement is still quite limited. As far as has been documented, this is the first study of cloning and expression stability of qRT-PCR reference genes in teak tissues. In total, eight candidate genes (*TgRP60S*, *TgCAC*, *TgACT*, *TgHIS3*, *TgSAND*, *TgB-TUB*, *TgUBQ*, and *TgEF-1 α*) were successfully identified.

One of the challenges of studying gene expression in trees is to ensure good quality of total RNA isolated from stems and branches, which are woody tissues with high lignin contents. In this study, the use of the Salzman protocol (SALZMAN; FUJITA; HASEGAWA, 1999) for total RNA extraction, followed by a DNase I (Promega) treatment, provided high quality RNA from all lignified tissues and was chosen as standard method for RNA extraction.

The stability of nine reference genes for qRT-PCR normalization was assessed by four methodological approaches, geNorm, NormFinder, BestKeeper and Delta Ct method, in teak lignified and non-lignified tissues at different developmental stages. In spite of some inconsistencies that are usually observed between these methods (LIN; LAI, 2010; CHANG et al., 2012; MARUM et al., 2012), our results were quite constant regardless of the algorithm used for analysis. When considering the rank of four most and least stable genes, *TgUbq* and *TgEF-1 α* were selected among the most stable in all methods, while *TgACT*, *TgHIS3* and *TgCAC* showed the least expression stability. The only clear discrepancy within the results was the inclusion of *Tg β -TUB* in the most stable group by NormFinder and in the least stable group by the other programs (Figure 4). In the NormFinder intra-tissues analysis, *Tg β -TUB* was the most stable gene in lignified tissues, whereas it was ranked eighth of the nine genes in non-lignified tissues (Table 4). These results suggest that *Tg β -TUB* is a suitable reference gene in lignified tissues and could be used as internal control for quantifying gene expression in them.

Among recent studies in trees searching for suitable reference genes, control genes such as *ACT*, *UBQ*, *EF-1 α* , *α -TUB*, *CAC*, *SAND*, *B-TUB* were considered to be stable in various tissues and different conditions (CHANG et al., 2012; HAN et al., 2012; MARUM et al., 2012; IMAI et al., 2014; WANG et al., 2014). *ACT*, *UBQ* and *EF-1 α* were shown to be suitable reference genes for normalization in lignified tissues of *Vernicia fordii* (HAN et al., 2012) and *Quercus suber* (MARUM et al., 2012). In our analysis, the most stable genes were *UBQ* and *EF-1 α* (Figure 7).

Genes encoding elongation factor-1 α and ubiquitin are frequently considered consistent reference genes under different experimental conditions. *EF-1 α* has been found to be one of the most stable reference genes in several plants and conditions such as *Nicotiana tabacum* (SCHMIDT; DELANEY, 2010), *Lolium perenne* (LEE et al., 2010) and *Capsicum annuum* (BIN et al., 2012). The *Ubq* gene showed high stability for qRT-PCR normalization in *Platycladus orientalis* (CHANG et al., 2012) and *Brachypodium distachyon* (HONG et al., 2008). In combination, *UBQ* and *EF-1 α* showed stable expression across different tissues of *Vernicia fordii* (HAN et al., 2012) and *Dimocarpus longan* (LIN; LAI, 2010). However, *EF-1 α* and *UBQ* were the most variable reference genes in *Lycopersicum esculentum* (EXPÓSITO-RODRÍGUEZ et al., 2008) and *Euphorbia esula* (CHAO et al., 2012), respectively, suggesting that these genes might not be suitable for qRT-PCR normalization in some plants and/or conditions. Although *ACT* is one of the most commonly used reference gene in plants, in teak it showed low stability when assessed in different tissue samples and with different statistical methods. Similar results were observed in *Nicotiana tabacum* plants with viral infections (LIU et al., 2012) and *Glycine max* (LIBAULT et al., 2008).

Studies have shown that the expression of reference genes can vary significantly under different experimental conditions (BARSALOBRES-CAVALLARI et al., 2009; LIN; LAI, 2010). To mitigate these variations, the use of multiple reference genes to assess target gene expression is appropriate. geNorm analysis of paired variable coefficients suggested the inclusion of a third reference gene (i.e. *TgUBQ*, *TgEF-1 α* and *TgGAPDH*) for normalization when considering the total number of samples (Figure 5). Although the cut off value ≤ 0.15 is frequently used to confirm the optimal number of reference genes (VANDESOMPELE et al., 2002), this is not an absolute number because small datasets require fewer reference genes than larger ones and previous studies have reported proper normalization with higher cut-off values (LIU et al., 2012). In this study, the combination of the two most stable reference genes (*TgUBQ* and *TgEF-1 α*) to evaluate expression stability in all samples provided a coefficient of 0.17 and, thus, can be sufficient for the normalization of qRT-PCR data in teak,

especially considering that the use of more than two reference genes in large scale gene expression profiles will significantly increase the costs of analysis.

To validate the utility of *TgUBQ* and *TgEF-1a* as reference genes, the expression profile of *TgCAD* was assessed in teak leaves and lignified tissues collected from plants in the field at different development stages (Figure 8). *CAD* functions in one of the final steps of monolignol biosynthesis in the phenylpropanoid pathway and its study is essential to understand the lignin deposition and cell wall formation in trees. It catalyzes the NADPH-dependent reduction of cinnamyl aldehydes to cinnamyl alcohols prior to their transport to the secondary cell wall for polymerization into the lignin heteropolymer (TRABUCCO et al., 2013). In plants, *CAD* expression may vary according to tissue and development stage (BARAKAT et al., 2010). In addition, it has been shown that *CAD/CAD*-like genes are differentially expressed in plants infected with pests and pathogens (COELHO et al., 2006; BHUIYAN et al., 2009).

Using *TgUBQ* or *TgEF-1a* as reference genes, qRT-PCR results showed that *TgCAD* was strongly expressed in leaves (average 133-fold), followed by stems from 1 year-old plants (40-fold), branch secondary xylem from 12 year-old trees (24-fold), branch secondary xylem from 60 year-old trees (5-fold) and stem secondary xylem from 60 year-old trees (*calibrator*) (Figure 8). We observed higher expression of *TgCAD* in younger lignified tissues compared to older ones, probably due to less lignin deposition and secondary wall formation in 60 year-old trees. The *TgCAD* expression in lignified tissues and leaves presented the same pattern whichever internal control used, indicating that the reference genes identified in this study are suitable for qRT-PCR normalization in different tissues and plant ages

This is the first attempt to identify qRT-PCR reference genes in several teak tissues. These results suggest the use of *TgUBQ* and *TgEF-1a* as the best combination of reference genes for gene expression assessment in leaves, flowers, seedlings, roots, and lignified stem and branch secondary xylem of varying ages in teak. In addition, we recommend the researchers to validate the reference genes in their samples of interest before performing any experiment. The different tissues show that *TgACT* is not a suitable reference gene to normalize gene expression in this tree, highlighting the need to evaluate commonly used reference genes for particular species, conditions, tissues and organs. Finally, they advise that the use of reference genes without validation may reduce precision or produce misleading results.

2.5 Conclusions

To the best of our knowledge, this study is the first attempt at cloning, sequencing and evaluating a set of commonly used candidate reference genes for the normalization of gene expression analysis using qRT-PCR in teak. Our data showed that expression stability varied considerably among the nine genes tested in the different samples of teak tissues. Stability analysis using NormFinder, Bestkeeper, geNorm and Delta Ct showed that *TgUBQ* and *TgEF-1a* are the most stable genes across different tissues and organs, while *TgACT* was deemed to be unsuitable as a reference gene. *TgCAD* expression analyses confirmed *TgUBQ* and *TgEF-1a* stability for correct normalization in teak. Consequently, they can be used in future gene expression studies of target genes in different teak tissues.

Additional Files

Additional File 1 - Information related to the orthologous plant sequences used in this study

Gene	Species	GenBank ID
RP60S	<i>Populus trichoparpa</i>	XM_002300027.1
	<i>Arabidopsis thaliana</i>	NM_117587.2
	<i>Glycine max</i>	XM_003531057.1
	<i>Pisum sativum</i>	U10046.1
	<i>Ricinus communis</i>	XM_002513364.1
	<i>Vitis vinifera</i>	XM_002277389.2
CAC	<i>Vitis vinifera</i>	XM_002281392.1
	<i>Arabidopsis lyrata</i>	XM_002894613.1
	<i>Populus trichoparpa</i>	XM_002318903.1
	<i>Ricinus communis</i>	XM_002512492.1
	<i>Glycine max</i>	XM_003535990.1
ACT	<i>Populus trichoparpa</i>	XM_002308329.1
	<i>Arabidopsis lyrata</i>	XM_002882721.1
	<i>Arabidopsis thaliana</i>	NM_112046.3
	<i>Glycine max</i>	NM_001254249.1
	<i>Ricinus communis</i>	XM_002530665.1
	<i>Vitis vinifera</i>	XM_002279636.1
HIS3	<i>Populus trichoparpa</i>	XM_002306258.1
	<i>Gossypium hirsutum</i>	AF024716.1
	<i>Lycopersicon esculentum</i>	X83422.1
	<i>Zea mays</i>	EU976723.1
SAND	<i>Populus trichoparpa</i>	XM_002314230.1
	<i>Arabidopsis thaliana</i>	NM_128399.3
	<i>Picea sitchensis</i>	EF676351.1
	<i>Vitis vinifera</i>	XM_002285134.1
B-TUB	<i>Populus trichoparpa</i>	XM_002298000.1
	<i>Gossypium hirsutum</i>	AF521240.1
	<i>Medicago truncatula</i>	XM_003630465.1
	<i>Nicotiana tabacum</i>	EF051136.2
	<i>Ricinus communis</i>	XM_002509755.1
	<i>Theobroma cacao</i>	GU570572.1
	<i>Vitis vinifera</i>	XM_002273478.2
UBQ	<i>Populus trichoparpa</i>	XM_002320914.1
	<i>Hevea brasiliensis</i>	EF120638.1
	<i>Medicago truncatula</i>	XM_003629847.1
	<i>Nicotiana tabacum</i>	DQ138111.1
	<i>Pyrus communis</i>	AF386524.1
	<i>Ricinus communis</i>	XM_002515167.1
	<i>Solanum tuberosum</i>	L22576.1
EF-1α	<i>Populus trichoparpa</i>	EF147714.1
	<i>Arabidopsis thaliana</i>	NM_100666.3
	<i>Elaeis guineensis</i>	AY550990.1
	<i>Gossypium hirsutum</i>	DQ174254.1
	<i>Malus domestica</i>	AJ223969.1
	<i>Nicotiana paniculata</i>	AB019427.1
	<i>Prunus persica</i>	FJ267653.1
	<i>Vitis vinifera</i>	XM_002284888.1

Additional File 2 - Clustal alignments used for designing primers to amplify orthologous sequences in teak. Grey squares mean forward and reverse primers, respectively

A) *RP60S* gene. Species and accession numbers used: *Populus trichoparpa* (XM_002300027.1), *Arabidopsis thaliana* (NM_117587.2), *Glycine max* (XM_003531057.1), *Pisum sativum* (U10046.1), *Ricinus communis* (XM_002513364.1), *Vitis vinifera* (XM_002277389.2)

B) CAC gene. Species and accession numbers used: *Vitis vinifera* (XM_002281392.1), *Arabidopsis lyrata* (XM_002894613.1), *Populus trichoparpa* (XM_002318903.1), *Ricinus communis* (XM_002512492.1), *Glycine max* (XM_003535990.1)

(Continue)

		10	20	30	40	50	60	70	80	90	100
<i>VvCac</i>	1	-----	ATGTTGCAGTGTAT	TTTCTCTTCT	GATTCT	CGAGAGGTAA	TGCT	SAGAAACAGCT	AC	CGGACACCGG	TTGATCGATCCATATGTGA
<i>AlCac</i>	1	----	AGAGATGCT	CAATGCTATATT	CTCTATCTCC	GATTCT	CGAGAGGTAA	TGCTAGAGAA	CGCTTAC	SGGCTAT	CGCGTTGATCGATCCATATGTGC
<i>PtCac</i>	1	-----	ATGTTGCAGTGTATATT	TTCTTTTCAGATTCCGG	CGGAGTAA	TGCTAGAGAAACAGCTAA	TGGGCA	TAAGTAGAG	AGATCCAT	TTGTGC	
<i>RcCac</i>	1	-----	ATGTTGCAGTGTATATT	CTCTCAGATTCCGG	CGGAGGTAA	CTAGAGAA	CGAGCTAACTGGG	CACCGAGTAGACAGATCCAT	TTGTGA		
<i>GmCac</i>	1	AGCGAAAC	ATGTTGCAGTGTAT	TTTCTCTCTCAGATT	CCGGAGAGGTAA	TGCTAGAGAAACAGCTAA	TGGG	CACCGCTAGATCGG	TCCATATGTGC		
		110	120	130	140	150	160	170	180	190	200
<i>VvCac</i>	93	TTGGTTCTGGGAA	CAATGCT	TTCCCAAGT	TCATTACCC	AAGCT	CTCC	ACCAGTAA	TGCTTCCACCAAC	CATTACAT	TTTCCAAATTACTCGGAGGGGA
<i>AlCac</i>	97	TTGGTTCTGGGAA	CAATGCT	TTCTTCAAGG	CGATTCCCTTTAAG	TTCTTCCAGC	ATTGCTTCCACCAAC	CATTACAT	TTTCCAAAT	TTACTCGGAG	GGGGA
<i>PtCac</i>	93	TTGGTTCTGGGAA	CAATGCT	TTCTTCAAGG	CGATTCCCTTTAAGCAACA	TTAGTATTGCT	TCACC	SAGCCATTAC	TTTCCAAAT	TTGTCGG	GAGGGGA
<i>RcCac</i>	93	TTGGTTCTGGGAA	CAATGCT	TTCTTCAAGG	CGATTCCCTTTAAGCAACA	TTAGTATTGCT	TCACC	SAGCCATTAC	TTTCCAAAT	TTGTCGG	GAGGGGA
<i>GmCac</i>	101	CTGGTTCTGGGAA	CAAGCCATT	TTCTCAAC	CTGATTCCCTTTAAGCAACA	AACAGTTATTGCTT	CTCC	TACCATTAT	TTTCCAAAT	TTTCCG	GAGGGGA
		210	220	230	240	250	260	270	280	290	300
<i>VvCac</i>	193	ATCAC	TTTCTTACGCTTC	CAAGTTGAAATGCC	CTTTTATGCT	CATTAGT	TTCTTTGCAGAGTAGC	AGATGCTCT	CTC	CAATTA	CTTGAGGGGT
<i>AlCac</i>	197	ATCAC	TTTCTTACGCTTC	CAAGTTGAAATGCC	CTTTTATGCT	CATTAGT	TTCTTTGCAGAGTAGC	AGATGCTCT	CTC	CAATTA	CTTGAGGGGT
<i>PtCac</i>	193	ATCAC	TTTCTTACGCTTC	CAAGTTGAAATGCC	CTTTTATGCT	CATTAGT	TTCTTTGCAGAGTAGC	AGATGCTCT	CTC	CAATTA	CTTGAGGGGT
<i>RcCac</i>	193	ATCAC	TTTCTTACGCTTC	CAAGTTGAAATGCC	CTTTTATGCT	CATTAGT	TTCTTTGCAGAGTAGC	AGATGCTCT	CTC	CAATTA	CTTGAGGGGT
<i>GmCac</i>	201	ATCAC	TTTCTTACGCTTC	CAAGTTGAAATGCC	CTTTTATGCT	CATTAGT	TTCTTTGCAGAGTAGC	AGATGCTCT	CTC	CAATTA	CTTGAGGGGT
		310	320	330	340	350	360	370	380	390	400
<i>VvCac</i>	293	TGAATGAAGA	CTTGAT	CAAGGATAA	CTTTGCT	CTATGAGCT	TTGGATGAGATGATAGACAAC	GGCTTCCCTCT	CAACACAGAACCT	TAACAT	CT
<i>AlCac</i>	297	TGAATGAAGA	CTTGAT	CAAGGATAA	CTTTGCT	CTATGAGCT	TTGGATGAGATGATAGACAAC	GGCTTCCCTCT	CAACACAGAACCT	TAACAT	CT
<i>PtCac</i>	293	TGAATGAAGA	CTTGAT	CAAGGATAA	CTTTGCT	CTATGAGCT	TTGGATGAGATGATAGACAAC	GGCTTCCCTCT	CAACACAGAACCT	TAACAT	CT
<i>RcCac</i>	293	TGAATGAAGA	CTTGAT	CAAGGATAA	CTTTGCT	CTATGAGCT	TTGGATGAGATGATAGACAAC	GGCTTCCCTCT	CAACACAGAACCT	TAACAT	CT
<i>GmCac</i>	301	TGAATGAAGA	CTTGAT	CAAGGATAA	CTTTGCT	CTATGAGCT	TTGGATGAGATGATAGACAAC	GGCTTCCCTCT	CAACACAGAACCT	TAACAT	CT
		410	420	430	440	450	460	470	480	490	500
<i>VvCac</i>	393	AAGGACATGAT	CTCT	ACCAAAAT	TCTTAGCAAA	CTTTGCT	CTCTCTCACTGCTAAGAGTTC	TAATCTAGCA	ACACTCTTCCAGGG	CAACAGGA	
<i>AlCac</i>	397	AAGGACATGAT	CTCT	ACCAAAAT	TCTTAGCAAA	CTTTGCT	CTCTCTCACTGCTAAGAGTTC	TAATCTAGCA	ACACTCTTCCAGGG	CAACAGGA	
<i>PtCac</i>	393	AAGGACATGAT	CTCT	ACCAAAAT	TCTTAGCAAA	CTTTGCT	CTCTCTCACTGCTAAGAGTTC	TAATCTAGCA	ACACTCTTCCAGGG	CAACAGGA	
<i>RcCac</i>	393	AAGGACATGAT	CTCT	ACCAAAAT	TCTTAGCAAA	CTTTGCT	CTCTCTCACTGCTAAGAGTTC	TAATCTAGCA	ACACTCTTCCAGGG	CAACAGGA	
<i>GmCac</i>	401	AAGGACATGAT	CTCT	ACCAAAAT	TCTTAGCAAA	CTTTGCT	CTCTCTCACTGCTAAGAGTTC	TAATCTAGCA	ACACTCTTCCAGGG	CAACAGGA	
		510	520	530	540	550	560	570	580	590	600
<i>VvCac</i>	493	TCTTGTCTTCCATCGAGAA	CTCAGAGSC	CAAGCTATG	CAACATGAG	CTTTATCTT	GATCTT	CTCAAGAAATGGATGCA	CTATAAA	AGGGATCGGA	
<i>AlCac</i>	497	TCTTGTCTTCCATCGAGAA	CTCAGAGSC	CAAGCTATG	CAACATGAG	CTTTATCTT	GATCTT	CTCAAGAAATGGATGCA	CTATAAA	AGGGATCGGA	
<i>PtCac</i>	493	TCTTGTCTTCCATCGAGAA	CTCAGAGSC	CAAGCTATG	CAACATGAG	CTTTATCTT	GATCTT	CTCAAGAAATGGATGCA	CTATAAA	AGGGATCGGA	
<i>RcCac</i>	493	TCTTGTCTTCCATCGAGAA	CTCAGAGSC	CAAGCTATG	CAACATGAG	CTTTATCTT	GATCTT	CTCAAGAAATGGATGCA	CTATAAA	AGGGATCGGA	
<i>GmCac</i>	501	TCTTGTCTTCCATCGAGAA	CTCAGAGSC	CAAGCTATG	CAACATGAG	CTTTATCTT	GATCTT	CTCAAGAAATGGATGCA	CTATAAA	AGGGATCGGA	
		610	620	630	640	650	660	670	680	690	700
<i>VvCac</i>	593	TTTGGTGAAT	GTGAGAT	TATCG	GAAGT	SGAGTGA	AGCTCCCA	CTTTGCT	CTCTGATT	TACAT	CTTCAATTGGCAACCC
<i>AlCac</i>	597	TTTGGTGAAT	GTGAGAT	TATCG	GAAGT	SGAGTGA	AGCTCCCA	CTTTGCT	CTCTGATT	TACAT	CTTCAATTGGCAACCC
<i>PtCac</i>	593	TTTGGTGAAT	GTGAGAT	TATCG	GAAGT	SGAGTGA	AGCTCCCA	CTTTGCT	CTCTGATT	TACAT	CTTCAATTGGCAACCC
<i>RcCac</i>	593	TTTGGTGAAT	GTGAGAT	TATCG	GAAGT	SGAGTGA	AGCTCCCA	CTTTGCT	CTCTGATT	TACAT	CTTCAATTGGCAACCC
<i>GmCac</i>	601	TTTGGTGAAT	GTGAGAT	TATCG	GAAGT	SGAGTGA	AGCTCCCA	CTTTGCT	CTCTGATT	TACAT	CTTCAATTGGCAACCC
		710	720	730	740	750	760	770	780	790	800
<i>VvCac</i>	693	TGATGTGAGATT	CATCC	TGTGTCGGTT	TCG	CCATGGGAAT	CAATAC	ATTCT	TCATTGTG	CCCTCCTGATGGA	CACTTAAAGCTCATGAGTTAC
<i>AlCac</i>	697	TGATGTGAGATT	CATCC	TGTGTCGGTT	TCG	CCATGGGAAT	CAATAC	ATTCT	TCATTGTG	CCCTCCTGATGGA	CACTTAAAGCTCATGAGTTAC
<i>PtCac</i>	693	TGATGTGAGATT	CATCC	TGTGTCGGTT	TCG	CCATGGGAAT	CAATAC	ATTCT	TCATTGTG	CCCTCCTGATGGA	CACTTAAAGCTCATGAGTTAC
<i>RcCac</i>	693	TGATGTGAGATT	CATCC	TGTGTCGGTT	TCG	CCATGGGAAT	CAATAC	ATTCT	TCATTGTG	CCCTCCTGATGGA	CACTTAAAGCTCATGAGTTAC
<i>GmCac</i>	701	TGATGTGAGATT	CATCC	TGTGTCGGTT	TCG	CCATGGGAAT	CAATAC	ATTCT	TCATTGTG	CCCTCCTGATGGA	CACTTAAAGCTCATGAGTTAC
		810	820	830	840	850	860	870	880	890	900
<i>VvCac</i>	793	AGGGT	CAAAAGTTGA	SGAGT	ACCCGATATAT	TGTAAGCC	CAAGTAA	ATCATCAGA	CTGGGACATG	CGAC	TCAGTGTCTTGGTTGGCATA
<i>AlCac</i>	797	AGGGT	CAAAAGTTGA	SGAGT	ACCCGATATAT	TGTAAGCC	CAAGTAA	ATCATCAGA	CTGGGACATG	CGAC	TCAGTGTCTTGGTTGGCATA
<i>PtCac</i>	793	AGGGT	CAAAAGTTGA	SGAGT	ACCCGATATAT	TGTAAGCC	CAAGTAA	ATCATCAGA	CTGGGACATG	CGAC	TCAGTGTCTTGGTTGGCATA
<i>RcCac</i>	793	AGGGT	CAAAAGTTGA	SGAGT	ACCCGATATAT	TGTAAGCC	CAAGTAA	ATCATCAGA	CTGGGACATG	CGAC	TCAGTGTCTTGGTTGGCATA
<i>GmCac</i>	801	AGGGT	CAAAAGTTGA	SGAGT	ACCCGATATAT	TGTAAGCC	CAAGTAA	ATCATCAGA	CTGGGACATG	CGAC	TCAGTGTCTTGGTTGGCATA
		910	920	930	940	950	960	970	980	990	1000
<i>VvCac</i>	893	ATCCTGGAAA	AGATTTGAC	TCA	TAAGCT	CTCA	TTCCAACT	ACCTCTTGATT	TATCAGAA	ATCTG	CTTCAATCATGGAAACAGT
<i>AlCac</i>	897	ATCCTGGAAA	AGATTTGAC	TCA	TAAGCT	CTCA	TTCCAACT	ACCTCTTGATT	TATCAGAA	ATCTG	CTTCAATCATGGAAACAGT
<i>PtCac</i>	893	ATCCTGGAAA	AGATTTGAC	TCA	TAAGCT	CTCA	TTCCAACT	ACCTCTTGATT	TATCAGAA	ATCTG	CTTCAATCATGGAAACAGT
<i>RcCac</i>	893	ATCCTGGAAA	AGATTTGAC	TCA	TAAGCT	CTCA	TTCCAACT	ACCTCTTGATT	TATCAGAA	ATCTG	CTTCAATCATGGAAACAGT
<i>GmCac</i>	901	ATCCTGGAAA	AGATTTGAC	TCA	TAAGCT	CTCA	TTCCAACT	ACCTCTTGATT	TATCAGAA	ATCTG	CTTCAATCATGGAAACAGT
		1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
<i>VvCac</i>	993	TGC	CAAT	AAGACCT	GTCT	TGGTCC	ATTGGG	CGAAT	CCCAAG	GAATAAGCC	CCCTTCA
<i>AlCac</i>	997	TGC	CAAT	AAGACCT	GTCT	TGGTCC	ATTGGG	CGAAT	CCCAAG	GAATAAGCC	CCCTTCA
<i>PtCac</i>	993	TGC	CAAT	AAGACCT	GTCT	TGGTCC	ATTGGG	CGAAT	CCCAAG	GAATAAGCC	CCCTTCA
<i>RcCac</i>	993	TGC	CAAT	AAGACCT	GTCT	TGGTCC	ATTGGG	CGAAT	CCCAAG	GAATAAGCC	CCCTTCA
<i>GmCac</i>	1001	TGC	CAAT	AAGACCT	GTCT	TGGTCC	ATTGGG	CGAAT	CCCAAG	GAATAAGCC	CCCTTCA

B) CAC gene. Species and accession numbers used: *Vitis vinifera* (XM_002281392.1), *Arabidopsis lyrata* (XM_002894613.1), *Populus trichoparpa* (XM_002318903.1), *Ricinus communis* (XM_002512492.1), *Glycine max* (XM_003535990.1)

(Conclusion)

C) *ACT* gene. Species and accession numbers used: *Populus trichoparpa* (XM_002308329.1), *Arabidopsis lyrata* (XM_002882721.1), *Arabidopsis thaliana* (NM_112046.3), *Glycine max* (NM_001254249.1), *Ricinus communis* (XM_002530665.1), *Vitis vinifera* (XM_002279636.1)

(Continue)

Figure 1 displays a multiple sequence alignment of the PtACT, AlACT, AtACT, GmACT, RcACT, and VvACT genes. The alignment is shown in a grid format, with the gene names listed on the left and the amino acid positions (1 to 1000) indicated at the top. The alignment is color-coded by amino acid type: A (green), C (blue), G (red), T (black), and others (grey). A black box highlights a conserved region from position 440 to 450. A scale bar at the top indicates positions from 10 to 1000. The alignment shows high conservation across the genes, particularly in the highlighted region.

[illegible]

D) *HIS3* gene. Species and accession numbers used: *Populus trichoparpa* (XM_002306258.1), *Gossypium hirsutum* (AF024716.1), *Solanum lycopersicon* (X83422.1), *Zea mays* (EU976723.1)

[illegible]

E) *SAND* gene. Species and accession numbers used: *Populus trichoparpa* (XM_002314230.1), *Arabidopsis thaliana* (NM_128399.3), *Picea sitchensis* (EF676351.1), *Vitis vinifera* (XM_002285134.1)

(Conclusion)

		1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
<i>PtSand</i>	1026	ACAAAGTC	CTTAATAGATGTTT	GAGAAAAATCC	CAAGTTTGATATGAC	CCATTGCTTGGAGGAC	CGATGTGTCTTCTCATCTCTCATCCATT	CA			
<i>AtSand</i>	907	ACAAATCAATAGAC	AGATGTTTGA	AAAGAAATCCAAAGTT	GATATGACACCCCTGCTTGGAGG	CGACATAT	TGCTCTTCTCATCTCTCTCCATT	CA			
<i>PsSand</i>	1184	ACAAAGTC	CAATAGATATCT	TTGAAAAAATCCAAAGTT	TGATATGACACCCCTT	TGGAGGAC	CGGATGTGTCTTCTCTCTCTATCATCT	CTT			
<i>VvSand</i>	853	ACAAAGTC	AGTAAATAGATGTTT	GAGAAATCCAAAGTT	TGATATGAC	CCATTGCTTGGAGGAC	AGATGTGTCTTCTCTCTCTCATCTCATCTCT				
		1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
<i>PtSand</i>	1126	TTAGTTGGAATCCAGCAACATTTCT	CATGCATCAACATTTGCT	TCCCTTGCCTTATG	GACAC	AGGCAAGCTGCAGGTGCTATACGTGA	ATGATGTGCGTA				
<i>AtSand</i>	1007	TTAGTTGGAACCCGCTACTTCTTCATGCT	CTACTGTTCTTCCCTTCCATATGCTT	TGAGCAAGCTACAGGA	CCATATTTGCAAGATTTGGCGC						
<i>PsSand</i>	1284	TCAGTTGGAATCCAGCAACATAT	CTCATGCATATATCTCT	TCCCTTGCCTTCC	CCATCCAAAGCAAGCTGCAGGACGATAT	CTCC	CAAGATGTGCGGA				
<i>VvSand</i>	953	TCATTTGGAACCCGCTACTTCTTCATGCATACACCTGCT	TCCCTTGCCTTATGCGACAAGGCAAGCT	TCAGGTGCGCATATTA	CAAGATGTGCTGTA						
		1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
<i>PtSand</i>	1226	TTCTGGTGTCTCTTTTGGCAATAT	TAATGTGCAAAACAAAT	CTTCTTAGTCTTCTTCTGGCG	CAAAAAGCTTC	TCTTCTCATCTCTGATCACATCTCGCTACTCT					
<i>AtSand</i>	1107	GTCCTGGTGTCTTATTCTCATCTACT	TAATGTGCGACACACAAGGTTGTCAGTCTTG	TGGTGACACA	AAAGCTCTCTCTCATCC	GATGACATGCTCTTCTACTCT					
<i>PsSand</i>	1384	TTCTGGTGTCTTATTGTAATCT	GTGCGACACA	AAGGTTATCAGCTT	TTGGCGCAAAAGCAATCT	TCATCCATGATGACATGCTTTTACTT					
<i>VvSand</i>	1053	TTGAGGCTCCCTTTTGGCAATACTAATGTG	CAACACAAAGTGATTAGTCTT	TGGTGCGACAAAAGCAATCTCTCTCT	CCTGATGATG	TATGCTGCTCTCTT					
		1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
<i>PtSand</i>	1326	TCGAACCTCAATATGCTTCAGAAATCATTTAGG	CAATTAATGGACCTTC	ATTAAGATCTCT	TTATCCCATTCGACATCTGCAATCTTTCTC						
<i>AtSand</i>	1207	TCAAATTTTGTCTATGCTCAGATTCATT	CAGGACATCAAGAATCTTCTCACC	ATTCGCTTCCAAAGATACAAAG	TCAGGCTTCTTCTGATGCTATG						
<i>PsSand</i>	1484	TCAAATTTTGTCTCTCATCGAATCTTTCAGACATCAGAAATCT	TTTCTGCGCAATCTTCCCAATCT	TGCGCAATCTCAATCTGCGCATCTCTTATGCTTATG							
<i>VvSand</i>	1153	TCGAACCTTTGTCTATGCTTCGAATCATT	TAGGACATCGAATCTTTCTCACC	ATTTTGGCTTCCCAAGATATAATCCATGGCATT	TTTATGCTTATG						
		1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
<i>PtSand</i>	1418	GCCAAATTTGCCCTCCAGATATACCAAGTGGCATTTT	GTATGCTTATGTCCTTATCTTGATGTT	GACACATCTCTCTTATCCGTTATTCGTAAT	TCGAATGCTT	CGAATGCTT					
<i>AtSand</i>	1307	TCGCAATTTCTTTGATGATGATACATATG	ATAATGTGCTTACACACCTTCAGATGCT	CTTCATCATCTCAAGATTTCTGAGG	CTGCTT	CGCTTGAGCTCTT					
<i>PsSand</i>	1584	TCGCAATTTCTTTGATGATGATACATATG	ATAATGTGCTTACACACCTTCAGATGCT	CTTCATCATCTCAAGATTTCTGAGG	CTGCTT	CGCTTGAGCTCTT					
<i>VvSand</i>	1253	TCGCAATTTCTTTGATGATGATACATATG	ATAATGTGCTTACACACCTTCAGATGCT	CTTCATCATCTCAAGATTTCTGAGG	CTGCTT	CGCTTGAGCTCTT					
		1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
<i>PtSand</i>	1518	CTTTTGAAGTCAAATGTTCTT	TAGCGAAGTTCAGAGGTCATGCTGGATGTTGGGATGCA	TGTTGAGGATTTGCGCTG	CCGATCCCATGTCCTGCTCTGGAT						
<i>AtSand</i>	1405	CTTCTCAAGTCAAATATCTTCAAGTCT	TGTTTCAAGTCAATCCGGGAT	GTGGGATGTTGGAATGAT	TACCAATGATGCAATCCGAGGGGCTCG						
<i>PsSand</i>	1682	CTTAGTCAAATCAAATGTTCTT	AGTCAAGTGGTTCAGAGGTCATGCTAGATGTTG	CTCTGCTGTTGGAGG	CCACTCTTGATGCAACATCTGCTGCTGAT						
<i>VvSand</i>	1351	CTTTTGAAGTCAAATGTTCTT	TAGCGAAGTTCAGAGATCC	TGCTAGATGTTGGGATGCGGTGTGGAGGATTTGCGCTT	TGTAATCTCTCTGCTCTGCTGTA						
		1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
<i>PtSand</i>	1618	CTGCTCTGCTCTCATTTTGGGGGACATCAGGA	CCGACGATCTCT	CTAGGCAATATAGGGAA	CCATATGCTGG	ATTGGTGTCTGCTGGACTTTG					
<i>AtSand</i>	1496	CTGCTCTGCTCTCATTTTGGGGGACATCAGGA	CCGACGATCTCT	CTAGGCAATATAGGGAA	CCATATGCTGG	ATTGGTGTCTGCTGGACTTTG					
<i>PsSand</i>	1782	CTGCTCTGCTCTCATTTTGGGGGACATCAGGA	CCGACGATCTCT	CTAGGCAATATAGGGAA	CCATATGCTGG	ATTGGTGTCTGCTGGACTTTG

F) *B-TUB* gene. Species and accession numbers used: *Populus trichoparpa* (XM_002298000.1), *Gossypium hirsutum* (AF521240.1), *Medicago truncatula* (XM_003630465.1), *Nicotiana tabacum* (EF051136.2), *Ricinus communis* (XM_002509755.1), *Theobroma cacao* (GU570572.1), *Vitis vinifera* (XM_002273478.2)

(Conclusion)

Figure 1. Multiple sequence alignment of the tubulin gene from *PtTub*, *GhTub*, *MtTub*, *NtTub*, *RcTub*, *TcTub*, and *VvTub*. The alignment is shown in blocks of 100 amino acids, with positions 1 to 1000 indicated at the top. The sequences are color-coded: *PtTub* (black), *GhTub* (grey), *MtTub* (light grey), *NtTub* (white), *RcTub* (light blue), *TcTub* (light green), and *VvTub* (light yellow). The alignment shows high conservation across all species, with some variations highlighted in red. The alignment is presented in a grid format, with the species names on the left and the amino acid positions at the top. The alignment is color-coded to show conserved regions and variations. The alignment is presented in a grid format, with the species names on the left and the amino acid positions at the top. The alignment is color-coded to show conserved regions and variations.

