UNIVERSITI PUTRA MALAYSIA

DE NOVO ASSEMBLY, ANNOTATION AND ANALYSIS OF TRANSCRIPTOME SEQUENCES OF CALLUS CULTURE FROM AQUILARIA MALACCENSIS LAM.

SIAH CHAI HAR

FH 2015 18

DE NOVO ASSEMBLY, ANNOTATION AND ANALYSIS OF TRANSCRIPTOME SEQUENCES OF CALLUS CULTURE FROM *AQUILARIA MALACCENSIS* LAM.

By

SIAH CHAI HAR

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

February 2015

SPECIALLY DEDICATED TO

My Beloved
Parents, Brothers & Sisters

'Except the Lord build the house, they labour in vain that build it:
Except the Lord keep the city, the watchman waketh but in vain.'
Psalm 127:1

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Master of Science

DE NOVO ASSEMBLY, ANNOTATION AND ANALYSIS OF TRANSCRIPTOME SEQUENCES OF CALLUS CULTURE FROM _AQUILARIA MALACCENSIS_ LAM.

By

SIAH CHAI HAR

February 2015

Chair: Rozi Mohamed, PhD
Faculty: Forestry

Aquilaria malaccensis is a major source of agarwood, a rare and highly priced wood product. Due to its high demand, it is endangered and listed in the Appendix II of Convention on International Trade in Endangered Species of Wild Fauna and Flora. Because of insufficient genomic and transcriptomic data available in public database for understanding the molecular basis of agarwood formation, the goal of this study was to obtain *A. malaccensis* expressed gene sequences using the transcriptome sequencing by next-generation sequencing (NGS). To obtain sufficient data from NGS sequencing, a high quality RNA sample is required. The RNA yield, purity, and integrity of six different extraction methods were compared. Conventional methods yielded RNA with good purity but the RNA integrity was poor. When using modified RNeasy Plant Mini kit, the highest yield was obtained while maintaining the integrity of RNA. This method was used to extract total RNA from callus samples and sent for transcriptome sequencing. Callus tissues were treated under stress condition by nutrient shortage of 1.1 µm 1-naphthaleneacetic acid (NAA), 2.2 µm 6-benzylaminopurine (BAP), and 15 g/L sucrose; collected during death phase when it producing brownish exudates and agarwood scent. Two cDNA libraries were constructed from mRNAs of untreated and treated callus tissues and sequenced using paired-end libraries on an Illumina HiSeq2000 platform. After filtering and trimming, a total of 200,062,275 and 166,544,202 clean reads were obtained for untreated and treated callus libraries, respectively. *De novo* assembly was carried out using SOAPdenovo-Trans and TGICL. A total of 231,594 unigenes were generated from the assembly. Assembled sequences were annotated using BLASTX against the NCBI non-redundant protein database where 41.5% unigenes showed significant alignment. The differential gene expression value between untreated and treated samples were calculated using DEGseq. A total of 14,029 genes were identified as differentially expressed. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations were reported using BLAST2GO software. Out of the 107,593 unigenes that showed significant alignment with non-redundant protein database, 96,743 unigenes were successfully annotated with at least one GO term. The

annotations were classified into the three main GO categories: biological processes (50.7%), molecular functions (24.0%) and cellular components (25.3%). A total of 144 KEGG pathways were identified from 46,076 unigenes. Enrichment analysis showed that the genes were enriched in the stress responses and sesquiterpene biosynthesis activity. This is the first comprehensive *de novo* transcriptome assembly and profiling for *A. malaccensis*. These transcriptome libraries provide valuable transcriptomic reference resource for future research.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Sarjana Sains

HIMPUNAN *DE NOVO*, ANOTASI DAN ANALISIS JUJUKAN TRANSKRIPTOM KULTUR KALUS DARI *AQUILARIA MALACCENSIS* LAM.

