ANALYSIS OF THE EXPRESSION OF MICRORNAS AND THEIR TARGETS IN DEVELOPING INFLORESCENCES OF OIL PALM (*Elaeis guineensis* Jacq.)

HO HUI LI

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

ANALYSIS OF THE EXPRESSION OF MICRORNAS AND THEIR TARGETS IN DEVELOPING INFLORESCENCES OF OIL PALM (*Elaeis guineensis* Jacq.)

HO HUI LI

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: HO HUI LI

Matric No: SHC080018

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):
ANALYSIS OF THE EXPRESSION OF MICRORNAS AND THEIR TARGETS IN DEVELOPING INFLORESCENCES OF OIL PALM (*Elaeis guineensis* Jacq.)

Field of Study: PLANT MOLECULAR BIOLOGY

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ANALYSIS OF THE EXPRESSION OF MICRORNAS AND THEIR TARGETS IN DEVELOPING INFLORESCENCES OF OIL PALM

(*Elaeis guineensis* Jacq.)

ABSTRACT

African oil palm (*Elaeis guineensis* Jacq.) is an important perennial oil crop with a long flower developmental phase. The formation of oil palm fruits relies on the successful progression of upstream flower developmental events in the inflorescences. As a monoecious plant, the sex ratio of the female to male inflorescences on each palm is important for breeding and commercial production. This study examined the gene expression profiles at the small RNA transcriptome level in developing female and male inflorescences of oil palm at the +6 and +15 stages of floral development, which correspond to the emergence of floral meristems and to the formation of floral organs respectively. This study also aimed to determine the differences in the mRNA transcriptome between the female and male inflorescences. As only the fertilized female inflorescences develop into fruit bunches, an integration of high-throughput sequencing of mRNA, small RNA and RNA degradome libraries was used to identify and to quantify miRNAs and their mRNA targets that were differentially expressed in +6 and +15 female inflorescences. From the female inflorescences, a consensus reference transcriptome was assembled. Mononucleotide repeats constituted the largest group of simple sequence repeats in the current datasets, followed by tri- and dinucleotide repeats. From the transcriptome data, 97 potential sex-specific transcripts were identified. Transcript expression patterns were validated using nCounter analysis for oil palm orthologs of acid phosphatase and *DEFICIENS*, which showed male-specific expression patterns whereas orthologs of bZIP transcription factor, late embryogenesis abundant protein and *TASSELSEED1* showed female-specific expression patterns. Transcripts for orthologs of acid phosphatase and late embryogenesis abundant protein were also strongly inflorescence-specific. Twenty-nine

putative oil palm-specific miRNAs were identified and different subsets of orthologous miRNA were expressed in female and male inflorescences, with miR535 as the most highly expressed miRNA in the female inflorescence while miR166 was the most highly expressed miRNA in the male inflorescence. From the 18 orthologous miRNA families and their mRNA targets with differential expression between +6 and +15 female inflorescences, two distinct subsets of orthologous miRNAs showed inverse expression patterns in +6 and +15 female inflorescences: miR163, miR168, miR172, miR1859, miR1873, miR396, miR4365, miR835 and miR858 showed an inverse expression pattern to at least one of their predicted targets and cleavage products of some of these predicted targets were identified in the degradome libraries. Many of the predicted targets are transcripts encoding protein kinases and components involved in DNA replication and repair. In summary, the present study showed that distinct subsets of miRNAs were differentially expressed in developing female and male inflorescences of oil palm. The inflorescence- and sex-specific transcripts for which the expression coincides with the early phase of sex differentiation in oil palm inflorescences may account at least in part, for the differences between female and male inflorescences. The miRNAs that were differentially expressed in the +6 and +15 female inflorescences are likely to act in concert with their mRNA targets to regulate the early phase of floral organ establishment.

Keywords: degradome, flower, sequencing, transcriptome

ANALISIS EKSPRESI MIKRORNA DAN SASARAN MRNA DALAM PERKEMBANGAN BUNGA KELAPA SAWIT (*Elaeis guineensis* Jacq.)

ABSTRAK

Kelapa sawit Afrika (*Elaeis guineensis* Jacq.) adalah sejenis tanaman saka yang penting dalam penghasilan minyak kelapa sawit dan mempunyai fasa perkembangan bunga yang panjang. Pembentukan buah kelapa sawit amat bergantung kepada aktiviti-aktiviti perkembangan bunga. Sebagai tumbuhan berumah satu (monoecious), nisbah jantina bunga pada setiap pokok sawit adalah penting untuk proses pembiakan dan pengeluaran komersial. Kajian ini telah menyiasat profil ekspresi untuk RNA bersaiz kecil dalam bunga betina dan bunga jantan kelapa sawit pada peringkat perkembangan bunga +6 dan +15, iaitu semasa kemunculan meristem bunga dan pembentukan organ-organ bunga. Kajian ini turut bertujuan untuk menentukan perbezaan ekspresi gen antara bunga betina dan bunga jantan. Memandangkan hanya bunga betina yang akan membentuk buah sawit, integrasi kaedah penjujukan berskala besar untuk mRNA, RNA bersaiz kecil dan "*degradome*" telah digunakan untuk mengenalpasti dan mengukur ekspresi miRNAs dan sasaran mRNAnya yang menunjukkan perubahan ekspresi yang ketara semasa perkembangan bunga betina (peringkat +6 dan +15). Sebuah transkriptom rujukan telah dibina dari bunga betina kelapa sawit. Mikrosatelit mononukleotida merupakan kelas mikrosatelit yang paling dominan, diikuti oleh mikrosatelit trinukleotida dan dinukleotida. Berdasarkan data transkriptom, 97 transkrip RNA telah dikenalpasti sebagai spesifik sama ada kepada bunga betina atau bunga jantan. Antara transkrip yang mana corak ekspresinya telah disahkan melalui analisis "*nCounter*", transkrip ortolog Acid Phosphatase dan *DEFICIENS* menunjukkan corak ekspresi yang spesifik kepada bunga jantan, manakala transkrip ortolog *bZIP*, Late Embryogenesis Abundant Protein dan *TASSELSEED1* menunjukkan corak ekspresi yang spesifik kepada bunga betina. Transkrip-transkrip ortolog Acid Phosphatase dan Late Embryogenesis Abundant Protein juga menunjukkan corak ekspresi yang spesifik kepada tisu bunga berbanding

dengan tisu-tisu kelapa sawit yang lain. Dua puluh sembilan miRNA yang spesifik kepada kelapa sawit telah dilaporkan. Subset miRNA ortolog yang berlainan diekpreskan dalam bunga kelapa sawit, di mana miR535 merupakan miRNA yang mempunyai ekspresi yang paling tinggi dalam bunga betina manakala miR166 merupakan miRNA yang mempunyai ekspresi yang paling tinggi dalam bunga jantan. Antara 18 miRNA ortolog dan sasarannya yang menunjukkan perubahan ekspresi yang ketara, dua kumpulan miRNA ortolog menunjukkan perubahan ekspresi yang bercorak songsang dalam bunga betina +6 and bunga betina +15: miR163, miR168, miR172, miR1859, miR1873, miR396, miR4365, miR835 dan miR858 menunjukkan perubahan tahap ekspresi secara songsang berbanding dengan sekurang-kurangya salah satu daripada sasaran mRNAnya dan sebahagian daripada hasil belahan sasaran mRNA telah dikenalpasti dalam perpustakaan "*degradome*". Kebanyakan sasaran mRNA ini adalah transkrip yang mengkodkan protein kinase dan komponen yang terlibat dalam replikasi dan perbaikan DNA. Kesimpulannya, kajian ini telah menunjukkan bahawa subset miRNA yang berlainan diekspreskan semasa perkembangan bunga betina dan bunga jantan kelapa sawit. Adalah dicadangkan bahawa transkrip yang spesifik kepada tisu bunga dan transkrip yang spesifik kepada jantina tertentu bersama-sama menyumbang kepada perbezaan antara bunga betina dan bunga jantan. Turut dicadangkan bahawa miRNA yang menunjukkan perubahan pada peringkat perkembangan bunga +6 dan +15 bertindak bersama sasaran mRNA untuk mengawalatur peringkat awal proses pembentukan organ-organ bunga kelapa sawit.

Kata kunci: "*degradome*", bunga, penjujukan, transkriptom

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Professor Dr. Jennifer Ann Harikrishna for her patience and guidance throughout my PhD study. It is her continued support and motivation that drive me towards the completion of this study.

I would like to acknowledge Ministry of Education Malaysia and University of Malaya for the sponsorship, research facilities and funding for my PhD study.

I would also like to thank all my friends, particularly Fauziah, Khai Swan, Su Ee, Wai Ting and Wan Sin for the great time we spent together inside and outside the lab. I gratefully acknowledge the contributions of Joel and Ranga for their knowledge and technical support in bioinformatics analyses.

Finally my deepest gratitude goes to my beloved family. I owe a special thanks to my husband, Raymond for being my personal cheerleader and supporting me throughout my PhD journey. I also dedicate this work to my baby girl Abigail− the greatest gift of all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

α	:	alpha
β	:	beta
°C	:	degree celsius
µg	:	microgram
µl	:	microliter
µM	:	micromolar
%	:	percentage
A	:	adenine (in DNA base sequence)
ATP	:	adenosine triphosphate
bp	:	base pair
C	:	cytosine (in DNA base sequence)
$CaCl_2$	:	calcium chloride
cDNA	:	complementary deoxyribonucleic acid
cm	:	centimeter
COG	:	Clusters of Orthologous Groups
contigs	:	contiguous sequences
Ct	:	cycle threshold
CTAB	:	cetyltrimethylammonium bromide
DNA	:	deoxy ribonucleic acid
DNase	:	deoxyribonuclease
dNTP	:	deoxy ribonucleotide triphosphate
dsRNA	:	double-stranded RNA
DTT	:	dithiothreitol
EDTA	:	ethylene diamine tetra acetic acid

EST	:	expressed sequence tag
e.g.	:	*exempli gratia*
FDR	:	false discovery rate
G	:	guanine (in DNA base sequence)
g	:	gram
g	:	relative centrifugal force
GO	:	gene ontology
hr	:	hour
i.e.	:	*id est*
kb	:	kilobase
kcal	:	kilocalorie
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
KOG	:	Eukaryotic Orthologous Groups of proteins
LB	:	Luria-Bertani
M	:	molar
MFE	:	minimum folding energy
mg	:	milligram
$MgCl_2$	:	magnesium chloride
$MgSO_4$	:	magnesium sulphate
min	:	minute
miRNA	:	microRNA
miRNA*	:	microRNA star strand
ml	:	milliliter
mM	:	millimolar
$MnCl_2$	:	manganese (II) chloride

mol	:	mole
mRNA	:	messenger RNA
N	:	ambiguous nucleotides (in DNA base sequence)
NaCl	:	sodium chloride
ng	:	nanogram
nm	:	nanometer
Nr	:	Genbank non-redundant protein database
nt	:	nucleotide
OD	:	optical density
PCI	:	phenol-chloroform-isoamylalcohol
PCR	:	polymerase chain reaction
PMRD	:	Plant MicroRNA Database
pre-miRNA	:	precursor miRNA
pri-miRNA	:	primary miRNA
PVP	:	polyvinylpyrrolidone
qRT-PCR	:	quantitative real-time reverse transcriptase PCR
RISC	:	RNA-induced silencing complex
RNA	:	ribonucleic acid
RNAi	:	RNA interference
RNase	:	ribonuclease
RNA-seq	:	high-thoughput RNA sequencing
RPKM	:	reads per kilobase per million mapped reads
rpm	:	revolutions per minute
rRNA	:	ribosomal RNA
RT-PCR	:	reverse-transcriptase PCR
s	:	second

siRNA	:	small interfering RNA
snRNA	:	small nuclear RNA
snoRNA	:	small nucleolar RNA
sp.	:	species
sRNA-seq	:	high-thoughput small RNA sequencing
SSR	:	simple sequence repeat
T	:	thymine (in DNA base sequence)
TBE	:	tris-borate EDTA
TE	:	tris-EDTA
TEMED	:	tetramethylethylenediamine
TPM	:	tags (or transcript) per million
Tris-HCl	:	trisaminomethane hydrochloride
tRNA	:	transfer RNA
U	:	unit
UV	:	ultra-violet
V	:	voltage
v/v	:	volume per volume
w/v	:	weight per volume

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

African oil palm (*Elaies guineensis* Jacq.) is a member of the palm subfamily Arecaceae. It produces two distinct types of oil, i.e., palm oil and palm kernel oil which are extracted from the fruit mesocarp and seed kernel, respectively.The thick fleshy mesocarp of oil palm fruit is exceptionally rich in oil (80% dry mass), making it one of the world's most efficient oil-crops (Murphy, 2009). First introduced to Malaysia in the early 1870's, oil palm has been commercially grown since 1917 (Cottrell, 1991). The primary variety of oil palm grown in Malaysia is the thin shell *tenera* (fruit shell of up to 3 mm thick), which is a hybrid of *dura* (3 to 8 mm thick) and *pisifera* (with no shell) (Wahid et al., 2004). For the year 2016, 23.29 million tonnes of palm oil products were exported, contributing an annual export value of RM64.58 billion to Malaysia (Puspadevi, 2017).

Oil palm has 32 chromosomes (2n=32) with a genome size of approximately 1.8 Gbp (Madon et al., 1998). In 2013, Singh et al. (2013) published the first assembled oil palm genome based on the AVROS (Algemene Vereniging van Rubberplanter terOostkust van Sumatra) which produces a *pisifera* fruit form. The total size of the assembly (a fifth genome build or P5-build) is 1.535Gb, which was represented by 16 genetic groups. More recently in 2016, Jin et al. (2016) published the draft genome sequence of an elite *dura* palm, with a total size of 1.701 Gb. The availability of oil palm reference genomes has provided a rich resource for subsequent studies in oil palm, including gene coexpression analysis (Guerin et al., 2016), genome-wide analysis of transposable elements (Beulé et al., 2015), genotyping analysis (Roberdi e al., 2015), genome sequencing/re-sequencing (Jin et al., 2016) and single nucleotide polymorphism (SNP) discovery (Pootakham et al., 2016).

As a monoecious plant, oil palm produces both female and male inflorescences on the same palm in alternate cycles, with occasional formation of mixed sex inflorescences during the transition between female and male batches (Biradar, 1978). In mature oil palms, the development of an inflorescence takes 32-36 months (Corley, 1976a), through a series of individual differentiation phases with the development of the inflorescence meristem being the longest phase in oil palm reproductive development (Adam et al., 2005). Each inflorescence primordium has the potential to develop into female and male inflorescences. While the application of growth regulators (plant hormones) (Corley, 1976b) and stress conditions (Durand-Gasselin et al., 1999) have been reported to affect the sex of oil palm inflorescences, the genetic factors regulating sex determination is not well understood.

MicroRNAs (miRNAs) are short single-stranded regulatory RNAs (generally 21-24 nucleotides) which derive from long stem-loop precursors transcribed from endogenous genomic DNA by RNA polymerase II (Bartel, 2004; Lee et al., 2004). MiRNAs selectively repress gene expression by binding to their target mRNA and cause degradation or translational silencing of that mRNA in a sequence-specific manner. RNA interference (RNAi) was first described as post-transcriptional gene silencing (PTGS) in plants (Llave et al., 2002; Bartel, 2004). The first plant miRNA was discovered in Arabidopsis (Reinhart et al., 2002). There are evolutionarily conserved miRNAs which can be found in species from different lineage and several recently diverged non-conserved miRNAs which are species-specific (Barakat et al., 2007; Subramanian et al., 2008; Sunkar et al., 2008; Lelandais-Briere et al., 2009; Zhang et al., 2010a). The plant miRNAs and their corresponding targets demonstrate a very high degree of sequence complementarity (Rhoades et al., 2002; Jones-Rhoades & Bartel 2004). In plants, miRNAs regulate the expression of genes that encode transcription factors, stress response proteins and other proteins required for growth,

development and physiology of plants (Borges & Martienssen, 2015; Li & Zhang, 2016). With respect to flower development, miRNAs have been reported to affect transition to flowering (Aukerman & Sakai, 2003; Achard et al., 2004; Spanudakis & Jackson, 2014), sex determination (Chuck et al., 2007) and floral organ development (Aukerman & Sakai, 2003; Zhu et al., 2009; Liu et al., 2010) via cleavage of target mRNAs (Guo et al., 2005; Li et al., 2014) and inhibition at the translation level (Chen, 2004; Li et al., 2013).

The overall flower maturation of oil palm is a long and complicated process, thus making conventional oil palm breeding a relatively slow process. The primary aim of the current study was to investigate the gene expression profiles at the sRNA transcriptome (sRNAome) level, in both female and male inflorescences of oil palm. Given that only the fertilized female inflorescences develop into fruit bunches which are desired for commercial oil and seed production, the study was further extended to specifically investigate the expression patterns of miRNAs and their mRNA target(s) during two key developmental stages of female inflorescence, i.e. +6 and +15 stages. The +6 and +15 stages fall within the period during which floral organs are forming. It was also the interest of this study to investigate the gene expression differences between female and male inflorescences at an early stage of development, i.e., leaf stage +6, as this is the earliest stage at which the sex of an inflorescence can be distinguished by morphology. The specific objectives of this study are summarized as follows:

1. To determine the differences in gene expression of +6 stage female and male inflorescences at the mRNA transcriptome level
2. To determine the differences in gene expression of developing female and male inflorescences at the sRNA transcriptome level

3. To determine the miRNAs and their mRNA targets that were differentially expressed at +6 and +15 stages of female inflorescence development.

CHAPTER 2

LITERATURE REVIEW

2.1 Current status of oil palm industry in Malaysia

Palm oil is the world's most produced and consumed vegetable oil (United States Department of Agriculture, 2015). Indonesia and Malaysia are the two largest palm oil producers, together accounting for more than 85% of total global palm oil production (United States Department of Agriculture, 2016). In 2015, oil palm products contributed to RM60.17 billion of Malaysia's total export revenue, with palm oil export alone accounting for RM41.26billion of export revenue (Malaysian Palm Oil Board [MPOB], 2016a). India, the European Union and China are the major importers of Malaysia palm oil, accounting for more than 48% of the country's export of palm oil (MPOB, 2016b). The oil palm cultivation area in Malaysia has increased steadily to 5.64 million hectares in 2015, as compared to 5.39 million hectares recorded for the previous year, with Sabah having the largest land area planted with oil palm (1.54 million hectares or 27% of the total oil palm planted area) (MPOB, 2016a). Among the largest local oil palm plantation companies are Sime Darby Berhad, IOI Corporation Berhad, Kuala Lumpur Kepong Berhad and Felda Global Ventures Holdings Berhad. The Malaysian Palm Oil Board (MPOB) is the leading government agency focused on oil palm, with a role to promote and develop national objectives, policies and priorities for the sustainability of the Malaysian oil palm industry.

Oil palm fruits produce two different types of oil: palm oil from the mesocarp and palm kernel oil from the seeds (Singh et al., 2009). Palm oil and palm kernel oil have different physical and chemical properties hence the variation in their applications (Mba et al., 2015). Palm oil containshigh amounts of saturated fat, vitamins and antioxidants, making it ideal for edible purposes (Koushki et al., 2015). With a higher content of saturated fats compared to palm oil, palm kernel oil is widely used in

commercial cooking since the higher saturated fat content allows greater thermal stability as well as a longer shelf life. For industrial application, both palm oil and palm kernel oil are also used for manufacturing a vast range non food items such as cosmetics, detergents, lubricants, soaps, toothpaste and waxes (Malaysian Palm Oil Council [MPOC], 2012).

Besides the aforementioned uses, palm oil also presents a high potential as a renewable source of energy. Biofuel is a environmental friendly renewable energy source, with functional properties resembling that of fossil fuels (Mekhilef et al., 2011). In 2006, MPOB has introduced a biodiesel named Envo Diesel, which is constituted by 5% of fully refined liquid palm oil and 95% of petroleum diesel (Bernama, 2006). This 5% substitution by palm oil in the Envo Diesel represents a reduction in the annual import of 500,000 tonnes of diesel which is equivalent to a saving of more than RM1 billion in foreign exchange each year (Basiron, 2006). The performance of Envo Diesel has been experimentally verified and compared against ordinary diesel fuel (Kalam et al., 2008). Since 2012, Malaysia has initiated a programme promoting the B5 biodiesel (a blend of 5% palm oil or palm methyl ester and diesel) among subsidised sectors such as retail stations, fleetcard, skid tanks and fisheries in limited areas of Peninsula Malaysia (Adnan, 2014). Despite its environmental and economic benefits, biodiesel is lacking polularity, possibly due to the low awareness and understanding of biodiesel, the relatively high palm oil prices as well as the abundant supply of petroleum-based counterparts.

While bioethanol production from the abundant cellulosic by-products of oil palm holds great potential (Alam et al., 2009; Gutiérrez et al., 2009; Jørgensen et al., 2010), the process is yet to be commercially up-scaled in Malaysia, most possibly due to the lack of infrastructure, insufficient research and development activities, high production and recovery costs. Besides bioethanol, biogas is another potential

renewable energy source generated during anaerobic fermentation of palm oil mill effluent (POME). Biogas comprises mainly methane (CH_4), carbon dioxide (CO_2), as well as traces of hydrogen sulphide (Subramaniam et al., 2010). Nevertheless currently less than 10% of palm oil mills in Malaysia are capturing their biogass as an energy source, also largely due to cost and infrastructure limitations (Subramaniam et al., 2010).

Being the largest crop by size and planted area in Malaysia, oil palm has the biggest potential in providing a huge volume of biomass, which includes fronds, trunks, empty fuit bunches, palm kernel shells, mesocarp fibers and POME. The organic by-products from oil palm industries can be converted into value-added products such as animal feeds (Ishida & Abu Hassan, 1997; Kume et al., 1998; Mohamed & Alimon, 2012) and fine chemicals via fermentation processes (Jørgensen et al., 2010; Motta et al., 2014; Ang et al., 2015; Eom et al., 2015; Kresnowatietal.,2016). The bacterium *Ralstonia eutropha* can effectively utilize plant oil and convert into high levels of biodegradable plastic (bioplastic) called polyhydroxyalkanoates (PHA) (Budde et al., 2010). The characteristics of PHA are similar to that of petroleum derived plastics such as polyethylene, polystyrene and polypropylene specifically in terms of amenability to thermal processing and moulding, which makes PHA a potential substitute for conventional plastics (Srivastava & Vishnoi, 2012). Organic acids generated from anaerobic digestion of POME could also be a potential feedstock for the production of PHA (Mumtaz et al., 2010). Anaerobically treated POME was shown to be a renewable carbon source of low cost for the bacterial strain, *Comamonas* sp. EB172 to efficiently produce PHA in both shake flasks and fermenters (Zakaria et al., 2010a, 2010b). Locally, SIRIM Berhad is actively participating in research of innovative bioplastic synthesis and production process of PHA and has successfully developed a commercially scalable automated system capable of transforming the by-products of

palm oil industry into versatile bioplastics (SIRIM Berhad, 2015). In 2011, SIRIM launched its pioneer bioplastics pilot plant located in Shah Alam, Selangor, which uses crude palm kernel oil and POME as feedstock to produce biodegradable plastic materials ("Malaysia's pioneer bioplastics,", 2011). The major challenges in commercialization of bioplastics are the high cost of production and recovery, hence the higher price of the final products compared to the conventional plastics.

Oil palm trunk (OPT) logs obtained during clearing and re-planting of plantation are also being used as a source of wood products for the plywood industry (Sulaiman et al., 2008). In order to reduce the dependence on forest logs, using OPT as an alternative raw material for the plywood industry presents a more renewable, sustainable and green approach at a competitive price. The annual supply of OPT in Malaysia is estimated at around 13.6 million logs based on 100,000 hectare of replanting area per year, which can then be converted into 4.5 million cubic metres of plywood (Mokhtar et al., 2011). The current challenge of the palm plywood industry is to improve the quality of materials produced from palm stems in terms of strength, stability and durability, to compete with rubberwood and other tropical hardwood products.

2.2 African oil palm (*Elaeis guineensis* Jacq.)

2.2.1 Oil palm genome

The genus *Elaeis* consists of two species, i.e., *Elaeis guineensis* which is native to Africa and is most widely used in plantations in Asia and Africa, and *Elaeis oleifera* which is native to South and Central America (Corley & Tinker, 2015). The African oil palm has 32 chromosomes (2n=32) with a genome size of approximately 1.8 Gbp (Madon et al., 1998). The first fully sequenced oil palm genome from an AVROS (Algemene Vereniging van Rubberplanter terOostkust van Sumatra) palm with

pisifera fruit form, comprised 1.535 Gb of assembled sequences (Singh et al., 2013); a more recent draft genome sequence of an elite *dura* palm comprised 1.701 Gb of assembled sequences (Jin et al., 2016). The closely related *Elaeis oleifera* has the same number of chromosomes (2n=32), and the genome size is similar to that of *Elaeis guineensis* (Singh et al., 2013).

The chromosomes have been ordered according to size (Figure 2.1) that correspond to the linkage groups identified by previous genetic mapping (Rajanaidu et al., 1989; Billotte et al., 2005) and fluorescence *in situ* hybridization studies (Castilho et al., 2000). The African oil palm genome has a guanine-cytosine content of 37%, similar to that of its close relative, date palm (Al-Dous et al., 2011). However, oil palm genes are enriched with a much higher guanine-cytosine content (50%). Computational gene prediction has estimated a total of 158,946 oil palm gene candidates (equivalent to 92 Mb of exonic gene space or 5% of the 1.8 Gb genome sequence), of which 34,802 were similar to known proteins at the peptide sequence level with 96% observed in transcriptome data (Singh et al., 2013).

Most large genomes contain high occurrences of repetitive sequences; in this case, more than half of the oil palm genome is constituted by repetitive DNA sequences (Singh et al., 2013). Castilho et al. (2000) has shown that different families of repetitive DNA sequence were distributed at different sites along the oil palm chromosomes, with simple sequence repeats being concentrated in the distal parts of chromosomal arms. Previous studies have identified di- motifs as the most abundant SSRs within available EST data of oil palm (Ting et al., 2010; Tranbarger et al., 2012). EST-SSRs with tri- motifs were the second most abundant class as reported by Ting et al. (2010), while the tetra- motifs were the second most abundant class as reported by Tranbarger et al. (2012). Such discrepancy between studies could be a consequence of: 1) different analytical algorithms/parameters used; and 2) different sources of input sequences used

for analysis. A more recent genome-wide analysis showed that additional to the high repeat content of the unmapped and unassembled contigs, more than 68% of the predicted oil palm gene candidates were also constituted of known retroelements and other transposons, with the most abundant class of repetitive elements being copia (33%) and gypsy (8%) retroelements, followed by other long terminal repeat (LTR) retrotransposons (6%) (Singh et al., 2013). While a high content of repetitive sequences has always been one of the major challenges in sequence alignment and assembly (Treangen & Salzberg, 2011), the current version of oil palm genome may require further improvement in the accuracy of assemblies, particularly the placement of repeats in the correct genome context. The methylation level of retrotransposon elements has been related to the 'mantled' phenotype oil palm fruit (Ong-Abdullah et al., 2015), supporting the hypothesis that epigenetic modifications are among the key mechanisms to maintain the stability of the repeat-rich genome of oil palm. Moreover, the presence of such exceptionally high repeat content of the oil palm genome may reflect extensive gene duplication events in oil palm. Interestingly, 73% of the repeat elements found in *E. guineensis* were absent in *E. oleifera* (Singh et al., 2013), reflecting a significant variation in gene duplication between plant species, even between very closely related species.

The chloroplast genome of oil palm was reported to be 156,973 bp in length and comprise 112 unique genes (Uthaipaisanwong et al., 2012), including 79 protein coding genes, 4 ribosomal RNA genes and 29 tRNA genes. Phylogenetic analysis of the chloroplastic protein-coding genes further supports that oil palm is closely related to date palm (*Phoenix dactylifera*) which also belongs to the Arecaceae subfamily.

Previously, gene discovery from the oil palm genome has been largely achieved by expressed sequence tag (EST) analysis. Collections of EST have been generated for organ-specific expression profiling (Jouannic et al., 2005; Ho et al., 2007) and to

identify genes associated with various developmental processes such as callogenesis and embryogenesis (Low et al., 2008; Lin et al., 2009). This method, however, has some disadvantages; specifically the high sequencing cost and that the ESTs generated represent only partial genomic information. More recently, gene discovery in oil palm has been accelerated by the next-generation sequencing (NGS) technologies. In 2013, the Malaysian Palm Oil Board published the first publicly available oil palm genome (Singh et al., 2013). Prior to and after the release of the oil palm genome sequence, several tissue-specific transcriptome studies have been reported, including those of fruit and seed (Bourgis et al., 2011; Tranbarger et al., 2011; Dussert et al., 2013; Shearman et al., 2013), inflorescence (Shearman et al., 2013; work reported in this thesis), leaf (Lei et al., 2014) and root (Ho et al., 2016), all of which are useful in complementing the genome sequences already available.

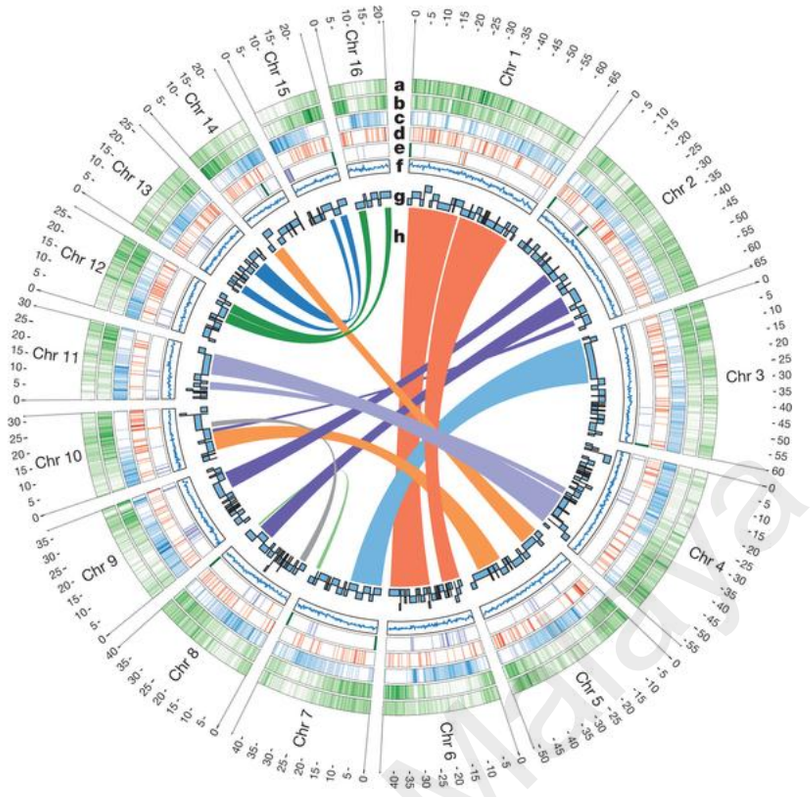


Figure 2.1: Oil palm genome. The oil palm genome is organized into 16 chromosome pairs that were numbered by size. a: gene density; b: methyl-filtered read density; c: retroelement density; d: simple sequence repeats; e: low copy number repetitive elements, including telomere repeat TTTAGGG (green), 5S rRNA (orange) and pericentromeric repeats (purple); f: regional G-C content (range 0.3–0.45); g: genetically mapped scaffolds from the P5-build; h: segmental duplications. Source: Singh et al. (2013)

2.2.2 Growth and development of oil palm

Oil palm is one of the larger palm species which have a single woody stem without secondary growth (Legros et al., 2009). The stem diameter of a mature palm is approximately 60 cm. In a mature oil palm, the new leaf primordia are produced about every two weeks and each takes about two years to develop. About 24 fronds are produced per year with 3-4 m being the average length of each frond. Primary roots of oil palm grow downwards from the base of the palm and radiate outward in a horizontal direction close to the surface of the ground with their length and depth depending on the soil condition (Jourdan et al., 2000).

An oil palm begins to flower when it is 4-6 years old (Cobley & Steele, 1976). As a monoecious plant, each oil palm produces separate female and male inflorescences (Figure 2.2) in alternate cycles throughout the year and occasional hermaphrodite inflorescences during the transition between female and male batches (Biradar, 1978). An inflorescence primordium is formed in the axil of each leaf during leaf initiation, potentially forming either male or female inflorescences (Verheye, 2002). The application of growth regulators (plant hormones) (Corley, 1976b) and stress conditions (Durand-Gasselin et al., 1999) have been reported to affect the sex of oil palm inflorescences. However, the mechanisms underlying regulation of sex determination of an inflorescence is still poorly understood.

The development of inflorescences occurs over two to three years via a series of individual differentiation phases (Corley & Gray, 1976). Each inflorescence can be related to the respective subtending leaves that are referred to as +number (or –number) of the emerged leaf fronds of a palm (Adam et al., 2005). The development of the inflorescence meristem is the longest phase in oil palm reproductive development; it takes about 20 months for the inflorescence meristem to develop and give rise to rachilla bracts, then to the rachilla meristem (Adam et al., 2005).

Each inflorescence comprises a central rachis that is composed of 100-300 rachillae in the male inflorescence, and around 150 rachillae in the female inflorescence, with the rachillae spirally arranged the rachis (Jacquemard, 1995). On both female and male inflorescences, the floral meristems are initiated between leaf stages +2 to +6. On the male inflorescence, each rachilla bears 400-1500 tiny staminate flowers. On the female inflorescence, each rachillae bear larger flowers, borne in 5-30 triads with two accompanying staminate flowers flanking the single pistillate flower which will develop into the fruits that are desired for commercial oil and seed production. Both staminate and pistillate flowers are trimerous, with a perianth distinguished into three sepals and three petals (Dransfield & Uhl, 1998). The staminate flowers each contain a rudimentary gynoecium and six functional stamens, whereas each pistillate flower contains a rudimentary androecium (staminode) and a functional gynoecium (Adam et al., 2005). At leaf stage +10, organ development of the accompanying staminate flowers initiates. At leaf stage +15, the development of reproductive organs in pistillate flowers initiates. By leaf stage +17, the anthers of the accompanying staminate flowers are fully developed. The microsporocytes that are contained within the pollen sacs degenerate and the stalk of the staminate flowers abscises before anthesis. The reproductive organs of the pistillate flowers continue to develop until maturity at leaf stage +18 (Adam et al., 2005).

In oil palm, a floral abnormality known as mantled, affects both *in vitro* culture- (Corley et al., 1986) and seed-derived (Hartley, 1988) palms. The stamens or staminodes of male and female mantled flowers, respectively, are transformed into carpels, which tend to lead to fruit bunch abortion and thus a reduction in palm oil yield. Notably, the long flower developmental phases also reflect a long window for exposure to various environmental influences (Corley, 1976a; Adam et al., 2011). While the formation of oil palm flower structures is known to be influenced by changes in gene

expression (Beulé et al., 2011), it has been suggested that the differential expression of floral genes is under the control of genetic regulatory mechanisms including epigenetic modification (Martin et al., 2009) and post-transcriptional regulation by microRNA (Blazquez et al., 1997; Kim et al., 2007), either independently or though interaction with environmental factors. After pollination, oil palm fruit set and ripening take another five to six months (Verheye, 2002).

Oil palm has an average productive life span of approximately 30 years (Verheye, 2002). As a perennial crop, oil palm can be harvested throughout the year thus ensuring an uninterrupted supply of palm oil. In each productive year, each palm produces 8 to 12 bunches of fruit. The most common varieties of oil palm fruit are *nigrescens* or *virescens* (Hartley, 1988). *Nigrescens* is deep violet to black at the apex and yellow at the base when unripe, and shows minimal color change during ripening. *Virescens* is green when unripe, and shows a distinct color change to reddish-orange during ripening. Generally, the hybrid "*tenera*" palms used in commercial plantations have mature fruit bunches that contain fruit varying in size from 2.5 to 5 cm long and 2 to 4 cm in diameter (Black, 2006). A bunch of fruits can weigh between 10 to 40 kilograms and each fruit contains a single seed (the palm kernel) protected by a wooden endocarp or shell (for *dura* and *tenera* fruit forms), surrounded by a soft oily pulp (the mesocarp). Oil palm fruit uniquely produces two types of oil: one extracted from the mesocarp called the crude palm oil (CPO) and the other from the kernel called the crude palm kernel oil (CPKO). CPO is rich in palmitic and oleic acids whereas CPKO contains predominantly the lauric fatty acids (Singh et al., 2009).

Figure 2.2: (A) Female inflorescence at anthesis after opening of the peduncular bract and prophyll; (B) Detail of female rachillae bearing mature pistillate flowers; (C) Male inflorescence at anthesis; (D) Detail of male rachillae bearing staminate flowers at anthesis. f: flower; p: prophyll; ra: rachilla; St: stigma. Source: Adam et al. (2005)

Oil palm has three main fruit forms, which are classified according to the shell thickness of the fruit: *dura* (3 to 8 mm thick), *tenera* (up to 3 mm thick) and *pisifera* (with no shell) (Figure 2.3A). The 'sex ratio' representing the proportion of female to total inflorescences is highest in *pisifera*, followed by *tenera* and *dura* palms (Sparnaaij et al., 1963). *Dura* and *tenera* are generally fertile, whereas most *pisifera* palms are usually unproductive and partially female sterile. *Tenera* palms that derive from crossings of *dura* and *pisifera* palms have high oil to bunch ratio (more than 20%) (Wahid et al., 2004) and therefore are the preferred commercial planting materials (Wonkyi-Appiah, 1987; Corley & Lee, 1992). Shell-less *pisifera* palms are commonly used as the male parent for breeding as they are homozygous recessive (sh^-/sh^-) for the shell gene. When crossed with a thick-shelled (2-8 mm) *dura* (Sh^+/Sh^+) female parent, the resultant *tenera* hybrid (Sh^+/sh^-) produces thin-shelled (0.5-4 mm) fruits with thicker oil-bearing mesocarp layer (Figure 2.3B) (Jalani et al., 2003; Singh et al., 2013). For optimized yields of commercial planting material, a higher proportion of female inflorescence in the *tenera* hybrid is preferred. For optimized seed production, a higher proportion of female inflorescence in the female *dura* parent over the male *pisifera* parent is preferred. The sex ratio however, can be altered by abortion of the entire inflorescence, a common phenomenon in oil palm that usually occurs at leaf stages +4 to +10 in response to stressful conditions (Adam et al., 2011). Inflorescence abortion is sex-dependent with female inflorescences usually more vulnerable to abortion than male inflorescences (Sparnaaij et al., 1963; Corley et al., 1995). The need for validation of a hybrid progeny performance, as a part of a selection process for elite *dura* and *pisifera* parental breeding materials, is another major reason for the long time period required for crop improvement.

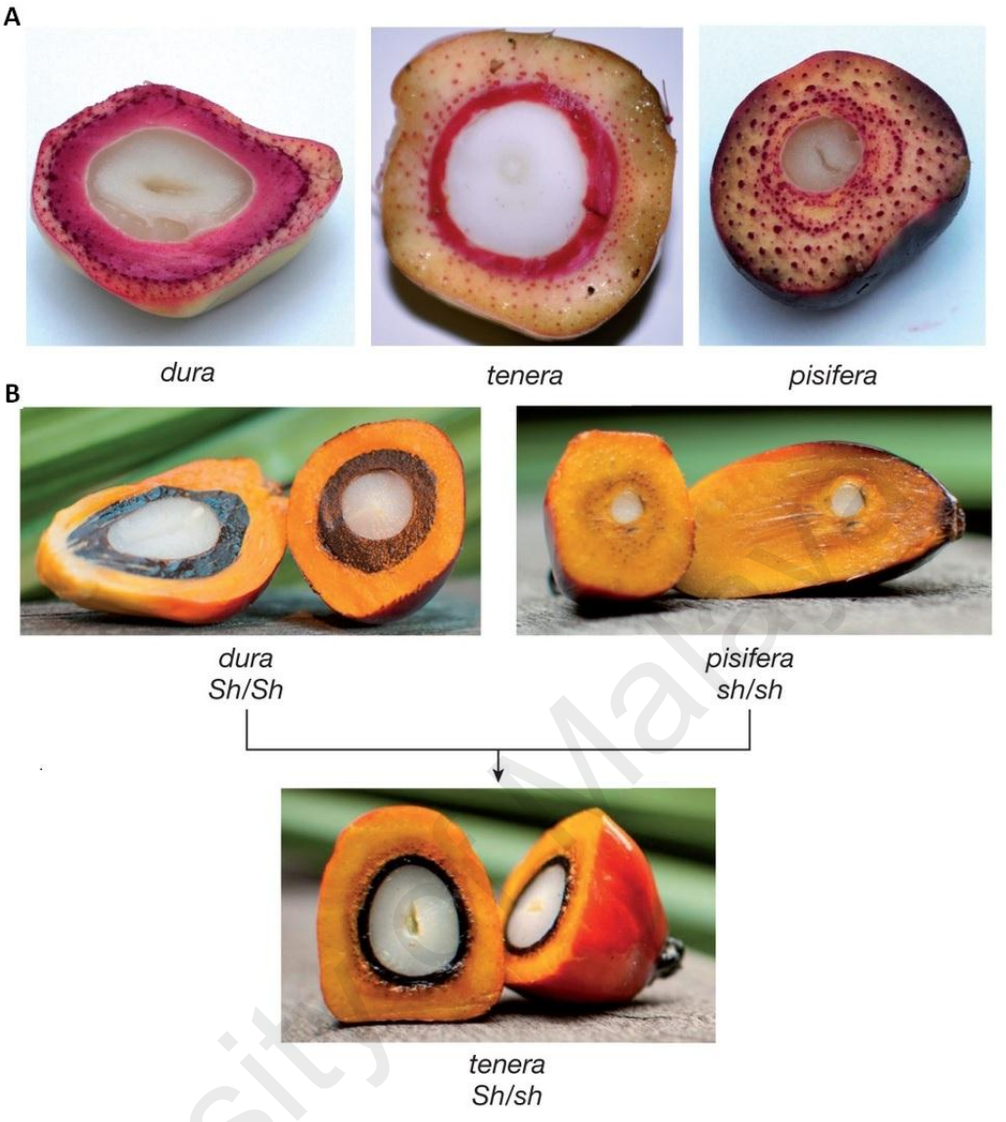


Figure 2.3: (A) Characteristics of *dura*, *tenera* and *pisifera* fruit forms; (B) Crossing between *dura* (Sh^{+}/Sh^{+}) and *pisifera* (sh^{-}/sh^{-}) palms produces the *tenera* (Sh^{+}/sh^{-}) hybrid. Source: Singh et al. (2013)

2.2.3 Molecular regulation of flower development of oil palm

Floral development in plants is under the control of a complex network of genes, of which many encode transcription factors (O'Maoileidigh et al., 2014). Generally, the classical 'ABC model' is used to model the structural development of the floral organs in plants (Coen & Meyerowitz, 1991; Theissen et al., 2000). The expression of a combination of floral homeotic gene(s) from different classes (A-, B- and C- function) collectively contributes to the formation of a specific floral organ identity, with variations occurring in some plant groups (Figure 2.4A). Class A genes specify sepals and together with class B genes they regulate petal formation. Class B and C genes collectively specify stamens. Class C genes alone regulate carpel formation. The 'ABC model' model was then modified into 'ABCDE model' with the additional class D genes specifying ovules (Colombo et al., 1995) and class E genes regulating the formation of all four major floral organs (Pelaz et al., 2000) (Figure 2.4B). In conjunction with A-, B- and C- function genes, class E genes regulate the formation of floral organs hence the 'ABC model' is also known as the 'ABCE model'. In oil palm, SQUAMOSA (SQUA) (A-function), GLOBOSA (GLO) (B-function), DEFICIENS (DEF) (B-function), AGAMOUS (AG) (C/D-functions) and AGAMOUS-like 2 (AGL2) (E-function) genes have been associated with floral organ identity determination, and the structural and functional data of these genes supports the conservation of the generic 'ABCE model' in oil palm, rather than the 'ABCDE model'(Adam et al., 2007a; Adam et al., 2007b).

Except for some class A genes such as APETALA2 of *Arabidopsis thaliana*, all floral organ identity genes identified to date encode MADS-box transcription factor family members (Murai, 2013). The MADS-box genes of angiosperms are extensively diverged to form a superfamily with subfamilies such as SQUA, DEF, GLO, AG, and SEPALLATA (Adam et al., 2006). Oil palm MADS-box gene families were shown to

be involved in the early stages of flowering as well as contributing to the determination of the floral organ identity (Syed Alwee et al., 2006; Adam et al., 2006, 2007b). From the EST collection of oil palm, Ho et al. (2007) identified several flowering-related homologs, including CONSTANS-like, AGL2, AGL20, LFY-like, SQUA, and SQUAMOSA binding protein (SBP). Other than the MADS-box genes, there is still a lack of information of genes likely to be involved in oil palm flower development.

MADS-box genes have been reported to regulate sex determination in several monocots and dicots, including *Rumex acetosa* (Ainsworth et al., 1995), *Betula pendula* (Elo et al., 2001), *Populus deltoides* (Sheppard et al., 2000), *Zea mays* (Heuer et al., 2000) and *Asparagus officinalis* (Park et al., 2003). In all these cases, a decrease in the expression of B- or C- function genes in carpels or stamens was observed in organs targeted for abortion which then led to the formation of unisexual flowers. Although it was previously reported that none of the MADS-box genes were involved in sex determination of oil palm inflorescences (Adam et al., 2007b), Ho et al. (2015) identified a *DEFICIENS* ortholog that was more highly expressed in oil palm male inflorescence, compared to female inflorescence.