- G) *UBQ* gene. Species and accession numbers used: *Populus trichoparpa* (XM_002320914.1), *Hevea brasiliensis* (EF120638.1), *Medicago truncatula* (XM_003629847.1), *Nicotiana tabacum* (DQ138111.1), *Pyrus communis* (AF386524.1), *Ricinus communis* (XM_002515167.1), *Solanum tuberosum* (L22576.1)

		10	20	30	40	50	60	70	80	90	100									
<i>PtUbq</i>	1		-----GCT	CTTTATACAAAACA	-----CATCA	AGCGAA	-----CCGCTGT	CCCTC	CTGCTCTG	CACTCCAAA										
<i>HbUbq</i>	1	-----	ACGCGGGGA	CGTTCAACCATTC	AAACCTTA	SCCAAA	CCCTCTT	CTGCTCTG	TTTCTG	CACTC	-----									
<i>MtUbq</i>	1	-----	-----	GAAC	CTAGAAGATAAATTC	AGAAAC	CTATG	SCCTTCA	CTCCAA	CACTTC	TTTGTGAGAG									
<i>NtUbq</i>	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----									
<i>PcUbq</i>	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----									
<i>RcUbq</i>	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----									
<i>StUbq</i>	1	CGAAGAAAAGGGCT	TGTAAACCCCTA	TAAAGTGGCA	TGGCAGAGT	TTACACTCTCAT	TTCCATCA	AAAGAAACCT	TAAAGCCGCA	CGGC	ACTGAT									
		110	120	130	140	150	160	170	180	190	200									
<i>PtUbq</i>	65	AACCC	AAG-AGGCG	-----	AGATGCAGATCTT	CGTGA	ACCTT	ACCGGGA	AAGACCATAA	CTC	GAGGTGAGTG	TCGACACATCGACAATGTC								
<i>HbUbq</i>	67	GGCAT	CAA	-----	AGATGCAGATCTT	CGTGA	ACCTT	ACCGGGA	AAGACCATAA	CTC	GAGGTGAGTG	TCGACACATCGACAATGTC								
<i>MtUbq</i>	69	TTG	CAGCAACCAACGA	GC	AGATGCAGATCTT	CGTGA	ACCTT	ACCGGGA	AAGACCATAA	CTC	GAGGTGAGTG	TCGACACATCGACAATGTC								
<i>NtUbq</i>	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----								
<i>PcUbq</i>	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----								
<i>RcUbq</i>	50	CTG	CAT	CGT	GGCGGA	-----	AGATGCAGATCTT	CGTGA	ACCTT	ACCGGGA	AAGACCATAA	CTC	GAGGTGAGTG	TCGACACATCGACAATGTC						
<i>StUbq</i>	101	TTCTCT	CTCCAGGCG	-----	AGATGCAGATCTT	CGTGA	ACCTT	ACCGGGA	AAGACCATAA	CTC	GAGGTGAGTG	TCGACACATCGACAATGTC	-----	-----						
		210	220	230	240	250	260	270	280	290	300									
<i>PtUbq</i>	161	AAGGCCAAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA						
<i>HbUbq</i>	156	AAGGCCAAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA						
<i>MtUbq</i>	169	AAGGCCAAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA						
<i>NtUbq</i>	79	AAGGC	AAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA					
<i>PcUbq</i>	84	AAGGCCAAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA						
<i>RcUbq</i>	147	AAGGCCAAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA						
<i>StUbq</i>	198	AAGGCCAAGAT	CCAGGACAAGGAG	GGCATCCC	CCGACCAAC	CAGCGG	CTCAT	TTTCGCT	GGAAAC	CACTC	GAGGACGGG	CCGACCTC	GC	GACTACA						
		310	320	330	340	350	360	370	380	390	400									
<i>PtUbq</i>	261	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>HbUbq</i>	256	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>MtUbq</i>	269	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>NtUbq</i>	179	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>PcUbq</i>	184	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>RcUbq</i>	247	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>StUbq</i>	298	ACATCCAGAAGGAGT	CCACTCTCCAC	CTGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
		410	420	430	440	450	460	470	480	490	500									
<i>PtUbq</i>	361	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>HbUbq</i>	356	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>MtUbq</i>	369	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>NtUbq</i>	279	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>PcUbq</i>	284	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>RcUbq</i>	347	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
<i>StUbq</i>	398	CAAGAAGAAGAGGT	TAAGTCCGCTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG					
		510	520	530	540	550	560	570	580	590	600									
<i>PtUbq</i>	461	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
<i>HbUbq</i>	456	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
<i>MtUbq</i>	469	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
<i>NtUbq</i>	379	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
<i>PcUbq</i>	384	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
<i>RcUbq</i>	447	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
<i>StUbq</i>	498	GGTGTGGG	ACTTTTCATGGT	TAATCATTTT	GATAGGCACT	ACTGTTGGT	TAAGTGTGGG	CTTACCTAT	GTTTACCAGA	GGCTGGTGGTGA	-----	TTAG	GCTCA							
		610	620	630	640	650	660	670	680	690	700									
<i>PtUbq</i>	559	TTTGAATG	CGCTTTT	CTT	TAAGT	TGGAACAAGGAT	-----	TAATTTATTT	TAGACAT	GGTTG	-----	CTG	TT	GAACCTTA	-----	TAGTTATGTTT	TATTCGCATT			
<i>HbUbq</i>	552	ATTGA	AT	CTCCCGCA	TTTTCGCTT	TAAAGGTT	TTT	GAGATG	CTTT	TATTTAT	TAACTAC	CGGTG	GACCT	CT	TAAC	CTTT	CTGTT	CGATT	ATGTTT	CT
<i>MtUbq</i>	569	TCAA	-----	TCATG	CTT	TAAGGTTT	-----	TCGTTT	TGTTAT	CTTT	TGTT	TAAGT	-----	TTT	TAAGT	TTT	TAAGT	TTT	TAAGT	TTT
<i>NtUbq</i>	471	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>PcUbq</i>	476	GA	AAATTT	GAGTTT	TAATTT	TCGAATTATAT	CGCCATGGAT	TTT	TAATTT	TGATGCTTT	TAGTGTG	TTT	TAAGT	TTT	TAAGT	TTT	TAAGT	TTT	TAAGT	TTT
<i>RcUbq</i>	544	ACAAAT	TTAACTCC	TTT	TAGATCTAT	GTGATCTAT	TTT	TAGATCTAT	GTGATCTAT	TTT	TAGATCTAT	GTGATCTAT	TTT	TAGATCTAT	GTGATCTAT	TTT	TAGATCTAT	GTGATCTAT	TTT	TAGATCTAT
<i>StUbq</i>	596	AG	-----	TTAGCAAT	CTTT	TATCAG	TTTCTT	TTT	TGTCGACGG	TAATTT	TAG	-----	GT	TTT	TTT	TTT	TTT	TTT	TTT	TTT
		710	720	730	740	750	760	770	780	790	800									
<i>PtUbq</i>	653	GG	ATTTT	GAGTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT
<i>HbUbq</i>	650	GG	ATTTT	GAGTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT	TAATTT
<i>MtUbq</i>	662	TT	ATATAT	AT	CTC	ATATAT	AT	CTC	ATATAT	AT	CTC	ATATAT	AT	CTC	ATATAT	AT	CTC	ATATAT	AT	CTC
<i>NtUbq</i>	471	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>PcUbq</i>	574	AT	GA	-----	GGCTT	TAGT	TTG	TTT	CAAT	TAAT	CCAGAT	GG	AAATG	TTT	TAAT	CCAGAT	GG	AAATG	TTT	TAAT
<i>RcUbq</i>	644	TGG	AT	GGT	ACTT	CT	TTT	TGA	AGTT	TAAT	GG	AT	GGT	ACTT	CT	TTT	TGA	AGTT	TAAT	GG
<i>StUbq</i>	688	GT	GA	CAACCC	CT	GGG	AT	GGT	ACTT	CT	TTT	TGA	AGTT	TAAT	GG	AT	GGT	ACTT	CT	TTT

H) *EF-1α* gene. Species and accession numbers used: *Populus trichoparpa* (EF147714.1), *Arabidopsis thaliana* (NM_100666.3), *Elaeis guineensis* (AY550990.1), *Gossypium hirsutum* (DQ174254.1), *Malus domestica* (AJ223969.1), *Nicotiana paniculata* (AB019427.1), *Prunus persica* (FJ267653.1), *Vitis vinifera* (XM_002284888.1)

(Continue)

Figure 1 displays a multiple sequence alignment of the EflA proteins from various species, including *P. tetrax* (*PteEflA*), *A. tetrax* (*AteEflA*), *E. laevis* (*ElgEflA*), *G. hirsuta* (*GheEflA*), *M. domestica* (*MdeEflA*), *N. peponis* (*NpeEflA*), *P. perfringens* (*PpeEflA*), and *V. vulnificus* (*VveEflA*). The alignment is shown in blocks of 100 amino acids, with positions 1 to 1000 indicated at the top and bottom. A black box highlights a conserved region from position 120 to 140. The alignment shows high conservation across all species, particularly in the regions highlighted by the black box and the scale bar.

H) *EF-1a* gene. Species and accession numbers used: *Populus trichoparpa* (EF147714.1), *Arabidopsis thaliana* (NM_100666.3), *Elaeis guineensis* (AY550990.1), *Gossypium hirsutum* (DQ174254.1), *Malus domestica* (AJ223969.1), *Nicotiana paniculata* (AB019427.1), *Prunus persica* (FJ267653.1), *Vitis vinifera* (XM_002284888.1)

(Conclusion)

- B) CAC gene. Species and translated sequences used: *Tectona grandis* (JZ515973), *Vitis vinifera* (XM_002281392.1), *Arabidopsis lyrata* (XM_002894613.1), *Populus trichocarpa* (XM_002318903.1), *Ricinus communis* (XM_002512492.1), *Glycine max* (XM_003535990.1)

		10	20	30	40	50	60	70	80	90	100						
<i>TgCac</i>	1	-----	PGATS	SCVPWR	TDPR	ASNEVYVDLVEEMDAI	INRDGL	LVKCEIYGEVQVNA	HLSGLPDLTL	FANPSIL	NDVRFHPCVR	RPWESNQIL					
<i>VvCac</i>	1	-----	NSSNVSDTLPGATAS	SCVPWR	TDPR	ASNEVYVDLVEEMDAI	INRDGL	LVKCEIYGEVQVNSHL	SGLPDLTL	SFANPSIL	NDVRFHPCVR	RPWESNQIL					
<i>AlCac</i>	1	-----	NSSNVSDTLPGATAS	SCVPWR	TDPR	ASNEVYVDLVEEMDAI	INRDGL	LVKCEIYGEVQVNSHL	SGLPDLTL	SFANPSIL	NDVRFHPCVR	RPWESNQIL					
<i>PtCac</i>	1	-----	NSSNVSDTLPGATAS	SCVPWR	TDPR	ASNEVYVDLVEEMDAI	INRDGL	LVKCEIYGEVQVNSHL	SGLPDLTL	SFANPSIL	NDVRFHPCVR	RPWESNQIL					
<i>RcCac</i>	1	-----	NSSNVSDTLPGATAS	SCVPWR	TDPR	ASNEVYVDLVEEMDAI	INRDGL	LVKCEIYGEVQVNSHL	SGLPDLTL	SFANPSIL	NDVRFHPCVR	RPWESNQIL					
<i>GmCac</i>	1	-----	NSSNVSDTLPGATAS	SCVPWR	TDPR	ASNEVYVDLVEEMDAI	INRDGL	LVKCEIYGEVQVNSHL	SGLPDLTL	SFANPSIL	NDVRFHPCVR	RPWESNQIL					
		110	120	130	140	150	160	170	180	190	200						
<i>TgCac</i>	92	SFVPPDGL	NLM	SYRVKKL	KSTPIYVKPOL	TS	SGTCRIS	VLVGIRND	PGKTIDS	ITVQF	RLP	PCVLS	SDLSSN	CGAVNVLA	KTC	SW	IGRIPKDKAPS
<i>VvCac</i>	101	SFVPPDGL	NLM	SYRVKKL	KSTPIYVKPOL	TS	SGTCRIS	VLVGIRND	PGKTIDS	ITVQF	RLP	PCVLS	SDLSSN	CGAVNVLA	KTC	SW	IGRIPKDKAPS
<i>AlCac</i>	101	SFVPPDGL	NLM	SYRVKKL	KSTPIYVKPOL	TS	SGTCRIS	VLVGIRND	PGKTIDS	ITVQF	RLP	PCVLS	SDLSSN	CGAVNVLA	KTC	SW	IGRIPKDKAPS
<i>PtCac</i>	101	SFVPPDGL	NLM	SYRVKKL	KSTPIYVKPOL	TS	SGTCRIS	VLVGIRND	PGKTIDS	ITVQF	RLP	PCVLS	SDLSSN	CGAVNVLA	KTC	SW	IGRIPKDKAPS
<i>RcCac</i>	101	SFVPPDGL	NLM	SYRVKKL	KSTPIYVKPOL	TS	SGTCRIS	VLVGIRND	PGKTIDS	ITVQF	RLP	PCVLS	SDLSSN	CGAVNVLA	KTC	SW	IGRIPKDKAPS
<i>GmCac</i>	101	SFVPPDGL	NLM	SYRVKKL	KSTPIYVKPOL	TS	SGTCRIS	VLVGIRND	PGKTIDS	ITVQF	RLP	PCVLS	SDLSSN	CGAVNVLA	KTC	SW	IGRIPKDKAPS
		210	220	230	240	250	260										
<i>TgCac</i>	192	MSGTLV	LETG	LERLHV	FPTF	QVGF	FRIMG	VALSGL	QIDK	LDL	KV	YMR	YKGF	RAL	TRAG	E	EVRS
<i>VvCac</i>	201	MSGTLV	LETG	LERLHV	FPTF	QVGF	FRIMG	VALSGL	QIDK	LDL	KV	YMR	YKGF	RAL	TRAG	E	EVRS
<i>AlCac</i>	201	MSGTLV	LETG	LERLHV	FPTF	QVGF	FRIMG	VALSGL	QIDK	LDL	KV	YMR	YKGF	RAL	TRAG	E	EVRS
<i>PtCac</i>	201	MSGTLV	LETG	LERLHV	FPTF	QVGF	FRIMG	VALSGL	QIDK	LDL	KV	YMR	YKGF	RAL	TRAG	E	EVRS
<i>RcCac</i>	201	MSGTLV	LETG	LERLHV	FPTF	QVGF	FRIMG	VALSGL	QIDK	LDL	KV	YMR	YKGF	RAL	TRAG	E	EVRS
<i>GmCac</i>	201	MSGTLV	LETG	LERLHV	FPTF	QVGF	FRIMG	VALSGL	QIDK	LDL	KV	YMR	YKGF	RAL	TRAG	E	EVRS

- C) ACT gene. Species and translated sequences used: *Tectona grandis* (JZ515974), *Populus trichoparpa* (XM_002308329.1), *Arabidopsis lyrata* (XM_002882721.1), *Arabidopsis thaliana* (NM_112046.3), *Glycine max* (NM_001254249.1), *Ricinus communis* (XM_002530665.1), *Vitis vinifera* (XM_002279636.1)

[illegible]

- D) *HIS3* gene. Species and translated sequences used: *Tectona grandis* (JZ515975), *Populus trichoparpa* (XM_002306258.1), *Gossypium hirsutum* (AF024716.1), *Lycopersicon esculentum* (X83422.1), *Zea mays* (EU976723.1)

```

      10      20      30      40      50      60      70      80      90     100
TgHis3  1  .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
PtHis3  1  MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAY
GhHis3  1  MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAY
LeHis3  1  MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAY
ZmHis3  1  MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRPGTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAY

      110     120     130
TgHis3  69  LVGLFEDTN.....|.....|.....|.....|.....|.....|
PtHis3  101  LVGLFEDTNLCATHAKRVTIMPKDIQLARRINGERA
GhHis3  101  LVGLFEDTNLCATHAKRVTIMPKDIQLARRINGERA
LeHis3  101  LVGLFEDTNLCATHAKRVTIMPKDIQLARRINGERA
ZmHis3  101  LVGLFEDTNLCATHAKRVTIMPKDIQLARRINGERA

```

E) *SAND* gene. Species and translated sequences used: *Tectona grandis* (JZ515976), *Populus trichoparpa* (XM_002314230.1), *Arabidopsis thaliana* (NM_128399.3), *Picea sitchensis* (EF676351.1), *Vitis vinifera* (XM_002285134.1)

Figure 1 displays a multiple sequence alignment of the TgSand, PtSand, AtSand, PsSand, and VvSand proteins. The alignment is presented in four blocks, with positions 10 to 400 marked at the top of each block. The proteins are aligned in rows, with TgSand at the top and VvSand at the bottom of each block. Conserved regions are indicated by black boxes with white text. The alignment shows high conservation between the species, particularly in the regions highlighted in black boxes.

- F) *B-TUB* gene. Species and translated sequences used: *Tectona grandis* (JZ515977), *Populus trichoparpa* (XM_002298000.1), *Gossypium hirsutum* (AF521240.1), *Medicago truncatula* (XM_003630465.1), *Nicotiana tabacum* (EF051136.2), *Ricinus communis* (XM_002509755.1), *Theobroma cacao* (GU570572.1), *Vitis vinifera* (XM_002273478.2)

```

      210      220      230      240      250      260      270      280      290      300
TgTub  1  -----TQQMWDAKNMM
PtTub  55  VLDNEALYDICFRTLKLTNPSPGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTS[GSQQYRALTIPELTQQMWDAKNMM
GhTub  201  VLDNEALYDICFRTLKLTNPSPGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTSRGSQQYRALTIPELTQQMWD[KNMM
MtTub  201  VLDNEALYDICFRTLKLTNPSPGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTSRGSQQY[SLTIPELTQQMWD[KNMM
NtTub  55  VLDNEALYDICFRTLKLT[PSFGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTSRGSQQYRAL[SVPELTQQMWDAKNMM
RcTub  201  VLDNEALYDICFRTLKLTNPSPGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTSRGSQQYRALTIPELTQQMWDAKNMM
TcTub  136  VLDNEALYDICFRTLKLT[PSFGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTSRGSQQYRALTIPELTQQMWDAKNMM
VvTub  201  VLDNEALYDICFRTLKLTNPSPGDLNHLISTTMSGVTCCLRFPQQLNSDLRKLAVNLIPFRLHFFMVGFAPLTSRGSQQYRALTIPELTQQMWDAKNMM

      310      320      330      340      350      360      370      380      390      400
TgTub  12  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLMSSTFVGNSTSIQEMFRRVS-----
PtTub  155  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLAMSS[RIY[-----K[MYYSRNV-----
GhTub  301  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLMSSTFMGNSTSIQEMFRRVSEQFTVMFRKAF[FLHWYTG
MtTub  301  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLMSSTFMGNSTSIQEMFRRVSEQFTVMFRKAF[FLHWYTA
NtTub  155  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLM[STF[GNSTSIQEMFRRVSEQFT[MFRRKAF[FLHWYTG
RcTub  301  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLMSSTFMGNSTSIQEMFRRVSEQFTVMFRKAF[FLHWYTG
TcTub  236  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLM[STF[GNSTSIQEMFRRVSEQFT[MFRRKAF[FLHWYTG
VvTub  301  CAADPRHGRYLTASAMFRGKMSTKEVDEQMINVQNKNSYFVEWIPNNVKSSVCDIPPTGLAMSSSTFMGNSTSIQEMFRRVSEQFTVMFRKAF[FLHWYTG

```

- G) *UBQ* gene. Species and translated sequences used: *Tectona grandis* (JZ515978), *Populus trichoparpa* (XM_002320914.1), *Hevea brasiliensis* (EF120638.1), *Medicago truncatula* (XM_003629847.1), *Nicotiana tabacum* (DQ138111.1), *Pyrus communis* (AF386524.1), *Ricinus communis* (XM_002515167.1), *Solanum tuberosum* (L22576.1)

```

      10      20      30      40      50      60      70      80      90     100
TgUbq  1  Q...DGDHNS...GI...R-----H...IDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKN-----
HbUbq  1  MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKKKKVKN-----
MtUbq  1  MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKK...HKKVKN
NtUbq  1  MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKKKKVKN
PcUbq  1  MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKK...HKKVKN
RcUbq  1  MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKKKKVKN
StUbq  1  MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGIPPDQQRLLIFAGKQLEDGRTLADYNIQESTLHLVLRRLGGAKKRRKKKTYTKPKKIKHKKKKVKN

```

F-1a gene. Species and translated sequences used: *Tectona grandis* (JZ5159), *Populus trichocarpa* (EF147714.1), *Arabidopsis thaliana* (NM_100666.3), *Elaeagnus angustifolia* (AY550990.1), *Gossypium hirsutum* (DQ174254.1), *Malus domestica* (AJ223969.1), *Nicotiana paniculata* (AB019427.1), *Prunus persica* (FJ267653), *Vitis vinifera* (XM_002284888.1)

[illegible]

Additional File 4 - Teak sequences (with accession numbers) used for designing qRT-PCR
primers (underlined)

(Continue)

>*Tectona grandis* ribosomal protein 60S (JZ515972)

GCAGATATGGTGAAGTTCTTGAAGCCGAACAAAGCCGTAATAATCCTGCAAGGCCGTTACGCCGGCCGTAAAGCA
GTGATCGTCCGCTCATTCGACGACGGCACTCGTGACCGGCCGTACGGCCATTGCTTGGTGGCGGGGCTGGCGAAG
TACCCGCGCAAGGTCATCCGCAAAGACTCGGGCGAAGAAGCAGGCGAAGAAATCACGAGTGAAATGCTTCATCAAA
TTGGTGAATTACAACCACATCATGCCACGCGCTACACGCTCGATGTGGATCTGAAGGATGTGGTCGCTCCGGAT
TGCTTGCACTCGAAAGATAAGAAGGTGACGGCGGCGAAGGAGACGAAGGCTCGGTTTCGAGGAGCGGTTCAAGACG
GGGAAAAACCGCTGGTTCCTTACCAAGCTC

>*Tectona grandis* Clathrin Adaptor Complex (JZ515973)

CCAGGGGCAACATCATCTTGTGTTCATGGAGAAAAACAGACCTTAAGCATGCCAGCAATGAAGTTTATGTTGAT
CTTGTGGAAGAAATGGATGCCACTATAACAGAGATGGGACTCTGGTAAAATGCGAGATATATGGTGAAGTTCAA
GTAAATGCTCATTTATCAGGTCTCCAGATCTCACTCTGTTGTTTGCGAACCCCTTCAATTCTTAATGATGTGAGA
TTTCACCCCTTGTGTTAGACTTCGTCCATGGGAATCAAACCAATTCTGTCTTTGTGCCACCTGATGGACAATTT
AATCTCATGAGTTACAGGGTAAAGAAGTTGAAGAGTACTCCAATTTATGTGAAGCCGCAATTGACCTCGGATTCA
GGGACATGTCTGATAAGTGTCTAGTTGGAATACGGAACGATCCTGGAAAGACGATTGACTCTATAACAGTTCAA
TTCCGATTACCACCTTGTGTTTTATCGTCTGATCTTTCATCAAATTTGTGGAGCTGTAAATGTTCTTGCTGACAAG
ACCTGCTCATGGACAATCGGACGAATACCAAAAGATAAAGCTCCTTCAATGTCTGCGACCTTGGTGCTTGAGACA
GGCATAGAGCGCCTTCATGTATTTCCC

>*Tectona grandis* Actin (JZ515974)

GTTAGCAATTGGGATGATATGGAGAAAACTGGGCATCACACATTCTACAACGAACCTTCGTGTTGCTCCAGAAGAG
CACCCAATTCTCTTGACAGATGCTCCTCTTAATCCCAAGGCCAACCCTGAAAAGATGACTCAAATTATGTTTGAG
ACCTTTAATGCCCTGCCATGTATGTTGCCATCCAGGCTGTTCTCTCCCTTTATGCCAGTGGTCGTACAACCTGGT
ATTGTTCTCGACTCTGGGGATGGTGTAGCCATACAGTGCCCATCTATGAAGGCTATGCACTCCCCATGCTATC
CTGCGTCTTGATCTTGCCGGGCGTGATCTCACTGACCACCTCATGAAGATATTGACAGAACGAGGCTACTCATTC
ACTACAACGTCAGAGCGGGAAATGTAAAGGACATAAAAGAAAAGCTGGCTTACATTGCTCTGGATTATGAGCAA
GAGCTAGAAACAGCAAAGACTAGCTCTGCTGTGGAGAAGAACTATGAACCTGCCTGATGGCAGG

>*Tectona grandis* Histone 3 (JZ515975)

ATTGGTGTCTTCGAAGAGACCAACAAGATACGCCTCCGCAGCTTCCTGGAGAGCGAGCACGGCGTGGCTTTGGAA
CCTCAAATCGGTCTTGAAATCCTGAGCGATTTACGAACCAACGCTGGAAAGGCAGCTCGCGGATGAGGAGTTC
TGTGCTCTTCTGATATTTCCGGATTTACGAAGAGCAACAGTTCCAGGGCGGTAACGATGGGGCTTCTTCACTCC
TCCCGT

>*Tectona grandis* Sand Family (JZ515976)

CATCGCTGCTTGGGGAACTGATGCTGTCTTCTCTCTCATCCATTCTTTTCAAGTTGGAATCCTGCCACTTTTC
CGCATGCCACTCATGTCTTCCGCTTGCTTATGCAACACGTCAAGCTGCAGGTGCCTTTTGAAGATGTAGCTGGT
CAGGAGTCCTATTTGCACTCTTATTGTGTAAACACAAGGTTATCAGTCTTGTGGCGCCCAAAAAGCATCTCTTC
ATCCTGATGATATATTGCTACTCTCCAATTTTATTATGTCCTCTGAATCTTTTAGGACATCTGAGTCCTTCTCAC
CAATTTGCCTGCCAAGATACAATTCCATGGCATTTCTTTATGCTTATGTGCATTATTTTGATATCGATACTTACC
TGATCTTGCTCACCAAGTTCTGATGCCTTTTATCATTTAAAAGATAGCAGGATTCCGATTGAAAATGTCCTTT
TAAAGTCAAATGTAAGTGAAGTTCAAAGATCCTTGGTAGATGGTGGTATGCATATTGAGGATTGCTTAGTG
ACCCCGCTTCTCGTCTGGGGCCATGTCTTCTCATCTAGGTCAACCAAGACCTGGTAGAGATTCTCCGGGGAGAA
TTAGAGGTGGAATTTGTTGAAATTGGTGGTCCAGCTGGACTTTGGCATTTCATG

>*Tectona grandis* Beta Tubulin (JZ515977)

ACACAGCAAATGTGGGATGCAAAGAACATGATGTGCGCCGCCGACCCCGCCACGGCCGCTACCTGACCGCCTCT
GCCATGTTCCGCGGCAAGATGAGCACGAAAGAAGTGGACGAACAAATGATCAACGTGCAGAACAAGAATTCATCC
TACTTCGTGCAATGGATCCCCAACAACGTTAAATCAAGTGTTTGTGACATTTCCCCCAACTGGGCTCTCAATGTCTG
TCGACTTTCGTGCGGAATTCGACGTCGATAACAGGAGATGTTCCGGCGCGTGTCG

>*Tectona grandis* Ubiquitin (JZ515978)

TTCAGCAGATTGACGGGTAAGACCATAACTCTGGAGGTTGAATCCTCCGACATATCGACAATGTCAAGGCCAAGA
TCCAGGACAAGGAAGGCATACCGCCGACCAGCAGCGCTCATCTTCGCTGGAAAACAGCTCGAGGACGGTCGCA
CCCTCGCCGACTACAACATCCAAAAGGAATCGACTCTCCACCTCGTTCTCCGCCTCCGCGGCGGCGCTAAGAAGC
GGAAGAAGAAGACCTACACCAAGCCGAAGAAAATCAAGCACAGAACA

Additional File 4 - Teak sequences (with accession numbers) used for designing qRT-PCR
primers (underlined)

(Conclusion)

>*Tectona grandis Elongation Factor 1 alpha* (JZ515979)

CTTTATCAAGAACATGATCACTGGTTATCACAGGCTGATTGTGCTGTCCTTATCATTGACTCTACCACTGGTGGT
TTTGAAGCTGGTATTTCCAAGGATGGTCAGACCCGTGAGCATGCATTGCTTTGCTTTCACTCTTGGTGTCAAGCAA
ATGATTTGTTGTTGCAACAAGATGGATGCCACCACACCAAAATACTCCAAGGCTAGGTATGATGAAATTGTGAAG
GAAGTGTCTTCCTACCTCAAGAAGGTTGGATACAACCCTGAAAAGATCCCATTTGTTCCCATTTCTGGTTTTGAG
GGAGACAACATGATTGAGAGGTCCACTAACTTGGACTGGTACAAGGGCCCAACCCCTCCTTGAGGCGCTTGACATG
GTCAGGAGCCCAAGAGGCCCTCAGACAAGCCTCTCCGTCTGCCACTTCAGGATGTTTACAAGATTGGTGGTATT
GGTACTGTCCCTGTTGGCCGAGTGGAGACCGGTATCCTCAAGCCTGGTATGGTTGTGACCTTTGGCCCGACTGGG
TTGACCACTGAAGTTAAGTCTGTTGAGATGCACCACGAAGCGTTGCAGGAGGCTCTTCCTGGTGATAATGTTGGG
TTCAACGTGAAGAATGTTGCTGTGAAGGATCTGAAACGTGGTTTTGTTGCCTCCAACCTCAAGGATGATCCTGCT
AAGGAAGCCGCCAACTTCCTCCCAAGTCATCATGAACCACCCCTGGCCAGATTGGA

> *Tectona grandis Glyceraldehyde 3-phosphate dehydrogenase* (FN431983.1)

GTA CTTTATTCACCATCTGAGTCTAATAAATGCTGCTCGCCGACTTCACCTTGTGATTCTGCTTTCTAGTAGTGT
TATCCTTTTGTGTTGCATTTGTGGTAGTATATTTTAAGGAGATGGATTGATTACTTGGTCTTCCCTTGATCTC
TGTTTTGTACATAATTTGTGTAGGCTGTTGGTAAAGTGCTTCAGCCCTAAATGGAAAGTTGACTGGTATGGCAT
TCCGAGTTCCAACAGTTGATGTTTCAGTCGTGGACCTCACTGTCAGGTTGGAGAAAGAGGCCACCTATGAGGAGA
TCAAAGCTGCCATCAAGTGAGTTTTTGAATTTGGTTTTGAACAATGTGAGCTAGTAATAGTTCTATGGATGTGTA
ATCTAAAGTTGCTTGGTTCTCTTAGGGAGGAGTCGGAGAACAAAGCTAAAGGGCATCTTGGGGTACACTGAAGATG
ATGTGGTGTCAACAGACTTTGTGGGCGATAGCCGGTAAATGTCTTTTCCTAATGGAGTACCTTTTGGTCTAGCAA
TTAGTCTCTCTTCGATTTGTAGGGTGCTCAAACCT

>*Tectona grandis Cinnamyl Alcohol Dehydrogenase* (JZ515980)

AGCCATGAAGTAATTGGAGAAGTTGTTGAATTGGGCTCAAAAGTGAAGAATTTCAAAGTGGGTGACATTATAGGA
GTTGGAGGAATTATTGGTTCTTGTGAAAAATGTACTCTCTGCAACTCCAATCTGGAGCAATACTGCAGCAACAGA
ATCTTTACCTACAATGACGTCTACAAAGATGGAACCTCAACTCAAGGGGGATATTCTTCTGCTATGGTTATTCAT
CACAGGTTTCCGAAATTTATCTTATTACATAAAAGCAAAAAACACGAAACCGAAACACTCAACAACGGATTAC
AGATTTGCAGTTAAAAATACCAGAAAAACTAGCACCAGAACAAGCAGCACCCTACTATGTGCCGGGGTGACAGCA
TACAGTCCCCCTCAAAGAGTTCATGGATTCCGGCAAGGTCTACAAAGGAGGAATATTAGGCTTGGGAGGAGTTGGT
CACCTGGCTGTGATGATAGCAAAGGCAATGGGTGTCATCATGTGACAGTAATAAGCTCTTCTGATAAGAAAAAGAG
GAGGCTATGGAGCATTTGCACGCAGACGCCTTCTTGGTGAGTCGTAGTGAGGATGAAATGAAGCAAGCGATAAAC
AGCCTCGACTATATACTCGACACCGTGCCTGTTGTTTCATCCTCTGCCATCATATATTTCACTTTTGAAAAC
TGAAAGCTGTTATTAGTAGCGGCAGTTCCTCAGCCACTTCAGTTTCTGGCTGCCGATATGATAAGAGGTATGGTA
TAT

Additional File 5 - Agarose gel (2%) electrophoresis showing amplification of a specific PCR product of the expected size for each gene. M represents 50 bp DNA ladder marker (GeneRuler™ 50bp DNA Ladder, Thermo Scientific, USA) and “-” represents negative control

ditional File 6 - Raw CP data used for statistical analysis in this study

Sample	<i>RP60S</i>	<i>SAND</i>	<i>ACT</i>	<i>CAC</i>	<i>GAPDH</i>	<i>B-TUB</i>	<i>HIS3</i>	<i>UBQ</i>	<i>EF-1α</i>
Flower 1	29.159	30.163	24.925	31.887	25.974	28.741	27.770	26.510	23.046
Flower 2	27.611	29.992	24.479	30.733	25.414	28.699	27.055	26.141	22.920
Flower 3	27.806	28.285	24.262	29.968	24.860	28.916	27.701	26.037	22.520
Flower 4	28.983	30.218	24.658	32.597	26.103	28.841	27.572	26.526	23.098
Flower 5	27.521	29.653	24.388	30.178	25.310	28.644	27.077	26.171	22.813
Flower 6	27.816	28.664	24.151	29.691	24.992	28.855	27.466	25.987	22.558
Leaf 1	29.820	30.447	27.498	32.098	26.482	28.904	27.554	27.794	24.410
Leaf 2	28.787	29.239	27.401	30.424	25.626	31.931	26.306	26.993	24.220
Leaf 3	28.847	29.426	25.849	30.955	25.932	28.147	26.543	26.642	22.884
Leaf 4	29.714	30.536	27.201	32.032	26.533	28.920	27.565	27.796	24.241
Leaf 5	28.888	29.037	27.158	29.953	25.635	32.017	26.560	26.976	24.194
Leaf 6	28.947	29.728	25.793	31.299	25.972	28.013	26.550	26.585	22.893
Root 1	27.900	29.061	25.016	30.590	25.232	28.649	26.951	26.033	22.504
Root 2	28.797	29.801	25.394	30.971	25.647	29.283	27.399	26.722	23.162
Root 3	28.319	29.789	25.790	31.423	25.802	29.618	28.157	26.823	23.336
Root 4	27.875	29.049	24.948	30.242	25.382	28.661	26.860	25.993	22.545
Root 5	28.821	29.923	25.515	30.612	25.691	29.088	27.414	26.616	23.021
Root 6	28.343	29.660	25.877	31.403	25.968	29.766	27.980	26.789	23.407
Seedling 1	27.679	30.160	24.804	31.406	24.638	29.793	29.507	26.157	22.903
Seedling 2	28.409	30.059	25.111	31.592	24.971	29.144	28.428	26.533	22.845
Seedling 3	28.719	30.913	24.926	32.432	25.728	28.731	28.702	26.399	22.926
Seedling 4	27.769	30.357	24.760	31.513	24.797	29.603	29.396	26.176	22.874
Seedling 5	28.273	29.673	25.068	31.469	25.032	29.083	28.465	26.480	22.855
Seedling 6	28.875	30.627	24.855	32.940	25.877	28.684	28.541	26.398	22.970
Branch secondary xylem 1	28.913	28.801	26.878	29.477	24.611	28.849	28.580	26.752	23.388
Branch secondary xylem 2	28.306	28.550	26.326	29.944	25.181	28.444	28.676	26.069	23.077
Branch secondary xylem 3	28.150	28.344	26.288	28.714	24.246	28.503	28.359	26.054	22.987
Branch secondary xylem 4	28.888	29.133	26.981	29.350	24.737	28.792	28.569	26.677	23.563
Branch secondary xylem 5	28.461	28.631	26.388	29.671	25.219	28.554	28.508	26.092	23.015
Branch secondary xylem 6	28.064	28.084	25.574	28.817	24.205	27.996	28.266	26.135	22.968
Stem secondary xylem 1	29.278	31.939	29.238	32.102	26.675	31.952	31.437	27.756	25.094
Stem secondary xylem 2	28.957	31.716	29.907	32.036	26.662	33.539	28.851	28.706	25.134
Stem secondary xylem 3	29.341	31.953	29.431	33.113	27.356	32.131	31.054	28.038	25.230
Stem secondary xylem 4	29.059	32.147	29.301	32.545	26.540	32.107	31.451	27.897	24.926
Stem secondary xylem 5	28.934	31.965	29.855	31.908	26.749	32.632	28.889	28.566	25.156
Stem secondary xylem 6	29.277	31.991	29.723	32.445	27.361	32.242	31.658	27.945	25.164

References

- ALTSCHUP, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic Local alignment search tool. **Journal of Molecular Biology**, London, v. 215, p. 403–410, 1990.
- ANDERSEN, C.L.; JENSEN, J.L.; ØRNTTOFT, T.F. Normalization of Real-Time Quantitative Reverse Transcription-PCR data : a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. **Cancer Research**, Baltimore, v. 64, p. 5245–5250, 2004.
- BARAKAT, A.; BAGNIEWSKA-ZADWORNA, A.; FROST, C.J.; CARLSON, J.E. Phylogeny and expression profiling of CAD and CAD-like genes in hybrid *Populus* (*P. deltoides* × *P. nigra*): evidence from herbivore damage for subfunctionalization and functional divergence. **BMC Plant Biology**, London, v. 10, p. 100, 2010.
- BARSALOBRES-CAVALLARI, C.F.; SEVERINO, F.E.; MALUF, M.P.; MAIA, I.G. Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. **BMC Molecular Biology**, London, v. 10, p. 1, 2009.
- BHUIYAN, N.H.; SELVARAJ, G.; WEI, Y.; KING, J. Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. **Journal of Experimental Botany**, Oxford, v. 60, n. 2, p. 509–521, 2009.
- BIN, W.S.; WEI, L.K.; PING, D.W.; LI, Z.; WEI, G.; BING, L.J.; GUI, P.B.; JIAN, W.H.; FENG, C.J. Evaluation of appropriate reference genes for gene expression studies in pepper by quantitative real-time PCR. **Molecular Breeding**, Dordrecht, v. 30, n. 3, p. 1393–1400, 2012.
- BRUNNER, A.M.; YAKOVLEV, I.A.; STRAUSS, S.H. Validating internal controls for quantitative plant gene expression studies. **BMC Plant Biology**, London, v. 4, p. 14, 2004.
- BUSTIN, S.A.; BENES, V.; GARSON, J.A.; HELLEMANS, J.; HUGGETT, J.; KUBISTA, M.; MUELLER, R.; NOLAN, T.; PFAFFL, M.W.; SHIPLEY, G.L.; VANDESOMPELE, J.; WITTEW, C.T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **Clinical Chemistry**, Baltimore, v. 55, n. 4, p. 611–622, 2009.
- CHAMBERS, J.P.; BEHPOURI, A.; BIRD, A.; NG, C.K.-Y. Evaluation of the use of the polyubiquitin genes, Ubi4 and Ubi10 as reference genes for expression studies in *Brachypodium distachyon*. **PloS One**, San Francisco, v. 7, n. 11, p. e49372, 2012.
- CHANG, E.; SHI, S.; LIU, J.; CHENG, T.; XUE, L.; YANG, X.; YANG, W.; LAN, Q.; JIANG, Z. Selection of reference genes for quantitative gene expression studies in *Platycladus orientalis* (Cupressaceae) using real-time PCR. **PloS One**, San Francisco, v. 7, n. 3, p. e33278, 2012.
- CHAO, W.S.; DOĞRAMACI, M.; FOLEY, M.E.; HORVATH, D.P.; ANDERSON, J.V. Selection and validation of endogenous reference genes for qRT-PCR analysis in leafy spurge (*Euphorbia esula*). **PloS One**, San Francisco, v. 7, n. 8, p. e42839, 2012.

COELHO, A.C.; HORTA, M.; NEVES, D.; CRAVADOR, A. Involvement of a cinnamyl alcohol dehydrogenase of *Quercus suber* in the defence response to infection by *Phytophthora cinnamomi*. **Physiological and Molecular Plant Pathology**, London, v. 69, n. 1/3, p. 62–72, 2006.

CONTE, I.; BANFI, S.; BOVOLENTA, P. Non-coding RNAs in the development of sensory organs and related diseases. **Cellular and Molecular Life Sciences**, Basel, v. 70, n. 21, p. 4141–4155, 2013.

CZECHOWSKI, T.; STITT, M.; ALTMANN, T.; UDVARDI, M.K.; SCHEIBLE, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization. **Plant Physiology**, Bethesda, v. 139, p. 5–17, Sept. 2005.

DHEDA, K.; HUGGETT, J.F.; CHANG, J.S.; KIM, L.U.; BUSTIN, S.A.; JOHNSON, M.A.; ROOK, G.A.W.; ZUMLA, A. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. **Analytical Biochemistry**, New York, v. 344, n. 1, p. 141–143, 2005.

EXPÓSITO-RODRÍGUEZ, M.; BORGES, A.A.; BORGES-PÉREZ, A.; PÉREZ, J.A. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. **BMC Plant Biology**, London, v. 8, p. 131, 2008.

GUERRERO-VÁSQUEZ, G.A.; ANDRADE, C.K.Z.; MOLINILLO, J.M.G.; MACÍAS, F. A. Practical first total synthesis of the potent phytotoxic ($\pm$)-naphthotectone, isolated from *Tectona grandis*. **European Journal of Organic Chemistry**, Weinheim, v. 2013, n. 27, p. 6175–6180, 2013.

HALLETT, J.T.; DIAZ-CALVO, J.; VILLA-CASTILLO, J.; WAGNER, M.R. Teak plantations : economic bonanza or environmental disaster? **Journal of Forestry**, Washington, v. 109, n. 5, p. 288–292, 2011.

HAN, X.; LU, M.; CHEN, Y.; ZHAN, Z.; CUI, Q.; WANG, Y. Selection of reliable reference genes for gene expression studies using real-time PCR in tung tree during seed development. **PloS one**, San Francisco, v. 7, n. 8, p. e43084, 2012.

HAUPTMAN, N.; GLAVAC, D. MicroRNAs and long non-coding RNAs: prospects in diagnostics and therapy of cancer. **Radiology and Oncology**, Ljubljana, v. 47, n. 4, p. 311–318, 2013.

HEALEY, S.P.; GARA, R.I. The effect of a teak (*Tectona grandis*) plantation on the establishment of native species in an abandoned pasture in Costa Rica. **Forest Ecology and Management**, Amsterdam, v. 176, n. 1/3, p. 497–507, 2003.

HONG, S.-Y.; SEO, P.J.; YANG, M.-S.; XIANG, F.; PARK, C.-M. Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR. **BMC Plant Biology**, London, v. 8, p. 112, 2008.