Oleh

SIAH CHAI HAR

Februari 2015

Pengerusi: Rozi Mohamed, PhD
Fakulti: Perhutanan

Aquilaria malaccensis merupakan sumber utama gaharu, hasil kayu yang jarang dijumpai dan sangat berharga. Berikutan dengan permintaannya yang tinggi, ia telah terancam dan disenaraikan dalam Lampiran II Konyensyen Perdagangan Antarabangsa Spesies Terancam Flora dan Fauna Liar. Oleh kerana kekurangan data genomik dan transkriptomik yang sedia ada dalam pangkalan data awam untuk memahami asas molecular pembentukan gaharu, matlamat kajian ini adalah untuk mendapatkan ekspresi gen *A. malaccensis* menggunakan penjujukan transkriptom dengan "next-generation sequencing" (NGS). Untuk mendapatkan data yang mencukupi daripada penjujukan NGS, sampel RNA yang berkualiti tinggi diperlukan. Hasil RNA, ketulenan, dan integriti dari enam kaedah pengekstrakan yang berbeza telah dibandingkan. Kaedah konvensional menghasilkan RNA dengan ketulenan bagus tetapi integriti RNA yang rendah. Semasa menggunakan RNeasy Plant Mini Kit yang telah diubahsuai, hasil yang paling tinggi telah diperolehi di samping mengekalkan integriti RNA. Kaedah ini telah digunakan untuk mengekstrak sampel RNA daripada sampel kalus dan dihantar untuk penjujukan transkriptom. Tisu kalus dirawat di bawah keadaan tekanan kekurangan nutrien 1.1 µm asid 1-naftalenaasetik (NAA), 2.2 µm 6-benzilaminopurina (BAP), dan 15 g/L sukrosa; dikumpul semasa fasa kematian apabila ia mengeluarkan eksudat keperangan dan bau gaharu. Dua perpustakaan cDNA telah dibina daripada mRNA tisu kalus tidak dirawat dan dirawat dan telah dijujuk menggunakan jujukan pasangan atas platform Illumina HiSeq2000. Selepas menapis dan merapi, sejumlah 200,062,275 dan 166,544,202 bacaan bersih telah diperolehi untuk perpustakaan kalus tidak dirawat dan dirawat, masing-masing. Perhimpunan *de novo* telah dijalankan dengan menggunakan SOAPdenovo-Trans dan TGICL. Sejumlah 231,594 unigen telah dijanakan daripada perhimpunan tersebut. Jujukan terhimpun telah dianotasi menggunakan BLASTX terhadap pangkalan data NCBI protein bukan berulangan, di mana 41.5% unigen menunjukkan penjajaran yang bermakna. Nilai perbezaan gen antara sampel kalus tidak dirawat dan dirawat telah dikira menggunakan DEGseq. Sejumlah 14,029 gen telah dikenal pasti sebagai mempunyai ekspresi berbeza. Anotasi Ontologi Gen (GO) dan Ensiklopedia Kyoto Gen

dan Genom (KEGG) telah dilaporkan menggunakan perisian BLAST2GO. Daripada 107,593 unigen yang menunjukkan penjajaran yang bermakna dengan pangkalan data protein bukan berulangan, 96,743 unigen telah berjaya dianotasikan dengan sekurang-kurangnya satu sebutan GO. Anotasi telah diklasifikasikan kepada tiga kategori GO utama: proses biologi (50.7%), fungsi molekular (24.0%) dan komponen selular (25.3%). Sejumlah 144 laluan KEGG telah dikenal pasti dari 46,076 unigen. Analisis pengkayaan menunjukkan bahawa gen telah diperkaya dalam tindak balas tekanan dan aktiviti biosintesis seskuiterpena. Ini adalah perhimpunan komprehensif *de novo* transkriptom dan pemprofilan pertama untuk *A. malaccensis*. Perpustakaan transkriptom ini menyediakan sumber rujukan transkriptomik yang bernilai untuk penyelidikan masa depan.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and encouragement given by a number of amazing people. First and foremost I would like to express my deep gratitude to my supervisor, Assoc. Prof. Dr. Rozi Mohamed for her supervision and guidance throughout the study. This appreciation also dedicated to my co-supervisor, Assoc. Prof. Dr. Parameswari Namasivayam, for her valuable feedback and research advices. I would also like to thank Prof. Charles Cannon, who has provided me an opportunity for a one-month research attachment at Xishuangbanna Tropical Botanical Garden, Chinese Academy of Science, Yunnan, China. Also a deep appreciation to Ms. Zhang Di for her bioinformatics guidance and life experience sharing during the attachment period.

I am thankful to past and present members of the Forest Biotechnology Lab, Wong Mun Theng, Lee Shiou Yih, Jong Phai Lee, Shashita a/p Jayaraman, Nurul Hazwani bt Daud and Azzarina bt Anor Basah for the advices, assistance and support that they have given me. I would also like to express my personal thanks to all my friends, especially Janice Lim Yanjun, Kwong Qi Bin and Chio Jhe Wei for their advices and encouragement throughout the study and thesis writing.

Last but not the least, I would like to thank my beloved family members for their understanding and moral support during my study.