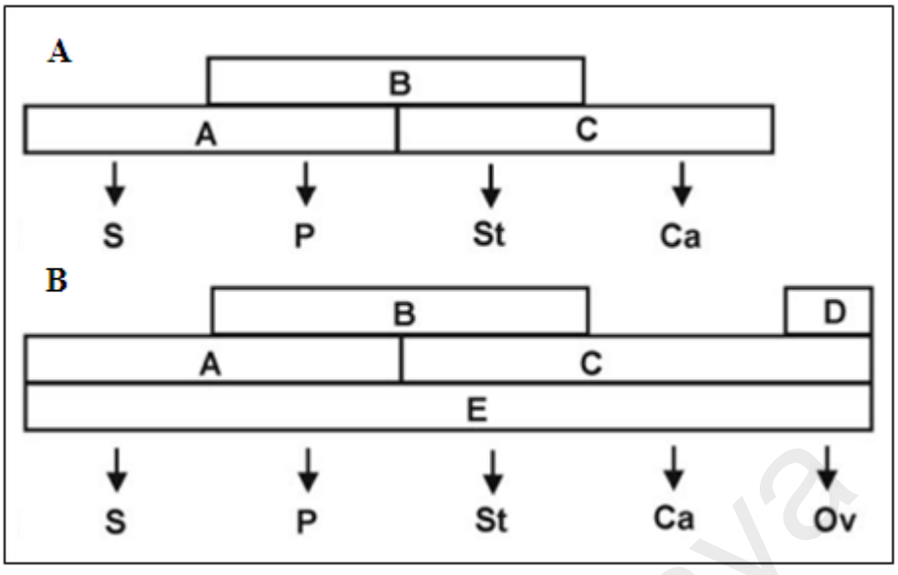


Figure 2.4: (A) Classical ABD model; (B) ABCDE model of floral organ identity in core eudicots. S: sepal; P: petal; St: stamen; Ca: carpel; Ov: ovule. Source: Coen & Meyerowitz (1991); Theissen (2001)

The 'mantled' phenotype of oil palm, is an apparent feminisation of male organs in female and male flowers (Corley et al., 1986) which apparently results from somaclonal variation during large-scale propagation of oil palm (Rival et al., 1999; Jaligot et al., 2000). Somaclonal variation describes epigenetic changes, which may be induced by stress inherent in cellular reprogramming during tissue culture regeneration (Morcillo et al., 2006). The abnormality in mantled oil palms, will lead to partial or complete sterility, resulting in less fruit production (Beule at al., 2011). Mantled flower abnormality has been observed in 5-10% of somatic embryo-derived regerated palms (Rival et al., 2008). Previously it has been suggested that the mantled phenotype may result from epigenetic regulation of MADS-box genes (Jaligot et al., 2011). Mantled phenotypes were also reported to be associated with differential expression of a complex network of genes, including genes from methylation (Shearman et al., 2013) and proteolytic signaling (Beulé et al., 2011) pathways. Interestingly, Rival et al. (1998) reported that mantled oil palms were able to revert to the normal phenotype over a period of years in the field, with the degree of the recovery depending on the severity of

the initial mantled abnormality. The mechanism of this reversion is, however, still unconfirmed although increased DNA methylation rate and a change in hormonal supplies have been suggested to be involved (Durand-Gasselin et al., 1990; Jaligot et al., 2000). A more recent study by Ong-Abdullah et al. (2015) on the other hand has shown that the loss of DNA methylation at *Karma* element in the intron of the homeotic gene *DEFICIENS*, and the resultant *Karma*-mediated alternative splicing, contribute to the mantled abnormality of oil palm. It remains to be confirmed whether the mantled abnormality is a consequent to the production of truncated DEFICIENS protein, accumulation of an aberrantly-spliced *DEFICIENS* transcript, a reduction in *DEFICIENS* transcript levels or by a combination of all these factors. Besides the role in flower development, the expression of many MADS-box genes was also linked to the different phases of mesocarp development, suggesting a functional divergence of MADS-box proteins in ripening oil palm fruit (Tranbarger et al., 2011).

2.3 MicroRNAs (miRNAs)

MicroRNAs (miRNAs) are short (generally 21-24 nucleotides), single-stranded and non protein-coding RNAs which derive from long stem-loop precursors transcribed from endogenous genomic DNA by RNA polymerase II (Bartel, 2004; Lee et al., 2004). The first miRNAs (lin-4 and let-7) were discovered as key regulators of developmental timing and cellular differentiation in *Caenorhabditis elegans* (Lee et al., 1993; Wightman et al., 1993). MiRNAs regulate gene expression by binding to their target mRNA and cause degradation or translational silencing of that mRNA in a sequence-specific manner. In plants, some miRNA are also processed into short interfering RNA (siRNA), via an RNA-dependent RNA polymerase, triggering transcriptional gene silencing (Ruiz-Ferrer & Voinnet 2009; Chellappan et al., 2010). Each miRNA may

target from one, up to several different, mRNA sequences, all of which must have complementarity to the mature miRNA sequence (Pasquinelli, 2012)

2.3.1 Biogenesis of miRNAs in plants

Plant miRNAs are predominantly 21 nucleotides in length (Voinnet, 2009). The biogenesis of plant miRNAs is well-characterized in *Arabidopsis thaliana* (Figure 2.5). Most plant genomes contain more than 100 miRNA genes (*MIR*) (Nozawa et al., 2012), most of which are located in intergenic regions (Reinhart et al., 2002). *MIR* genes are transcribed by DNA-dependent RNA polymerase II (pol II) (Bartel, 2004). This non-protein-coding RNA transcript is capped, spliced, polyadenylated and folded into a long hairpin stem-loop primary miRNA (pri-miRNA). Nuclear cap-binding proteins, CBP20 and CBP80, are added to the 5' terminal pri-miRNA to facilitate the maturation of miRNA (Kim et al., 2008; Laubinger et al., 2008). The role of DAWDLE (DDL) protein was proposed to stabilize pri-miRNA and facilitate the recognition by an RNase III named Dicer-like 1 (DCL1) enzyme (Yu et al., 2008). Interestingly, three miRNA families, miR822, miR839 and miR869 are processed by DCL4 instead of DCL1, possibly due to some structural differences in their hairpin precursors (Rajagopalan et al., 2006; Ramachandran & Chen, 2008; Amor et al., 2009; Zhang et al., 2010). DCL1 cleavage first takes place in a specialized subnuclear region called a D-body, in the presence of HYPONASTIC LEAVES1 (HYL1), SERRATE (SE) and TOUGH (TGH) proteins to form a shorter hairpin structure known as a precursor miRNA (pre-miRNA) (Chen, 2005; Lobbes et al., 2006; Yang et al., 2006; Zhang et al., 2006b; Fang & Spector, 2007; Song et al., 2007; Dong et al., 2008). Next, DCL1 further trims the pre-miRNA into a shorter double-stranded RNA (dsRNA) forming a miRNA: miRNA* duplex. The 3′ termini of the initial miRNA: miRNA* duplex are then methylated by the nuclear HUA ENHANCER1 (HEN1) protein (Yu et al., 2005) before it is

transported out from the nucleus via HASTY, a plant homolog of the mammalian Exportin-5 (Papp et al., 2003). In the cytoplasm, together with cytoplasmic ARGONAUTE (AGO) proteins (Park et al., 2005), the active strand of the miRNA duplex called the mature miRNA loads onto a RNA-induced silencing complex (RISC), which is then guided to its complementary target mRNA and directs translational inhibition or cleavage of the target mRNA transcript (Chen, 2005; Yang & Li, 2012).

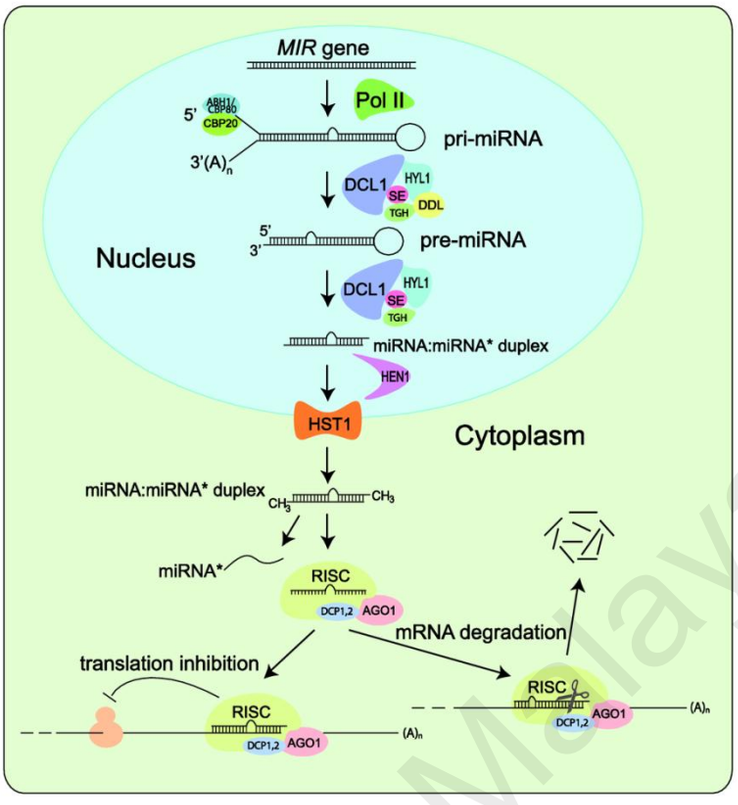


Figure 2.5: Biogenesis pathway of miRNA in plants. *MIR* genes are transcribed by RNA polymerase II (pol II). The transcript is capped by CBP20 and CBP80 and polyadenylated. Next, the pri-miRNA is associated with DDL and subsequently cleaved by DCL1 with the help of HYL1, SE and TGH, producing a pre-miRNA. Pre-miRNA is further cleaved by DCL1, releasing a duplex consisting of miRNA/miRNA* which is later methylated by HEN1. Then, the duplex is transported via HASTY to the cytoplasm and the mature miRNA is incorporated into the AGO1-containing RISC that directs translational inhibition or claevage of the target mRNA transcript. AGO: ARGONAUTE; CBP: CAP-BINDING PROTEIN; DDL: DAWLDLE; DCL1: DICER-LIKE1; HYL1: HYPONASTIC LEAVES1; SE: SERRATES; HEN1: HUA ENHANCER1; RISC: RNA-INDUCED SILENCING COMPLEX; TGH: TOUGH. Source: Yang & Li (2012)

2.3.2 Plant miRNAs and their functions

The first plant microRNA was discovered in Arabidopsis (Reinhart et al., 2002). Following that, the discovery and naming of most plant miRNAs have been based on nucleotide similarity with their homologs in Arabidopsis and rice. These evolutionarily conserved miRNAs can be found in species from different lineages together with several recently diverged non-conserved miRNAs which are species-specific (Barakat et al., 2007; Subramanian et al., 2008; Sunkar et al., 2008; Lelandais-Briere et al., 2009; Zhang et al., 2010a). In contrast to animal miRNAs which are conserved for most of the

complete precursor sequence, only the mature miRNA and its passenger sequence miRNA star (miRNA*) are conserved in plants while the hairpin precursors (pre-miRNAs) vary significantly (Bartel, 2004).

MiRNAs are classified into gene families according to the sequence of the mature miRNAs, which will define a miRNA's target and function. The members of a miRNA family contain an identical or nearly identical mature miRNA sequence that differ at most at four positions, and each of the family members usually derive from a highly variable pre-miRNA sequences (Griffiths-Jones et al., 2006; Jones-Rhoades et al., 2006). In some exceptional cases, the pre-miRNA sequences have been shown to be conserved between dicots and monocots as reported for miR159 and miR319 (Palatnik et al., 2007). In accordance to miRBase criteria and conventions for miRNA identification and naming, at least one member of each family has to be validated experimentally, while the others may be obtained by bioinformatic predictions based on sequence similarity of the mature miRNA sequence and the ability of the flanking region to fold into a characteristic hairpin structure (Ambros et al., 2003).

In plants, the mechanism of miRNA-dependent control of gene expression was described as post-transcriptional gene silencing (PTGS) (Llave et al., 2002; Bartel, 2004). Most plant miRNAs either directly or indirectly regulate the expression of genes that encode transcription factors, stress response proteins and other proteins required for growth, development and physiology of plants (Rogers & Chen, 2013), by binding to their highly complement target sequence within coding regions and occasionally in the 3' untranslated region (3' UTR), at multiple binding sites (Bonnet et al., 2004; Jones-Rhoades et al., 2004; Sunkar et al., 2005; Jones-Rhoades et al., 2006). Other than regulating gene expression via translational inhibition and/or mRNA cleavage, previous studies have reported that small RNAs of varying sizes, namely MIR-derived siRNAs (Chellappan et al., 2010) and long miRNAs (lmiRNAs) (Wu et al., 2010), are able to

direct DNA methylation at loci from which they are produced as well as in *trans* at their target genes in Arabidopsis and rice. More recently, miRNAs have been reported to regulate epitranscriptomic modification of mRNAs in animal cells (Chen et al., 2015a), further complicating the mechanisms of miRNA-mediated regulation.

With respect to flower development, miRNAs have been reported to affect transition to flowering (Achard et al., 2004; Aukerman & Sakai, 2003; Spanudakis & Jackson, 2014), sex determination (Chuck et al., 2007) and floral organ development (Aukerman & Sakai, 2003; Liu et al., 2010; Zhu et al., 2009) via cleavage of target mRNAs (Guo et al., 2005; Li et al., 2014) and inhibition at the translation level (Chen, 2004; Li et al., 2013). In oil palm, a previous study used a combined homology and structural analysis approach to computationally identify orthologous miRNAs from the EST sequences derived from various tissue types (Md Nasaruddin et al., 2007). Multiple precursor paralogs of oil palm miR172 have been reported to show tissue-specific expression patterns, some of which have been predicted to cleave the APETALA2-like targets (Mehrpooyan et al., 2012). Different families of miRNAs have also been reported to be differentially expressed in developing mesocarp tissues of oil palm fruit, with some of the miRNAs proposed to be related to the regulatory pathway of fatty acid metabolism (Fang et al., 2013). More recently, in an analysis of gene models in *E. guineensis* and *E. oleifera*, the authors reported 40 miRNAs homologous to previously known miRNAs; several of these were found in at least one of the two species (Low et al., 2014).

2.3.3 Approaches for miRNA discovery

The initial discovery of the first miRNA in *C. elegans* was achieved through genetic screens (Lee et al., 1993; Wightman et al., 1993). As a general discovery method, however, genetic screening is cost-ineffective, time consuming and is limited

by the samples selected, as the expression level of miRNAs often varies significantly across different tissues and/or under different environmental conditions (Lai, 2003; Zhang et al., 2006c). To date, only a few plant miRNAs have been discovered by genetic screens, including miR319 that regulates leaf morphogenesis (Palatnik et al., 2003) and miR164c that regulates flower development in Arabidopsis (Baker et al., 2005).

Alternatively, plant miRNAs have been identified via direct cloning and sequencing of small RNAs (Llave et al., 2002; Park et al., 2002; Reinhart et al., 2002; Sunkar et al., 2005). Such small RNA data is searched for sequence and structural homologies with previously reported miRNAs. A major challenge of this method is annotation of non-conserved miRNA families that have not been previously described, as a majority of the clones obtained may represent degradation products of rRNAs, tRNAs and snRNAs, resulting in a low discovery rate compared to the cloning and sequencing cost incurred.

MiRNA identification using bioinformatics tools has also contributed significantly to the prediction of new miRNAs in both animal and plant systems. Computational prediction of miRNAs generally involves the use of algorithms which are based on the sequence similarity to known miRNA families and the ability of the corresponding genomic loci of the miRNA to fold into an extended hairpin structure as RNA (Lagos-Quintana et al., 2001; Lau et al., 2001; Jones-Rhoades, 2004; Adai et al., 2005). The secondary stucture of miRNA precursors is highly conserved from species to species (Axtell & Bartel, 2005; Zhang et al., 2006c) and has a low minimum folding free energy (Zhang et al., 2006d). Some mRNA folding programs which are commonly used to predict the hairpin precursors of miRNAs include mfold (Zuker & Jacobson, 1998; Waugh et al., 2002; Zuker, 2003), RNAfold (Zuker & Stiegler, 1981; McCaskill,

1990; Hofacker et al., 1994), Sfold (Ding & Lawrence, 2003; Ding et al., 2004), and UNAFold (Markham & Zuker, 2008).

Various studies have successfully identified miRNAs by computational prediction from EST, transcriptome and genome datasets for several plant species (Jones-Rhoades & Bartel 2004; Wang et al., 2004a; Adai et al., 2005; Lindow & Krogh 2005; Zhang et al., 2005; Jin et al., 2008; Yin & Shen, 2010); each adopting different prediction algorithms. Notably, this method is only useful to identify homologs of a known miRNA which is highly conserved across the plant kingdom (Jones-Rhoades & Bartel 2004; Wang et al., 2004a; Axtell & Bartel, 2005; Sunkar et al., 2005; Zhang et al., 2006b). Table 2.1 summarizes the results from previous studies, which report computationally predicted miRNAs in various plants, including several important crop species.

Table 2.1: Summary of studies on approaches used to identify plant miRNA genes and targets in different plant species. [a, b] refer to the number of miRNAs and miRNA targets reported in each study.

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Arabidopsis thaliana (Arabidopsis)	PatScan	Complementarity between miRNAs and their target	14	49	Rhoades et al., 2002
	MIRcheck	Comparative genomics	92	19	Jones- Rhoades & Bartel, 2004
	MIRFINDER	Comparative genomics	91	58	Bonnet et al., 2004
	Genome-wide analysis	Comparative genomics	83	371	Wang et al., 2004a
	EST analysis	Homology search	17	-	Zhang et al., 2005
	findMiRNA	Comparative genomics	37	-	Adai et al., 2005
	WUBLAST	Short and perfectly	96	102	Li & Zhang, 2005
	-	Homology search	20	-	Li et al., 2005
	-	Complementarity between miRNAs and their target	592	656	Lindow & Krogh, 2005
	454 Life Science	Comparative genomics	38	-	Rajagopalan et al.,
	EST analysis	Homology search	21	1	Zhang et al., 2006a
	Illumina sequencing	Principal component analysis	113	-	Wang et al., 2011
Arachis hypogaea (Peanut)	Illumina sequencing	Comparative genomics	14	30	Zhao et al., 2010

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Camellia sinensis (Tea)	EST analysis	Homology search	4	30	Prabu & Mandal, 2010
Capsicum annuum (Chili)	EST analysis	Homology search	3	-	Zhang et al., 2005
	EST analysis	Homology search	3	-	Zhang et al., 2006a
Elaeis guineensis (Oil palm)	EST analysis	Homology search	5	-	Md Nasaruddin et al., 2007
	454 Life Science	Homology search	3	-	Shearman et al., 2013
	MatureBayes	Homology search	8	1	Low et al., 2014
Glycine max (Soybean)	EST analysis	Homology search	45	-	Zhang et al., 2005
	-	Complementarity between miRNAs and their target	21	-	Dezulian et al., 2005
	EST analysis	Homology search	57	-	Zhang et al., 2006a
	EST analysis	Homology search	82	72	Subramanian et al., 2008
	EST analysis	Homology search	32	31	Wang et al., 2009a
	Illumina sequencing	Comparative genomics	87	603	Joshi et al., 2010
	Illumina sequencing	Comparative genomics	236	38	Kulcheski et al., 2011
	Illumina sequencing	Homology search	306	1,219	Li et al., 2011
	Illumina sequencing	Comparative genomics	26	126	Song et al., 2011
	Illumina sequencing	Comparative genomics	636	15	Sun et al., 2016

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Glycine max (Soybean)	Illumina sequencing	Homology search	437	898	Xu et al., 2016
Gossypium hirsutum (Cotton)	EST analysis	Homology search	1	-	Zhang et al., 2005
	EST analysis	Homology search	1	-	Zhang et al., 2006a
	EST & GSS analysis	Homology search	37	96	Qiu et al., 2007
	EST analysis	Homology search	30	139	Zhang et al., 2007
	EST analysis	Homology search	15	-	Khan Barozai et al.,
	Genome-wide analysis	Comparative genomics	25	223	Pang et al., 2009
	Illumina sequencing	Homology search	34	-	Ruan et al., 2009
	Illumina sequencing	Homology search	60	556	Yang et al., 2013
	Illumina sequencing	Homology search	264	567	An et al., 2015
	Illumina sequencing	Homology search	198	107	Zhang et al., 2015a
Hordeum vulgare (Barley)	EST analysis	Homology search	17	-	Zhang et al., 2005
	EST analysis	Homology search	24	1	Zhang et al., 2006a
	-	Complementarity between miRNAs and their target	8	22	Dryanova et al., 2008
	EST analysis	Homology search	28	445	Kantar et al., 2010
	EST analysis	Homology search	156	121	Colaiacovo et al., 2010

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Jatropha curcas (Nettlespurge)	Direct cloning & sequencing	Homology search	52	28	Wang et al., 2012a
	EST and GSS analysis	Homology search	24	78	Vishwakarma & Jadeja,
	Illumina sequencing	Homology search	196	57	Galli et al., 2014
Musa acuminata (Banana)	Illumina sequencing	Homology search	151	815	Bi et al., 2015
	EST & GSS analysis	Homology search	32	121	Chai et al., 2015
	Illumina sequencing	Homology search	170	173	Ghag et al., 2015
	Illumina sequencing	Homology search	181	108	Lee et al., 2015
Nicotiana benthamiana (Tobacco)	EST analysis	Homology search	4	-	Zhang et al., 2005
	EST analysis	Homology search	5	-	Zhang et al., 2006a
	Genome-wide analysis	Comparative genomics	65	1,225	Frazier et al., 2010
	Illumina sequencing	Homology search	356	15	Qi et al., 2012
Oryza sativa (Rice)	-	Homology search	40	-	Li et al., 2005
	EST analysis	Homology search	25	-	Zhang et al., 2005
	EST analysis	Homology search	40	1	Zhang et al., 2006a
	miRanda	Global computational analysis	34	70	Archak & Nagaraju,
	454 Life Science	Comparative genomics	39	95	Zhu et al., 2008
	454 Life Science	Comparative genomics	23	20	Sunkar et al., 2008
	PHRED	Comparative genomics	6	-	Lacombe et al., 2008

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Oryza sativa (Rice)	SplamiR	Complementarity between miRNAs and their target	37	-	Thieme et al., 2011
	Illumina sequencing	Homology search	341	602	Baldrich et al., 2015
	Illumina sequencing	Homology search	143	110	Cheah et al., 2015
	Illumina sequencing	Homology search	1,177	698	Bakhshi et al., 2016
	Illumina sequencing	Homology search	393	399	Lian et al., 2016
	Illumina sequencing	Homology search	799	5,575	Xu et al., 2017
Phoenix dactylifera L. (Date palm)	Genome-wide analysis	Comparative genomics	81	199	Xiao et al., 2013
	SOLiD™ System 4.0	Homology search	238	874	Xin et al., 2015a
	Illumina sequencing	Homology search	333	1,071	Yaish et al., 2015
	Genome-wide analysis	Comparative genomics	183	731	da Silva et al., 2016
Physcomitrella patens (Moss)	EST analysis	Homology search	2	-	Zhang et al., 2005
	EST analysis	Homology search	2	-	Zhang et al., 2006a
	MIRcheck	Comparative genomics	205	29	Axtell et al., 2007
	EST analysis	Homology search	48	59	Fattash et al., 2007
	EST analysis	Gene-oriented strategy	25	19	Wan et al., 2011
Solanum lycopersicum (Tomato)	EST analysis	Homology search	9	-	Zhang et al., 2005
	EST analysis	Homology search	9	1	Zhang et al., 2006a
	EST analysis	Homology search	21	57	Yin et al., 2008

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Solanum lycopersicum (Tomato)	EST analysis	Homology search	13	23	Zhang et al., 2008
	-	Comparative genomics	11	3	Itaya et al., 2008
	454 Life Science	Comparative genomics	4	62	Moxon et al., 2008a
	miRCat	Comparative genomics	20	7	Mohorianu et al., 2011
	Illumina sequencing	Homology search	33	-	Gao et al., 2015a
	Illumina sequencing	Homology search	174	54	Jin & Wu, 2015
	Illumina sequencing	Homology search	1,067	856	Cheng et al., 2016
	Illumina sequencing	Homology search	753	362	Zhou et al., 2016a
Solanum tuberosum (Potato)	EST analysis	Homology search	13	-	Zhang et al., 2005
	EST analysis	Homology search	17	-	Zhang et al., 2006a
	EST analysis	Homology search	22	75	Guo et al., 2007
	EST & GSS analysis	Homology search	48	186	Zhang et al., 2009a
	EST analysis	Homology search	202	1,094	Xie et al., 2011
Sorghum bicolor (Sorghum)	-	Complementarity between miRNAs and their target	64	-	Dezulian et al., 2005
	EST analysis	Homology search	14	-	Zhang et al., 2005
	EST analysis	Homology search	17	-	Zhang et al., 2006a
	Genome-wide analysis	Comparative genomics	82	-	Paterson et al., 2009
	miRDeep	Comparative genomics	9	14	Calviño et al., 2011

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Sorghum bicolor (Sorghum)	-	Complementarity between miRNAs and their target	64	-	Dezulian et al., 2005
	EST & GSS analysis	Homology search	31	72	Katiyar et al., 2012
	Illumina sequencing	Homology search	623	1,300	Katiyar et al., 2015
Triticum aestivum (Wheat)	EST analysis	Homology search	19	-	Zhang et al., 2005
	EST analysis	Homology search	31	1	Zhang et al., 2006a
	EST analysis	Homology search	23	46	Yao et al., 2007
	-	Complementarity between miRNAs and their target	17	31	Dryanova et al., 2008
	GenomicSVM	Homology search	79	59	Jin et al., 2008
	EST analysis	Homology search	34	46	Yin & Shen, 2010
	Illumina sequencing	Homology search	78	26	Tang et al., 2012a
	Illumina sequencing	Homology search	873	1,282	Meng et al., 2013
	Illumina sequencing	Homology search	323	524	Sun et al., 2014
	Illumina sequencing	Homology search	239	4,151	Zhou et al., 2015
	Illumina sequencing	Homology search	256	1,988	Chu et al., 2016
	Illumina sequencing	Homology search	201	629	Song et al., 2017

Table 2.1, continued

Plant	Approach/Programme	Principle	[a]miRNA	[b]miRNA target	References
Zea mays (Maize)	-	Complementarity between miRNAs and their target	57	-	Dezulian et al., 2005
	EST analysis	Homology search	23	-	Zhang et al., 2005
	GSS analysis	Homology search	188	115	Zhang et al., 2006e
	EST analysis	Homology search	30	1	Zhang et al., 2006a
	EST & GSS analysis	Homology search	11	26	Chen et al., 2009
	Genome-wide analysis	Comparative genomics	61	247	Zhang et al., 2009b
	SplamiR	Complementarity between miRNAs and their target	14	-	Thieme et al., 2011
	Illumina sequencing	Homology search	142	278	Sheng et al., 2015
	Illumina sequencing	Homology search	75	105	Xin et al., 2015b
	Illumina sequencing	Homology search	55	137	Wu et al., 2016
	Illumina sequencing	Homology search	422	3,945	Zhou et al., 2016b
	Illumina sequencing	Homology search	399	2,431	Xu et al., 2017

More recently, small RNA cloning and sequencing techniques were adapted for 'second-generation' DNA sequencing (also known as the next-generation sequencing, NGS) technologies (Goodwin et al., 2016) and has significantly accelerated the discovery and expression profiling of miRNA from plants (summarised in Table 2.1). NGS platforms offer high-throughput sequencing capacity at remarkably lower cost, compared to the conventional Sanger sequencing. Currently, several examples of leading next-generation DNA sequencing systems are delivered by 454 Life Sciences (Roche), Illumina, Inc., and Thermo Fisher Scientific Inc. (Table 2.2). Each of these sequencing platforms is characterized by distinct sequencing principles, read length, sequencing throughput, output data etc. Suzuki et al. (2011) has evaluated the accuracy and sequence bias of three next-generation sequencing platforms, i.e., Genome Sequencer FLX System (454 Life Sciences, Roche), Genome Analyzer (Illumina, Inc.), and SOLiD system (Thermo Fisher Scientific Inc.), by analyzing quality of sequence reads obtained from *Escherichia coli* strain DH1. Glenn (2011) has summarized the major characteristics of all commercially available second and third generation sequencing platforms. The output of most NGS platforms is as a large volume of short reads. When analyzing genomes that contain high occurrences of repeats, the presence of repetitive sequences may create ambiguities in sequence alignment and *de novo* assembly of short reads (Treangen & Salzberg, 2011). While several strategies have been utilized to resolve mis-assemblies resulting from repetitive sequences (Treangen & Salzberg, 2011), the more promising solution seems to be the use of longer read lengths that can span repeats (Cahill et al., 2010; Ummat & Bashir, 2014; Rhoads & Au, 2015), as offered by some of the third-generation sequencing technologies.

The latest developments of the high-throughput sequencing industry are known as the 'third-generation' sequencing systems, which are differentiated by single-molecule resolution, long reads, short run time and lower overall cost, providing the flexibility to conduct sequencing projects in different modes depending on the need

(Heather & Chain, 2016). In contrast to NGS systems that are mostly based on sequencing by synthesis (SBS) technologies, the 'third-generation' sequencers allow direct sequencing of single DNA molecules, such that the synchronization step is no longer required thereby eliminating the biases introduced by PCR amplification and dephasing (Whiteford et al., 2009). These new sequencing systems are also expected to exhibit enhanced capacity performance along with improvement in the data quality and the types of data produced at a much reduced cost (Levy & Myers, 2016). Nevertheless, the overall cost, ease of use, capacity performance, post-sequencing data storage and analysis remain as the determining factors for assessing the best application of these sequencing systems.

Table 2.2: Advances in DNA sequencing technologies. Note that the applications listed here are those relevant to small RNA discovery and profiling. The features and specifications are correct at the time of writing and are subject to improvement or revision. *The product is currently under development.

Sequencing platform	Company	Sequencing technology	Read length	Sequencing application						
				Genome	Metagenome	Exome	RNA	sRNA	Degradome	Epigenome
Second-generation sequencing platform										
Genome Sequencer Junior System Genome Sequencer FLX+ System	454 Life Sciences (Roche)	Pyrosequencing in high-density picoliter reactors	400 bp - 1 kb	✓	✓	✓	✓	✓	✓	✓
HiSeq Series HiSeq X Series MiniSeq System MiSeq Series NextSeq Series	Illumina, Inc.	Sequencing by synthesis (SBS) of single-molecule arrays with reversible terminators	150 - 300 bp	✓	✓	✓	✓	✓	✓	✓

Table 2.2, continued

Sequencing platform	Company	Sequencing technology	Read length	Sequencing application						
				Genome	Metagenome	Exome	RNA	sRNA	Degradome	Epigenome
Second-generation sequencing platform										
Ion Personal Genome Machine™ Ion PGM™ System Ion Proton™ System	Thermo Fisher Scientific Inc.	Incorporation of nucleotides is recorded by measuring the hydrogen ions released during nucleotide chain elongation and the signal is translated into digital information on a semiconductor chip	200 - 400 bp	✓	✓	✓	✓	✓		✓
Polonator Genome Analyzer	Azco Biotech	Sequencing by ligation of polony libraries	26 bp	✓			✓			
MaxSeq™	Intelligent Biosystems	SBS chemistry	35 - 55 bp	✓			✓			
SOLiD™ Systems	Thermo Fisher Scientific Inc.	Massively parallel sequencing by hybridization-ligation	50 - 75 bp	✓	✓	✓	✓	✓	✓	✓

Table 2.2, continued

Sequencing platform	Company	Sequencing technology	Read length	Sequencing application						
				Genome	Metagenome	Exome	RNA	sRNA	Degradome	Epigenome
Third-generation sequencing platform										
Helicos® Genetic Analysis System	Helicos BioSciences	Sequencing single molecule using SBS chemistry	25 - 60 bp	✓			✓	✓	✓	
GridION System*	Oxford Nanopore Technologies	Uses nanopores for the direct, electronic analysis of single molecules including DNA, RNA, proteins and other molecules	>5 kb	✓			✓			
Sequel™ System PacBio RS II	Pacific Biosciences	Single molecule, real-time, or SMRT™, detection of biological processes	>1,000 bp	✓	✓	✓	✓			✓

2.3.4 Plant miRNA target prediction

The major challenge in understanding miRNA functions is to identify their regulatory mRNA target(s). In plants, the characteristic of the perfect or near perfect complementarity between miRNA and target site (Pillai, 2005) has been employed as the key feature for target gene analysis. Several web-based and locally-installed tools have been developed to predict miRNA targets in plants (Table 2.3), each with different optimization thresholds, hence inconsistencies in the performance (Srivastava et al., 2014). Analysis of complementarity requirements for miRNA targeting in *Nicotiana benthamiana* revealed that complementarity between the miRNA 5' regions and the target sites is critical for repression efficacy (Liu et al., 2014a). It is therefore the primary criterion for most, if not all miRNA target prediction tools to 'score' the degree of complementarity between miRNA and mRNA target according to their customized scoring schema. As many of the currently available tools are being trained and optimized based on datasets derived from model plant species in particular, the default parameters may not always be suitable as thresholds for analysis of datasets from non-model plants (Srivastava et al., 2014).

Table 2.3: Computational tools for plant miRNA target prediction. [a] web-based server; [b] local installation required; [c] no longer available; [1]miRNA-target complementary; [2]target site conservation; [3]free energy of hybridization; [4]accessibility of target site; [5]multiplicity of target sites

Tool	Prediction criteria				URL	Reference
	Comp[1]	Cons[2]	Hyb[3]/ Acc[4]	Mul[5]		
PatScan[b]	*				N/A	Dsouza et al., 1997
MicroInspector[a]	*		*		http://bioinfo1.uni-plovdiv.bg/cgi-bin/microinspector/	Rusinov et al., 2005
miRU[c]	*	*			N/A	Zhang, 2005
miRNAassist[b]	*			*	N/A	Xie et al., 2007
IntaRNA[a]	*		*	*	http://rna.informatik.uni-freiburg.de	Busch et al., 2008; Wright et al., 2014
UEA sRNA toolkit (plant version)[a,b]	*		*		http://srna-tools.cmp.uea.ac.uk/plant/ http://srna-workbench.cmp.uea.ac.uk	Moxon et al., 2008b; Stocks et al., 2012
WMD3[a,b]	*		*		http://wmd3.weigelworld.org/cgi-bin/webapp.cgi	Ossowski et al., 2008
TAPIR[a]	*	*	*		http://bioinformatics.psb.ugent.be/webtools/tapir/	Bonnet et al., 2010
Targetfinder[b]	*	*			https://github.com/carringtonlab/TargetFinder	Fahlgren & Carrington, 2010
Target-align[c]	*				N/A	Xie & Zhang, 2010

Table 2.3, continued

Tool	Prediction criteria				URL	Reference
	Comp[1]	Cons[2]	Hyb[3]/ Acc[4]	Mul[5]		
MultiMiTar [a,b]	*	*			http://www.isical.ac.in/~bioinfo_miu/multimitar.htm	Mitra & Bandyopadhyay, 2011
psRNATarget[a]	*	*	*	*	http://plantgrn.noble.org/psRNATarget/	Dai & Zhao, 2011
p-TAREF[b]	*	*	*		http://scbb.ihbt.res.in/new/p-taref/index.html	Jha & Shankar, 2011
miRTour[a]	*	*			http://bio2server.bioinfo.uni-plovdiv.bg/miRTour/	Milev et al., 2011
Target_Prediction[b]	*		*		N/A	Sun et al., 2011
imiRTP[b]	*	*	*	*	http://admis.fudan.edu.cn/projects/imiRTP.htm	Ding et al., 2012
C-mii[b]	*		*		http://www.biotec.or.th/isl/c-mii	Numnark et al., 2012
psRobot[a,b]	*	*		*	http://omicslab.genetics.ac.cn/psRobot/	Wu et al., 2012
miRNA target plot[b]	*				N/A	Abhinand et al., 2013
iMir[b]	*	*	*	*	http://www.labmedmolge.unisa.it/inglese/research/imir	Giurato et al., 2013
comTAR[a]	*	*	*		http://rnabiology.ibr-conicet.gov.ar/comtar	Chorostecki & Palatnik, 2014

2.3.5 Degradome sequencing

Confident identification of a novel plant miRNA requires both the computational prediction of the target(s), usually based on either perfect or nearly perfect sequence complementarity between a miRNA and its target mRNA(s), followed by experimental validation of the target(s). Typically a single miRNA may regulate from one up to several target mRNA sequences, however only a minority of these predicted miRNA targets have been verified, mostly by 5' Rapid Amplification of cDNA Ends (RACE) in a small scale basis (Alves-Junior et al., 2009; Jiao et al., 2011; Chorostecki et al., 2012; Shamimuzzaman & Vodkin, 2012; Song et al., 2012; Wan et al., 2012; Xia et al., 2012).

Modified from 5'-RACE, Degradome Sequencing (Degradome-Seq) is a parallel analysis of RNA ends (PARE) that runs on high-throughput DNA sequencing platforms (Table 2.2). Degradome-Seq allows global analyses of RNA decay by producing data from the sequencing the 5' ends of uncapped RNAs. In plants, upon base pairing of the 5' end of the miRNA with the target mRNA, the cleavage of target mRNA usually occurs at the position between the 10th or 11th nucleotide from the 5' end of the miRNA (Palatnik et al., 2003; Jones-Rhoades & Bartel, 2004). The endonuclease activity of RISC-associated AGO protein cleaves target mRNA and generates a free 5'-monophosphate remaining on the 3′ fragment to which an RNA adaptor can be ligated. The ligation products are subsequently enriched by reverse transcription and second strand synthesis, prior to high-throughput sequencing and cleavage site mapping. In a degradome library, the tags which represent cleaved mRNA targets often produce the most abundant and distinct peaks at the predicted cleavage site relative to other regions of the transcript (Addo-Quaye et al., 2008; German et al., 2008). By matching the cleavage sites to the corresponding miRNA sequences, this relates miRNAs to their respective mRNA targets.

Degradome-Seq has been adapted to screen miRNA-mediated mRNA degradation by identifying the miRNA cleavage sites in various plants including *Arabidopsis thaliana* (German et al., 2008), *Brassica napus* (Xu et al., 2012a), *Cucumis sativus* (Mao et al., 2012), *Glycine max* (Song et al., 2011; Hu et al., 2013), *Gossypium hirsutum* (Yang et al., 2013) *Hordeum vulgare* (Curaba et al., 2012), *Nicotiana tabacum* (Tang et al., 2012b) and *Triticum aestivum* (Tang et al., 2012a). Data from the degradome (cleaved mRNAs) derived from Arabidopsis and rice has been conveniently collected into online databases (Arabidopsis PARE Database, http://mpss.udel.edu/at_pare and Rice PARE Database, http://mpss.udel.edu/rice_pare/) to facilitate the study of miRNA-target mRNA pairs in these plants (Nakanoet al., 2006). The limitation of Degradome-Seq, however, is that it is restricted to the mRNA targets which are regulated via miRNA-directed cleavage, but not those which are translationally repressed by miRNAs.

To date, a number of prediction programs dedicated for plant degradome analysis have been developed, each of these differ in approach and implementation (Table 2.4). Most of these programs are linked to and exploit information from miRNA, gene, protein and biological pathway resources such as miRBase (http://www.mirbase.org/), Ensembl (www.ensembl.org/), Swiss-Prot (www.ebi.ac.uk/swissprot/), UCSC genome browser (http://genome.ucsc.edu/), KEGG pathway (http://www.genome.jp/kegg/pathway.html), and other databases to allow functional annotation of the predicted targets. Other than transcriptional suppression, widespread translational regulation by miRNAs was also observed in plants (Brodersen et al., 2008; Lanet et al., 2009; Yu & Wang, 2010), which further complicates miRNA target identification.

Table 2.4: Plant degradome data analysis tools. [a] web-based server; [b] local installation required. SeqTar is not available in a publicly downloadable package and is available for non-commercial purposes upon request.

Server/ Tool	Functions	URL	Reference
CleaveLand[b]	Uses degradome data to find cleaved small RNA targets	http://axtell-lab-psu.weebly.com/cleaveland.html	Addo-Quaye et al., 2008
starBase[a]	Explores miRNA-target interactions from degradome sequencing (Degradome-Seq, PARE) data	http://starbase.sysu.edu.cn/	Yang et al., 2010
SeqTar[b]	Identifying miRNA-guided cleavage sites from degradome of polyadenylated transcripts in plants	-	Zheng et al., 2011
PAREsnip[b]	Identifies genome-wide networks of sRNA/target interactions resulting in transcript cleavage through the degradome	http://srna-workbench.cmp.uea.ac.uk	Folkes et al., 2012
SoMART[a]	A collection of web server tools to identify miRNAs or tasiRNAs that potentially regulate genes of interest	http://somart.ist.berkeley.edu	Li et al., 2012
sPARTA[b]	A tool for plant miRNA target prediction and PARE validation	http: //mpss.udel.edu/tools/mirna apps/download.php	Kakrana et al., 2014
StarScan[a]	An integrated web-based tool to identify sRNA targets from degradome sequencing data	http://mirlab.sysu.edu.cn/starscan/	Liu et al., 2015
MTide[b]	An integrated tool for exploring miRNA-target interaction in plants	http://bis.zju.edu.cn/MTide	Zhang et al., 2015b

CHAPTER 3

MATERIALS AND METHODS

3.1 Oil palm samples

The Malaysian Palm Oil Board (MPOB) provided the oil palm female inflorescence tissue samples for the small RNA, mRNA and degradome (parallel analysis of RNA ends (PARE)) sequencing. Each of the small RNA, mRNA and degradome libraries was constructed using RNA pooled from equal amounts of RNA (20 µg) extracted from each of the female inflorescences from 2 palms and at the respective leaf stages (i.e. +6 and +15 leaf stages). Sime Darby Technology Centre Sdn Bhd provided additional oil palm female inflorescence tissue samples for the quantitative RT-PCR experiments. All the oil palm female inflorescences were sampled from *Elaeis guineensis* Jacq. *tenera* hybrid palms (*Dura* x *Pisifera*). The female inflorescence tissue samples provided by Sime Darby Technology Centre Sdn Bhd were collected on the same day from 6 different, twenty-year-old oil palms from a single estate in Carey Island, Selangor, Malaysia. The female inflorescences were sampled by removing the leaves from the felled palms then dissecting the crown to reach the tissues, which were immediately placed in liquid nitrogen. The maturity stages of the inflorescences were determined by the length of the inflorescence, together with the correlation of the maturity of the flower to their axillary leaves (Adam et al., 2005). In this study, +6 leaf stage female inflorescences were 3-4cm in length while +15 leaf stage female inflorescences were 15cm in length. The numbering attributed to inflorescence stage corresponds to the number of the inflorescence's subtending emerged leaves. In inflorescences at the +6 leaf stage, the new floral meristem arises, and subsequently develops into triads of two staminate flowers flanking a pistillate flower (Adam et al., 2007a). In inflorescences at the +15 leaf stage, the floral triads are composed of two staminate flowers bearing six initiated stamens

fully covered by the perianth, and a pistillate flower composing of only the perianth (Adam et al., 2007a).

3.2 Total RNA extraction

A CTAB method (Murray & Thompson, 1980) was modified to isolate total nucleic acid from oil palm tissue samples. The CTAB extraction buffer was supplemented with PVP-40 to remove poly-phenolic compounds (Liu et al., 1998; Kiefer et al., 2000). The use of lithium chloride was excluded from the protocol to avoid the loss of low molecular weight RNA, which was the product of interest.

First, the mortar and pestle were pre-chilled in liquid nitrogen. Approximately 0.5 g of oil palm tissue was ground to a fine powder in liquid nitrogen using the pre-chilled mortar and pestle. The extraction buffer contained 2% (w/v) CTAB, 100 mM Tris-Cl (pH 8.0), 2% (w/v) PVP, 2 M NaCl, 25 mM EDTA (pH 8). A 2% (v/v) β-mercaptoethanol (Sigma-Aldrich, USA) was freshly added to the extraction buffer just before use. The ground tissue was mixed well in 1 ml of extraction buffer on a vortex mixer. An equal volume of phenol-chloroform-isoamylalcohol (PCI) (25:24:1; v/v/v) was added to the mixture and mixed well by vortex.

The mixture was centrifuged at 10,000 *g* for 10 min. The upper layer of clear supernatant was transferred into a new 2 ml tube. The extraction with PCI (25:24:1; v/v/v) was repeated until no precipitate was found at the inter-phase layer. After the last extraction, an equal volume of chloroform-isoamylalcohol (24:1; v/v) was added to the clear supernatant to remove the residual PCI. The mixture was mixed well by repeated inversion before it was centrifuged at 10,000 *g* for 10 min. The clear supernatant (upper layer) was transferred into a new 1.5 ml tube which was then precipitated overnight in an equal volume of pre-chilled isopropanol at -80°C.

The overnight precipitation product was centrifuged (10,000 *g*) at 4°C for 15 min. The supernatant was discarded and the pellet was air-dried for 30 min. The pellet was then dissolved in 50 µl of RNase-free water. The nucleic acids were separated on a 2% (w/v) agarose gel. The integrity of the RNA was measured by using Bioanalyzer (Agilent Technologies, USA). The minimum RNA quality required was RNA integrity number, RIN ≥ 8.

3.3 DNase treatment

RQ1 RNase-free DNase (Promega Corporation, Madison, WI, USA) was used to remove DNA contamination in the total RNA samples according to the manufacturer's protocols.

3.4 Small RNAs isolation from total RNAs

Total RNAs were separated on a 5% (w/v) low melting preparative grade agarose gel (Promega Corporation, Madison, WI, USA). The electrophoresis was performed at 50 V for 2 hr. The gel was stained in ethidium bromide solution (4 µg/ml) (Promega Corporation, Madison, WI, USA) for 5 min. A fragment of gel containing small RNAs (15-35 nucleotides according to a 10 bp DNA marker) was excised and eluted overnight in 500 µl 0.3 M NaCl at 4°C. After the overnight elution, the gel fragment was discarded and the elution product was then precipitated overnight in an equal volume of pre-chilled isopropanol at -80°C (Ho et al., 2006). The overnight precipitation product was centrifuged (10,000 *g*) at 4°C for 15 min. The supernatant was discarded and the pellet was air-dried for 30 min. The pellet was then dissolved in 30 µl RNase-free water.

3.5 Small RNA profiling

3.5.1 cDNA library construction using hybrid adapters

A small RNA cDNA library was constructed using the method as described by Ho et al. (2006) (Figure 3.1). The hybrid adapters (Integrated DNA Technologies, Inc., USA) are: 5' adapter (5'-N6ACCCTCTTGGCACCCACTaaa-3') and 3' adapters (5'-uuuACCAGGCACCCAGCAATGN3-3') in which the lower case is RNA; upper case is DNA; N6 is an amino 6-carbon spacer for blocking the 5' end; N3 is an amino modifier for blocking the 3' end.

The hybrid adapters were ligated to the small RNAs using T4 RNA ligase (Promega Corporation, Madison, WI, USA) and the ligation product was amplified by RT-PCR. Each reaction tube contained 1 U AMV Reverse Transcriptase (Promega Corporation, Madison, WI, USA), 1 U *Tfl* DNA Polymerase (Promega Corporation, Madison, WI, USA), 1X AMV/*Tfl* Reaction Buffer, 1 mM $MgSO_4$, 2 mM dNTP mix, 1 µM forward primer (5' AGTGGGTGCCAAGAGGGT 3') (Integrated DNA Technologies Pte. Ltd., Singapore), 1 µM reverse primer (5' CATTGCTGGGTGCCTGGT 3') (Integrated DNA Technologies Pte. Ltd., Singapore) and 1 µg RNA sample. First strand cDNA synthesis was performed by incubating the reaction mixture at 45^{o}C for 45 min and followed by thermal inactivation of reverse transcriptase at 94^{o}C for 2 min. Next, the synthesis of second strand cDNA was carried out according to the following temperature settings (Table 3.1). Step 1-3 were repeated for 30 cycles.