- IMAI, T.; UBI, B.E.; SAITO, T.; MORIGUCHI, T. Evaluation of reference genes for accurate normalization of gene expression for real time-quantitative PCR in *Pyrus pyrifolia* using different tissue samples and seasonal conditions. **PloS One**, San Francisco, v. 9, n. 1, p. e86492, 2014.
- KUBISTA, M.; ANDRADE, J.M.; BENGTSSON, M.; FOROOTAN, A.; JONÁK, J.; LIND, K.; SINDELKA, R.; SJÖBACK, R.; SJÖGREEN, B.; STRÖMBOM, L.; STÅHLBERG, A.; ZORIC, N. The real-time polymerase chain reaction. **Molecular Aspects of Medicine**, Elmsford, v. 27, n. 2/3, p. 95–125, 2006.
- LEE, J.M.; ROCHE, J.R.; DONAGHY, D.J.; THRUSH, A.; SATHISH, P. Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (*Lolium perenne* L.). **BMC Molecular Biology**, London, v. 11, p. 8, 2010.
- LIBAULT, M.; THIBIVILLIERS, S.; BILGIN, D.D.; RADWAN, O.; BENITEZ, M.; CLOUGH, S.J.; STACEY, G. Identification of four soybean reference genes for gene expression normalization. **The Plant Genome**, Madison, v. 1, n. 1, p. 44, 2008.
- LIN, Y.L.; LAI, Z.X. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. **Plant Science**, Limerick, v. 178, n. 4, p. 359–365, 2010.
- LIU, D.; SHI, L.; HAN, C.; YU, J.; LI, D.; ZHANG, Y. Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. **PloS one**, San Francisco, v. 7, n. 9, p. e46451, 2012.
- LIVAK, K.J.; SCHMITTGEN, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. **Methods**, San Diego, v. 25, n. 4, p. 402–408, 2001.
- LUKMANDARU, G.; TAKAHASHI, K. Variation in the natural termite resistance of teak (*Tectona grandis* Linn. fil.) wood as a function of tree age. **Annals of Forest Science**, Les Ulis, v. 65, p. 708, 2008.
- MACKAY, I.M.; ARDEN, K.E.; NITSCHKE, A. Real-time PCR in virology. **Annals of Forest Science**, Les Ulis, v. 30, n. 6, p. 1292–1305, 2002.
- MARUM, L.; MIGUEL, A.; RICARDO, C.P.; MIGUEL, C. Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*. **PloS one**, San Francisco, v. 7, n. 4, p. e35113, 2012.
- MIRANDA, I.; SOUSA, V.; PEREIRA, H. Wood properties of teak (*Tectona grandis*) from a mature unmanaged stand in East Timor. **Journal of Wood Science**, Tokyo, v. 57, n. 3, p. 171–178, 2011.
- OHDAN, T.; FRANCISCO, P.B.; SAWADA, T.; HIROSE, T.; TERAOKA, T.; SATOH, H.; NAKAMURA, Y. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. **Journal of experimental botany**, Oxford, v. 56, n. 422, p. 3229–3244, 2005.

PFAFFL, M.W. Quantification strategies in real-time PCR Michael W. Pfaffl. In: BUSTIN, S.A. (Ed.). **A–Z of quantitative PCR**. La Jolla: International University Line, 2004. p. 87–112.

PFAFFL, M.W.; TICHOPAD, A.; PRGOMET, C.; NEUVIANS, T.P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. **Biotechnology Letters**, Dordrecht, v. 26, n. 6, p. 509–515, 2004.

RADONIĆ, A.; THULKE, S.; MACKAY, I.M.; LANDT, O.; SIEGERT, W.; NITSCHKE, A. Guideline to reference gene selection for quantitative real-time PCR. **Biochemical and Biophysical Research Communications**, Orlando, v. 313, n. 4, p. 856–862, 2004.

SALZMAN, R.A.; FUJITA, T.; HASEGAWA, P.M. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. **Plant Molecular Biology Reporter**, Athens, v. 17, n. 765, p. 11–17, 1999.

SCHMIDT, G.W.; DELANEY, S.K. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. **Molecular Genetics and Genomics**, Berlin, v. 283, n. 3, p. 233–241, 2010.

SILVER, N.; BEST, S.; JIANG, J.; THEIN, S.L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. **BMC Molecular Biology**, London, v. 7, p. 33, 2006.

TRABUCCO, G.M.; MATOS, D.A.; LEE, S.J.; SAATHOFF, A.J.; PRIEST, H.D.; MOCKLER, T.C.; SARATH, G.; HAZEN, S.P. Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. **BMC Biotechnology**, London, v. 13, n. 1, p. 61, 2013.

VANDESOMPELE, J.; PRETER, K. de; PATTYN, F.; POPPE, B.; ROY, N. van; PAEPE, A. de; SPELEMAN, F. Accurate normalization of real-time quantitative RT -PCR data by geometric averaging of multiple internal control genes. **Genome Biology**, London, v. 3, n. 7, p. 34, 2002.

VERHAEGEN, D.; FOFANA, I.J.; LOGOSSA, Z.A.; OFORI, D. What is the genetic origin of teak (*Tectona grandis* L.) introduced in Africa and in Indonesia? **Tree Genetics & Genomes**, Davis, v. 6, n. 5, p. 717–733, 2010.

WANG, H.-L.; CHEN, J.; TIAN, Q.; WANG, S.; XIA, X.; YIN, W. Identification and validation of reference genes for *Populus euphratica* gene expression analysis during abiotic stresses by quantitative real-time PCR. **Physiologia Plantarum**, Copenhagen, v. 152, n. 3, p. 529–545, 2014.

XU, M.; ZHANG, B.; SU, X.; ZHANG, S.; HUANG, M. Reference gene selection for quantitative real-time polymerase chain reaction in *Populus*. **Analytical Biochemistry**, New York, v. 408, n. 2, p. 337–339, 2011.

ZHU, J.; ZHANG, L.; LI, W.; HAN, S.; YANG, W.; QI, L. Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. **PloS One**, San Francisco, v. 8, n. 1, p. e53196, 2013.

3 CHARACTERIZATION OF CINNAMYL ALCOHOL DEHYDROGENASE GENE FAMILY IN LIGNIFYING TISSUES OF *Tectona grandis* L.f.

Abstract

Lignin, a phenolic compound formed by monolignols is present in secondary cell walls and confers support, pathogens defense and fluids conduction in plants. The cinnamyl alcohol dehydrogenase (CAD) enzyme catalyzes the last step of monolignols synthesis in the phenylpropanoid pathway. There is more than one gene that characterize the CAD family, of which the number and functions may vary between plants. This work is the first to characterize genes of this family in *Tectona grandis* (teak), a tropical tree with high valuable timber used for furniture, flooring and shipbuilding. *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera* and *Populus trichocarpa* have the complete CAD gene family described. We characterized four members of CAD gene family in teak; the complete *TgCAD1* has 1071 bp, 357 amino acids, 38.96 kDa and matched with AtCAD5 three-dimensional model and the partial *TgCAD2*, *TgCAD3* and *TgCAD4* have 891, 657 and 753 bp, corresponding to 297, 219 and 251 amino acids, respectively. Also, the four CAD members in teak exhibited conserved residues for catalytic zinc action, structural zinc ligation and NADPH binding and substrate specificity, consistent with the mechanism of alcohol dehydrogenases. Phylogenetic analysis using *Arabidopsis*, rice, perennial ryegrass, corn, tobacco, alfalfa, sugarcane, Norway spruce, loblolly pine, eucalyptus, cocoa tree, black cottonwood and grape protein CAD members showed that TgCADs are present in three main classes and seven groups. Quantitative real time PCR revealed that *TgCAD4* is expressed in different vegetative tissues except leaves while *TgCAD1* is highly expressed in leaves and could be related with pathogen defense. *TgCAD3* and *TgCAD4* were highly expressed in juvenile and mature sapwood. Phylogeny and expression profiles suggest that *TgCAD2*, *TgCAD3* and *TgCAD4* are involved in wood development, with tissue-specialized expression profiles. *TgCAD3* and *TgCAD4* seem to be duplicated and highly related to lignin biosynthesis, and *TgCAD4* could be related with teak maturation. This is one of the first studies in tropical trees on the CAD gene family including structure, phylogeny and expression.

Keywords: Tropical tree; Gene characterization: Phylogeny; Relative expression; Sapwood

3.1 Introduction

Lignin, a three-dimensional phenolic heteropolymer composted by *p*-coumaryl (H), coniferyl (G) and sinapyl (S) alcohols has essential roles in plant structural rigidity, pathogen defense and conduction of water and nutrients due to its hydrophobicity (BONAWITZ; CHAPPLE, 2010; XU et al., 2013; TANG et al., 2014). Lignin usually presents G and H units in gymnosperms and G, H and S units in angiosperms. It is a major component of plant cell walls and constitutes the last step of cell division, expansion and elongation before cell death (EUEDES et al., 2014; LAURICHESSE; AVÉROUS, 2014; ZENG et al., 2014).

Several biotic and abiotic stresses can activate the polymerization process which leads to lignin production (BONAWITZ; CHAPPLE, 2010; ZENG et al., 2014). In the last decade, lignin has been a target for several studies due to its agricultural and economic importance, suggesting that at least ten enzymes are required for lignin biosynthesis (XU et al., 2013). Lignin is the final product of the phenylpropanoid pathway, with cinnamyl alcohol dehydrogenase (CAD, EC 1.1.1.195) being the enzyme that catalyzes the conversion of cinnamyl aldehydes to cinnamyl alcohols, which is part of the medium-chain dehydrogenases (MDRs) (LI; LU; CHIANG, 2006; TANG et al., 2014). The MDRs catalyze the oxidation of alcohols to aldehydes with a NADP⁺ reduction and use zinc in their catalytic reaction in the motif called zinc-containing ADH, usually with the residues GHEX2GX5[GA]X2[IVACS] (LARROY et al., 2002). In the phenylpropanoid metabolism, the first step is the production of cinnamic acid with a deamination of phenylalanine by the phenylalanine ammonia-lyase (PAL) enzyme, with subsequent hydroxylations and methylations with a final production of the lignin monomers (BARAKAT et al., 2009; BONAWITZ; CHAPPLE, 2010).

The same authors reported that *p*-coumaric acid, *p*-coumaroyl-CoA, *p*-coumaroyl-CoA shikimate, caffeoyl shikimate, caffeoyl-CoA, feruloyl CoA, coniferaldehyde, *p*-coumaraldehyde, 5-Hydroxy- coniferaldehyde, sinapyl aldehyde, coniferil alcohol, sinapyl alcohol and *p*-coumaryl alcohol can be produced by the action of different enzymes. An exporting of lignin monomers to the apoplast after its synthesis is presumed, followed by oxidation mediated by laccases and peroxidases leading to reactive radical species and finally bimolecular coupling and polymer elongation (LI; LU; CHIANG, 2006; BONAWITZ; CHAPPLE, 2010).

Mansell et al. (1974) were the first to purify the CAD enzyme, and since then several homologues have been isolated, even obtaining complete gene families, such as *Arabidopsis* with 9 members (KIM et al., 2004), rice with 12 *CAD* genes (TOBIAS; CHOW, 2005) and *Populus* with 15 members (BARAKAT et al., 2009, 2010). *Brachypodium* (BUKH; NORD-LARSEN; RASMUSSEN, 2012), camellia, (DENG et al., 2013) and melon (JIN et al., 2014) also present several *CAD* genes characterized. Nearly, 80 *CAD* genes in 35 plants have been studied, usually present in a multi-gene family and with differential expression during plant development and environmental conditions (GUO; RAN; WANG, 2010; TANG et al., 2014).

Among different *CAD* gene families, a strong correlation between evolutionary pattern and gene function exist, even with several *CAD* orthologs being present, both in angiosperms as in gymnosperms (GUO; RAN; WANG, 2010). The same authors related that *CAD* genes with specific functions experience rapid duplication and those with conserved functions have

little copy number. The first bona fide *CAD* gene (involved in monolignol biosynthesis) came from lycophytes (as *Selaginella*) and subsequently it has been present in the main lineage of vascular plants; Class II has important roles in lignin biosynthesis and plant stress resistance and the functions of Class II genes are not conserved as bona fide *CAD* (GUO; RAN; WANG, 2010); Class III is a monophyletic clade responsible for plant development but not related with lignin biosynthesis. Recently, *CAD* gene has been explored through downregulation by reverse genetics and *CAD* mutants in several species, including dicot and monocot plants and coniferous species, to observe its impact on lignin amount and composition (VANHOLME et al., 2010; BOUVIER D'YVOIRE et al., 2013; PREISNER et al., 2014). *CAD* antisense gene expression in poplar (PILATE et al., 2002; BAUCHER et al., 2003) and eucalyptus (VALÉRIO et al., 2003) trees allowed a suitable woody tissue usage for agro-industrial purposes without compromising tree growth. Also, some researchers have altered lignin content in plants such as *Medicago truncatula* (ZHAO et al., 2013) and *Phoenix dactylifera* (SAIDI et al., 2013). Beyond this, the substantial regulation of this gene by *MYB* transcription factors is known (RAHANTAMALALA et al., 2010; MA; WANG; ZHU, 2011).

Plantation forests are essential for carrying the world's demand for wood and environmental sustainability, so it appears that biotechnology application to tree improvement will aid in speeding up this process (BONAWITZ; CHAPPLE, 2010; EUDES et al., 2014). Teak (*Tectona grandis* Linn. F.), from Lamiaceae family, is a deciduous tree with the most valuable commercial timber in the tropics, due to its high durability, dimensional stability and resistance to external environmental factors (LUKMANDARU; TAKAHASHI, 2008). It is used for furniture, buildings, finishes, cabinets, sleepers, decorative veneers, lamination, house walls flooring, joinery, carpentry, vehicles, mining and shipbuilding (BAILLÈRES; DURAND, 2000; BHAT; PRIYA; RUGMINI, 2001). Natural populations of teak are found in Laos, Myanmar, Thailand, Java Islands and India, with a worldwide planted area and natural forest of 33,381 ha (2,5 million m³ of wood) (KOLLERT; CHERUBINI, 2012).

Teak is also an interesting species for having one of the most expensive woods presenting naphthoquinones and anthraquinones (significant antifungal and antitermitic extractives); it has several environmental roles and can be used in agroforestry systems and forest recovery, making this species a profitable tree around the world (HEALEY; GARA, 2003; HALLETT et al., 2011). Unfortunately, despite being so important, there is a lack of genetic studies regarding gene expression and characterization (ALCÂNTARA; VEASEY, 2013; GALEANO et al., 2014). Consequently, to obtain *CAD* members related to lignin

biosynthesis in teak, this study amplified, cloned and characterized three partial *CAD* genes (*TgCAD2*, *TgCAD3* and *TgCAD4*) and a complete bona fide gene (*TgCAD1*). Their amino acid sequences were then analyzed and clustered by phylogenetical analyses. Finally, expression profiles were performed in several organs focusing in lignified tissues, including sapwood, thus making this the first study to provide *CAD* members of *T. grandis*.

3.2 Materials and Methods

3.2.1 Plant material

To amplify the *CAD* gene family, stem secondary xylem was used from 60-year-old trees, located in Piracicaba, São Paulo State, Brazil, after removing the bark, secondary phloem and vascular cambium (1,5 cm approx.) in order to reach the sapwood (Figure 1), which contains functioning vascular tissue. To perform quantitative real-time PCR, in vitro leaf blades and roots of two months were collected, along with branch and stem secondary xylem from 12-year-old trees, and sapwood from 12- and 60-year old trees was collected with a Pressler borer at DBH (DEEPAK; SINHA; RAO, 2010) (Figure 1) from fifteen trees in plantations (lat. 22°42'23"S, long. 47°37'7"W, 650 Meters Above Sea Level) located in Piracicaba, São Paulo State, Brazil. All samples were immediately frozen in liquid nitrogen and stored at -80°C.

Figure 1 - Methodology for collection of lignifying tissues in “Monte Alegre” teak plantations, Piracicaba, São Paulo, Brazil. A) Pressler borer. B) Cross section from the inner stem. C) 12-year-old borer. D) 60-year-old borer. P=pressler, Co=Cortex, S=Sapwood, H=Hardwood, Pi=Pith

3.2.2 RNA extraction and cDNA synthesis

Using mortar and pestle, frozen samples (1.0 gr) were ground using liquid nitrogen, followed by RNA extraction (mixing five samples as one replicate) using the protocol developed by Salzman et al. (1999). Total RNA (2 µg) from each sample was treated with DNase I (Promega). RNA quality was assessed running an agarose gel and using Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., USA), followed by a PCR reaction to ensure absence of DNA contamination, running an electrophoresis on a 1% (w/v) agarose gel with ethidium bromide staining. Treated RNA (1,0 µg) was used to synthesize cDNA using the SuperScriptTM III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer’s instructions.

3.2.3 Amplification of *CAD* family in *Tectona grandis*

3.2.3.1 Amplification of *TgCAD1*

Clustal alignment (<http://www.ebi.ac.uk/Tools/msa/clustalw2>) of *Citrus sinensis* (HQ841075.1), *Eucalyptus urophylla* (FN393570.1), *Populus trichocarpa* (XM_002313839.1), *Acacia auriculiformis* x *Acacia mangium* (EU275981.1), *Arabidopsis thaliana* (Z31715.1), *Bambusa multiplex* (FJ787493.1) and *Picea sitchensis* (BT071217.1) was performed to detect conserved domains of *Cinnamyl alcohol dehydrogenase* (*CAD*) gene (Additional File 1) in order to manually design the degenerate primers (Additional File 2), followed by a PCR using 100 ng of cDNA of stem secondary xylem from 60-year-old trees.

The single fragments corresponding to the proper size were excised, purified using Fragment CleanUp® (Invisorb, USA) and cloned using the CloneJet™ PCR Cloning Kit (Thermo Scientific, USA) and *DH5α*™ competent cells (Life Technologies, USA). Colonies were sequenced with the 3100 Genetic Analyzer (Applied Biosystems, USA) using the specific primers of the pJET1.2/Blunt vector.

Sequences were blasted (blastx), translated to amino acid sequences (<http://web.expasy.org/translate/>) and submitted to PFAM (<http://pfam.sanger.ac.uk/search>) to confirm *CAD* domains followed by the design of specific primers to amplify the internal region (Table 1). In order to obtain the extremes of the *TgCAD1* gene, the 3' and 5' RACE System for Rapid Amplification of cDNA Ends Kit (Life Technologies, USA) was performed, using the partial *cinnamyl alcohol dehydrogenase* sequence.

For the 5'- end, the first strand of cDNA was prepared with 5µg total RNA, GSP1 primer (100 nM; 5'-GCA CAC GAG ATC CAC TAT TTC A-3'), MgCl₂ (2.5 mM), dNTPs (400 µM), DTT (10 mM), SuperScript II RT (200 units), incubated 42 °C for one hour, 70 °C to stop reaction and finally added 1 µl RNase mix. SNAP column was used to purify cDNA, followed by an addition of Terminal deoxynucleotidyl transferase (TdT) and dCTP (200 µM) to create binding tails for the abridged anchor primer (AAP). PCR of tailed cDNA was performed using nested GSP2 primer (400 nM; 5'-CCC TGG AAC CAA AGG GTA TT-3'), Abridged Anchor Primer (400 nM), MgCl₂ (1.5 mM), dNTPs (200 µM), Taq DNA polymerase (2.5 units) and tailed cDNA (5 µl). For the 3'- extreme, the first strand of cDNA was obtained using 5µg total RNA, MgCl₂ (2.5 mM), dNTPs (500 µM), DTT (10 mM), SuperScript II RT (200 units), incubated 42 °C for one hour, 70 °C to stop reaction and finally added 1 µl RNase H.

PCR reaction was performed adding MgCl₂ (1.5 mM), dNTPs (200 µM), Taq DNA polymerase (2.5 units), Gene-Specific Primer GSP (200 nM; 5'-TTC ATC AGG TCA GGG GTG AG-3'), Universal Amplification Primer UAP (200 nM; 5'-CUA CUA CUA CUA GGC CAC GCG TCG ACT AGT AC-3') and cDNA (2 µl). Both 5' and 3' RACE were cloned and sequenced to complete the *TgCAD1* gene.

3.2.3.2 Amplification of *TgCAD2*, *TgCAD3*, *TgCAD4*

The cDNA sequences from thirteen *Populus trichocarpa* CAD genes were used (*PtrCAD1*, *PtrCAD2*, *PtrCAD3*, *PtrCAD5*, *PtrCAD6*, *PtrCAD7*, *PtrCAD8*, *PtrCAD9*, *PtrCAD10*, *PtrCAD12*, *PtrCAD13*, *PtrCAD14*, *PtrCAD15*) (Additional File 4) (BARAKAT et al., 2010; SHI et al., 2010) available at JGI (<http://genome.jgi.doe.gov/>) to design specific primers above the conserved domains (Additional File 5) in order to amplify the other teak CAD genes.

A pool of cDNA was obtained with stem secondary xylem from 60-year-old trees, branch secondary xylem from 12-year-old trees, in vitro leaves and roots with two months and used for the amplifications by PCR reactions using 100 ng of the pool. Gel fragments were purified (Fragment CleanUp®, Invisorb, USA), cloned (CloneJet™ PCR Cloning Kit, Thermo Scientific, USA), sequenced (3100 Genetic Analyzer, Applied Biosystems, USA) and analyzed by blastx. Translation (<http://web.expasy.org/translate/>) and domain evaluation (<http://pfam.sanger.ac.uk/search>) were also performed.

3.2.4 Characterization and modeling of the *TgCAD1* protein

TgCAD1 was aligned with the initial sequences used to obtain the gene in teak (*Citrus sinensis*, *Eucalyptus urophylla*, *Populus trichocarpa*, *Acacia auriculiformis* x *Acacia mangium*, *Arabidopsis thaliana*, *Bambusa multiplex* and *Picea sitchensis*). Elements of the secondary structure, domains and relevant residues, catalytic and structural sites were found following Larroy et al. (2002), Youn et al. (2006), Jin et al. (2014) and Tang et al. (2014) annotations. ProtParam Tool (<http://web.expasy.org/protparam/>) was used to explore protein characteristics. Using the *TgCAD1* protein sequence, the three-dimensional structure was predicted with the binary complex AtCAD5. The “.pdb” file was used for *TgCAD1* structural comparing and the amino acids were submitted to MODELLER (<http://www.salilab.org/modweb>). Figures were produced and edited in PyMOL program

(<http://www.pymol.org>) (GUO; RAN; WANG, 2010) and statistical validation (Additional File 3) was performed with PDBSUM (<http://www.ebi.ac.uk>).

3.2.5 Phylogeny and characterization of CAD family in *Tectona grandis*

Domains of TgCAD1, TgCAD2, TgCAD3, TgCAD4, and several members of the CAD family from *Arabidopsis thaliana*, *Oryza sativa*, *Lolium perenne*, *Zea mays*, *Nicotiana tabacum*, *Medicago sativa*, *Saccharum officinarum*, *Picea abies*, *Pinus taeda*, *Eucalyptus globulus*, *Theobroma cacao*, *Populus trichocarpa* and *Vitis vinifera* were aligned with Clustal W, run with neighbor joining method using 10,000 bootstrap replicates, poisson model and pairwise deletion, all in Mega 6 program, in order to build the phylogenetic tree.

Secondary structure elements, important relevant residues, and catalytic and structural sites were found following Larroy et al. (2002), Youn et al. (2006), Jin et al. (2014) and Tang et al. (2014) annotations. Peptide signals and protein subcellular localization was followed as in Jin et al. (2014) using Signal IP4.1 program (<http://www.cbs.dtu.dk/services/SignalP/>) and CELLO program v.2.5 (<http://cello.life.nctu.edu.tw/>), respectively.

3.2.6 Gene expression of the CAD family in *Tectona grandis* by qRT-PCR

Primers for qRT-PCR (Additional File 6) were designed above *TgCAD1*, *TgCAD2*, *TgCAD3* and *TgCAD4* sequences, using OligoPerfect™ Designer (Life Technologies, USA). Primer specificity was evaluated with the melting curve and the amplification efficiencies with the standard curve using three cDNA dilutions from the leaf (Additional File 7).

The mix for qRT-PCR contained cDNA from each sample (125 ng), primers (50 µM), SYBR Green PCR Master Mix (Applied Biosystems, USA) (12.5 µl), finally adding water for a total volume of 25 µl. The StepOnePlus™ System (Applied Biosystems, USA) was used for the PCR reactions as follows: 2 min at 50°C, 2 min at 95°C, 45 cycles of 15 s at 95°C, 1 min at 65°C, using 96-well optical reaction plates (Applied Biosystems, USA).

Each reaction was done with technical replicates, with negative control (absence of template), leaf was used as calibrator and *EF1α* as control gene (GALEANO et al., 2014). Statistical analyses were performed in SAS at 95% confidence level using F-test for ANOVA and LSD for pair comparison.

3.3 Results

3.3.1 Amplification of CAD family in *Tectona grandis*

3.3.1.1 Amplification of *TgCAD1*

Although lignified tissues are difficult to macerate, protocol from Salzman et al. (1999) was efficient to perform RNA extraction from stem secondary xylem, and consequently its cDNA was used to amplify the first *Cinnamyl Alcohol Dehydrogenase* gene in teak using degenerate primers designed above conserved domains from *Citrus sinensis*, *Eucalyptus urophylla*, *Populus trichocarpa*, *Acacia auriculiformis* x *Acacia mangium*, *Arabidopsis thaliana*, *Bambusa multiplex* and *Picea sitchensis*. An EST fragment of 985 bp was amplified using the primers TgCAD1f and TgCAD1r (Additional File 2), performing the standard RT-PCR process, followed by sequencing the internal portion of the gene using internal primers inTgCAD1f and inTgCAD1r (Additional File 2). After verifying the domains of the partial sequence, the principle of the Rapid Amplification of cDNA ends (RACE) was used. It is based in the amplification of the 5'- and 3'- ends of the gene from the partial cDNA using special adapters in the 5'CAP and poli (A) regions and designing specific primers. Therefore, from the partial *TgCAD1* EST, we obtained the complete sequence of *TgCAD1*, composed by 1071 bp, which could be translated into a protein sequence encoding 357 amino acid residues using ExPASy tool.

3.3.1.2 Amplification of *TgCAD2*, *TgCAD3*, *TgCAD4*

From the primers designed using thirteen *CAD* genes from *Populus trichocarpa* (Additional File 4 and 5), three *CAD* members in teak were amplified, called as *TgCAD2*, *TgCAD3* and *TgCAD4* (Additional File 8 and 9) from *PtCAD2*, *PtCAD5* and *PtCAD15*, respectively. We obtained 891, 657 and 753 nucleotides for *TgCAD2*, *TgCAD3* and *TgCAD4*, respectively. The deduced amino acid sequences of *TgCAD2*, *TgCAD3* and *TgCAD4* (297, 219 and 251 amino acids, respectively) (Additional File 9) showed, by blastp, the highest homology with *Mimulus guttatus* (86, 79 and 73% of identity, respectively), followed by *Eucalyptus grandis* (79% identity with *TgCAD2*), *Perilla frutescens* (77% identity with *TgCAD3*) and *Olea europaea* (72% identity with *TgCAD4*). The *CAD* protein sequences of teak were aligned against each other by clustal, but only *TgCAD3* and *TgCAD4* showed high identity (68%). The other comparisons were between 42-48% of identity.

3.3.2 Characterization and modeling of the TgCAD1 protein

Using ProtParam tool from ExPASy, TgCAD1 presented 38.96 kDa of molecular mass, a theoretical pI of 5.7 and a protein formula of $C_{1739}H_{2771}N_{459}O_{519}S_{17}$. Also, the CAD1 protein from teak was classified as stable (instability index of 27.07) and positive value for hydrophobicity (hydropathicity value of -0.020). Hydrophobic and hydrophilic residues represented 44.4% and 30.9%, respectively. The remaining residues were neutral. *TgCAD1* exhibited the alcohol dehydrogenase GroES-like domain” and the “Zinc-binding dehydrogenase” domains (Figure 2).

With blastp, TgCAD1 protein showed highest identity with *Ricinus communis* and *Prunus mume* (71%) followed by *Vitis vinifera* and *Jatropha curcas* (70%). When the *TgCAD1* gene is aligned with CADs from *Arabidopsis thaliana*, *Citrus sinensis*, *Eucalyptus urophylla*, *Populus trichocarpa*, *Acacia auriculiformis* x *Acacia mangium*, *Bambusa multiplex* and *Picea sitchensis*, several motifs are conserved among all species, although some features appear to be specific to teak, which are described later.

TgCAD1 protein presents a dimer with two zinc ions per subunit (YOUN et al., 2006a) when compared with AtCAD4 and AtCAD5, with several amino acids involved in catalytic and structural zinc ligation (LARROY et al., 2002).

Therefore, the TgCAD1 nucleotide binding domain is composed by six parallel β -sheet chains (β A to β F) flanked by five α -helices (α A to α E), while the catalytic domain consists in a core of antiparallel β -sheet chains (β 1 to β 9) with six helical segments (α 1 to α 6) on the surface of the TgCAD1 molecule (Figure 2 and 3). The ramachandran statistics showed 92% of most favored regions (Additional File 3).

3.3.3 Phylogeny and characterization of CAD family in *Tectona grandis*

When performing ClustalW with all of the TgCADs proteins with the sequences from *Populus trichocarpa* used to amplify in teak (PtCAD2, PtCAD5, PtCAD15) as well as two *Arabidopsis thaliana* protein sequences (AtCAD1, AtCAD8), we found several conserved characteristics in teak proteins which were identified previously in the alignments performed by Jin et al. (2014) and Tang et al. (2014). TgCAD1 and TgCAD2 exhibited the Zn1 binding motif GHEX₂GX₅GX₂V and the catalytic sites for zinc ion action (C47, H69 and C163 with TgCAD1 as reference) (Figure 4).

Figure 2 - Comparison of amino acid sequences of the *Cinnamyl alcohol dehydrogenase* gene in the species *Citrus sinensis* (*CsCAD1*, Genebank accession number ABM67695.1), *Eucalyptus urophylla* (*EuCAD2*, Genebank accession number ACU77870.1), *Populus trichocarpa* (*PtrCAD1*, Genebank accession number XP_002313875.1), *Acacia auriculiformis* x *Acacia mangium* (*AaCAD3*, Genebank accession number ABX75855.1), *Arabidopsis thaliana* (*AtCAD4*, Genebank accession number CAP09029.1), *Bambusa multiplex* (*BmCAD1*, Genebank accession number ADG02378.1), *Picea sitchensis* (*PsCAD1*, Genebank accession number ABK27071.1) and *Tectona grandis* (*TgCAD1*, Genebank accession number ABK27071.1). Elements of the secondary structures (alpha and beta) are defined above sequences. Domains of nucleotide ligation are squared. Conserved residues of glycine (GXGGXG) are determined by red dots. Identical amino acids are shaded with grey. Conserved residues which determine substrate ligation are shaded in brown. Green dots define places of catalytic zinc action while blue dots define places of structural zinc ligation

Figure 3 - Protein structures of (A) *TgCAD1* and (B) *AtCAD5*. α and β subunits, catalytic (red dots) and structural (blue dots) residues for zinc ion action and conserved glycine residues for substrate specificity (GXGGXG) are delimited

All the four CAD proteins of *T. grandis* had the Zn² structural motif with the coordinating cysteine residues for zinc ion ligation (C100, C103, C106 and C114 with TgCAD1 as reference) and the NADPH coenzyme binding motif GLGG[VL]G (usually called Rossman Fold) (Figure 4), which indicate that these CAD proteins from teak are alcohol dehydrogenases (zinc-dependent), belonging to the medium-chain dehydrogenase/reductase family. Only TgCAD1 presented the Phe299 and Asp123 residues as determinants of substrate specificity and binding, respectively (JIN et al., 2014).

Also, functional divergence of CAD members in teak can be understood through phylogenetic neighbor joining method using deduced amino acid sequences and CAD proteins from other species. The phylogenetic tree revealed seven groups and three classes of CAD with all the sequences used, including the localization of Monocots (Figure 5). TgCADs were classified in three groups and the three classes, supported by high bootstrap values (Figure 5). TgCAD1 was located in the group I, Class I (bona fide CAD).

TgCAD2 was positioned in the group V class III. TgCAD3 and TgCAD4 belonged to group II class II (A-SAD). Class I has been associated with lignin biosynthesis containing genes such as *AtCAD5*, *PoptrCAD4*, *OsCAD2* (BARAKAT et al., 2010; TANGet al., 2014), with *AtCAD5* being involved the three *p*-coumaryl, coniferyl and sinapyl alcohol (KIM et al., 2004).

Figure 4 - Amino acid sequence alignment of the *Tectona grandis* CAD proteins and other experimentally proved CADs from *Arabidopsis thaliana* (Genebank database) and *Populus trichocarpa* (JGI database). AtCAD1 (AY288079), AtCAD8 (AY302080), PtCAD2 (LG_XVI0159), PtCAD5 (LG_XVI2049), PtCAD15 (LG_IX000475). Multiple alignments were performed with the ClustalW2 software program. Zn1, Zn2, and NADPH binding motifs are squared. Conserved glycine residues for substrate specificity (GXGGXG) are determined by red dots. Identical amino acids are shaded with grey. Green dots define coordinating residues of catalytic zinc ion action while blue dots define coordinating cysteine residues involved in the structural zinc ion ligation. White star indicates key Phe299 residue for substrate specificity and black star indicates key Asp123 residue for substrate binding. Most of the alignment information was identified in Jin et al. (2014) and Tang et al. (2014)

Figure 5 - Phylogenetic tree of monocots, dicots and *Tectona grandis* CAD proteins. The neighbor-joining method was used with 10000 bootstraps. Teak CAD proteins are represented with a diamond. The bar indicates the evolutionary distance of 0.05. NCBI Accession numbers of sequences used to build the tree are *Arabidopsis thaliana*: AtCAD1 (AY288079), AtCAD2 (AY302077), AtCAD3 (AY302078), AtCAD4 (AY302081), AtCAD5 (AY302082), AtCAD6 (AY302075), AtCAD7 (AY302079), AtCAD8 (AY302080) and AtCAD9 (AY302076); *Oryza sativa*: OsCAD1 (AAN09864), OsCAD2 (BK003969), OsCAD3 (AAP53892), OsCAD4 (BK003970), OsCAD5 (BK003971), OsCAD7 (CAE05207), OsCAD8 (BK003972) and OsCAD9 (AAN05338); *Lolium perenne*: LpCAD1(AAL99535), LpCAD2 (AAL99536) and LpCAD3 (AAB70908); *Zea mays*: ZmCAD1 (AJ005702) and ZmCAD2 (Y13733); *Nicotiana tabacum*: NtCAD1 (X62343) and NtCAD2 (X62344); *Medicago sativa*: MsCAD1 (AAC35846) and MsCAD2 (AAC35845); *Saccharum officinarum*: SoCAD1(CAA13177); *Picea abies*: PaCAD1(CAA51226); *Pinus taeda* PtaCAD1(CAA86072); *Eucalyptus globulus*: EglCAD1 (AF038561); *Theobroma cacao*: TcCAD9 (EOY15101.1). *Populus trichocarpa* (JGI database): PtCAD1 (estExt_fgenes4_pg.C_LG_I2533), PtCAD2 (estExt_fgenes4_pg.C_LG_XVI0159), PtCAD3 (estExt_fgenes4_pg.C_LG_VI0462), PtCAD4 (estExt_Genewise1_v1.C_LG_IX2359), PtCAD5 (estExt_Genewise1_v1.C_LG_XVI2049), PtCAD6 (eugene3.00011775), PtCAD7 (eugene3.00020162), PtCAD8, (eugene3.00091019) PtCAD9 (eugene3.20690001), PtCAD10 (grail3.0004034803), PtCAD11 (gw1.VI.1869.1), PtCAD12 (gw1.XI.816.1), PtCAD13 (LG_I002927), PtCAD14 (LG_III001697), PtCAD15 (LG_IX000475) and PtCAD16 (LG_IX000970). *Vitis vinifera* (Plant Genome Database): VviCAD1 (TA39092_29760_D1b), VviCAD2 (GSVIVP00000463001), VviCAD3 (GSVIVP00002954001), VviCAD4 (GSVIVP00008718001), VviCAD5 (GSVIVP00008719001), VviCAD6 (GSVIVP00011478001), VviCAD7 (GSVIVP00011479001), VviCAD8 (GSVIVP00011484001), VviCAD9 (GSVIVP00011638001), VviCAD10 (GSVIVP00011639001), VviCAD11 (GSVIVP00011640001), VviCAD12 (GSVIVP00013365001), VviCAD13 (GSVIVP00014356001), VviCAD14 (GSVIVP00019568001), VviCAD15 (GSVIVP00029747001), VviCAD16 (GSVIVP00036661001) and VviCAD17 (GSVIVP00036664001). Groups (colored) were defined by BARAKAT et al. (2010). Classes were identified by Guo et al. (2010)

3.3.4 Gene expression of the CAD family in *Tectona grandis* by qRT-PCR

In order to explore the expression levels of the four *TgCADs*, teak leaves, roots, stem secondary xylem from 12- and 60-year-old trees and branch secondary xylem from 12-year-old trees were collected. By quantitative real-time PCR, expression levels of all *TgCAD* genes were detected for all tissues sampled with several variations, except for *TgCAD4* which showed either no gene expression or expressed at very low level in leaves (Figure 6). *Efla* was the endogenous control for this experiment (GALEANO et al., 2014). *TgCAD1* is three-fold more expressed in leaves than root and branch secondary xylem and ten-fold more expressed than stem secondary xylem at both ages.

Distinctively, *TgCAD2* is almost six-fold more expressed in branch secondary xylem and root than leaf and stem secondary xylem. *TgCAD3* and *TgCAD4* genes presented similar expression, with more expression in roots compared to the other tissues. *TgCAD4*, particularly, exposed almost 160-fold more expression compared to the rest of the tissues, with weak transcript level. All four *TgCAD* genes presented significant expression in roots. As lignin biosynthesis genes are expected to be expressed in secondary xylem, all *TgCADs* had transcript level in at least secondary xylem coming from branches of 12-year-old trees (Figure 6), with *TgCAD2* showing the highest expression level, followed by *TgCAD4*, *TgCAD3* and *TgCAD1*, although the last three do not show statistical significance with the F-test. *TgCAD4* had 20-fold more expression in branch secondary xylem than leaf.

3.3.5 Gene expression in sapwood from mature and young *T. grandis* trees

Furthermore, for a better understanding of the *TgCADs* genes function, sapwood tissue was collected and contrasted with leaves in the quantitative real-time PCR experiments (Figure 7). *TgCAD1* and *TgCAD2* genes have low expression in sapwood coming from both plant ages (12- and 60-year-old teak trees) with a statistical grouping difference evaluated by F-test. In contrast, *TgCAD3* and *TgCAD4* presented high expression levels in sapwood in both ages, but only *TgCAD4* showed statistical differences, with 60-year-old trees presenting almost 300-fold and two-fold more expression than leaf and 12-year-old trees, respectively. *TgCAD3* presented approximately 9-fold more expression in sapwood from mature than young teak trees.

Figure 6 - Expression of teak *CAD* gene family. Relative quantification of gene expression was studied in different tissues (leaf, root, stem and branch secondary xylem from different ages), shown at the bottom of the diagrams. The name of each gene is specified at the top of the histograms. $\pm$ means SE of the biological triplicates and technical replicates. * shows $p < 0.05$ according to F-test. Y axis indicates the relative expression level of each gene compared to the control tissue (leaves). *EF1a* was the endogenous control used according to Galeano et al. (2014)

Figure 7 - Expression of teak *CAD* gene family in sapwood at different developmental stages. The name of each gene is specified at the top of the histograms. $\pm$ means SE of the biological triplicates and technical replicates. * shows $p < 0.05$ according to F-test. Y axis indicates the relative expression level of each gene compared to the control tissue (leaves). *EF1a* was the endogenous control used according to Galeano et al. (2014)

3.4 Discussion

3.4.1 Characterization of CAD gene family

In our study, a bona fide CAD cDNA and protein of 1071 bp and 357 amino acids, respectively, were obtained from *Tectona grandis*. Its molecular mass and hydrophobicity are similar to CAD from other species. Referring to protein structure, α -helix and β -sheet were the main alternate features in TgCAD1, being interlaced with coils and turns (Figure 2 and 3), which were similar to previous predictions in *Arabidopsis thaliana* AtCAD5, *Sorghum bicolor* SbCAD4 and *Ginkgo biloba* GbCAD1 (YOUN et al., 2006b; SATTLER et al., 2009; CHENG et al., 2013; TANG et al., 2014), including the Rossmann foldable structure of $\beta\alpha\beta$ (ROSSMAN; MORAS; OLSEN, 1974) which is part of the NADP(H) binding domain.