I certify that a Thesis Examination Committee has met on 6 February 2015 to conduct the final examination of Siah Chai Har on her thesis entitled "*De Novo* Assembly, Annotation and Analysis of Transcriptome Sequences of Callus Culture from *Aquilaria malaccensis* Lam." in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

Tan Soon Guan, PhD
Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Chairman)

Mohd. Puad bin Abdullah, PhD
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner)

Ho Chai Ling, PhD
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner)

Wickneswari Ratnam, PhD
Professor
Faculty of Science and Technology
Universiti Kebangsaaan Malaysia
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 15 April 2015

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Rozi Mohamed, PhD
Associate Professor
Faculty of Forestry
Universiti Putra Malaysia
(Chairman)

Parameswari Namasivayam, PhD
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Member)

BUJANG KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BAP	6-benzylaminopurine
BLAST	Basic Local Alignment Search Tool
cDNA	complementary deoxyribonucleic acid
CDS	coding sequences
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CM	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol
CMK	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase
CMP	2-C-methyl-D-erythritol 2,4-cyclodiphosphate
CMP	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol synthase
CoA	coenzyme A
CTAB	cetyltrimethylammonium bromide
DEG	differentially expressed genes
DEPC	diethylpyrocarbonate
DXP	1-deoxy-D-xylulose-5-phosphate
DXPR	1-deoxy-D-xylulose-5-phosphate reductoisomerase
DXPS	1-deoxy-D-xylulose-5-phosphate synthase
EC	enzyme commission
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tags
EtBr	ethidium bromide
FDR	false discovery rate
FPP	farnesyl diphosphate
Gb	gigabases
GB	gigabyte
GO	gene ontology
HCl	hydrochloric acid
HDR	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
HDS	4-hydroxy-3-methylbut-2-enyl-diphosphate synthase
HMBPP	1-hydroxy-2-methyl-2-butenyl 4-diphosphate
HMG	3-hydroxy-3-methyl-glutaryl
HMGR	HMG-CoA reductase
IPP	isopentenyl diphosphate
IUCN	International Union for Conservation of Nature
KEGG	Kyoto Encyclopedia of Genes and Genomes
MARS	MA-plot with Random Sampling model
Mb	megabases
MCT	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase

MEP	2-C-methyl-D-erythritol 4-phosphate
MK	mevalonic acid kinase
mRNA	messenger ribonucleic acid
MS	Murashige and Skoog
MVA	mevalonic acid
MVAP	mevalonate acid 5-phosphate
MVAPP	mevalonate acid 5-diphosphate
MVPD	MVAPP decarboxylase
MVPK	mevalonate acid 5-phosphate kinase
NAA	1-naphthaleneacetic acid
NaCl	Sodium Chloride
NCBI	National Center of Biotechnology Information
ncRNA	non-coding ribonucleic acid
NGS	next-generation sequencing
NR	non-redundant
NTES	Sodium-Tris-EDTA-SDS
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PIR	Protein Information Resource
PPCM	2-Phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol
PRF	Protein Research Foundation
PVP	polyvinylpyrrolidone
RAM	randome-access memory
RefSeq	Reference Sequence
RIN	RNA integrity number
RNA	ribonucleic acid
RNA-Seq	RNA sequencing
rRNA	ribosomal ribonucleic acid
SAGE	serial analysis of gene expression
SDS	sodium dodecyl sulfate
SE	standard error
SOAP	Short Oligonucleotide Analysis Package
SSTE	Sodium dodecyl sulfate–Tris-HCl–EDTA
TGICL	TIGR Gene Indices Clustering Tools
tRNA	transfer ribonucleic acid
UPM	Universiti Putra Malaysia
UV	ultraviolet
WEGO	Web Gene Ontology Annotation Plot

CHAPTER 1

INTRODUCTION

Aquilaria malaccensis, locally known as karas, is a tropical tree belonging to the family Thymelaeaceae. It has been listed on the Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, 2014). Karas is one of the principle sources of agarwood, a highly commercial resinous wood and one of the valuable non-timber products in Asian tropical forest. Agarwood (also called gaharu, jinko, eaglewood, aloes wood or kalamabak, depending on the region) is the trade name for fragrant wood, which is used for perfume, incense, traditional medicines and other products (Chakrabarty *et al*., 1994).

Agarwood is formed as a result of pathological processes in response to wounding and fungal infection (Persoon, 2008; Ng *et al.*, 1997). This involves plant defense mechanism and induction of different defense genes (Nath *et al*., 2005). For better understanding of agarwood formation, it is important to study the molecular mechanism of *A. malaccensis* in response to stresses.