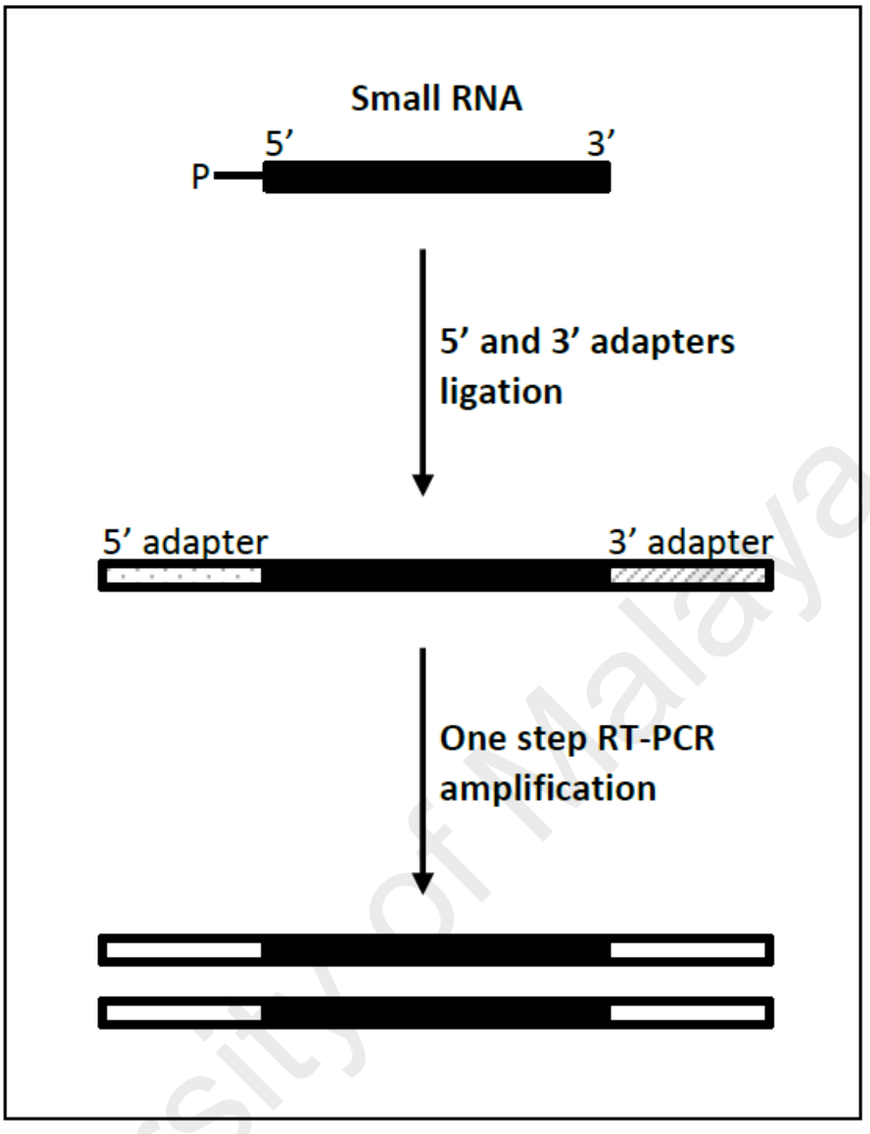


Figure 3.1: Small RNA cDNA construction. 5’ and 3’ adapters are hybrid adapters. Source: Ho et al. (2006)

Table 3.1: Temperature settings for PCR amplification

Step	Temperature	Duration
1. Denaturation	94^{o}C	30 s
2. Annealing	50^{o}C	1 min
3. Extension	68^{o}C	2 min
4. Final extension	68^{o}C	7 min
5. Soak	4^{o}C	30 min

The cDNA was then cloned into pGEM-T Easy vector (Promega Corporation, Madison, WI, USA) according to the manufacturer's protocols. Transformation into competent *Escherichia coli* (DH5α strain) was carried out by adding 10 µl of the ligation products to competent *E. coli*, and the mixture was incubated on ice for 30 min. The heat shock process was carried out by placing the mixture tube in a water bath at 42^{o}C for 90 s. The tube was immediately placed on ice for 2 min. 700 µl of Luria broth (LB) was added into the microcentrifuge tube. The tube was incubated at 37^{o}C with shaking at 250 rpm for one and a half hr. The transformed cells were plated out on LB agar plates containing 50 µg/ml of ampicillin and 50 µl of X-gal (50 mg/ml). The plate was incubated overnight at 37^{o}C. The bacterial colonies were selected based on blue/white screening.

Using PCR, white colonies were screened to identify the clones which contained insert of the desired size. The primers used for PCR screening were M13 forward (5' CGCCAGGGTTTTCCCAGTCACGAC 3') and reverse (5' TCACACAGGAAACAGCTATGAC 3´) primers (Integrated DNA Technologies Pte. Ltd., Singapore). The cells from each white colony were picked and dipped into each respective 0.2 ml PCR tube. Each 50 µl PCR reaction mixture contained 1X Taq reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix, 0.2 µM forward M13 primer, 0.2 µM reverse M13 primer and 0.5 U Taq DNA polymerase. The PCR products were separated on a 1% (w/v) agarose gel at 120V for 30 min. The clones containing an

insert of desirable size were re-cultured and the plasmid was extracted using Wizard® Plus Minipreps DNA Purification System (Promega Corporation, Madison, WI, USA) according to the user manual. The insert-containing plasmids were sequenced using Sanger sequencing. The small RNA sequences were analyzed using bioinformatics tools via databases of NCBI GenBank (http://www.ncbi.nlm.nih.gov/), Sanger miRBase (http://www.mirbase.org/) and Sanger Rfam (http://rfam.sanger.ac.uk/).

3.5.2 cDNA library construction by polyadenylation of 3' end of the small RNAs

A cDNA library was also constructed by polyadenylating the 3' end of small RNAs before ligating the 5' adapter to small RNAs. Polyadenylation was performed by incubating the small RNA sample and poly(A) polymerase (Takara Bio Inc., Japan) at 37°C for 30 min. Reaction buffer for the poly(A) polymerase contains 50 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 250 mM NaCl, 1 mM DTT, 0.05% bovine serum albumin (BSA), 400 µg/ml tRNA and 0.1 mM [^{3}H]ATP.

The polyadenylated small RNAs were recovered by phenol-chloroform extraction and isopropanol precipitation (as described in section 3.2). The purified polyadenylated small RNAs were then ligated with a 5' adapter (5'-N6ACCCTCTTGGCACCCACTaaa-3') (Integrated DNA Technologies, Inc., USA) using T4 RNA Ligase (Promega Corporation, Madison, WI, USA). The reaction mixture was prepared according to the user manual and was incubated at 37°C for 30 min. Next, first strand cDNA was synthesized using Reverse Transcriptase (Promega Corporation, Madison, WI, USA) with a poly-d(T) adapter as RT primer (5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(T)30) (Integrated DNA Technologies Pte. Ltd., Singapore). After adding the reverse transcriptase enzyme, the reaction mixture was incubated at 45^{o}C for 45 min and followed by 94°C for 2 min. Subsequently, PCR was carried out using primers targeting 5' adapter and seed

sequence on the poly-d(T) adapter. PCR amplification was carried out according to the temperature settings as shown in Table 3.1.

The PCR products were separated on a 1% (w/v) agarose gel. PCR products with a size of about 100 bp were purified and ligated into pGEM-T Easy vector (Promega Corporation, Madison, WI, USA) according to the user manual. Next, all the ligation product (10 µl) was transformed into competent *Escherichia coli* (DH5α strain) by heat shock. The tube which contained the mixture of competent cells and the ligation product was placed in a water bath at 42°C for 60 s. The tube was immediately placed back on ice for 2 min. 700 µl of LB broth was added into each tube. The tubes were incubated at 37°C with shaking at 250 rpm for 2 hr. 100 µl of transformed cells were plated out on LB plates containing 50 µg/ml of ampicillin and to which 50 µl of X-gal had been spread onto the gel for blue-white colony selection. The plates were incubated overnight at 37°C. White bacterial colonies were selected for PCR screening (section 3.5.1). The clones which carried an insert of the desirable size were re-cultured and the plasmid was extracted using Wizard® Plus Minipreps DNA Purification System (Promega Corporation, Madison, WI, USA) according to the user manual. The insert-containing plasmids were sequenced using a Sanger sequencing service provided by First BASE Laboratories Sdn Bhd, Malaysia.

3.5.3 Small RNA high-throughput sequencing

Small RNA samples were processed and sequenced using Illumina (Solexa) next generation high-throughput DNA sequencing technology at the Beijing Genomics Institute (BGI), Shenzhen, China (http://www.genomics.cn/index.php).

3.5.3.1 Small RNA cDNA library construction and sequencing

Purified small RNAs (18-30 nucleotides) were ligated to 5' and 3' adapters (Figure 3.2) using T4 RNA ligase (Promega Corporation, Madison, WI, USA). The ligated products were amplified by RT-PCR and sequenced using a Hiseq 2000 (Illumina Inc., San Diego, CA).

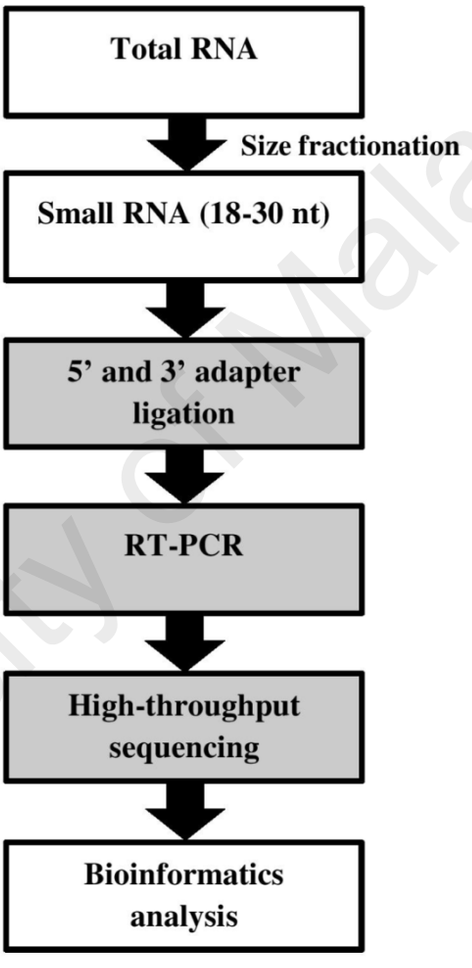


Figure 3.2: General workflow of small RNA cDNA library construction for sequencing using Illumina high-throughput sequencing platform. The white-filled boxes indicate the procedures that were performed using workstations and technical supports in the University of Malaya and the grey-filled boxes indicate the procedures that were performed by BGI.

3.5.3.2 Small RNA sequencing output

The output reads generated from the Illumina GA IIx (Illumina Inc., San Diego, CA) were filtered to remove the adapter sequences used for small RNA cDNA library construction, contaminant sequences and low quality (< 20 Q score) sequences (containing ambiguous nucleotides, 'N') (these preliminary data filtering procedures were performed by BGI). The following bioinformatics analyses were performed using workstations and technical supports in the University of Malaya. The sequences ranging from 18-30 nucleotides were collected for the analysis of size distribution. Short Read Mapping Package (SHRiMP version 2.2.3, http://compbio.cs.toronto.edu/shrimp/) (Rumble et al., 2009) was used to map the filtered reads to Rfam 11.0 (http://rfam.xfam.org/) (Griffiths-Jones et al., 2005) with no mismatch allowed. The sequences that matched rRNA, tRNA, small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) in Rfam were removed from the subsequent analyses. Sequences less than 19 or more than 24 nucleotides were also removed. The remaining sequences were called the clean reads.

3.5.3.3 Functional annotation of small RNAs

To annotate plant orthologous miRNAs, the clean reads were used to search against the non-redundant orthologous plant miRNA reference set from the Plant microRNA Database (PMRD June 11, 2012 update, http://bioinformatics.cau.edu.cn/PMRD/) (Zhang et al., 2010b). An in-house Python script (Appendix A) was used to align the clean reads to the non-redundant orthologous plant miRNA reference set, with only one mismatch allowed for each alignment. The orthologous miRNA matches were named according to the orignal gene name except for the part that indicates the species name, which was replaced with egu-, denoting *Elaeis guineensis*. Nucleotide bias at each position of the orthologous miRNAs was

determined using an online RNA base composition calculator (http://www.currentprotocols.com/WileyCDA/CurPro3Tool/toolId-7.html).

To annotate oil palm-specific miRNAs, the remaining small RNA sequences (after annotation of plant orthologous microRNAs) were used to search against the oil palm reference genome (Singh et al., 2013) and the transcriptomes produced from oil palm female inflorescences (section 3.6.3). The criteria applied for the annotation of oil palm-specific miRNA were: (i) transcripts (unigenes) form a stem-loop structure of 75 nucleotides with a bulge-loop size less than 6 nucleotides; (ii) small RNA reads were located within the stem region of the stem-loop structure; (iii) a maximum of 3 mismatches was allowed between the miRNA:miRNA* duplex; (iv) miRNA and miRNA* formed a duplex with 3' overhangs; (v) predicted minimum folding energy (MFE) ranged from -15 kcal/mol to -47.2 kcal/mol. Using Bowtie (Langmead et al., 2009) in the miRDeep2 tool (Friedlander et al., 2012), the small RNA sequences were mapped to the oil palm reference genome and transcriptomes. RNAfold (Hofacker, 2003) was used to predict the stem-loop structure from the sequences of 300 nucleotides spanning the matched small RNA sequences in the reference genome and transcriptomes. The p-values were calculated for the miRNA precursors predicted by miRDeep2 using Randfold (Bonnet et al., 2004). The oil palm-specific miRNAs were arbitrarily named starting at '1' according to the miRBase gene nomenclature (Griffiths-Jones et al., 2006).

3.5.3.4 MicroRNA quantification and differential expression

For all small RNA libraries, the miRNA read counts were normalized to tags per million (TPM) as follows:

Normalized expression, TPM

= (miRNA read count/total clean read) x 1,000,000

DEGSeq (Wang, et al., 2010) was used to calculate the differential expression changes of miRNAs, using cutoff values of Log_2fold change ≥1 and false discovery rate (FDR) value < 0.05. gplots R package (http://cran.r-project.org/web/packages/gplots/index.html) was used to plot heatmaps to show the differential expression changes.

3.5.3.5 MicroRNA target prediction

The reference transcriptome produced from oil palm female inflorescences was used as reference data to predict miRNA targets. psRNAtarget online server (http://plantgrn.noble.org/psRNATarget/) (Dai & Zhao, 2011) with default parameters were used to predict miRNA targets.

3.6 Transcriptome sequencing

The RNA samples were processed and sequenced using Illumina (Solexa) next generation high-throughput DNA sequencing technology at Beijing Genomics Institute (BGI), Shenzhen, China (http://www.genomics.cn/index.php).

3.6.1 Transcriptome cDNA library construction and sequencing

The cDNA library construction and sequencing were performed by BGI. Purified mRNA samples extracted from the oil palm female inflorescence samples, i.e. +6 (female inflorescence at +6 leaf stage) and +15 (female inflorescence at +15 leaf stage) were sequenced using an Illumina HiSeq^{TM} 2000 (Illumina Inc., San Diego, CA)

according to the manufacturer's protocols. Poly(A) mRNA was isolated from 20 µg of total RNA using Sera-mag Magnetic Oligo (dT) Beads (Illumina Inc., San Diego, CA). mRNA was fragmented into short sequences of 100 to 400 nucleotides at 94°C for 5 min in 1 X Fragmentation Buffer (Ambion, Inc. Austin, TX, USA). Random hexamer primers (Illumina Inc., San Diego, CA) and SuperScript II reverse transcriptase (Invitrogen™, Life Technologies, Thermo Fisher Scientific Corporation, Waltham, MA, USA) were used to synthesize the first-strand cDNA from the fragmented short sequences. The second-strand cDNA was synthesized in reaction buffer containing RNaseH, DNA polymerase I and dNTPs. The cDNA was purified using QIAQuick PCR extraction kit (QIAGEN, Hilden, Germany) and was then end-repaired, phosphorylated and 3'-adenylated using Illumina's kit according to the manufacturer's protocols, before ligating to Illumina sequencing adapters. After gel electrophoresis, cDNA fragments 200 ± 25 bp were excised from agarose gel and were enriched by 15 cycles of PCR amplification. Both ends of the libraries were sequenced using an Illumina HiSeq™ 2000 (Illumina Inc., San Diego, CA).

3.6.2 Transcriptome sequencing output

The90-bp paired-end raw reads were filtered to remove adapter sequences, low quality reads with ambiguous sequences ('N') and sequence reads containing more than 10% bases with Phred score, Q<20. The remaining sequences called the clean reads were provided in FASTQ format by BGI. The following bioinformatics analyses were performed using workstations and technical supports in University of Malaya.

3.6.3 *De novo* sequence assembly

SOAPdenovo V1.05 (http://soap.genomics.org.cn/soapdenovo.html) (Li et al., 2010) was used to assemble the clean reads into contigs, scaffolds and unigenes (Figure 3.3). Different kmer sizes ranging from 25 to 63 were tested using the 63mer version of SOAP denovo V1.05. Contigs represented the longest assembled sequences containing no Ns (unknown nucleotides). The reads were mapped back to contigs and by combining paired-end information, contigs were linked into scaffolds. Scaffold length was estimated according to average segment length of each pair of reads. Unknown bases were filled with Ns. Paired-end reads were used to fill the gaps in scaffolds by using GapCloser V.12 for SOAPdenovo (http://soap.genomics.org.cn/soapdenovo.html). After filling gaps in scaffolds using paired-end reads, the product sequences were referred as unigenes. TGI clustering tool V2.1 (Pertea et al., 2003) was used to assemble all the unigenes from the two samples (+6 and +15 female inflorescences) to form a single set of non-redundant reference transcriptome. Unigene sequences that were less than 100 nucleotides in length were discarded. Unigene sequences were aligned with NCBI Blast Database using blastx (E-value <1E-6). Sequence orientations were determined according to the best hit in the database. Orientation and coding DNA sequence (CDS) of sequences which have no hit in blast were predicted using ESTScan (Iseli, 1999).

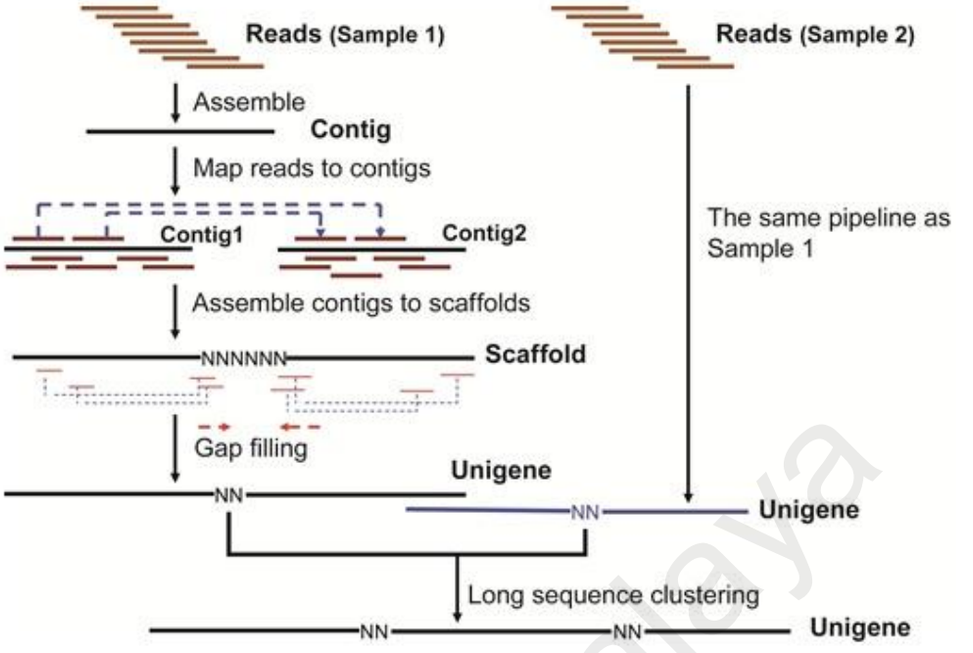


Figure 3.3: *De novo* assembly of short reads into contigs, scaffolds and unigene. Source: Huang et al. (2012)

3.6.4 Functional annotation of transcripts

Using blastx with an E-value <1E-6, the *de novo* assembled unigenes were annotated by similarity searches against the publicly available reference protein databases, namely NCBI non-redundant (nr) protein database (http://www.ncbi.nlm.nih.gov/protein), Swiss-Prot protein database (http://www.ebi.ac.uk/uniprot), NCBI COGdatabase (http://www.ncbi.nlm.nih.gov/COG/), Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/), and Gene Ontology (GO) database (http://www.geneontology.org/). With nr annotation, Blast2GO (Conesa et al., 2005) was used to obtained the GO annotation. WEGO software (Ye et al., 2006) was used to perform GO functional classification for all unigenes and to understand the distribution of gene functions of the species from the macro level.

In order to validate the quality of the current assembly and annotation, the unigenes were compared against a list of 85 plant housing keeping genes (Yue et al., 2010), using blastx with an E-value <1E-6. These randomly selected house keeping

genes are conserved in many plants, including *O. sativa*, *A. thaliana*, *M. truncatula*, *V. vinifera* and *P. trichocarpa*.

3.6.5 Quantification of differentially expressed transcripts

Library normalization was conducted prior to expression comparison of transcripts between the two transcriptome libraries (+6 and +15 female inflorescences). Discovery rate (FDR) cut-off < 0.05 and Log_2 ratio ≥ 1 were used as threshold values to select for differentially expressed transcripts. Transcript counts were normalized to transcripts per million (TPM) (Audic & Claverie, 1997). An R package, DEGSeq (version 1.15, http://bioinfo.au.tsinghua.edu.cn/software/degseq) (Wang et al., 2010) was used to identify differentially expressed transcripts based on read count. The transcripts normalization and differential gene expression were calculated using an R script. Only transcripts with at least 10 read counts present in the library were used for transcript quantification. False discovery rate (FDR) cut-off < 0.05 and Log_2 ratio ≥ 1 were used as threshold values to select for differentially expressed transcripts.

3.6.6 Identification of simple sequence repeats (SSR)

The method and parameters described in Wang et al. (2010) and Franchini et al. (2011) were used to identify simple sequence repeats (SSRs) from unigenes. The assembled unigenes were screened for SSRs using a Perl script known as MIcroSAtellite identification tool (MISA, http://pgrc.ipk-gatersleben.de/misa/misa.html). The minimum lengths for SSRs were at least 8 for dinucleotide, 12 for trinucleotide, 16 for tetranucleotide, 20 for pentanucleotide and 24 for hexanucleotide repeats.

3.7 Identification of sex-specific transcripts in female and male inflorescences

This section of work (methods described throughout section 3.7) was performed in collaboration with Sime Darby Technology Centre Sdn Bhd. A consensus 454 sequence assembly named EGrefseq, was built by combining the sequences derived from male inflorescence, female inflorescence, apical meristem, mesocarp, leaf and root of oil palm. The following bioinformatics analyses were performed using workstations and technical support (Ranganath Gudimella and Joel Zi-Bin Low) at University of Malaya and at Sime Darby Technology Centre Sdn Bhd.

The reads generated by Roche 454 GS-FLX were used as input to the Newbler program package V2.5 (454 Life Sciences, Roche Diagnostics Corporation, Branford, CT, USA). Sequences originating from organelles and rRNA were removed. The assembly process used default settings with the addition of the '-urt' option that assembles transcripts with low-read coverage. An in-house Perl scripts (Ho et al., 2015) was used to remove redundant sequences and sequences smaller than 200 bp. The resulting assembled sequences were known as isotigs or contigs. The Newbler program was used to cluster the isotigs into isogroups. The notion of a gene corresponds to an isogroup, while the splice variants were represented as isotigs or contigs. In the following sections, both contigs and isotigs are referred to as transcripts.

3.7.1 Functional annotation of EGrefseq sequences

Using BLASTX with an E-value threshold of 1E-10, the EGrefseq sequence functions were annotated using the Swiss-Prot database (October 2013 release, http://www.expasy.ch/sprot), the TAIR10 database (ftp://ftp.arabidopsis.org/home/tair/), the RGAP 7 database (http://rice.plantbiology.msu.edu/) and the KEGG database (http://www.genome.jp/kegg/). The best match (minimum length of 200 bp) with the highest score in bits was used to annotate each transcript.

3.7.2 Quality assessment EGrefseq sequences

The reliability of our assembly was evaluated by adapting a computational algorithm, Core Eukaryotic Genes Mapping Approach (Parra et al., 2007). This uses a defined set of 458 core eukaryotic genes (CEG) that are highly conserved across a wide range of eukaryotic taxa to evaluate the completeness of a genome or annotations and in this case our transcriptomes.

3.7.3 Transcriptome reference alignment and sample read counts

The 454 reads were mapped to EGrefseq using bwasw of the BWA package (Li & Durbin, 2010). Uniquely mapped reads were counted from the reference alignment using SAMtools (Li et al., 2009) in each sample dataset.

3.7.4 Differential transcript expression analysis

Transcript expression profiles of oil palm male (Mi) and female (Fi) inflorescences were compared and scored by the presence or absence of a transcript in a sample dataset. A transcript was considered unique when it had at least 10 counts in one sample and a zero count in the other sample.

3.7.5 Prediction of open reading frame

The longest open reading frame (ORF) was predicted for each transcript using OrfPredictor (Min et al., 2005). As the sequencing was performed on a non-directional basis, ORF search was performed on both strands of the transcripts and only the longest frame was considered. The standard codon table was used, with ATG as the start codon. The predicted ORFs were annotated using BLAST2GO (V2.6.6) (Conesa & Götz, 2008) against GenBank nr (non-redundant) protein database with the E-value cut-off of 1E-5. The annotation was refined using Inter-ProScan, ANNEX, Enzyme Code and KEGG annotations.

3.7.6 Validation of 454 expression data

An nCounter analysis system (NanoString Technologies, Seattle, WA, USA) was used to determine the expression level of 40 sex-specific transcripts in six tissues, i.e., male inflorescence, female inflorescence, shoot apical meristem, mesocarp, leaf and root of oil palm. These 40 transcripts were randomly selected from a total of 97 potential sex-specific transcripts (Table 5.2), all of which had ≥10 counts in at least one of the tissues studied. Target-specific oligonucleotide probes were designed using an in-house software (nDesign Gateway) developed by NanoString and were synthesized by Integrated DNA Technologies, Inc., USA (Appendix C, Supplementary Table C1).

In order to obtain sufficient RNA for analysis, pooled RNA samples were used. The RNA for each tissue type was extracted from six different palms (of the same genetic background as the palms used for the generation of 454 transcriptome libraries described in section 3.7), and was then combined in equal amounts. A transcript was considered inflorescence-specific when it showed an inflorescence-predominant expression pattern (Log_2-transformed fold-change ratios ≥0.50). A transcript was considered sex-specific when it showed a sex-predominant expression pattern (Log_2-transformed fold-change ratios ≥0.50) between female and male inflorescences. Only the transcripts that had at least 10 counts in at least one of the oil palm tissues were analyzed. Pre-mRNA splicing factor SLU7 (*SLU7*) and glutaredoxin genes that had been shown to stably express in different oil palm tissues (Yeap et al., 2013) were used as the housekeeping gene controls. Standard negative and positive controls were spiked into the samples according to the manufacturer's protocol. Four technical replicates for each tissue type were used. The raw counts were normalized using the geometric means of the positive controls and the two housekeeping genes in the nSolver Analysis Software provided by NanoString Technologies, Seattle, WA, USA.

3.8 Degradome sequencing

An RNA degradome represents a collection of mRNA-derived degradation products. The aim of the degradome analysis was to identify the target mRNA(s) of a miRNA in a high-throughput scale using the high-throughput sequencing technology. The degraded fragments of mRNA resulting from miRNA cleavage were mapped to the reference transcriptome derived from female inflorescences. Total RNA samples were processed and sequenced using Illumina (Solexa) next generation high-throughput DNA sequencing technology at Beijing Genomics Institute (BGI), Shenzhen, China (http://www.genomics.cn/index.php).

3.8.1 Degradome library construction and sequencing

Using the service provided by BGI, two degradome (parallel analysis of RNA ends, PARE) libraries were constructed: one from the +6 female inflorescences and one from the +15 female inflorescence. Poly(A)-enriched RNA was isolated and was ligated to an RNA oligonucleotide adaptor containing a 3′ MmeI recognition site. The ligation product was used as the template for first-strand cDNA synthesis by reverse transcription (RT) and the cDNA was later amplified using PCR. After purification and digestion with restriction enzyme MmeI, the PCR product was ligated to a double-stranded DNA adaptor, and was then gel purified for PCR amplification. The final cDNA library was purified and sequenced on an Illumina HiSeq^{TM} 2000 (Illumina Inc., San Diego, CA) according to the manufacturer's protocols.

3.8.2 Sequence generation and processing

Illumina's Pipeline V1.5 software was used to remove adaptor sequences and low quality sequencing reads. The following bioinformatics analyses were performed using workstations and technical supports in University of Malaya. The filtered sequencing reads of 20 and 21 nucleotides were used to identify potentially cleaved

targets by using CleaveLand (Brousse et al., 2014). The degradome reads were mapped to the reference transcriptome derived from oil palm female inflorescences. Only the perfect matching alignment(s) for a given read were extended to 35-36 nucleotides by adding 15 nucleotides upstream of the sequence. All resulting reads (t-signature) were reverse-complemented and aligned to the miRNAs identified in this study. A maximum of five mismatches of the alignments were allowed. Alignments where the 5′ degradome sequence nucleotide was coincident with the tenth nucleotide of miRNAs were scored using the method described by Allen et al. (2005). The potential targets were selected and categorized as I, II, or III, as described by Addo-Quaye et al. (2008). Using blastx with an E-value <1E-6, all potential targets were matched to NCBI Genbank (http://www.ncbi.nlm.nih.gov/genbank/) for functional annotation.

3.9 Quantitative PCR analysis of microRNAs

Three biological replicates each made from pooling total RNA extracted from female inflorescences of two palms (i.e. six palms sampled in total) were used for reverse transcription (RT) and quantitative PCR (qPCR) analysis. Stem-loop RT primers for miRNAs were designed and prepared according to the method described by Kramer (2011).

Reverse transcription was performed using SuperScript® III First-Strand Synthesis System (Invitrogen™, Life Technologies, Thermo Fisher Scientific Corporation, Waltham, MA, USA). The cDNA was amplified using qPCR with gene-specific forward primers and universal reverse primer (Table 3.2). A total of 20µl reaction mix was prepared using Power SYBR® Master Mix (Life Technologies, Thermo Fisher Scientific Corporation, Waltham, MA, USA). The reactions were run on a QuantStudio 12K Flex real-time PCR platform (Applied Biosystems®, Life Technologies, Thermo Fisher Scientific Corporation, Waltham, MA, USA) following a protocol of 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min.

Dissociation curves for all target genes were established at the end of PCR cycle at 95°C for 15 s, 60°C for 1 min, followed by 95°C for 15 s. *Cyclophilin 2* (*Cyp 2*) was used as the reference gene for analysis by the $\Delta\Delta C_T$ method (Livak & Schmittgen, 2001). All reactions were run in three technical replicates for each of the three biological replicates.

Table 3.2: Primer sequences used in quantitative RT-PCR analysis. All the primers were synthesized by Integrated DNA Technologies Pte. Ltd., Singapore.

Primer	Sequence 5'-3'
Stem-loop reverse transcription (RT) primer	
miR10	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCCTGA
miR13	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCCTCTC
miR15	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACATACCA
miR156	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGTGCTC
miR160	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTATGCT
miR166	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGAATG
miR167	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTAGATC
miR168	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTTCCCG
miR172	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACATGCAG
miR396	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTTCAAG
miR528	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCTCCTC
miR535	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGCGTGC

Table 3.2, continued

Primer	Sequence 5'-3'
Stem-loop reverse transcription (RT) primer	
miR827	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCGTTTG
miR894	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGGTGA
miR1859	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTTGGCA
Forward qPCR primer	
miR10	CACGCATGGGAATGGGGG
miR13	CGGCGGAGGGAAAGTGGA
miR15	GACGCAGGGCGTTTGGTC
miR156	GCCGCCGTGACAGAAGAGA
miR160	CAAACGGCGTGCGAGGAG
miR166	CGGGACTCTCGGACCAGG
miR167	GCACGGTGAAGCTGCCAG
miR168	GCAGCATCGCTTGGTGCA
miR172	CGCGCGGCAGAATCTTGATG
miR396	GCGGCGTTCCACAGCTTT

Table 3.2, continued

Primer	Sequence 5'-3'
Forward qPCR primer	
miR528	CGAGGCTGGAAGGGGCAT
miR535	CGCGCCTGACAACGAGAG
miR827	GCGCGGCTTAGATGACCAT
miR894	GCACCGGTTTCACGTCGG
miR1859	GCGCGCGTTTCCAATACCT
Universal reverse	CCAGTGCAGGGTCCGAGGTA
Housekeeping gene, *Cyclophilin 2* (*Cyp2*)	
*Cyp 2*_forward	CTCGTCTGATGTCGTCTA
*Cyp 2*_reverse	CTGCTGGTACTCTGGTAA

CHAPTER 4

INTEGRATION OF RNA SEQUENCING AND MIRNA SEQUENCING: BUILDING A CONSENSUS REFERENCE TRANSCRIPTOME FOR MIRNA IDENTIFICATION AND TARGET PREDICTION

4.1 Introduction

RNA that is of sufficient high-quality in terms of total yield, purity and integrity is a primary requirement for RNA to be assayed by RNA and small RNA high-throughput sequencing. Depending on the sample's origin, different RNA extraction methods or commercial kits may produce varying yield, quality and integrity of the RNA. Oil palm tissues are rich in phenolic compounds, polysaccharides and secondary metabolites (Habib et al., 2014), and the leaves have a waxy cuticle and high fibre content (Corley & Tinker, 2015). These compounds can easily co-precipitate with RNA during the extraction process and may cause oxidation and degradation of the RNA (Xiao et al., 2012; Habib et al., 2014). In addition, the extraction of intact RNA can be difficult due to the enzymatic degradation of RNA by endogenous and/or exogenous RNase (Shi & Bressan, 2006; Rubio-Piña & Vázquez-Flota, 2008), thus obtaining a high quality RNA sample from oil palm tissues can be technically challenging. In this thesis, female inflorescences, male inflorescences and leaf of oil palm were used to produce the RNA for miRNA and mRNA sequencing and expression studies.

RNA sequencing (RNA-Seq) is a highly sensitive and accurate deep-sequencing tool for transcriptome analysis (Wang et al., 2009b). Successful application of RNA-Seq in transcriptome profiling is not limited to the plant species with a reference genome or reference transcripts (Dugas et al., 2011; Xu et al., 2012b; Zhang et al., 2015c; Shankar et al., 2016; Wang et al., 2016a), but can also be applied for species without a reference genome (Tranbarger et al., 2011; Cardoso-Silva et al., 2014; Mantello et al., 2014; Jain et al., 2016). Unlike hybridization-based methods, *de novo* assembly of RNA-Seq data allows detection of sequences without genome

sequence information hence is particularly useful for non-model species (Wang et al., 2009b; Wilhelm & Landry, 2009).

Currently, several examples of leading deep-sequencing technologies are delivered by 454 Life Sciences (Roche), Illumina, Inc., Pacific Biosciences and Thermo Fisher Scientific Inc., with each of these sequencing technologies being characterized by distinct sequencing principles, read length, sequencing throughput, output data etc. RNA-Seq has facilitated several studies in oil palm, including studies of alternative gene spliced transcripts (Ong-Abdullah et al., 2015), changes in gene expression (Bourgis et al., 2011; Tranbarger et al., 2011; Shearman et al., 2013; Lei et al., 2014; Ho et al., 2015; Ajambang et al., 2016; Guerin et al., 2016; Ho et al., 2016), mutations/SNPs (Pootakham et al., 2013; Lei et al., 2014; Ong-Abdullah et al., 2015; Guerin et al., 2016) and simple sequence repeat polymorphisms (Xiao et al., 2014).

The aim of this part of the thesis study was to generate a set of consensus reference sequences from oil palm +6 and +15 female inflorescences. This involved the preparation of high quality RNA, followed by deep sequencing, assembly and functional annotation of transcriptomes (+6 and +15 female inflorescences). The study was extended to investigate the simple sequence repeat (SSR) motif distribution in the tissues studied. The consensus reference sequences were later used for the subsequent miRNA study, for identification of miRNA precursors, for determining expression patterns of the corresponding target mRNA(s), and to determine the cleavage patterns of the target mRNA by the regulating miRNA(s) (which was coupled with degradome sequencing).

4.2 Results

4.2.1 An efficient method for extracting high-quality RNA from oil palm tissues

Intact total nucleic acids extracted from female inflorescences, male inflorescences and leaf of oil palm were separated on a 2% (w/v) agarose gel (Figure 4.1). Intact bands of 28S rRNA (approximately 1200 bp) and 18S (approximately 750 bp) rRNA were observed, together with bands of small RNAs. All samples showed A_{260nm}/A_{280nm} greater than 1.9, A_{260nm}/A_{230nm} greater than 2.0, 28S:18S rRNA ratio greater than 1.3 and the RNA Integrity Number (RIN) greater than 9 (Table 4.1). Bioanalyzer (Agilent 2100) RNA analysis showed three distinct peaks at 47 s (representing 28S rRNA), 42.5 s (representing 18S rRNA) and small spikes in fluorescence between 25-27 s (representing the small RNA species) (Figure 4.2). The small RNA from all five tissues (Figure 4.1) was used for high-throughput sequencing of small RNA (sRNA-seq) (section 6.2.4).

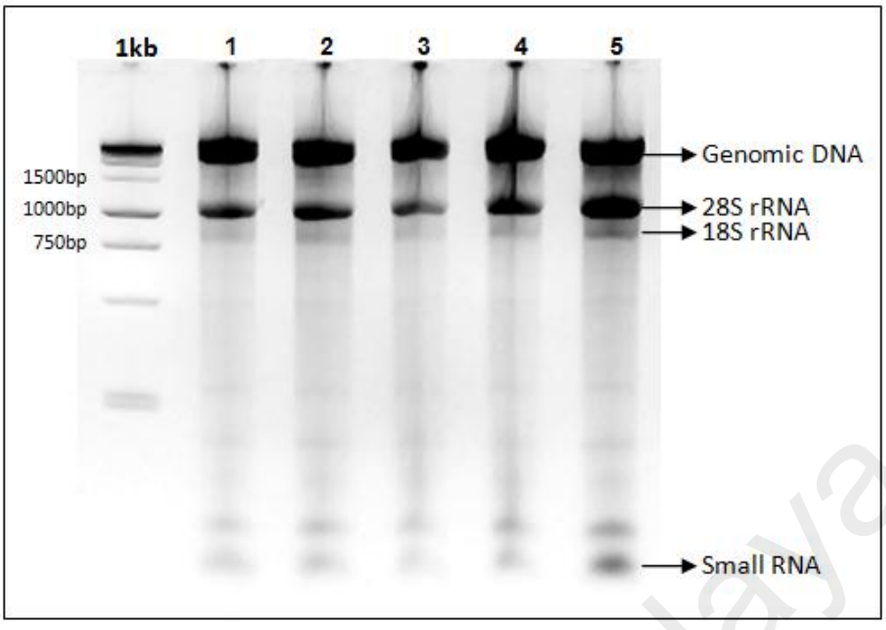


Figure 4.1: Gel electrophoresis of total nucleic acids extracted from female inflorescences, male inflorescences and leaf of oil palm, on a 2% (w/v) agarose gel. 1: +6 female inflorescence; 2: +15 female inflorescence; 3: +6 male inflorescence; 4: +15 male inflorescence; 5: leaf; 1kb: 1kb DNA marker (Fermentas)

Table 4.1: RNA quality for female inflorescences, male inflorescences and leaf of oil palm. All samples were re-suspended in 60 µl of DEPC-treated distilled water. The RNA integrity number (RIN) is based on a numbering system from 1 to 10, with 1 being the most degraded quality and 10 being the most intact.

Sample	Concentration (ng/µl)	Total yield (µg)	A_{260nm}/A_{280nm}	A_{260nm}/A_{230nm}	RIN	28S:18S
F^{+6}	349.84	20.99	1.92	2.11	9.5	1.37
F^{+15}	344.75	20.69	1.91	2.08	9.4	1.36
M^{+6}	337.29	20.24	1.93	2.10	9.5	1.45
M^{+15}	367.33	22.04	1.92	2.09	9.5	1.44
L	331.39	19.88	1.91	2.11	9.5	1.39

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence; M^{+6}: +6 male inflorescence; M^{+15}: +15 male inflorescence; L: leaf

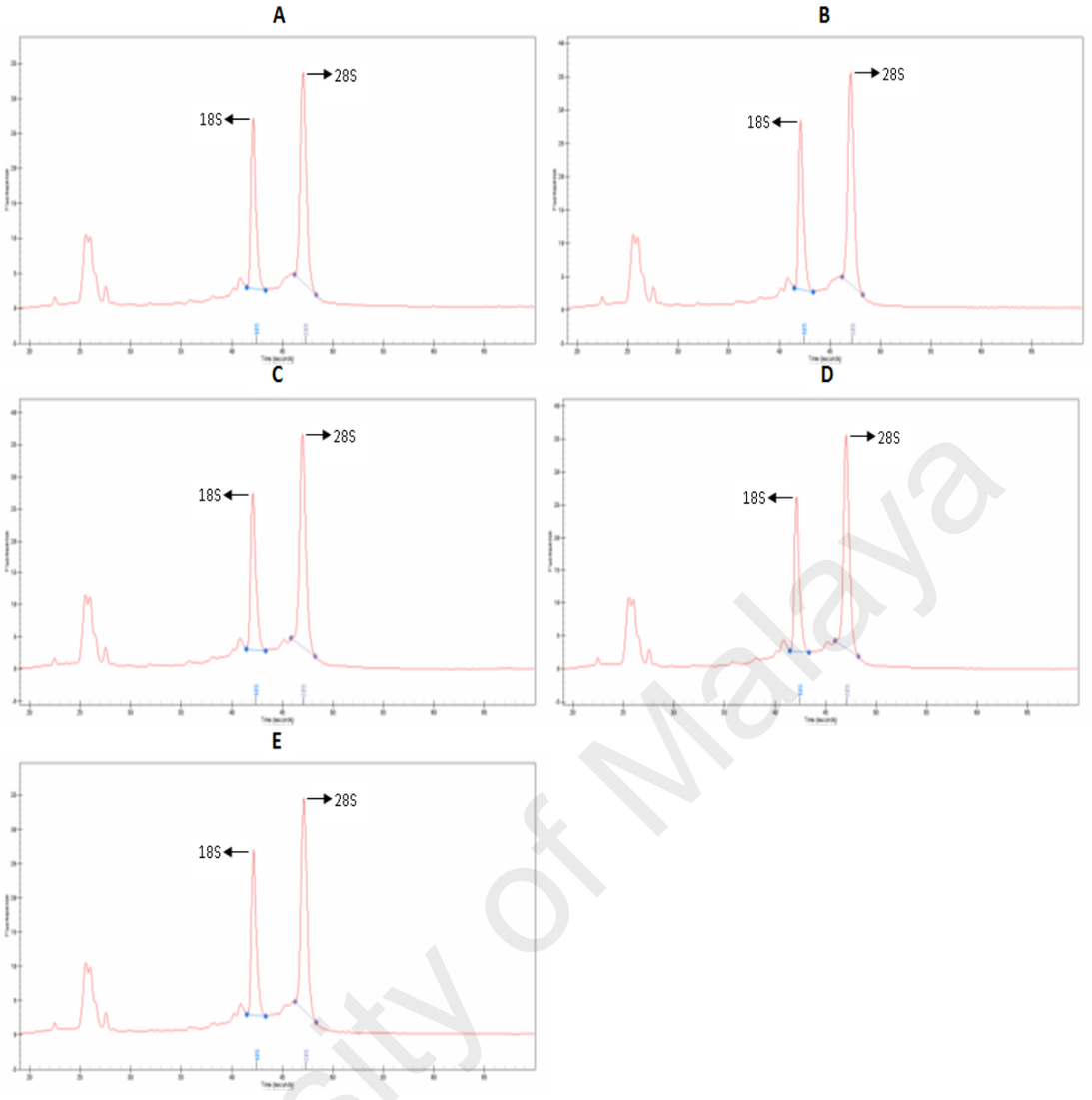


Figure 4.2: Electropherograms of total RNA extracted from female inflorescences, male inflorescences and leaf of oil palm. A: +6 female inflorescence; B: +15 female inflorescence; C: +6 male inflorescence; D: +15 male inflorescence; E: leaf

4.2.2 RNA-seq output statistics

The total RNA extracted from +6 and +15 female inflorescences (section 4.2.1) was used for RNA-Seq. Over 37 million and 38 million 90 bp paired-end reads were generated for +6 and +15 female inflorescences, respectively (Table 4.2), which in total accounted for more than 6.8 Gb of transcriptomic sequence data with a GC content of 47-48%. Over 94% of the sequencing reads had base call accuracy, i.e. Phred score of Q20.

Table 4.2: Output statistics of sequencing for +6 and +15 female inflorescences

	F^{+6}	F^{+15}
Total number of reads	37,713,472	38,941,380
Total number of bases	3,394,212,480	3,504,724,200
Q20 percentage	95.08%	94.46%
GC percentage	47.02%	48.27%

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence

4.2.3 Generation of a consensus reference assembly from oil palm female inflorescences

Using SOAPdenovo V1.05 (Li et al., 2010) and the dataset generated from the +15 female inflorescence, sequence assembly using different k-mer sizes (25, 41, 51 and 61) were assessed (Table 4.3). A k-mer size of 61 was chosen as the optimum k-mer size for the assembly of +6 and +15 female inflorescences transcriptomes, as it gave the lowest number of contigs (48,542 contigs), yet the mean length of the contigs was longer (465 nt) as compared to those produced by other k-mer sizes tested. The clean reads were first assembled into 69,418 (for +6 female inflorescence) and 48,542 (for +15 female inflorescence) contigs. Contigs with <100 nt in length were discarded for further analysis. The contigs from these two individual datasets were overlapped against each other and were clustered into a set of consensus reference assembly

comprising of 31,181 non-redundant unigenes (Table 4.4). In the consensus reference assembly, sequences of 100-500 nt represented 46.61% of the total unigenes. Over 50% of the unigenes were longer than 1000 nt (N50= 1,092) and only a very small number (18 unigenes) were longer than 5,000 nt.