Also, SKL motif is usually present in the carboxy terminus of several CAD proteins (TOBIAS; CHOW, 2005), which is a target sequence for peroxisomes and plays a role in subcellular localization. Our amino acid sequence analysis found no homology between the SKL residues of TgCAD1 with other species (Figure 2), advising absence of this protein in the peroxisome, as was found by Cheng et al. (2013) for *Lolium perenne*. The cartoon modeling style by PyMOL showed similar orientation in most of the residues of both AtCAD5 and TgCAD1 and allowed the visual localization of glycine residues along with the catalytic and structural zinc ion action (Figure 3). There are ten residues to be involved in stabilizing the aromatic ring of cinnamaldehydes with pi-bonding, related by Youn et al. (2006b) and Tang et al. (2014). TgCAD1 shows the presence of T49, Q53, L58, F299 and I300 residues when compared in alignment (Figure 2), but differences were found in N60, F119, A276, A286 and R290 TgCAD1 residues instead of M60, W119, V276, P286 and L290 AtCAD4 residues, respectively.

After using Signal IP 4.1 software, no typical signal peptides were found in all TgCADs. Several CADs have an evolutionary conserved SKL sequence at the C-terminus, which aids as a signal sequence to locate enzymes in peroxisomes (JIN et al., 2014). TgCAD1 has a SNL instead, suggesting that this protein may not be located in the peroxisomes (Figure 2). Subcellular localization prediction of TgCADs (Additional File 10) showed that they may exist in the cytoplasm. Several CADs contained in all proteins highly conserved GLGGVG motif, all the Zn-catalytic centre, Zn-binding site and key residues for correct activity of monolignols, such as *Pennisetum purpureum* CAD proteins (Tang et al. 2014), seven *Brachypodium distachyon* CAD genes (BUKH; NORD-LARSEN; RASMUSSEN, 2012;

TRABUCCO et al., 2013), BcCAD1 and BcCAD2 from *Brassica chinensis* (ZHANG et al., 2010) and all CAD genes from *Oryza sativa* (TOBIAS; CHOW, 2005).

Plants such as *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera*, *Brachypodium distachyon*, *Sorghum bicolor* and *Populus trichocarpa* present the entire standardized CAD gene family, whereby they can be used for reliable phylogenetic studies with other plants in order to understand functional differentiation and grouping. The four teak CAD proteins are part of three groups and three classes using amino acid sequences (Figure 5).

Usually, Class I is comprised by *bona fide* CAD genes, Class III is poorly understood (TANG, X. et al., 2014) and Class II is comprised by A-SAD type genes. *TgCAD1* is in group I and Class I, *TgCAD2* in group V class III. *TgCAD3* and *TgCAD4* are in group II class II. It is known that Class II and III contain monocots (groups III and IV in class II and group V in class III) and eudicots angiosperm sequences (Figure 5), which means that their evolution happened before monocot and eudicot differentiation (BARAKAT et al., 2009; TANG et al., 2014).

Gene duplication in the land plants ancestor of *CAD-like* genes in Class II and III (such as *PtCAD10* and *AtCAD6*) has been described previously (BARAKAT et al., 2010). *TgCAD1*, a *bona fide* CAD, is grouped close to *AtCAD4* and *AtCAD5*, which, as was previously described, participate in lignin biosynthesis and pathogen-infected tissues (TRONCHET et al., 2010). Several genes involved in lignin biosynthesis and content belong to Class II, such as *OsCAD7* (LI et al., 2009), *AtCAD7* and *AtCAD8* (KIM et al., 2004) and *PtCAD10* (BARAKAT et al., 2009).

It is possible that *TgCAD3* and *TgCAD4* are involved in lignin biosynthesis. *TgCAD2* belongs to Class III, where important genes such as *PtCAD2*, *AtCAD1*, *OsCAD1* and *OsCAD4* are located (Figure 5). Unfortunately, this class is not completely understood. A single CAD gene cannot be responsible for the lignin biosynthesis in plants, but rather several members of the family work together and may respond differently according to the external signals. There is a functional redundancy in which some members compensate functions from other genes, due to metabolic networks complexities (KIM et al., 2004; TANG, R. et al., 2014), meaning that some *CAD-like* genes such as *TgCAD2*, *TgCAD3* and *TgCAD4* could perform *bona fide* CAD gene functions.

3.4.2 Differential expression of *TgCAD* gene family

We evaluated the relative expression of four *CAD* genes in different teak tissues. They have expression in all the tissues analyzed, except for *TgCAD4* which did not present significant transcripts level in leaves (Figure 6 and 7). Presumably, *TgCAD* genes, which all are expressed in several tissues, especially in root, are involved in several functions, as lignin biosynthesis and secondary cell wall formation.

Although *TgCAD1* is part of the *bona fide CAD* group and is close to *AtCAD4* and *AtCAD5* (*Arabidopsis* genes related to lignin biosynthesis, biotic stress and high expression in root) (TRONCHET et al., 2010), it seems to have low relation with lignin biosynthesis due to its low expression in lignified tissues (Figure 6 and 7), at least without stress or treatments. As leaves are susceptible to pathogens and herbivorous insects, some genes such as *AtCAD4* and *AtCAD5* are highly induced by biotic attack (TRONCHET et al., 2010). Presumably, *TgCAD1* could be responsible for pathogen defense in leaves and highly expressed with biotic stress, since this gene is clustered with *AtCAD4* and *AtCAD5* and has high expression in leaves, but further studies with pathogen treatment must be done.

TgCAD2 and *TgCAD3* were more expressed in root and branch secondary xylem, while *TgCAD4* showed a significant expression level in root, followed by branch. Curiously, when evaluating expression in sapwood (which is composed primary of secondary xylem), *TgCAD1* and *TgCAD2* had no significant expression levels while *TgCAD3* and *TgCAD4* presented substantial transcripts level at both ages, with *TgCAD4* being 300-fold more expressed in 60-year-old teak trees when compared with leaves. Presumably, *TgCAD4* could be related to teak maturation and secondary wall deposition at latter stages of the tree.

In melon, five *CAD* genes were studied, with the exception of *CmCAD4* (similar to *TgCAD1*), and all of them presented high expression in roots, young stems and during sapwood development, as did our results (JIN et al., 2014). Also, *CmCAD1*, *CmCAD2* and *CmCAD3* presented differences in parts of the flower and during melon fruit developing stages. Additionally, *LtuCAD1* from *Liriodendron tulipifera* (XU et al., 2013) and only *BdCAD1* among the other six *CAD* genes in *Brachypodium distachyon* (TRABUCCO et al., 2013; BOUVIER D'YVOIRE et al., 2013) were predominantly expressed in xylem and roots.

As well, *Ginkgo biloba GbCAD1* (CHENG et al., 2013) and *Lolium perenne LpCAD1*, *LpCAD2*, *LpCAD3* (LYNCH et al., 2002) were expressed in lignifying tissues, similar to *TgCAD4*. In *Populus*, the *PtCAD10* gene, which is close to *TgCAD3* and *TgCAD4* phylogenetically (Figure 5), was 100-times more expressed in xylem compared to other *CAD*

genes, however *PtCAD9*, from the same group II, was preferentially expressed in leaves and xylem (BARAKAT et al., 2009). Above all, *AtCAD8* and *AtCAD7* (KIM et al., 2004), the most closely related amino acid sequences to *TgCAD3* and *TgCAD4* (Figure 5) have been suggested to be involved in lignin biosynthesis.

Albeit some species (as *Brachypodium distachyon*) have showed the existence of several *CAD* members with only one related with monolignol biosynthesis, others exhibit synergistic control over lignin production like *AtCAD4* and *AtCAD5* (TRABUCCO et al., 2013), which could be happening with *TgCAD3* and *TgCAD4*.

PtCAD12, closely related with *TgCAD2*, presented expression in all tissues but was greatly expressed in leaves (BARAKAT et al., 2009). It is noteworthy that the same authors found that *PtCAD2*, *PtCAD3*, *PtCAD5*, *PtCAD6*, *PtCAD11*, *PtCAD14* and *PtCAD15* have no expression differences between tissues without treatments. *CAD* gene duplications have been reported due to close clustering between sequences (BARAKAT et al., 2009, 2010), as *TgCAD3* and *TgCAD4* seem to be.

Notably, *Populus CAD* genes can change their functions and expression profiling under different stress conditions due to innumerable specialized motifs (MeJA, wound, defense responsiveness, ethylene) present in their sequences (BARAKAT et al., 2009). Therefore, the function of bona fide *CAD* genes (as *TgCAD1*) in lignin biosynthesis can be compensated, hence some *CAD/CAD-like* genes not associated with secondary xylem formation could be induced with biotic or abiotic stress, as herbivore damage in *Populus* leaves (BARAKAT et al., 2009).

In other cases, high expression levels of a *CAD* gene from *Picea sitchensis* have been observed in sapwood compared to bark after pathogen inoculation, meaning a fast response by this tissue (DEFLORIO et al., 2011). Also, *CAD* members can present highly divergent functions as is the case with *Camellia sinensis* genes (DENG et al., 2013). Finally, a functional specialization in *CAD* teak family can be suggested due to differences in clustering along with the expression profiles in the tissues tested, with a possible duplication in *TgCAD3* and *TgCAD4* genes but maintaining similar functions.

3.5 Conclusions and Perspectives

The methodology was successful in the identification and characterization of conserved domains of *TgCAD1* gene and modeling along with validation of 3D *TgCAD1* protein structure. *TgCAD3* and *TgCAD4* are potential enzymes responsible in lignin biosynthesis, which exhibited proper Zn-catalytic center and binding sites along with cofactor

binding motif. *TgCAD1* could be related with pathogen defense and biotic stress. Also, since *TgCAD3* and *TgCAD4* genes are clustered with lignin-related genes, they have high relative expression in secondary xylem and in sapwood during teak maturation, and they seem to be duplicated and be responsible for the synergistical control of the monolignol biosynthesis. Finally, *CAD* gene family in teak seems to have a crucial role in wood formation of this tropical tree, but studies of gene expression activating *CAD* genes in teak under some stress conditions or treatments would be interesting to investigate. Understanding how lignin biosynthesis occurs appears to be essential for future studies of plant transformation to improve growth rates in teak and adaptability, upward aiming transcriptomic studies, genetic improvement and selection, along with conservation of genetic resources. *TgCAD2*, *TgCAD3* and *TgCAD4* characterization of the complete sequence and functional studies *in planta* are essential for understanding their relation to secondary xylem formation and cell wall lignin deposition. *CAD* gene family characterization in teak supports the idea that monolignol formation process is conserved among different species of vascular plants.

Additional Files

Additional File 1 - Alignment of the *Cinnamyl Alcohol Dehydrogenase* gene of *Arabidopsis thaliana* (Z31715.1), *Acacia auriculiformis* x *Acacia mangium* (EU275981.1), *Populus trichocarpa* (XM_002313839.1), *Citrus sinensis* (HQ841075.1), *Eucalyptus urophylla* (FN393570.1), *Bambusa multiplex* (FJ787493.1) and *Picea sitchensis* (BT071217.1) (Continue)

Additional File 1 - Alignment of the *Cinnamyl Alcohol Dehydrogenase* gene of *Arabidopsis thaliana* (Z31715.1), *Acacia auriculiformis* x *Acacia mangium* (EU275981.1), *Populus trichocarpa* (XM_002313839.1), *Citrus sinensis* (HQ841075.1), *Eucalyptus urophylla* (FN393570.1), *Bambusa multiplex* (FJ787493.1) and *Picea sitchensis* (BT071217.1) (Conclusion)

Additional File 2 - Primers used to amplify *TgCAD1* gene

Pair Primer Characteristics	Name	Sequence	Primer size	Amplicon size
Amplify the beginning and end of the <i>CAD</i> gene of several species	TgCAD1f*	5'-GGATGGGCTGCAAGRGA-3'	17 bp	985 bp
	TgCAD1r*	5'-ACGAA <u>Y</u> CTGTACCT <u>V</u> ACATCG T-3'	21 bp	
Amplify the internal sequence of the <i>CAD</i> gene in teak	inTgCAD1f	5'-TTCATCAGGTCAGGGGTGAG-3'	20 bp	900 bp
	inTgCAD1r	5'-ACGCTTCCGATAAACTTGC-3'	20 bp	

*Degenerate bases: R=A+G; Y=C+T; V=A+C+G.

Additional File 3 - Ramachandran graphic to evaluate statistically the protein structure of
TgCAD1

Additional File 4 - Primers above *Populus CAD* gene family used to amplify *Tectona grandis*
CAD members

Gene name and accession number (JGI database) *	Primer Name	Sequence	Primer size (pb)	Amplicon
<i>PoptrCAD1</i> (LG_XVI0159)	Pt1F	5'-ATGAAATTGTTGGTATTGTGACA-3'	23	967pb
	Pt1R	5'-CATTATCCATGCCTCGGAAC-3'	20	
<i>PoptrCAD2</i> (LG_I2533)	Pt2F	5'-GATGTTGGGAAAGATGACATTTC-3'	23	826pb
	Pt2R	5'-TATGTTTTGTGCCTCCTGTTGC-3'	22	
<i>PoptrCAD3</i> (LG_VI0462)	Pt3F	5'-GCACTTCATTGTTCTGATTCCA-3'	22	574pb
	Pt3R	5'-CATGGCAGTGTTGACATAGTCC-3'	22	
<i>PoptrCAD5</i> (LG_XVI2049)	Pt5F	5'-AAAAATTCAAGGTCGGAGACAA-3'	22	650pb
	Pt5R	5'-GCAACTACCTCCCACTGTCTTC-3'	22	
<i>PoptrCAD6</i> (eugene3.00011775)	Pt6F	5'-GTAGGCAGCAAGGTTGAAAAGT-3'	22	754pb
	Pt6R	5'-AATCCATTGGAATCACCTCAAC-3'	22	
<i>PoptrCAD7</i> (eugene3.00020162)	Pt7F	5'-TCGCATGAGAATTATTGTGACC-3'	22	594pb
	Pt7R	5'-TTCTTTCGTTGATCCTGTCATG-3'	22	
<i>PoptrCAD8</i> (eugene3.00091019)	Pt8F	5'-CCCTTGAGATTTTACGGTCTTG-3'	22	491pb
	Pt8R	5'-ATGGCAGTGTTACATAATCCA-3'	22	
<i>PoptrCAD9</i> (eugene3.20690001)	Pt9F	5'-AGGATGTGAGATTCAAGGTGCT-3'	22	969pb
	Pt9R	5'-TTCAAATCTTCATTGTGTTGCC-3'	22	
<i>PoptrCAD10</i> (grail3.0004034803)	Pt10F	5'-ACCATCACTTACGGTGGCTACT-3'	22	839pb
	Pt10R	5'-CCAGGGTATTGCCGATGTTAAG-3'	22	
<i>PoptrCAD12</i> (gw1.XI.816.1)	Pt12F	5'-TCACTTTTAATGGCATTGATGC-3'	22	586pb
	Pt12R	5'-AGCAGCACAGAAGTCCAACATC-3'	22	
<i>PoptrCAD13</i> (LG_I002927)	Pt13F	5'-AATAGGCTCGTGTGGAAAATGT-3'	22	643pb
	Pt13R	5'-CATTTCTTGCGTCTCCTTTACC-3'	22	
<i>PoptrCAD14</i> (LG_III001697)	Pt14F	5'-AAGAAGTAGGCTCCAATGTCCA-3'	22	698pb
	Pt14R	5'-TTCTTGTGTCAATTTTCGTCCC-3'	21	
<i>PoptrCAD15</i> (LG_IX000475)	Pt15F	5'-GTTGGGGAAGTGACAGAGGTAG-3'	22	789pb
	Pt15R	5'-CATGGCTGTGTTACATAATCC-3'	2	

* *Populus* genes were mentioned in Barakat et al. (2010).

Additional File 5 - *CAD* sequences from *Populus trichocarpa* family. Primers used in this study are underlined. The “I” between nucleotides defines the position of an intron in the DNA of the *CAD* gene family

(Continue)

>PoptrCAD1

TATATACACACAGATCACTGGGTTATTGTTCTGCACTGAGTTTGAACATCATCAAGGAAGTGAAAAGCAGAGAGAATGGCAAAATCACCAGAAG
TAGAACATCCACATAAGGCTTTTGGCTGGGCTGCCAAAGATAGTTCTGGGGTCTTTCTCCCTTTTCATTTCTCAAGAAGG**I**GACAATGGAGTT
GAAGATGTGACCATAAAAAATCCTGTACTGTGGAGTTTGCCATTGGAGCTTGCACGCTGCCAAGAATGAATGGGGGTTTCCAGATATCCTTTGG
TTCCTGGG**I**CATGAAATTTGTTGGTATTGTGACAAAAATTGGAAGCAATGTGAAGAAGTTCAAAGTGGACGATCAGGTTGGTGTGGAGTACTG
GTGAACCTCCTGTAAGTCATGCGAGTATTGCGACCAGGACTTGGAGAATTACTGCCCTAAAATGATATTTACATACAATGCCCAAAACCATGATG
GGACAAAACTTATGGTGGTTATTCTGATACAATTGTGGTTGACCAGCACTTTGTACTCCGTATTCTGATAGCATGCCTGCTGATGGGGCTGC
ACCACTATTATGTGCTGGGATCACAGTGTACAGCCCAATGAAATATTATGGAATGACAGAACCAGGGAAGCATTGGGAATCGTAGGATTGGGG
GGGCTTGGACATGTGCTGTGAAGATTGGTAAGGCCTTTGGTTTGAAAGTTACAGTCATCAGTTTCATCATCAAGAAAGGAGAGCGAAGCACTTG
ATAGACTTGGTGCTGATTCAATTCCTTGTGAGCAGTGACCCTGAGAAAATGAAGG**I**CAGCATTTGGCACTATGGATTACATCATTGACACTGTG
TCTGCAGTTCATGCCTTGGCTCCACTTCTTAGTCTGCTGAAGACAAATGGAAAACCTGTTACTTTGGGCTTGCCCTGAGAAGCCCTTGAGCTGC
CTATCTTCCCTTTGGTCTTGGG**I**GCGAAAGCTAGTTGGTGGAGTGATATTGGAGGGGTGAAAGAGACTCAAGAGATGTTGGACTTCTGTGCG
AAGCACAATATTACCTCAGATGTTGAGGTGATCCGAATGGATCAAATCAACACAGCCATGGATAGGCTTGCCAAATCGGATGTCAGGTACCGGT
TTGTGATTGATGTGGCAACTCCCTGTGACAATCTCAGTTATGAAGCCCTCGATTCTCTTGGTCTTTTTCAGCTAAACCAAGTCACGATATG
CTTCGTATACTGTCCCTTCTCTGTTCCGAGGCATGGATAATGGAATAAGGTTTAATAAGGTTTTAGCACCAGATCTTTTACATTTTGCGCTC
CTTGGAGTGGTAGAACATTTGAAGTTCTGTGTCTCGGACTTGTGTTCTGTTTGTGTTTAAATTATATCTGATCGCCTAGTAATGATTTCGATAAC
AAGCA

>PoptrCAD2

ACAGCTGTGCTTATCCCTTCTCTTCTGCTCTACGCTGAGCATTTATCCTAGCTCTAGAGCATTTTACAGGTC**I**ATTGTTGAAGTTCTGCTT
CATTTGCTTTGAGAGAAGTAGAGGATGGCTTCAGATAAGAGCGAGAATTGCCTTGCCCTGGGCTGCAAAAGATGAATCAGGAGTCTGTCCCGTA
TAAATTTAAAGAAGG**I**GATGTTGGGAAAGATGACATTT**CA**GTAAAAATAACACACTGTGGAATTGCTATGCTGATGTTCTCCTTACCAGGA
ACAAATTCGGAAGTCATTATACCCAGTAGTGCCAGG**I**TCATGAGATAGTTGGAAGTGTTCAGAGGTTGGATCTGATGTTCAACGCTTCAAA
ATCGGCGACCATGTTGGAGTGGGAACATTTATCAATTCATGCAGGATTCGAGATATTGTAATGATGGGCTTGAAGTGCATTGTGCAATGGAA
TTATTACCACCATTAACAGTGTGTGATGTCGATGGCACCATCAGAAAAGGAGGGTACTCCAGTTTTATTGTTGTTTACAGAAAG**I**ATACTGCCAC
AGAATACCTGACGGTTATCCGCTAGCTTTAGCAGCACCATTGCTGTGCGCTGGCATCACAGTCTACACCCCATGATTCGTGACAAGATGAACC
AACCTGGAATTCCTCGGGGTGATTGGACTGGGCGGCTTGGTCACATGGCTGTGAAGTTGGGAAGGCTTTTGGGAATGAATGTCACAGTTTT
CAGCACAAGCATATCGAAAAAGGAGGAAGCCTTGAATCTGCTTGGAGCAGACAACTTTGTAGTCTCGTCGGACACGGAGCAATGAAGG**I**CTC
TAGATAAATCCCTGGACTTCATTATTGACACAGCATCAGGTGAACATCCATTGATCCATACATTACAACCTGAGACCGCTGGAGTCTGGT
CCTGGTGGGCGCTCCAAGTGAATGAAGCTCACTCCTCTGAAGCTGCTTCTTGGTA**I**TGATATCCATTTCTGGAAGT**GCAACAGGAGGCACAA**
AACATACACAGGAATGCTGGACTTTTGTGGTACTCACAAATCTACCCCAAAGTAGAAGTCATACCAATTCAGTCTGTGAATGAAGCACTTGA
GAGGTTGATAAAGACGATGTGAATAACCGATTGTGATTGACATTGGAATCTCCCTCAAGTGA

>PoptrCAD3

TTGAGTGCATATTGATCAGGAAGCCAAATCTTGATCTTACAGTTGTATTTTATAGAGAAAAATGGTAGCCAAATTGCCAGAGGAAGAGCATC
CAAAACAGGCCTTCGGATGGGCAGCAAGAGACCAATCTGGGGTCTCTCTCCCTTCAAATCTCCAGGAG**I**AGCAACAGCAGAAAAAGGATGTG
GCATTCAAGGTGTTGATTGTGGGATATGCCACTCCGACCTTCACATGGCCAAGAATGAATGGGGCGTTACTCAATACCCTCTTGTCTCTGGG
ICATGAGATTGTGGGAATAGTGACAGAGGTGGGGAGCAAAGTAGAAAAATTCAAGGTTGGAGACAAAGTGGGTGTAGGGTGCATGGTTGGATC
ATGCCACTCTTGCGATAGTTGTGACGACAATCTTGAGAATTACTGTCCAAAAATGATACTTACCTATGGTGCCAAGAAATTATGATGGCACCATC
ACATATGGAGGCTACTCAGACCTTATGGTTGCCGAAGAGCACTTCATTGTTTCGTATTCAGATAATCTATCTCTTGTGCTGGTGTCTCTCTCT
TGTGTGCTGGCATCACAGTATATAGCCCTTGAGGTATTTTGGACTTGACAAACCCGGTATGCATGTGGGTGTAGTCGGCTTGGTGGTCTAGG
TCACGTAGCAGTGAAATTTGCAAGGCTATGGGGGTCAAGGTGACAGTGATTAGCACCTCTCCTAACAAGAGCAAGAGGCTGTAGAGAATCTT
GGTGTGACTCTTTTTTGGTTAGTAGTGACCAGGGTCAGATGCAG**I**CTGCAATGGGCACATTGGATGGTATCATTGATACAGTGTCCGCAGT
TCACCCTATGTTGCCCTTATTACTCTATTGAAGTCTCATGGAAGCTAGTTTGGTTGGTGTCTCCAGAGAAGCCTCTTGAATTACCTGTCTTT
CCTTTGATCGGCGG**I**GAGGAAGATGGTGGGAGGTAGTTGCATCGGAGGAATGAAGGAGACACAAGAGATGATTGATTTTGCAGCCAAACACAA
TATAACAGCCGACGTCGAGGTTATTCCGAT**GGACTATGTCAACACTGCCATG**GAGCGCATGCTAAAGGAGAGCTTAGATATCGATTGTCTATC
GACGTGCCAAACCACTAAACCCTTAG

>PoptrCAD5

ATACATAGAGAAAAATGGCAGACAAATGCCAGAGGAAGAGCATCCAAACCCGGCTTTGGATGGGCTGCAAGAGACCAGTCTGGGGTCTCTCT
CCCCCTTCAAATTTCCAGGAG**I**AGCAACAGGGGAAAAGGACGTGGCATTCAAGGTGTTGATTGTGGGATATGCCACTCCGATCTTCACATG
GTCAAGAATGAATGGGGCGTTACTCAATACCCTCTTATCCCTGGG**I**CATGAGATTGTGCGAGTAGTGACAGAGGTGGGGAGCAAAGTAGAAAA
ATTCAAGGTGCGAGACAAAGTGGGTGTAGGGTGCATGGTTGGATCATGCCGCTCTTGCGATAGCTGTGACAAACATCTTGAGAATTACTGTTCA
AAAAAGATACCTTACTTACCGTGCCAAGTACTATGACGGAACCGTCACATATGGAGGCTACTCAGATAATATGGTTGCTGATGAACATTCATTG
TTCGTATTTCAAATAACCTACCTCTCGATGCTGGTGTCTCTCTCTGTTGTGCTGGAAATCACAGTCTATAGCCCTTGAGATATTTTGGACTTGA
CAAACCTGGTATGCATGTGGGCATTGTTGGCCCTCGGTGGTCTAGGTCATGTAGCAGTGAAATTTGCAAGAGCTATGGGCGTCAAGGTGACGGTG
ATTAGTACCTCTCCTAATAAAAAAGCAGGAGGCTCTAGAAAACTTTGGAGCTGACTCGTTTTTGGTCAGTCGTGACCAAGATCAGATGCAGG**I**C
TGCAATGGGAACATTTGGATCATAGTGTACAGTGTCCGAGTTCACCTCTGTTGCTTGTGCTCTATTGAAGTCTCATGGAAAGCTA
GTTTTGGTGGTGTCTCCGAGAAGCCTCTTGAAGTCCCGCTCTTTCTTTGATCACCGG**I**GAGGAAGACAGTGGGAGGTAGTTGCGTTGGAG

Additional File 5 - *CAD* sequences from *Populus trichocarpa* family. Primers used in this study are underlined. The “I” between nucleotides defines the position of an intron in the DNA of the *CAD* gene family

(Continuation)

```
>PoptrCAD6
GCATTACCATCCACAACCTTAATTCGTGCAAAAACAGAGGAAAAAGAAATGGCAGCAAAATCTTGTACAGGAAGGCATCCCACCTGAGGCTTTTGGG
TGGGCAGCAAGAGACCACTCCGGGGTCTCTCTCCTTTCAAATTTCTCTCGGAGGIGCAACAGGAGAGAAGGACGTCGCATTCAAGGTGCTGTT
CTGTGGAATATGTCACCTCGGACCTTCATATGATCAAGAATGAGTGGGGTATCTCTTCTACCTGTTGTCCCCGGGICATGAGATTGTGGGAC
AAGTGACAGGGGTAGGCAGCAAGGTTGAAAAGTTCAAAGTTGGAGATAAAGTTGGGGTAGGGTACATGGTTGGATCATGCCAATCTTGCATAG
TTGTATGACGATCTCGAAAATTACTGCCCAGACACAATAGTCACCAAGTGGTGGCAAGTACCATGATGGAGGCCACCACATACGGAGGCTTCTCA
GACATTATGGTCGCAGATGAGCACTATGTAATTCGAATTCAGAGAATTTGCCTCTTGATGCCGGTGCTCCTCTCCTATGTGCTGGGATTACAG
TGTATAGCCCCCTTGAAATATTATGGCCTTGACAAACCAGGTATGCATGTGGGTGTAGTCGGGCTTGGTGGGCTGGGTGATGTAGCTGTAAAGTT
TGCAAAAGCTATGGGGATCAAGGTGACAGTGATCAGTACCTCTCCAAAAAAGAACAGGAGGCTCTTGAGCATCTTGGCGCTCATTCATTTTGG
GTTAGTCGTGACCCCGATCAGATGCAGGICTGCAATAGGCACAATGGATGGTATAATTGACACGGTCTCGACGATGCACCTCTCTTCCCTTT
GATTGGTCTGTGTAAGACTCAGGGAAAGCTGGTTTTTGGTTGGTGCGCCGGAGAAGCCACTTGAGCTACCAAGTGTTCCTCTTATCATGGGIAA
GGAAGATAGTGGGTGGTAGTAGCATCGGAGGAATAAAGGAAACACAAGAGATGATTGATTTTGCAGCCAAGAACACAATAACAGCAGACGTTGA
GGTGATTCCAAATGGATTATGTGAACACTGCCTTGGAGCGGCTATCGAAATCAGATGTTAGGTACAGATTTGTGATTGACATTGGCAATACATTG
AAGAATTGATGTTTGTCTTCCATATACCTCAGTGGAGAAATGGCCTGTGTTTTGAGGATTCGTAGGTCGAATCATTTTTCAGCTTTTCAGCAAT
AAGTTTTTGTCTGTTAACGAAACATATATCGGCTTGTAATGCTATGGTGGGCAAGGGAATTGGAAGGCCAACCCCTGTGGTTCTCTTTCTTC
CAATCCATTACTTTGGATCTACAAATATTGCTCGCCTTATTTTAAATACTAATTTG
>PoptrCAD7
ATACTCGAGAGAGCATCTTCTTTTCAAGCGACCCATTTTCCCTGCTCTGTCCGAGCAGAGCCTCGAGTCTCAGGTTATGGCTCAAACAAC
TCAAACCACACGCAGACCGTTTCTGGCTGGGCAGCCCTCGACTCTTCTGGCAAAGTCGTCCTTACACTTTCAAAGAAGGIGAGAATGGTGT
CAACGATGTGACCATAGAGATTATGTATTGCGGCATATGTCTACTGATCTCCATTTTGCAAGAAGCATTGGGGCATTTCCATGTACCCAGTT
GTCCCCGGGICATGAAATTACTGGGATAATCACGAAGGTGGGGAGCAACGTGAACAAGTTCAAGTTGGGAGACAGAGTTGGAGTGGGGTGCTT
AGCAGCTTCATGTTTGGAGTGTGATTCTGCAAGAGCTCGCATGAGAATTATTGTGACCAATACAGTTGACTTACAATGGTATTTCTGCGGAT
GGTAGCATCACTTATGGTGGCTACTCTAAATTCCTGGTTGCAGATCACAGGTATGITGGTGGCGGATACCAGAAAACCTGCCAATGGATGCAAC
AGCGCCCTCTCTGTGCAAGAAATTACTGTATTACGCGCTTTCAAGGATAGCAACTTGGTCGACACACCAGGAAAAAGGGTAGGGGTGGTTGGT
CTCGGAGGTCTAGGACATGTAGCTGTCAAGTTTGGCAAGGCATTTCCGTCAACATGTAACTGTGATCAGCACCTCTCCGTCTAAAGAAAAAGAGG
CCAGAGAACGCTTGGGAGCTGATGATTTTATTGTGACGACTAATGCCCAAGAACTTCAGGICAGCGAGAAGAACTCTAGATTTTATCGTGGAT
ACAGTGTGAGCTAAGCATTCTCTGGGGCCAACTAGAAATTACTCAAAGTAAACGGGACATTGGCAGTGGTATGCGCACCCAGACCAGCCGATGG
AACTGCCAGCTTTTCTCTGATATTCGGICAAGAGATCTGTGAGGGGCAGCATGACAGGATCAACGAAAGAAACACAAGAGATGCTGGATGTG
TGTGGCAAGCACAACATAACTTGTGACATTGAGCTTGTGAAAACAGACAACATTAATGAGGCTTGGGACCGCTTGAAGAAACGATGTCAGAT
ATCGTTTTGTGATGACATTGCTGGAAAGTCTTCAAATCTTTAGAGATACAGAAGTGTCCCTCACATAATAGTATATTCCTTTTCTTTTAAAT
TTTGTGGGCATATTTCTTTTGTTCGTGTGTCGCCGAAATAAACAGTTATTGCAAAATTCGGTTGGATCATT
>PoptrCAD8
ATGGCAGAAAAATCTTACGAGGAAGAACATCCTACCAAGGCTTTTGGATGGGCAGCCAGAGACCAATCCGGGGTCTCTCTCTCTTCAAATTTCT
CCAGGAGGTICAACAGGAGAGAAGGATGTGAGATTGAAGGTGCTGTTTTGTGGAATATGTCACACAGACCTTCATATGGCCAAGAATGAGTGG
GGTAATTCCACGTACCTCTAGTTCTCTGGGICATGAGATTGTTGGGCAAGTGACAGGAGTAGGAAGCAAAGTTGAAAAGTTGAGATTGGAGA
CAAAGTGGGGTGGCGGGCATGATTGGATCATGCCACTCTTGCATAGTTGCAGCAACAATCTTGAATAATTACTGCTCGGAAGTGATAATCACA
TATGGTGCAAAATACCTAGATGGAACCAACACATATGGGGGCTACTCGGACATTATGGTCGTAGATGAGCACTTCGTAGTTTCATATCCAGACA
ACCTATCTCTTGATGACGCCGACCTCTCCTATGCGCTGGAATTACAGTGTATAGCCCTTGAGATTTTACGGTCTTGACAAACCGGCTATGCA
TGTGGGTCTGTTGGCTGAGTGTGCTGAGTGTAAAGTTTGCAAAAGCTATGGGGGTCAAGGTGACAGTTATTAGCAACCTCAGCC
AACAGAAACAAGAAGCCCTCGAGCATCTTGGTGCCGACTCATTTTTGGTTAGCCGTGATCAGGATCAGATGCAGGICTGCAATGGGCACAAT
GGATGGTGAATTGACACGGTGTGCGCCATGCATCTATCTTGCCTTTGATTAGTCTATTGAAGACTCAAGGAAAGCTGGTTTTGTTGGTGGTGG
CCTGCGAAACCACTTGAGCTACCAAGTGTTCCTTTGATCGTGGGIAAGAAAAATAGTGGGTGGGAGTGTGGTGGAGGAATGCAGGAAACACA
AGAGATGATTGATTTTGTCTGCTAAGAAACAACATAACAGCAGATATCGAGTTGATTTCATGGATTATGTGAACACTGCCATGGAGCGGCTATTG
AAAACATGATGTTAGGTATCGATTTGTCATTGATATTGGCAACACAATGAAAAATTGA
>PoptrCAD9
GGCAGTACCAAATAAAGCAAAAAAGAGGGTGGAAATGGCAGAAAAATCTTACGAGGAAGAACATCCTACCAAGGCTTTTGGATGGGCAGCCAGA
GACCAATCCGGGGTCTCTCTCCTTTCAAATTTCTCCAGGAGGTICTACAGGAGAGAAGGATGTGAGATTCAAGGTGCTCTTTTGTGGAATATG
TCACTCAGACCTTCATATGGCCAAGAATGAGTGGGGTACTGCCACGTACCTCTAGTTCCCGGGICATGAGATTGTTGGGGAAGTGACAGAGG
TAGGAAGCAAAGTTGAGAAGTTCAAAGTTGGAGACAAGTGGGGTGGGTGCTTGGTTGGATCATGCCACTCTTGGCATAGTTGCAACAACAA
TCTCGAGAATTATTGCCCAAAAAATGATACTCACCTACAGTACCAAAATACCAGATGGAACCAACAGTACGGAGGCTACTCAGACAGCATGGTC
ACGGATGAGCACTTCTGTAATTCGTATTCCAGACAACCTACCTCTAGATGCCGCTGCACCTCTCCTATGTGCTGGGATTACAGTTTACAGCCCC
TGAGGTTTTTTAATCTTGACAAACCGGGTATGCATGTGGGCGTGGTTGGGCTCGGCGGGCTAGGTCATGTAGCTGTAAAGTTTGGCAAGGCCAT
GGGGTCAAGGTTACAGTTATTAGCACCTCTCCCAAGAAGAACAAGAAGCCCTTGAGCATCTTGGTGCTGACTCGTTTCTAGTTAGTCGTGAC
CAGGATGAGATGCAGGICTGCAATGGGCACAATGGATGGTGAATTGACACGGTGTGCGCCATTATCCTATCTTGCCTTTGATTAGTCTATT
GAAGACTCAAGGAAAGCTGGTCTTGGTTGGTGCCTGAAAAGCCACTTGAGTACCAAGTGTTCCTCTGATCATGGGIAAGAAAAATAGTGG
GAGGAGTACCATAGGAGGAATGAAGGAAACACAAGAGATGATTGATTTTGTGCGCAAGAACAACATCACGGCAGACATTGAGGTTATCTCGAT
GGATTATGTGAACACAGCCATGGAGCGGCTTTCGAAAACAGATGTGAGATACCGATTGTTTATCGACATCGGCAACACAATGAAGATTGAAAT
CTCTACCTCATAAACACTGTACGAATAAAACGCATGTAAAGTGAGATTTCGTAGTTTGAAAAATCTTAAGATTTTTCGGTGCCTCTATGAGCAT
TGAGCACAATTTAATTATGGTTTGAAATTTGAAGCAATGATGTGACGTTATTAGTACCTCTATTTTGTGTTAAATTTCTGTTCAAGTATTTACG
CTTTTAT
```

Additional File 5 - *CAD* sequences from *Populus trichocarpa* family. Primers used in this study are underlined. The “I” between nucleotides defines the position of an intron in the DNA of the *CAD* gene family

(Conclusion)

>PoptrCAD12

ATGAGCTCCGAGGGTGTAAAGATGACTGTCTTGCTTGGGCAGCGAGAGACCCCTCTGGAGTCTTATCTCCTTACAAGTTTAGTCGCAGG**I**GC
TCTTTGGAAAAGATGATGTTTCGCTAAAAATAACGCACTGTGGAGTTTGCTATGCTGATGTTATCTGGAGTAAGAACAAGCATGGAGATTCGCGC
TATCCATTGGTGCCCGG**I**ACATGAGATTGCTGGAATTGTAAAGGAAGTTGGATCCAGTGTGACGAACCTCAAGGTTGGTGACCATGTTGGAGT
AGGCACTTACGTTAATCTTGCAGAGAATGTGAGCATTGCAATGACAAGGAAGAAGTTAGTTGTGAAAAAGGATCAGTTTTCTACTTTTAATGGC
ATTGATGCTGATGGTTCGATTACAAAGGGTGGATATTCTAGCTACATCGTTGTCCATGAAAAGGTA**I**CTGCTTTTCGGATACCTGATGGTTATCC
TTTGGCCTCTGCAGCACCTCTGCTTTGTGCTGGAATCACTGTGTACAACCCCATGATGCGACATAAGATGAACCAACCTGGTAAATCTCTTGA
GTGATTGGGCTCGTGGTCTGGGTACATGGCAGTGAAGTTTGGCAAGGCTTTTGGCTTGAAAGTAACTGTTTAAAGCACAGCGTATCTAAAA
AGGAGGAAGCCCTGAGTGTGCTCGGCGCAGACAATTTCTGATTACATCCGATCAAGCGCAGATGAAGG**I**CCTTGACAAATCACTAGACTTC
ATAATCGACACAGCATCCGCGCATACCCATTTGATCCATACTTGCTCTTTTGAAGACTGCTGGTGTGTTTTGTTCTGTTGGTTCCCAAGTGA
AGTCAAATTCAGTCTGCGAGCCTCAATATCGGTA**I**TGAAAACCTGTAGCTGGTAGCATAACAGGTGGTACAAGAGTGATCCAAGAGATGTTGG
ACTTCTGTGCTGCTAATAAAATTTACCCCGGGATCGAAGTAATCCAATTCAGTATATAAATGAAGCTCTTGAGAGGATGGTAAAGAACGACGT
GAAGTACCGTTTGTGATTGATATTGAGAACTCCTTGAAG