Karas is a timber species which takes a considerable long time to grow and form the resinous wood. Since the study using fresh plants is time consuming and difficult, *in vitro* plant tissue culture has provided an alternative to study the agarwood formation mechanism. The major compounds of agarwood are mostly identified as sesquiterpenoids and phenylethyl chromone derivatives (Naef, 2011; Yagura *et al.*, 2003). These agarwood fragrant constituents are reported being produced in *Aquilaria* calli and cell suspension cultures (Okudera and Ito, 2009). Therefore, study of agarwood formation mechanism under controlled environment using *Aquilaria* callus culture has become possible.

During the last decades, DNA/RNA sequencing has completely changed the vision of biology and particularly in plant biology. It has been possible to characterize a large number of genes by their nucleotide sequences, thus providing an alternative shortcut to study corresponding protein sequences and their functions (Delseny *et al.*, 2010; Morozova and Marra, 2008). As next-generation sequencing (NGS) technologies become increasingly mature and cost-effective, they are now being applied to the study of gene expression (Fox *et al*., 2009).

Gene expression or RNA analysis has been an important part in an increasingly large number of studies. Extraction of total RNA from eukaryotic cells has been a basic and crucial stage in molecular biology studies. Obtaining intact, high quality RNA is an essential step to ensure accuracy in analyzing gene expression. Thus it is essential to use the most appropriate method that maximizes the yield and maintains the integrity of extracted RNA. It has been difficult to extract good quality total RNA from woody plants such as *Aquilaria* because they often contain high levels of phenolic compounds,

carbohydrates and other secondary metabolites that bind and/or co-precipitate with RNA (Kansal *et al.*, 2008; MacRae, 2007).

Currently, there is no commercial kit available, specifically for the extraction of high-quality RNA from woody plants. Furthermore, because of the large amounts of polysaccharides, the common protocols for RNA extraction usually result in poor yields when applied to woody plants. Thus, it is crucial to optimize RNA extraction protocol for *A. malaccensis* for downstream application.

Although agarwood is in high demand and highly valuable, the molecular mechanism of agarwood formation and related defense response in *A. malaccensis* still remains poorly understood. Currently, there is no complete transcriptomic database available for *A. malaccensis*.

The main objectives of this study were:

1. To compare six RNA extraction methods and determine the optimal efficient method to extract RNA from wood tissues of *A. malaccensis*
2. To analyze the transcriptome from two libraries namely untreated and treated callus cultures
3. To compare the expression profiles of untreated and treated callus cultures, annotate their functions and assign their enzymatic pathways

BIBLIOGRAPHY

Agrios, G.N. (2005). *Plant Pathology*. 5th ed., pp. 208-209. San Diego: Academic Press.

Anonymous. (2003). Review of significant trade *Aquilaria malaccensis*. http://www.cites.org/sites/default/files/eng/com/pc/14/E-PC14-09-02-02-A2.pdf.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics.* 25(1): 25-29.

Azim, M. K., Khan, I. A., and Zhang, Y. (2014). Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. *Plant Molecular Biology*. 85: 193-208.

Banerjee, A.K., Laya, M.S., and Poon, P.S. (2006). Sesquiterpenes classified as phytoalexins. In: Attaur, R. (ed). *Studies in natural products chemistry*. Vol 33. pp. 193-237. Elsevier, Amsterdam.

Barden, A., Noorainie, A., Mulliken, T., and Song, M. (2000). *Heart of the matter: Agarwood use and trade and CITES implementation for Aquilaria malaccensis*. TRAFFIC International.

Basah, A.A. (2015). *Cloning and Characterization of Sesquiterpene Synthase 1 and δ-Guaiene Synthase 1 Genes Associated with 'Gaharu' Formation in Aquilaria Malaccensis Lam.* Unpublished dissertation, Universiti Putra Malaysia, Malaysia.

Bentley, D.R. (2006). Whole-genome re-sequencing. *Current Opinion in Genetics and Development*. 16: 545–552.

Bergemann, T.L., and Wilson, J. (2011). Proportion statistics to detect differentially expressed genes: a comparison with log-ratio statistics. *BMC Bioinformatics.* 12(1): 228.

Blüthgen, N., Brand, K., Čajavec, B., Swat, M., Herzel, H., and Beule, D. (2005). Biological profiling of gene groups utilizing Gene Ontology. *Genome Information*. 16(1): 106–115.