Table 4.3: Optimization of k-mer size for *de novo* sequence assembly. Different k-mer sizes were used to assemble the dataset derived from the +15 female inflorescence.

k-mer size	N50	Longest contig (nt)	Mean length (nt)	Total number of contig
25	312	4,677	271	171,706
41	600	5,248	403	95,873
51	715	6,049	422	74,661
61	712	5,311	465	48,542

Table 4.4: *De novo* sequence assembly of oil palm female inflorescence transcriptomes

Sample	N50	Longest contig (nt)	Mean length (nt)	Total number of
F^{+6}	654	6,192	451	69,418
F^{+15}	713	5,310	465	48,542
Consensus*	1,092	11,077	751	31,181

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence
Consensus*: datasets derived from +6 and +15 female inflorescence were clustered into a non-redundant consensus reference assembly

4.2.4 Purine metabolism, starch and sucrose metabolism and pyrimidine metabolism were the most represented pathways in oil palm female inflorescences

In total, 21,933 unigenes had significant matches to protein sequences in GenBank NR protein database. Similarly, 14,579 unigenes had significant matches to sequences in the Swiss-Prot database. By comparing against the KEGG database, 6,129 unigenes had significant matches in the database and were assigned to 134 KEGG pathways. The three most represented pathways are purine metabolism (352 unigenes or 5.74%), starch and sucrose metabolism (337 unigenes or 5.50%) and pyrimidine metabolism (178 unigenes or 2.90%). A total of 23,867 of unigenes were annotated and classified into 12 KOG categories (Figure 4.3). A majority of the unigenes were classified under nuclear structure (41%) and intracellular trafficking/secretion/vesicular transport (41%).

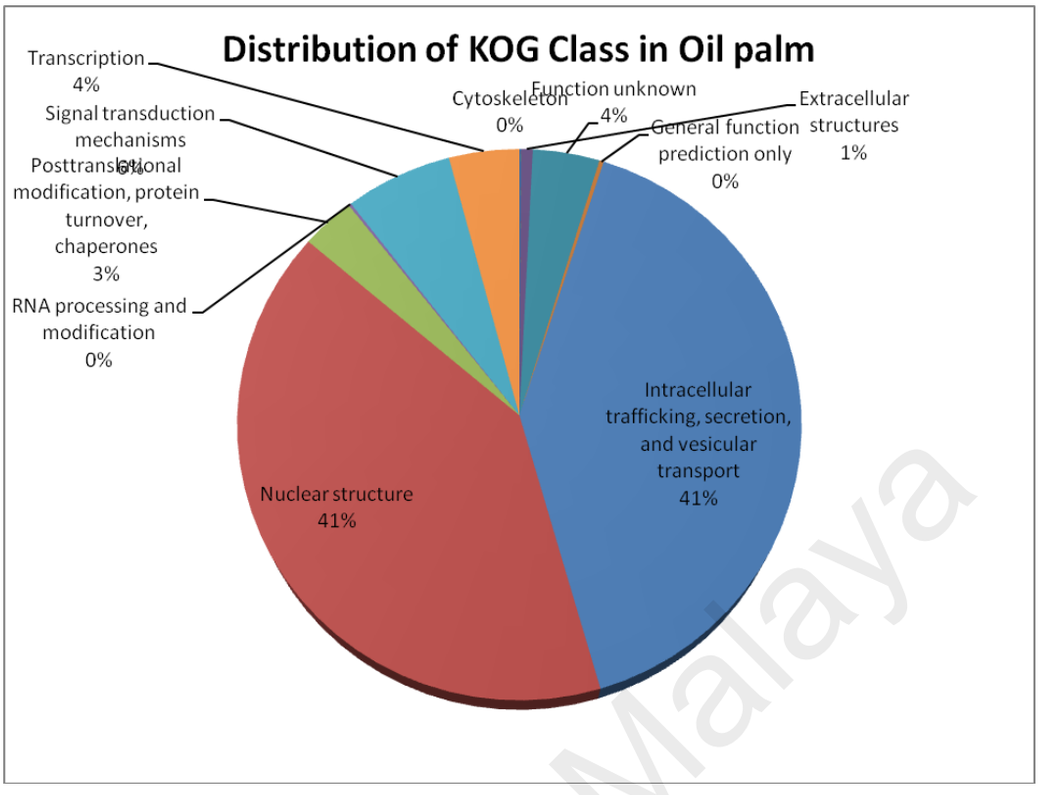


Figure 4.3: Distribution of oil palm unigenes according to KOG categories. The KOG categories are cell wall/membrane/envelope biogenesis, chromatin structure and dynamics, cytoskeleton, extracellular structures, general function prediction only, intracellular trafficking/secretion/vesicular transport, nuclear structure, post-translational modification/protein turnover/chaperones, RNA processing and modification, signal transduction mechanisms, transcription and function unknown. The percentage value (%) indicates the percentage of unigenes that was associated with each KOG category.

4.2.5 Broad categories of genes were differentially expressed in +6 and +15 female inflorescences

Differential expression analysis revealed that 2,867 or 9.19% of the *de novo* assembled unigenes were differentially expressed in +6 and +15 female inflorescences. These unigenes were assigned into broad categories of pathway groups with most of the differentially expressed unigenes associated with various metabolic processes (Figure 4.4).

4.2.6 Mononucleotide repeats were the the most abundant class of simple sequence repeats (SSRs) in oil palm female inflorescences

From the 31,181 non-redundant unigenes examined, a total of 4,119 sequences were identified to contain 5,060 SSRs (Table 4.5). About 18% of these sequences contained multiple types of SSRs. Mononucleotide repeats constituted the largest group of SSRs (48.5%) in the current datasets, followed by tri- (27.5%) and dinucleotide (22.5%) repeats (Table 4.6).

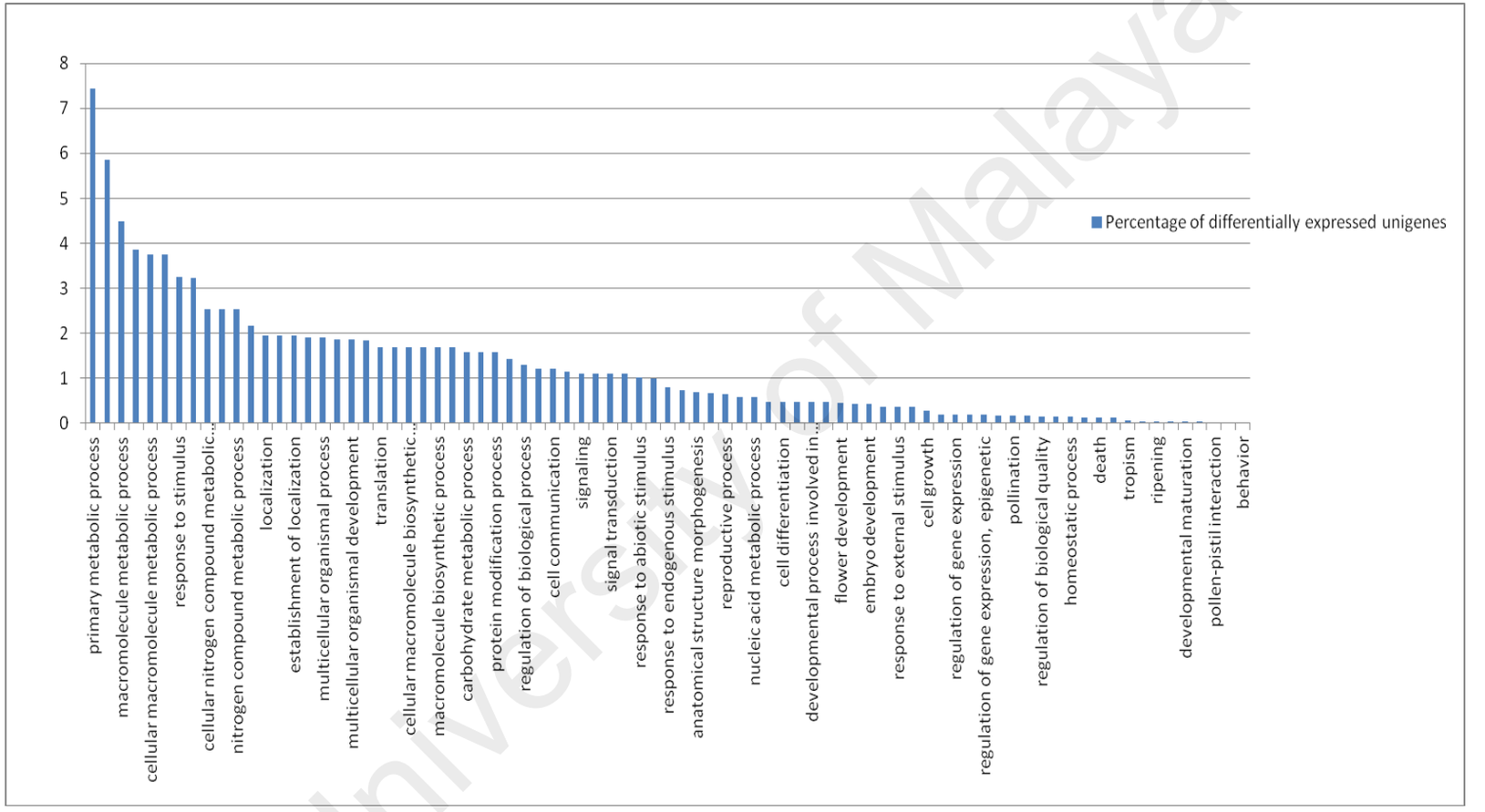


Figure 4.4: GO classification of differentially expressed unigenes. One unigene could be annotated into more than one GO term.

Table 4.5: Summary of microsatellite mining

Total number of sequences examined	31,181
Total size of examined sequences (bp)	23,427,455
Total number of identified SSRs	5,060
Number of SSR containing sequences	4,119
Number of sequences containing more than 1 SSR	745
Number of SSRs present in compound formation	396

Table 4.6: Types of SSRs identified in transcriptomes of oil palm female inflorescences

SSR type	Number of repeats												Total
	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**>15**	
Mono-nt (N)	-	-	-	-	-	615	401	316	227	170	139	587	2,455 (48.5%)
Di-nt (NN)	-	333	196	164	124	81	66	39	27	28	19	59	1,136 (22.5%)
Tri-nt (NNN)	709	344	177	75	48	19	8	8	4	1	-	1	1,394 (27.5%)
Tetra-nt (NNNN)	37	8	2	1	1	-	-	-	1	-	-	-	50 (1.0%)
Penta-nt (NNNNN)	11	2	-	-	1	-	-	-	-	-	-	-	14 (0.3%)
Hexa-nt (NNNNNN)	7	4	-	-	-	-	-	-	-	-	-	-	11 (0.2%)
Grand total													5060 (100%)

4.3 Discussion

4.3.1 High quality RNA is essential for high-throughput sequencing and gene expression analyses

For recalcitrant inflorescence and leaf tissues from oil palm, the RNA extraction method was optimized with an additional of 3-4 extraction steps using phenol-chloroform-isoamylalcohol for an improved quality of the isolated nucleic acids, indicating a high content of proteins. Besides the total yield of RNA, high RNA purity is essential for successful cDNA library construction as the multiple enzymatic steps involved in the library construction are sensitive to polyphenol, polysaccharide and protein contaminants. In this study, the extracted RNA samples have A_{260nm}/A_{280nm} ratio> 1.9, indicating low protein contamination, and A_{260nm}/A_{230nm} ratio >2.0, indicating low contamination of salts and organic compounds (Table 4.1).

The measure of RNA integrity (RIN) is another criterion that correlates with transcript quantification in RNA-seq experiments (Gallego Romero et al., 2014). Poor RNA integrity may complicate the assembly of contigs into larger scaffolds (Johnson et al., 2012). In the current study, all the extracted RNA samples had a RIN>9 (Table 4.1) which was higher than the typical cutoff value for RNA-Seq, i.e. RIN>8 (Van Verk et al., 2013). Degradation of larger RNAs will cause higher proportions of shorter RNA sequences and hence the over-representation of degraded RNA species in the resultant small RNA cDNA library. While the 28S and 18S ratio of 2:1 has long been considered the benchmark for intact RNA (Skrypina et al., 2003), the 28S:18S peak area ratio of the RNA samples extracted from oil palm inflorescences and leaf was between 1.3 and 1.5 (Table 4.1), which could be partly due to structural instability of the 28S rRNA relative to the 18S RNA (Venkov & Hadjiolov, 1969), or could possibly a characteristic that correlates to the tissue of origin or a variation across different plant species.

4.3.2 Generation of a consensus reference assembly from oil palm female inflorescences

Considering the high cost of RNA-seq and the fact that only the female inflorescence that will develop into the fruit bunch (which is the product of interest), RNA-seq was performed on only two of the oil palm tissues studied, i.e. the +6 and the +15 female inflorescences. Low et al. (2014) has predicted over 5000 of oil palm gene models that were mappable to the genome (Singh et al., 2013). The consensus reference assembly produced in the current study consists of 31,181 non-redundant unigenes, many of which may represent "isotigs", i.e. splice variants of the same gene. Splicing variations of precursor mRNAs often result in functional diversity of proteins and these splice variants could be cell-, tissue-, developmental stage-specific and can be produced in response to environmental signals (Syed et al., 2012; Reddy et al., 2013). While the increasing number of plant genome sequences and transcriptome data has facilitated *de novo* identification of splice variants, a major challenge is to accurately predict alternative splicing events and splice variants, which eventually needs to be supported by biological evidence. Besides isotigs, some of the unigenes may be chimeric contaminants, partially processed RNAs or genomic contaminants.

4.3.3 Broad categories of genes were expressed in developing oil palm female inflorescence

Notwithstanding the availability of reference genome for oil palm (Singh et al., 2013), high quality transcriptome data is necessary to complement the current genome assembly, especially for the genomic regions that were still not/ partially assembled. In this study, a total of 85 nuclear housekeeping genes which have been reported to be conserved from moss to flowering plants (Yue et al., 2010) were used to evaluate the quality of the consensus transcriptome. Out of the 85 housekeeping genes, 78 orthologous sequences were identified in the consensus transcriptome produced in the

current study, representing a coverage of over 91%, which in turn validates the consensus transcriptome as likely to have good representation of the genome and as a comprehensive reference set.

Functional annotation of unigenes revealed that transcripts involved in broad categories of pathways/ processes were expressed in developing female inflorescences, as it would be expected that floral development does not exist in isolation, but rather in the form of crosstalk between diverse cellular and metabolic pathways via various mediators including miRNAs, transcription factors and signal transducers. The most represented KEGG pathways in oil palm female inflorescences include purine and pyrimidine metabolism. KOG analysis showed that a majority of the unigenes were classified under nuclear structure (41%) (Figure 4.3). The two distinct stages of female inflorescence development studied here, i.e. +6 and +15 stages, fall within the period during which floral meristems are emerging followed by floral organ formation. Highly represented purine and pyrimidine metabolism pathways have also been reported in developing flower of olive (*Olea europaea* L.) (Alagna et al., 2016) and *Phalaenopsis* orchid (Huang et al., 2015). Highly represented purine and pyrimidine metabolism pathways may reflect active DNA replication machinery for cell division, differentiation and organogenesis in +6 and +15 female inflorescences, where purine and pyrimidine are used as precursors for nucleotide biosynthesis.

While a high starch and sucrose metabolism is commonly reported for starch- and/or sucrose -rich fruits/organs such as mango (*Mangifera indica* Linn) (Wu et al., 2014), pineapple (*Ananas comosus*) (Zhang et al., 2012), potato (*Solanum tuberosum* L.) (Ferreira et al., 2010) and sweet potato (*Ipomoea batatas* L.) (Firon et al., 2013), the same pathway was among the most dominant pathways in oil palm female inflorescence that will eventually develop into oil-rich fruits. It is known that starch and sucrose metabolism produces substrates for lipid metabolism (Rawsthorne, 2002). However, not much is known about the regulation of the switch from a high starch and

sucrose metabolism to a high lipid metabolism in oil palm female inflorescences. In plant, besides as providers of carbon and energy, sugars also act as signaling molecules, which coordinate with hormonal signaling pathways to regulate various plant physiological processes including growth, development, and stress-related responses (Rolland et al., 2006; Wind et al., 2010). As sucrose-mediated signaling has been reported to extensively cross-talk with many important factors regulating flowering networks and the plant circadian clock (Bolouri Moghaddam & Van den Ende, 2013), it would be interesting to investigate if similar mechanisms are involved in the regulation of floral development of oil palm.

Besides nuclear structure, a large proportion of the unigenes were associated with intracellular trafficking/secretion/vesicular transport functions (41%) (Figure 4.3). Lipids, polysaccharides, secretory proteins and other macromolecules are transported to various intracellular and extracellular destinations during cell growth (Cheung & de Vries, 2008; Toyooka et al., 2009). In developing female inflorescence, the rapidly dividing and expanding cells require active bidirectional trafficking of these molecules to and from the plasma membrane and extracellular space and hence the high representation of the functionally-related unigenes in the current transcriptomes. Notably, many of the unigenes have unknown functions, indicating an urgent need to extract functional information from the unannotated regions of the oil palm reference genome.

4.3.4 Primary metabolic pathway genes were differentially expressed in developing oil palm female inflorescences

Genes differentially expressed in +6 and +15 female inflorescences were enriched in functions related to diverse metabolic processes (Figure 4.4). The two distinct stages of inflorescence development studied here, i.e. the +6 and +15 stages, can be phenotypically distinguished by the size of the inflorescence (Appendix B,

Supplementary Figure B1), which resulted from active cell division, differentiation and organogenesis along the development. The increase in size of female inflorescence from +6 stage to +15 stage requires the production of primary metabolites such as building blocks of nucleic acids (i.e. purine and pyrimidine as discussed earlier), macromolecular compounds and energy (i.e. starch as discussed earlier), highlighting the importance of primary metabolic pathways during the growth of the inflorescence tissues.

Unlike primary metabolites, secondary metabolites do not play a role in growth, development and reproduction, and are typically produced during the end or near the stationary phase of growth, mostly for defense against herbivores, pests and pathogens (Wink, 1988; Bennett & Wallsgrove, 1994). In the current study, only a minimal number of differentially expressed genes was related to secondary metabolism (classified under pollination and pollen-pistil interaction) (Figure 4.4), which can be expected as the +6 and +15 stages of inflorescence fall within the period during which fundamental primary metabolic pathways are dominant. In *Brunfelsia calycina,* flowers changed colour from purple to white due to anthocyanin degradation, parallel to an increase in fragrance and petal size (Bar-Akiva et al., 2010). In *Antirrhinum majus,* flowers produced secondary metabolites such as scent compounds and anthocyanin pigments to attract pollinators (Muhlemann et al., 2012). While similar change of flower colour and production of scent compounds is not evident during the maturation of oil palm flowers (Adam et al., 2005), the mechanism for oil palm flowers to attract its insect pollinators is still not known (Meléndez & Ponce, 2016).

In addition to inherent metabolic variability that is likely to exist across different plant species, gibberellins (Zhuang et al., 2015), heat/drought (Li et al., 2015a) and light conditions (Christiaens et al., 2016) have been reported to induce metabolic changes in developing flowers. The long flower developmental phase of oil palm inflorescence presents a long window of exposure to various environmental stimuli,

which can act either independently or jointly. Therefore it is important to take these elements into consideration before dissecting the distinct developmental regulation of secondary metabolism in developing inflorescence and the related upstream primary metabolic pathways.

4.3.5 Potential SSR markers for marker-assisted selection

Similar to most plants, the oil palm genome is largely constituted by repetitive DNA sequences (Singh et al., 2013). Castilho et al. (2000) has characterized a range of repetitive DNA sequence families from oil palm genome which showed variation in distribution, length and nucleotide composition. Mononucleotide repeats was the most abundant class of SSR detected in unigenes derived from oil palm female inflorescences (Table 4.6). Similar to those found in Arabidopsis and rice (Lawson & Zhang, 2006), poly-A/T mononucleotide SSRs were significantly more abundant than poly-G/C. A study by Ting et al. (2010) analyzing the expressed sequence tags (EST) derived SSR has otherwise shown that dinucleotide repeats formed the largest group of SSRs (45.6%), followed by trinucleotide repeats (34.5%). Characterization of EST-derived SSRs by Tranbarger et al. (2012) showed that dinucleotide repeats formed the largest group of SSRs, followed by tetranucleotide repeats. The discrepancies between the abundance of each type of SSR in the current and the previous studies (Ting et al., 2010; Tranbarger et al., 2012) (Figure 4.5) could be due to: 1) different analytical algorithms/parameters used, 2) different number of sequence analyzed, and 3) different source of the input sequences used for analysis. It is noted that the current study was limited to unigenes which might represent only a partial gene sequence, thus the data provided in this work should be considered preliminary in reflecting the SSR content of oil palm genome. Nonetheless, the large collection of SSRs identified in this study provides potential markers that could be relevant to applications such as genetic mapping, molecular breeding and quantitative trait loci (QTL) analysis, as well as for

assessment of genetic diversity within the palm family. However, before any of these SSRs can be applied for large-scale marker-assisted selection in oil palm, much downstream functional studies are needed to understand the role of the SSR-containing transcripts in transcriptional and post-transcriptional regulation and also the phenotypic traits that co-segregate with these markers.

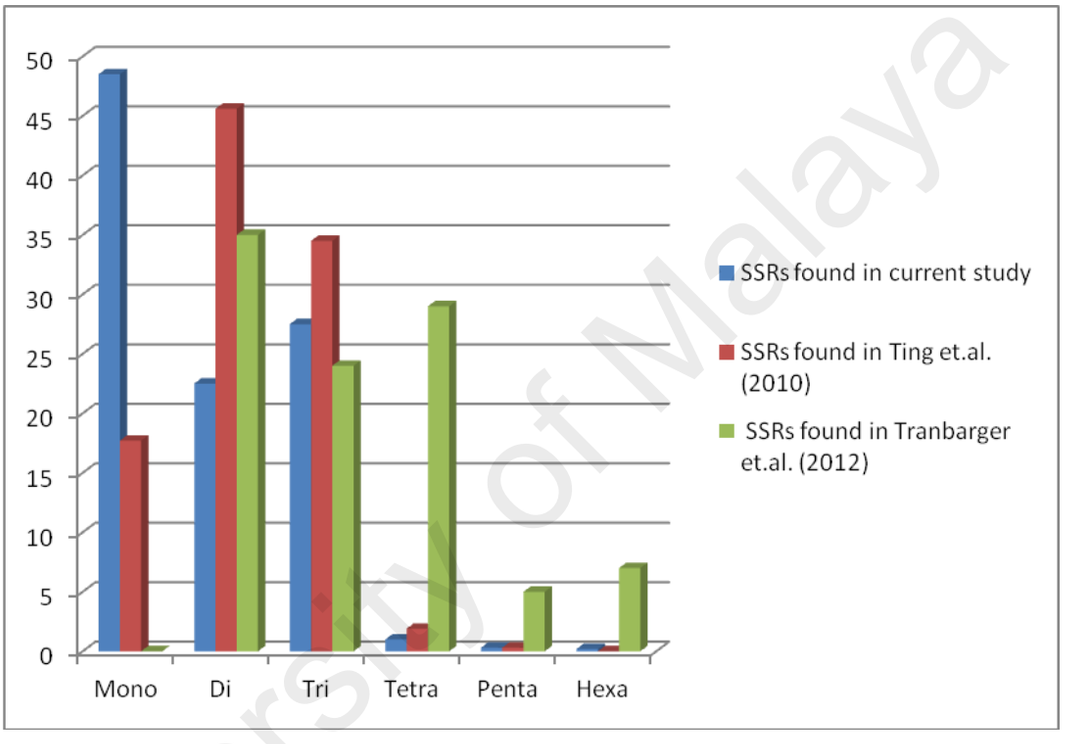


Figure 4.5: Comparison of the abundance of SSR type in oil palm. Blue, red and green bars represent the relative abundance of each SSR type found in the current study, study by Ting et al. (2010) and study by Tranbarger et al. (2012), respectively. The x-axis represents SSRs that have different SSR units and the y-axis represents the abundance (in percentage) of each type of SSR.

CHAPTER 5

EXPRESSION PATTERNS OF INFLORESCENCE- AND SEX-SPECIFIC TRANSCRIPTS IN FEMALE AND MALE INFLORESCENCES OF OIL PALM

5.1 Introduction

In monoecious plants, floral sex allocation has evolved in response to natural selection, to enable alteration of the sex ratio (Willson, 1979). For oil palm, the sex ratio of the female to male inflorescences on each palm is important for both breeding and commercial production. In the interest of commercial cultivation, optimum fruit yield requires a higher ratio of female inflorescences in the production plants, whereas for hybrid breeding, a different floral sex ratio may be required, for example in the male parent line used for breeding purposes (Durand-Gasselin et al., 1999). Oil palm floral sex ratio has been reported to be regulated by environmental (Corley, 1976a), anthropogenic (Corley, 1976b; Durand-Gasselin et al., 1999) and very likely, genetic factors. Although the morphological differences between female and male inflorescences of oil palm have been well characterized (Adam et al., 2005), the genetic factors of floral sex expression are still poorly understood.

Sex determination of oil palm inflorescences usually initiates in the developing tissues adjacent to the frond stage −20 (Durand-Gasselin et al., 1999), which at that point lie deep within the palm stem. It is challenging to extract these developing materials as the sampling procedure is destructive to both the tissues and the palm hence it is impossible to know the final sex of the tissue once the tissues are sampled. Moreover, sexual identity of an inflorescence might not be stably determined at the initiation of primordium (Kater et al., 2001). At the +6 stage (approximately 10 months after initiation of inflorescence meristem), the morphology of female and male inflorescences can be reliably distinguished (Adam et al., 2005). The male inflorescence at leaf stage +6 displayed a series of long, finger-like and cylindrical spikelets spirally arranged around the rachis, whereas the female inflorescence

displayed shorter spiny spikelets. The +6 stage is also within the period where abortion of entire inflorescences often occurs, which will alter the sex ratio of inflorescences of a palm (Durand-Gasselin et al., 1999).

The aim of this section of the study was to determine the differences between female and male inflorescences of oil palm using whole genome expression analysis. This involved deep sequencing, assembly and functional annotation of transcriptomes of female and male inflorescences at the earliest stage at which the female and male tissues can be reliably distinguished, i.e., +6 leaf stage.

5.2 Results

5.2.1 A broad and dense consensus transcriptome from oil palm female and male inflorescences

The assembled transcriptomes of multiple oil palm tissues, consisting of the male inflorescence, female inflorescence, apical meristem, leaf, mesocarp and root were combined in order to generate a comprehensive reference transcriptome, which was named EGrefseq. Over 20,000,000 454 reads with an average read length of 400 bp were used to assemble 60,210 non redundant transcripts comprising 70M bp (Table 5.1). The transcriptome sequences were deposited at DDBJ/EMBL/GenBank under the accession GCKD00000000. About 45% of EGrefseq transcripts are longer than 1000 nt. By matching the transcripts to TAIR10 and RGAP 7, 37,737 and 40,162 transcripts had significant sequence matches against the respective databases. A total of 30,192 transcripts had BLAST matches in Swiss-Prot database and 40,208 transcripts had significant matches with entries in the KEGG database.

Transcripts for all 458 CEG were represented in EGrefseq and 453 out of the total (98.91%) had alignments with lengths exceeding 60% of either the CEG or the EGrefseq sequence (Figure 5.1). This validates the EGrefseq as likely to have good representation of the genome and as a comprehensive reference set.

Table 5.1: Overview of EGrefseq consensus assembly

Number of transcripts	60,210
Number of genes	38,981
Total size of EGrefseq (bp)	70,422,832
Longest length of transcript (bp)	11,413
Shortest length of transcript (bp)	200
Average length of transcript (bp)	1,169
N50 (bp)	1,652

Bp, base pair.

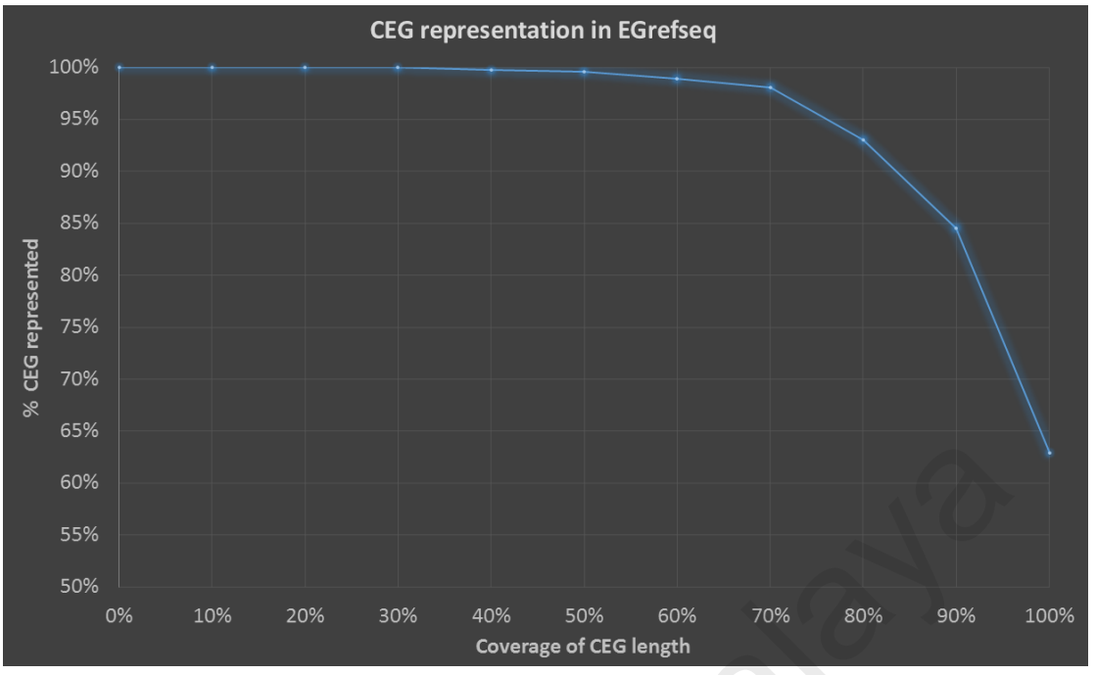


Figure 5.1: Quality assesment of the current assembly and annotations. The assembled unigenes were aligned against 458 core eukaryotic genes (Parra et al., 2007) that are highly conserved across a wide range of eukaryotic taxa. CEG: core eukaryotic genes

5.2.2 Common and unique transcripts in female and male inflorescences

Datasets for male (Mi) and female (Fi) inflorescences shared over 11,000 common transcripts with at least 10 counts in both tissues (Figure 5.2). Based on a strict selection criterion which a transcript must be present at 10 or more counts in the transcriptome of one tissue and absent in the other, 97 potential sex-specific transcripts that were differentially expressed in female and male inflorescences were identified (Figure 5.2and Table 5.2) (the complete expression data and annotation for these 97 transcripts are shown in Appendix C, Supplementary Table C2). These sex-specific transcripts were mapped to the oil palm reference genome (Singh et al., 2013) with no less than 98% of their length. Among the sex-specific transcript, 86 transcripts mapped across different chromosomes, whereas the remaining 11 transcripts mapped to unlinked contigs. The expression level of orthologs of several known floral genes were further examined (Table 5.3). Orthologous transcripts of *FLOWERING LOCUS T* (*FT*), *LEAFY* (*LFY*), *APETALA1* (*AP1*) and *SUPPRESSOR OF CONSTANS*

OVEREXPRESSION1 (*SOC1*) were commonly expressed in both female and male inflorescences of oil palm. A MADS-box gene, *DEFICIENS*, showed male-specific expression pattern whereas *TASSELSEED1* (*TS1*) showed female-specific expression pattern. Several of the previously reported candidate sex-determining genes such as *ANTHER EAR*, *DWARF*, *SUPERMAN* and *TASSELSEED2* had counts that were below the threshold of selection criteria and therefore were not considered as sex-specific (Table 5.3).

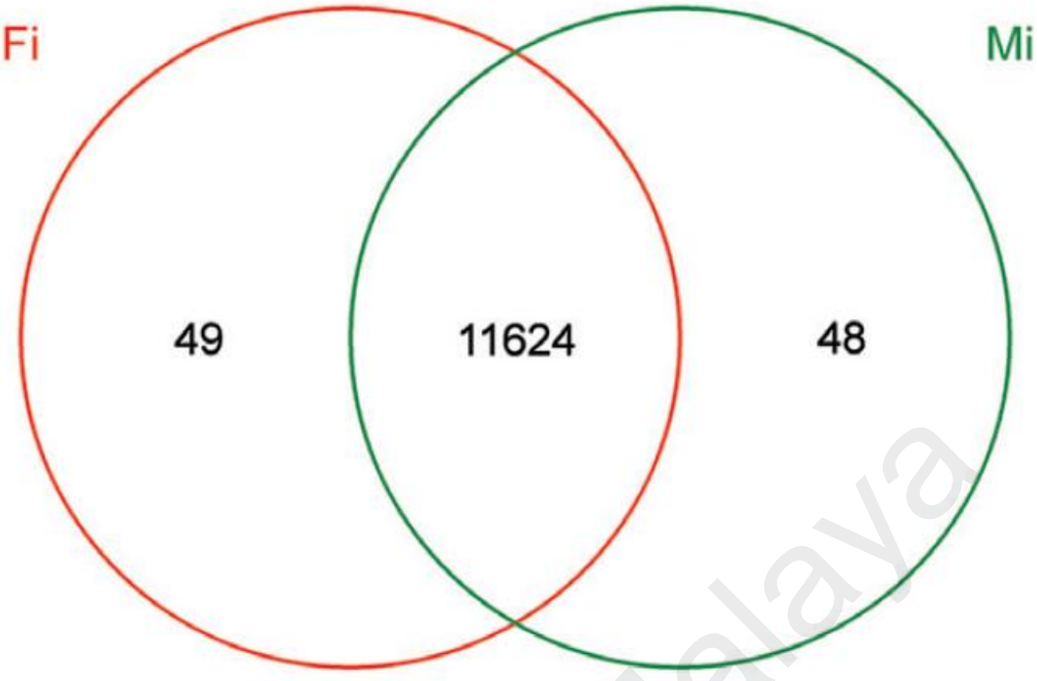


Figure 5.2: Common and unique transcripts between female and male inflorescences. The comparison was based on the 454 transcriptome datasets. A transcript was considered unique when it had at least 10 counts in one sample and a zero count in the other sample. Fi, female inflorescence; Mi, male inflorescence

Table 5.2: Functional annotation of sex-specific transcripts

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig05793	0	64	N/A	N/A	N/A	N/A
isotig05961	23	0	Ribosome biogenesis regulatory protein (RRS1)	Ribosome biogenesis regulatory protein	Ribosome biogenesis regulatory protein homolog	Ribosome biogenesis regulatory protein homolog-
isotig08322	15	0	N/A	N/A	N/A	N/A
isotig08594	13	0	Nuclear factor Y, subunit C11	Histone-like transcription factor and archaeal histone	Dr1-associated corepressor	Repressor protein
isotig08668	0	19	N/A	N/A	N/A	N/A
isotig12716	0	11	ETO1	BTB protein	ETO1-like protein 1	ETO1-like protein 1-like
isotig14530	0	19	N/A	N/A	N/A	N/A
isotig15839	0	11	N/A	N/A	N/A	N/A
isotig17449	0	21	N/A	N/A	N/A	N/A
isotig19349	0	14	SAUR-like auxin-responsive protein	Auxin-responsive SAUR gene family member	Auxin-induced protein X10A	Uncharacterized protein
isotig22971	11	0	Alpha/beta-hydrolases superfamily protein	Palmitoyl-protein thioesterase 1 precursor	Palmitoyl-protein thioesterase 1	Palmitoyl-protein thioesterase 1
isotig23091	0	10	Lipoxygenase	Lipoxygenase protein	Putative lipoxygenase 5	Putative lipoxygenase 5-like
isotig23145	12	0	Protein kinase superfamily protein	ELMO/CED-12 family protein	N/A	Uncharacterized protein
isotig25955	61	0	COP1 SUPPRESSOR 1	Zinc finger, C3HC4 type domain containing protein	Nitric oxide synthase-interacting protein homolog	Nitric oxide synthase-interacting protein-like
isotig26131	11	0	N/A	N/A	N/A	N/A

Table 5.2, continued

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig26605	0	10	Xyloglucan endotransglycosylase-related	Glycosyl hydrolases family 16	Brassinosteroid-regulated protein BRU1	Probable xyloglucan endotransglucosylase/hydrolase
isotig28587	0	11	N/A	N/A	N/A	N/A
isotig30794	0	11	*Arabidopsis* THO/TREX	Cold-induced protein	THO complex subunit 5B	THO complex subunit 5
isotig31484	0	11	Clathrin adaptor complexes medium subunit family protein	Adaptor complexes medium subunit family domain	AP-4 complex subunit mu	Hypothetical protein
isotig32315	0	11	N/A	Expressed protein	N/A	Uncharacterized protein
isotig32456	11	0	N/A	Expressed protein	N/A	N/A
isotig32475	12	0	Histone superfamily protein	Expressed protein	N/A	Centromere protein S-like
isotig33665	0	14	Fasciclin-like arabinogalactan protein 4	Fasciclin-like arabinogalactan protein	Fasciclin-like arabinogalactan protein 4	Fasciclin-like arabinogalactan protein 4
isotig33843	0	30	N/A	60S ribosomal protein	N/A	60S ribosomal protein L34-like
isotig34682	0	12	N/A	N/A	N/A	N/A
isotig35216	10	0	Gamma subunit of ATP synthase	ATP synthase gamma chain	ATP synthase subunit gamma	Hypothetical protein
isotig37359	10	0	N/A	DEAD-box ATP-dependent RNA helicase	N/A	DEAD-box ATP-dependent RNA helicase 28-like
isotig37496	10	0	Pre-mRNA splicing factor SF3a60	ZOS3-17 - C2H2 zinc finger protein	Splicing factor 3A subunit 3	Splicing factor 3A subunit 3-like
isotig38100	27	0	Phototropic-responsive NPH3 family protein	Phototropic-responsive NPH3 family protein	Root phototropism protein 3	Signal transducer, putative
isotig38200	10	0	Nodulin21	Auxin-induced protein	Protein WALLS ARE THIN 1	Auxin-induced protein 5NG4-like

Table 5.2, continued

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig40056	0	10	DNA repair protein	Auxin response factor	Cell cycle checkpoint control protein RAD9A	DNA repair protein rad9, putative
isotig40383	0	39	Heat shock protein	hsp20/alpha crystallin family protein	22.0 kDa class IV heat shock protein	Heat-shock protein, putative
isotig40414	0	14	bZIP protein	bZIP transcription factor	Protein FD	Protein FD-like
isotig40710	10	0	MADS-box transcription factor	OsMADS16 - MADS-box family gene	MADS-box transcription factor 16	MADS-box transcription factor 16-like
isotig40869	0	11	Pentatricopeptide repeat (PPR) superfamily protein	Pentatricopeptide	Pentatricopeptide repeat-containing protein At2g35030	Pentatricopeptide repeat-containing protein, putative
isotig42197	11	0	N/A	Proline-rich family protein	N/A	N/A
isotig42516	23	0	N/A	BBTI13 - Bowman-Birk type bran trypsin inhibitor precursor	Bowman-Birk type trypsin inhibitor	Bowman-Birk type trypsin inhibitor-like
isotig43008	0	14	N/A	N/A	N/A	N/A
isotig43317	0	35	N/A	N/A	N/A	N/A
isotig48344	16	0	PCMP (plant combinatorial and modular protein)	Pentatricopeptide	Pentatricopeptide repeat-containing protein At3g46790	Pentatricopeptide repeat-containing protein At2g27610-
isotig48345	13	0	N/A	Retrotransposon protein	Putative ribonuclease H protein At1g65750	Putative ribonuclease H protein At1g65750-like
isotig48468	0	10	Unknown protein	Retrotransposon protein	Exopolyphosphatase	Exopolyphosphatase-like
isotig49139	0	10	N/A	N/A	N/A	N/A
isotig49719	11	0	WRKY transcription factor	WRKY11	Probable WRKY transcription factor 28	WRKY transcription factor, putative
isotig50300	0	17	N/A	N/A	N/A	N/A

Table 5.2, continued

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig50416	14	0	RbcX protein	Expressed protein	N/A	N/A
isotig50646	0	10	Unknown protein	Expressed protein	N/A	Uncharacterized protein
isotig51070	10	0	N/A	Albumin-2	N/A	Pierisin
isotig51466	0	11	SKOR	Potassium channel SKOR	Potassium channel KOR1	Potassium channel KOR1-like
isotig51755	0	10	N/A	N/A	N/A	N/A
isotig51940	0	11	Unknown protein	Expressed protein	N/A	Uncharacterized protein
isotig51958	22	0	Enhancer of polycomb-like transcription factor protein	Expressed protein	N/A	Uncharacterized protein
isotig52132	13	0	N/A	Retrotransposon protein	Transposon Ty3-G Gag-Pol polyprotein	N/A
isotig52249	0	15	Mnd1 family protein	Meiotic coiled-coil protein	Meiotic nuclear division protein 1 homolog	Hypothetical protein
isotig53408	10	0	HAD superfamily, subfamily IIIB acid phosphatase	HAD superfamily phosphatase	Acid phosphatase 1	Acid phosphatase 1-like
isotig54309	0	12	N/A	Embryonic protein	N/A	N/A
isotig54707	0	75	Heat shock protein	hsp20/alpha crystallin family protein	17.1 kDa class II heat shock protein	17.9 kDa class II heat shock protein-like
isotig54830	10	0	N/A	N/A	N/A	N/A
isotig55379	0	10	N/A	N/A	N/A	N/A
isotig55812	0	17	N/A	N/A	N/A	N/A

Table 5.2, continued

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig55854	10	0	RmlC-like cupins superfamily protein	Cupin domain containing protein	N/A	Vicilin-like antimicrobial peptides 2-2-like
isotig56382	0	11	N/A	N/A	N/A	N/A
isotig56629	0	14	N/A	N/A	N/A	N/A
isotig58680	12	0	Phospholipase D	Phospholipase D	Phospholipase D delta	Phospholipase D delta-like
isotig58939	12	0	MADS-box transcription factor	OsMADS6 - MADS-box family gene	MADS-box transcription factor 6	Hypothetical protein
isotig59228	10	0	N/A	N/A	N/A	N/A
isotig59236	0	13	N/A	N/A	N/A	N/A
isotig60447	0	10	N/A	N/A	N/A	N/A
isotig62514	10	0	N/A	N/A	N/A	N/A
isotig62606	0	15	N/A	N/A	N/A	N/A
isotig62742	10	0	N/A	N/A	N/A	N/A
isotig62743	15	0	N/A	N/A	N/A	N/A
isotig62839	10	0	N/A	N/A	N/A	N/A
isotig63768	0	14	N/A	N/A	N/A	N/A
isotig66709	10	0	N/A	N/A	N/A	N/A

Table 5.2, continued

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig66790	23	0	N/A	N/A	N/A	N/A
isotig67331	11	0	N/A	N/A	N/A	N/A
isotig67623	0	11	MuDR family transposase	Transposon protein	N/A	Uncharacterized protein
isotig67634	12	0	N/A	N/A	N/A	N/A
isotig68647	15	0	N/A	N/A	N/A	N/A
isotig68874	0	14	N/A	ATP/GTP/Ca++ binding protein	N/A	N/A
isotig69051	0	11	N/A	N/A	N/A	N/A
isotig69329	0	10	Dehydroquinate-shikimate dehydrogenase	Bifunctional 3-dehydroquinate dehydratase/shikimate	Bifunctional 3-dehydroquinate dehydratase/shikimate	Bifunctional 3-dehydroquinate dehydratase/shikimate
isotig69416	12	0	N/A	N/A	N/A	N/A
isotig69569	0	14	N/A	N/A	N/A	N/A
isotig69858	13	0	N/A	N/A	N/A	N/A
isotig70164	11	0	N/A	N/A	Curculin-2	N/A
isotig70298	0	12	N/A	N/A	N/A	N/A
isotig70475	0	12	1-phosphatidylinositol-3-phosphate 5-kinase	Phosphatidylinositol 3- and 4-kinase family protein	Serine/threonine-protein kinase ATM	Serine/threonine-protein kinase ATM-like
isotig70480	16	0	N/A	N/A	N/A	N/A

Table 5.2, continued

Transcript	Count		Annotation			
	Mi	Fi	TAIR10	RGAP 7	Swiss-Prot	KEGG
isotig70872	12	0	N/A	N/A	N/A	N/A
isotig70874	11	0	N/A	N/A	N/A	N/A
isotig71149	51	0	N/A	N/A	N/A	N/A
isotig71267	0	13	Propyzamide hyper sensitive 1	PHS1	Dual specificity protein phosphatase PHS1	Propyzamide hyper sensitive 1
isotig71550	18	0	N/A	N/A	N/A	N/A
isotig72098	22	0	N/A	N/A	N/A	N/A
contig91491	0	11	N/A	N/A	N/A	N/A

Mi, male inflorescence; Fi, female inflorescence; N/A, no match found.

Table 5.3: Expression level of orthologs of known floral genes in female and male inflorescences of oil palm

Gene	Transcript	Transcriptome count		References
		Mi	Fi	
AGAMOUS-like *62*[a]	isotig58939	12	0	Adam et al., 2006
ANTHER EAR	isotig48988	4	4	Kim et al., 2007
APETALA1[a]	isotig12218	9	16	Jaeger et al., 2006
	isotig12219	20	12	
	isotig12303	748	879	
	isotig23316	729	961	
DEFICIENS[a]	isotig40710	10	0	Adam et al., 2006
DWARF	isotig45754	9	2	Calderon-Urrea and Dellaporta, 1999
FLOWERING LOCUS T	isotig05923	14	3	Yoo et al., 2005
LEAFY	isotig20825	202	118	Weigel et al., 1992
	isotig20827	15	7	
	isotig54325	52	19	
SUPERMAN	isotig52692	0	0	Kazama et al., 2009
SUPPRESSOR OF CONSTANS OVEREXPRESSION1[a]	contig15488	61	37	Yoo et al., 2005
TASSELSEED1	isotig23091	0	10	Acosta et al., 2009
TASSELSEED2	isotig39747	0	0	Acosta et al., 2009
	isotig39748	0	0	
WIP-zinc finger protein	isotig39321	107	48	Martin et al., 2009
	isotig39322	26	22	
	isotig39379	47	38	
	isotig39380	28	18	
	isotig51093	10	19	

Mi, male inflorescence; Fi, female inflorescence.
[a]MADS-box genes.