>PoptrCAD13

ATGTCAAGGTTACAAGCAGAAGAAGAACTCAAAGGCTTTTGGATGGGCAGCTAGAGACTCTTCAGGAATTTTATCCCTTTCCACTTTACTA
GAAGG**I**GTTAATGGAGATAACGATATTACAATCAAGATTTTATATTGTTGGGATTGGCATTCTGACCTGCACATAGCTCGGAATGATTTTGA
ATTCTATCTATCTGTTGTTCCCGGG**I**CATGAGATCGTTGGTGTGGTGACCCAGGTTGGGAAAAAGGTAGACAGGTTCAAGGTCGGAGACAA
AGCAGGTGTGGGATGCTTAAATAGGCTCGTGTGGAAAATGTGAGAATTGCCAAAACACATGGAGAGTTACTGCTCCAAAACCTGTTTACTTTT
ACTATTTTTTATGACGCGGAGACAGAATTATGGTGGATATTCTGATGTATATGTTGTTAACGAGCACTTTGCCATTCTGTTTCCGGATAATC
TCTCATTAGGAGGTGGAGCTCCGCTGCTCTGTGCTGGGATTACAGTTTTTAGTCCCATGAAGTATTTTGGGCTCGACAAGGCTGGGATGCATCT
GGGGTGGTGGTCTTGGTGGCCTGGGTCAATTTGGCGGTGAAGTTTGCGAAGGCATTGGAATGAAAGTGACTGTGATCAGCACATCTCCGAGC
AAACAGCATGAAGCAATTGAGCAACTCAAAGCTGACTCGTTTATAGTTAGTCATGACATGAAGCAATGGAGG**I**CTGCTACTGGCACCATGAA
CGGTATCATAGACTGTGCTGCACTTCACTCTAAAGCCATTGCTTGATCTCTTAAAGACTAATGGAATAATATTGGTGGGTGCCCAA
AGTCTTGAAAAGCCCTTGGAGGTGCTGCAATGCCCTCTACGG**I**AAGAAAGCTAGTGTGAGGAAGCATGGCAGGGGGGTAAAGGAGACGCA
AGAAATGATTGATTTTGTGCTGCCGAACACACATAGAGGCTAACGTTGAAGTTATCCCATGGATTATGTGAACAAGGCCATGGACCGTCTTGCA
AAAGGAGCTGTCCGTATCGATTTGTTATTGACATAGCAACACCCCTTTAA

>PoptrCAD14

ATGAGTTCCTTTCAAGAGGGAAAGGACTGCCTTGGGTGGGCAGCAAGAGATGCCTCTGGAGTTCTGTACCTTACCATTTTAAGCGAAGG**I**GC
AATCGGTGCTGATGACATTTCAAGTGAAGATTACATACTGTGGAATATGTTATGGTGACATAGTTTATACCAGGAACAACATGAAGACTCAAAA
TATCCTGTAGTTCAGG**I**ACATGAGATTGCTGGAATCGTGAAGAAGTAGGCTCCAATGTCCAGCGCTTCAAACTGGTGACCTGTTGGAGT
GGGAACATATATTAATTCATGCAGAAATTGTGATGAATGTAAGGGCTAGAAGTGCACTGCCCAAATGGAATGGTTCCCACAATTAATGCT
GTGGATGTAGATGTTACAATCACAAAGGGAGGATACTCTAGTTTCTATTGTTGTTTCATGAAAAGG**I**ACTGCTACAAGATACCTGAAAACCTACCC
TTTAGCTCTAGCAGCACCTTTGCTTTGTTTCAAGGAATTACTGTTTACTCTCCATGATCCACTACAAGATGAACCAACCTGGTAAATCTCTAGGG
GTGATCGGGCTAGGAGGCCCTTGGTCACATGGCAGTGAAGTTTGGCAAGGCTTTTGGACTGAATGTAAACAGTTTTCAGTACAAGTATATCCAAGA
AAGAGGAAGCCTTGAATGTCTTGGAGCAGACAAATTTATTGTCTCGACTGATGAGGAAGAAATGAAG**I**ACCTTGTCTAGAACTTTGGACTTC
ATAATTGACTCAGCATCAGGAGATCATCCATTCGATCCATACATGTCCCTCTGAAGACTAATGGCTGTTGCTCATGGTGTGCTATCCAAAAG
AAGTTAAACTCGATCCTCTGAGCCTTTTTACAGGTAT**I**GAGATCGATTACTGGAAGTTTCACTGGTGGGACGAAATTGACACAAGAAATGTTG
GAGTCTGTGCTGCCACAAAATATATCCGGAGATCGAAGTGATACCAATGAATACGCTTATGAAGCCTTTGAGAGGATGTTGAAGGGAGATG
TCAAGTATCGATTGTTGATCGACATCGAGAATCTTTGAAGTGA

>PoptrCAD15

ATGGCAGAAAAATCTTACAGGAAGAACATCCTAACAAAGGCTTTTGGATGGGCAGCCAGAGACCAATCCGGGGTCTCTCTCTCTTTCAAATTCT
CCAGGAGGT**I**CTACAGGAGAGAAGGATGTGCGATTCAAGGTGCTCTTTTGTGGAATATGTCACTCAGACCTTCATATGGCCAAGAATGAGTGG
GGTACTGCCACGTACCTCTAGTTCCTGGG**I**CATGAGATTGTTGGGGAAGTGACAGAGGTAGGAAGCAAAGTTGAGAAGTTCAAAGTTGGAGA
CAAAGTGGGGGTGGGGTGTGTTGGATCATGCCACTCTTGGCATAGTTGCAACAACAATCTCGAGAATTATTGCCAAAAATGATACTCACC
TACAGTACCAAAATACCACGATGGAACCAACACGTACGGAGGCTACTCAGACAGCATGGTTCAGGATGAGCACTTCGTAATTCGTATTCCAGACA
ACCTACCTCTAGATGCCGCTGCACCTCTCCTATGTGCTGGGATTACAGTTTACAGCCCTTGAGGTTTTTAAATCTTGACAAACCGGGTATGCA
TGTGGCGTGGTTGGGCTCGGCGGCTAGGTCATGTAGCTGTAAAGTTTGGCAAGGCCATGGGGGTCAAGGTTACAGTTATTAGCACCTCTCCC
AAGAAGAAACAAGAAGCCCTTGAGCATCTTGGTGTGACTCGTTTCTAGTTAGTCGTGACCAGGATGAGATGCAGG**I**CTGCACTGGGCACAAT
GGATGGTGTAAATTGACACGGTGTGCGCCATTATCCTATCTTGCTTTGATTAGTCTATTGAAGACTCAAGGAAAGCTGGTCTTGGTTGGTGGC
CCTGAAAAGCCACTTGAGCTACCAAGTGTTCCTCTGATCATGGG**I**AAGAAAAATAGTGGGAGGGAGTACCATAGGAGGAATGAAGGAAACACA
AGAGATGATTGATTTTGTGCTGCCAAGAACAACATCACGGCAGATATTGAGGTTATCTCGATGGATTATGTGAACACAGCCATGGAGCGGCTTTG
AAAACAGATGTCAGATACCGATTTGTTATCGACATCGGCAACACAATGAAGATTTGA

Additional File 6 - Primers for quantitative real-time PCR above the *CAD* gene family in teak

Gene	Sequence	Primer size	Amplicon size
<i>TgCAD1</i>	5'-TTCATCAGGTCAGGGGTGAG-3'	20 bp	260 pb
	5'-CCCTGGAACCAAAGGGTATT-3'	20 bp	
<i>TgCAD2</i>	CATCTCTTCGGATGAAAAGG	20 bp	174 bp
	GCAAGGCTTCCTGGATAAAACT	22 bp	
<i>TgCAD3</i>	CTCACTTACAACAGCGTTTTGC	22 bp	180 bp
	GTCGAGCCCAAAATATCTCAAC	22 bp	
<i>TgCAD4</i>	CATTAGCACATCTTCAAACA	20 bp	257 bp
	CTATCGCCTTCCCCCCTAGAATC	23 bp	

Additional File 7 - Melting curves and efficiencies of *CAD* teak primers for quantitative real-time PCR

Additional File 8 - Nucleotide sequences for *CAD* family in *Tectona grandis*

>TgCAD1

ATGGGCAGTCTTGAGGTAGAGAGGACCACCATCGGTTGGGCTGCAAGGGACCCGTCTGGCGTTCTCTCTCCTTAC
 ACTTATAGCCTCAGGGAAACAGGTCCTGATGACATTCTCTTGAGAGTGTGTACTGTGGAGTAGACCACACAGAC
 CTTTCATCAGGTCAGGGGTGAGCTTGGCAACACCAAATACCCTTTGGTTCCAGGGCATGAAGTAATTGGAGAAGTT
 GTTGAATTGGGCTCAAAAGTGAAGAATTTCAAAGTGGGTGACATTATAGGAGTTGGAGGAATTATTGGTTCTTGT
 GAAAAATGTACTCTCTGCAACTCCAATCTGGAGCAATACTGCAGCAACAGAATCTTTACCTACAATGACGTCTAC
 AAAGATGGAACCTCAACTCAAGGGGGATATTCTTCTGCTATGGTTATTCATCACAGATTTGCAGTTAAAAATACCA
 GAAAACTAGCACCAGAACAAGCAGCACCACACTACTATGTGCCGGGTGACAGCATACAGTCCCCTCAAAGAGTTT
 ATGGATTCCGGCAAGGTCTACAAAGGAGGAATATTAGGCTTGGGAGGAGTTGGTCACCTGGCTGTGATGATAGCA
 AAGGCAATGGGTCATCATGTGACAGTAATAAGCTCTTCTGATAAGAAAAAGAGGAGGCTATGGAGCATTTCGAC
 GCAGACGCCCTTCTTGGTGAGTCGTAGTGAGGATGAAATGAAGCAAGCGATAAACAGCCTCGACTATATACTCGAC
 ACCGTGCCCTGTTGTTTCATCTCTGCCATCATATATTTCACTTTTGAAAAGCTGAAGGAAAGCTGTTATTAGTAGCG
 GCAGTTCCCTCAGCCACTTCAGTTTCTGGCTGCCGATATGATAAGAGGTAAGAAAGCAATCATGGCAAGTTTTATC
 GGAAGCGTGAAGGATACAGAGGAATTACTCAACTTTTGGGAGGAGAAGGGGTTGACAACTATGATAGAGGTGGTG
 AAGATAGACGTGGTTAACGAAGCATTTGAAAGAATGAAAGAAACGATGTAAGATACAAGTTTGTATTGGATGTG
 GCTGGCAGCAATCTTCAGTGA

>TgCAD2

GACGTTTATTTCGGAGGCTCGAGTTTTAGCAGATGATGTTGGGAAAGATGACATTTCTATAAACATAGTGTACTGT
 GGGGTTTGTCTTGGCTGATGTTGCTTGGACGAGGAACAAATTGGGCAATTCAAAGTATCCTTTAGTGCCCGGACAT
 GAGATTGTTGGGATTGTGACAGAAGTCGGATCCGATGTTGATCGCTTTATAGTTGGTGACTATGTGCGAGTCGGA
 ACTTATGTTAATACTTGCAGAGAGTGCGAGTATTGTGATAGTGAATTAGAAGTTCTTTGCTCAAAGGGGCCAGTC
 TTGACGTTTGTATGGTGTAGATGTTGATGGTACCATCACTAAGGGAGGATATTCTAGCTACATCGTAGTTTCATCA
 AGGTACTGTTTCAAGATACCTGAGAATACCCTCCACAGTTAGCAGCACCTCTGCTCTGTGCTGGAATTACCGTT
 TACACCCCATGATACGGCATAACATGAATCAACCTGGAATACTTTAGGGGTAATTGGGCTAGGTGGGCTTGGC
 CACTTGGCTGTGAAGTTCGGCAAGGCTTTTGGATTGAAAGTAACAGTTTTTAGCACCAGCATATCCAAGAGGGA
 AAGGCATTAATCTTCTTGGGGCAGACAATTTTGTCTCTCTTCGGATGAAAAGGAGATGAAGGCTTTGGATAAG
 TCACCTTGACTTCATCATAAACACAGCATCAGGAGATATACCATTGATTTATATTTGTCACTGTTGAAGAGCACT
 GGTGTGCTTGTCTTGGTTGGATTTCCTCAAGTGAAGTGAAGTTTTATCCAGGAAGCCTTGCTATTGGTGCAAAAACA
 ATTACAGGAAGTGCAACAGGAGGCACAAAACATAATCTTCTAGAAGCTCCACAATTACGCCCTAG

>TgCAD3

AAATTCAAGGTCGGAGACAAAGTCGGCGTCGGTTGTTTGGTCAATTCGTGCCGGAATGCGAACAATGCTCTAAC
 GATCTCGAGAATTACTGCCCTCAAATTGTGCTCACTTACAACAGCGTTTTGCCCCGACGGGTCCGTCACTTACGGC
 GGCTACTCTGATATTATGGTGTGCGGACGAGGATTTTATTATCCGGTGGCCCGAAAATTTCCCCCTCGACAAAGGA
 GCTCCGTTGCTCTGTGCTGGGATCACCACGTACAGCCCGTTGAGATATTTTGGGCTCGACAAGCCCGGCCTCCAC
 GTGGGAGTTGCCGGGCTCGGCGGGCTGGGCCATGTTGCAGTTAAATTCGCCAAGGCTTTTCGGAACAAAAGTTACG
 GTGATCAGCACCTCCGCCGTTAAAAAGAAGGAAGCCATCGAAGCCCTCGGAGCAGACGCGTTTCTCATAAGCCGT
 GATCCGGCGGAAATTACGGCGGCGGCTGGGACATTGGACGGGATCATAGACACTGTATCGGCGCAACACCCGCTT
 CCGCCATTATTAAGCTTGTGAAGCCGCATGGGAAGCTGGTCTGTTGGAGCGCCGGAAGCCGCTTGAGCTG
 CCAGTTTTCCCGCTGATCTCCTCCCGGAAGACAGTGGGAGGTAGTTGCATCTTCTAG

>TgCAD4

GGTGGCTCGAGTTTTTCAAGCAAGATGTTGGGGAAGTGACAGAGGTAGGTAGCAAGGTGGAGAAATACAAGATTGGG
 GACAAAGTAGGTGTTGGATGCTTGGTTGGATCGTGTGCCAGTGTGAGGAGTGTACCAACAATGAGGAAAGTTAC
 TGCCCAAGCAAGTACTCTCAATCAACGCGCGTTACTATGATGGTACCATCACATATGGAGGTTTCTCTAACCTC
 ATGGTTACTGATGAACATTTTCATCATTCGTTGGCCTGAGAATTTGCCCTTGATAGCGGTGCCCTCTGCTGTGT
 GCCGGGATTACAACCTTACAGCCCATTTAGGCGCTTTGGGCTGGACAAACCTGGAGTGAATGTTGGCGTTGTAGGT
 CTTGGTGGGATTGGCCATCTCTCCGTGAGGTTTGCTAAGGCCTTAGGGAGTAAGGTGACAGTCATTAGCACATCT
 TCAAAACAAAAAGAAGGAAGCAATTGAAACTTTTCTGCTGACGACTTTTGGTTAGCCACGACCAAGAGCAGATG
 CAGGCTGCTGCAGGCACCTTGGATGGTATCATCGATACTGTCTCTGCAAATCATTCCCTATTACCATTAGTAAAT
 TTGTTAAAGCCTCATGGCAAGCTTATCTTGGTTGGTCTTCCACAAAACTTGAGGTGCCTACCTTTCCCTGATT
 CTAGGGGGGAAGGCATAGTCGGAAGTGAAGCGGAGGGGTGAAAGAGACGCAAGAGATGATCGAAATTTCGCAGC
 TAA

Additional File 9 - Amino acid sequences for CAD family in *Tectona grandis*

>TgCAD1

MGSLEVERTTIGWAARDPSGVLSPYTYSLRETGPDDILLRVLYCGVDHTDLHQVRGELGNTKYPLVPGHEVIGEV
VELGSKVKNFKVGDIIGVGGIIGSCEKCTLCNSNLEQYCSNRIFTYNDVYKDGTPQTGGYSSAMVIHHRFAVKIP
EKLAPQAAAPLLCAGVTAYSPLKEFMDSGKVYKGGILGLGGVGHVLAVMIAKAMGHHVTVISSDCKKKEEAMEHLH
ADAFVLSRSEDEMKQAINSLDYILDTPVHVHPLPSYISLLKTEGKLLLVAAPQPLQFLAADMIRGKKAIMASFI
GSVKDTEELLNFWEEKGLTTMIEVVKIDVVNEAFERMERNDVRYKFVLDVAGSNLQ-

>TgCAD2

DVYSEARVLADDVVGKDDISINIVYCGVCFADVAWTRNKLGNKYPLVPGHEIVGIVTEVGSVDVDRFIVGDYVGVG
TYVNTCRECEYCDSELEVLCSEKGPVLTDFGVDVDGTITKGGYSSYIVVHQRYCFKIPENYPPQLAAPLLCAGITV
YTPMIRHNMNQPGKSLGVIGLGGGLGHLAVKFGKAFGLKVTVFSTISISKREEALNLLGADNFVISSDEKEMKALDK
SLDFIINTASGDIPFDLYLSLLKSTGVLALVGFPPSEVKFYPGSLAIGAKTITGSATGGTKHNLLEAPQLRP-

>TgCAD3

KFKVGDVGVGCLVNSCRKCEQCSNDLENYCPQIVLTYNVLPDGSVITYGGYSDIMVSEDFIIRWPENFPLDKG
APLLCAGITTYSPRYFGLDKPGLHVGAVAGLGGGLGHVAVKFAKAFGTKVTVISTASAGKKKEAIEALGADAFLISR
DPAEIQAAAGTLDGIIDTVSAQHPLPPLLSLLKPHGKLVVVGAPKPLELPVFPLISSRKTVGGSCEF-

>TgCAD4

GGSSFQQDVGEVTEVGSKEKYKIGDKVGVGCLVGSQRQCEECTNNEESYCPKQVLSINARYYDGTITYGGFSNL
MVTDEHFIIRWPENLPLDSGAPLLCAGITTYSPRRFGLDKPGVNVGVVGLGGIGHLSVRFAKALGSKVTVIST
SNKKKEAIEETFRADDFLVSHDQEQMAAAGTLDGIIDTVSANHSLPLVNLLKPHGKLILVGLPQKLEVPFPLI
LGGKAIVGTASGGVKETQEMIEIRS-

Additional File 10 - TgCADs subcellular localization prediction

CAD	CELLOV 2.5 program	
	Score	Localization Class
TgCAD1	4,727	cytoplasmic
TgCAD2	2,608	cytoplasmic
TgCAD3	3,110	cytoplasmic
TgCAD4	3,079	cytoplasmic

References

- ALCÂNTARA, B.K.; VEASEY, E.A. Genetic diversity of teak (*Tectona grandis* L. f.) from different provenances using microsatellite markers. **Revista Árvore**, Viçosa, v. 37, n. 4, p. 747–758, 2013.
- BAILLÈRES, H.; DURAND, P.Y. Non-destructive techniques for wood quality assessment of plantation grown teak. **Bois et Forêts des Tropiques**, Nogent-sur-Marne, v. 263, n. 1, p. 17–29, 2000.
- BARAKAT, A.; BAGNIEWSKA-ZADWORNA, A.; CHOI, A.; PLAKKAT, U.; DILORETO, D.S.; YELLANKI, P.; CARLSON, J.E. The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. **BMC Plant Biology**, London, v. 9, p. 26, 2009.
- BARAKAT, A.; BAGNIEWSKA-ZADWORNA, A.; FROST, C.J.; CARLSON, J.E. Phylogeny and expression profiling of CAD and CAD-like genes in hybrid *Populus* (*P. deltoides* × *P. nigra*): evidence from herbivore damage for subfunctionalization and functional divergence. **BMC Plant Biology**, London, v. 10, p. 100, 2010.
- BAUCHER, M.; HALPIN, C.; PETIT-CONIL, M.; BOERJAN, W. Lignin: genetic engineering and impact on pulping. **Critical Reviews in Biochemistry and Molecular Biology**, Boca Raton, v. 38, n. 4, p. 305–350, 2003.
- BHAT, K.M.; PRIYA, P.B.; RUGMINI, P. Characterisation of juvenile wood in teak. **Wood Science and Technology**, New York, v. 34, n. 6, p. 517–532, 2001.
- BONAWITZ, N.D.; CHAPPLE, C. The genetics of lignin biosynthesis: connecting genotype to phenotype. **Annual Review of Genetics**, Palo Alto, v. 44, p. 337–363, 2010.
- BOUVIER D'YVOIRE, M.; BOUCHABKE-COUSSA, O.; VOOREND, W.; ANTELME, S.; CÉZARD, L.; LEGÉE, F.; LEBRIS, P.; LEGAY, S.; WHITEHEAD, C.; MCQUEEN-MASON, S.J.; GOMEZ, L.D.; JOUANIN, L.; LAPIERRE, C.; SIBOUT, R. Disrupting the cinnamyl alcohol dehydrogenase 1 gene (BdCAD1) leads to altered lignification and improved saccharification in *Brachypodium distachyon*. **The Plant Journal: for Cell and Molecular Biology**, Oxford, v. 73, n. 3, p. 496–508, 2013.

BUKH, C.; NORD-LARSEN, P.H.; RASMUSSEN, S.K. Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. **Journal of experimental botany**, Oxford, v. 63, n. 17, p. 6223–6236, 2012.

CHENG, H.; LI, L.; XU, F.; CHENG, S.; CAO, F.; WANG, Y.; YUAN, H.; JIANG, D.; WU, C. Expression patterns of a cinnamyl alcohol dehydrogenase gene involved in lignin biosynthesis and environmental stress in *Ginkgo biloba*. **Molecular Biology Reports**, Dordrecht, v. 40, n. 1, p. 707–721, 2013.

DEEPAK, M.S.; SINHA, S.K.; RAO, R.V. Tree-ring analysis of teak (*Tectona grandis* L. f.) from Western Ghats of India as a tool to determine drought years. **Emirates Journal of Food and Agriculture**, Al Ain, v. 22, n. 5, p. 388–397, 2010.

DEFIORIO, G.; HORGAN, G.; WOODWARD, S.; FOSSDAL, C.G. Gene expression profiles, phenolics and lignin of Sitka spruce bark and sapwood before and after wounding and inoculation with *Heterobasidion annosum*. **Physiological and Molecular Plant Pathology**, London, v. 75, n. 4, p. 180–187, 2011.

DENG, W.-W.; ZHANG, M.; WU, J.-Q.; JIANG, Z.-Z.; TANG, L.; LI, Y.-Y.; WEI, C.-L.; JIANG, C.-J.; WAN, X.-C. Molecular cloning, functional analysis of three cinnamyl alcohol dehydrogenase (CAD) genes in the leaves of tea plant, *Camellia sinensis*. **Journal of plant physiology**, Stuttgart, v. 170, n. 3, p. 272–282, 2013.

EUDES, A.; LIANG, Y.; MITRA, P.; LOQUE, D. Lignin bioengineering. **Current Opinion in Biotechnology**, London, v. 26, p. 189–198, 2014.

GALEANO, E.; VASCONCELOS, T.S.; RAMIRO, D.A.; MARTIN, V.D.F. de; CARRER, H. Identification and validation of quantitative real-time reverse transcription PCR reference genes for gene expression analysis in teak (*Tectona grandis* L.f.). **BMC Research Notes**, London, v. 7, p. 464, 2014.

GUO, D.-M.; RAN, J.-H.; WANG, X.-Q. Evolution of the Cinnamyl/Sinapyl Alcohol Dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of Bona Fide CAD. **Journal of Molecular Evolution**, New York, v. 71, n. 3, p. 202–218, 2010.

HALLETT, J.T.; DIAZ-CALVO, J.; VILLA-CASTILLO, J.; WAGNER, M.R. Teak plantations : economic bonanza or environmental disaster? **Journal of Forestry**, Washington, v. 109, n. 5, p. 288–292, 2011.

HEALEY, S.P.; GARA, R.I. The effect of a teak (*Tectona grandis*) plantation on the establishment of native species in an abandoned pasture in Costa Rica. **Forest Ecology and Management**, Amsterdam, v. 176, n. 1/3, p. 497–507, 2003.

JIN, Y.; ZHANG, C.; LIU, W.; QI, H.; CHEN, H.; CAO, S. The cinnamyl alcohol dehydrogenase gene family in melon (*Cucumis melo* L.): bioinformatic analysis and expression patterns. **PloSone**, San Francisco, v. 9, n. 7, p. e101730, 2014.

- KIM, S.-J.; KIM, M.-R.; BEDGAR, D.L.; MOINUDDIN, S.G.A; CARDENAS, C.L.; DAVIN, L.B.; KANG, C.; LEWIS, N.G. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, n. 6, p. 1455–1460, 2004.
- KOLLERT, W.; CHERUBINI, L. **Teak resources and market assessment 2010 (*Tectona grandis* Linn. F.)**. Rome: FAO, 2012. 42 p.
- LARROY, C.; FERNANDEZ, M.R.; GONZALEZ, E.; PARES, X.; BIOSCA, J.A. Characterization of the *Saccharomyces cerevisiae* YMR318C (ADH6) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: relevance in aldehyde reduction. **Biochemical Journal**, London, v. 361, p. 163–172, 2002.
- LAURICHESSE, S.; AVÉROUS, L. Progress in polymer science chemical modification of lignins: towards biobased polymers. **Progress in Polymer Science**, Elmsford, v. 39, p. 1266–1290, 2014.
- LI, L.; LU, S.; CHIANG, V. A genomic and molecular view of wood formation. **Critical Reviews in Plant Sciences**, London, v. 25, n. 3, p. 215–233, 2006.
- LI, X.; YANG, Y.; YAO, J.; CHEN, G.; LI, X.; ZHANG, Q.; WU, C. FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. **Plant Molecular Biology**, Dordrecht, v. 69, n. 6, p. 685–697, 2009.
- LUKMANDARU, G.; TAKAHASHI, K. Variation in the natural termite resistance of teak (*Tectona grandis* Linn. fil.) wood as a function of tree age. **Annals of Forest Science**, Les Ulis, v. 65, p. 708, 2008.
- LYNCH, D.; LIDGETT, A.; MCINNES, R.; HUXLEY, H.; JONES, E.; MAHONEY, N.; SPANGENBERG, G. Isolation and characterisation of three cinnamyl alcohol dehydrogenase homologue cDNAs from perennial ryegrass (*Lolium perenne* L.). **Journal of Plant Physiology**, Stuttgart, v. 159, p. 653–660, 2002.
- MA, Q.-H.; WANG, C.; ZHU, H.-H. TaMYB4 cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes. **Biochimie**, Paris, v. 93, n. 7, p. 1179–1186, 2011.
- MANSELL, R.L.; GROSS, G.G.; STÖCKIGT, J.; FRANKE, H.; ZENK, M.H. Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. **Phytochemistry**, New York, v. 13, n. 11, p. 2427–2435, 1974.
- PILATE, G.; GUINEY, E.; HOLT, K.; PETIT-CONIL, M.; LAPIERRE, C.; LEPLÉ, J.; POLLET, B.; MILA, I.; WEBSTER, E.A.; MARSTORP, H.G.; HOPKINS, D.W.; JOUANIN, L.; BOERJAN, W.; SCHUCH, W.; CORNU, D.; HALPIN, C. Field and pulping performances of transgenic trees with altered lignification. **Nature Biotechnology**, New York, v. 20, p. 607–612, June 2002.

PREISNER, M.; KULMA, A.; ZEBROWSKI, J.; DYMINSKA, L.; HANUZA, J.; ARENDT, M.; STARZYCKI, M.; SZOPA, J. Manipulating cinnamyl alcohol dehydrogenase (CAD) expression in flax affects fibre composition and properties. **BMC Plant Biology**, London, v. 14, p. 50, 2014.

RAHANTAMALALA, A.; RECH, P.; MARTINEZ, Y.; CHAUBET-GIGOT, N.; GRIMA, J.; PACQUIT, V. Coordinated transcriptional regulation of two key genes in the lignin branch pathway - CAD and CCR - is mediated through MYB- binding sites. **BMC Plant Biology**, London, v. 10, p. 130, 2010.

ROSSMAN, M.G.; MORAS, D.; OLSEN, K.W. Chemical and biological evolution of a nucleotide-binding protein. **Nature**, London, v. 250, p. 194–199, July 1974.

SAIDI, M.N.; BOUAZIZ, D.; HAMMAMI, I.; NAMSI, A.; DRIRA, N.; GARGOURI-BOUZID, R. Alterations in lignin content and phenylpropanoids pathway in date palm (*Phoenix dactylifera* L.) tissues affected by brittle leaf disease. **Plant science: an International Journal of Experimental Plant Biology**, Limerick, v. 211, p. 8–16, 2013.

SALZMAN, R.A.; FUJITA, T.; HASEGAWA, P.M. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. **Plant Molecular Biology Reporter**, Athens, v. 17, n. 765, p. 11–17, 1999.

SATTLER, S.E.; SAATHOFF, A.J.; HAAS, E.J.; PALMER, N.A.; FUNNELL-HARRIS, D.L.; SARATH, G.; PEDERSEN, J.F. A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum brown midrib6 phenotype. **Plant Physiology**, Bethesda, v. 150, n. 2, p. 584–595, 2009.

SHI, R.; SUN, Y.-H.; LI, Q.; HEBER, S.; SEDEROFF, R.; CHIANG, V. L. Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. **Plant & Cell Physiology**, Kyoto, v. 51, n. 1, p. 144–163, 2010.

TANG, R.; ZHANG, X.-Q.; LI, Y.-H.; XIE, X.-M. Cloning and in silico analysis of a cinnamyl alcohol dehydrogenase gene in *Pennisetum purpureum*. **Journal of Genetics**, Bangalore, v. 93, n. 1, p. 145–158, 2014.

TANG, X.; XIAO, Y.; LV, T.; WANG, F.; ZHU, Q.; ZHENG, T.; YANG, J. High-throughput sequencing and assembly of the *Isatis indigotica* transcriptome. **PloSone**, San Francisco, v. 9, n. 9, p. e102963, 2014.

TOBIAS, C.M.; CHOW, E.K. Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. **Planta**, Berlin, v. 220, n. 5, p. 678–688, 2005.

TRABUCCO, G.M.; MATOS, D.A.; LEE, S.J.; SAATHOFF, A.J.; PRIEST, H.D.; MOCKLER, T.C.; SARATH, G.; HAZEN, S.P. Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. **BMC biotechnology**, London, v. 13, n. 1, p. 61, 2013.

TRONCHET, M.; BALAGUÉ, C.; KROJ, T.; JOUANIN, L.; ROBY, D. Cinnamyl alcohol dehydrogenases-C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in *Arabidopsis*. **Molecular Plant Pathology**, London, v. 11, n. 1, p. 83–92, 2010.

VALÉRIO, L.; CARTER, D.; RODRIGUES, J. C.; TOURNIER, V.; GOMINHO, J.; MARQUE, C.; BOUDET, A.; MAUNDERS, M.; PEREIRA, H.; TEULIERES, C. Down regulation of cinnamyl alcohol dehydrogenase, a lignification enzyme, in *Eucalyptus camaldulensis*. **Molecular Breeding**, Dordrecht, v. 12, p. 157–167, 2003.

VANHOLME, R.; DEMEDTS, B.; MORREEL, K.; RALPH, J.; BOERJAN, W. Lignin biosynthesis and structure. **Plant Physiology**, Bethesda, v. 153, n. 3, p. 895–905, 2010.

XU, Y.; THAMMANNAGOWDA, S.; THOMAS, T.P.; AZADI, P.; SCHLARBAUM, S.E.; LIANG, H. LtuCAD1 is a cinnamyl alcohol dehydrogenase ortholog involved in lignin biosynthesis in *Liriodendron tulipifera* L., a basal angiosperm timber species. **Plant Molecular Biology Reporter**, Athens, v. 31, n. 5, p. 1089–1099, 2013.

YOUN, B.; CAMACHO, R.; MOINUDDIN, S.G.A.; LEE, C.; DAVIN, L.B.; LEWIS, N.G.; KANG, C. Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. **Organic & Biomolecular Chemistry**, Cambridge, v. 4, n. 9, p. 1687–1697, 2006.

ZENG, Y.; ZHAO, S.; YANG, S.; DING, S. Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. **Current Opinion in Biotechnology**, London, v. 27, p. 38–45, 2014.

ZHANG, L.; WANG, G.; CHANG, J.; LIU, J.; CAI, J.; RAO, X.; ZHANG, L.; ZHONG, J.; XIE, J.; ZHU, S. Effects of 1-MCP and ethylene on expression of three CAD genes and lignification in stems of harvested Tsai Tai (*Brassica chinensis*). **Food Chemistry**, London, v. 123, n. 1, p. 32–40, 2010.

ZHAO, Q.; TOBIMATSU, Y.; ZHOU, R.; PATTATHIL, S.; GALLEG0-GIRALDO, L.; FU, C.; JACKSON, L.A.; HAHN, M.G.; KIM, H.; CHEN, F.; RALPH, J.; DIXON, R.A. Loss of function of cinnamyl alcohol dehydrogenase 1 leads to unconventional lignin and a temperature-sensitive growth defect in *Medicago truncatula*. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 110, n. 33, p. 13660–13665, 2013.

4 RNA-SEQ REVEALS TRANSCRIPT PROFILING AND *MYB* TRANSCRIPTION FACTORS OF LIGNIFIED TISSUES IN *Tectona grandis*

Abstract

Background: Currently, *Tectona grandis* is one of the most valuable trees in the world and no gene dataset related to secondary xylem is available. Considering how important the secondary xylem and sapwood transition from young to mature trees is, little is known about the expression differences between those successional processes and which transcription factors could regulate lignin biosynthesis in this tropical tree. Although *MYB* transcription factors are one of the largest superfamilies in plants related to secondary metabolism, it has not yet been characterized in teak. These results will open new perspectives for studies of diversity, ecology, breeding and genomic programs aiming to understand deeply the biology of this species. **Results:** We present a widely expressed gene catalog for *T. grandis* using Illumina technology and the de novo assembly. A total of 462,260 transcripts were obtained, with 1,502 and 931 genes differentially expressed for stem and branch secondary xylem, respectively, during age transition. Analysis of stem (vertical growth) and branch (horizontal growth) secondary xylem indicates substantial similarity in gene ontologies including carbohydrate enzymes, response to stress, protein binding, but interestingly allowed us to find transcription factors and heat-shock proteins differentially expressed. TgMYB1 displays a MYB domain and a predicted coiled-coil (CC) domain, while TgMYB2, TgMYB3 and TgMYB4 showed R2R3-MYB domain and phylogenetically grouped with several gymnosperms and flowering plants. *TgMYB1* and *TgMYB4* presented more expression in mature secondary xylem and sapwood, in contrast with *TgMYB2* whose expression is higher in young lignified tissues. *TgMYB3* is expressed at lower level in secondary xylem. **Conclusions:** Expression patterns of *MYB* transcription factors in lignified tissues are dissimilar when tree development was evaluated, obtaining more expression of *TgMYB1* and *TgMYB4* in lignified tissues of 60-year-old trees, opening a door for further functional characterization by reverse genetics and a possible evaluation and improvement of tree growth rate using molecular markers with those genes, once sapwood is a perfect indicator of wood quality. The obtained transcriptome represents a new sequence data of *T. grandis* deposited in public databases, representing an unprecedented opportunity to discover several related-genes associated with secondary xylem such as transcription factors and stress-related genes in a tropical tree.

Keywords: Gene expression; *MYB* transcription factors; Heat-shock proteins; Secondary xylem; Sapwood

4.1 Introduction

Teak (*Tectona grandis* Linn. f.) (Lamiaceae) is the most important and highly valued commercial hardwood timber in the tropics due to its high durability, dimensional stability, heartwood-sapwood proportions, weightlessness and resistance to weathering. Also, it is used for carpentry, floors, shipbuilding and agroforestry, thus becoming a high-class furniture and

a standard timber in end-use classification of other tropical timbers (BHAT; PRIYA; RUGMINI, 2001; JAIN; ANSARI, 2013; SHUKLA; VISWANATH, 2014). It is a deciduous species presenting natural populations in Thailand, Laos, Myanmar, India and Java Islands.

Teak grows properly within 25-38°C, between 1,250 and 2,500 mm/year of rainfall, presenting the best yields under 600 meters above sea level and produces better wood quality with long dry periods, from 3 to 5 month long (BHAT et al., 2005; GOH; MONTEUUIS, 2005; KEOGH, 2009; KOLLERT; CHERUBINI, 2012). This species is the major component of the forest economies of many tropical countries. It is the only valuable hardwood that constitutes a globally emerging forest resource with a planted area of 4,346 million ha (0,5 million m³ of wood) and natural forest of 29,035 million ha (2 million m³ of wood) around the world, and Brazil presents the largest teak reforestation in South America (KOLLERT; CHERUBINI, 2012).

Due to its importance, many efforts have focused on the study of teak populations variability (SHRESTHA; VOLKAERT; STRAETEN, 2005; VERHAEGEN et al., 2005; FOFANA et al., 2009; SREEKANTH et al., 2012; LYNGDOH et al., 2013; MINN; PRINZ; FINKELDEY, 2014). However, there are no genetic studies nor next-generation sequencing regarding wood formation in teak. Wood comes from secondary growth, starting with the vascular cambium expansion and cell division in stems of young trees, followed by a differentiation of secondary xylem and several events such as xylem cells expansion, secondary cell wall deposition and programmed cell death (CHAFFEY, 2002; DHARMAWARDHANA; BRUNNER; STRAUSS, 2010; LIU; FILKOV; GROOVER, 2014).

In most tropical America, including Brazil, harvesting occurs at 20 years, producing small-dimension logs, which are not in demand on the international market (BHAT et al., 2005; KOLLERT; CHERUBINI, 2012). Teak is not a fast growing species but can produce a timber of optimum strength in relatively short rotations of 21 years (BHAT; INDIRA, 1997) depending of the sapwood-heartwood percentages. The timber quality produced will be the overriding commercial factor for the near future (GOH et al., 2007), and usually relates to the amount, color and durability of the heartwood (BHAT et al., 2005).

For that reason, techniques such as ESTs and microarrays have been used extensively to understand wood formation in trees such as *Pinus* (YANG et al., 2004) and *Populus* (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010). However, today, large-scale studies of biological phenomena are unthinkable without the use of next-generation sequencing technologies (NGS), such as RNA sequencing (RNA-seq), which encourages

developmental and genomics research of woody growth in trees (LIU; FILKOV; GROOVER, 2014), especially for species without a sequenced genome and no molecular information available (GORDO et al., 2012; SCHLIESKY et al., 2012) as teak. In tropical trees, the use of next-generation sequencing in order to find differentially expressed unigenes involved in secondary xylem is restricted to some species (UENO et al., 2013).

Availability of nondestructive wood analysis methods such as core sampling would provide a valuable way to study teak wood in different aspects and avoid depletion of both natural and plantation teak resources (GOH; MONTEUUIS, 2005). Heartwood and sapwood percentage are not easily assessed on standing trees, but can be determined from a bore core (BHAT et al., 2005). In teak, it is certainly needed to identify genes such as those controlling secondary xylem, vessel formation, sapwood and heartwood differentiation, volume growth and abiotic stress.