Bradford, J.R., Hey, Y., Yates, T., Li, Y., Pepper, S.D., and Miller, C.J. (2010). A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics*. 11: 282.

Brooks, G. (1998). *Biotechnology in Healthcare: An Introduction to Biopharmaceuticals*. Pharmaceutical Press, London, UK.

Brown, T.A. (2002). *Genomes*. 2nd Edition. Oxford: Wiley-Liss.

Brunner, A.M., Yakovlev, I.A., and Strauss, S.H. (2004). Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology*. 4: 14.

Bryant Jr., D.W., and Mockler, T.C. (2012). *De novo* short-read assembly. In *Bioinformatics for High Throughput Sequencing*. pp. 85-105. Springer, New York.

Buckingham, L. (2011). *Molecular diagnostics: fundamentals, methods and clinical applications*. FA Davis, USA.

Bullard, J.H., Purdom,E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics.* 11: 9.

Bustin, S.A., Benes, V., Nolan, T., and Pfaffl, M.W. (2005). Quantitative real-time RT-PCR--a perspective. *Journal of Molecular Endocrinology*. 34: 597-601.

Chakrabarty, K., Kumar, A. and Menon, V. (1994). *Trade in Agarwood*. In: Barden, A., Noorainie Awang Anak, T. Mulliken, and M. Song. (2000). *Heart of the matter: Agarwood use and trade and CITES implementation for Aquilaria malaccensis*. TRAFFIC International, India.

Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*. 11: 113-116.

CITES. (2014). Appendix II of Convention on International Trade in Endangered Species of Wild Fauna and Flora. http://www.cites.org/eng/app/appendices.php.

Clarke, K., Yang, Y., Marsh, R., Xie, L., and Zhang, K.K. (2013). Comparative analysis of *de novo* transcriptome assembly. *Science China Life Sciences*. 56(2): 156-162.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B., Faulkner, G., Brown, M., Taylor, D., Steptoe, A., Wani, S., Bethel, G., Robertson, A., Perkins, A., Bruce, S., Lee, C., Ranade, S., Peckham, H., Manning, J., McKernan, K., and Grimmond, S. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*. 5(7): 613-619.

Cones, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21: 3674-3676.

Cseke, L.J., Kirakosyan, A., Kaufman, P.B., and Westfall, M.V. (2011). *Handbook of Molecular and Cellular Methods in Biology and Medicine.* CRC Press, USA.

Delseny, M., Han, B., and Hsing, Y.I. (2010). High throughput DNA sequencing: The new sequencing revolution. *Plant Science*. 179: 407-422.

Emrich, S.J., Barbazuk, W.B., Li, L., and Schnable, P.S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research*. 17: 69–73.

Faldt, J., Martin, D., Miller, B., Rawat, S., and Bohlmann, J. (2003). Traumatic resin defense in Norway spruce (*Picea abies*): methyl jasmonate-induced terpene synthase gene expression, and cDNA cloning and functional characterization of (+)-3-carene synthase. *Plant Molecular Biology*. 51: 119–133.

Faridah-Hanum, I., Mustapha, M. Z., Lepun, P., Tuan Marina, T. I., Nazre, M., Alan, R., and Mohamed, R. (2009). Notes and the distribution and ecology of *Aquilaria* Lam. (Thymelaeaceae) in Malaysia. *The Malaysian Foresters*. 72(2): 247-259.

Fleige, S., and Pfaffl, M.W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*. 27: 126-139.

Fox, S., Filichkin, S., and Mockler, T.C. (2009). Applications of ultra-high- throughput sequencing. *Methods in Molecular Biology*. 553: 79-108.

Garg, R., Patel, R.K., Tyagi, A.K., and Jain, M. (2011). *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*. 18(1): 53-63.

Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nature Chemical Biology*. 3(7): 408–414.

Góngora-Castillo, E., and Buell, C.R. (2013). Bioinformatics challenges in *de novo* transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Natural Product Reports*. 30(4): 490-500.

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., Hacohen, N., Gnirke, A., Rhind, N., Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 29(7): 644-652.

Howe, G.T., Yu,J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., Lorenz, W.W., and Dean, J.F. (2013). A SNP resource for Douglas-fir: *de novo* transcriptome assembly and SNP detection and validation. *BMC Genomics*. 14(1): 137.

Huang, X., and Madan, A. (1999). CAP3: a DNA sequence assembly program. *Genome Research*. 9: 868–877.