5.2.3 nCounter analysis validated the expression levels of sex-specific transcripts in floral and non-floral tissues of oil palm

As transcriptome data from high-throughput sequencing methods is generally only accepted as predictive, a probe-based method called nCounter analysis system (NanoString Technologies, Seattle, WA, USA) was used to re-examine the expression levels of 40 sex-specific transcripts in floral and non-floral tissues of oil palm. By applying similar selection criteria for the threshold level of gene expression (≥10 counts in at least one of the tissues), 16 transcripts that had at least 10 counts in at least one of the oil palm tissues analyzed were identified (Appendix C, Supplementary Table C3). Of these, four transcripts were more abundant in female inflorescence, four were more abundant in male inflorescence and eight were considered not significantly different between female and male inflorescences (Log_2-transformed fold-change ratios ≤0.50), while seven of the transcripts showed significantly higher expression levels in the floral tissues compared with those of shoot apical meristem, mesocarp and leaf (Figure 5.3). Among the 16 transcripts, isotig40710 (putative *DEFICIENS*), isotig53408 (putative acid phosphatase), isotig59228 (unannotated) and isotig67634 (unannotated), were more highly expressed in male inflorescence in comparison with female inflorescence whereas isotig23091 (putative *TASSELSEED1*), isotig28587 (unannotated), isotig40414 (putative bZIP transcription factor) and isotig54309 (unannotated), were more highly expressed in female inflorescence in comparison to male inflorescence (Figure 5.4).

Open reading frame analysis of the unannotated transcripts showed that several sex-specific transcripts contain putative ORFs (E-value threshold of 1E-10) which encode proteins regulating lipid and starch metabolism (Appendix C, Supplementary Table C4). The unannotated isotig28587 and isotig54309 that were highly expressed in female inflorescence, contain putative ORFs encoding a maltose excess protein and a late embryogenesis abundant protein, respectively. The male-abundant isotig67634

contains a putative ORF encoding a protein that is similar to a bacterial acyltransferase (Figure 5.4).

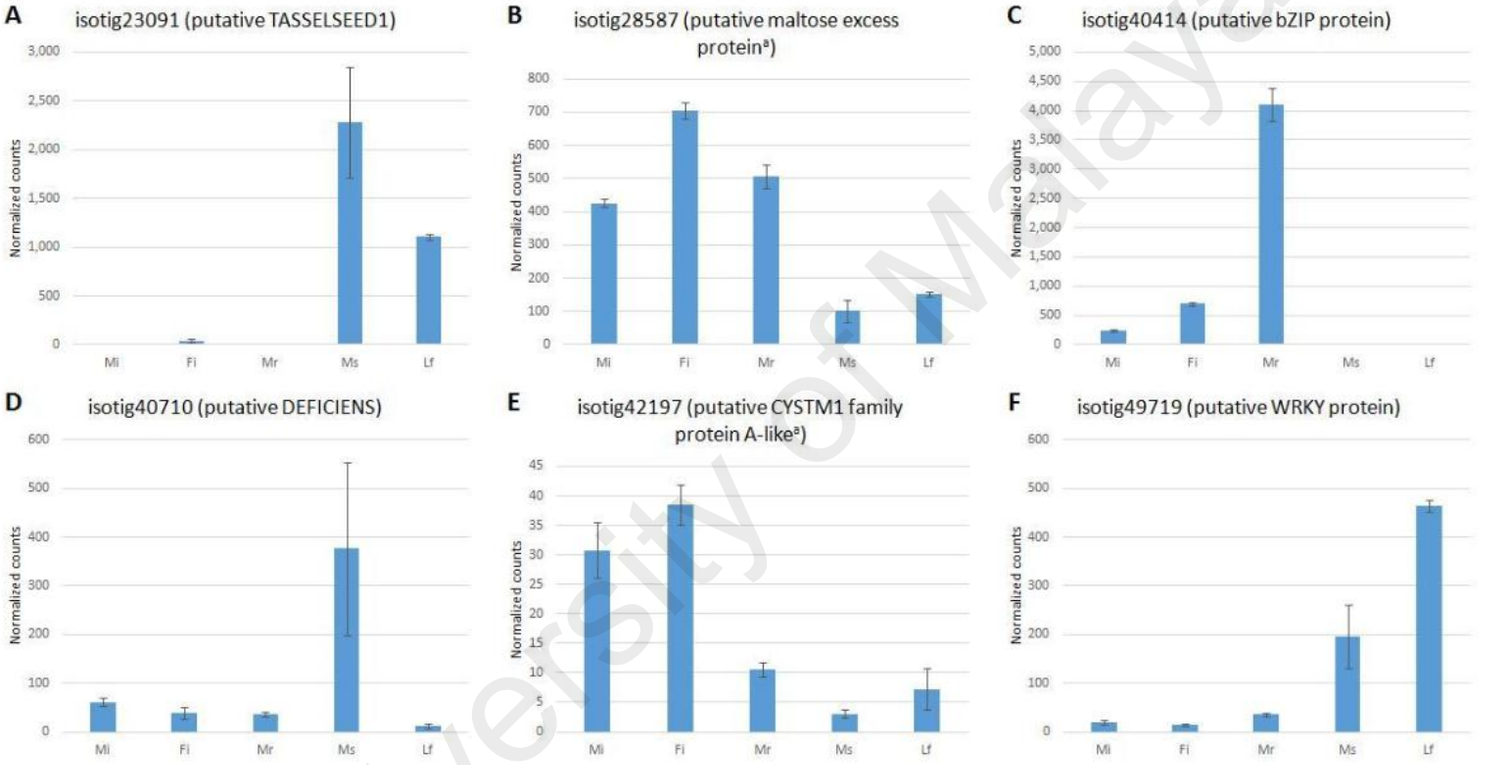


Figure 5.3: Differential expression levels of inflorescence- and sex-specific transcripts in different oil palm tissues. Transcript counts (panels A to P) were normalized using two housekeeping genes (panels Q and R). The error bars represent the standard deviation from the mean expression level calculated from four technical replicates. Mi, male inflorescence; Fi, female inflorescence; Mr, shoot apical meristem; Ms, mesocarp; Lf, leaf

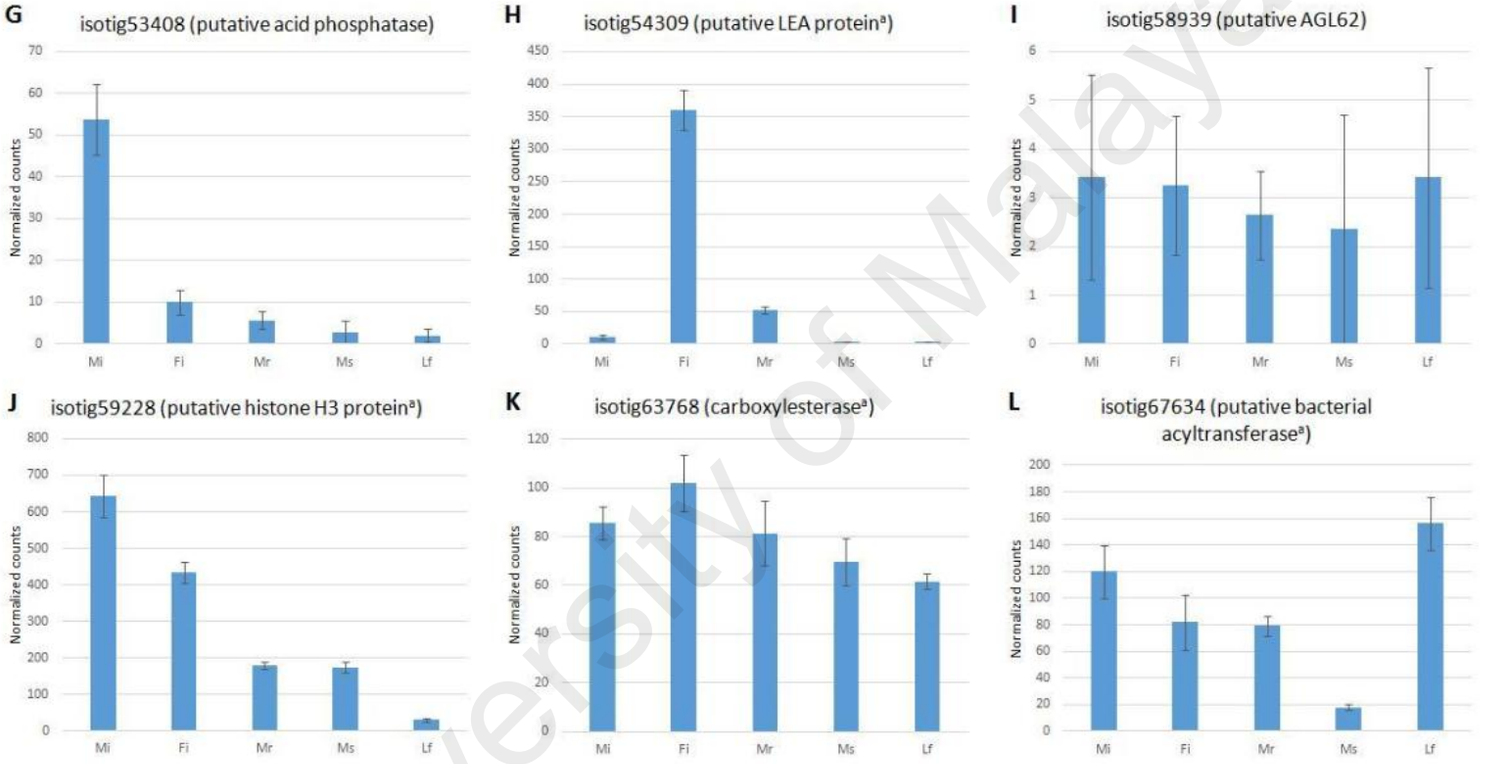


Figure 5.3, continued

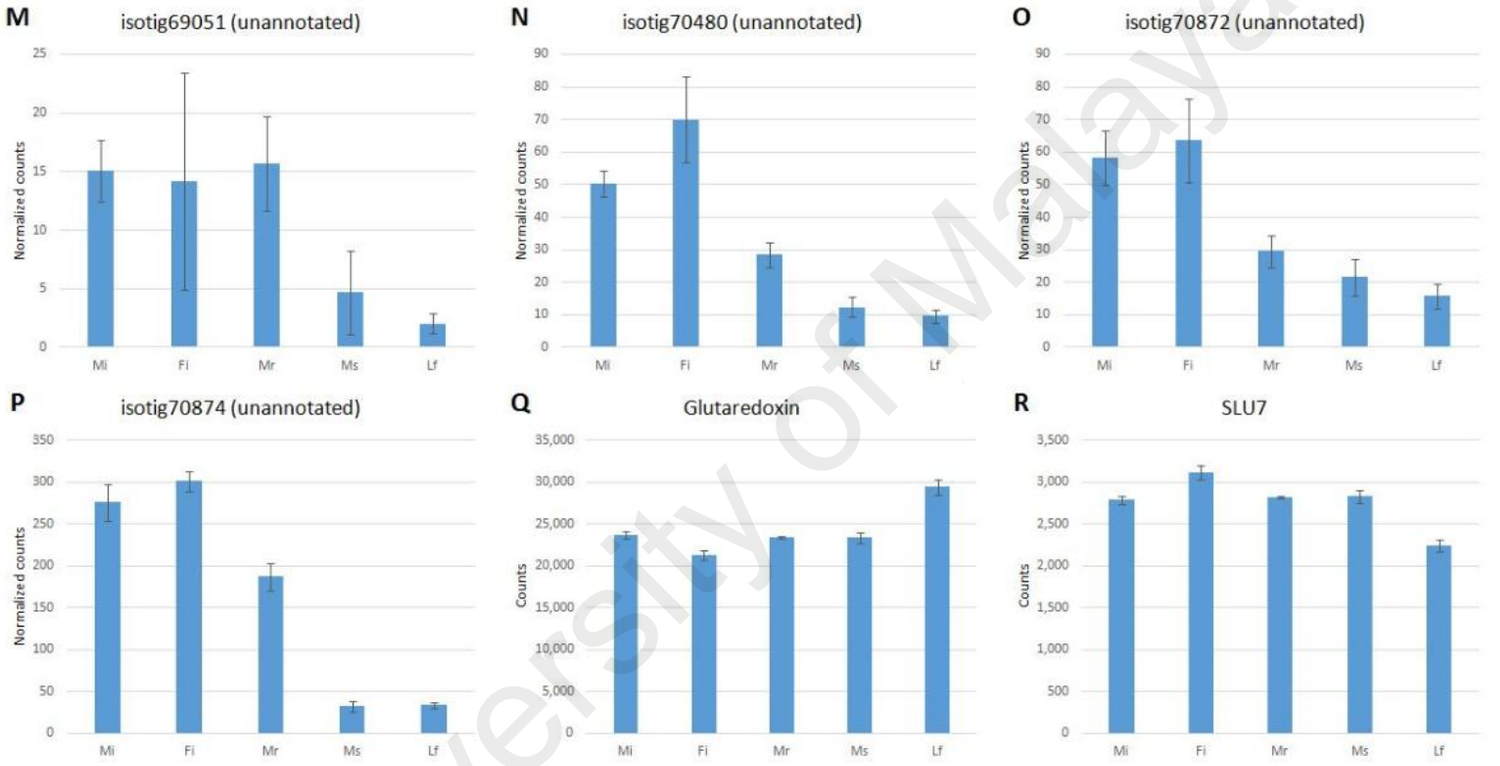


Figure 5.3, continued

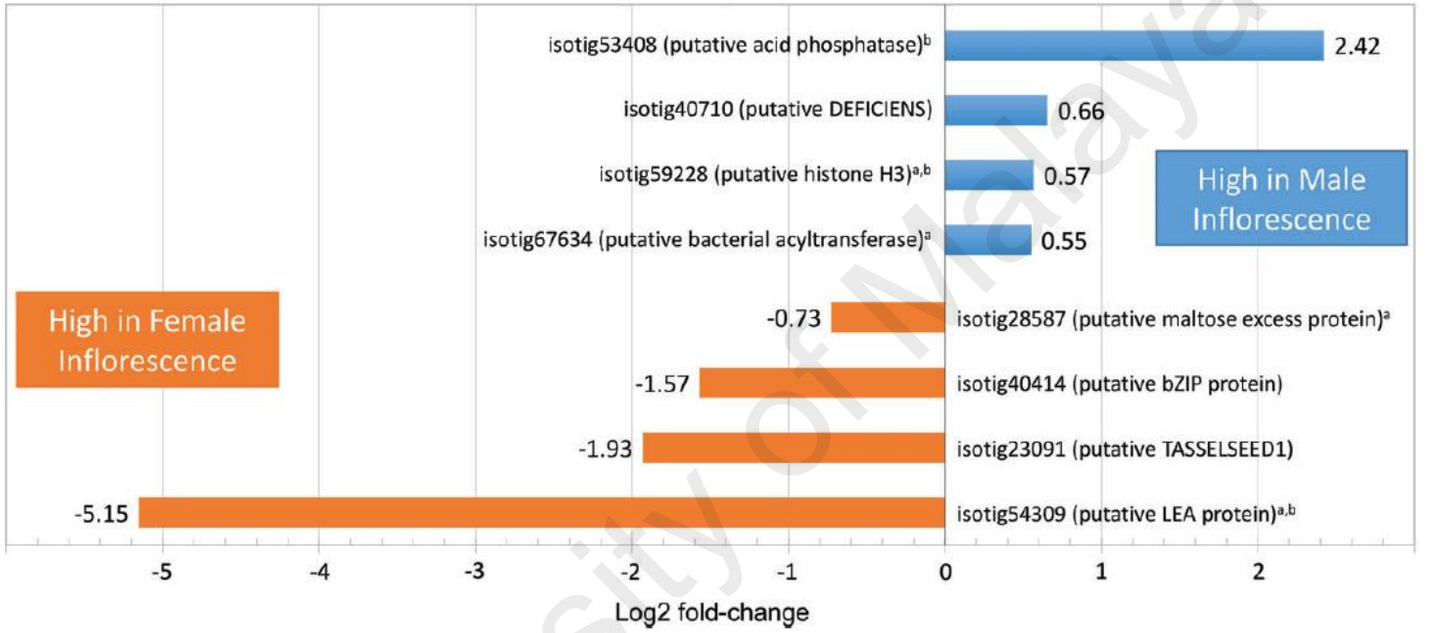


Figure 5.4: Differential expression levels of sex-specific transcripts in female and male inflorescences of oil palm. Values represent Log_2-transformed fold-change ratios of relative expression between female and male inflorescences. Only the transcripts that have Log_2-transformed fold-change ratios ≥0.50 were shown. [a] Indicates putative function based on ORF prediction. [b] Indicates that the transcript also showed inflorescence-specific expression

5.3 Discussion

5.3.1 +6 female and male inflorescences showed distinct gene expression patterns

Distinct patterns of differential expression of transcripts were observed in female and male inflorescences at the +6 stage of development (Figure 5.4), suggesting that these are sex-specific and may have a role in differential development of the inflorescence. At +6 stage, only the bracteoles are formed (Adam et al., 2005). The genes expressed at this stage are possibly involved in the upcoming development of floral organ primordia into perianth organs, followed by the gynoecium or androeciumin female and male flowers, respectively. Among the sex-specific transcripts are those coding for MADS-box proteins, WIP-zinc finger protein and *TASSELSEED1*, along with many genes that could not be annotated (Table 5.2). Analysis of the current transcriptome data for several known master regulators of flower development such as *FLOWERING LOCUS T* (*FT*), *LEAFY* (*LFY*), *APETALA1* (*AP1*) and *SUPPRESSOR OF CONSTANS OVEREXPRESSION1* (*SOC1*) showed them to be expressed in both female and male inflorescences of oil palm at the +6 stage of inflorescence development (Table 5.3). These genes are very likely a part of the conserved regulatory components in flower development of oil palm, possibly acting via similar mechanisms as in the model species *Arabidopsis thaliana* (Ó'Maoiléidigh et al., 2014). The genes with multiple oil palm transcript variants showing different levels of expression in the same tissue, but for which there has only been a single locus identified in the genome, such as *LEAFY*, might be interesting for future study as these could represent true splice variants. As the current study aimed to determine genes that were truly male or female specific, the transcripts which were common in both female and male inflorescences were excluded from the expression validation studies.

In contrast to the above group of floral developmental genes, and to an earlier report (Adam et al., 2007b) a differential expression for some of the MADS-box genes was observed in the +6 stage inflorescence (Tables 5.2 and 5.3). A previous *in planta*

study using RT-PCR (Adam et al., 2007b) did not identify any sex-specific expression patterns for oil palm MADS-box genes, thus none of them have implicated in sex differentiation of the oil palm inflorescence. While *DEFICIENS* is known to function in both petal and stamen organ identity (Zahn et al., 2005), some *DEFICIENS* family members are predominantly expressed in stamens (Laitinen et al., 2006). The higher expression level of a *DEFICIENS* ortholog (isotig40710) in the male inflorescence compared with the female inflorescence (Figure 5.4 and Table 5.2) might be indicating a male organ-specific function. The oil palm male inflorescence is comprised of solely staminate flowers thus would have a higher proportion of male organs compared to the female inflorescence, which has both rudimentary staminate flowers as well as pistillate flowers. Male-specific expression of *DEFICIENS* has also been reported for *R. acetosa* (Ainsworth et al., 1995) and *Populus trichocarpa* (Sheppard et al., 2000). The sex-specific expression pattern of the oil palm *DEFICIENS* could be temporal, as reported for *P. trichocarpa* where *DEFICIENS* was initially expressed throughout the inner whorl meristem in both female and male flowers, only becoming male-specific during the later stage of flower development (Sheppard et al., 2000). Conclusive evidence, however, will require a temporal expression study starting from the initiation of inflorescence primordia until the fully matured inflorescences are formed, a study that would be difficult and highly costly due to the long maturation phase of oil palm inflorescences and the destructive nature of sampling procedures.

The expression of *DEFICIENS* was not exclusive to the floral tissues of oil palm (Figure 5.3). Consistent with the high expression levels of MADS-box genes reported for ripening oil palm fruit (Tranbarger et al., 2011), *DEFICIENS* was also highly expressed in mesocarp, suggesting functional divergence of MADS-box proteins in different tissue types. One of the oil palm WIP orthologs, isotig39321, was more highly expressed in the male compared to the female inflorescence (Table 5.3). The function of WIP in oil palm inflorescences may be similar to that in melon, in which

high levels of CmWIP1 expression cause carpel abortion, leading to the formation of unisexual male flowers (Martin et al., 2009). While several floral sex-related transcripts have been reported in other plants (Calderon-Urrea & Dellaporta, 1999; Kazama et al., 2009), their expression patterns differed in the oil palm inflorescences studied. Using the selection criteria for sex-specificity (≥10 counts in one tissue and absent in the other), orthologous transcripts of *ANTHER EAR*, *DWARF* and *SUPERMAN* did not show sex-specific expression patterns in oil palm inflorescences (Table 5.3). This may reflect a difference in oil palm development or could be a result of the sample being outside of a short window of expression, over the very long floral development phase in oil palm, as compared to model plants like Arabidopsis and maize.

5.3.2 Inflorescence- and sex-specific transcripts in female and male inflorescences

In maize, *TASSELSEED* (*TS*) genes encode lipoxygenase, which controls the formation of unisexual flowers via selective abortion of female floral organs (Calderon-Urrea & Dellaporta, 1999; Acosta et al., 2009). The current study showed that a *TS1* transcript (isotig23091) was otherwise more highly expressed in the female inflorescence whilst *TASSELSEED2* was not expressed in neither male nor female inflorescences (Figure 5.4 and Table 5.2). As lipoxygenase has been associated with jasmonic acid-regulated female fertility (Li et al., 2001), a high level of expression of lipoxygenase may be essential for the formation of fully fertile female reproductive organs and the subsequent fruit set in oil palm. Another sex-specific candidate identified in this study, i.e., a putative bZIP transcription factor (isotig40414) showed a higher expression in the female inflorescence compared with male inflorescences (Figure 5.4). Given that bZIP transcription factors play diverse roles in plant development, the sex-biased expression pattern of this specific bZIP transcription factor suggests that it has at least some sex-specific functions in oil palm. In agreement with

the floral induction signaling mechanism of Arabidopsis in which bZIP transcription factor interacts with *FLOWERING LOCUS T* and activates the expression of *APETALA1* (Jaeger et al., 2006), our results showed that upregulation of the bZIP is coupled with a high level of expression of several *APETALA1* members in female inflorescence of oil palm (Table 5.3), supporting the potential AP1-activating properties of bZIP in oil palm.

Among the validated transcripts, a transcript coding for a putative acid phosphatase (isotig53408) was highly expressed in male inflorescence but not in all other tissues (Figures 17G and 18). In blackberry (*Morus nigra* L.), higher levels of acid phosphatase activity have been detected in male flower buds compared with female flower buds (Lal & Jaiswal,1988), and acid phosphatase activity has been correlated with pollen tube growth (Ibrahim et al., 2002). Thus, the higher level of expression of acid phosphatase in the male inflorescences of oil palm may reflect a higher phosphate mobilization rate, possibly to cater for pollen development in the staminate flowers. In contrast to the male inflorescence-specific acid phosphatase expression, a strong female- and inflorescence-specific expression of isotig54309 that contained an ORF encoding a late embryogenesis abundant (LEA) protein was observed (Figures 17H and 18). LEA proteins are commonly associated with dehydration tolerance (Hundertmark & Hincha, 2008). Under water deficit conditions, oil palms predominantly abort the female inflorescences (Corley, 1976a) thus the higher expression level of LEA protein in the female inflorescence may represent one of the physiological responses that results from the higher sensitivity of the female inflorescence to water stress. As the data from this study can only give a preliminary indication of the role of these genes in inflorescence development, it would be interesting to investigate these two transcripts further, if a feasible model system for functional validation of oil palm genes can be developed.

CHAPTER 6

DIFFERENTIAL EXPRESSION OF MIRNAS AND THEIR MRNA TARGETS DURING INFLORESCENCE DEVELOPMENT IN OIL PALM

6.1 Introduction

MicroRNAs (miRNAs) are short single-stranded RNAs (21-24 nt) which selectively repress gene expression by binding to the target mRNA, causing degradation or translational silencing of the target mRNA in a sequence-specific manner (Lee et al., 2004; Bartel, 2004). Found in almost all eukaryotic organisms, miRNAs comprise one of the key players in RNA-dependent regulation of gene expression, known as RNA interference, or RNAi, and first described as post-transcriptional gene silencing (PTGS) in plants (Llave et al., 2002; Bartel, 2004).

One of the early approaches for miRNA discovery was by direct cloning after the isolation of an enriched pool of small RNAs (usually between 18 to 25 nt in size), which are expressed in different tissues, developmental stages and environmental conditions in an organism and followed by conventional Sanger sequencing (Berezikov et al., 2006). A number of plant miRNAs have been identified via direct cloning and sequencing of small RNAs (Llave et al., 2002; Park et al., 2002; Reinhart et al., 2002; Sunkar et al., 2005). Direct cloning of small RNAs involves the use of oligo-nucleotide adapters that are ligated to each of the 5' and 3' ends of isolated small RNA by T4 RNA ligase, after which the ligation products are amplified by RT-PCR. The PCR amplicons can either be directly ligated to a cloning vector (Wang et al., 2004b; Arazi et al., 2005; Sunkar et al., 2005) or concatemerized (Ho et al., 2006; Mohanpuria & Yadav, 2012; Yew & Kumar, 2012) before ligating to a cloning vector prior to sequencing.

High-throughput sequencing or the next-generation sequencing systems are typically represented by Genome Analyzer/HiSeq 2000/MiSeq from Illumina, GS FLX Titanium/GS Junior from Roche and SOLiD/Ion Torrent PGM from Life Sciences

(summarized in Table 2.2). Next-generation sequencing technologies have the advantages of substantial reduction of cost per sequence, high throughput, high sensitivity and accuracy in genome scale analysis of miRNA repertoire (Liu et al., 2012) and have been widely applied to sequence the small RNA population in many plant species, including Arabidopsis (Fahlgren et al., 2007; Liang et al., 2015), maize (Zhang et al., 2009b; Liu et al., 2014b), rice (Wu et al., 2009a; Guoet al., 2012), poplar (Li et al., 2013; Song et al., 2013), tobacco (Guo et al., 2011; Gao et al., 2015b), tomato (Gao et al., 2015a; Zhou et al., 2016a).

The aim of this part of the thesis study was to characterize the miRNA expression profiles in developing female and male inflorescences of oil palm. This involved the deep sequencing of the small RNA population in the inflorescence tissues, followed by functional annotation of the small RNAs and miRNA differential expression analyses between female and male inflorescences. By integrating the high-throughput sequencing data from mRNA, small RNA and RNA degradome libraries, the study was further extended to specifically examine the miRNAs and their mRNA targets that were differentially expressed at two key stages (+6 and +15) of female inflorescence development, which fall within the period during which floral meristems are emerging followed by floral organ formation.

6.2 Results

6.2.1 No orthologous miRNA was identified in the small RNA cDNA library constructed via ligation of small RNAs to hybrid 3’ and 5’ adapters

The results described in sections 6.2.1 and 6.2.2 were obtained using total RNA extracted from the +6 female inflorescence of oil palm. Small RNAs of size ranging from 15-35 nt were purified from a 5% (w/v) low melting preparative grade agarose gel (Figure 6.1). The small RNAs were ligated to 3’ and 5’ hybrid adapters. RT-PCR was carried out to amplify the ligation product. The primers used for PCR amplification

were designed based on the adapters' sequences. PCR products of 60-70 bp (small RNA ligated with 3' and 5' adapters) were purified from the 1% (w/v) agarose gel (Figure 6.2).

After library construction and colony selection, 300 insert-containing pGEM-T easy vectors (randomly selected from 685 white colonies) were sequenced (Table 6.1). The library of 300 insert-containing pGEM-T easy vectors contained 42% (126/300) unannotated sequences, i.e. with no match to any known sequence in NCBI Genbank, miRBase and Rfam. Classes of small RNA found in this library were mostly tRNAs, 5.8S rRNA, 18S rRNA, 26S rRNA, chloroplast sequences and mitochondrial sequences. No orthologous miRNA sequence was found from the clones sequenced from the library.

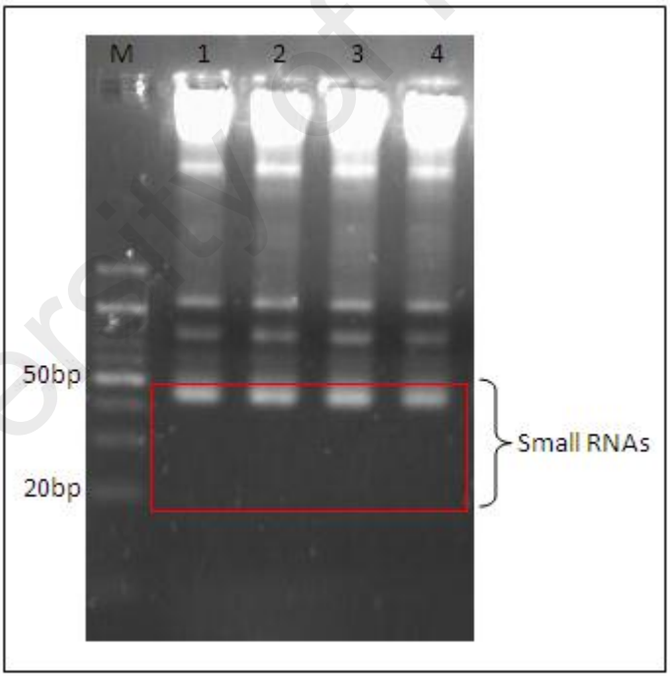


Figure 6.1: Electrophoresis of total RNA extracted from oil palm's +6 female inflorescence on a 5% (w/v) low melting preparative grade agarose gel. M: O'RangeRuler^{TM}10 bpDNAladder (Thermo Scientific^{TM}); 1-4: Four replicates (labeled as 1-4) for small RNA purification

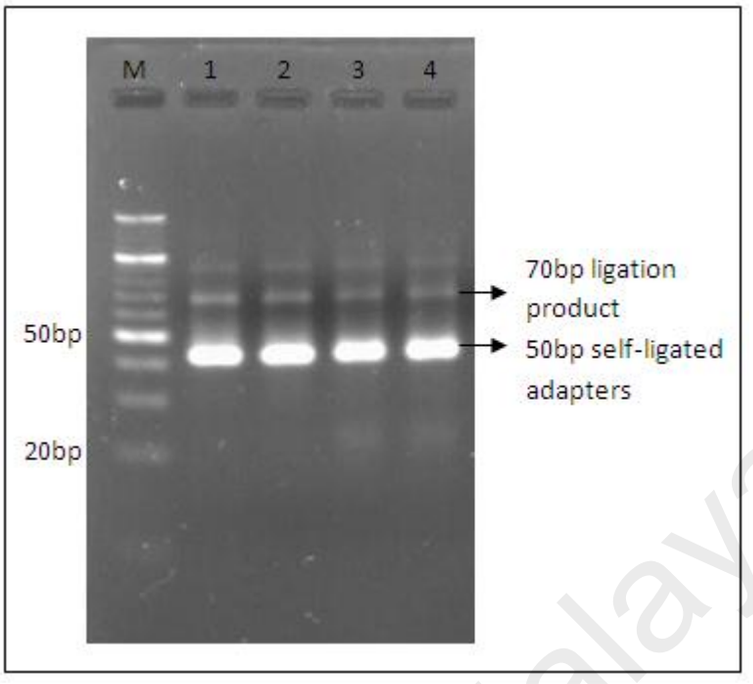


Figure 6.2: RT-PCR products of small RNAs ligated to 3' and 5' adapters. M: O'RangeRuler^{TM}10 bpDNAladder (Thermo Scientific^{TM}); 1-4: four technical replicates for RT-PCR

Table 6.1: Sequence identity of the small RNA cDNA library constructed via ligation of small RNAs to hybrid 3' and 5' adapters

Sequence identity	Number of sequence	Percentage (%)
mRNA	36	12
miRNA	0	0
rRNA	27	9
tRNA	17	5.7
Adapter sequence	58	19.3
Chloroplast sequence	15	5
Mitochondrial sequence	21	7
Unknown sequence	126	42
Total	300	100

6.2.2 miR159 and miR535 were identified via direct cloning method

PCR product with size about 110 bp was purified from electrophoresis gel (Figure 6.3). A total of 300 insert-containing pGEM-T easy vectors (randomly selected from 605 white colonies) were sequenced (Table 6.2). Similar to the previous approach (section 6.2.1), the small RNA cDNA library constructed by polyadenylation of 3' end of small RNAs, was largely composed of unannotated short sequences (155/300). Two orthologous miRNAs namely miR159 and miR535 were found in this library. Other small RNAs found in this library included fragments of tRNAs, 5.8S rRNA, 18S rRNA, 26S rRNA, chloroplast sequences and mitochondrial sequences.

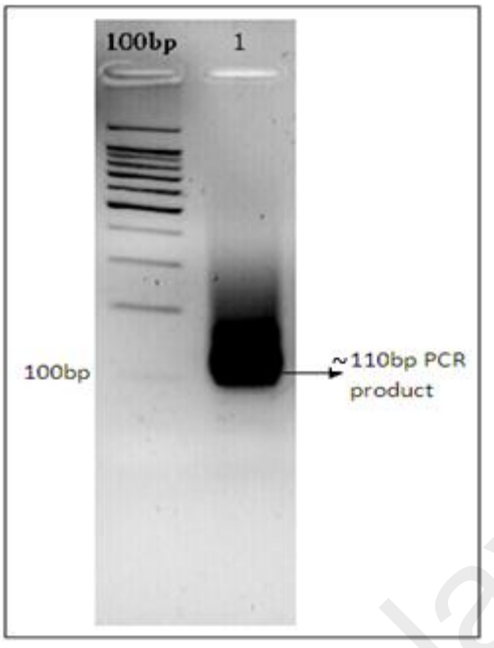


Figure 6.3: RT-PCR product of small RNAs ligated to 5' adapter and polyadenylated at the 3' end. 100bp: VC 100bp Plus DNA Ladder (Vivantis Technologies Sdn Bhd); 1: PCR product amplified from polyadenylated small RNA which was ligated to a 5' hybrid adapter

Table 6.2: Sequence identity of the small RNA cDNA library constructed via polyadenylation of 3' end of small RNAs

Sequence identity	Number of sequence	Percentage (%)
mRNA	33	11
miRNA	2	0.7
rRNA	29	9.7
tRNA	28	9.3
Adapter sequence	19	6.3
Chloroplast sequence	19	6.3
Mitochondrial sequence	15	5
Unknown sequence	155	51.7
Total	300	100

6.2.3 High-throughput sequencing of small RNA (sRNA-seq)

6.2.3.1 Small RNA population in oil palm female inflorescences, male inflorescences and leaf

The output data generated by the Genome Analyzer, GA IIx platform (Illumina Inc., San Diego, CA, USA) is summarized in Table 6.3. After removing the null inserts, adapter contaminants inserts of less than 18 nt and polyA sequences, the remaining "clean reads" were used for following bioinformatic analysis. The small RNA population in all five small RNA libraries showed similar length distribution with the 24 nt small RNA being the most predominant species, which accounted for at least 38% of the sequenced small RNA (Figure 6.4). The 21 nt small RNA was the second most abundant class of small RNA. Among the five libraries, +6 male inflorescence and leaf contained the highest percentage of 21 nt small RNA, which were 29.53% and 25.01%, respectively. Nucleotide bias analysis at each position showed that at least 70% of the orthologous miRNAs in small RNA libraries of +6 female inflorescence, +15 female inflorescence, +15 male inflorescence and leaf initate with guanine (G) whereas nearly 80% of the orthologous miRNAs in +6 male inflorescence initate with uracil (U) (Figure 6.5).

Table 6.3: Output of small RNA high-throughput sequencing

Sequencing output	Samples									
	F^{+6}		F^{+15}		M^{+6}		M^{+15}		L	
	Count	%	Count	%	Count	%	Count	%	Count	%
Total reads	18,457,329	-	16,863,063	-	12,959,297	-	12,458,031	-	12,576,041	-
High quality reads	17,031,613	100	15,760,053	100	12,689,138	100	12,196,958	100	12,298,676	100
3’ adapter null	8,011	0.05	10,113	0.06	35,277	0.28	45,656	0.37	41,079	0.33
Insert null	25,868	0.15	5,205	0.03	2,801	0.02	1,315	0.01	7,045	0.06
5’ adapter contaminant	196,438	1.15	239,128	1.52	41,245	0.33	25,124	0.21	36,765	0.30
Smaller than 18nt	365,769	2.15	500,201	3.17	582,186	4.59	71,618	0.59	227,958	1.85
PolyA	9,690	0.06	2,223	0.01	2,433	0.02	2,897	0.02	836	0.01
Clean reads*	16,425,837	96.44	15,003,183	95.20	12,025,196	94.77	12,050,348	98.8	11,984,993	97.45

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence; M^{+6}: +6 male inflorescence; M^{+15}: +15 male inflorescence; L: leaf
*After removing the null inserts, adapter contaminants, inserts less than 18 nt and polyA sequences

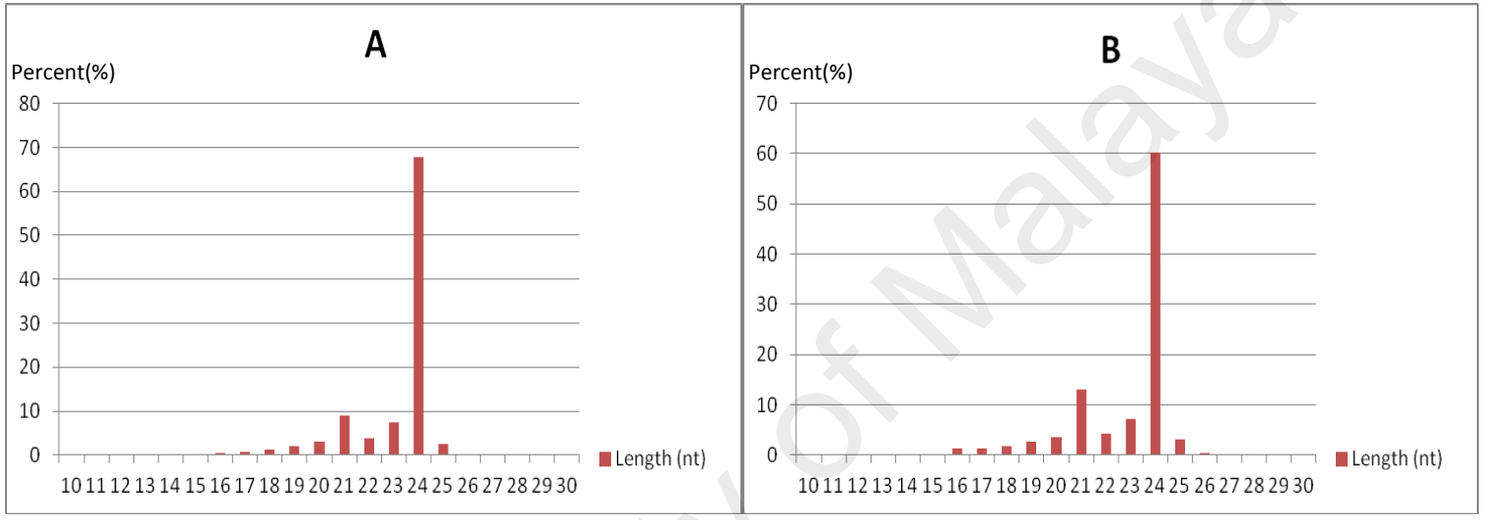


Figure 6.4: Length distribution of small RNAs. Small RNA libraries derived from (A) +6 female inflorescence; (B) +15 female inflorescence; (C) +6 male inflorescence; (D) +15 male inflorescence; and (E) leaf

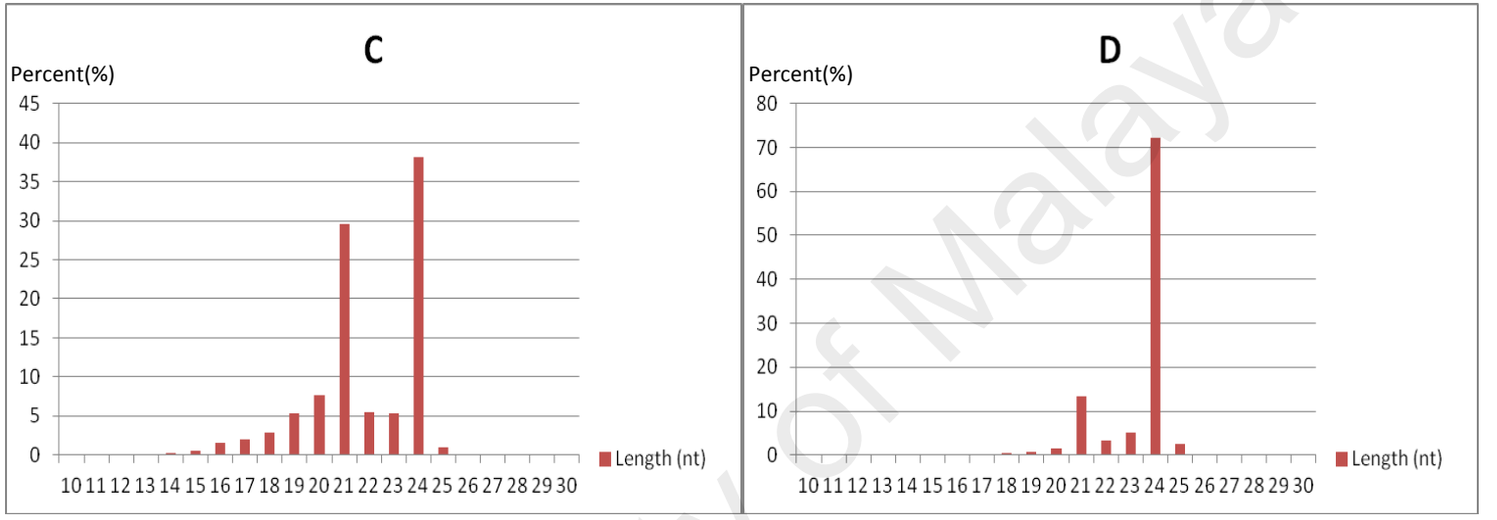


Figure 6.4, continued

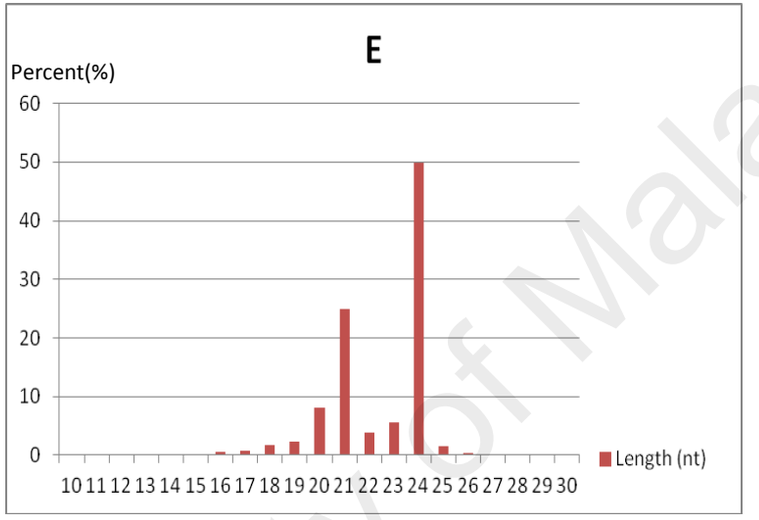


Figure 6.4, continued

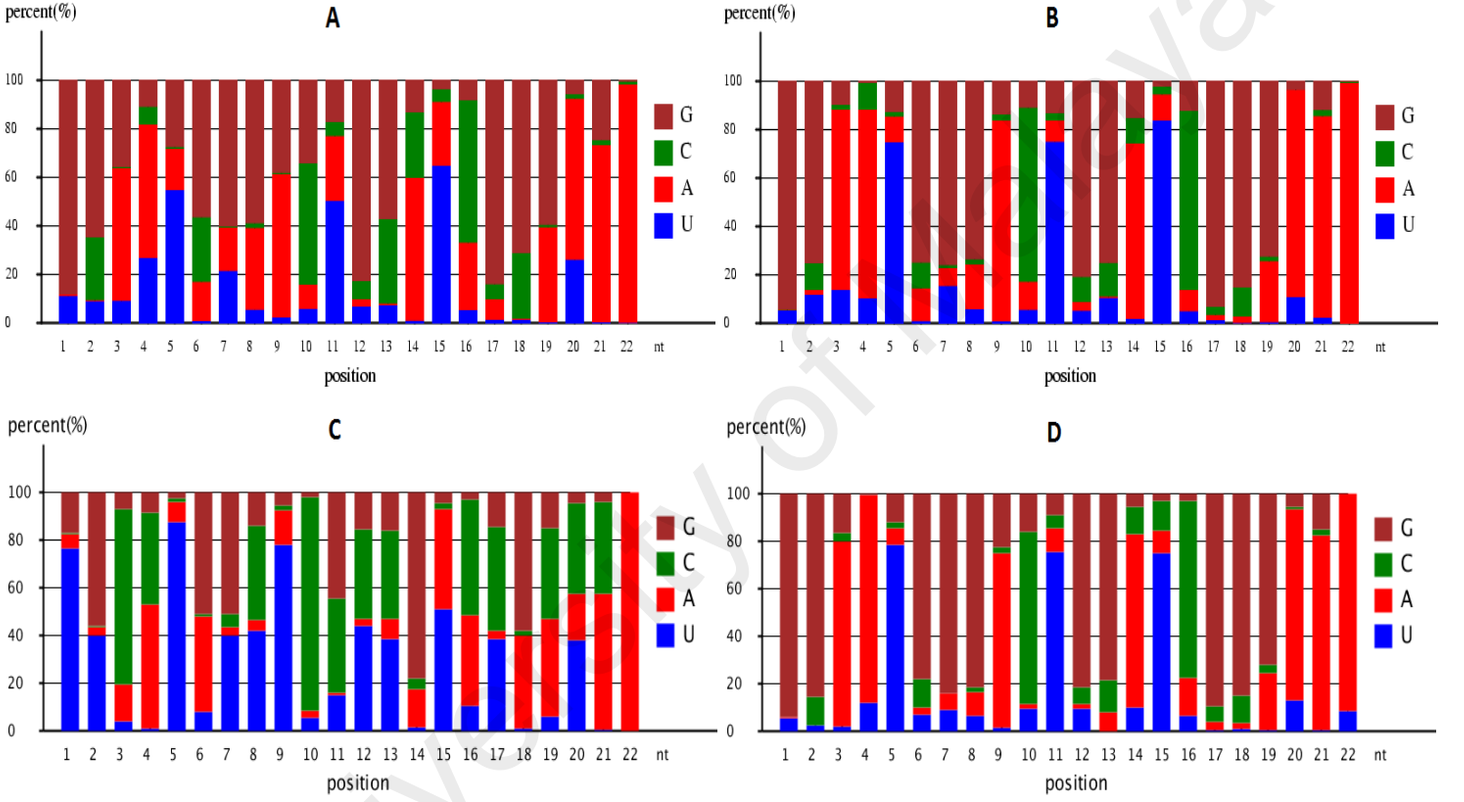


Figure 6.5: Nucleotide bias at each position of orthologous miRNAs. (A) +6 female inflorescence; (B) +15 female inflorescence; (C) +6 male inflorescence; (D) +15 male inflorescence; and (E) leaf

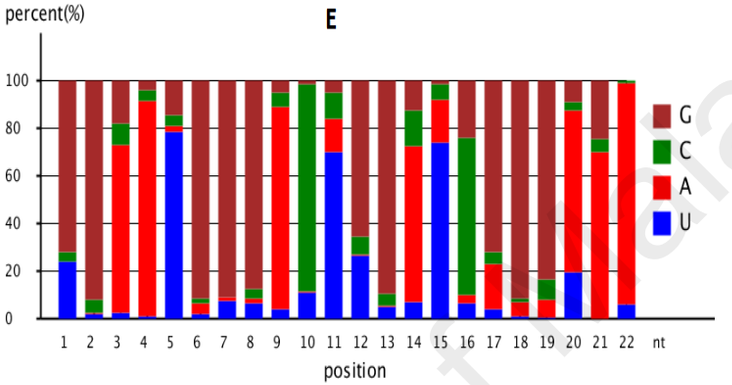


Figure 6.5, continued

6.2.3.2 Different subsets of miRNA families were expressed in female inflorescence, male inflorescence and leaf of oil palm

Over 97% of the small RNA (sRNA) in the five libraries was unannotated sequences with no known function (Figure 6.6). Among the annotated sequences, siRNA was the most abundant class of sRNA, accounting for 0.90-1.12% of the total sRNA. Orthologous miRNA sequences constituted only 0.15-0.27% of the total sRNA. The other classes of annotated sRNAs collectively constituted 0.40-0.96% of the sRNA libraries.