Those studies have been documented in *Populus tremula* (SCHRADER et al., 2004), *Populus euphratica* (QIU et al., 2011), *Populus trichocarpa* (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010; BAO et al., 2013), eucalyptus (MIZRACHI et al., 2010), conifers (BEDON; GRIMA-PETTENATI; MACKAY, 2007; PAVY et al., 2008), and *Fraxinus spp.* (BAI et al., 2011), but it needs to be done in teak to help improving wood quality, growth speed and environmental adaptability (BHAT et al., 2005). Several genes have shown an expression driven by the wood formation processes, and some families of transcription factors are key to activate and orchestrate such steps (ZHONG et al., 2008).

The *MYB* transcription factors have been related to the coordination of genes which drive the lignin biosynthesis, with a great range of regulation and operating at all points of the phenylpropanoid pathway (ZHAO; DIXON, 2011). The R2R3-MYB proteins (characterized by two imperfect conserved repeats of ~50 amino acids) belong to a large family of transcription factors with over 120 members in angiosperms, also defined by an N-terminal DNA-binding domain (DBD), a C-terminal modulator region with regulatory activity; also R2R3-MYB proteins show a potential of binding AC elements (representative of lignin biosynthetic genes), which belong to the most abundant type in plants with essential roles in vascular organization (ROGERS; CAMPBELL, 2004; BEDON; GRIMA-PETTENATI; MACKAY, 2007).

Therefore, genetic examination of the superior growth (MIZRACHI et al., 2010) of a prized woody plant such as *T. grandis* would provide a collection of expressed genes from several tissues. A better understanding of secondary xylem formation is essential not only as a

fundamental part of plant biology (anatomy, biochemistry and at the genetic level), but also because it is crucial to obtain solutions for problems in forest conservation, improving the offerings of woody products (LIU; FILKOV; GROOVER, 2014). Also, it is hoped that through genetic selection and plant transformation, the non-durable core could be reduced or eliminated, the growth could be increased and the epicormic branches could be controlled, making the so-called “juvenile wood” problem a thing of the past (KEOGH, 2009).

Sapwood/hardwood characteristics are reliable predictors of overall genetic improvement of timber strength (BHAT; INDIRA, 1997). Therefore, this is the first RNA sequencing in this tropical woody plant, covering the transcriptome of *T. grandis* during young to mature transition and in several tissues. We detected more than 462.260 transcripts and found that more than 2000 unigenes were differentially expressed between ages and tissues. We also supplied several heat-shock proteins and analyzed the expression of some *MYB*-related transcription factors differentially expressed in teak secondary xylem, including sapwood tissue.

4.2 Materials and Methods

4.2.1 Plant material

Removal and discarding of the *T. grandis* bark of the trunk and the outer suberized layer (secondary phloem and vascular cambium) of approximately 1.5 cm thickness was performed, with a subsequent collection of a yellow blade of 5 mm located after removal, taking a heterogeneous tissue which includes priority secondary xylem. Usually, cells of the cambial zone have thin cell walls and can be easily removed from the stem (LIU; FILKOV; GROOVER, 2014). Branch (from the base and recent ones) and secondary xylem on the main stem at DBH (Diameter at breast height) were sampled from twelve-years-old and sixty-years-old *T. grandis* trees from experimental field, Agriculture College “Luiz de Queiroz”, University of São Paulo, in Piracicaba, São Paulo State, Brazil. Additionally, seedlings after two weeks of seed germination, leaves and roots from two month-old *in vitro* teaks were sampled. Flowers at different stages were collected from the twelve year-old teak trees. All tissues/organs were harvested in ten randomized trees, (joining five samples as one replicate), immediately frozen by immersion in liquid nitrogen and stored at -80°C until RNA extraction. For quantitative Real Time PCR, sapwood from 12- and 60-year-old trees were also collected at the same location, with two replicates, each one coming from five trees, using an increment

borer at DBH (DEEPAK; SINHA; RAO, 2010) (Additional File 17), followed by immediate nitrogen immersion and RNA extraction.

4.2.2 Total RNA extraction and Illumina sequencing

Frozen tissue samples of 1.0 g were weighed and ground into fine powder in liquid nitrogen using a sterilized mortar and pestle. Total RNA was extracted following the protocol standardized by Salzman et al. (1999). 2 µg of total RNA from each sample were treated with DNase I (Promega), and the treated samples were analyzed in agarose gels to ensure absence of DNA and no degradation. In addition, PCR control reactions to examine for genomic DNA contamination were performed using total RNA without reverse transcription as template, and negative results (absence of bands) were assessed by electrophoresis on a 1% (w/v) agarose gel with ethidium bromide staining. The Agilent RNA 6000 Nano kit (Agilent, Santa Clara, CA) was used to verify the total RNA quality by the RIN factor in a 2100 Bioanalyzer (Agilent, Santa Clara, CA). Then, the TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA) was used to prepare two libraries from 1 µg of total RNA. For clustering the libraries, the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA) was used. To verify the size of the libraries, the Agilent DNA 1000 kit (Illumina, San Diego, CA) was used. For sequencing, the TruSeq SBS Kit v3-HS (Illumina, San Diego, CA) was used, with 200 cycles, using the Illumina HiSeq 1000 (Illumina, San Diego, CA) located at “Escola Superior de Agricultura Luiz de Queiroz”, Universidade de São Paulo (Brazil).

4.2.3 Mapping data against closely-related genomes

The closest available genomes to teak were used to validate and compare reads, such as *Solanum lycopersicum*, *Populus trichocarpa* and *Eucalyptus grandis* (BESSEY, 1915; CHASE et al., 2009; BELL; SOLTIS; SOLTIS, 2010). We used TopHat2 (KIM et al., 2013) and Bowtie2 (LANGMEAD; SALZBERG, 2012) to map teak reads against those genomes, available at NCBI (<http://www.ncbi.nlm.nih.gov/>) followed by the use of Cufflinks (TRAPNELL et al., 2012) to obtain transcript fragments in .fasta format. Blast2Go (CONESA et al., 2005) was used to annotate and compare transcripts and also to obtain functional categories. Common annotations between the three genomes and teak were subsequently used to perform clustalw with Clustalw2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) to check similarity, to design primers in the conserved domains (<http://www.premierbiosoft.com/>) and sequenced to corroborate

consistency of those mapped genes. All mapping was performed in the “Ohio Super Computer Center” (OSC), Ohio State University (USA), using clustering processes to improve the performance of TopHat2, Bowtie2 and Cufflinks. Sequencing was performed with the 3100 Genetic Analyzer (Applied Biosystems, USA) in the “CEBTEC” Center, Agriculture College “Luiz de Queiroz”, University of São Paulo, Piracicaba, São Paulo state, Brazil.

4.2.4 Cleaning and *de novo* assembly

Raw reads of the twelve samples were “grepped” and “trimmed” to increase the quality and further be used in the *de novo* assembly (BLANKENBERG et al., 2010). The *de novo* assembly was performed for the twelve samples with the cleaned reads using the Trinity program, version 2013 (GRABHERR et al., 2011; HAAS et al., 2013) at the “Ohio Super Computer Center” (OSC), Ohio State University (USA). Then, the reference transcriptome was prepared and RSEM tool was used to estimate abundance of reads for subsequent differential expression.

4.2.5 Detection and annotations of differentially expressed unigenes between twelve- and sixty-year-old trees

We used DESeq, an R Bioconductor package (ANDERS; HUBER, 2010), to perform the differential expression of unigenes between lignified tissues and the different ages at the “Ohio Super Computer Center” (OSC), Ohio State University, USA. Abundance estimation and FPKM value was obtained using RSEM (HAAS et al., 2013). Next, two matrixes were generated, one containing the counts of RNA-seq fragments and used for differential expression by DESeq and the other one performing the TMM normalization in order to generate graphics. The lignified groups for comparison were: (1) Branch secondary xylem of 12-year-old trees against Branch secondary xylem of 60-year-old trees, (2) Stem secondary xylem of 12-year-old trees against stem secondary xylem of 60-year-old trees, (3) Branch vs. Stem secondary xylem, (4) Other tissues (flower, leaf, root, seedling) vs. Branch secondary xylem (5) Other tissues (flower, leaf, root, seedling) vs. Stem secondary xylem. The results were represented in “MA” and “volcano” plots from pairwise comparisons using both replicates for branch and stem secondary xylem and a cutoff of false discovery rate (FDR) ≤ 0.05 . Subsequently, differentially expressed unigenes were exported with the “cdbfasta” tool (<http://compbio.dfci.harvard.edu/tgi/software/>) with the contig name from assemblies of

Trinity database in .fasta format, annotated using Blast2Go (CONESA et al., 2005) and KEGG metabolic pathways were obtained.

4.2.6 Phylogeny of *MYB* transcription factors differentially expressed in teak

MYB transcription factors with complete coding sequence were selected manually from the annotated differentially expressed genes of stem secondary xylem. The phylogenetic trees were built with Clustal W amino acid alignments and following the neighbor joining tree method in Mega 6, using 10,000 bootstrap replication for the tree nodes, poisson model, amino acid substitution type, uniform rates and pairwise deletion. The first phylogenetic tree was built using sequences of all 126 *Arabidopsis* R2R3 *MYB* proteins downloaded from the TAIR *Arabidopsis* genome annotation (MATUS; AQUEA; ARCE-JOHNSON, 2008; ZHONG et al., 2008; ZHAO; DIXON, 2011). The second phylogenetic tree was constructed with several predicted *MYB* protein sequences from white spruce, loblolly pine and diverse *Arabidopsis* *MYB* sequences (BEDON; GRIMA-PETTENATI; MACKAY, 2007).

4.2.7 Gene expression of *MYBs* along the lignified teak tissues by qRT-PCR

Three cDNA samples were synthesized from each tissue (branch, stem secondary xylem and sapwood from twelve- and sixty-years-old *T. grandis* trees, leaves and roots from two month-old in vitro teaks), each replicate coming from five trees (see Plant Material), using 1.0 µg of the treated RNA using the SuperScriptTM III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer's instructions. cDNA concentration was determined with the Ultrospec 2100 PRO Spectrophotometer (Amersham Biosciences, USA). The primers for qRT-PCR were designed flanking *TgMYB1*, *TgMYB2*, *TgMYB3* and *TgMYB4* teak sequences (Additional File 18), followed by determining the standard curve with four cDNA dilutions and the melting curve (Additional File 19). The qRT-PCR mixture contained 125 ng of cDNA from each sample, primers to a final concentration of 50 µM each, 12.5 µl of the SYBR Green PCR Master Mix (Applied Biosystems, USA) and PCR-grade water up to a total volume of 25 µl. Each gene reaction was performed in technical replicate. PCR reactions without template were also done as negative controls for each primer pair. The quantitative real time PCRs were performed employing the StepOnePlusTM System (Applied Biosystems, USA). All PCR reactions were performed under the following conditions: 2 min at 50°C, 2 min at 95°C, and 45 cycles of 15 s at 95°C and 1 min at 65°C in 96-well optical reaction plates (Applied Biosystems, USA). Leaf sample was used as calibrator to normalize the values

between different plates and *EF1 α* as control gene, following previous studies in teak (GALEANO et al., 2014). All statistically significant differences between the means were performed in SAS program at 95% confidence level with the F-test, and the pair comparison procedure was performed with LSD at 95% confidence level.

4.3 Results

4.3.1 Quality of the RNA and the reads

Based on the bioanalyzer results (Additional File 1), all samples showed appropriate RIN factor. The libraries had a size of 280 bp, approximately. We generated almost 193 million paired-end reads, covering 233 Gb of sequence data with a sequence length of 100 bp (Table 1). The dataset of raw reads was deposited in NCBI database under SRA accession number PRJNA 269536. After cleaning the data with “grepped” and “trimmed” procedures (BLANKENBERG et al., 2010), the “base sequence quality” and the “per sequence GC content” were improved, losing 9.5% of the reads (Table 1), and between 3.8% (branch of 60-year-old teak trees) and 11.14% (seedling) (Additional File 2), obtaining more than 174 million sequence reads with a size of 75 Gb (Table 1). Indeed, with this quality it was possible to continue the subsequent analyses.

4.3.2 Read mapping against the tomato, *Populus* and *Eucalyptus* genomes

Solanum lycopersicum was used to map *T. grandis* reads because both are in Solanales and Lamiales orders, respectively, located in the same clade (Lamiidae) with a final divergence inside angiosperm diversification in the late cretaceous (BELL; SOLTIS; SOLTIS, 2010). *Populus trichocarpa* and *Eucalyptus grandis* (Malpighiales and Myrtales orders, respectively) were used to map teak reads because they are the closest trees with available genomes (BESSEY, 1915; CHASE et al., 2009; BELL; SOLTIS; SOLTIS, 2010). TopHat2 and Bowtie2 are two programs developed to map reads against genomes and cufflinks to obtain transcript fragments, and they were useful for teak reads. Focusing on teak lignified tissues, it was found in branch secondary xylem of 12-year-old trees appropriate reads mapped in all three genomes (Additional File 3a) and consequently we found genes such as *hydrolase* and *YCF68*. In branch secondary xylem of 60-year-old trees, several reads mapped (Additional File 3b) in the described genomes and *atp synthase alpha II* gene and *mitochondrial protein* were found. *Hydrolase*, *YCF68*, *atp synthase alpha II* and *mitochondrial protein* nucleotide sequences were translated into proteins, their function

analyzed, and primers over the clustal were designed to amplify by PCR followed by sequencing. We obtained 503 bp, 289 bp, 271 bp and 98 bp for *Hydrolase*, *YCF68*, *atp synthase alpha II* and *mitochondrial protein*, respectively. After sequencing, all blasts showed 100% similarity with the original sequences from *Populus*, Eucalyptus and tomato, exhibiting the effectiveness of those bioinformatics programs to map and finally the purity and good quality of the teak reads. The sequences obtained for teak are displayed in Additional File 4. Afterwards, we performed the *de novo* assembly of teak transcriptome.

Table 1 - Overview of sequencing, assembly, differential expressed genes and annotations

Raw data	
Total number of reads without cleaning	192,841,634
Size without cleaning (Gb)	233.1
Sequence length without cleaning (bp)	100
Total number of reads after cleaning	174,528,668
Size after cleaning (Gb)	75.4
Sequence length after cleaning (bp)	75-100
% Erased reads	9.5
Assembly	
Number of transcripts obtained with Trinity	462,260
N50 length (bp)	2140
Differential Expression	
Number of most expressed transcripts in Stem second. Xyl.	1502
Number of most expressed transcripts in Branch second. Xyl.	931
Annotations by Blast2Go	
Number of Predicted CDS (partial/complete) in Stem second. Xyl.	669
Number of Predicted CDS (partial/complete) in Branch second. Xyl.	603

4.3.3 *De novo* assembly

The assembly of the transcriptome from the leaf, root, seedling, flower, secondary xylem of teak branch and stem was performed using the Trinity assembler (HAAS et al., 2013). For lignified tissues such as branch secondary xylem of both tree ages (12- and 60-year-old trees), we used between 9,622,608 and 16,324,986 reads, and for stem secondary xylem of both tree ages (12- and 60-years-old) we used between 9,417,573 and 10,963,888 reads (Additional File 2). Flower, leaf, root and seedling were 10,080,256, 12,955,867, 11,564,402 and 13,241,021 reads, respectively. Unpaired reads were from 1,508,503 (branch) to 3,699,463 (stem) in all samples. When using all those reads as input for Trinity (HAAS et al., 2013), we obtained a total of 462,260 transcripts with a mean for N50 length of 2140 bp (Table 1), with 112,850, 139,535, 129,126 and 80,749 contigs for stem secondary xylem, branch secondary xylem, non-lignified tissues (root, flower, seedling, leaf) and unpaired reads, respectively (Additional File 2). Contigs coming from lignified samples were subsequently used for differential expression analyses.

4.3.4 Unigenes differentially expressed in lignified tissues between 12- and 60-year-old trees

Differentially expressed transcripts in all the comparison groups with DESeq program obtained a 95% confidence level (Figure 1-2). In the case of the branch secondary xylem transcripts differentially expressed from both 12- and 60-year-old teak trees with repetitions, the dispersion plot (Figure 1a) showed the presence of significant genes differentially expressed between both ages, showing a normalized grouping tendency in most of the transcripts with the fitted curve. Also, in figure 1b all the differentially expressed transcripts are exposed in red dots. The dispersion plot (Figure 1c) of stem secondary xylem transcripts differentially expressed from both 12- and 60-year-old teak trees (with repetitions) showed a normalized grouping tendency with a fitted curve. Several differentially expressed transcripts in stem secondary xylem were also obtained (red dots, Figure 1d). Additionally, looking for differentially expressed genes between all branch and stem samples (Additional File 5), the contrast between both tissues is clear. As well, Figure 1 exhibited almost the same quantity of differentially expressed and shared genes between both tissues. When plotting stem and branch against non-lignified tissues (flower, seedling, leaf and root) (Figure 1e-f), still stem exhibited more genes differentially expressed compared to branch.

Finally, with DESeq, we obtained 1,502 and 931 differentially expressed genes for stem and branch secondary xylem, respectively, when comparing 12- and 60-year-old trees (Table 1, Figure 2). Also, differential expression between branch and stem secondary xylem, stem secondary xylem against non-lignified tissues (leaf, flower, root and seedling) and branch secondary xylem against non-lignified tissues provided 28,022, 14,293 and 10,783 genes, respectively (Figure 2).

4.3.5 Functional annotations of unigenes differentially expressed in lignified tissues

From the 1,502 and 931 differentially expressed transcripts for stem and branch secondary xylem, respectively (Figure 2), an annotation of 669 (44.5%) and 603 (65%) genes was achieved with a known function by Blast2Go, respectively (Table 1). Among the 669 genes annotated for stem secondary xylem, 48% (Figure 3a) exhibited strong homology (E-value smaller than $1e-50$). Also, for the same tissue, the similarity distribution showed that 89% of the genes have more than 60% identity with other plants (Figure 3b) and for the species distribution, *T. grandis* had the greatest number of matches with *Vitis vinifera*, followed by *Glycine max*, *Theobroma cacao* and *Populus trichocarpa* (Figure 3c and 3f).

Figure 1 - Differential expression of log2 ratio (fold change) versus mean between different conditions with DESeq program. a) Dispersion plot for branch secondary xylem transcripts. b) Significantly differentially expressed transcripts scatterplot for branch secondary xylem transcripts. c) Dispersion plot for stem secondary xylem transcripts. d) Significantly differentially expressed transcripts scatterplot for stem secondary xylem transcripts. e) Significantly differentially expressed transcripts scatterplot for branch secondary xylem against flower, seedling, leaf and root. f) Significantly differentially expressed transcripts scatterplot for stem secondary xylem against flower, seedling, leaf and root. Fitted curve of the spots is in red. Red dots indicate transcripts differentially expressed at 10% false discovery rate and black spots transcripts are expressed in common (ANDERS; HUBER, 2010)

Figure 2 - Venn Diagram showing number of differentially expressed genes in the different tissues and ages. For the diagram, we used leaf, flower, root, seedling, stem and branch secondary xylem, comparing young (12-years-old) and mature (60-years-old) trees for the last two tissues

On the other hand, from the 603 genes annotated for branch secondary xylem, 33% (Figure 3d) revealed an homology with e-value smaller than $1e-50$, and in the identity comparison showed that 92% of the genes have more than 60% identity with other plants (Figure 3e).

Most of the differentially expressed genes had a size between 1,000 and 4,000 bp (Additional File 6 and 7). Gene ontology (GO) tool classified the unigenes in several sub-categories for biological process, cellular component and molecular function. In stem secondary xylem (Figure 4), catabolic process (9%), cellular protein modification process (8%), response to stress (8%) and carbohydrate metabolic process represented the most abundant sub-categories in the biological process category (Figure 4a), indicating the expression of genes related to catabolic activities and stress, where several heat-shock proteins were found.

Under the molecular function category, the top 2 sub-categories were nucleotide and protein binding (29% and 24%, respectively) (Figure 4b), where four *MYB* transcription factors were found and used for subsequent analysis. In the cellular component category, plastid (21%) and protein complex (14%) were the most abundant (Figure 4c). The last two categories explain the regulation as one of the most important aspects of these differentially expressed genes in stem secondary xylem.

In branch secondary xylem (Additional File 8), all categories showed similar results to stem secondary xylem, except for the protein transport through plasma membrane function.

Further, several heat shock proteins with significant up-regulation were found in stem secondary xylem (Additional File 9-10). Presumably, the change of the season when the material was collected (first week of October 2012), with an increase of almost 3°C and 35.3 mm of rainfall (month average) from winter to spring could activate stress-related genes for both tissues in *T. grandis* (Additional File 11).

4.3.6 Metabolic pathways of unigenes

Beyond finding transcription factors, heat-shock proteins and annotating genes from secondary xylem from teak, we searched for pathways related to those genes. For branch secondary xylem 57 paths were identified in the annotated genes (Additional File 12), the most relevant of which, due to gene size, were aminobenzoate degradation, glycerolipid metabolism, phenylpropanoid biosynthesis, ascorbate and aldarate metabolism, glycosaminoglycan biosynthesis and alpha-linolenic acid metabolism (Additional File 13).

Of those pathways, some relevant differentially expressed genes with significant sizes were *NADPH oxidase* (4,755 bp), *diphosphoinositol-pentakisphosphate kinase 1-like* (3,825 bp) and *diacylglycerol kinase 1-like isoform X1* (3,731 bp) (Additional File 12).

In the case of stem secondary xylem, 88 metabolic pathways were identified for all annotated differentially expressed genes (Additional File 14), with some relevant metabolisms exhibited in Additional File 15, such as selenocompound metabolism, porphyrin and chlorophyll metabolism, drug metabolism, glycan degradation, sesquiterpenoid and triterpenoid biosynthesis, carotenoid biosynthesis, zeatin biosynthesis, and the most interesting ones, irinotecan (Figure 5a) and azathioprine-mercaptopurine metabolisms (Figure 5b), with the genes located inside the pathway.

The *ali-esterase* (Figure 5a) (which produces the irinotecan) has 3,050 bp, while the *Hypoxanthine-guanine phosphoribosyltransferase* has 1,981 bp (Figure 5b). Other relevant genes obtained from the gene ontologies and metabolic pathways are *methionine-tRNA ligase-like* (3,183 bp), *phosphoribosyltransferase* involved in drug metabolism (3,050), another *ali-esterase* blasted with *Populus* with 89% similarity (3,578 bp), *beta-galactosidase 17-like* involved in glycan degradation (4,591 bp) and *adipocyte plasma membrane-associated protein-like* involved in zeatin biosynthesis (2,908 bp) (Additional File 15).

Figure 3 - Homology analysis of *T. grandis* differentially expressed unigenes. Branch secondary xylem : a) E-value distribution. b) Similarity distribution. c) Species distribution. Stem secondary xylem: d) E-value distribution. e) Similarity distribution. f) Species distribution

Figure 4 - Gene ontology (GO) assignment for the unigenes differentially expressed of *T. grandis* stem secondary xylem. GO assignments (multilevel pie chart with term filter value 5) as predicted for (a) biological process, (b) molecular function and (c) cellular components. The number of unigenes assigned to each GO term is shown behind semicolon

Figure 5 - (a) Irinotecan metabolism with the teak **ali-esterase** enzyme (EC 3.1.1.1). (b) Azathioprine-mercaptopurine metabolism with the teak **phosphoribosyltransferase** enzyme (EC 2.4.2.8). Enzymes are denoted in grey

4.3.7 Phylogenetic analysis of the teak R2R3-MYB gene family

TgMYB1 showed a predicted coiled-coil (CC) domain (*MYB-CC* family) (Additional File 16), a subtype within the *MYB* superfamily, as defined by Rubio et al. (2001). *TgMYB2*, *TgMYB3* and *TgMYB4* were consistent with the consensus DNA-binding domain sequences (DBDs) defined for *R2R3-MYB* family, finding R2R3 motifs similar to those found in Arabidopsis, gymnosperm and angiosperm plants (BEDON; GRIMA-PETTENATI; MACKAY, 2007). *TgMYB2*, *TgMYB3* and *TgMYB4* presented the WT_{X1}EED_{X2}L_{X3}V_{X4}G_{X6}W and the R_{X4}C_{X1}LRW_{X3}L_{X1}P conserved motifs within the R2 region (Additional File 16).

TgMYB2 and *TgMYB4* presented the T_{X2}EE_{X2}LI_{X2}H_{X3}GNKW motif, *TgMYB3* presented the bHLH protein-binding motif ([DE]L_{X2}[RK]_{X3}L_{X6}L_{X3}R) and *TgMYB2*, *TgMYB3* and *TgMYB4* presented the PGR_{X2}N_{X1}IK_{X2}WN motif, all in the R3 region (Additional File 16). Using the complete *R2R3-MYB* family from Arabidopsis, a phylogenetic tree was obtained to elucidate functional grouping which could also be present in the teak *MYB* family (Figure 6).

TgMYB3 is located in the epidermal cell fate group, and closely-related to the flavonol glycosides group and C2 repressor motif group, the members of which participate in bHLH interactions and promoter repression (MATUS; AQUEA; ARCE-JOHNSON, 2008). *TgMYB2* were found together in the axillary meristem group and close to the *AtMYB83* (secondary wall biosynthesis).

TgMYB4 is inside the GAMYB-like genes group, which are microRNA-regulated genes that facilitate anther development (MILLAR; GUBLER, 2005). Additionally, *TgMYB1* seems to share a common ancestor with *AtMYB55*, which do not have related function yet, and can be considered an external group for not being part of the *R2R3-MYBs* family. However, using gymnosperm and angiosperm protein sequences to characterize teak *MYBs* transcription factors, we schemed the three major groups (A, B, C) and subgroups (2, 4, 8, 9, 13, 21, 22) of *R2R3-MYBs* as described by Bedon et al. (2007).

Therefore, *TgMYB2* fell into group A (pine and spruce *MYB7*, pine *MYB6*, *MYB9* and *AtMYB44* orthologous) subgroup 22 (Figure 7), which presents motifs involved in protein or DNA interactions. *TgMYB4* is found in group B (*AtMYB101* orthologous) subgroup 13 (Figure 7).

Figure 6 - Integrated phylogram of the 126 *Arabidopsis* R2R3 MYB proteins with teak MYB proteins. Consensus circular tree was conducted by neighbor-joining method and 10000 bootstraps using Mega6 software. Teak MYB proteins are denoted with red dots. Each functional group is colored. References for MYB gene functions are defined by previous reports (MATUS; AQUEA; ARCE-JOHNSON, 2008; ZHONG et al., 2008; ZHAO; DIXON, 2011)

The same author previously described group B as being present only in angiosperms. *TgMYB3* is located in group C and is closely-related to subgroups 2, 4, 8, 9, 13 and lignin biosynthesis genes *AtMYB61*, *AtMYB46*, *PgMYB4*, *PgMYB2*, *PtMYB2*. *TgMYB1* is still apart from the R2R3 MYB proteins (Figure 7), as found in the *Arabidopsis* grouping (Figure 6), as expected. Altogether, although R2R3 motifs have several differences in *T. grandis* sequences, they grouped closely to secondary wall biosynthesis genes from other species.

Figure 7 - Phylogenetic tree of gymnosperm and angiosperm R2R3-MYB proteins. The neighbor-joining method was used using 10000 bootstraps with several spruce, pine, *Arabidopsis* and teak protein MYB sequences. Teak MYB proteins are denoted with a diamond. The bar indicates the evolutionary distance of 0.2%. *Arabidopsis* proteins were chosen as landmarks indicating the three main groups (circles A, B and C) and subgroups (Sg next to bracket; nd, not determined) defined by BEDON; GRIMA-PETTENATI; MACKAY (2007)

4.3.8 Gene expression of MYB transcription factors in teak

Quantitative real time PCR analysis showed that four teak *MYBs* are differentially expressed in lignified tissues, being *TgMYB1*, *TgMYB2*, *TgMYB4* up-regulated and *TgMYB3* down-regulated (Figure 8-9).

In leaves and roots, *TgMYB1*, *TgMYB2* and *TgMYB4* showed almost no expression levels compared to lignified tissues. *TgMYB3* was expressed much higher in leaves than the other tissues, and stem secondary xylem of both ages is shown as down-regulated. The up-regulated genes *TgMYB1* and *TgMYB4* showed comparatively higher expression in stem secondary xylem and sapwood (3-fold and 2-fold, respectively) (Figure 9-10) in mature (60-years-old) compared to young (12-years-old) trees.

Figure 8 - Expression patterns of some genes from the DESeq analysis. We chose five *MYB* transcription factors and one *NAC* gene from the differentially expressed unigenes obtained when comparing stem secondary xylem from mature and young trees. The fold changes of the genes were calculated as the log2 value

Figure 9 - Expression of teak *MYB* genes. Relative quantification of expression was examined in different tissues (leaf, root, stem and branch secondary xylem from different ages). The name of each gene is indicated at the top of each histogram. Tissues considered are shown at the bottom of the diagrams. $\pm$ means SE of three biological replicate samples. * $p < 0.05$ according to F-test. Y-axis indicates the relative expression level of each gene compared to the control tissue (leaves). *EF1 α* was the endogenous control used according to GALEANO et al. (2014)

Inversely, *TgMYB2* expression is 2-fold higher (Figure 9) and 60-fold higher (Figure 10) in stem secondary xylem and sapwood, respectively, of young teak trees. The down-regulated gene *TgMYB3* showed similar expression pattern in stem secondary xylem and sapwood of trees from both ages (Figure 9-10), although in the DESeq expression level stem

secondary xylem from 60-year-old trees showed almost 150-fold less expression compared to 12-year-old trees. Branch secondary xylem of 12-year-old trees seems to have considerable expression levels in *TgMYB1* and *TgMYB4* genes compared to leaves (3- and 6- fold, respectively), but similar expression compared to stem secondary xylem at both ages, with a 95% statistical confidence level. These results confirm that the unigenes obtained from the transcriptome assembly were differentially expressed, with differences between both ages, once the real time PCR (Figure 9) is in agreement with DESeq results (Figure 8).

Figure 10 - Relative expression levels of teak *MYB* genes in sapwood. The name of each gene is indicated at the top of each histogram. Tissues considered are shown at the bottom of the diagrams. $\pm$ means SE of three biological replicate samples. * $p < 0.05$ according to F-test. Y-axis indicates the relative expression level of each gene compared to the control tissue (leaves). *EF1a* was the endogenous control used according to (GALEANO et al., 2014)

4.4 Discussion

4.4.1 *T. grandis* transcriptome

The high sensitivity of sequencing technologies presents the RNA-Seq as the preferred choice for transcriptome studies (ZENONI et al., 2010), widely replacing the microarray-based gene expression technology (ROBERTS et al., 2011; MUTZ et al., 2013), the sequencing of cDNA libraries, the SAGE and SuperSAGE analysis (GORDO et al., 2012). Despite the forestry and economic importance of *T. grandis* around the world, it is very poorly characterized, with only 134 gene sequences deposited in Genbank (access 30/10/2014), most of them being alleles used for molecular markers (GANGOPADHYAY et

al., 2003; SHRESTHA; VOLKAERT; STRAETEN, 2005; VERHAEGEN et al., 2005; FOFANA et al., 2008, 2009; ALCÂNTARA; VEASEY, 2013; LYNGDOH et al., 2013).

Also, previous genetic studies have focused on proteomic analysis and kinetics of *T. grandis* (TIWARI et al., 2006; LACRET et al., 2011; QUIALA et al., 2012; BALOGUN; LASODE; MCDONALD, 2014). In this study, we have generated more than 192 million sequence reads (100 pb) corresponding to 233 Gb of raw sequence data from several tissues (Table 1). *T. grandis* without a sequenced genome and a lack of a sequenced genome in the Lamiales order makes analysis of the teak RNAseq dataset more difficult. Teak reads mapped against tomato, eucalyptus and populus genomes. *Tectona grandis* is a diploid species with $2n=36$ chromosomes (GILL; YEDI; BIR, 1983). Ohri e Kumar (1986) estimated the size of its genome by cytogenetic studies, finding about 465 Mbp ($1C=0.48$ pg), which is about the same and 2-fold larger than the genome of *Populus trichocarpa* and *Arabidopsis thaliana*, respectively.

A. thaliana has at least 1,533 transcription factor genes (approximately 6% of the coding capacity of its genome) (GONG et al., 2004), and assuming a similar proportion of transcription for *T. grandis*, all the transcription factors could be estimated in 27.9 Mbp. Comparatively, 270 million reads were obtained from *Phaseolus vulgaris* (WU et al., 2014), 71 million reads were generated from stem-root of *Piper nigrum* (GORDO et al., 2012), 59 million reads were generated from *Vitis vinifera* (ZENONI et al., 2010), 42 million reads were obtained in *Camellia sinensis* (WEI et al., 2014) and close to 20 million reads were obtained from *Petroselinum crispum* (LI et al., 2014) and *Isatis indigotica* (TANG et al., 2014). In eucalyptus, pyrosequencing gave 1.1 million reads (VILLAR et al., 2011). In that sense, Trinity appears as a good choice to assemble de novo full-length transcripts for species without reference genome (GRABHERR et al., 2011) because it corrects almost 99% of the sequencing errors. Trinity is a strategy which assembles a set of unique sequences from reads aided by the creation of independent de Bruijn graphs, each representing one group of sequences and assembles isoforms within the groups, running in parallel in a computational cluster (MARTIN; WANG, 2011; COMPEAU; PEVZNER; TESLER, 2011). We obtained 462,260 transcripts from all tissues using the Trinity platform. Recent studies found 33,238 unigenes in *Isatis indigotica* (TANG et al., 2014), 62,828 unigenes from *Phaseolus vulgaris* representing 49 Mb (WU et al., 2014), 50,161 unigenes from *Petroselinum crispum* (LI et al., 2014) and 60,000 unigenes in *Camellia sinensis* (WEI et al., 2014). Several trees have generated significant genes, such as *Salix matsudana* with 106,403 unigenes (RAO et al.,

2014), *Populus trichocarpa* with 36,000 unigenes (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010), *Populus euphratica* with 86,777 unigenes (QIU et al., 2011) and *Fraxinus spp.* with 58,673 unigenes (BAI et al., 2011).

4.4.2 RNAseq provided several useful unigenes differentially expressed in lignified tissues of *T. grandis*

From the transcriptome obtained, we were able to identify differentially expressed genes with DESeq program, obtaining an invaluable gene dataset of lignified tissues of teak. DESeq method is a parametric approach which works with technical replicates, with the variance and mean linked by local regression, and uses the negative binomial distribution (a natural extension of the Poisson distribution) to visualize the intensity-dependent ratio of expression data (ANDERS; HUBER, 2010; WANG et al., 2010; GARBER et al., 2011; KVAM; LIU; SI, 2012). Our analysis for differentially expressed genes is based in biological replicates, which allow a solid biological interpretation. We found 1,502 and 931 differentially expressed genes in stem and branch secondary xylem, respectively, between young and mature teak trees. Recent studies have shown substantial differences obtaining differentially expressed genes. Gordo et al. (2012) obtained 22,363 transcripts from stem-root of *Piper nigrum*. In stem, almost 3,000, 8,266 and 1,042 differentially expressed genes were obtained in *Populus trichocarpa* (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010), alfalfa (YANG et al., 2011) and *Brassica juncea* (SUN et al., 2012), respectively. In eucalyptus, 50,000 contigs were obtained (VILLAR et al., 2011) and in *Salix matsudana* 292 miRNA stress-related differentially expressed genes (RAO et al., 2014).

It is common to find in some treatments no more than 1,000 differentially expressed genes, as the case of *Camellia sinensis* (WEI et al., 2014). To compare between two general tissue types that are of interest for woody biomass production (MIZRACHI et al., 2010) such as stem and branch, along with the comparison between young (12-years-old) and mature (60-years-old) trees, we properly performed the differential expression procedure with DESeq program (Figure 1-2). All the differentially expressed genes in both tissues presented high homology (by lower p-values), matched with lignified plants, presented sizes between 1,000-4,000 bp (Figure 3), and after annotations, catabolic processes, response to stress, carbohydrate metabolism, protein binding, transport and plastid localization were the most abundant sub-categories, consistent with biopolymers production, transport, storage and xylogenic-related genes in the transcriptome of *E. grandis* × *E. urophylla* hybrid clone

(MIZRACHI et al., 2010), *Picea glauca* (PAVY et al., 2008) and *Populus trichocarpa* (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010).

Several differentially expressed genes in the transition between young to mature trees in secondary xylem include glycosaminoglycan biosynthesis, glycan degradation cell wall carbohydrate (galactose, starch, sucrose) metabolic genes (Additional File 12 and 14), *diacylglycerol kinase*, *ali-esterase*, *pectin*-related genes and *galactosyl transferase* (Additional File 8) likely involved in cell wall synthesis and extension, plant defense, cellulose, hemicellulose, lignin and pectin formation were found. In *Pinus taeda* (YANG et al., 2004; PRASSINOS et al., 2005) and in aspen (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010), several pectin esterases, carbohydrate genes and transcription factors highly expressed in woody tissues were found. Additionally, studies with drought have found differentially expressed genes from cell wall and carbohydrate biosynthetic processes which respond greatly to drought stress and enhance mechanical resistance of drought-exposed cells (WU et al., 2014). Also, several kind of stress in different plants have shown up- and down-regulation of metabolic pathways such as carbon metabolism, sucrose and starch synthesis in maize with drought stress (KAKUMANU et al., 2012). Both, stem and branch secondary xylem indicated a high proportion of predicted genes localized in plastids and plasma membrane in *T. grandis*, as was found in *P. nigrum* stem (GORDO et al., 2012).

4.4.3 Activation of stimulus response genes and heat-shock proteins with local environmental changes

Differentially expressed genes included several stimulus response genes, cell death-associated genes and phenylpropanoid biosynthetic genes (Additional File 13-15) and their presence correlated with both changes in the environmental conditions and maturation of the tree. Consequently, several heat shock proteins were observed in stem secondary xylem with a noticeable expression by DESeq (Additional File 9-10). Similar to our results, during ecodormancy of *Quercus petraea* several stress-related genes were found, including one heat shock protein (HSP18.2), as one of the most expressed genes among all, which is regulated by ABA (UENO et al., 2013). Ecodormancy state occurs when temperatures rise from late winter to early spring to prevent bud burst, so heat shock proteins show chaperone activity in order to maintain the proteins in their functional conformation and prevent degradation and damage during heat stress (UENO et al., 2013). Curiously, genes encoding enzymes related to heat

stress and heat-shock proteins showed differential expression between climacteric treatments in *Pyrus ussuriensis* fruits (HUANG et al., 2014).

Also, Schrader et al. (2004) compared regulatory networks between primary and secondary meristems, finding common regulatory mechanisms between both stages, with several stress-related genes playing a role in protecting the secondary xylem against different circumstances, and acting with different sugars (such as sucrose synthase and glycosylases) to transport them into the cambial zone, as also observed by YANG et al. (2004) with one heat-shock protein and cell-wall related genes in *Pinus taeda*. All the heat-shock proteins and almost all the differentially expressed genes showed in young secondary xylem 2-fold more expression (square root of the log2 value of DESeq) than mature ones (Additional File 10), suggesting elevated rates of protein turnover in younger stages of teak, as might be expected for actively dividing cells compared to mature tissues (60-years-old).