Illumina. (2010). *mRNA Sequencing Sample Preparation Guide*. http://supportres.illumina.com/documents/myillumina/274f5b53-5e28-4a5c-a8a1-93eaba3c318b/mrna-seq_sampleprep_1004898_d.pdf.

Imbeaud, S., and Auffray, C. (2005). 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments. *Drug Discovery Today*. 10: 1175-1182.

Iorizzo, M., Senalik, D.A., Grzebelus, D., Bowman, M., Cavagnaro, P.F., Matvienko, M., Ashrafi, H., Deynze, A.V., and Simon, P.W. (2011). *De novo* assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics*. 12(1): 389.

Ishihara, M., Tsuneya, T., and Uneyama, K. (1993). Fragrant sesquiterpenes from agarwood. *Phytochemistry*. 33: 1147–1155.

IUCN Red List. (2014). The IUCN Red List of Threatened Species. Version 2014.2. <www.iucnredlist.org>.

Jantan, I. (1990). *Gaharu*. *Timber Digest*, 107, December. In: Barden, A., Noorainie, A., Mulliken, T., and Song, M. (2000). *Heart of the matter: Agarwood use and trade and CITES implementation for Aquilaria malaccensis.* TRAFFIC International.

Jayaraman, S., Daud, N.H., Halis, R., and Mohamed, R. (2014). Effects of plant growth regulators, carbon sources and pH values on callus induction in *Aquilaria malaccensis* leaf explants and characteristics of the resultant calli. *Journal of Forestry Research*. 25(3): 535-540.

Jayaraman, S., and Mohamed, R. (2015). Crude extract of *Trichoderma* elicits agarwood substances in cell suspension culture of the tropical tree, *Aquilaria malaccensis* Lam. *Turkish Journal of Agriculture and Forestry*. doi:10.3906/tar-1404-63.

Jiang, H., and Wong, W.H. (2009). Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics*. 25(8): 1026-1032.

Jones, J.D.G., Dunsmuir, P., and Bedbrook, J. (1985). High level expression of introduced chimaeric genes in regenerated transformed plants. *EMBO Journal*. 4: 2411-2418.

Jong, P.L., Tsan, P., and Mohamed, R. (2014). Gas chromatography-mass spectrometry analysis of agarwood extracts from mature and juvenile *Aquilaria malaccensis*. *International Journal of Agriculture and Biology*. 16: 644–648.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Research*. 36: D480-D484.

Kansal, R., Kuhar, K., Verma, I., Gupta, R.N., Gupta, V.K., and Koundal, K.R. (2008). Improved and convenient method of RNA isolation from polyphenols and polysaccharide rich plant tissues. *Indian Journal of Experimental Biology*. 47: 842-845.

Kumeta, Y. and Ito, M. (2010). Characterization of δ-guaiene synthases from cultured cells of *Aquilaria*, responsible for the formation of the sesquiterpenes in agarwood. *Plant Physiology*. 154: 1998–2007.

La Frankie, J. (1994). Population dynamics of some tropical trees that yield non-timber forest products. *Economic Botany*. 48(3): 301-309.

Langmead, B., Hansen, K., and Leek, J. (2010). Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biology*. 11(8):R83.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 10: R25.

Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J.A., Stewart, R., and Dewey, C.N. (2014). Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. *Genome Biology*. 15(12): 553.

Li, C., Wang, Y., Huang, X., Li, J., Wang, H., and Li, J. (2013). *De novo* assembly and characterization of fruit transcriptome in *Litchi chinensis* Sonn and analysis of differentially regulated genes in fruit in response to shading. *BMC Genomics*. 14(1): 552.

Li, D., Deng, Z., Qin, B., Liu, X., and Men, Z. (2012). *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics.* 13(1): 192.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 25: 2078-2079.

Liu, Y., Chen, H., Yang, Y., Zhang, Z., Wei, J., Meng, H., Chen, W., Feng, J., Gan, B., Chen, X., Gao, Z., Huang, J., Chen, B., and Chen, H. (2013). Whole-tree agarwood-inducing technique: an efficient novel technique for producing high-quality agarwood in cultivated *Aquilaria sinensis* trees. *Molecules*. 18(3): 3086-3106.

Lorenz, W.W., Yu, Y., and Dean, J.F.D. (2010). An improved method of RNA isolation from loblolly pine (*P. taeda* L.) and other conifer species. *Journal of Visualized Experiment.* 36: e1751.

MacRae, E. (2007). Extraction of plant RNA. *Methods in Molecular Biology*. 353: 15-24.