Matching the small RNA sequences to Plant MicroRNA Database (PMRD), 418 orthologous miRNA families were identified in +6 female inflorescence, 451 orthologous miRNA families were identified in +15 female inflorescence, 468 orthologous miRNA families were identified in +6 male inflorescence, 338 orthologous miRNA families were identified in +15 male inflorescence and 324 orthologous miRNA families were identified in leaf. Table 6.4 shows the partial list of miRNAs with highest expression in the five tissues studied. In both datasets of +6 and +15 female inflorescences, miR535 was the most abundant orthologous miRNA family, whereas in both datasets of +6 and +15 male inflorescences, miR166 was the most abundant orthologous miRNA family (Table 6.4). In oil palm leaf, miR156 was the most highly expressed orthologous miRNA family. A total of 29 oil palm-specific miRNAs that have not been reported yet in plants other than oil palm were predicted (Table 6.5). Among the predicted oil palm-specific miRNAs, only egu-miR1 was sequenced in all five sRNA libraries.

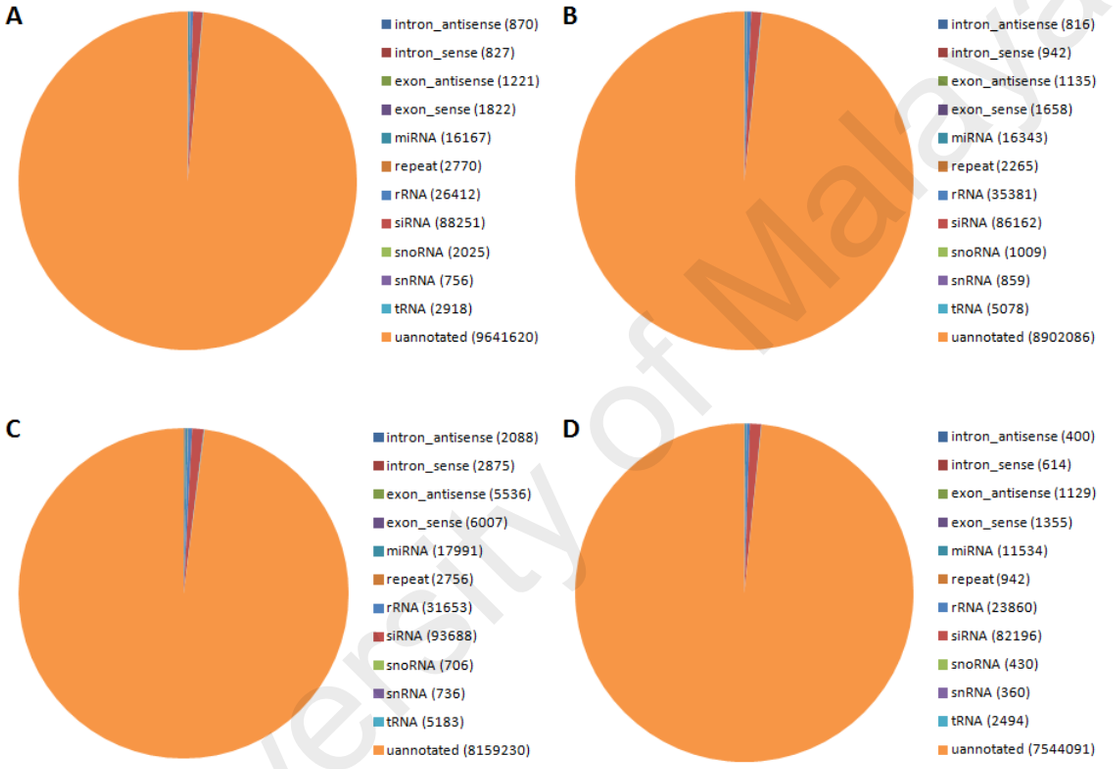


Figure 6.6: Classification of small RNA using PMRD and Rfam databases as reference. Small RNA libraries of (A) +6 female inflorescence; (B) +15 female inflorescence; (C) +6 male inflorescence; (D) +15 male inflorescence; and (E) leaf. Small RNAs that did not match to both PMRD and Rfam were classified as "unannotated".

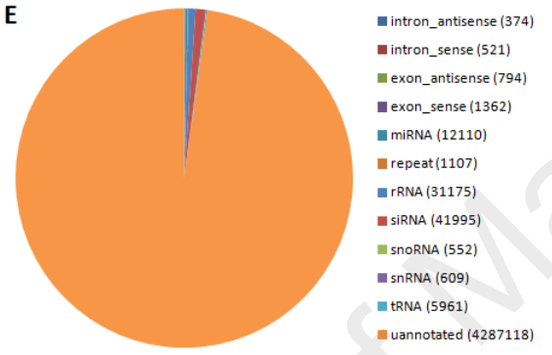


Figure 6.6, continued

Table 6.4: Most highly expressed orthologous miRNAs in female inflorescences, male inflorescences and leaf of oil palm

miRNA	Read count				
	F^{+6}	F^{+15}	M^{+6}	M^{+15}	L
miR156	15,373	13,561	1,822	1,940	2,127,576
miR159	3,730	3,483	8,735	3,383	4,545
miR160	581	11	1,205	656	924
miR162	1,123	1,955	2,816	1,589	4,423
miR164	5,189	3,133	878	2,327	13,731
miR166	2,389	18936	39,090	169,636	121,528
miR167	5,088	2,788	19,579	33,634	29,901
miR168	9,176	56,780	30,822	42,739	61,099
miR169	636	233	1,524	1,163	625
miR171	721	488	807	993	3,893
miR172	4,365	16,914	8,906	3,598	1,052
miR319	705	90	8,701	1,481	2,420
miR396	991	4,063	1,026	1,035	8,704
miR528	8,470	31,909	937	2,138	6,969
miR529	1,128	1,528	53	101	1,916
miR535	32,929	146,808	14,680	38,172	50,554
miR775	18	902	3,147	1	1
miR827	3,298	769	155	212	211
miR894	182	3,995	2,464	6,321	6,719
miR1507	9,069	13,153	6,228	8,388	1,802
miR1534	0	1,470	0	2,616	1,150
miR1852	7,329	7,677	7,908	4,987	3,403
miR1859	6,159	314	0	432	956
miR2118	1,266	356	327	865	786
miR2639	0	0	5,278	2,684	0

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence; M^{+6}: +6 male inflorescence; M^{+15}: +15 male inflorescence; L: leaf

Table 6.5: Oil palm-specific miRNAs predicted from female inflorescence

miRNA	Sequence (5'-3')	F^{+6}	F^{+15}	M^{+6}	M^{+15}	L
egu-miR1*	GGAAUGGGACUGGAUCGGGAAA	2,049	2,786	312	1,390	19,066
egu-miR2*	GAUUAGGCCAGAAGAAAGAAGC	6	8	0	0	0
egu-miR3	UGCAGCGUUGAUGAGCUCCUA	0	0	8	20	10
egu-miR4	UGCCUGGCUCCCUGUAUGCCA	0	0	453	488	976
egu-miR5	UCUUGGUAGAAGCAUGAUAUU	0	0	10	6	0
egu-miR6	AAGGGUCUGGGCAAGGGCGG	0	0	12	0	20
egu-miR7	GCGGGUGGGACGCGGAGGGGGC	7	0	0	0	0
egu-miR8	AAGAGACAGCUGACCUGAAGGU	9	0	0	0	0
egu-miR9	UGGGAGAAUGAUGAGACUGCA	162	0	0	0	0
egu-miR10*	UGGGAAUGGGGGCAUCAGGA	931	0	0	0	0
egu-miR11	UGAACAAGCAGGACUCGAGAA	0	5	0	0	0
egu-miR12	AGGGAAAGUGGAGGAGGGAGAGG	0	18	0	0	0
egu-miR13	UUUCAUUGAGAGGAAUCAGCA	0	0	6	0	0
egu-miR14	UGGGACUCGCCAAGCUUUUGA	0	0	6	0	0
egu-miR15	AAGGGAUUGGGGAAGGGCGGG	0	0	8	0	0
egu-miR16*	UGAAGCCGUGACUCCUCGACC	0	0	8	0	0
egu-miR17	CUCCAUGAUCGGAUGAUGUGG	0	0	12	0	0
egu-miR18	CAGUUCGGUAGAAAGAGGUUG	0	0	12	0	0
egu-miR19	GAAGAUCCGGGUAAGAUAUGC	0	0	36	0	0
egu-miR20	AUGCGAAGUAGAGGGUCUGCA	0	0	56	0	0
egu-miR21*	UUUGAAACGAUGUGGUAGAAA	0	0	102	0	0
egu-miR22*	UUCUUUCCUUUGCCCUAGACUG	0	0	0	8	0
egu-miR23	AAGGACUGUUUUGAGGAUGAG	0	0	0	12	0
egu-miR24	AGCUGCUGGUUUAUGGAUCCC	0	0	0	22	0
egu-miR25	GGGAAUGGGGGCAUCAGGAAAU	0	0	0	388	0
egu-miR26	UCUAGUUAGUAGAUCUUAGUA	0	0	0	0	14
egu-miR27	GGGUAGACUCUUAGAAGAAAAGA	0	0	0	0	18
egu-miR28*	CGCACAUGGCGUGGAACGAUC	0	0	0	0	110
egu-miR29	CGGGUGGGACGCGGAGGGGGCU	0	0	0	0	684

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence; M^{+6}: +6 male inflorescence; M^{+15}: +15 male inflorescence; L: leaf
miRNA-star sequence (miRNA) was identified

6.2.3.3 High abundance of 24 nt small interfering RNA (siRNA) in female and male inflorescences

Putative siRNAs were predicted by pairing any two unannotated reads which were complementary to each other and were able form 2-nt 5' and 3' overhangs. Each of the small RNA libraries of female and male inflorescences contained over 80,000 siRNA candidates, whereas the small RNA library of leaf contained over 40,000 siRNA candidates (Table 6.6), of which 95% of these siRNA candidates were 24 nt in size. Some of the siRNA candidates can be mapped to different classes of repeat elements (Figure 6.7).

Table 6.6: Overview of siRNA predicted from oil palm tissues

Sample	No. of unique siRNA
F^{+6}	88,251
F^{+15}	86,162
M^{+6}	93,688
M^{+15}	82,196
L	41,995

F^{+6}: +6 female inflorescence; F^{+15}: +15 female inflorescence;
M^{+6}: +6 male inflorescence; M^{+15}: +15 male inflorescence;
L: leaf

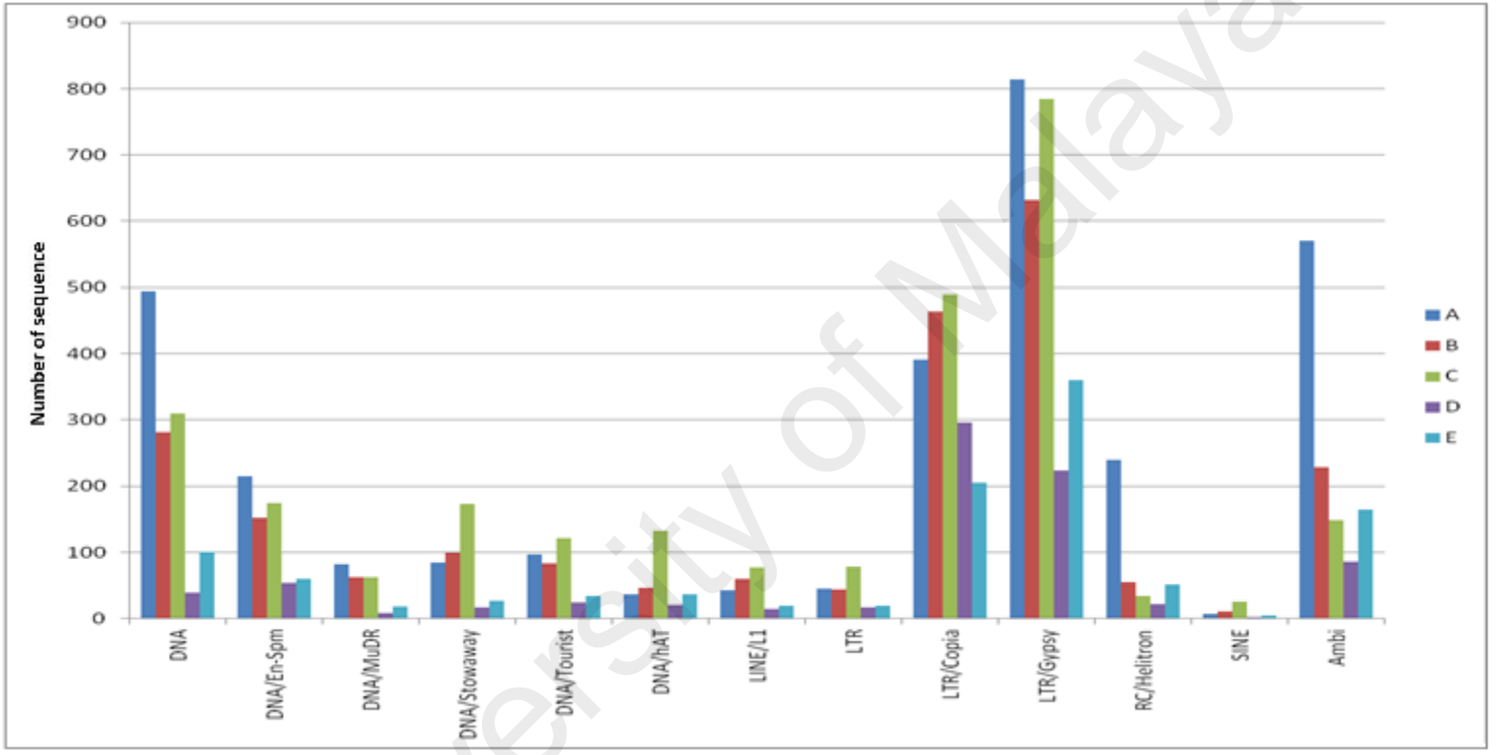


Figure 6.7: Distribution of repeat-associated 24 nt sequences in +6 and +15 female inflorescences. Ambi (ambiguous) denotes sRNAs that overlap with more than one type of repeat. (A) +6 female inflorescence; (B) +15 female inflorescence; (C) +6 male inflorescence; (D) +15 male inflorescence; and (E) leaf

6.2.3.4 Predicted miRNA precursors and miRNA targets

The transcriptomes of +6 and +15 female inflorescences were used to identify potential miRNA precursors and target mRNAs for miRNA. The precursor sequences were predicted to fold into hairpin structures with a minimal folding energy (MFE) ranging from -19.6 kcal/mol to -124.20 kcal/mol. Multiple precursor variants were predicted for several miRNA families, including miR319, miR529 and miR2118 (Table 6.7). Many of the miRNAs, both orthologous and oil palm-specific, were predicted to regulate multiple target transcripts (Table 6.7).

Table 6.7: Precursor sequences of orthologous miRNAs and oil palm-specific miRNAs

miRNA families	Number of precursors*	Number of target mRNAs*
156	1	34
159	1	17
160	0	3
164	1	14
166	0	3
167	0	8
168	1	4
169	1	3
171	1	12
172	0	11
319	3	6
396	0	8
528	0	9
529	3	16
535	0	3
827	0	1
894	0	1
1507	0	1
1859	0	3
2118	3	1
egu-miR1	1	1
egu-miR2	1	5
egu-miR7	1	3
egu-miR8	1	4
egu-miR9	1	4
egu-miR10	1	6
egu-miR11	1	1
egu-miR12	1	11

*The precursor sequences and miRNA targets were predicted from the consensus transcriptome assembled from +6 and +15 female inflorescences of oil palm.

6.2.3.5 MiRNAs and their targets were differentially expressed in developing female inflorescences

Orthologous miRNAs belonging to 381 miRNA families were differentially expressed (absolute Log_2 fold change ≥ 1 or $\leq$-1, FDR cut-off < 0.05) in +6 and +15 female inflorescences (Figure 6.8). Between +6 and +15 female inflorescences, over 90% of the differentially expressed miRNAs showed two- to one thousand-fold difference in expression levels.

Gene ontology (GO) analysis of the annotated consensus transcriptome predicted a majority of the miRNA targets as functionally related to metabolic process (biological process), binding (molecular function) and cell (cellular component) (Figure 6.9). Among the predicted targets of the differentially expressed miRNAs were transcripts coding for floral regulators such as SQUAMOSA promoter binding protein-like transcription factor (targeted by egu-miR156), auxin response factor (targeted by egu-miR167), floral homeotic protein APETALA 2 (targeted by egu-miR172) and class III HD-ZIP proteins (targeted by egu-miR166); metabolic enzyme, galacturonosyltransferase (targeted by egu-miR164). The predicted targets for oil palm-specific miRNAs included transcripts coding for aromatic amino acid family biosynthetic enzyme, phospho-2-dehydro-3-deoxyheptonate aldolase 1 (targeted by egu-miR1); transcriptional corepressor, TOPLESS-related 2 protein (targeted by egu-miR9); fatty acid biosynthetic enzyme, palmitoyl-acyl carrier protein thioesterase (targeted by egu-miR10); MOTHER of FLOWERING LOCUS T (FT)/TERMINAL FLOWER 1 (TF 1)-like proteins (targeted by egu-miR10). More than 10% of the predicted targets of the orthologous miRNAs were of unknown function (uncharacterized), whereas more than 30% of the predicted targets of the oil palm-specific miRNAs were of unknown function (uncharacterized).

Among the differentially expressed miRNAs, 18 miRNA families had TPM ≥100 in at least one of the small RNA libraries (Figure 6.10). These included miR163, miR168, miR172, miR1859, miR1873, miR396, miR4365, miR835 and miR858 which showed an inverse expression pattern to at least one of their predicted targets. Among the predicted targets of the differentially expressed miRNAs, many are transcripts coding for protein kinases and components involving in DNA replication and repair.

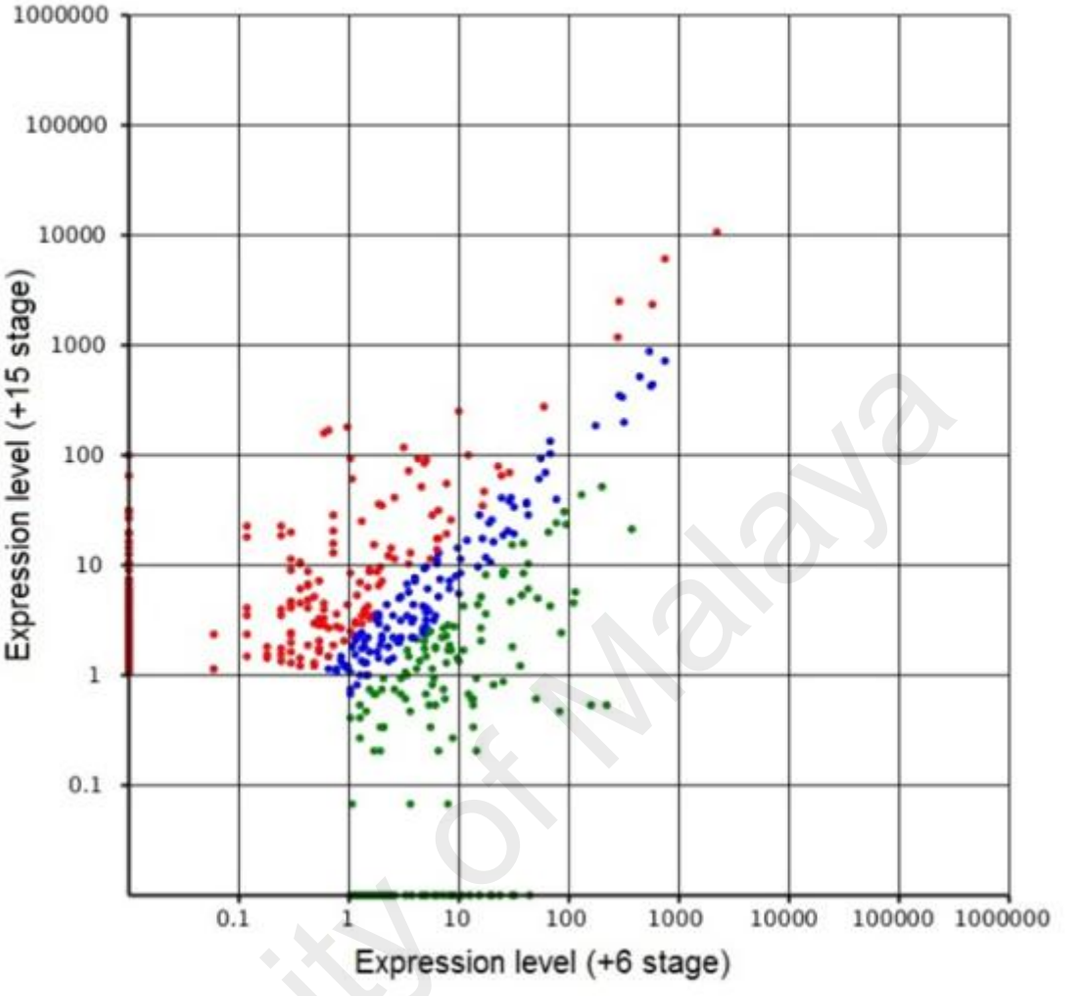


Figure 6.8: Differentially-expressed orthologous miRNAs in +6 and +15 female inflorescences. Green filled-circles indicate miRNAs that were more highly expressed in +6 female inflorescence (absolute Log_2 fold change ≥1 and FDR cut-off <0.05), red filled-circles indicate miRNAs that were more highly expressed in +15 female inflorescence (absolute Log_2 fold change ≥1 and FDR cut-off < 0.05) and blue filled-circles indicate non-differentially expressed miRNAs.

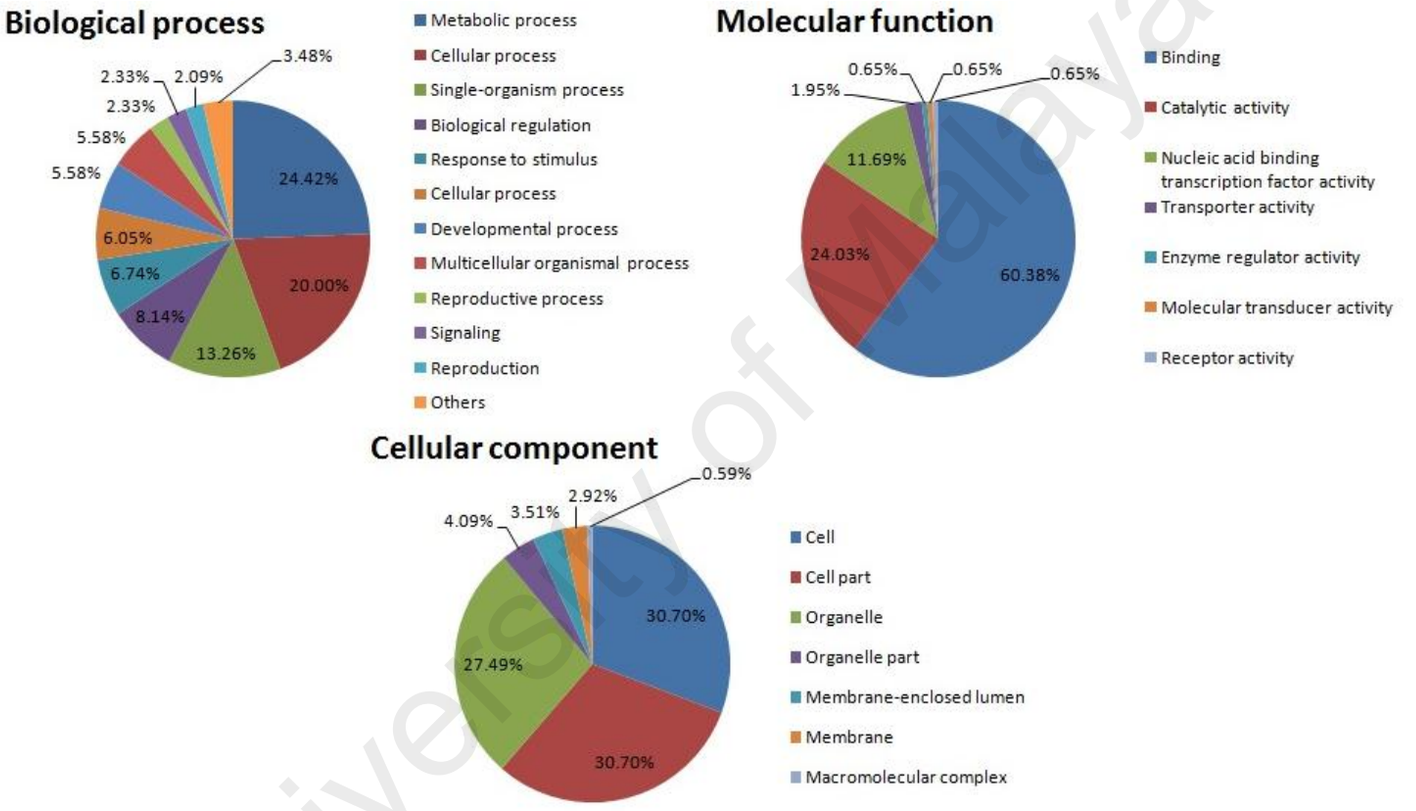
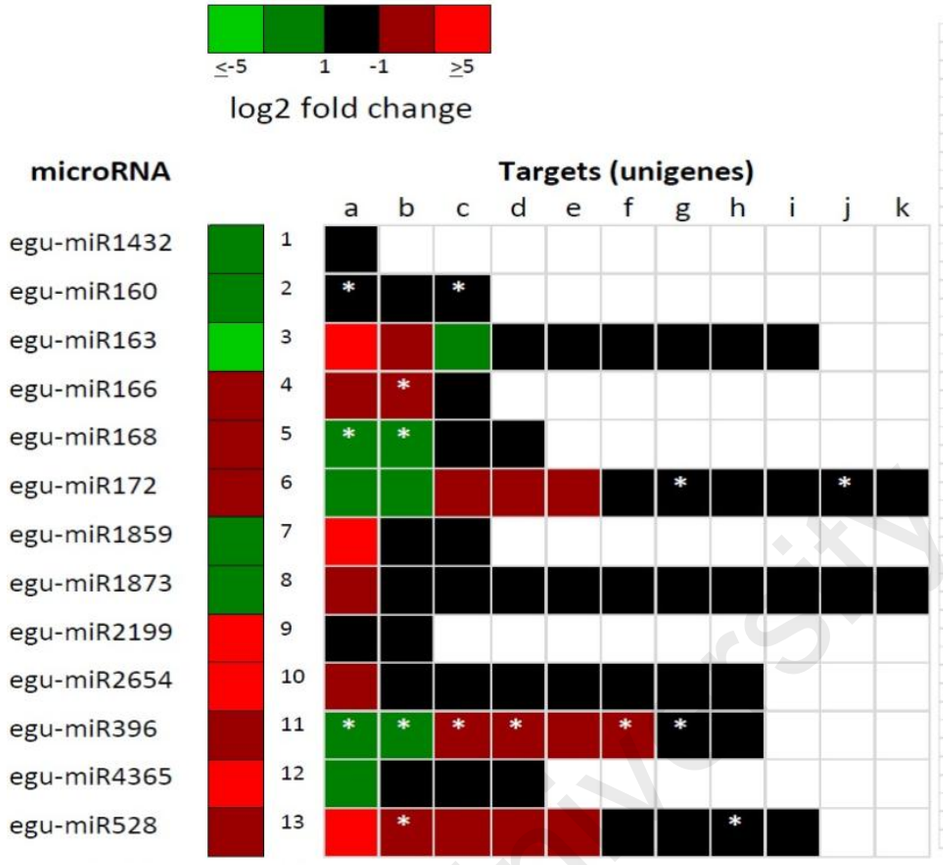


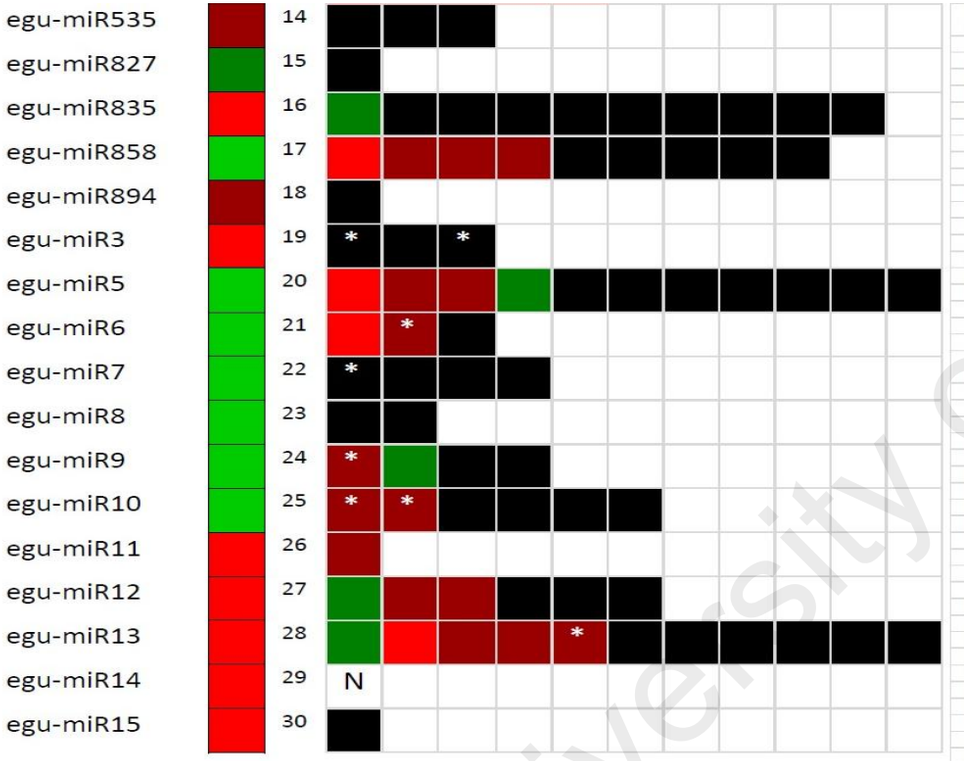
Figure 6.9: Gene ontology (GO) classification of the mRNA targets for all orthologous miRNAs identified in +6 and +15 female inflorescences. The mRNA targets of orthologous miRNAs were classified into biological process, molecular function and cellular component at the second level of GO classification.



Indicator:

1a: Mitochondrial import inner membrane translocase subunit protein (CL1Contig455)
2a, 16d, 23a, 26a: Auxin response factor (CL1416Contig1; scaffold3108; CL1Contig6581; CL1Contig504)
2b: Dolichyl-phosphate β-glucosyltransferase (CL5834Contig1)
2c: Heterogeneous nuclear ribonucleoprotein (CL1Contig1002)
3a: Flavine-containing monoxygenase (CL1Contig4016)
3b, 27f: MYB protein (CL1Contig1790; CL1Contig761)
3c: Glucomannan 4-β-mannosyltransferase 9-like (CL325Contig2)
3d: Pre-mRNA-splicing factor (CL2251Contig1)
3e: B3 domain-containing protein (CL5815Contig1)
3f: Sigma factor (C63690)
3g: Glycine-rich RNA-binding protein (CL7075Contig1)
3h: Cell number regulator 8-like (CL7664Contig1)
3i: COP9 signalosome complex subunit 6a (scaffold3993)
4a, 10d, 10e: RNA helicase (CL1874Contig1; scaffold1910; CL1Contig1573)
4b: Class III HD-ZIP protein (CL1Contig6721)
4c, 8e: Pentatricopeptide repeat-containing protein (scaffold3020; CL334Contig1)
5a, 5b: Protein argonaute 1-like (C182317; CL1Contig3683)
5c: Katanin p80 WD40 repeat-containing subunit B1 homolog 1-like (CL1Contig3430)
5d: Proton pump interactor (scaffold8460)
6a, 6b: Rolled expressed (CL1Contig7429, CL37Contig3)
6c: Polyadenylate-binding protein (Contig215)
6d, 28d: Ribosomal protein (CL1Contig1492; CL5702Contig1)
6e: Chromosome-associated kinesin (CL1Contig3061)
6f: Ethylene-responsive transcription factor (CL1Contig5600)
6g, 6h: Leaf and flower related protein (scaffold9904; CL2371Contig1)
6i, 6j, 6k, 28k: MADS-box transcription factor (CL184Contig1; CL1Contig4708; CL79Contig4; CL1Contig2657)
7a: NADH-plastoquinone oxidoreductase subunit 2 (Contig144)
7c: D111/G-patch domain containing protein (CL1Contig8760)
8a: Brefeldin A-inhibited guanine nucleotide-exchange protein 1-like (CL1349Contig1)
8b: Nucleoside diphosphate kinase-like protein (CL3360Contig1)
8c: Regulatory particle non-ATPase 13 (CL1Contig867)
8d: AMSH-like ubiquitin thioesterase 2-like (CL1239Contig1)
8f: Aminophospholipid ATPase (Contig930)
8g: β-galactosidase 10 (CL1Contig3464)
8h: Condensin-2 complex subunit d3-like (C112940)
8i: Secretory carrier membrane protein (CL7027Contig1)
8j, 22c: Protein phosphatase (CL2209Contig1; CL1Contig6820)
8k: NAD-malate dehydrogenase (scaffold7909)
9a: GAMYB-binding protein (Contig1238)
9b: BTB/POZ domain-containing protein POB1-like (CL557Contig2)
10b: L-ascorbate peroxidase chloroplastic-like isoform 2 (CL6969Contig1)
10c: TLC ATP/ADP transporter (CL5576Contig1)
10f: ARIADNE protein (CL1977Contig1)
10g: SAM domain family protein (CL1Contig8197)
10h: LETM1-like protein (scaffold2288)

Figure 6.10: Differentially expressed miRNAs in oil palm female inflorescences. The relative expression levels of orthologous and oil palm-specific miRNAs shown as the ratio of +15/+6 inflorescence RNA (absolute Log_2 fold change ≥ 1 or ≤-1 and FDR cut-off < 0.05) and their corresponding targets (+15/+6) were calculated using the RNA-seq data. The numbers and letters refer to the predicted target(s) as shown in the bottom panel. White asterisk (*) indicates target for which cleavage products were identified in the degradome data.



No annotation: 10a, 17b, 24b, 24d, 25c, 28a, 28c (scaffold9681; CL346Contig1; CL3559Contig1; CL284Contig2; CL1Contig1094; CL1Contig4610; CL7543Contig1)

Protein/ uncharacterized protein: 7b, 13a, 13c, 13h, 13i, 16a, 21a, 22d, 25e, 25d, 28b, 28e (CL1Contig8073; CL14Contig4; CL1070Contig1; CL1459Contig1; scaffold9206; CL1Contig6857; CL1Contig4691; C88392; CL6500Contig1; CL5996Contig1; CL4085Contig1; CL1Contig9019)

No target match: 29a (N)

11a, 11b, 11c, 11d, 11f:: Growth-regulating factor (CL2200Contig1; CL2840Contig1; CL1Contig7158; CL1Contig193; CL3049Contig1)
11e, 16g: GDSL esterase lipase (CL893Contig1; C38072)
11g: Galactoside 2-α-l-fucosyltransferase-like (CL1226Contig2)
11h: PAP-associated domain-containing protein 5-like (CL1Contig8141)
12a: ATP-dependent helicase RHP16-like (CL1796Contig1)
12b: GDP-mannose transporter (C117016)
12c: DNA repair protein homolog (C56372)
12d: Integral membrane protein (scaffold2487)
13b: RNA recognition motif-containing protein (CL1Contig7172)
13d: Probable membrane-associated kinase regulator 2-like (CL1Contig1057)
13e: Pseudo response regulator (CL1Contig8870)
13f: D-lactate dehydrogenase (CL2994Contig1)
13g: Regulator of chromosome condensation repeat-containing protein (CL1213Contig2)
14a: Potassium channel (C186299)
14b: Nucleoprotein TPR (CL1630Contig1)
14c: UDP-N-acetylglucosamine 1-carboxyvinyltransferase (CL1Contig7939)
15a: BWF1-like protein (CL2490Contig1)
16b: Putative transcription factor (Contig288)
16c: Sulfite reductase (CL1Contig3003)
16e, 17i: BRI1-KD interacting protein (scaffold2494; C128680)
16f, 17h, 20d, 25a, 27b, 27c, 27d, 27e: Protein kinase (scaffold7867; CL459Contig1; CL4957Contig1; CL1Contig5727; CL7477Contig1; CL37Contig4; CL3328Contig1; CL1Contig5184)
16h: Fimbrin-like protein 2-like (CL1Contig7669)
16i: ATP citrate lyase α-subunit (CL1Contig8649)
16j: Glycosyl hydrolase family 38 protein (CL1Contig1958)
17a: Nucleolysin TIAR-like (Contig182)
17c: Calmodulin-binding protein (CL1Contig1095)
17d: RNA polymerase α subunit (Contig155)
17e: Kinase interacting (KIP1-like) protein (CL7463Contig1)
17f: Nuclease HARBI1-like (CL1Contig7091)
17g: Flap endonuclease (C63946)
18a: Choline transporter-like protein (CL1Contig7359)
19a, 23b: GTP-binding protein (CL1Contig5623; CL1Contig3120)
19b: Xaa-Pro aminopeptidase 1 (scaffold8985)
19c: Formate dehydrogenase (CL1Contig8779)
20a: Nascent polypeptide-associated complex subunit α-like (CL1Contig873)
20b: ADP-ribosylation factor (CL1Contig143)
20c: Saposin-like type region 1 family protein (C127132)
20e: CONSTANS interacting protein 5-like protein (CL1Contig6636)
20f, 22b: Chromatin assembly factor (CL1Contig8690)
20g: Anaphase-promoting complex subunit 7 (CL1Contig603)
20h: DEK domain-containing chromatin associated protein (Contig315)
20i: ATP synthase subunit A chloroplastic (CL1Contig6099)
20j: Cytochrome C oxidase assembly protein (Contig998)
20k: Mediator of RNA polymerase II transcription subunit 31 (scaffold10854)
21b: Cyclin dependent kinase inhibitor (CL1Contig825)
21c: Transparent testa 1 (CL715Contig1)
22a: Histone acetyltransferase type b catalytic subunit (CL1Contig452)
24a: Growth-regulating factor (CL1Contig193)
24c: Diphosphoinositol polyphosphate (CL4925Contig1)
25b: Proteasome subunit α (CL990Contig1)
25f: Acyl-ACP thioesterase (CL2452Contig1)
27a: Cyclin-dependent kinase f-4-like (scaffold11359)
28f: Global transcription factor group (CL1Contig3508)
28g: F-box protein PP2-A13-like (scaffold8278)
28h: Pectinesterase inhibitor (CL1Contig8908)
28i: Somatic embryogenesis receptor kinase (scaffold2072)
28j: Eukaryotic translation initiation factor 1A-like (CL1Contig8494)
30a: 26S proteasome non-ATPase regulatory subunit 13 (CL1Contig8847)

Figure 6.10, continued

6.3 Discussion

6.3.1 Direct cloning and sequencing of small RNAs has a low rate of miRNA identification

The direct cloning method via ligation of the sRNAs to hybrid 3' and 5' adapters (Ho et al., 2006) did not capture any orthologous miRNA in the small RNA cDNA library produced (Table 6.1). In contrast, similar method has been successfully applied and indentified a total of 22 miRNA families in Arabidopsis (Xie et al., 2005), possibly because a much larger number of clones (a total of more than 5,000) were sequenced and analyzed. The resultant library contained an exceptionally high percentage (42%) of sequences with unknown identity, a feature which was also found in the high-throughput sequencing data (Figure 6.6). Such sequences may derive from unannotated regions of oil palm genome, the unassembled contigs of oil palm genome assembly, or could represent contaminants of unconfirmed origin. More than 19% of the clones contained adapter dimers resulted from self-ligation concatenation of 3' and 5' adapters (Figure 6.2 and Table 6.1). Efficiency of ligation of adapters to the templates has been reported to be improved by modifications of the templates or adapters, such as by polyadenylation of 3' end of the templates (Wang et al., 2004b), dephosphorylation of the 5'-end of the templates and the use of adenylated DNA adapters (Zhuang et al., 2012), with such modifications incurring additional cost.

Due to the self-ligation of the 3' and 5' adapters, instead of ligating the small RNAs to a 3' adapter, the 3' end of small RNAs was polyadenylated using poly(A) polymerase before ligating to a 5' adapter (an RNA-DNA hybrid). The resultant small RNA cDNA library contained sequences that are similar to that produced by the direct cloning method as described by Ho et al. (2006), except for two orthologous miRNAs (miR159 and miR535). The method of polyadenylating the 3' end of small RNAs had a low frequency of miRNA detection, i.e. only 0.7% of the content was represented by orthologous miRNAs (Table 6.2). The success rate of cloning miRNAs may be

proportional to the expression level of a miRNA, hence the success rate may be expected to be even lower for those low expressed miRNAs. The percentage of adapter sequence in the library was much lower compared to the cloning method described by Ho et al. (2006), indicating an improvement in the efficiency of adapters ligating to the 3' end- polyadenylated templates. Both cloning methods have similar limitations: the process is technically laborious when dealing with large numbers of clones and the high Sanger sequencing cost incurred, making these methods less effective and less cost-efficient for large scale miRNA discovery studies. As the current study was limited to only 300 clones, the performance in terms of frequency of orthologous miRNA detection may not be directly comparable to that of the high-throughput sequencing which produces millions of sequence reads for each sRNA cDNA library. Nonetheless, for the interest of the current study, high-throughput sequencing of small RNA (sRNA-seq) was chosen as an alternative for large scale discovery and profiling of miRNA expression in oil palm tissues.

6.3.2 Oil palm miRNAs showed tissue-specific nucleotide bias

Due to the low miRNA detection rate of the direct cloning approaches, high-throughput sequencing method was used to enable large scale identification and expression profiling of small RNA in various oil palm tissues. Although deriving from the same plant, different patterns of nucleotide bias were observed in the five oil palm tissues studied. In small RNA library of +6 male inflorescence, uracil (U) dominated the start of the majority of orthologous miRNAs (Figure 6.5). Uracil (U) has been reported to dominate the first position of the 5' end of miRNAs in a number of plants, including Arabidopsis (Mi et al., 2008), legume (Formey et al., 2014), maize (Liu et al., 2014b), rice (Bakhshi et al., 2016) and wheat (Chen et al. 2015b), with the pattern of nucleotide bias varies according to the length of miRNAs (Liu et al., 2014b; Guo et al., 2016). Different argonaute (AGO) proteins preferentially recruit small RNAs with

different 5′ terminal nucleotide (Mi et al., 2008); previous report has shown that AGO1 proteins have more affinities towards miRNAs with a 5’ terminal uracil, thus resulting in uracil bias in the first position of cloned miRNAs (Mi et al., 2008).

In small RNA libraries of both +6 and +15 female inflorescences, +15 male inflorescence and leaf, guanine (G) dominated the first nucleotide of orthologous miRNAs (Figure 6.5), as similarly seen in zoonotic parasites, *Fasciolopsis buski* (Chen et al., 2016), *Taenia solium* and *Taenia asiatica* (Ai et al., 2014). The ribonuclease Dicer excises mature miRNAs from miRNA precursors (pre-miRNAs) (Rogers & Chen, 2013). Starega-Roslan et al. (2015) has shown that Dicer tends to avoid cleaving at sites that produce miRNAs with a guanine at the 5′ terminus. The heterogeneity of 5’ terminus in different oil palm tissues as observed in the current study suggests that Dicer activity may exhibit some tissue-specific variation, probably to enable assortment of miRNAs among different classes of AGO proteins as well as to allow miRNAs to target on a different range of target mRNAs in different tissues. While the seed region of miRNAs has been reported to determine miRNA-mediated silencing efficacy in human HeLa cells (Hibio et al., 2012), analysis of complementarity requirements for miRNA targeting in *Nicotiana benthamiana* revealed that mismatches to the miRNA 5’ regions significantly reduce or eliminate silencing efficacy (Liu et al., 2014a). The nucleotide bias patterns for the starting position(s) of the miRNAs as observed in this study may therefore play a role in determining the silencing efficiency of the oil palm miRNAs, as well as regulating miRNA actions, such as by binding to different mRNA or different positions on the same mRNA for gene regulation.