4.4.4 MYB transcription factors revealed phylogenetic grouping and distinct expression during maturity

Transcription factors differentially expressed during vascular development and secondary growth are of high interest due to the economic importance of wood, because they play roles as regulators which control response networks, being reflected in the modification of wood and fiber qualities, and several studies have identified regulated transcription factors with the onset of secondary growth (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010). *MYB* transcription factor family plays a fundamental role in xylem development in different plant species and it is a critical regulator of phenylpropanoid pathway (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010) such as *Arabidopsis thaliana* (ZHONG; RICHARDSON; YE, 2007; MATUS; AQUEA; ARCE-JOHNSON, 2008; KO; KIM; HAN, 2009; BHARGAVA et al., 2010; KIM et al., 2012), maize (FORNALÉ et al., 2010), wheat (MA; WANG; ZHU, 2011), and trees such as *Picea glauca* (BEDON; GRIMA-PETTENATI; MACKAY, 2007), *Pinus taeda* (PATZLAFF et al., 2003; BOMAL et al., 2008), Eucalyptus genera (GOICOECHEA et al., 2005; LEGAY et al., 2007) and populus genera (KARPINSKA et al., 2004; PRASSINOS et al., 2005; DHARMAWARDHANA; BRUNNER; STRAUSS, 2010; MCCARTHY et al., 2010).

GO process annotation and manual verification of transcription factors led to finding *MYB* transcription factors tissue-specific, and whose function is clearly linked to the teak maturation. To classify and predict the biological role of four *MYB* transcription factors,

domain protein sequence was analyzed (Additional File 15) and phylogenetic distances were calculated comparatively with all *MYB* transcription factors from *Arabidopsis thaliana* and some trees. In that sense, *TgMYB1* is part of the MYB-CC family; *TgMYB2*, *TgMYB3* and *TgMYB4* are part of the R2R3-MYB family with *TgMYB3* displaying the bHLH motif (Additional File 15).

Our data show that the DNA-binding domains (DBDs) of *T. grandis* are conserved. However, *TgMYB3* was found in the arabidopsis *MYB* group which participates in bHLH interactions, promoter repression and lignin biosynthesis genes, while *TgMYB4* is in the GAMMYB-like group and inside the group “B” which is only present in angiosperms. Also, *TgMYB2* is close to secondary wall biosynthesis function and protein or DNA interactions (Figure 6-7). *TgMYB1* is outside the groups and need to be more elucidated.

This diversity between *T. grandis*, Arabidopsis and some trees might give different roles in the secondary xylem formation. It has been identified in poplar 297 *MYB* members (DHARMAWARDHANA; BRUNNER; STRAUSS, 2010) and 126 R2R3-MYB transcription factors in Arabidopsis (MATUS; AQUEA; ARCE-JOHNSON, 2008). But, with the transcript expression levels by DESeq (log2-ratio) and through qRT-PCR analysis of four of the *MYB* transcription factors in *T. grandis*, it was found that *TgMYB1* and *TgMYB4* showed more expression in secondary xylem and sapwood of mature trees than young ones, *TgMYB2* less expression levels in lignified tissues of mature than young trees and *TgMYB3* a down-regulation in secondary xylem and sapwood at both ages. High expression of the Arabidopsis *AtMYB103*, *AtMYB85*, *AtMYB52*, *AtMYB54*, *AtMYB69*, *AtMYB42*, *AtMYB43*, *AtMYB20*, *AtMYB58*, *AtMYB63*, *AtMYB75*, as a simplified example, has been associated with secondary wall thickening (ZHONG et al., 2008; ZHAO; DIXON, 2011).

In *Picea glauca*, *PgMYB2*, *PgMYB4* and *PgMYB8*, closely-related with *TgMYB3* by the phylogenetic analysis (Figure 7), were expressed in stem and root (Bedon et al. 2007), curiously expressed preferentially in the secondary differentiating xylem of both juvenile and mature trees. The same authors described that some *MYB* genes were highly expressed in apical stem, such as *PgMYB6*, *PgMYB7* and *PgMYB11*, being grouped with *TgMYB2* (Figure 7). To conclude, the *T. grandis* *MYB* family structure and expression is not all that divergent from the gymnosperm and small flowering plants, such as *Arabidopsis thaliana*.

Even though there is only a 5% increase in wood density going from 50- to 51-year-old trees compared to trees going from 8- to 9-year-old trees (when teak responds to fertilization and cultural operations in the initial years), Bhat et al. (2005) speculated that

much of the growth characteristics and biological changes related to wood traits (noticed in early ages) should be absent in later years when sapwood gives way to the comparatively stable heartwood.

In our results, *TgMYB1* and *TgMYB4* are differentially expressed in secondary xylem, and highly expressed in sapwood of 60-year-old trees compared to young ones, presumably because they are key in conferring some woody properties that 12-year-old sapwood does not have. Presumably, *TgMYB1* and *TgMYB4* could explain the transition from sapwood (usually called "baby teak") to heartwood and they are probably clues in enhancing the heartwood content and natural resistance as a genetic character, something desirable for teak producers.

4.5 Implications and Perspectives of this study

These results, the first dataset of sequences of the Lamiales order and *Tectona* genus, will open new perspectives for studies of diversity, ecology, breeding and genomic programs aiming to understand deeply the biology of this species. In tropical zones, woody plants go through seasonal cycles with two stages: a growing period when environmental conditions are favorable and a period of non-growth in winter, and these phenological cycles have been shown to be strongly affected by an increase in the temperature, which has an impact on the biological processes (UENO et al., 2013).

Heat-shock proteins have a crucial role in maintaining the proteins in their functional conformation when temperatures rise, preventing degradation and damage during heat stress, from late winter to early spring (UENO et al., 2013). Indeed, heat-shock proteins aid defending *T. grandis* against those environmental changes in the region sampled and need to be more studied. Similarly, the molecular mechanism underlying regulation of wood formation in tropical forest trees remains poorly understood. Our transcriptomic study reported changes in the accumulation of up-and down-regulated genes through the maturation of *T. grandis*.

Among all these genes, four were chosen, quantified and validated by qRT-PCR. The up-regulation of *TgMYB1*, *TgMYB2* and *TgMYB4* in teak secondary xylem (*TgMYB1* and *TgMYB4* in mature and *TgMYB2* in young trees) may also be triggered by other transcription factors, especially NAC master regulators (PAVY et al., 2008), in response to cell wall thickening, regulation of phenylpropanoid genes, changing environmental conditions prevailing between winter and spring and as a possible response to other biotic and abiotic stimuli.

It is important to take into account how the maturation of teak can influence the expression of the *TgMYB1* and *TgMYB4* transcription factors and a decrease of *TgMYB2*, once they are selectively expressed in mature sapwood. The drastic differences in wood quality comparing young to mature trees are well known, and heartwood and sapwood are considered high heritability characters, so they seem to be important features to be included in breeding programs (BHAT et al., 2005), particularly when short rotations, such as the Brazilian ones (20 years) are targeted. Also, the quality of the juvenile wood itself will be an important target for improvement, and this can be assessed at an earlier stage, along with seeking trees that keep up fast juvenile growth speed for more years reducing the rotation age and yielding higher percentage of heartwood (BHAT et al., 2005).

Globally, the current study provides several novel observations: (i) it contributes an extensive transcriptome analysis for a tropical wood with respects to secondary growth; (ii) we achieved transcription (gene expression) disparity from a gradient of young to mature secondary xylem and sapwood, identifying several tissue- and developmental stage-specific genes; (iii) the secondary growth has unique molecular biology processes, which includes DNA interacting proteins, regulators of lignin pathway, multitude of stress-related proteins, peptide transporters, carbohydrate metabolic genes, pectin formation and teak-specific genes such as irinotecan; (iv) our results provide for the first time differentially expressed heat-shock proteins and *MYB* transcription factors in teak (*MYB-CC* and *R2R3-MYB* types), contributing to the understanding of the molecular mechanisms in tropical wood, incentives to conduct reverse genetics and plant transformation in *T. grandis*, and they will aid in understanding regulatory networks of wood formation.

4.6 Conclusions

The transcriptome of *T. grandis* was assembled using about 192 million reads without a reference genome. More than 2,000 differentially expressed genes, including highly expressed heat-shock proteins, carbohydrate metabolic genes and *MYB* transcription factors were obtained, with two biological replicates of 12 and 60-year-old trees. Analyses using DESeq revealed that there are transcriptome changes in maturation of teak secondary xylem from 12- to 60-year-old trees, while enriched GO groups for branch and stem secondary xylem were found similar. In addition, this is the first attempt to assemble transcripts and characterize *MYB* transcription factors from secondary xylem of *T. grandis*. Four *MYB* transcription factors were classified and characterized, finding three of them with high

expression and one down-regulated in lignified tissues, with significant correlation between DESeq and qRT-PCR expression analysis. The understanding of gene function of woody tissues in forest tree species is highly challenging due to the lack of standard tree transformation, also, due to plant size, slow growth and long generation time, which make breeding programs a very long process. In order to contribute to assist selection of highly productive trees, next-generation sequencing has become the closest technology to identify target genes among thousands of candidates. In conclusion, the data obtained can be used in applied and basic science along with biotechnological approaches to improve tropical trees.

Additional Files

Additional File 1 - RIN factor of all samples used for Illumina sequencing

Additional File 2 - Raw data, cleaning data and assembly

Tissues	Raw data	Clean data*	% Errased reads	Total trinity transcripts	Total trinity components	Contig N50
Stem secondary xylem 12yoR1	14168695	12640036	10,79	112850	48633	2291
Stem secondary xylem 12yoR2	16166720	14439726	10,68			
Stem secondary xylem 60yoR1	16412620	14618307	10,93			
Stem secondary xylem 60yoR2	16207144	14402146	11,14			
Branch secondary xylem 12yoR1	14185715	12838285	9,5	139535	59771	2365
Branch secondary xylem 12yoR2	14133086	12698822	10,15			
Branch secondary xylem 60yoR1	18783055	18081842	3,73			
Branch secondary xylem 60yoR2	15990913	15384693	3,79			
Flower	13725131	12348918	10,03	129126	65592	2178
Leaf	17947895	16010790	10,79			
Root	16248866	14411320	11,3			
Seedling	18871794	16653783	11,75			
Unpaired	-	-	-	80749	53522	1725
TOTAL	192841634	174528668	-	462260	227518	
Media	-	-	9,55	-	-	2140

*Includes unpaired data R1= replicate 1 R2= replicate 2

Additional File 3a - Common contigs between Populus, Eucalyptus and tomato genomes for
branch of 12-year-old teak trees

Blas2Go_ID	% Similarity	Accession Number	length
<i>rrna intron-encoded homing endonuclease</i>	91	XP_003614385	872
<i>atp synthase subunit alpha</i>	98	CAK18872	341
<i>cell wall-associated hydrolase</i>	76	XP_003637074	431
<i>atp synthase subunit beta</i>	77	XP_003614388	784
<i>cell wall-associated hydrolase</i>	84	XP_003637074	1272
<i>ycf68 protein</i>	68	XP_003610227	1504
<i>ycf68 protein</i>	67	XP_003610227	1539
<i>cell wall-associated hydrolase</i>	76	XP_003637074	1554
<i>rrna intron-encoded homing endonuclease</i>	93	EJY66653	503
<i>hydrolase</i>	69	XP_003588355	1550
<i>retrotransposon protein</i>	100	AFK37255	307
<i>atp synthase subunit alpha</i>	98	XP_003588326	1120

Additional File 3b - Common contigs between Populus, Eucalyptus and tomato genomes for
branch of 60-year-old teak trees

Blas2Go_ID	Similarity %	Accession Number	length
<i>nadh dehydrogenase subunit b</i>	98	ESQ30846	480
<i>atp synthase subunit alpha</i>	98	CAK18872	358
<i>nadh dehydrogenase subunit 7</i>	97	EPS74730	553
<i>nadh dehydrogenase subunit 2</i>	94	XP_002439577	449
<i>nadh dehydrogenase subunit 5</i>	100	XP_003588311	547
<i>nadh dehydrogenase subunit</i>	83	CAN64512	354
<i>ribosomal protein l2</i>	98	YP_006503856	1037
<i>ycf68 protein</i>	67	XP_003610227	2170
<i>cell wall-associated partial</i>	62	XP_003637074	2477
<i>atp synthase subunit alpha</i>	97	AGC78945	124
<i>atp synthase subunit alpha</i>	98	XP_003588326	1137

Additional File 4 - Genes found with the mapping

>*Tghyd*

AATGCTGCACCCTAGATGGCGAAAGTCCAGTAGCCGAAAGCATCACTAGCTTACGCTCTGACCCGAGTAGCATG
GGACACGTGGAATCCCGTGTGAATCAGCAAGGACCACCTTGCAAGGCTAAATACTCCTGGGTGACCGATAGCGA
AGTAGTACCGTGAGGGAAGGGTGAAGAAGACCCCATCGGGGAGTGAAATAGAACATGAAACCGTAAGCTCCCA
AGCAGTGGGAGGAGCCAGGGCTCTGACCGCGTGCCTGTTGAAGAATGAGCCGGCGACTCATAGGCAGTGGCTTG
GTTAAGGGAACCCACCGGAGCCGTAGCGAAAGCGAGTCTTCATAGGGCAATTGTCACTGCTTATGGACCCGAAC
CTGGGTGATCTATCCATGACCAGGATGAAGCTTGGGTGAACTAAGTGGAGGTCCGAACCGACTGATGTTGAAG
AATCAGCGGATGAGTTGTGGTTAGGGGTGAAATGCCACTCAAACCTCT

>*TgYcf68*

CCCGCGGGTAAGACAGAGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCTGTAGGTGGCTTTTTAA
GTCCGCCGTCAAATCCAGGGCTCAACCCTGGACAGGCGGTGGAACTACCAAGCTGGAGTACGGTAGGGGCAG
AGGGAATTTCCGGTGGAGCGGTGAAATGCGTAGAGATCGGAAAGAACCAACGGCGAAAGCACTCTGCTGGGC
CGACACTGACACTGAGAGACGAAAGCTAGGGGAGCGAATGGGATTAGATACCC

>*TgAtpsynth*

TCGAGATAACACATGCTCACCGCTTGTGCGGGCCCCCGTCAATTCCCTTTGAGTTTCATTCTTGCGAACGTACTC
CCCAGGCGGGATACTTAACGCGTTAGCTACAGCACTGCACGGGTCGATACGCACAGCGCCTAGTATCCATCGTT
TACGGCTAGGACTACTGGGGTATCTAATCCCATTCGCTCCCCTAGCTTTTCGTCTCTCAGTGTCAAGTGTGCGCCC
AGCAGAGTGCTTTCGCCGTTGGTGTTCTTTCCGATCTCTACGCATT

>*Tgmitprot*

GACTGCCGGAGCTTGATACGGTTTCCCGATCGGAGATCCATGGATCACAGACGGTATCTCCCCATGGCCTTTC
GCCTCTGAAAGCGTCCCTTCA

Additional File 5 - Significantly differentially expressed transcripts plot for stem-branch genes. Grey plots indicate transcripts differentially expressed and black spots transcripts expressed in common

Additional File 6 - Length and number of sequences for stem differentially expressed genes

Additional File 7 - Length and number of sequences for branch differentially expressed genes

Additional File 8 - Gene ontology (GO) assignment for the unigenes differentially expressed of *T. grandis* branch secondary xylem. GO assignments (multilevel pie chart with term filter value 5) as predicted for (a) biological process, (b) molecular function and (c) cellular components. The number of unigenes assigned to each GO term is shown behind semicolon

Additional File 9 - 43 genes highly differentially expressed between stem secondary xylem
from 12- and 60-year-old trees

Gene	baseMean Stem secondary xylem 12yo	baseMean Stem secondary xylem 60yo	p value
<i>transmembrane bax inhibitor motif-containing</i>	6097,3	868	0
<i>NTGP3 putative rac protein</i>	2858,4	482,2	0,00186308
<i>kda class i heat shock</i>	5769,4	898,1	0,00065701
<i>splicing factor U2af small subunit A-like</i>	2229,9	266,8	0,0001194
<i>heat shock 70 kda</i>	1147,2	244,6	0,02134347
<i>voltage-gated potassium channel subunit beta-like</i>	1591,5	122,3	3,39E-06
<i>atp binding cassette subfamily b4 isoform 2</i>	2252,4	148,1	3,63E-07
<i>galactinol--sucrose galactosyltransferase 2-like isoform X1</i>	1299	15568,5	9,87E-07
<i>carboxylesterase 8-like</i>	6,8	4352,9	3,17E-25
<i>chaperone</i>	1790,9	289,7	0,00191109
<i>chlorophyll a b binding</i>	4640,4	676,8	0,00041139
<i>protein mizu-kussei 1-like</i>	1629,2	318,2	0,00859593
<i>oligopeptide transporter 1</i>	3507	658,8	0,00399448
<i>glucose-6-phosphate phosphate translocator chloroplastic-like</i>	163	1158,9	0,00115311
<i>oxygen-evolving enhancer protein 2, chloroplastic-like</i>	1696,1	393,8	0,02730737
<i>desumoylating isopeptidase 1-like</i>	4061,1	782,2	0,00456294
<i>nuclear pore complex protein nup98-nup96-like isoform x2</i>	1849,9	70,4	2,04E-09
<i>kda class i heat shock protein</i>	146647	38425,2	0,03359285
<i>heat shock protein 83-like</i>	3620,1	376,1	2,02E-05
<i>ribulose biphosphate carboxylase oxygenase activase chloroplastic-like</i>	1890,4	186,3	2,60E-05
<i>large proline-rich protein bag6-like isoform x1</i>	141,1	1527,4	2,13E-05
<i>f-box protein skip27-like</i>	1201,4	269,8	0,02854693
<i>kda heat shock peroxisomal-like</i>	22363	3960,7	0,00159461
<i>heat shock protein 83-like</i>	90973,5	22710	0,02386294
<i>heat shock protein 83-like</i>	11170	1237,1	2,12E-05
<i>pre-mrna-splicing factor cef1-like</i>	1674,8	359,6	0,01647392
<i>Inositol-3-phosphate synthase</i>	3625	458,8	0,00012983
<i>nuclear pore complex protein nup98-nup96-like isoform x1</i>	1682,8	72,6	9,07E-09
<i>bi1-like</i>	13792,6	1735,6	7,21E-05
<i>peptidyl-prolyl cis-trans isomerase FKBP65-like</i>	1870	259,3	0,00053044
<i>kda class i heat shock</i>	31868,3	8141,1	0,02882191
<i>nadh dehydrogenase</i>	1759,6	427,8	0,03588352
<i>nucleoporin</i>	8824,3	292,8	2,71E-11
<i>diacylglycerol kinase 1-like isoform X1</i>	1724	252,6	0,00089485
<i>protein vip1-like isoform X2</i>	1444,5	234	0,00244917
<i>protein dj-1 homolog b-like</i>	2427,3	564,1	0,02270032
<i>glucosidase 2 subunit beta-like</i>	1862,5	148,9	3,66E-06
<i>kda class i heat shock</i>	1588,1	87,8	1,41E-07
<i>trans-alpha-bergamotene synthase</i>	197,5	1281,1	0,00191109
<i>glutathione s-</i>	3156,2	423	0,00024414
<i>60s ribosomal protein l10-like</i>	4243	1091,4	0,03743841
<i>luminal binding protein</i>	4809,8	1082,3	0,01417015

Additional File 10 - Other relevant differentially expressed genes from secondary xylem. We chose other genes with the highest expression between young (12-years-old) and mature (60-years-old) trees, and performed a transformation of root square in order to visualize their values

Additional File 11 - Monthly average of relative humidity, rainfall and temperature in Piracicaba 2012-2013

Additional File 12 - Branch secondary xylem pathways found by Kegg

Pathways	Number of Sequences	Number of enzymes
1. Starch and sucrose metabolism	16	10
2. Amino sugar and nucleotide sugar metabolism	6	3
3. Purine metabolism	6	2
4. Methane metabolism	4	3
5. Porphyrin and chlorophyll metabolism	4	2
6. Thiamine metabolism	4	2
7. Glyoxylate and dicarboxylate metabolism	4	2
8. Galactose metabolism	4	4
9. Aminobenzoate degradation	3	2
10. Glycolysis / Gluconeogenesis	3	2
11. Glutathione metabolism	3	2
12. Oxidative phosphorylation	3	3
13. Terpenoid backbone biosynthesis	3	2
14. T cell receptor signaling pathway	3	1
15. Pentose phosphate pathway	3	2
16. Phosphatidylinositol signaling system	3	2
17. Fructose and mannose metabolism	3	3
18. Glycerolipid metabolism	3	2
19. Aminoacyl-tRNA biosynthesis	2	2
20. Pentose and glucuronate interconversions	2	2
21. Carbon fixation in photosynthetic organisms	2	1
22. Riboflavin metabolism	2	1
23. Carotenoid biosynthesis	2	1
24. Pantothenate and CoA biosynthesis	2	1
25. Glycosphingolipid biosynthesis - globo series	2	1
26. Sphingolipid metabolism	2	1
27. Sulfur metabolism	2	1
28. Biosynthesis of terpenoids and steroids	2	1
29. Inositol phosphate metabolism	2	1
30. Phenylalanine, tyrosine and tryptophan biosynthesis	2	2
31. Carbon fixation pathways in prokaryotes	2	2
32. Biotin metabolism	1	1
33. Toluene degradation	1	1
34. Pyruvate metabolism	1	1
35. Phenylpropanoid biosynthesis	1	1
36. Butanoate metabolism	1	1
37. Arginine and proline metabolism	1	2
38. Phenylalanine metabolism	1	1
39. Fatty acid degradation	1	1
40. Glycine, serine and threonine metabolism	1	1
41. Caprolactam degradation	1	1
42. Propanoate metabolism	1	1
43. Fatty acid elongation	1	1
44. Fatty acid biosynthesis	1	1
45. Tryptophan metabolism	1	1
46. N-Glycan biosynthesis	1	1
47. Geraniol degradation	1	1
48. Valine, leucine and isoleucine degradation	1	1
49. Nicotinate and nicotinamide metabolism	1	1
50. Primary bile acid biosynthesis	1	1
51. Lysine degradation	1	1
52. Ascorbate and aldarate metabolism	1	1
53. Glycosaminoglycan biosynthesis - heparan sulfate / heparin	1	1
54. Glycerophospholipid metabolism	1	1
55. alpha-Linolenic acid metabolism	1	1
56. Linoleic acid metabolism	1	1
57. Cysteine and methionine metabolism	1	1

Additional File 13 - Relevant enzymes found for differentially expressed genes in branch. In grey, sequences higher than 3000 bp

Metabolism	Enzyme code	Size	Seq description	% Similarity
Aminobenzoate degradation	ec:3.1.3.41 - nitrophenyl phosphatase	1530	phosphoglycolate phosphatase 1B, chloroplastic-like	93% [Citrus sinensis]
	ec:3.1.3.2 - phosphatase	3825	diphosphoinositol-pentakisphosphate kinase 1-like	91% [Vitis vinifera]
Glycerolipid metabolism	ec:2.7.1.107 - kinase (ATP)	3731	diacylglycerol kinase 1-like isoform X1	86% [Solanum tuberosum]
Phenylpropanoid biosynthesis	ec:1.11.1.7 - lactoperoxidase	4755	NADPH oxidase	86% [Nicotiana tabacum]
Ascorbate and aldarate metabolism	ec:1.3.2.3 - dehydrogenase	2252	L-galactono-1,4-lactone dehydrogenase	81% [Arabidopsis thaliana]
Glycosaminoglycan biosynthesis - heparan sulfate / heparin	ec:2.4.1.224 - 4-alpha-N-acetylglucosaminyltransferase	2885	unnamed protein product	88% [Vitis vinifera]
alpha-Linolenic acid metabolism	ec:1.13.11.12 - 13S-lipoxygenase	2652	lipoxygenase	88% [Actinidia arguta]

Additional File 14 - Stem secondary xylem pathways found by Kegg

(Continue)

	Pathways	Number of Sequences	Number of enzymes
1	Starch and sucrose metabolism	17	10
2	Glycerolipid metabolism	16	4
3	Purine metabolism	15	8
4	Glycerophospholipid metabolism	15	5
5	Phosphatidylinositol signaling system	15	4
6	Glycolysis / Gluconeogenesis	11	5
7	Carbon fixation in photosynthetic organisms	10	4
8	Fructose and mannose metabolism	10	3
9	Pyruvate metabolism	9	7
10	Pentose phosphate pathway	9	3
11	Nicotinate and nicotinamide metabolism	9	1
12	Methane metabolism	8	2
13	Selenocompound metabolism	8	4
14	Galactose metabolism	8	7
15	Aminoacyl-tRNA biosynthesis	7	4
16	Cysteine and methionine metabolism	7	6
17	Valine, leucine and isoleucine degradation	6	5
18	Terpenoid backbone biosynthesis	6	3
19	Propanoate metabolism	6	4
20	Pyrimidine metabolism	6	4
21	Carbon fixation pathways in prokaryotes	5	4
22	Inositol phosphate metabolism	5	4
23	Thiamine metabolism	5	2
24	Pentose and glucuronate interconversions	5	2
25	Oxidative phosphorylation	4	3
26	Streptomycin biosynthesis	4	3
27	Glutathione metabolism	4	3
28	Porphyrin and chlorophyll metabolism	4	4
29	Sphingolipid metabolism	4	2
30	Glyoxylate and dicarboxylate metabolism	4	4
31	Drug metabolism - other enzymes	3	2
32	Arginine and proline metabolism	3	3
33	Amino sugar and nucleotide sugar metabolism	3	3
34	Other glycan degradation	3	2
35	Citrate cycle (TCA cycle)	3	3
36	Lysine degradation	3	3
37	Glycine, serine and threonine metabolism	3	2
38	Fatty acid biosynthesis	3	2
39	alpha-Linolenic acid metabolism	2	2
40	Sulfur metabolism	2	1
41	Drug metabolism - cytochrome P450	2	1
42	Metabolism of xenobiotics by cytochrome P450	2	1
43	Aminobenzoate degradation	2	1
44	Tryptophan metabolism	2	2
45	Pantothenate and CoA biosynthesis	2	2
46	Sesquiterpenoid and triterpenoid biosynthesis	2	2
47	Carotenoid biosynthesis	2	1
48	Monoterpenoid biosynthesis	2	2
49	Penicillin and cephalosporin biosynthesis	2	2
50	Glycosphingolipid biosynthesis - ganglio series	2	1
51	Glycosphingolipid biosynthesis - globo series	2	1
52	Fatty acid degradation	2	2
53	T cell receptor signaling pathway	2	2
54	Biosynthesis of terpenoids and steroids	2	1
55	Flavonoid biosynthesis	2	2
56	Phenylpropanoid biosynthesis	2	2
57	Aflatoxin biosynthesis	2	1
58	Tetracycline biosynthesis	2	1
59	Alanine, aspartate and glutamate metabolism	2	2
60	Riboflavin metabolism	2	1
61	Ascorbate and aldarate metabolism	2	2
62	Glycosaminoglycan degradation	2	1
63	Linoleic acid metabolism	1	1

Additional File 14 - Stem secondary xylem pathways found by Kegg

(Conclusion)

64	Biosynthesis of unsaturated fatty acids	1	2
65	Arachidonic acid metabolism	1	1
66	Taurine and hypotaurine metabolism	1	1
67	Insect hormone biosynthesis	1	1
68	Chloroalkane and chloroalkene degradation	1	1
69	Butirosin and neomycin biosynthesis	1	1
70	Zeatin biosynthesis	1	2
71	Limonene and pinene degradation	1	1
72	beta-Alanine metabolism	1	1
73	Indole alkaloid biosynthesis	1	1
74	D-Glutamine and D-glutamate metabolism	1	1
75	beta-Lactam resistance	1	1
76	Synthesis and degradation of ketone bodies	1	1
77	Novobiocin biosynthesis	1	1
78	Phenylalanine, tyrosine and tryptophan biosynthesis	1	2
79	Cyanoamino acid metabolism	1	1
80	Benzoate degradation	1	1
81	Phenylalanine metabolism	1	1
82	Steroid biosynthesis	1	1
83	Butanoate metabolism	1	1
84	Peptidoglycan biosynthesis	1	1
85	Flavone and flavonol biosynthesis	1	1
86	Histidine metabolism	1	1
87	Glycosaminoglycan biosynthesis - heparan sulfate / heparin	1	1
88	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	1	1

Additional File 15 - Relevant enzymes found for differentially expressed genes in stem. In
grey, sequences higher than 3000 bp

(Continue)

Metabolism	Enzyme code	Teak sequence	Size	Seq description	% Similarity	
Selenocompound metabolism	ec:2.7.7.4 - adenylyltransferase	comp22952_c0_seq14	2193	eukaryotic peptide chain release factor GTP-binding subunit ERF3A-like	81% [Solanum tuberosum]	
		comp22952_c0_seq19	2505	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A-like isoform X1	79% [Glycine max]	
		comp23531_c0_seq7	2959	methionine--tRNA ligase-like	93% [vitis vinifera]	
	ec:6.1.1.10 - ligase	comp23531_c0_seq26	3183	methionine--tRNA ligase-like	92% [Cucumis sativus]	
		comp23531_c0_seq16	3070	methionine--tRNA synthetase-like	93% [vitis vinifera]	
		comp23531_c0_seq22	1719	methionine--tRNA synthetase-like	86% vitis vinifera]	
		comp20538_c0_seq1	1939	isopenicillin N epimerase-like	90% [Solanum lycopersicum]	
	ec:4.4.1.1 - gamma-lyase	comp23104_c0_seq3	1104	heme oxygenase 1, chloroplastic-like	85% Solanum lycopersicum]	
	Porphyrin and chlorophyll metabolism	ec:1.14.99.3 - oxygenase (biliverdin-producing)	comp23104_c0_seq3	1104	heme oxygenase 1, chloroplastic-like	92% [Citrus sinensis]
		comp15255_c0_seq1	1418	protochlorophyllide reductase, chloroplastic-like	89% [Solanum lycopersicum]	
ec:1.3.1.33 - reductase		comp24236_c0_seq1	1452	chlorophyllide a oxygenase, chloroplastic-like	98% [Citrus sinensis]	
ec:1.14.13.122 - oxygenase		comp24729_c1_seq15	1981	Hypoxanthine-guanine phosphoribosyltransferase isoform 2	80% [Theobroma cacao]	
Drug metabolism - other enzymes	ec:2.4.2.8 - phosphoribosyltrans ferase	comp24729_c1_seq8	1480	hypoxanthine-guanine phosphoribosyltransferase-like	89% [Solanum tuberosum]	
	comp19743_c0_seq12	3050	hypothetical protein POPTR_0006s26730g	89% [Populus trichocarpa]		
	ec:3.1.1.1 - ali-esterase	comp23081_c0_seq6	3578	lysosomal alpha-mannosidase-like	80% [Vitis vinifera]	
Other glycan degradation	ec:3.2.1.24 - alpha-D-mannosidase	comp23446_c0_seq76	4591	beta-galactosidase 17-like	82% [Solanum lycopersicum]	
	ec:3.2.1.23 - lactase (ambiguous)	comp22867_c0_seq18	1503	beta-galactosidase isoform 1	80% [Vitis vinifera]	
		comp5905_c0_seq2	1157	glutathione S-transferase parA-like	90% [Solanum lycopersicum]	
		ec:2.5.1.18 - transferase	comp17040_c0_seq1	744	squalene synthase 2	90% [Salvia miltiorrhiza]
Sesquiterpenoid and triterpenoid biosynthesis	ec:2.5.1.21 - synthase	comp24494_c1_seq20	1008	dihydroflavonol-4-reductase-like	87% [Solanum lycopersicum]	
	ec:1.1.1.216 - dehydrogenase (NADP+)	comp18980_c0_seq15	1459	phytoene synthase	96% [Osmanthus fragrans]	
Carotenoid biosynthesis	ec:2.5.1.32 - synthase	comp18980_c0_seq8	1794	phytoene synthase	92% [Nicotiana tabacum]	
		comp24674_c0_seq2	1656	isopentenyltransferase	79% [Solanum lyconersicum]	

Additional File 15 - Relevant enzymes found for differentially expressed genes in stem. In
grey, sequences higher than 3000 bp

(Conclusion)

Zeatin biosynthesis	ec:2.5.1.27 - dimethylallyltransfe rase	comp24941_c0_seq2	2908	Adipocyte plasma membrane-associated protein-like	81% [Solanum lycopersicum]
Drug Metabolism- Other enzymes	ec:3.1.1.1 - ali- esterase	comp19743_c0_seq12	3050	hypothetical protein POPTR_0006s26730g	87% [Populus trichocarpa]
Drug Metabolism- Other enzymes	ec:2.4.2.8 - phosphoribosyltrans ferase	comp24729_c1_seq15	1981	Hypoxanthine-guanine phosphoribosyltransferase isoform 2	80% [Theobroma cacao]

Additional File 16 - Predicted MYB domain protein sequences from *Tectona grandis*. Amino acid sequences of the four *MYB* transcription factors were obtained with EXPASY tool (<http://web.expasy.org/>). Grey shading indicates identical amino acid residues that agree with the motifs referenced by Bedon et al. (2007). MYB-CC type transfactor domain (*TgMYB1*) and R2R-MYB DNA-binding domains (MYBR2R3-DBDs) (*TgMYB2*, *TgMYB3*, *TgMYB4*) are indicated. bHLH motif ([DE]L × 2 [RK] × 3L × 6L × 3R) is indicated in *TgMYB3*

>TgMYB1

Motif of the MYB-CC Type transfactor

MKVLYLYSGMQITEALKLQMEVQKRLEHQLEVQRQLQLRIDAQGKYLKKIIEEQHLSGVLSEMPGSGVSVSGTDDICPDSSNKTDPATPATSEPPFLDKPGKEHAPAKSLSVDESHSSHHEPQTPDSDCRVAPSVVSPNERPEKKQRGNVVTCTKSEMLNNSILESSFSPYHLPHSIFLTSEHFDHSSVGSNQLERVSGGNP

>TgMYB2

R2 region (DNA-binding domain)

WT.eEDE.L...V...G..W RcgKSCRLRW.NyLrp...RG

MLAGLRFSFSFFLFNFRELLLNTFSHAIYSYLSYCLCRRTSGPRFCSPRQWTAEEDETLRMAVQCFFGRKWKKIAECLNDRTVLQCLLRWKRVLHPDLVKGFWWSKEEDGVLIELVNKYGLKRWSTIASNLPGRR

R3 region (DNA-binding domain)

t...EE...Li.Lh...LGNKWs PGRTDN.IKN.WN

GMQCQARWYNHLKPNIEKGAWTEAEELALIRAHQSYGNKWAELTKFFPPGRDENAIKTHWNSSVEEKLDMYLASGLLPKFQGLSLLSCPSHPAASSSSKAQQSSADNSVVKGGIEVEEAFECSSQGLNIASSDAWTLQK

>TgMYB3

R2 region (DNA-binding domain)

WT.eEDE.L...V...G..W RcgKSCRLRW.NyLrp...RG

MKEKQRPSKRELNRGHWTAEEDRKLAQVDIHGAQWTTIAAKAGLARCAKSCRLRWNYLRPNIKRGWISDQERDLIIIRLHKLIGNRWSLIAGRIPGRTDNEIKNYWNTHLSKKILDKGVLVAGISTKDMGSKSDQQTVEEKTQSVTSSGAEDSKAKVDDDDADFFDFSNESSPTLEWVTKFLEFSNS

R3 region (DNA-binding domain)

DL..R...L.....L...R PGRTDN.IKN.WN

bHLH protein binding motif

>TgMYB4

R2 region (DNA-binding domain)

WT.eEDE.L...V...G..W RcgKSCRLRW.NyLrp...RG

MSVTSESNEKMPKNCIDSPAADDANSGRNVGGNDRLLKKGFWTSVEDAILVEYVTRKHGEGNWNNAVQRHSGLARCGKSCRLRWANHLRPDLKKGARSPPEEYLIIEELHAKMGNKWARMMAAEIPGRTDNEIKNYWNTRIKRRQRAGLPVYPPDICLOASNENQKGNISTFSCGDPHYLDFMEVNNFEIIPAVEFNLEVDKQVYPPAFLDIPGSSLLPQGFHSSYPDKSFISTHPSRRLRGSEPLLHGVSATMSNTIPGGSQYRNVSQVNAQSFYSSAYYHNLTFDHASSSVLSGSHADLNGNPSSSEPTWAMKLELPQLTQMGNWGSPSPFLPPLESVDTLIQTPPTHTLSCHLSPQNSGLLDAVLHESETIKNSRDSSHWQSSHASSMAVNVMDASSQVIHETGWESHGELTSPLGHSSLFSEGTPTSGDSFDEPESVEAIPGFRVKEEATFRGSMQSDNKVETTNQMFSPRDLALLF

Additional File 17 - Increment core sampling in teak. A) Use of pressler core barrel at Diameter Breast High (DBH). B) Core sample containing “s” (sapwood) and “h” (heartwood), which subsequently was placed on aluminum and transported in liquid nitrogen for the RNA extraction

Additional File 18 - Primers for quantitative real time PCR

Gene	Sequence	Primer size	Amplicon size
<i>TgMYB1</i>	5' GCTACAGTTGCGGATAGATG 3'	20 bp	145 bp
	5' ATTACTGGAGTCAGGGCAAATG 3'	22 bp	
<i>TgMYB2</i>	5' TCCAAAATTCCAAGGTCTGTCT 3'	22 bp	128 bp
	5' AAGCCTCCTCCACTTCTATTCC 3'	22 bp	
<i>TgMYB3</i>	5' CGGAAACAGATGGTCACTGATA 3'	22 bp	239 bp
	5' CAGCATCATCATCATCAACCTT 3'	22 bp	
<i>TgMYB4</i>	5' GGATCAGAACCTTTGTTACATGG 3'	23 bp	175 bp
	5' TGCCAGAAAGTACACTTGAGGA 3'	22 bp	

Additional File 19 - Melting curves and efficiencies of primers for quantitative real time PCR

Additional File 20 - Teak *MYB* transcription factor sequences obtained from RNA-seq in *Tectona grandis*

>TgCAD1
 ATGAAAAGTTCTTTATCTTTATAGCGGAATGCAAATAACTGAAGCGCTCAAGCTGCAGATGGAGGTTCAAAGCGATTGCATGAGCAATTGGAGG
 TGCAAAGACAGCTACAGTTGCGGATAGATGCCAAGGGAAGTATTTAAAAAGATAATTGAAGAACAACAACATTTAAGTGGAGTTCTTTCAGA
 AATGCCTGGCTCAGGGGTTTCTGTATCTGGAACAGATGACATTTGCCCTGACTCCAGTAATAAACTGACCCAGCAACCCCTGCTGCAACATCA
 GAGCCACCTTTTCTAGACAAGCCTGGCAAAGAACATGCTCCAGCCAAGAGTCTTTCTGTTGATGAATCCCACTCCTCACACCATGAGCCACAAA
 CCCCTGATTCTGATTGTCGTGTGGCTCCATCAGTTGTGAGCCAAATGAGAGACCAGAGAAAAAGCAGCGTGGAACAATGTTGTCACATGCAC
 TAAATCAGAAATGGTCTGAACAACCTCAATACTGGAGTCAAGCTTTAGTCTCTTACCATCTGCCGCATTCAATTTTCTTGACAAGCGAGCAC
 TTTGATCATTCATCTGTTGGCAGCGAAAAATCAGTTAGAAAAGAGTCTCTGGTGGCAATCCGTAA