Mardis, E.R., and Wilson, R.K. (2009). Cancer genome sequencing: a review. *Human Molecular Genetics.* 18: 163-168.

Marguerat, S., and Bähler, J. (2010). RNA-Seq: from technology to biology. *Cellular and Molecular Life Sciences.* 67(4): 569-579.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-Seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 18: 1509-1517.

Martin, J.A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*. 12(10): 671-682.

Metzker, M.L. (2009). Sequencing technologies-the next generation. *Nature Reviews Genetics*. 11(1): 31-46.

Miller, J.R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*. 95(6): 315-327.

Mizrachi, E., Hefer, C.A., Ranik, M., Joubert, F., and Myburg, A.A. (2010). *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics*. 11(1): 681.

Mohamed, R., Jong, P.L., and Zali, M.S. (2010). Fungal diversity in wounded stems of *Aquilaria malaccensis*. *Fungal Diversity*. 43(1): 67-74.

Morozova, O., and Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*. 92: 255-264.

Morozova, O., Hirst, M., and Marra, M.A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics*. 10: 135-151.

Naef, R. (2011). The volatile and semi-volatile constituents of agarwood, the infected heartwood of *Aquilaria* species: a review. *Flavour and Fragrance Journal*. 26: 73-87.

Nagarajan, N., and Pop, M. (2013). Sequence assembly demystified. *Nature Reviews Genetics*. 14(3): 157-167.

Nath, P., Sane, V.A., Sane, A.P., and Trivedi, P.K. (2005). Plant Gene Expression, Regulation of. In: Meyers RA (ed) *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, 2nd edn, vol 10. Wiley-VCH, Weinheim, pp. 349-350.

Ng, L.T., Chang, Y.S., and Kadir, A.A. (1997). A review on agar (gaharu) producing *Aquilaria* species. *Journal of Tropical Forest Products*. 2(2): 272–285.

Novaes, E., Drost. D.R., Farmerie, W.G., Pappas, G.J., Grattapaglia, D., Sederoff, R.R., and Kirst, M. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics*. 9: 312.

Okudera, Y., and Ito, M. (2009). Production of agarwood fragrant constituents in *Aquilaria* calli and cell suspension cultures. *Plant Biotechnology*. 26(3): 307–315.

Oldfield, S., Lusty, C., and MacKiven, A. (1998). *The World List of Threatened Trees*. World Conservation Press, Cambridge.

Opitz, L., Salinas-Riester, G., Grade, M., Jung, K., Jo, P., Emons, G., Ghadimi, B.M., Beissbarth, T., and Gaedcke, J. (2010). Impact of RNA degradation on gene expression profiling. *BMC Medical Genomics.* 3: 36.

Oshlack, A., Robinson, M.D. and Young, M.D. (2010) From RNA-seq reads to differential expression results. *Genome Biology*. 11: 220.

Ossowski, S. Schennberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Research.* 12: 2024–2033.

Ozsolak, F., and Milos, P.M. (2010). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*. 12(2): 87-98.

Persoon, G.A. (2008). Growing 'The Wood of The Gods': Agarwood Production in Southeast Asia. *Smallholder Tree Growing for Rural Development and Environmental Services.* pp. 245-262. USA: Springer.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., and Quackenbush, J. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 19: 651-652.

Pojanagaroon, S., and Kaewrak, C. (2005). Mechanical methods to stimulate aloes wood formation in *Aquilaria crassna* Pierre ex H. LEC. (Kritsana) trees. In: Jatisatienr, A., Paratasilpin, T., Elliott, S., Anusarnsunthorn, V., Wedge, D., Craker, L.E., and Gardner, Z.E. (eds) WOCMAP III congress on medicinal and aromatic plants: conservation, cultivation and sustainable use of medicinal and aromatic plants. vol 2. pp. 161–166.

Qi, S.Y., He, M.L., Lin, L.D., Zhang, C.H., Hu, L.J., and Zhang, H.Z. (2005). Production of 2-(2-phenylethyl) chromones in cell suspension cultures of *Aquilaria sinensis*. *Plant Cell, Tissue and Organ Culture*. 83(2): 217-221.

Rahman, M.A., and Basak, A.C. (1980). Agar production in agar trees by artificial inoculation and wounding. *Bano Bigan Patrika*. 9: 86-93.

Reis-Filho, J. S. (2009). Next-generation sequencing. *Breast Cancer Research*. 11(3): S12.