6.3.3 High-throughput sequencing identified orthologous and 'novel' oil palm-specific miRNAs

In comparison to the conventional methods for cloning and sequencing of small RNA, high-throughput small RNA sequencing allows large scale identification orthologous miRNAs (Table 6.4) as well as prediction of "novel" miRNA candidates (Table 6.5) that had not been previously reported yet in plants other than oil palm, i.e. the "oil palm-specific miRNAs". In female inflorescence, miR535 was the most highly expressed orthologous miRNA family (Table 6.4). Due to its high sequence similarity to miR156/157 family, miR535 may regulate *SPL* (SQUAMOSA Promoter Binding Protein Like) genes in conjunction with miR156/157 members (Carra et al., 2009). *SPL* genes encode transcription factors which are involved in various plant growth and developmental processes including flower and fruit development (Chen et al., 2010). Mica et al. (2010) has suggested that miR535 could possibly be involved in maturation of berries (*Vitis vinifera*) although its regulatory mechanism remains unknown. In male inflorescence, miR166 was the most highly expressed orthologous miRNA family (Table 6.4). The members of miR166 family have been shown to regulate shoot apical meristem and lateral organ formation of Arabidopsis by regulating the homeodomain-leucine zipper (HD-ZIP) gene family (Williams et al., 2005). Zhu et al. (2011) has shown that AGO10 specifically interacts with miR166/165 to maintain the shoot apical meristem. In oil palm leaf, miR156 was the most highly expressed orthologous miRNA family. Regulation of the SBP-box transcription factors (*SPL9* and *SPL15*) by miR156 has been demonstrated to affect floral transition which the over-expression of miR156 resulted in delay of flowering in Arabidopsis (Schwarz et al., 2008). The repression of *SPL9* and *SPL* in *MIR156* over-expressing transgenic Arabidopsis has caused phenotypic changes including altered inflorescence architecture and enhanced branching.

Other than sequencing reads that are allocated to known RNA groups, there were large numbers of sequencing reads mapped to genomic regions without annotation. These reads potentially originate from miRNA genes which are not previously identified. In this study, a total of 29 'novel' oil palm-specific miRNA candidates were predicted (Table 6.5). These miRNA candidates were sequenced in at least one of the five small RNA libraries, with at least five read counts. While the highly expressed miRNAs in the tissue samples studied were all orthologous miRNAs, most of the oil palm-specific miRNAs (except for egu-miR1) were minimally expressed. Except for egu-miR1, most oil palm-specific miRNA candidates were expressed at a low level (< 1000 read counts) in one, or few of the small RNA libraries, suggesting temporal- and/or tissue specific- expression. Expression of miRNAs could be confined to specific layers of cells (Boualem et al., 2008); the expression level of a miRNA may be high only within the local cells distributed over a confined area such as the floral meristem, but not the entire inflorescence tissue. The presence of a diverse set of miRNAs (orthologous and oil palm-specific), with each having different expression levels at different developmental stages, shows the complexity and heterogeneity of miRNA-mediated gene regulation inherent to oil palm.

6.3.4 24 nt sRNA is the most abundant class of sRNA in oil palm's inflorescences and leaf

In the five oil palm tissues studied, the 24 nt sRNA represented the most abundant class of sRNA (Figure 6.4), many of which were predicted to be short interfering RNAs (siRNAs). In contrast to miRNAs (generally 21-24 nucleotides) which derive from long stem-loop precursors transcribed from endogenous genomic DNA by RNA polymerase II (Bartel, 2004; Lee et al., 2004), siRNAs are processed from double-stranded RNAs, (introduced directly into the cytoplasm or taken up from

the environment) by Dicer-like 3 (DCL3) (Meister & Tuschl, 2004; Mello & Conte, 2004).

A majority of the 24 nt sRNAs mapped to the loci of repeats and transposable elements, with the highest number of sequences matching long terminal repeats (LTRs) (Figure 6.7). For a number of plants with high repetitive genome content, such as rice (*Oryza sativa*) (Morin et al., 2008), legumes (*Medicago truncatula* (Szittya et al., 2008) and *Arachis hypogeal* L. (Chi et al., 2011)) and wheat (*Triticum aestivum* L.) (Sun et al., 2014), similar high abundance of 24 nt sRNA population have also been reported. However this is not always the case, as in banana the 21 nt long RNA was the most abundant sRNA (Lee et al., 2015). The oil palm genome is rich in repeats, retroelements and transposons (Singh et al., 2013); the high abundance of 24 nt sRNAs in our small RNA libraries may reflect a high level of RNA-mediated epigenetic processes to maintain genome stability. As the methylation level of retrotransposon elements has been reported to contribute to the mantled phenotype of oil palm fruit (Ong-Abdullah et al., 2015), a process which is mediated via siRNA in plants, this illustrates the functional and potentially economically important impact of siRNA-mediated gene silencing during plant development. Furthermore, it has also been shown that methylation of miRNA genes regulates gene expression in flowers of poplar (Song et al., 2015), indicating a link between siRNAs and miRNAs in the genetic regulatory circuits of flower development. As revealed by transcriptome data, the expression levels of transcripts coding for RNA-dependent RNA polymerase (CL1Contig4817, CL1Contig8572, CL3Contig2, scaffold4316) and argonaute proteins (C182317, CL1723Contig1, CL1Contig3683, CL1Contig3831, CL1Contig4924, CL1Contig5909, CL1Contig6775, CL1Contig7916, CL1Contig8914, CL85Contig2 and Contig187) were differentially expressed in +6 and +15 female inflorescences, reflecting temporal variations in the biosynthesis and activity of siRNAs, and hence the developmental stage-specific methylation pattern of the genome.

The current study predicted a large number of siRNAs, presenting interesting candidates as epigenetic and post-transcriptional gene silencing effectors in the regulation of inflorescence development. These siRNA candidates, however, require further bioinformatics analysis by mapping back to the oil palm reference genome, and support with downstream functional studies before their roles can be confirmed.

6.3.5 Orthologous miRNAs target flowering-related genes

The current transcriptome datasets were derived from only the +6 and +15 female inflorescences, thus the differential expression analysis of miRNAs and their targets was limited to these two tissues. A large number of miRNAs were differentially expressed in +6 and +15 female inflorescences, reflecting a high degree of miRNA-mediated regulation of gene expression in the tissues studied. Among the differentially expressed miRNAs, several flowering-related miRNAs showed expression patterns that are similar to those of model plants (Huijser & Schmid, 2011), albeit the magnitude of change in miRNA expression during reproductive development may vary from plant to plant. Taking miR160 and miR172 as examples which showed opposite expression patterns in developing oil palm female inflorescence, miR160 was highly expressed at the +6 stage whereas miR172 was highly expressed at the +15 stage. While sequentially operating miRNAs (miR156 and miR172) have been reported to regulate juvenile-to-adult transition in Arabidopsis (Wu et al., 2009b), validation using RT-qPCR did not show significant difference in the expression levels of miR156 in +6 and +15 female inflorescences of oil palm (Figure 7.2). It is likely that downregulation of miR156 had occurred at a much earlier stage, probably during the switch from vegetative meristem to inflorescence meristem. The sequential expression patterns of distinct subsets of miRNAs observed in this study could possibly coordinate the synchronization of developmental timing in oil palm female inflorescences. On the other hand, the high expression of miR172 in the developing female inflorescences

appeared to continue in subsequent phases, i.e. the emerged and mature female flowers of oil palm (Mehrpooyan et al., 2012), reflecting an active requirement for miR172 across inflorescence and floral development. In agreement with the previous report by Somyong et al. (2016), we found a high expression of miR159 in both +6 and +15 female inflorescences, supporting their potential involvement in female flower determination. Despite the high sequence similarity between miR159 and miR319, they do not co-regulate the same targets (Spanudakis & Jackson, 2014). Similarly, although there is a high sequence similarity between miR156/157 and miR535, miR156/157 was more highly expressed at the +6 stage, while miR535 was more highly expressed at the +15 stage (Figure 6.10), suggesting a divergence of roles for the different family members.

A single miRNA can regulate the expression of multiple genes (Lim et al., 2005). Functional annotation of miRNA targets revealed that a miRNA may regulate not only flower homeotic genes, but also functionally diverse genes involved in broad categories of pathways (Figure 6.9). A previous study has demonstrated that miRNAs may selectively regulate their targets in a dose-dependent and non linear manner (Shu et al., 2012), thus the preferred mRNA targets may vary over the developmental stages, depending on the expression level of the regulating miRNAs. Nevertheless, it would be expected that floral developmental processes do not exist in isolation, but rather in the form of crosstalk between diverse cellular and metabolic pathways via mediators including miRNAs, transcription factors and signal transducers. Receptor protein kinases are important regulators of signaling cascades for growth and development, including cell specification in plants (De Smet et al., 2009). Supporting this, members of serine threonine protein kinase appeared to be differentially regulated by different miRNAs (miR858, oil palm-specific miR4 and miR12) (Figure 6.10), highlighting the importance of signal transduction in the control of inflorescence development.

Among the important flowering-related regulators that were predicted to be under miRNA regulation include transcripts coding for auxin response factor (targeted by miR160 and miR835), ethylene-responsive transcription factor (targeted by miR172), growth-regulating factor (targeted by miR396), MADS-box transcription factor (targeted by miR172) and MYB protein (targeted by miR163), some of which were differentially expressed in +6 and +15 female inflorescences (Figure 6.10). Also included as miRNA targets are those that have not previously directly associated with flower development, with many related to DNA replication and repair, highlighting the importance of tightly regulated DNA replication machinery for cell division and differentiation during inflorescence development. Interaction between DNA replication components influences the expression of key flowering genes (Sanchez Mde et al., 2012). Thus miRNA-regulated expression of DNA replication components may add a second layer of regulation for the expression of flowering genes.

Many of the predicted miRNA targets in our current study have also been reported in other plant species, some of which may represent highly conserved components in flowering-related pathways. In addition, several transcripts coding for unannotated proteins, e.g. CL1Contig6857 (targeted by miR835) and CL346Contig1 (targeted by miR858) showed inverse expression pattern with their regulating miRNAs, making them the candidate regulators of inflorescence development in oil palm.

CHAPTER 7

SMALL RNA-SEQ, RNA-SEQ, DEGRADOME-SEQ AND QRT-PCR COMPLEMENT EACH OTHER IN MIRNA STUDY

7.1 Introduction

Validation has been an important follow-up strategy to show consistency in the expression trends and/or fold changes determined from high-throughput sequencing and that from microarray (Kang et al., 2012; Liu et al., 2014b), northern blotting (Yang et al., 2011; Cao et al., 2013; Barciszewska-Pacak et al., 2015) or qRT-PCR (Wang et al., 2012b; Lee et al., 2015; Nie et al., 2015; Wang et al., 2015). Combined with small RNA sequencing,degradome sequencing has been used for large scale identification of small RNA targets in many plants (Liu et al., 2014b; Xing et al., 2014; Li et al., 2015b; Fang et al., 2016; Hu et al., 2016; Wang et al., 2016b; Song et al., 2017). Degradome sequencing allows global identification of the miRNA-directed cleavage products and prediction of cleavage site within miRNA targets (Addo-Quaye et al., 2008; German et al., 2008; Zhou et al., 2010).

In contrast to animal miRNAs, the high degree of complementarity between plant miRNAs and their targets has allowed prediction of targets with relatively high confidence (Jones-Rhoades et al., 2006). The prediction of miRNA targets is usually followed by experimental validation to confirm the miRNA-target interaction. In this study, RNA degradome sequencing (Degradome-Seq), a parallel analysis of RNA ends (PARE) that runs on high-throughput DNA sequencing platforms was used for high-throughput analyses of RNA decay by profiling the 5' end of uncapped RNA fragments in oil palm +6 and +15 female inflorescences. The aim was to globally identify the truncated transcripts resulted from endonucleolytic cleavage guided by miRNAs or siRNAs in the two female inflorescence tissues.

In this study qRT-PCR and degradome sequencing were used as the post-sequencing validation tools; qRT-PCR was used to validate the miRNA expression

profiles determined by the high-throughput sequencing, whereas degradome sequencing was used to confirm the miRNA targets in oil palm female inflorescences. It is important to note that validation using the same RNA samples assayed in the high-throughput sequencing experiments only validates the technology but not the conclusion about the treatments/conditions evaluated. It is the validation using different biological replicates from the same populations/genetic backgrounds that can support the biological conclusions derived from the high-throughput sequencing experiments (Allison et al., 2006).

7.2 Results

7.2.1 Degradome sequencing identified potential target(s) for both orthologous and oil palm-specific miRNAs

The output statistics for degradome sequencing of +6 and +15 female inflorescences were summarized in Table 7.1. The degradome short reads have been deposited in NCBI's Sequence Read Archive (SRA) database (accession number: SRR5189965 and SRR5189968). The cleavage products of the predicted targets for orthologous and oil palm-specific miRNAs were identified (results published in Ho et al., 2017). GO analysis showed that a majority of the cleaved miRNA targets was functionally related to the GO terms; biosynthetic process, metabolic process (biological process), binding, catalytic (molecular function), cell and organelle (cellular component) (Figure7.1). Based on the signature abundance at the target sites, these cleaved targets were classified according to Cleaveland4 version 4.3 (Addo-Quaye et al., 2008). Forty-nine of the predicted orthologous miRNA-target pairs (~5% of the total predicted miRNA-target pairs) were classified under Cleaveland category 0 and 1 (Addo-Quaye et al., 2008). Category 0 denotes > 1 raw read at a position with complementarity to the mature miRNA, and the abundance at that position equals to the maximum (where "maximum" means the maximum number of matches of any mature

miRNA on the transcript and only one maximum on the transcript). Category 1 denotes > 1 raw read at the position, with abundance at the position equal to the maximum on the transcript, and more than one maximum position on the transcript.

Table 7.1: Output statistics of degradome sequencing for +6 and +15 female inflorescences

Type	+6 female inflorescence		+15 female inflorescence	
	Unique sRNAs (%)	**Total sRNAs (%)**	**Unique sRNAs (%)**	**Total sRNAs (%)**
total	3,237,523 (100)	20,273,827 (100)	2,444,724 (100)	19,495,821 (100)
cDNA_antisense	262,677 (8.11)	1,606,448 (7.92)	174,046 (7.12)	1,405,702 (7.21)
cDNA_sense	1,730,738 (53.46)	15,177,125 (74.86)	1,232,244 (50.40)	14,058,948 (72.11)
rRNA	7,954 (0.25)	90,950 (0.45)	5,709 (0.23)	81,210 (0.42)
snRNA	3,141 (0.10)	13,009 (0.06)	1,713 (0.07)	7,430 (0.04)
snoRNA	2,621 (0.08)	15,278 (0.08)	1,647 (0.07)	12,088 (0.06)
tRNA	1,305 (0.04)	20,768 (0.10)	855 (0.04)	5,533 (0.03)
unannotated	1,229,087 (37.96)	3,350,249 (16.53)	1,028,510 (42.07)	3,924,910 (20.13)

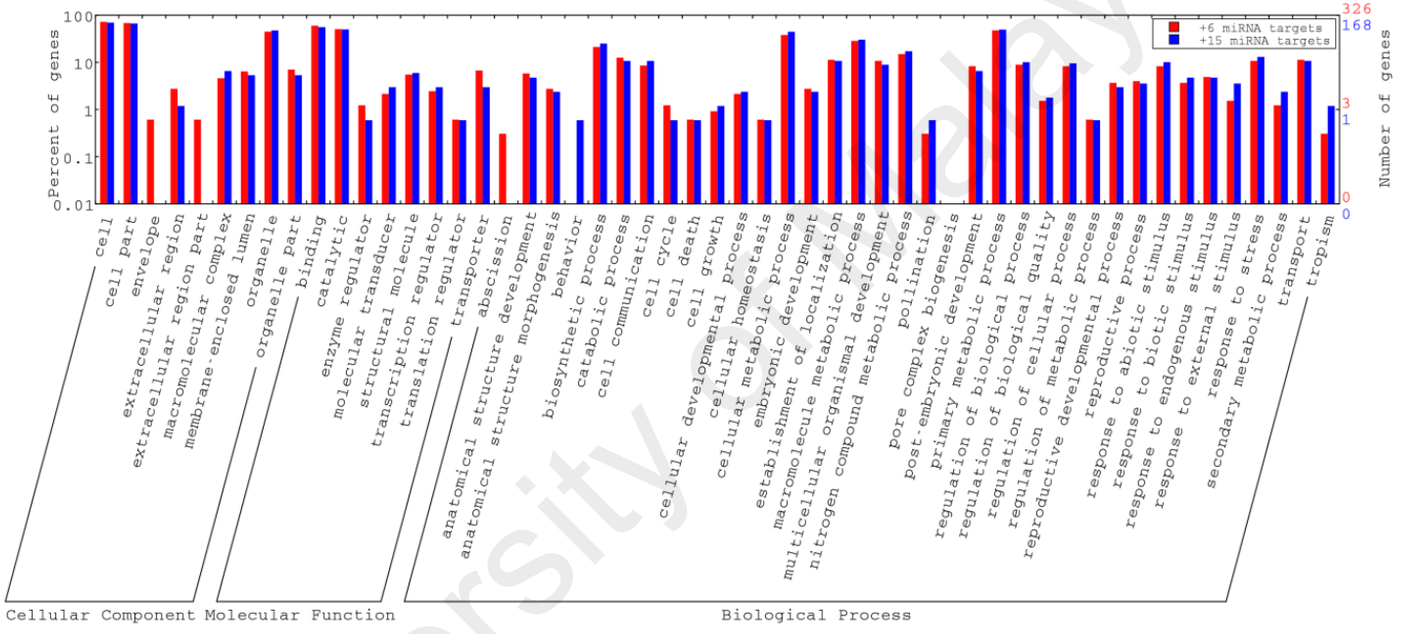


Figure 7.1: Gene ontology (GO) classification of the mRNA targets for orthologous miRNAs. The mRNA targets were predicted from degradomes of +6 and +15 female inflorescences. The mRNA targets were classified into biological process, molecular function and cellular component at the second level of GO classification.

7.2.2 RT-qPCR validates the sRNA-seq results

Using RT-qPCR, the expression levels of 16 randomly selected miRNAs, including 12 orthologous miRNAs and four oil palm-specific miRNAs were validated. Most of the RT-qPCR profiles (represented by green bars) were similar to the gene expression profiles determined by RNA-seq (represented by red bars); miR160, miR167, miR396, miR827 and miR1859 were more highly expressed in the +6 female inflorescences whereas miR166, miR168, miR172, miR528, miR535 and miR894 were more highly expressed in +15 female inflorescences (Figure 7.2). Oil palm-specific egu-miR10, egu-miR13 and egu-miR15, which had low read counts in the small RNA libraries, produced CT values ranging from 28 to 32 (i.e. low transcript levels). Among the miRNAs validated, egu-miR10, miR156, miR160, miR167, miR827 and miR1859 showed opposite expression profiles as measured by RT-qPCR and RNA-seq (Figure 7.2).

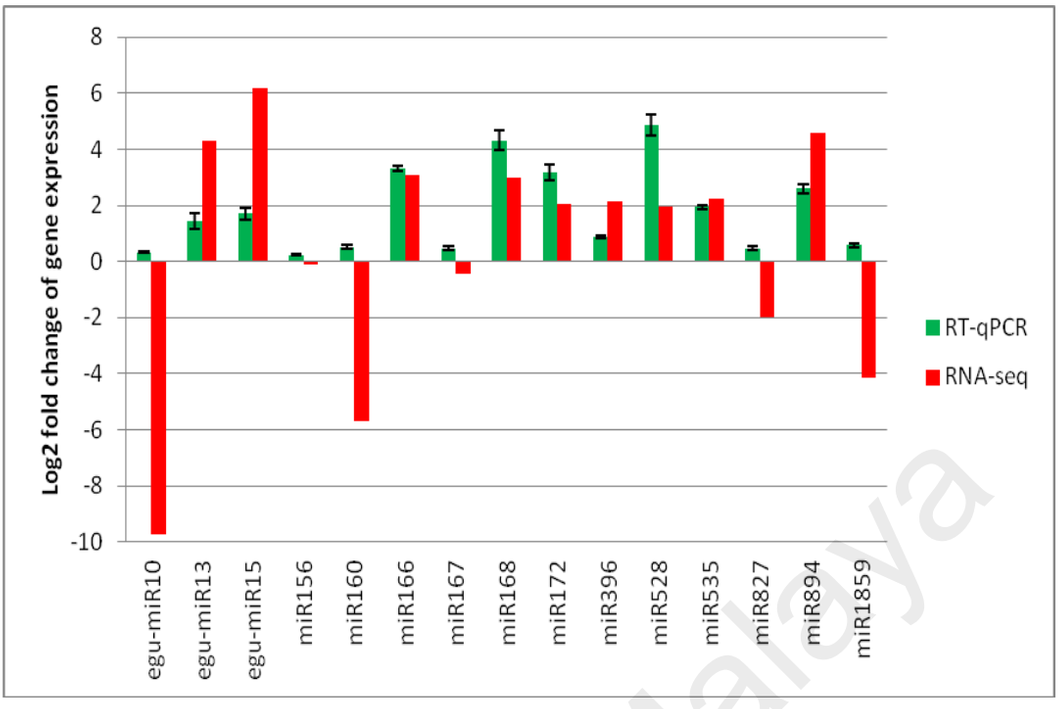


Figure 7.2: Validation and comparative expression of orthologous and oil palm-specific miRNAs in +6 and +15 female inflorescences by RNA-seq and RT-qPCR. The fold change of expression level was determined using: i) the ratio of normalized read counts (TPM_{+15}/ TPM_{+6}) (RNA-seq); for miRNAs that had zero count in either one of the sRNA libraries, their expression levels were arbitrarily set to 0.01; ii) the $2^{-\Delta\Delta CT}$ method (RT-qPCR); the expression level of miRNAs determined using RT-qPCR was normalized to the level of endogenous *Cyp 2* and was shown as mean ± standard deviation (S.D.) of three independent biological replicates.

7.3 Discussion

7.3.1 MiRNAs inversely correlated with the expression of predicted target gene(s)

The data from RNA degradome sequencing was consistent with the miRNAs regulating their predicted targets by cleaving their corresponding mRNAs and hence the inverse correlation between the expression levels for several of the miRNA-target pairs (Figure 6.10). When the expression level of a miRNA increases, the expression level of the mRNA target would be expected to decrease if cleavage of the mRNA target occurs. For instance, the expression level of miR163 was low in the +15 stage female inflorescence, while two of its predicted targets, the transcripts of a *MYB* (CL1Contig1790) and a flavine containing monooxygenase (CL1Contig4016) genes, showed an increase in expression in the same tissue as compared to the +6 stage female inflorescence. Mir168-mediated AGO1 homeostasis is crucial for meristem formation and organ polarity in Arabidopsis (Vaucheret et al., 2004; Kidner & Martienssen, 2005). Similar mechanism of regulation and relationship between AGO1 and miR168 have also been reported to affect growth rate, floral phase transition, leaf epinasty and fruit development in tomato (Xian et al., 2014). The current degradome data confirmed the miR168-guided cleavage of *ARGONAUTE1* (*AGO1*)-like transcripts (C182317 and CL1Contig3683) and hence the negative correlation between the expression level of miR168 and the *AGO1*-like transcripts (Figure 6.10), suggesting that similar mechanism was involved during the development of floral meristem and floral organs in oil palm female inflorescences. MiR535, the most abundant class of orthologous miRNA in oil palm female inflorescences (Table 6.4) and has been predicted to regulate the *SPL* (*SQUAMOSA* Promoter Binding Protein Like) gene family in conjunction with miR156/157 members in Arabidopsis (Carra et al., 2009). As *SPL* genes encode transcription factors important for plant growth and development (Chen et al., 2010), *SPL* genes may be dynamically regulated by multiple distinct miRNAs. Supporting this, the degradome data showed that a *SPL* transcript (scaffold8787) was

cleaved by three different miRNAs, i.e. miR156, miR529 and miR535, with each targeting distinct cleavage sites on the same transcript (Appendix C, Supplementary Table C5). These miRNAs could be synergistic, or acting sequentially, or simply having the same mutual targets.

Besides mRNA cleavage, miRNAs can also induce translational repression via imperfect complementarities with target sites in coding regions and 3' UTRs at variable efficacies (Hausser et al., 2013; Iwakawa & Tomari, 2013). In emerged and mature flowers of oil palm, the expression level of miR172 was inversely correlated with its *AP2*-like mRNA target in emerged and mature flowers of oil palm (Mehrpooyan et al., 2012). However in the present study, although the cleavage products of the predicted targets were detected, a clear inverse correlation between the expression level of miR172 and the targets (flower homeotic protein and MADS-box transcription factors) in the preceding stages, i.e. +6 and +15, was not found. Besides cleaving the mRNA targets, some miR172 paralogs have been proposed to act via translational inhibition in a tissue and developmental stage-dependent manner (Aukerman & Sakai, 2003; Chen, 2004). This may be the case for the differentially expressed miRNA such as miR172 that did not show an inverse expression pattern with its predicted targets (flower homeotic protein and MADS-box transcription factors) in the tissues studied. In *Physcomitrella patens*, it has been reported that miRNA-mediated cleavage of mRNA could be conditionally switched to miRNA-mediated DNA methylation, depending on the ratio of a miRNA to its target (Khraiwesh et al., 2010). In rapeseed (*Brassica napus* cv. Drakkar) and pumpkin (*Cucurbita maxima*), miR399 has been reported to be transported from the shoots to the roots via the phloem conduit, and to down-regulate the target transcript (*PHO2*) in the roots (Pant et al., 2008). Considering the potential of long distance travelling of miRNAs, it is possible that such mobile miRNAs which are expressed in the oil palm inflorescences but then transmitted systemically to distant tissues. It is also worth noting that some of the phloem proteins and RNA-binding

proteins identified in the transcriptomes produced in this study could be associated with the transport of mobile miRNAs.

RNA degradation is ubiquitous in all cells with many different pathways involved, collectively regulating RNA maturation and turnover in the cells (Houseley & Tollervey, 2009). Thus it can be expected that additional to the cleavage products of small RNAs, the degradome data produced in this study also contained other degraded RNA species. Also it is not surprising that the degradome profile of a cell/tissue may vary across the developmental life span, subjecting to the regulation of RNA degradation (including that of miRNA-mediated RNA cleavage) in response to endogenous and/or exogenous stimuli, further increasing the complexity of plant degradome.

7.3.2 RT-qPCR results were consistent with the miRNA expression trends determined by sRNA-seq

Despite being the most highly expressed oil palm-specific miRNA and was sequenced in all five tissues studied (Table 6.5), the amplification of egu-miR1 using RT-qPCR was not successful. The stem-loop RT primer used was designed such that it contains only six 3' nucleotides that are complementary to the target miRNA (Kramer, 2011). Changes in the stem-loop primer sequence may significantly change the melting temperature (T_m) and the binding specificity of the primers, thus restricting modification and optimization of the primer sequence for a higher sensitivity. Instead of the traditional oligonucleotide-based primers, an alternative option would be locked nucleic acids (LNA) oligonucleotides, which typically consist of a mixture of LNA™ and DNA or RNA. It has previously been reported that inclusion of LNAs into PCR primers improves both sensitivity and performance of RT-qPCR, by increasing template binding strength and specificity for DNA amplification (Ballantyne et al., 2008) even for those miRNAs of low abundance

(Jensen et al., 2011), which probably would be beneficial for optimal detection of low expressed transcripts such as the oil palm-specific miRNAs.

For egu-miR10, miR156, miR160, miR167, miR827 and miR1859, dicrepancies between the fold-change values measured by RT-qPCR and high-throughput sequencing of small RNA were observed. Low transcript levels can result in inefficient amplification, causing significant distortions in fold-changes of the transcripts (Bhargava et al., 2014), which may be the case for low and moderately expressed miRNAs such as egu-miR10 and miR160 (Figure 7.2). Such discrepancies between different methods of expression analysis may also be due to the fact that RNA-seq data in the current study was derived from a single pooled-sample measurement whilst the RT-qPCR data was derived from three independent biological replicates. Replication is important for a more precise and reliable summary statistics when assessing variation in measurements, with the effect of biological variability usually substantially greater than technical variability (Blainey et al., 2014). It has been recently established that at least six biological replicates should be used for RNA-Seq experiments and at least 12 for the identification of significantly differentially expressed genes of all fold changes (Schurch et al., 2016). The combination of the number of replicates and the type of differential expression tools also affects the true positive and false positive performances (Schurch et al., 2016). However, the exact number of biological replicates required to elicit a biologically significant conclusion may vary, depending on the objective, sample type, experimental design etc. Although it has been suggested that sequence polymorphisms between biological replicates can result in a gain or a loss in reads from different biological replicates depending on their sequence consistency with the reference genome sequence (Degner et al., 2009), the effect may not be substantial miRNAs, for which the sequences from the same family are mostly conserved across different plant species.

Despite the many advantages of high-throughput sequencing, it is yet to completely replace current RT-qPCR methods, but rather as a complementary technique depending on the needs of the experiment and the resources available. The results of the high-throughput sequencing will provide preliminary identification of those interesting candidates that need further evaluation using RT-PCR methods. When cost is a limiting factor, an alternative experimental design would be to use fewer biological replicates for high-throughput sequencing and to compensate with a higher number of biological replicates for RT-qPCR validation, at the expense of labor, time and the number of transcript to study in parallel. In the current study that compared measurements of a high-throughput sequencing experiment (for a single pooled-sample) and three RT-qPCR experiments (for three independent biological replicates), nine miRNAs displayed consistent expression trend (i.e. up-/down-regulation) in both RNA-seq and RT-qPCR analyses (Figure 7.2), again highlighting the importance for post-sequencing validation step.

CHAPTER 8

GENERAL DISCUSSION

Oil palm is monoecious, producing alternate cycles of female and male inflorescences throughout the year, with occasional hermaphrodite inflorescences during the transition between female and male batches (Biradar, 1978). The development of an inflorescence occurs over approximately 32-36 months via a series of individual differentiation phases (Corley, 1976a; Corley & Gray, 1976) and each developing oil palm inflorescence can be related to its respective subtending leaves that are referred to as +number (or –number) of the emerged leaf fronds of a palm (Adam et al., 2005).

Notwithstanding the detailed characterization of morphological changes throughout the reproductive development of oil palm (Adam et al., 2005), the underlying genetic factors during the slow development of oil palm inflorescence are still little understood. In model plants such as Arabidopsis and rice, the regulatory role of miRNAs in flower development has been comprehensively investigated (Luo et al., 2013). As genetic variations across plant lineages may result in species-specific functional diversification of a gene, it cannot be assumed that the regulatory function of these miRNAs will be identical for oil palm. For example, *EgNAC1* was most highly expressed in both flower and fruit samples of oil palm, as opposed to its homolog *ATAF2*, which was highly expressed in root and leaves and had lowest expression in flower buds of Arabidopsis (Shearman et al., 2013), probably reflecting functional divergence between plant species.

In response to natural selection, floral sex allocation in monoecious plants has evolved to enable alteration of the sex ratio (Willson, 1979). For commercial cultivation of oil palm, optimum fruit yield requires a higher ratio of female inflorescences in the production palms, whilst for hybrid breeding, a different floral sex ratio may be required, for example in the male parent line (Durand-Gasselin et al.,

1999). Oil palm floral sex ratio has been reported to be regulated by environmental (Corley, 1976a), anthropogenic (Corley, 1976b; Durand-Gasselin et al., 1999) and very likely genetic factors, either independently or in combination. In oil palm, sex determination of the inflorescences usually initiates in the developing tissues adjacent to the frond stage −20 (Durand-Gasselin et al., 1999) which at that point, lies deep within the palm stem. The sampling procedure is destructive to both the developing inflorescence tissues and the palm, making it impossible to know the final sex of the tissue once it is sampled. Moreover, sexual identity of an inflorescence might not be stably determined at initiation of primordium (Kater et al., 2001). Therefore, in order to determine the differences between female and male inflorescences, the +6 stage (approximately 10 months after initiation of inflorescence meristem) was selected, when the morphology of female and male inflorescences can be reliably distinguished (Adam et al., 2005). The male inflorescence at leaf stage +6 displayed a series of long, finger-like and cylindrical spikelets spirally arranged around the rachis, whereas the female inflorescence displayed shorter spiny spikelets (Appendix B, Supplementary Figure B2). No hermaphrodite inflorescence was observed among the inflorescence materials sampled in this study possibly due to the age of the palms as the frequency of hermaphrodite inflorescences has been reported to decrease with the age of the palms (Williams & Thomas, 1970). The +6 stage is also within the period where abortion of entire inflorescences often occurs, which will alter the sex ratio of inflorescences of a palm (Durand-Gasselin et al., 1999).

The primary aim of the current study was to investigate the gene expression profiles at the sRNA transcriptome (sRNAome) level, in both female and male inflorescences of oil palm. Given that only the fertilized female inflorescences develop into fruit bunches which are desired for commercial oil and seed production, the study was further extended to specifically investigate the expression patterns of miRNAs and their mRNA target(s) during two key developmental stages of female inflorescence, i.e.

+6 and +15 stages. The +6 and +15 stages fall within the period during which floral organs are forming. It was also the interest of this study to investigate the gene expression differences between female and male inflorescences at an early stage of development, i.e., leaf stage +6, as this is the earliest stage at which the sex of an inflorescence can be distinguished by morphology (Adam et al., 2005). The +6 stage is also within the period where abortion of whole inflorescences often occurs, which will alter the sex ratio of inflorescences of a palm (Durand-Gasselin et al., 1999).

8.1 Different orthologous miRNAs were expressed in female and male inflorescences of oil palm, regulating the expression of flowering-related genes

At the sRNAome level, female and male inflorescences displayed distinct expression patterns of miRNAs. In +6 and +15 male inflorescences, miR166 was the most abundant class of orthologous miRNA. In +6 and +15 female inflorescences, miR535 was the most abundant class of orthologous miRNA. In the out-group, i.e., the leaf, miR156 was the most highly expressed orthologous miRNA (Table 6.4). Such differences were not unexpected considering the distinct differences in the morphology of these tissues. This is however not always the case; despite of the distinct morphological differences, miR166 was the most abundant class of miRNA in both male (Wang et al., 2015) and female (Wang et al., 2012b) flowers of hickory (*Carya cathayensis* Sarg.). Similar high expression of miR166 was also reported for both female and male flowers of andromonoecious *Populus tomentosa* (Song et al., 2013).

In this study, an exceptionally high expression level of miR166 was found in the oil palm male inflorescences. A study by Grigg et al. (2005) has shown that mutations in the *SERRATE* (*SE*) gene encoding a zinc-finger protein result in an over-accumulation of miR165/166 precursor and simultaneously reducing the accumulation of mature miR165/166, indicating that *SE* is required for post-transcriptional regulation

of *MIR165* and *MIR166* genes (Grigg et al., 2005). In situ hybridizations have demonstrated that miR165/166 accumulates in a tissue-specific manner, suggesting that these miRNAs are not mobile or at least with limited mobility within a given tissue (Kidner & Martienssen, 2004). Li et al. (2017) has shown that the members of miR166 family were differentially expressed among different tissue types in soybean. The high expression level of miR166 in male inflorescences suggests a corresponding down-regulation of its targets, the *HOMEODOMAIN-LEUCINE ZIPPER* (*HD-ZIP*) family genes. Previous genetic analysis in Arabidopsis has revealed that *HYPONASTIC LEAVES1* (*HYL1*) modulates inner microsporangia and stamen architecture by repressing the expression of *HD-ZIP III* genes and enhancing the expression of *FILAMENTOUS FLOWER* (*FIL*) gene through miR165/166 (Lian et al., 2013), reflecting the complexity of a dynamically regulated mechanism for stamen development. Besides the role in stamen, members of HD-ZIP III transcription factor family redundantly specify upper (adaxial) cell fate in the leaf of Arabidopsis (Emery et al., 2003) and maize (Juarez et al., 2004). While it has been reported in maize that miR166 predominantly accumulates immediately below (abaxial) the incipient leaf (Juarez et al., 2004), a conclusive evidence of miR166 regulating the adaxial-specific expression of *HD-ZIP III* is still lacking. As the current study was limited such that it did not include transcriptome data deriving from male inflorescences, the possible negative regulation of *HD-ZIP III* by miR166 cannot be confirmed. It will also be useful to further elucidate whether spatial expression and interaction between miR166 and *HD-ZIP III* regulate the symmetry of the oil palm inflorescences.

The high expression levels of miR535 in the female inflorescences as compared to the male inflorescences and the leaf contradict with the previous reports that have identified miR535 as leaf- and root-biased in apple (Xia et al., 2012), and as leaf-biased in mangrove (Wen et al., 2016). On the other hand, An et al. (2011) has described miR535 as a miRNA induced by low temperatures, in contrast to the high expression

levels of miR535 in female inflorescences of oil palm which grows in warm climates. While the thermal regulation of miR535 gene is still unclear, it is important to understand the regulation and function of miR535 in oil palm female inflorescence which is highly sensitive to heat and/or drought stresses.

Based on the high sequence similarity to miR156/157, miR535 has been predicted to regulate the *SPL* (*SQUAMOSA* Promoter Binding Protein Like) gene family in conjunction with miR156/157 members in Arabidopsis (Carra et al., 2009). In support of this prediction, the current degradome data showed that a *SPL* transcript (scaffold8787) was cleaved by three different miRNAs, i.e., miR156, miR529 and miR535, with each targeting different cleavage sites on the same transcript (Appendix C, Supplementary Table C5). Besides inducing the expression of flowering-promoting genes (Wang et al., 2009a; Bratzel & Turck, 2015), *SPL* genes have been reported to regulate the development of female reproductive tract in Arabidopsis (Xing et al., 2013), floral organ size and ovule production in cotton (Liu et al., 2017). While the combination of sRNA-seq, RNA-seq and degradome data confirms miRNA-mediated regulation of *SPL* transcript, it is not known whether miR156, miR529 and miR535 are acting synergistically, sequentially or simply co-regulating *SPL* genes during the development of female reproductive organs. Besides *SPL* genes, miR535 has also been predicted to regulate cysteine protease in apple (Xia et al., 2012), calcium-transporting ATPase 1, mitogen-activated protein kinase, pentatricopeptide repeat-containing protein and a putative S-locus-specific glycoprotein in mulberry (Jia et al., 2014), cytochrome P450 in citrus (Ma et al., 2016) and guanosine monophosphate synthetase in mangrove (Wen et al., 2016). It should be noted that the target(s) of a miRNA may vary across plant taxa, developmental stages and environmental conditions.

A recent study by Somyong et al. (2016) has identified *EGmiR159a* and *EgACCO1* (encoding aminocyclopropane carboxylate (ACC) oxidase), both of which were most highly expressed in ovaries of female flowers but were not expressed in the

male flowers, to be the most linked QTL genes for female flower determination and, to a lesser extent, fruit development of oil palm. In contrast to the previous study (Somyong et al., 2016) which was limited to few mature floral organs, the present study showed that a high expression of miR159 was also found in the less mature tissues, in this case the +6 and +15 female and male inflorescences. The predicted targets of miR159, i.e., the transcripts coding for ACC oxidase (CL223Contig2 and CL1Contig85) were more highly expressed in the +15 female inflorescence compared to the +6 female inflorescence (Appendix C, Supplementary Table C6), supporting its female organ-determining role prior to the initiation of female reproductive organs in the pistillate flowers. The expression levels of these transcripts, however, cannot be determined for male inflorescences due to the lack of corresponding transcriptome data. Taken together, we speculate that the roles of miR159 and *EgACCO1* may not be limited to the ovary and fruit, and thus have a role in non-reproductive parts of the flowers such as the developing sepals and petals in both female and male inflorescence.

In comparison to female and male inflorescences, an exceptionally high expression of miR156 was evident in oil palm leaf. A study in rice (*Oryza sativa*) has found that miR156 level gradually increased from young leaf to old leaf after the juvenile stage (Xie et al., 2012). Despite the increasing miR156 level during leaf growth, the miR156-targeted rice SQUAMOSA-promoter binding-like (SPL) transcription factors were predominantly expressed in young leaves and did not significant change over the stages of leaf development. It is however not known whether similar changes in the expression level of miR156 also occurs during leaf growth in oil palm, as the leaf tissues used in this study were all collected from the 17th frond of each sampled palm. In Arabidopsis, miR156 regulates the expression of the juvenile phase, and acts sequentially with miR172 to regulate the timing of the juvenile-to-adult transition (Wu et al., 2009b). MiR172 acts downstream of miR156 to regulate the transitions between developmental stages and in specifying floral organ

identity, by targeting *APETALA2*-like (*AP2*) transcription factors. A previous genetic screening of Arabidopsis mutants defective in several floral regulators has highlighted the need for *AP2* genes in stamen emergence of Arabidopsis (Nag et al., 2007). An up-regulation of miR172 coupled with a down-regulation of several *AP2* transcripts (CL1Contig6918, CL746Contig1, Contig1254, Contig1293, scaffold11436 and scaffold8532) were observed in the +15 female inflorescence, possibly reflecting a mechanism to suppress stamen development hence the formation of rudimentary staminate flowers in oil palm female inflorescences. Besides mRNA cleavage, miR172 also regulates *AP2* via translational inhibition (Chen, 2004). Further studies of miR172-*AP2* interactions via mRNA cleavage and/or translational inhibition, as well as identifying the true target transcript(s) for miR172 are necessary for a better understanding of the functionality of the male reproductive organs in oil palm inflorescences.

In addition to the known flowering regulators that were predicted to be under miRNA regulation, several novel miRNA targets including the transcripts coding for cyclin dependent kinase inhibitor (targeted by egu-miR6), protein kinase (targeted by egu-miR10) and proteosome subunit (targeted by egu-miR10), all of which showed an inverse expression pattern between the regulating miRNA and the predicted target and were seen to have a matching cleavage products in the degradome data (Figure 6.10) (Appendix C, Supplementary Table C7). Among the predicted miRNA targets, many are involved in DNA replication and repair, highlighting the importance of tightly-regulated DNA replication machinery for cell division and differentiation during inflorescence development. While the interaction between DNA replication components influences the expression of key flowering genes (Sanchez Mde et al., 2012), miRNA-regulated expression of DNA replication components may add a second layer of regulation for the expression of flowering genes.

Most of the predicted miRNA targets in our current study have also been reported in other plant species, some of which may represent highly conserved components in flowering-related pathways. In addition, the current study identified several transcripts coding for unannotated proteins, e.g. CL1Contig6857 (targeted by miR835) and CL346Contig1 (targeted by miR858), which showed inverse expression pattern with their regulating miRNAs, making them the candidate regulators of inflorescence development in oil palm. Additional studies will be required to elucidate the roles of these unannotated proteins in the developing inflorescences.

Previous finding have shown that evolutionarily conserved miRNAs exert more functions *in vivo* (Axtell & Bartel, 2005). In the current study, orthologous miRNAs such as miR163 and miR172 have multiple predicted targets, each with distinct biological functions (Figure 6.10). On the other hand, oil palm-specific miRNAs such as miR5 and miR13, each has over 10 predicted targets. The *in vivo* functions of these oil palm-specific miRNAs would, however, require further investigation before their roles as miRNA can be concluded.

8.2 Different subsets of orthologous miRNAs were differentially expressed at +6 and +15 stages of inflorescence development

Variation in miRNA expression profiles across different developmental stages has been reported in developing flowers for a number of plant species (Wang et al., 2012b; Chen et al., 2013; Liu et al., 2014b; Wang et al., 2015) (Figure 8.2). In oil palm, a large number of miRNAs were differentially expressed in female and male inflorescences, reflecting a high degree of miRNA-mediated regulation of gene expression in these tissues. In female inflorescences, differential expression of distinct subsets of miRNAs (Figure 6.10 and Table 6.4) and their mRNA targets (Figure 6.10) were observed at the two developmental stages studied, implying their role in the regulation of developmental progression during the formation of floral meristems at the

+6 stage and development of floral triads at the +15 stage (summarized in Figure 8.1). Among the differentially expressed miRNAs, the upregulation of miR172 in the two developing female inflorescences studied appeared to continue into the subsequent phases, i.e. the emerged and mature female flowers of oil palm (Mehrpooyan et al., 2012), reflecting an active requirement for miR172 as early as +6 stage until the maturity of the female flowers. However it is not known exactly when the upregulation of miR172 begins in oil palm inflorescence. In maize, there has been evidence that miR172 controls sex determination and meristem cell fate by targeting an *AP2* gene (Chuck et al., 2007). In the case of oil palm, sex determination of the inflorescences usually initiates in the developing tissues adjacent to the frond stage −20 (Durand-Gasselin et al., 1999) which at that point, lies deep within the palm stem. While a recent study has reported the gene expression differences between female and male inflorescences (Ho et al., 2015), the detailed mechanism of sex determination of an inflorescence is still unclear. In order to determine if oil palm miR172 (or any other miRNAs) plays sex-determining role(s), it requires a temporal expression study beginning from the initiation of inflorescence primordia until the fully matured inflorescences are formed, a study that would be difficult and highly costly due to the long maturation phase of oil palm inflorescences and the destructive nature of sampling required.

Among the differentially expressed miRNAs, several flowering-related miRNAs showed expression patterns that are similar to those reported in model plants (Huijser & Schmid, 2011), albeit the magnitude of change in miRNA expression during reproductive development may vary from plant to plant. Taking miR160 and miR172 as examples which showed opposite expression patterns in developing oil palm female inflorescence, miR160 was highly expressed at the +6 stage whereas miR172 was highly expressed at the +15 stage. Despite that sequentially operating miRNAs, i.e. miR156 and miR172, have been reported to regulate juvenile-to-adult transition in

Arabidopsis (Wu et al., 2009b), validation using RT-qPCR did not show significant difference in the expression levels of miR156 in +6 and +15 female inflorescences of oil palm (Figure 7.2). It is speculated that downregulation of miR156 might have occurred at a much earlier stage, probably during the switch from vegetative meristem to inflorescence meristem (Figure 8.1). Nonetheless, the expression of distinct subsets of miRNAs together with their mRNA targets during the two different inflorescence stages, supports their role in the regulation of developmental progression during the formation of floral meristems at the +6 stage and development of floral triads at the +15 stage of oil palm (summarized in Figure 8.1).