>TgCAD2
 ATGCTGGCTGGCCTTAGATTTTCATTTTCTTTTCTTTTAAATTTAGAGAACTTTTACTCAACACATTTAGTCATGCTATATACTCTTACT
 ACCTTTCGTATTGTTTATGCAGGAGGACGAGTGGCCCAAGATTTTGTTCCTTAGACAATGGACTGCAGAAGAGGATGAGACATTGAGAATGGC
 TGTTCATGCTTCGAAGGGAGAAAAATGAAAAAGATAGCGGAGTGTCTCAATGATCGAACAGTTCTTCAGTGCCTGCTTAGTGGAAGAGAGTT
 CTTTCATCCGATCTGTGAAAGGGCCATGCTCGAAAGAGGAGGATGGAGTATTAATTGAATGGTCAACAAATATGGTCTAAAAAGGTGGTCTA
 CCATAGCATCAAATCTTCTCGACGCGAGGGGAATGCAATGCCAAGCAAGGTGGTACAATCATCTTAAGCCCAACATAGAAAAAGGAGCTTGGAC
 AGAGGCTGAGGAATTGGCTTTGATCCGCGCCCATCAGAGTTATGGAAACAAATGGGCAGAGTTAACTAAGTCTTCCCTGGGAGAGATGAGAAC
 GCCATTAAAAACCACTGGAATAGCTCCGTTGAGGAGAAATGGACATGTATTTGGCATCAGGATTACTTCCAAAATTCGAAGGTCTGTCTCTC
 TGAGCTGCCCTAGTCACTCCCTGCAGCTTCTCTTCTTCCAAGGCACAGCAAAGTAGTGGCGATAATAGTGTTGTTAAAGGTGGAATAGAAGTGA
 GGAGGCTTTTGTAGTGCAGTCAAGGTTTGAACATTGCCAGCTCTGATGCTTGGACACTTCAGAAATGA

>TgCAD3
 ATGAAAAGAGAAGCAGCGGCCATCAAAGAGGGAACCAACCGAGGGCCATGGACGCGGAGGAGGATCGGAAACTAGCCAAAGCCGTCGACATCC
 ACGGCGCTAAGCAGTGGACCACCATTGCTGCAAAAGCAGGGCTAGCGCGTTGCGCCAAGAGTTGCAGACTAAGATGGATGAATTATCTGAGGCC
 GAACATCAAGAGAGGCAATATATCTGATCAAGAAGAGGACTTGATCATCCGGCTCCATAAACTCCTCGGAAACAGATGGTCACTGATAGCAGGA
 AGATTGCCGGGTGCAACAGACAATGAGATCAAGAACTACTGGAATATCATTTGAGCAAGAAGATATTGGACAAAGGGGTATTAGTTGCAGGAA
 TTTTCGACGAAGACATGGGCTCCAAAAGTGATCAGCAAACTGTAGAGAGAGACACAAAGTGTTACTAGCAGTGGTGCAGAGGATTCAAAAGC
 AAAGGTTGATGATGATGATGCTGATTTCTTTGATTCTCCAATGAGAGCCCTTCACTTTGGAGTGGGTACCAAAATTTCTTGAATTTAGTAAT
 AGTTGA

>TgCAD4
 ATGAGTGTGACAAGTGAAAGCAATGAAAAGATGATGCCTAAGAATTGCATAGACTCACCAGCTGCAGACGATGCTAACAGTGAAGAAATGTTG
 GAGGGAACGATCGACTGAAAAAGGGTCTTGGACTTCTGTGGAAGATGCAATTTTGTAGTTGAATATGTTACCAAAACACGGAGAGGGGAACCTGGAA
 TGCTGTTTCAGAGACACTCGGGGCTCGCCCGTTGTGGCAAAAGTTGTCGTTTGGAGTGGGCAAAATCACCTGAGACCTGATCTAAAGAAAGGTGCA
 TTTAGTCCAGAGGAGATCTTTATCATTTGAACCTCATGCAAGATGGGAAATAAATGGGCTCGAATGGCTGCTGAGTTACCTGGCCGCAACAG
 ATAATGAGATAAAAACTACTGGAACACTAGAATCAAGAGAAGACAACGGGCGGGCTTACCAGTCTATCCACCTGATATCTGTTTACAAGCATC
 AAATGAGAACCACAAAAAGGCAATATAAGCACTTTCTCTTGCGGGGATCCACATTTATCTAGACTTCATGCCAGTTAACAACCTTTGAGATTCCA
 GCTGTGGAGTTCAAAAACTTGAAGTGGATAAGCAGGTATACCCACCAGCATTTCTTGATATCCCTGGTAGTAGCTTGCTGCCACAAGGTTTTTC
 ACTCTTCTTACCCAGACAAGTCTTTTATCTCAACAACCTATCCATCCAGGCGCCTTCGAGGATCAGAACCCTTGTGATGTTGTAAGTGCCAC
 AATGAGCAACACTATTCCGGGAGGAAGTCAATATCGAAATGTTAGTTATGTGCAGAAATGCTCAATCTTTTATATACTCTTCTGCATATTATCAT
 AATTTAACTTTTGATCATGATCATCTCTCAAGTGTACTTTCTGGCAGCCATGCTGATTTAAATGGCAATCCTTCTTCTTCAGAGCCCACTTGGG
 CAATGAAGTTGGAGTCCCTTCACTCCAACTCAAATGGGCAATTTGGGGCTCACCTCCTTCCCATTGCCTCCCCCTTGAATCTGTTGATACCTT
 GATCCAAACCCCTCCAACCTGAACACACTCTATCATGTCACTTTTCAACCCAAAACAGTGGCCTATTGGATGCAGTACTGCATGAGTCAGAAACC
 ATAAAAAATTCAGGGACAGCTCTCACTGGCAAGTTCACATGCTTCCAGTATGGCCGTGAATGTGATGGATGCTTCATCTCAAGTTATCCATG
 AGACGGGATGGGAATCACATGGGGAACCTCCCTTTGGGTCACTTCTTCTTTGTTGAGTGAAGGCACCCCTACCAGTGGGGATTCAATTTGA
 TGAACCCGAATCTGTAGAGGCAATACCAGATTAGAGTTAAAGAAGAAGCAACCTTCCGGGGTTCAATGCAATCCGACAACAAGGTTGAGACG
 ACAACACAGATGTTTCAGCAGGCCAGATTTGTTGCTTGCTTACTGTTCTAA

References

- ALCÂNTARA, B.K.; VEASEY, E.A. Genetic diversity of teak (*Tectona grandis* L. f.) from different provenances using microsatellite markers. **Revista Árvore**, Viçosa, v. 37, n. 4, p. 747–758, 2013.
- ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biology**, London, v. 11, n. 10, p. R106, 2010.
- BAI, X.; RIVERA-VEGA, L.; MAMIDALA, P.; BONELLO, P.; HERMS, D.A.; MITTAPALLI, O. Transcriptomic signatures of ash (*Fraxinus spp.*) phloem. **PloS One**, San Francisco, v. 6, n. 1, p. e16368, 2011.

- BALOGUN, A.O.; LASODE, O.A; MCDONALD, AG. Devolatilisation kinetics and pyrolytic analyses of *Tectona grandis* (teak). **BioresourceTechnology**, Essex, v. 156, p. 57–62, 2014.
- BAO, H.; LI, E.; MANSFIELD, S.D.; CRONK, Q.C.B.; EL-KASSABY, Y.A; DOUGLAS, C.J. The developing xylem transcriptome and genome-wide analysis of alternative splicing in *Populus trichocarpa* (black cottonwood) populations. **BMC Genomics**, London, v. 14, n. 1, p. 359, 2013.
- BEDON, F.; GRIMA-PETTENATI, J.; MACKAY, J. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). **BMC Plant Biology**, London, v. 7, p. 17, 2007.
- BELL, C.D.; SOLTIS, D.E.; SOLTIS, P.S. The age and diversification of the angiosperms revisited. **American Journal of Botany**, Columbus, v. 97, n. 8, p. 1296–303, 2010.
- BESSEY, C.E. The phylogenetic taxonomy of flowering plants. **Annals of the Missouri Botanical Garden**, Saint Louis, v. 2, n. 1, p. 109–164, 1915.
- BHARGAVA, A.; MANSFIELD, S.D.; HALL, H.C.; DOUGLAS, C.J.; ELLIS, B.E. MYB75 functions in regulation of secondary cell wall formation in the Arabidopsis inflorescence stem. **Plant Physiology**, Bethesda, v. 154, n. 3, p. 1428–1438, 2010.
- BHAT, K.M.; INDIRA, E.P. **Effect of faster growth on timber quality of teak**. Thrissur: Kerala Forest Research Institute, 1997. 60 p. (KFRI Research Report, 132).
- BHAT, K.M.; NAIR, K.K.N.; BHAT, K.V; MURALIDHARAN, E.M.; SHARMA, J.K. Quality timber products of teak from sustainable forest management. In: INTERNATIONAL CONFERENCE ON QUALITY TIMBER PRODUCTS OF TEAK FROM SUSTAINABLE FOREST MANAGEMENT PEECHI, 2003, Peechi. **Proceedings...** Peechi: Kerala Forest Research Institute, 2005. p. 669.
- BHAT, K.M.; PRIYA, P.B.; RUGMINI, P. Characterisation of juvenile wood in teak. **Wood Science and Technology**, New York, v. 34, n. 6, p. 517–532, 2001.
- BLANKENBERG, D.; GORDON, A.; KUSTER, G. von; CORAOR, N.; TAYLOR, J.; NEKRUTENKO, A. Manipulation of FASTQ data with Galaxy. **Bioinformatics**, Oxford, v. 26, n. 14, p. 1783–5, 2010.
- BOMAL, C.; BEDON, F.; CARON, S.; MANSFIELD, S.D.; LEVASSEUR, C.; COOKE, J.E.K.; BLAIS, S.; TREMBLAY, L.; MORENCY, M.-J.; PAVY, N.; GRIMA-PETTENATI, J.; SÉGUIN, A.; MACKAY, J. Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. **Journal of Experimental Botany**, Oxford, v. 59, n. 14, p. 3925–3939, 2008.
- CHAFFEY, N. Why is there so little research into the cell biology of the secondary vascular system of trees? **New Phytologist**, Cambridge, v. 153, n. 2, p. 213–223, 2002.

CHASE, M.W.; FAY, M.F.; REVEAL, J.L.; SOLTIS, D.E.; SOLTIS, P.S.; PETER, F.; ANDERBERG, A.A.; MOORE, M.J.; OLMSTEAD, R.G.; RUDALL, P.J.; KENNETH, J. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. **Botanical Journal of the Linnean Society**, London, v. 161, p. 105–121, 2009.

COMPEAU, P.E.C.; PEVZNER, P.A.; TESLER, G. How to apply de Bruijn graphs to genome assembly. **Nature Biotechnology**, New York, v. 29, n. 11, p. 987–991, 2011.

CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; TALÓN, M.; ROBLES, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, Oxford, v. 21, n. 18, p. 3674–3676, 2005.

DEEPAK, M.S.; SINHA, S.K.; RAO, R.V. Tree-ring analysis of teak (*Tectona grandis* L. f.) from Western Ghats of India as a tool to determine drought years. **Emirates Journal of Food and Agriculture**, Al Ain, v. 22, n. 5, p. 388–397, 2010.

DHARMAWARDHANA, P.; BRUNNER, A.M.; STRAUSS, S.H. Genome-wide transcriptome analysis of the transition from primary to secondary stem development in *Populus trichocarpa*. **BMC Genomics**, London, v. 11, p. 150, 2010.

FOFANA, I.J.; LIDAH, Y.J.; DIARRASSOUBA, N.; N'GUETTA, S.P.A.; SANGARE, A.; VERHAEGEN, D. Genetic structure and conservation of Teak (*Tectona grandis*) plantations in Côte d'Ivoire, revealed by site specific recombinase (SSR). **Tropical Conservation Science**, México, v. 1, n. 3, p. 279–292, 2008.

FOFANA, I.J.; OFORI, D.; POITEL, M.; VERHAEGEN, D. Diversity and genetic structure of teak (*Tectona grandis* L.f) in its natural range using DNA microsatellite markers. **New Forests**, Dordrecht, v. 37, n. 2, p. 175–195, 2009.

FORNALÉ, S.; SHI, X.; CHAI, C.; ENCINA, A.; IRAR, S.; CAPELLADES, M.; FUGUET, E.; TORRES, J.-L.; ROVIRA, P.; PUIGDOMÈNECH, P.; RIGAU, J.; GROTEWOLD, E.; GRAY, J.; CAPARRÓS-RUIZ, D. ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. **The Plant Journal: for Cell and Molecular Biology**, Oxford, v. 64, n. 4, p. 633–644, 2010.

GALEANO, E.; VASCONCELOS, T.S.; RAMIRO, D.A.; MARTIN, V.D.F. de; CARRER, H. Identification and validation of quantitative real-time reverse transcription PCR reference genes for gene expression analysis in teak (*Tectona grandis* L.f.). **BMC Research Notes**, London, v. 7, p. 464, 2014.

GANGOPADHYAY, G.; GANGOPADHYAY, S.B.; PODDAR, R.; GUPTA, S.; MUKHERJEE, K.K. Micropropagation TEAK genetic fidelity.pdf. **Biologia Plantarum**, Praha, v. 46, n. 3, p. 459–461, 2003.

GARBER, M.; GRABHERR, M.G.; GUTTMAN, M.; TRAPNELL, C. Computational methods for transcriptome annotation and quantification using RNA-seq. **Nature Methods**, New York, v. 8, n. 6, p. 469–77, 2011.

GILL, B.; YEDI, Y.; BIR, S. Cytopalynological studies in woody members of family Verbenaceae from north-west and central India. **Journal of the Indian Botanical Society**, Madras, v. 62, p. 235–244, 1983.

GOH, D.K.S.; MONTEUUIS, O. Rationale for developing intensive teak clonal plantations , with special reference to Sabah. **Bois et Forêts des Tropiques**, Nogent-sur-Marne, v. 285, n. 3, p. 5–15, 2005.

GOH, D.K.S.; CHAIX, G.; BAILLÈRES, H.; MONTEUUIS, O. Mass production and quality control of teak clones for tropical plantations: the Yayasan Sabah Group and CIRAD Joint Project as a case study. **Bois et Forêts des Tropiques**, Nogent-sur-Marne, v. 293, n. 3, p. 65–77, 2007.

GOICOECHEA, M.; LACOMBE, E.; LEGAY, S.; MIHALJEVIC, S.; RECH, P.; JAUNEAU, A.; LAPIERRE, C.; POLLET, B.; VERHAEGEN, D.; CHAUBET-GIGOT, N.; GRIMA-PETTENATI, J. EgMYB2, a new transcriptional activator from Eucalyptus xylem, regulates secondary cell wall formation and lignin biosynthesis. **The Plant Journal: for Cell and Molecular Biology**, Oxford, v. 43, n. 4, p. 553–567, 2005.

GONG, W.; SHEN, Y.; MA, L.; PAN, Y.; DU, Y.; WANG, D.; YANG, J.; HU, L.; LIU, X.; DONG, C.; MA, L.; CHEN, Y.; YANG, X.; GAO, Y.; ZHU, D.; TAN, X.; MU, J.; ZHANG, D.; LIU, Y.; DINESH-KUMAR, S.; LI, Y.; WANG, X.; GU, H.; QU, L.; BAI, S.; LU, Y.; LI, Y.; ZHAO, J.; ZUO, J.; HUANG, H.; DENG, X.W.; ZHU, Y. Genome-Wide ORFeome Cloning and Analysis of Arabidopsis Transcription Factor Genes. **Plant Physiology**, Bethesda, v. 135, n. 1, p. 773–782, 2004.

GORDO, S.M.C.; PINHEIRO, D.G.; MOREIRA, E.C.O.; RODRIGUES, S.M.; POLTRONIERI, M.C.; LEMOS, O.F. de; SILVA, I.T. da; RAMOS, R.T.J.; SILVA, A.; SCHNEIDER, H.; SILVA, W.A.; SAMPAIO, I.; DARNET, S. High-throughput sequencing of black pepper root transcriptome. **BMC Plant Biology**, London, v. 12, n. 1, p. 168, 2012.

GRABHERR, M.G.; HAAS, B.J.; YASSOUR, M.; LEVIN, J.Z.; THOMPSON, D.A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACHOEN, N.; GNIRKE, A.; RHIND, N.; PALMA, F. di; BIRREN, B.W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature Biotechnology**, New York, v. 29, n. 7, p. 644–652, 2011.

HAAS, B.J.; PAPANICOLAOU, A.; YASSOUR, M.; GRABHERR, M.; BLOOD, P.D.; BOWDEN, J.; COUGER, M.B.; ECCLES, D.; LI, B.; LIEBER, M.; MACMANES, M.D.; OTT, M.; ORVIS, J.; POCHET, N.; STROZZI, F.; WEEKS, N.; WESTERMAN, R.; WILLIAM, T.; DEWEY, C.N.; HENSCHER, R.; LEDUC, R.D.; FRIEDMAN, N.; REGEV, A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols**, New York, v. 8, n. 8, p. 1494–512, 2013.

HUANG, G.; LI, T.; LI, X.; TAN, D.; JIANG, Z.; WEI, Y.; LI, J.; WANG, A. Comparative transcriptome analysis of climacteric fruit of Chinese pear (*Pyrus ussuriensis*) reveals new insights into fruit ripening. **PloS One**, San Francisco, v. 9, n. 9, p. e107562, 2014.

JAIN, A.; ANSARI, S.A. Quantification by allometric equations of carbon sequestered by *Tectona grandis* in different agroforestry systems. **Journal of Forestry Research**, Pekin, v. 24, n. 4, p. 699–702, 2013.

KAKUMANU, A.; AMBAVARAM, M.M.R.; KLUMAS, C.; KRISHNAN, A.; BATLANG, U.; MYERS, E.; GRENE, R.; PEREIRA, A. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. **Plant Physiology**, Bethesda, v. 160, n. 2, p. 846–867, 2012.

KARPINSKA, B.; KARLSSON, M.; SRIVASTAVA, M.; STENBERG, A.; SCHRADER, J.; STERKY, F.; BHALERAO, R.; WINGSLE, G. MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. **Plant Molecular Biology**, Dordrecht, v. 56, n. 2, p. 255–270, 2004.

KEOGH, R.M. **The future of teak and the high-grade tropical hardwood sector**. Rome: FAO, 2009. 47 p.

KIM, D.; PERTEA, G.; TRAPNELL, C.; PIMENTEL, H.; KELLEY, R.; SALZBERG, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biology**, London, v. 14, n. 4, p. R36, 2013.

KIM, W.-C.; KO, J.-H.; KIM, J.-Y.; KIM, J.-M.; BAE, H.-J.; HAN, K.-H. MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis*. **The Plant Journal: for Cell and Molecular Biology**, Oxford, v. 73, n. 4, p. 26–36, 2012.

KO, J.-H.; KIM, W.-C.; HAN, K.-H. Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. **The Plant Journal: for Cell and Molecular Biology**, Oxford, v. 60, n. 4, p. 649–665, 2009.

KOLLERT, W.; CHERUBINI, L. **Teak resources and market assessment 2010 (*Tectona grandis* Linn. F.)**. Rome: FAO, 2012. 42 p.

KVAM, V.M.; LIU, P.; SI, Y. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. **American Journal of Botany**, Columbus, v. 99, n. 2, p. 248–256, 2012.

LACRET, R.; VARELA, R.M.; MOLINILLO, J.M.G.; NOGUEIRAS, C.; MACÍAS, F.A. Anthrathectone and naphthothectone, two quinones from bioactive extracts of *Tectona grandis*. **Journal of Chemical Ecology**, New York, v. 37, p. 1341–1348, 2011.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, New York, v. 9, n. 4, p. 357–359, 2012.

LEGAY, S.; LACOMBE, E.; GOICOECHEA, M.; BRIÈRE, C.; SÉGUIN, A.; MACKAY, J.; GRIMA-PETTENATI, J. Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. **Plant Science**, Limerick, v. 173, n. 5, p. 542–549, 2007.

LI, M.-Y.; TAN, H.-W.; WANG, F.; JIANG, Q.; XU, Z.-S.; TIAN, C.; XIONG, A.-S. De Novo Transcriptome sequence assembly and identification of AP2/ERF transcription factor

related to abiotic stress in parsley (*Petroselinum crispum*). **PloS One**, San Francisco, v. 9, n. 9, p. e108977, 2014.

LIU, L.; FILKOV, V.; GROOVER, A. Modeling transcriptional networks regulating secondary growth and wood formation in forest trees. **Physiologia Plantarum**, Copenhagen, v. 151, n. 2, p. 156–163, 2014.

LYNGDOH, N.; JOSHI, G.; RAVIKANTH, G.; VASUDEVA, R.; SHAANKER, R.U. Changes in genetic diversity parameters in unimproved and improved populations of teak (*Tectona grandis* L.f.) in Karnataka state, India. **Journal of Genetics**, Bangalore, v. 92, n. 1, p. 141–145, 2013.

MA, Q.-H.; WANG, C.; ZHU, H.-H. TaMYB4 cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes. **Biochimie**, Paris, v. 93, n. 7, p. 1179–1186, 2011.

MARTIN, J.A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews. Genetics**, London, v. 12, n. 10, p. 671–682, 2011.

MATUS, J.T.; AQUEA, F.; ARCE-JOHNSON, P. Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. **BMC Plant Biology**, London, v. 8, p. 83, 2008.

MCCARTHY, R.L.; ZHONG, R.; FOWLER, S.; LYSKOWSKI, D.; PIYASENA, H.; CARLETON, K.; SPICER, C.; YE, Z.-H. The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. **Plant & Cell Physiology**, Kyoto, v. 51, n. 6, p. 1084–1090, 2010.

MILLAR, A.A.; GUBLER, F. The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. **The Plant Cell**, Rockville, v. 17, n. 3, p. 705–721, 2005.

MINN, Y.; PRINZ, K.; FINKELDEY, R. Genetic variation of teak (*Tectona grandis* Linn. f.) in Myanmar revealed by microsatellites. **Tree Genetics & Genomes**, Davis, v. 10, n. 5, p. 1435–1449, 2014.

MIZRACHI, E.; HEFER, C.A.; RANIK, M.; JOUBERT, F.; MYBURG, A.A. De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. **BMC Genomics**, London, v. 11, n. 1, p. 681, 2010.

MUTZ, K.-O.; HEILKENBRINKER, A.; LÖNNE, M.; WALTER, J.-G.; STAHL, F. Transcriptome analysis using next-generation sequencing. **Current Opinion in Biotechnology**, London, v. 24, n. 1, p. 22–30, 2013.

OHRI, D.; KUMAR, A. Nuclear DNA amounts in some tropical hardwoods. **Caryologia**, Firenze, v. 39, n. 3/4, p. 303–307, 1986.

PATZLAFF, A.; MCINNIS, S.; COURTENAY, A.; SURMAN, C.; NEWMAN, L. J.; SMITH, C.; BEVAN, M.W.; MANSFIELD, S.; WHETTEN, R.W.; SEDEROFF, R.R.; CAMPBELL, M.M. Characterisation of a pine MYB that regulates lignification. **The Plant Journal**, London, v. 36, n. 6, p. 743–754, 2003.

PAVY, N.; BOYLE, B.; NELSON, C.; PAULE, C.; GIGUÈRE, I.; CARON, S.; PARSONS, L. S.; DALLAIRE, N.; BEDON, F.; BÉRUBÉ, H.; COOKE, J.; MACKAY, J. Identification of conserved core xylem gene sets: conifer cDNA microarray development, transcript profiling and computational analyses. **The New Phytologist**, Cambridge, v. 180, n. 4, p. 766–786, 2008.

PRASSINOS, C.; KO, J.-H.; YANG, J.; HAN, K.-H. Transcriptome profiling of vertical stem segments provides insights into the genetic regulation of secondary growth in hybrid aspen trees. **Plant & Cell Physiology**, Kyoto, v. 46, n. 8, p. 1213–1225, 2005.

QUIALA, E.; CAÑAL, M.J.; RODRÍGUEZ, R.; YAGÜE, N.; CHÁVEZ, M.; BARBÓN, R.; VALLEDOR, L. Proteomic profiling of *Tectona grandis* L. leaf. **Proteomics**, Weingem, v. 12, n. 7, p. 1039–1044, 2012.

QIU, Q.; MA, T.; HU, Q.; LIU, B.; WU, Y.; ZHOU, H.; WANG, Q.; WANG, J.; LIU, J. Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*. **Tree Physiology**, Oxford, v. 31, n. 4, p. 452–461, 2011.

RAO, G.; SUI, J.; ZENG, Y.; HE, C.; DUAN, A.; ZHANG, J. De Novo transcriptome and small RNA analysis of two Chinese Willow cultivars reveals stress response genes in *Salix matsudana*. **PloS One**, San Francisco, v. 9, n. 10, p. e109122, 2014.

ROBERTS, A.; PIMENTEL, H.; TRAPNELL, C.; PACHTER, L. Identification of novel transcripts in annotated genomes using RNA-Seq. **Bioinformatics**, Oxford, v. 27, n. 17, p. 2325–2329, 2011.

ROGERS, L.A.; CAMPBELL, M.M. The genetic control of lignin deposition during plant growth and development. **The New Phytologist**, Cambridge v. 164, n. 1, p. 17–30, 2004.

RUBIO, V.; LINHARES, F.; SOLANO, R.; MARTÍN, A.C.; IGLESIAS, J.; LEYVA, A.; PAZ-ARES, J. A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. **Genes & Development**, Cold Spring Harbor, v. 15, n. 16, p. 2122–2133, 2001.

SALZMAN, R.A.; FUJITA, T.; HASEGAWA, P.M. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. **Plant Molecular Biology Reporter**, Athens, v. 17, n. 765, p. 11–17, 1999.

SCHLIESKY, S.; GOWIK, U.; WEBER, A.P.M.; BRÄUTIGAM, A. RNA-Seq assembly - are we there yet? **Frontiers in Plant Science**, Lausanne, v. 3, p. 220, Sept. 2012.

SCHRADER, J.; NILSSON, J.; MELLEROWICZ, E.; BERGLUND, A.; NILSSON, P.; HERTZBERG, M. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. **The Plant Cell**, Rockville, v. 16, p. 2278–2292, Sept. 2004.

- SHRESTHA, M.K.; VOLKAERT, H.; STRAETEN, D. van der. Assessment of genetic diversity in *Tectona grandis* using amplified fragment length polymorphism markers. **Canadian Journal of Forest Research**, Ottawa, v. 35, n. 4, p. 1017–1022, 2005.
- SHUKLA, S.R.; VISWANATH, S. Comparative study on growth, wood quality and financial returns of teak (*Tectona grandis* L.f.) managed under three different agroforestry practices. **Agroforestry Systems**, Dordrecht, v. 88, n. 2, p. 331–341, 2014.
- SREEKANTH, P.M.; BALASUNDARAN, M.; NAZEEM, P.A.; SUMA, T.B. Genetic diversity of nine natural *Tectona grandis* L.f. populations of the Western Ghats in Southern India. **Conservation Genetics**, Dordrecht, v. 13, n. 5, p. 1409–1419, 2012.
- SUN, Q.; ZHOU, G.; CAI, Y.; FAN, Y.; ZHU, X.; LIU, Y.; HE, X.; SHEN, J.; JIANG, H.; HU, D.; PAN, Z.; XIANG, L.; HE, G.; DONG, D.; YANG, J. Transcriptome analysis of stem development in the tumorous stem mustard *Brassica juncea* var. *tumida* Tsen et Lee by RNA sequencing. **BMC plant biology**, London, v. 12, n. 1, p. 53, 2012.
- TANG, X.; XIAO, Y.; LV, T.; WANG, F.; ZHU, Q.; ZHENG, T.; YANG, J. High-throughput sequencing and de novo assembly of the *Isatis indigotica* transcriptome. **PloS One**, San Francisco, v. 9, n. 9, p. e102963, 2014.
- TIWARI, A.; KUMAR, P.; CHAWHAAN, P.H.; SINGH, S.; ANSARI, S.A. Carbonic anhydrase in *Tectona grandis* : kinetics, stability, isozyme analysis and relationship with photosynthesis. **Tree Physiology**, Oxford, v. 26, p. 1067–1073, 2006.
- TRAPNELL, C.; ROBERTS, A.; GOFF, L.; PERTEA, G.; KIM, D.; KELLEY, D.R.; PIMENTEL, H.; SALZBERG, S.L.; RINN, J.L.; PACHTER, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, London, v. 7, n. 3, p. 562–578, 2012.
- UENO, S.; KLOPP, C.; LEPLÉ, J. C.; DERORY, J.; NOIROT, C.; LÉGER, V.; PRINCE, E.; KREMER, A.; PLOMION, C.; PROVOST, G. le. Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. **BMC Genomics**, London, v. 14, p. 236, 2013.
- VERHAEGEN, D.; OFORI, D.; FOFANA, I.; POITEL, M.; VAILLANT, A. Development and characterization of microsatellite markers in *Tectona grandis* (Linn. f). **Molecular Ecology Notes**, Oxford, v. 5, n. 4, p. 945–947, 2005.
- VILLAR, E.; KLOPP, C.; NOIROT, C.; NOVAES, E.; KIRST, M.; PLOMION, C.; GION, J.-M. RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. **BMC Genomics**, London, v. 12, n. 1, p. 538, 2011.
- WANG, L.; FENG, Z.; WANG, X.; WANG, X.; ZHANG, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. **Bioinformatics**, Oxford, v. 26, n. 1, p. 136–138, 2010.

WEI, K.; WANG, L.-Y.; WU, L.-Y.; ZHANG, C.-C.; LI, H.-L.; TAN, L.-Q.; CAO, H.-L.; CHENG, H. Transcriptome analysis of indole-3-butyric acid-induced adventitious root formation in nodal cuttings of *Camellia sinensis* (L.). **PloS One**, San Francisco, v. 9, n. 9, p. e107201, 2014.

WU, J.; WANG, L.; LI, L.; WANG, S. De novo assembly of the common bean transcriptome using short reads for the discovery of drought-responsive genes. **PloS One**, San Francisco, v. 9, n. 10, p. e109262, 2014.

YANG, S.S.; TU, Z.J.; CHEUNG, F.; XU, W.W.; LAMB, J.F.S.; JUNG, H.-J.G.; VANCE, C.P.; GRONWALD, J.W. Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. **BMC Genomics**, London, v. 12, n. 1, p. 199, 2011.

YANG, S.-H.; ZYL, L. VAN; NO, E.-G.; LOOPSTRA, C.A. Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). **Plant Science**, Limerick, v. 166, n. 5, p. 1185–1195, 2004.

ZENONI, S.; FERRARINI, A.; GIACOMELLI, E.; XUMERLE, L.; FASOLI, M.; MALERBA, G.; BELLIN, D.; PEZZOTTI, M.; DELLEDONNE, M. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq 1 [W]. **Plant Physiology**, Bethesda, v. 152, p. 1787–1795, Apr. 2010.

ZHAO, Q.; DIXON, R.A. Transcriptional networks for lignin biosynthesis: more complex than we thought? **Trends in Plant Science**, Oxford, v. 16, n. 4, p. 227–233, 2011.

ZHONG, R.; RICHARDSON, E.A; YE, Z.-H. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. **The Plant Cell**, Rockville, v. 19, n. 9, p. 2776–2792, 2007.

ZHONG, R.; LEE, C.; ZHOU, J.; MCCARTHY, R.L.; YE, Z.-H. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. **The Plant Cell**, Rockville, v. 20, n. 10, p. 2763–2782, 2008.

5 GENERAL CONCLUSIONS

It is the first study of molecular characterization of teak genes. During the work, it was performed RNA sequencing, bioinformatics, cloning and gene expression analyses of two gene families involved in essential tree metabolisms such as *Cinnamyl alcohol dehydrogenase* (*CAD*) genes and *MYB* transcription factors. For gene expression analyses, it was found that *TgEF-1a* and *TgUBQ* are the most stable genes across different tissues while *TgACT* was judged to be unsuitable. After amplifying several *CAD* genes in teak, it was found four members belonging to the medium-chain dehydrogenase/reductase (MDR) family. TgCAD1 is structurally similar to AtCAD5 but does not show significant expression in lignified tissues. In contrast, TgCAD3 and TgCAD4 are involved in lignin biosynthesis for being clustered with lignin-related genes and presenting high expression in secondary xylem and in sapwood during teak maturation.

The transcriptome of *T. grandis* was obtained, embracing 192 million reads and more than 2,000 genes as differentially expressed with two biological replicates, including highly expressed heat-shock proteins, carbohydrate metabolic genes and four characterized *MYB* transcription factors. Of these, several transcriptomic changes in maturation of teak secondary xylem using RNAseq platform were observed; as a relevant result, high expression of *TgMYB1* and *TgMYB4* in lignified tissues of 60-year-old trees was obtained.

Understanding gene function of woody tissues and how lignin biosynthesis occurs in forest tree species are highly challenging (but essential) due to the plant size, slow growth, lack of standard tree transformation, and long generation times of this type of species. In that sense, comparative genomics appears as the closest approach to understand the function of thousands of genes in trees with a subsequent selection of target transcripts with potential applications, including marker-aided breeding and natural polymorphisms, as well as plant transformation looking for improved tree growth rates, adaptability and wood quality, whenever discrete loci affecting wood structure and composition could be recognized.

A few remarks: (i) the extensive transcriptome (generated by Illumina paired-end sequencing) and gene expression profiles analyzed in this research are considerably relevant, since no expression information about teak was available in databases (especially related to secondary growth and lignin biosynthesis) before this study, and the data obtained can be used in applied and basic science along with biotechnological approaches; (ii) transcription (gene

expression) disparity from a gradient of young to mature secondary xylem and sapwood was achieved, identifying several tissue- and developmental stage-specific genes; (iii) my results provide for the first time reference genes for quantitative real-time PCR, differentially expressed *cinnamyl alcohol dehydrogenase* genes, heat-shock proteins and *MYB* transcription factors in teak (MYB-CC and R2R3-MYB types), contributing to the understanding of the molecular mechanisms in tropical woods, incentivizing to conduct reverse genetics and plant transformation in *T. grandis*, and they will aid in understanding regulatory networks of wood formation.

6 IMPLICATIONS AND PERSPECTIVES OF THIS STUDY

These results, the first dataset of expressed sequences of the Lamiales order and *Tectona* genus, will open new perspectives for studies of diversity, ecology, breeding and genomic programs aiming to understand deeply the biology of this species.

6.1 Heat-Shock proteins

Heat-shock proteins have a crucial role in maintaining the proteins in their functional conformation when temperatures or other environmental conditions rise, preventing degradation and damage during heat stress, from late winter to early spring. In tropical zones, woody plants go through phenological cycles, which are strongly affected in the biological processes by an increase in temperature. Indeed, heat-shock proteins could aid in defending *T. grandis* against those environmental changes in the region sampled, and therefore need to be more studied.

6.2 Regulation in teak

Regulation of wood formation in tropical forest trees remains poorly understood. This transcriptomic study reported changes in transcripts through *T. grandis* maturation. Presumably, *TgMYB1*, *TgMYB2* and *TgMYB4* may also be triggered in teak secondary xylem by other transcription factors, especially *NAC* master regulators, in response to cell wall thickening, regulation of phenylpropanoid genes, environmental conditions changes and as a response to biotic and abiotic stimuli.

6.3 It is essential to understand genes involved in teak wood

Teak wood demand is growing exponentially, and it is known that the quality of the teak wood produced is the predominant commercial factor for the near future. Usually this quality relates to the quantity, color and durability of the heartwood. Consequently, it is necessary to discover genes in teak that control the secondary xylem vessels' formation, sapwood and heartwood differentiation, the volume growth and abiotic stress-related genes. It needs to be done in teak wood in order to help improve the quality, growth time and environmental adaptability. Whereas the whole teak transcriptome will be available to the scientific community in a few months, it is essential to realize that several genes from this

database could be used for functional analyses to ensure the pioneering research in this tropical forest species with high economic value.

6.4 The transcriptomes lead to discover gene families, understand lignification processes and establish transcriptional regulatory networks

The preliminary results that comprise the teak transcriptome are the first set of expressed data sequences of Lamiales order and *Tectona* genus, which opens up great prospects for genetic diversity studies and breeding programs in order to deeply understand the biology of this species. There are unique molecular networks within the secondary growth of teak that includes proteins interacting with DNA, several regulators of phenylpropanoid pathway, multiplicity of proteins related to stress, peptide transporters, metabolic genes of carbohydrates, pectin-related genes and specific teak genes such as the irinotecan protein. Likewise, the molecular mechanisms involved in regulating the formation of tropical wood in the forest trees are unknown. My study reported transcriptomic changes in the increase and decrease of gene expression during teak maturation.

The up-regulation of *TgMYB1*, *TgMYB2* and *TgMYB4* in secondary xylem could also be regulated at higher levels by other master transcription factors, particularly *NAC* genes, in response to cell wall thickening, regulation of phenylpropanoid genes, the environmental conditions in the season changes, and as a possible response to biotic and abiotic stimuli. And, although a large number of expressed genes have been identified during the radial growth of woody trees (second growth), it is not yet known how they interact to influence the woody growth. Similarly, as the formation of the timber is a developing process that includes the differentiation, elongation and secondary wall thickness of programmed cell death of the wood cells, it will be interesting to identify transcription factors in trees that are directly involved in the regulation of the secondary wall biosynthesis with the aim of regulatory networks. Some cases of regulatory networks in woody tissues have found the transcription factor *NAC* mediating several pathways of the secondary wall biosynthesis in different vascular plants, concluding that this network is conserved among such species.

6.5 The next-generation sequencing in breeding programs

Unfortunately, despite having transcriptomes available for different forest species, the establishment of molecular markers from transcriptomes is limited. In the case of teak,

thousands of transcripts in different tissues were found, with over 2,000 differentially expressed genes, including several members of different families of transcription factors, many of them related to lignin biosynthesis. Although the projects in the area of transcriptomics are currently based on the simple discovery of differentially expressed genes, they have potential applications, including breeding programs assisted by molecular markers and studies of natural polymorphisms seeking benefit in tree growth and wood quality, whenever discrete *loci* could be detected and regulate the wood structure and composition. And despite the importance of teak, efforts have been focused on the study of genetic variability among populations of this species using microsatellite markers, but none correlated with the wood characteristics and SNPs discovery from ESTs.

6.6 Wood and heartwood quality as characteristics in breeding programs

It is important to consider how teak maturation can influence the expression increase of the transcription factors *TgMYB1* and *TgMYB4* and a decrease in *TgMYB2* in mature sapwood (60 years), a tissue traditionally considered as a high heritability character. Therefore, it seems to be an important feature to be included in teak improvement programs, particularly when short rotations, such as the Brazilian ones (20 years), are the target. The quality of juvenile wood itself will be an important feature to improve, and this can be estimated at an early stage of teak plants by looking for trees with a faster youth growth rate and for more years of development, which will result in a significant reduction in the periods of crops rotation with higher percentage yield of heartwood.

Teak offers potential for wood production with optimum strength and with relatively short rotation of 21 years, and fast-growing clones can be selected for planting the species without reducing the wood density. Still, the development of teak commercial plantations around the world is encouraged by its high market price and high demand. Unfortunately, most of the teak plantations are still produced from non-selected genotypes seeds, which can result in poor stands and low-quality wood; in the traditional markets of Thailand, Singapore, China and Brazil, there is a great concern for the future supply of teak wood.

6.7 Three questions to be answered in future research projects

What is the function of the differentially expressed genes TgCAD3, TgCAD4, TgMYB1 and TgMYB4? Genetic transformation, microscopy and regulatory networks could assist in the detailed understanding of the role of these genes and their relevance.

Could the genes involved in the expression of secondary xylem between young and mature sapwood and the establishment of its regulatory and molecular interactions networks contribute to breeding programs? The standardization of markers such as EST-SNPs would provide sufficient information to assess the potential of the genes obtained in this study when selecting genotypes. Previous studies of multigenic association in trees with lignin-related features and EST-SNPs characterization have shown their viability in forest species. This application may be the first report of EST-SNPs markers in *Tectona grandis*, which could be used in conservation genetics and reproduction studies in the near future.

Given our current understanding of the of gene expression of TgCAD1 to TgCAD4 and TgMYB1 to TgMYB4 in sapwood, can this tissue along with the heartwood be used as a basis for diagnosis of quality teak seedlings? Both are key factors in the production and marketing of teakwood and will be the targets in breeding programs of this species in the near future. The growth in trees is synonymous with wood and productivity, which represent a goal for the intensive production of biomass when the final application is looking for structural wood and wood products and forestry environmental conservation.