Reymond, P., and Farmer, E.E. (1998). Jasmonate and salicylate as global signals for defense gene expression. *Current Opinion in Plant Biology*. 1(5): 404-411.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 11: R25.

Romero, I.G., Pai, A.A., Tung, J., and Gilad, Y. (2014). RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biology*. 12(1): 42.

Rounsley, S., Marri, P.R., Yu, Y., He, R., Sisneros, N., Goicoechea, J.L., Lee, S.J., Angelova, A., Kudrna, D., Luo, M., Affourtit, J., Desany, B., Knight, J., Niazi, F., Egholm, M., and Wing, R.A. (2009). *De Novo* Next-Generation Sequencing of Plant Genomes. *Rice.* 2(1):35–43.

Sambrook, J., and Russel, D.W. (2001). *Molecular cloning: A laboratory manual*. Third edition. Cold Spring Harbor Laboratory Press, New York.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*. 7(1): 3.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. *Nature Methods*. 5: 16–18.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*. 26(10): 1135-1145.

Surget-Groba, Y., and Montoya-Burgos, J.I. (2010). Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Research*. 20: 1432-1440.

Swee, L.C. (2008). *Agarwood (Aquilaria malaccensis) in Malaysia*. In: International Expert Workshop on CITES Non-Detriment Findings, Mexico, November 2008.

Tabata, Y., Widjaya, E., Mulyaningsih, T., Parman, I., Wiriadinata, H., Mandang, Y.I., and Itoh, T. (2003). Structural survey and artificial induction of aloeswood. *Wood Research:Bulletin of the Wood Research Institute Kyoto University*. 90: 11–12.

Tan, S.C., and Yiap, B.C. (2009). DNA, RNA, and protein extraction: the past and the present. *Journal of Biomedicine and Biotechnology*. 2009: 574398.

Trapnell, C., and Salzberg, S.L. (2009). How to map billions of short reads onto genomes. *Nature Biotechnology*. 27(5): 455.

Varshney, R.K., Nayak, S.N., May, G.D., and Jackson, S.A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology*. 27(9): 522-530.

Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I., and Marden, J.H. (2008). Rapid transcriptome characterization for a non-model organism using 454 pyrosequencing. *Molecular Ecology*. 17: 1636-1647.

Vijay, N., Poelstra, J.W., Künstner, A., and Wolf, J.B. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular Ecology*. 22(3): 620-634.

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology*. 64: 665-700.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics.* 26(1): 136-138.

Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., Chen, X., and Li, Y. (2010). *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics*. 11(1): 726.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 10: 57-63.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani,V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Trzyk, G.P., Lupski, J.R., Chinault, C., Song, X., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., and Rothberg, J.M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 452: 872–876.

Wicker, T., Narechania, A., Sabot, F., Stein, J., Vu, G.T., Graner, A., Ware, D., and Stein, N. (2008). Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC Genomics*. 9: 518.

Wilhelm, B.T., and Landry, J.R. (2009). RNA-Seq--quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*. 48: 249-257.

Wong, M.M.L., Cannon, C.H., and Wickneswari, R. (2011). Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. *BMC Genomics*. 12: 342.

Wong, M.T., Siah, C.H., Faridah, Q.Z., and Mohamed, R. (2013). Characterization of wound responsive genes in *Aquilaria malaccensis*. *Journal of Plant Biochemistry and Biotechnology*. 22(2): 168-175.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T., Li, Y., Xu, X., Wong, G.K., and Wang, J. (2014). SOAPdenovo-Trans: d*e novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. 30: 1660-1666.

Xu, Y., Zhang, Z., Wang, M., Wei, J., Chen, H., Gao, Z., Sui, C., Luo, H., Zhang, X., Yang, Y., Meng, H., and Li, W. (2013). Identification of genes related to agarwood formation: transcriptome analysis of healthy and wounded tissues of *Aquilaria sinensis*. *BMC Genomics*. 14(1): 227.

Yagura, T., Ito, M., Kiuchi, F., and Honda, G. (2003). Four new 2-(2-phenylethyl) chromone derivatives from withered wood of *Aquilaria sinensis*. *Chemical and Pharmaceutical Bulletin*. 51: 560–564.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*. 30(4): e15.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., and Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*. 34: W293–297.

Yu, F., and Utsumi, R. (2009). Diversity, regulation, and genetic manipulation of plant mono-and sesquiterpenoid biosynthesis. *Cellular and Molecular Life Sciences*. 66(18): 3043-3052.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using *de Bruijn* graphs. *Genome Research*. 18(5): 821-829.