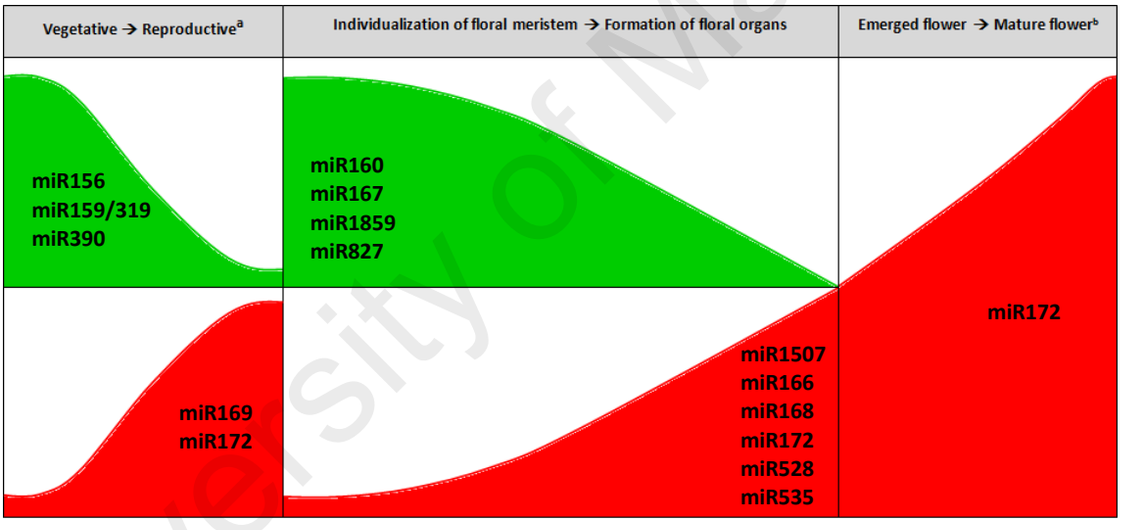


Figure 8.1: Model of miRNA expression during inflorescence development in oil palm. [a] miR156, miR159/319, miR169, miR172 and miR390 regulate juvenile-adult transition and vegetative-reproduction transition of plant development (Spanudakis & Jackson, 2014; Hong & Jackson, 2015). [b]The expression level of miR172 increases as oil palm matures, with the highest expression in the mature flowers (Mehrpooyan et al., 2012). Green color indicates miRNAs upregulated in vegetative/immature tissues and red color indicates miRNAs upregulated in reproductive/mature tissues.

8.3 Flowering-related miRNAs showed unique expression patterns in oil palm female inflorescences

Consistent with a previous report that the expression miR172 was higher in the mature female flowers compared to flowers of earlier stages (Mehrpooyan et al., 2012), the expression of miR172 was higher in the +15 female inflorescence compared to the +6 female inflorescence (Figure 6.10). Whilst flowering regulators such as miR172 and miR396 of oil palm showed expression profiles similar to those reported during floral development in other plant species (Wang et al., 2012b; Wang et al., 2015), miRNAs such as miR163 and miR535 appeared to be differentially expressed in developing oil palm female inflorescence but not in plants such as hickory, rice and maize (Figure 8.2). Such difference would be expected among genetically diverged plants, for which miRNA expression may be further complicated by different growing environments. Furthermore, the distinct miRNA expression profiles observed in different plant species may represent regulatory mechanisms governing floral sexual identity that are unique to a specific floral developmental stage of each plant. The biogenesis and regulatory networks of miRNAs involve complex processes regulated by multiple genetic factors (Rogers & Chen, 2013), which themselves are regulated via mechanisms that are still largely unclear. Thus any changes in the expression levels of these regulatory factors, for example, in response to environmental changes such as water or mineral deprivation, would be expected to affect the expression of miRNAs and as well as their downstream activity.

In perennial plants such as hickory and oil palm, the long flower developmental phase presents a long window of exposure to various environmental influences. In these long-cycle plants, the differential expression of miRNAs occurs over a much larger temporal scale than in shorter-cycle plants. In oil palm, the maturation of inflorescences take 32–36 months (Corley, 1976a), the period during which the developing inflorescences are exposed to environmental influences over several seasons. In

contrast, for plants with short reproductive cycles such as cucurbits, the entire flower development, starting from time of primordial initiation to anthesis, only takes about 20 days (Bai et al., 2004), giving a much shorter term of exposure to environmental cues as the flowers differentiate. Also, it is not yet known if, in long-cycle plants, the entire reproductive development is regulated by recurrent waves of miRNAs over time. Therefore, the detailed patterns of miRNA expression over developmental stages in long-cycle plants would require additional sampling points spanning the entire reproductive cycle. In addition to previously mentioned inter-specific genetic variation among different plants and the effect of environmental stimuli, it should be noted that the differential expression profiles of miRNAs may also vary among tissue types, stages of development, types of gene expression analysis platforms and the use of different cut-off thresholds for differential expression.

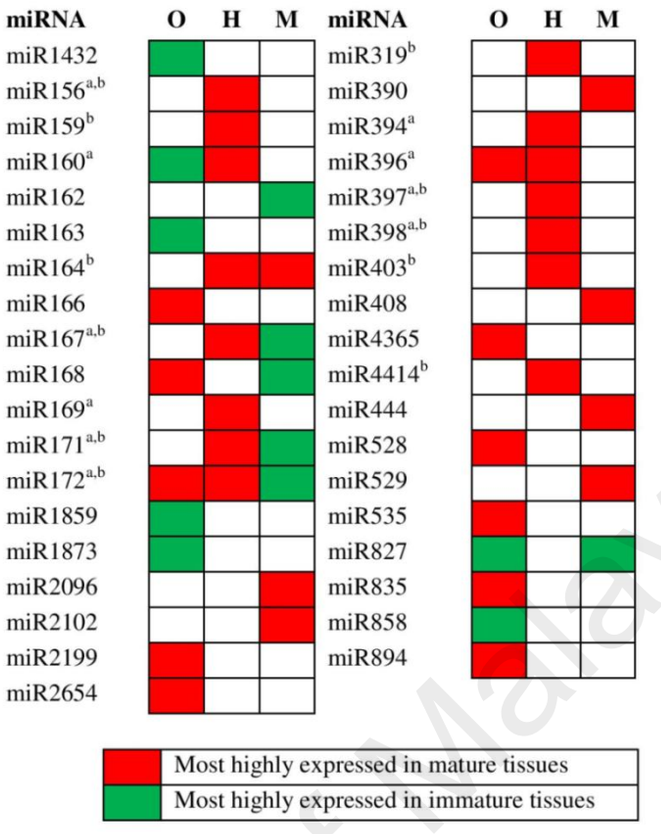


Figure 8.2: ***Differentially expressed miRNAs identified in inflorescence tissues of African oil palm, hickory and maize.*** These putative flowering-related miRNAs are the statistically significant differentially-expressed miRNAs reported in the following studies: O, African oil palm (*Elaeis guineensis* Jacq.); female inflorescences at +6 and +15 stages; sRNA-seq; Log_2 fold change ≥ 1 or ≤ -1 and FDR cut-off < 0.05; [current study]; H, Hickory (*Carya cathayensis* Sarg.); flower buds at the vegetative and reproductive stages; sRNA-seq; fold change ≥2 or ≤0.5 and *P*-value < 0.05 (Wang et al., 2012b; Wang et al., 2015)[a,b]; M, Maize (*Zea mays*); maize ears at the growth point elongation phase, spikelet differentiation phase, floret primordium differentiation phase and floret organ differentiation phase; miRNA microarray; one-way ANOVA and Fisher's LSD, *P*-value <0.01, n=4 (Liu et al., 2014b)

8.4 Expression of inflorescence- and sex-specific transcripts coincides with the early phase of sex differentiation in oil palm inflorescences

At the transcriptome level, distinct expression patterns were observed for a set of sex-specific transcripts in +6 female and male inflorescences (Figure 5.4 and Table 5.3). At +6 stage, the bracteoles are formed from the floral meristems (Adam et al., 2005). The genes expressed at this stage are likely to play a role in the upcoming development of floral organ primordia into perianth organs, followed by the gynoecium or androeciumin female and male flowers, respectively.

In plants, it can be expected that the differences in the environmental exposure during the reproductive cycles may result in variation in the expression patterns and timing of flower-related miRNAs and/or genes. Moreover, different gene expression patterns for sex-related genes in oil palm could also result from species-specific genetic regulatory mechanisms such as epigenetic control (Martin et al., 2009) and post-transcriptional regulation by microRNA (Chuck et al., 2007; Aryal et al., 2014) either independently or through interaction with environmental factors. Future studies on epigenetic- and microRNA-mediated regulation of gene expression may help to elucidate novel regulatory networks interacting with environmental stimuli during floral development.

The current analysis revealed that multiple genes were differentially expressed, with each regulating specific processes that are likely to contribute to the sex-specific features of an inflorescence. Previous transcriptome-based studies for the dioecious plants, *Populus tomentosa* (Song et al., 2013) and *Salix suchowensis* (Liu et al., 2013), and the trioecious plant, *Carica papaya* (Urasaki et al., 2012) have shown gene expression changes between sexes, with most of these relating to genes within a sex-related chromosome or loci. On the other hand, no clear common gene-expression patterns have been reported for monoecious plants such as *Cucumis sativus* (Guo et al., 2010) and *Quercus suber* (Rocheta et al., 2014). For the current study, there is no evidence for a sex chromosome or sex-related loci, as the known floral genes and sex-specific transcripts described here mapped to different locations on separate chromosomes of oil palm (Appendix B, Supplementary Figure B3), suggesting unique sex differentiation pathways that may involve activation and/or suppression of genes randomly distributed across the oil palm genome. It is also likely that a combined effect of multiple genes is required for the initiation and development of floral organs in female and male inflorescences. It should be noted that the sex-specific transcript candidates reported in this study were based on a stringent criterion of at least 10

counts in one tissue and zero in the other tissue, with an aim to identify transcripts that were strongly predominant in either male or female inflorescences. As a result, this may have excluded some transcripts with marginal differential expression patterns and yet contributing to the sex expression of an inflorescence. Nevertheless, these criteria enabled the identification of several transcripts with clear inflorescence- and sex-specific expression patterns, which were likely to play a role in the synchronized coordination of different programs of genetic control that drives sex differentiation in oil palm inflorescences.

8.5 Understanding gene regulation during inflorescence development of oil palm: Challenges and promises

Oil palm inflorescences, especially the immature ones lie deep inside the crown of the palm, thus the sampling procedure for these developing materials is destructive to both the tissues of interest as well as the palm. Due to the fact that it is costly and technically difficult to assay samples from single palms, pooled inflorescence tissues were used in this study. Nevertheless, through the selection of palms of the same age and genetic background, grown in the same commercial estate and sampled at the same time, the current study design has minimized the variations due to circadian rhythm effects, diurnal changes in the environment and genetic background. Hence, the expression data reported here should reflect the central tendency and provide a good representation of the oil palm population used. It is noted that the variation which could have arisen from the genetic background and environmental effects cannot be determined and hence more robust conclusions on gene function and specificity would require further studies involving a larger number of palms from different genetic backgrounds grown under different environments. While the transcriptome study was limited to oil palm female inflorescences at two developmental time points, i.e., +6 and +15 stages, this provides a basis for future studies that may be extended to include

female inflorescences at different stages of development, or other types of oil palm tissues (such as male inflorescence, leaf and root), as well as to functionally elucidate the regulatory role(s) of the novel miRNAs and their mRNA targets.

Many of the sequences reported in this study are still functionally unknown, indicating an urgent need to extract functional information from the unannotated regions of the oil palm reference genome. Among the sequences that could not be annotated, some of which might represent contaminants of unknown origin or it could be simply due to the incomplete oil palm genome assembly and annotation. The unmapped and unassembled contigs in the current version of oil palm genome (the fifth genome build or P5-build) have a high repeat content of approximately half of the 1.53-Gb P5-build, or 57% of the 1.8-Gb *E. guineensis* genome (Singh et al., 2013). In this study, a huge number of siRNA candidates were predicted, of which many can be mapped to the loci of repeats and transposable elements (Appendix B, Supplementary Figure B4). Hence it is speculated that some of these predicted repeat-associated siRNAs might match the unmapped and unassembled contigs, with the exact genetic locus in the genome still unknown. More recently, the trend in NGS has been moving towards the use of longer reads that span longer distances/range, with the aim to more accurately reconstruct highly repetitive, heterozygous and even polyploidy genomes. While the use of longer reads may be useful to complement the current oil palm genome model, Quail et al. (2012) has shown that longer reads did not increase mappability of the reads significantly, although a beneficial effect was obtained from having mate-pair information. Nevertheless, the unannotated or uncharacterized proteins and unigenes predicted to be regulated by miRNAs in female inflorescences (Figure 6.10) are potential floral regulators. Functional verifications, such as via overexpression and/or silencing, are required to elucidate their roles in oil palm.

Following the generalization of the use of next generation sequencing technologies (HTS) in various aspects of oil palm research (Bourgis et al., 2011;

Tranbarger et al., 2011; Uthaipaisanwong et al., 2012; Dussert et al., 2013; Shearman et al., 2013; Singh et al., 2013; Lei et al., 2014; Ho et al., 2015; Ho et al., 2016), one of the major bottlenecks has shifted from HTS data generation to rapid and accurate analysis and interpretation of an overwhelming amount of output data. The demand for extensive computational analyses of high-throughput sequencing data has rapidly increased, resulting in the development of various analysis algorithms and pipelines, with each being optimized to suit different types of sequence data (Conesa et al., 2016; Escalona et al., 2016; Hintzsche et al., 2016). This, in turn, has created an unprecedentedly high demand for bioinformatics expertise for HTS data analysis. For biologists who have no prior experience in HTS, it would be challenging to embark on the analysis of the complex data without support and collaboration from researchers who have been working on HTS projects, software scientists, as well as bioinformaticians.

For important plant/crop species like Arabidopsis (Schmitz et al., 2013; Berardiniet al. 2015), maize (Andorf et al., 2015; Zhu et al., 2015), sorghum (Makita et al., 2015; Tian et al., 2016) and rice (Kawahara et al., 2013; Narsai et al., 2013), dedicated databases that contains extensive epigenetic/genomic/proteomic/transcriptomic information have been established. In contrast, oil palm sequence data is mostly dispersed across different public and private databases, thus limiting accessibility. The initiatives to establish a centralized and systematically curated repository that houses oil palm-derived sequences for sharing and dissemination has been very limited, such as the Oil Palm Database hosted by the Genome Institute, Thailand National Center for Genetic Engineering and Biotechnology (BIOTEC) (http://www4a.biotec.or.th/oilpalmdb/index.php). Nevertheless, the establishment of such repository is of great importance, especially in the interest of efficient and broad dissemination of data resources to the community within this field.

The current study has shown that distinct subsets of miRNA families were differentially expressed in the female and male inflorescences, prompting an interest in identifying key miRNAs that regulate the formation of pistillate (in female inflorescence) and functional staminate (in male inflorescence) flowers. Overexpressing or silencing these key miRNAs may potentially alter the process of inflorescence development in oil palm and the ultimate fruit formation.

CHAPTER 9

CONCLUSION

This study aimed to provide new information on the genetic basis for floral development in African oil palm, an important economic crop in Malaysia and other tropical countries. There were three primary aims of this study and the new information that can be concluded for each aim was as follows:

9.1 +6 female and male inflorescences showed distinct expression patterns of inflorescence- and sex-specific transcripts

This study has identified 2 inflorescence- and 97 sex-specific transcripts for which the expression coincides with the early phase of sex differentiation in oil palm inflorescences. Notable among these, acid phosphatase and *DEFICIENS*, showed male-specific expression patterns whereas orthologs of bZIP transcription factor, late embryogenesis abundant protein and *TASSELSEED1* were female-specific. Orthologous transcripts of acid phosphatase and late embryogenesis abundant protein were also strongly inflorescence-specific. The results presented in this study indicate that differential expression of different subsets of genes, each regulating specific processes, are likely to contribute to the sex-specific features of an inflorescence. However, many of the transcripts reported in this study are still functionally unknown, indicating an urgent need to extract functional information from the unannotated regions of the oil palm reference genome. Further studies using more individuals from different genetic background and growing environment are needed so that the potential use of these transcripts as tissue- and/or sex-specific markers can be assessed. Future studies on the regulation of these candidate genes may provide means of controlling or selecting specific sexual expression patterns in oil palm.

9.2 Gene expression profiles at the sRNAome level showed variation in the expression pattern of miRNAs in developing female and male inflorescences

This study revealed that different populations of miRNAs were expressed in oil palm female and male inflorescences, with these miRNAs predicted to regulate the expression of different target genes. Among the orthologous miRNAs, miR535 was the most highly expressed miRNA in the female inflorescence whereas miR166 was the most highly expressed miRNA in the male inflorescence. A total of 29 'novel' oil palm-specific miRNA candidates were also reported and will be worthy of future study as the current degradome data confirms their active cleavage of mRNAs involved in various processes during the growth and development of female inflorescence. Besides miRNAs, the presence of a huge number of 24 nt siRNAs in all the oil palm tissues studied highlights their importance, most likely reflecting a high level of RNA-mediated epigenetic processes to maintain genome stability. These siRNA sequences are interesting as candidates for understanding their functions as epigenetic and post-transcriptional gene silencing effectors.

9.3 MiRNAs and their targets were differentially expressed in +6 and +15 female inflorescences

Distinct subsets of miRNAs and their target(s) were differentially expressed in +6 and +15 female inflorescences, reflecting a high degree of miRNA-mediated regulation of gene expression during these two key developmental stages. Some of these miRNAs were highly expressed at the +6 stage while some were highly expressed at the +15 stage, implying different roles in the regulation of developmental progression during the formation of floral meristems at the +6 stage and development of floral triads at the +15 stage of female inflorescence. RNA degradome showed that miRNAs negatively regulate their mRNA targets by cleavage and hence the inverse

expression patterns between the regulating miRNAs and the predicted targets. The degradome data also supports that a single miRNA may simultaneously regulate multiple targets and an mRNA may be regulated by multiple miRNAs, reflecting a complex and fine-tuned interaction network between miRNAs and their targets at the post-transcriptional level. It is noted that the use of degradome sequencing in this study is however limited to the identification of miRNA-cleaved targets but not those that are translationally inhibited. The detailed mechanism by which oil palm miRNAs regulate their target(s) remains to be elucidated and is a subject for further research.

The development of oil palm inflorescences is a long process that involves coordinated changes in expression levels of a large number of miRNAs and also their target transcript(s) across different developmental stages. As in other species, it can be concluded that subsets of miRNAs were expressed and act in concert with their mRNA targets to coordinate the developmental events in developing inflorescences, by involving floral transcription factors and signal transduction pathways mediated by protein kinases. The functional genomic data provided in this study, including the sRNAomes and a dense consensus transcriptome assembled from multiple tissues, may serve as genetic resources for exploring this plant's biology, specifically for tracking down key genetic regulators in oil palm reproductive development. Nevertheless, the miRNA-target regulatory modules predicted in this study provide a new insight into the role of the miRNA-mediated regulatory mechanisms in oil palm.

REFERENCES

Abhinand, C. S., Santhosh Mithra, V. S., & Sreekumar, J. (2013). An algorithm for plant microRNA target prediction. *Journal of Root Crops*, *39*(1), 22-27.

Achard, P., Herr, A., Baulcombe, D. C., & Harberd, N. P. (2004). Modulation of floral development by a gibberellin-regulated microRNA. *Development*, *131*(14), 3357-3365.

Acosta, I. F., Laparra, H., Romero, S. P., Schmelz, E., Hamberg, M., Mottinger, J. P., … Dellaporta, S. L. (2009). Tasselseed1 is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science*, *323*, 262-265.

Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V., & Sundaresan, V. (2005). Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Research*, *15*, 78-91.

Adam, H., Collin, M., Richaud, F., Beule, T., Cros, D., Omore, A., … Tregear, J. W. (2011). Environmental regulation of sex determination in oil palm: Current knowledge and insights from other species. *Annals of Botany*, *108*, 1529-1537.

Adam, H., Jouannic, S., Escoute, J., Duval, Y., Verdeil, J. L., & Tregear, J. W. (2005). Reproductive developmental complexity in the African oil palm (*Elaeis guineensis*, Arecaceae). *American Journal of Botany*, *92*, 1836-1852.

Adam, H., Jouannic, S., Morcillo, F., Richaud, F., Duval, Y., & Tregear, J. W. (2006). MADS box genes in oil palm (*Elaeis guineensis*): Patterns in the evolution of the SQUAMOSA, DEFICIENS, GLOBOSA, AGAMOUS, and SEPALLATA subfamilies. *Journal of Molecular Evolution*, *62*, 15-31.

Adam, H., Jouannic, S., Morcillo, F., Verdeil, J. L., Duval, Y., & Tregear, J. W. (2007a). Determination of flower structure in *Elaeis guineensis*: Do palms use the same homeotic genes as other species? *Annals of Botany*, *100*, 1-12.

Adam, H., Jouannic, S., Orieux, Y., Morcillo, F., Richaud, F., & Duval, Y. (2007b). Functional characterization of MADS box genes involved in the determination of oil palm flower structure. *Journal Experimental Botany*, *58*, 1245-1259.

Addo-Quaye, C., Miller, W., & Axtell, M. J. (2008). CleaveLand: A pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, *25*, 130-131.

Adnan, H. (2014, April 26). Palm oil biodiesel programme to cover all of Malaysia by July. *The Star*. Retrieved from http://www.thestar.com.my/business/business-news/2014/04/26/b5-goes-nationwide/

Ai, L., Chen, M., Zhang, Y., Chen, S., Zhou, X., & Chen, J. (2014). Comparative analysis of the miRNA profiles from *Taenia solium* and *Taenia asiatica* adult. *African Journal of Microbiology Research, 8*(9), 895-902.

Ainsworth, C., Crossley, S., Buchanan-Wollaston, V., Thangavelu, M., & Parker, J. (1995). Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. *Plant Cell, 7*, 1583-1598.

Ajambang, W., Volkaert, H., & Sudarsono, S. (2016). Carbohydrate deprivation upsurges the expression of genes responsible for programmed cell death in inflorescence tissues of oil palm (*Elaeis guineensis* Jacq.). *Turkish Journal of Biology, 40*, 1320-1327.

Alagna, F., Geu-Flores, F., Kries, H., Panara, F., Baldoni, L., O'Connor, S. E., & Osbourn, A. (2016). Identification and characterization of the iridoid synthase involved in oleuropein biosynthesis in olive (*Olea europaea*) fruits. *Journal of Biological Chemistry, 291*, 5542-5554.

Alam, M. Z., Kabbashi, A. A., Nahdatul, S., & Hussin, I. S. (2009). Production of bioethanol by direct bioconversion of oil-palm industrial effluent in a stirred-tank bioreactor. *Journal of Industrial Microbiology and Biotechnology, 36*, 801-808.

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., … Malek, J. A. (2011). *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology, 29*, 521-527.

Allen, E., Xie, Z., Gustafson, A. M., & Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell, 121*(2), 207-221.

Allison, D. B., Cui, X., Page, G. P., & Sabripour, M. (2016). Microarray data analysis: From disarray to consolidation and consensus. *Nature Reviews Genetics, 7*(1), 55-65.

Alves-Junior, L., Niemeier, S., Hauenschild, A., Rehmsmeier, M., & Merkle, T. (2009). Comprehensive prediction of novel microRNA targets in *Arabidopsis thaliana*. *Nucleic Acids Research, 37*(12), 4010-4021.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., … Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA*, *9*(3), 277-279.

Amor, B. B., Wirth, S., Merchan, F., Laporte, P., d'Aubenton-Carafa, Y., Hirsch, J., … Crespi, M. (2009). Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Research*, *19*(1), 57-69.

An, F. M., Hsiao, S. R., & Chan, M. T. (2011). Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in *Phalaenopsis* orchid. *PLoS ONE*, *6*(5), e18937.

An, W., Gong, W., He, S., Pan, Z., Sun, J., & Du, X. (2015). MicroRNA and mRNA expression profiling analysis revealed the regulation of plant height in *Gossypium hirsutum*. *BMC Genomics*, *16*, 886.

Andorf, C. M., Cannon, E. K., Portwood, J. L., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., … Lawrence-Dill, C. J. (2016). MaizeGDB update: New tools, data and interface for the maize model organism database. *Nucleic Acids Research*, *44*(D1), D1195-201.

Ang, S. K., Yahya, A., Abd Aziz, S., & Md Salleh, M. (2015). Isolation, screening, and identification of potential cellulolytic and xylanolytic producers for biodegradation of untreated oil palm trunk and its application in saccharification of lemongrass leaves. *Preparative Biochemistry & Biotechnology*, *45*(3), 279-305.

Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., & Baulcombe, D. C. (2005). Cloning and characterization of micro-RNAs from moss. *The Plant Journal*, *43*, 837-848.

Archak, S., & Nagaraju, J. (2007). Computational prediction of rice (*Oryza sativa*) miRNA targets. *Genomics Proteomics Bioinformatics*, *5*(3-4), 196-206.

Aryal, R., Jagadeeswaran, G., Zheng, Y., & Ming, R. (2014). Sex specific expression and distribution of small RNAs in papaya. *BMC Genomics*, *15*, 20.

Audic, S., & Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Research*, *7*(10), 986-995.

Aukerman, M. J., & Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *Plant Cell*, *15*(11), 2730-2741.

Axtell, M. J., & Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell*, *17*, 1658-1673.

Axtell, M. J., Snyder, J. A., & Bartel, D. P. (2007). Common functions for diverse small RNAs of land plants. *Plant Cell*, *19*, 1750-1769.

Bai, S. L., Peng, Y. B., Cui, J. X., Gu, H. T., Xu, L. Y., Li, Y. Q., … Bai, S. N. (2004). Developmental analyses reveal early arrests of the spore-bearing parts of reproductive organs in unisexual flowers of cucumber (*Cucumis sativus* L.). *Planta*, *220*(2), 230-240.

Baker, C. C., Sieber, P., Wellmer, F., & Meyerowitz, E. M. (2005). The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in *Arabidopsis*. *Current Biology*, *15*(4), 303-315.

Bakhshi, B., Fard, E. M., Nikpay, N., Ebrahimi, M. A., Bihamta, M. R., Mardi, M., & Salekdeh, G. H. (2016). MicroRNA signatures of drought signaling in rice root. *PLoS ONE*, *11*(6), e0156814.

Baldrich, P., Campo, S., Wu, M. T., Liu, T. T., Hsing, Y. I., & San Segundo, B. (2015). MicroRNA-mediated regulation of gene expression in the response of rice plants to fungal elicitors. *RNA Biology*, *12*(8), 847-863.

Ballantyne, K. N., van Oorschot, R. A., & Mitchell, R. J. (2008). Locked nucleic acids in PCR primers increase sensitivity and performance. *Genomics*, *91*(3), 301-305.

Barakat, A., Wall, K., Leebens-Mack, J., Wang, Y. J., Carlson, J. E., & DePamphilis, C. W. (2007). Large-scale identification of microRNAs from a basal eudicot (*Eschscholzia californica*) and conservation in flowering plants. *The Plant Journal*, *51*(6), 991-1003.

Bar-Akiva, A., Ovadia, R., Rogachev, I., Bar-Or, C., Bar, E., Freiman, Z., … Oren-Shamir, M. (2010). Metabolic networking in *Brunfelsia calycina* petals after flower opening. *Journal of Experimental Botany*, *61*(5), 1393-1403.

Barciszewska-Pacak, M., Milanowska, K., Knop, K., Bielewicz, D., Nuc, P., Plewka, P., … Szweykowska-Kulinska, Z. (2015). *Arabidopsis* microRNA expression regulation in a wide range of abiotic stress responses. *Frontiers in Plant Science*, *6*, 410.

Bartel, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, *116*(2), 281-297.

Basiron, Y. (2006, September 9). Pushing the case for Envodiesel. *Malaysian New Straits Times*, pp. 22.

Bennett, R. N., & Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytologist*, *127*(4), 617-633.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis Information Resource: Making and mining the 'gold standard' annotated reference plant genome. *Genesis*, *53*(8), 474-485.

Berezikov, E., Cuppen, E., & Plasterk, R. H. (2006). Approaches to microRNA discovery. *Nature Genetics*, *38*, S2-S7.

Bernama. (2006, March 21). PM launches 'Envo Diesel' biofuel. *Bernama*. Retrieved from http://www.bernama.com/bernama/v3/news_lite.php?id=187164

Beulé, T., Agbessi, M. D. T., Dussert, S., Jaligot, E., & Guyot, R. (2015). Genome-wide analysis of LTR-retrotransposons in oil palm. *BMC Genomics*, *16*, 795.

Beulé, T., Camps, C., Debiesse, S., Tranchant, C., Dussert, S., Sabau, X., … Tregear, J. (2011). Transcriptome analysis reveals differentially expressed genes associated with the *mantled* homeotic flowering abnormality in oil palm (*Elaeis guineensis*). *Tree Genetics & Genomes*, *7*, 169-182.

Bhargava, V., Head, S. R., Ordoukhanian, P., Mercola, M., & Subramaniam, S. (2014). Technical variations in low-input RNA-seq methodologies. *Scientific Reports*, *4*, 3678.

Bi, F., Meng, X., Ma, C., & Yi, G. (2015). Identification of miRNAs involved in fruit ripening in Cavendish bananas by deep sequencing. *BMC Genomics*, *16*, 776.

Billotte, N., Marseillac, N., Risterucci, A. M., Adon, B., Brottier, P., Baurens, F. C., … Charrier, A. (2005). Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics*, *110*, 754-765.

Biradar, N. V. (1978). Unusual inflorescence in *Elaeis guineensis*. *Principes*, *22*, 115.

Black, A. C. (2006). *Dictionary of agriculture* (3rd ed.). London, UK: A & C Black Publishers Ltd.

Blainey, P., Krzywinski, M., & Altman, N. (2014). Points of significance: Replication. *Nature Methods*, *11*(9), 879-880.

Blazquez, M. A., Soowal, L. N., Lee, I., & Weigel, D. (1997). LEAFY expression and flower initiation in *Arabidopsis*. *Development*, *124*, 3835-3844.

Bolouri Moghaddam, M. R., & Van den Ende, W. (2013). Sugars, the clock and transition to flowering. *Frontiers in Plant Science*, *4*, 22.

Bonnet, E., He, Y., Billiau, K., & Van de Peer, Y. (2010). TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinfomatics*, *26*(12), 1566-1568.

Bonnet, E., Wuyts, J., Rouze, P., & Van de Peer, Y. (2004). Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*, *20*, 2911-2917.

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*, *16*, 727–741.

Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J. P., … Frugier, F. (2008). MicroRNA166 controls root and nodule development in *Medicago truncatula*. *The Plant Journal*, *54*, 876-887.

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G. F., Drira, N., Ohlrogge, J. B., & Arondel, V. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12527-12532.

Bratzel, F., & Turck, F. (2015). Molecular memories in the regulation of seasonal flowering: From competence to cessation. *Genome Biology*, *16*, 192.

Brodersen, P., Sakvarelidze-Archard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., & Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, *320*, 1185-1190.

Brousse, C., Liu, Q., Beauclair, L., Deremetz, A., Axtell, M. J., & Bouche, N. (2014). A non-canonical plant microRNA target site. *Nucleic Acids Research*, *42*, 5270-5279.

Budde, C. F., Mahan, A. E., Lu, J., Rha, C., & Sinskey, A. J. (2010). Roles of multiple acetoacetyl coenzyme A reductases in polyhydroxybutyrate biosynthesis in *Ralstonia eutropha* H16. *Journal of Bacteriology*, *192*(20), 5319-5328.

Busch, A., Richter, A., & Backofen, R. (2008). IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, *24*(24), 2849-2856.

Cahill, M. J., Köser, C. U., Ross, N. E., & Archer, J. A. C. (2010). Read length and repeat resolution: Exploring prokaryote genomes using next-generation sequencing technologies. *PLoS ONE, 5*(7), e11518.

Calderon-Urrea, A., & Dellaporta, S. L. (1999). Cell death and cell protection genes determine the fate of pistils in maize. *Development*, *126*, 435-441.

Calviño, M., Bruggmann, R., & Messing, J. (2011). Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genomics*, *12*, 356.

Cao, S., Zhu, Q., Shen, W., Jiao, X., Zhao, X., Wang, M., … Liu, Q. (2013). Comparative profiling of miRNA expression in developing seeds of high linoleic and high oleic safflower (*Carthamus tinctorius* L.) plants. *Frontiers in Plant Science*, *4*, 489.

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Canesin, L. E. C., Pinto, L. R., … Vicentini, R. (2014). *De novo* assembly and tanscriptome analysis of contrasting sugarcane varieties. *PLoS ONE*, *9*(2), e88462.

Carra, A., Mica, E., Gambino, G., Pindo, M., Moser, C., Enrico Pè, M., & Schubert, A. (2009). Cloning and characterization of small non-coding RNAs from grape. *The Plant Journal*, *59*(5), 750-763.

Castilho, A., Vershinin, A., & Heslop-Harrison, J. S. (2000). Repetitive DNA and the chromosomes in the genome of oil palm (*Elaeis guineensis*). *Annals of Botany*, *85*, 837-844.

Chai, J., Feng, R., Shi, H., Ren, M., Zhang, Y., & Wang, J. (2015). Bioinformatic identification and expression analysis of banana microRNAs and their targets. *PLoS ONE*, *10*(4), e0123083.

Cheah, B. H., Nadarajah, K., Divate, M. D., & Wickneswari, R. (2015). Identification of four functionally important microRNA families with contrasting differential expression profiles between drought-tolerant and susceptible rice leaf at vegetative stage. *BMC Genomics*, 16, 692.

Chellappan, P., Xia, J., Zhou, X., Gao, S., Zhang, X., Coutino, G., … Jin, H. (2010). siRNAs from miRNA sites mediate DNA methylation of target genes. *Nucleic Acids Research*, *38*, 6883-6894.

Chen, F., Zhang, X., Zhang, N., Wang, S., Yin, G., Dong, Z., & Cui, D. (2015b). Combined small RNA and degradome sequencing reveals novel miRNAs and their targets in the high-yield mutant wheat strain Yunong 3114. *PLoS ONE*, *10*(9), e0137773.

Chen, M. X., Hu, W., Li, J., He, J. J., Ai, L., & Chen, J. X. (2016). Identification and characterization of microRNAs in the zoonotic fluke *Fasciolopsis buski*. *Parasitology Research*, *115*(6), 2433-2438.

Chen, T., Hao, Y. J., Zhang, Y., Li, M. M., Wang, M., Han, W., … Zhou, Q. (2015a). m(6)A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell*, *16*(3), 289-301.

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science*, *303*, 2022-2025.

Chen, X. (2005). MicroRNA biogenesis and function in plants. *FEBS Letters*, *579*(26), 5923-5931.

Chen, X., Li, W. C., & Fu, F. L. (2009). Bioinformatic prediction of microRNAs and their target genes in maize. *Yi Chuan*, *31*(11), 1149-1157.

Chen, X., Zhang, Z., Liu, D., Zhang, K., Li, A., & Mao, L. (2010). SQUAMOSA promoter-binding protein-like transcription factors: Star players for plant growth and development. *Journal of Integrative Plant Biology*, *52*(11), 946-951.

Cheng, H., Wang, Y., Tao, X., Fan, Y., Dai, Y., Yang, H., & Ma, X. (2016). Genomic profiling of exogenous abscisic acid-responsive microRNAs in tomato (*Solanum lycopersicum*). *BMC Genomics*, *17*, 423.

Cheung, A. Y., & de Vries, S. C. (2008). Membrane trafficking: Intracellular highways and country roads. *Plant Physiology*, *147*(4), 1451-1453.

Chi, X., Yang, Q., Chen, X., Wang, J., Pan, L., Chen, M., … Yu, S. (2011). Identification and characterization of microRNAs from peanut (*Arachis hypogaea* L.) by high-throughput sequencing. *PLoS ONE*, *6*, e27530.

Chorostecki, U., Crosa, V. A., Lodeyro, A. F., Bologna, N. G., Martin, A. P., Carrillo, N., … Palatnik, J. F. (2012). Identification of new microRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Research*, *40*, 8893-8904.

Chorostecki, U., & Palatnik, J. F. (2014). comTAR: A web tool for the prediction and characterization of conserved microRNA targets in plants. *Bioinformatics*, *30*(14), 2066-2067.

Christiaens, A., De Keyser, E., Pauwels, E., De Riek, J., Gobin, B., & Van Labeke, M. (2016). Suboptimal light conditions influence source-sink metabolism during flowering. *Frontiers in Plant Science*, *7*, 249.

Chu, Z., Chen, J., Xu, H., Dong, Z., Chen, F., & Cui, D. (2016). Identification and comparative analysis of microRNA in wheat (*Triticum aestivum* L.) callus derived from mature and immature embryos during *in vitro* culture. *Frontiers in Plant Science*, *7*, 1302.

Chuck, G., Meeley, R., Irish, E., Sakai, H., & Hake, S. (2007). The maize *tasselseed4* microRNA controls sex determination and meristem cell fate by targeting *Tasselseed6/indeterminate spikelet1*. *Nature Genetics*, *39*, 1517-1521.

Cobley, L. S., & Steele, W. M. (1976). *An introduction to the botany of tropical crops* (2nd ed.). London, UK: Longman Group Ltd.

Coen, E. S., & Meyerowitz, E. M. (1991). The war of the whorls: Genetic interactions controlling flower development. *Nature*, *353*(6339), 31-37.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J., Angenent, G. C., & van Tunen, A. J. (1995). The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell*, *7*, 1859-1868.

Colaiacovo, M., Subacchi, A., Bagnaresi, P., Lamontanara, A., Cattivelli, L., & Faccioli, P. (2010). A computational-based update on microRNAs and their targets in barley (*Hordeum vulgare* L.). *BMC Genomics*, *11*, 595.

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, *2008*, 619832.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., … Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13.

Corley, R. H. V. (Ed) (1976a). *Inflorescence abortion and sex differentiation*. Amsterdam, Netherlands: Elsevier.

Corley, R. H. V. (1976b). Sex differentiation in oil palm: Effects of growth regulators. *Journal of Experimental Botany*, *27*, 553-558.

Corley, R. H. V., & Gray, B. S. (1976). Growth and morphology. In R. H. V. Corley, J. J. Hardon & B. J. Wood (Eds.), *Developments in crop*. Amsterdam, Netherlands: Elsevier.

Corley, R. H. V., & Lee, C. H. (1992). The physiological basis for genetic improvement of oil palm in Malaysia. *Euphytica*, *60*, 179-184.

Corley, R. H. V., Lee, C. H., Law, I. H., & Wong, C. Y. (1986). Abnormal flower development in oil palm clones. *Planter (Kuala Lumpur)*, *62*, 233-240.

Corley, R. H. V., Ng, M., & Donough, C. R. (1995). Effects of defoliation on sex determination in oil palm clones. *Experimental Agriculture*, *31*, 177-190.

Corley, R. H. V., & Tinker, P. B. H. (2015). *The oil palm* (5th ed.). Hoboken, NJ: Wiley- Blackwell.

Cottrell, R. C. (1991). Introduction: Nutritional aspects of palm oil. *The American Journal of Clinical Nutrition*, *53*(4 Suppl), 989S-1009S.

Curaba, J., Spriggs, A., Taylor, J., Li, Z., & Helliwell, C. (2012). miRNA regulation in the early development of barley seed. *BMC Plant Biology, 12*, 120.

Da Silva, A. C., Grativol, C., Thiebaut, F., Hemerly, A. S., & Ferreira, P. C. G. (2016). Computational identification and comparative analysis of miRNA precursors in three palm species. *Planta*, *243*(5), 1265-1277.

Dai, X., & Zhao, P. X. (2011). psRNATarget: A plant small RNA target analysis server. *Nucleic Acids Research*, *39*, W155-159.

Degner, J. F., Marioni, J. C., Pai, A. A., Pickrell, J. K., Nkadori, E., Gilad, Y., & Pritchard, J. K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, *25*(24), 3207-3212.

De Smet, I., Voss, U., Jurgens, G., & Beeckman, T. (2009). Receptor-like kinases shape the plant. *Nature Cell Biology*, *11*, 1166-1173.

Dezulian, T., Palatnik, J. F., Huson D., & Weigel, D. (2005). Conservation and divergence of microRNA families in plants. *Genome Biology*, *6*, P13.

Ding, J., Li, D., Ohler, U., Guan, J., & Zhou, S. (2012). Genome-wide search for miRNA-target interactions in *Arabidopsis thaliana* with an integrated approach. *BMC Genomics*, *13*(Suppl 3), S3.

Ding, Y., & Lawrence, C. E. (2003). A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Research*, *31*, 7280-7301.

Ding, Y., Chan, C. Y., & Lawrence, C. E. (2005). RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA*, *11*(8), 1157-1166.

Dong, Z., Han, M. -H., & Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate *in vitro* processing of pri-miRNA by DCL1. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 9970-9975.

Dransfield, J., & Uhl, N. W. (1998). Palmae. In K. Kubitzki (Ed.), *Families and genera of vascular plants, flowering plants: Monocotyledons*. Berlin, Germany: Springer-Verlag.

Dryanova, A., Zakharov, A., & Gulick, P. J. (2008). Data mining for miRNAs and their targets in the *Triticeae*. *Genome*, *51*, 433-443.

Dsouza, M., Larsen, N., & Overbeek, R. (1997). Searching for patterns in genomic data. *Trends in Genetics*, *13*, 497-498.

Dugas, D. V., Monaco, M. K., Olson, A., Klein, R. R., Kumari, S., Ware, D., & Klein, P. E. (2011). Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics*, *12*, 514.

Durand-Gasselin, T., Noiret, J. M., Kouamé, R. K., Cochard, B., & Adon, B. (1999). Availability of quality pollen for improved oil palm (*Elaeis guineensis* Jacq.) seed production. *Plantations, Recherche, Développement*, *6*, 264-276.

Dussert, S., Guerin, C., Andersson, M., Joët, T., Tranbarger, T. J., Pizot, M., … Morcillo, F. (2013). Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiology*, *162*(3), 1337-1358.

Elo, A., Lemmetyinen, J., Turunen, M. L., Tikka, L., & Sopanen, T. (2001). Three MADS-box genes similar to APETALA1 and FRUITFULL from silver birch (*Betula pendula*). *Physiologia Plantarum*, *112*, 95-103.

Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., … Bowman, J. L. (2003). Radial patterning of *Arabidopsis* shoots by class III *HD-ZIP* and *KANADI* genes. *Current Biology*, *13*(20), 1768-1774.

Eom, I. Y., Yu, J. H., Jung, C. D., & Hong, K. S. (2015). Efficient ethanol production from dried oil palm trunk treated by hydrothermolysis and subsequent enzymatic hydrolysis. *Biotechnology for Biofuels*, *8*, 83.

Escalona, M., Rocha, S., & Posada, D. (2016). A comparison of tools for the simulation of genomic next-generation sequencing data. *Nature Reviews Genetics*, *17*(8), 459-469.

Fahlgren, N., & Carrington, J. C. (2010). MiRNA target prediction in plants. *Methods in Molecular Biology*, *592*, 51-57.

Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., … Carrington, J. C. (2007). High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS ONE*, *2*(2), e219.

Fang, L., Liang, Y., Li, D., Cao, X., & Zheng, Y. (2013). Dynamic expression analysis of miRNAs during the development process of oil palm mesocarp. *Plant Science Journal*, *31*, 304-312.

Fang, Y., & Spector, D. L. (2007). Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Current Biology*, *17*, 818-823.

Fang, Y. N., Zheng, B. B., Wang, L., Yang, W., Wu, X. M., Xu, Q., & Guo, W. W. (2016). High-throughput sequencing and degradome analysis reveal altered expression of miRNAs and their targets in a male-sterile cybrid pummelo (*Citrus grandis*). *BMC Genomics*, *17*, 591.

Fattash, I., Voß, B., Reski, R., Hess, W. R., & Frank, W. (2007). Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *BMC Plant Biology*, *7*, 13.

Ferreira, S. J., Senning, M., Sonnewald, S., Kessling, P. M., Goldstein, R., & Sonnewald, U. (2010). Comparative transcriptome analysis coupled to X-ray CT reveals sucrose supply and growth velocity as major determinants of potato tuber starch biosynthesis. *BMC Genomics*, *11*, 93.

Firon, N., LaBonte, D., Villordon, A., Kfir, Y., Solis, J., Lapis, E., … Nadir, L. A. (2013). Transcriptional profiling of sweetpotato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC Genomics*, *14*, 460.

Folkes, L., Moxon, S., Woolfenden, H. C., Stocks, M. B., Szittya, G., Dalmay, T., & Moulton, V. (2012). PAREsnip: A tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. *Nucleic Acids Research*, *40*(13), e103.

Formey, D., Sallet, E., Lelandais-Brière, C., Ben, C., Bustos-Sanmamed, P., Niebel, A., … Crespi, M. (2014). The small RNA diversity from *Medicago truncatula* roots under biotic interactions evidences the environmental plasticity of the miRNAome. *Genome Biology*, 15(9), 457.

Franchini, P., Van der Merwe, M., & Roodt-Wilding, R. (2011). Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. *BMC Research Notes*, *4*, 59.

Frazier, T. P., Xie, F., Freistaedter, A., Burklew, C. E., & Zhang, B. (2010). Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). *Planta*, *232*(6), 1289-1308.

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research, 40,* 37-52.

Gallego Romero, I., Pail, A. A., Tung, J., & Gilad, Y. (2014). RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biology, 12,* 42.

Galli, V., Guzman, F., de Oliveira, L. F. V., Loss-Morais, G., Körbes, A. P., Silva, S. D. A., … Margis, R. (2014). Identifying microRNAs and transcript targets in *Jatropha* seeds. *PLoS ONE, 9*(2), e83727.

Gao, C., Ju, Z., Cao, D., Zhai, B., Qin, G., Zhu, H., … Zhu, B. (2015a). MicroRNA profiling analysis throughout tomato fruit development and ripening reveals potential regulatory role of RIN on microRNAs accumulation. *Plant Biotechnology Journal, 13*(3), 370-382.

Gao, J., Yin, F., Liu, M., Luo, M., Qin, C., Yang, A., … Pan, G. (2015b). Identification and characterisation of tobacco microRNA transcriptome using high-throughput sequencing. *Plant Biology* (Stuttgart, Germany: Online), *17*(3), 591-598.

German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., … Green, P. J. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnology, 26,* 941-946.

Ghag, S. B., Shekhawat, U. K. S., & Canapathi, T. R. (2015). Small RNA profiling of two important cultivars of banana and overexpression of miRNA156 in transgenic banana plants. *PLoS ONE, 10*(5), e0127179.

Giurato, G., De Filippo, M. R., Rinaldi, A., Hashim, A., Nassa, G., Ravo, M., … Weisz, A. (2013). iMir: An integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. *BMC Bioinformatics, 14,* 362.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources, 11*(5), 759-769.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17,* 333-351.

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Research, 34,* D140-144.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005). Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Research, 33*, D121-4.

Grigg, S. P., Canales, C., Hay, A., & Tsiantis, M. (2005). SERRATE coordinates shoot meristem function and leaf axial patterning in *Arabidopsis*. *Nature*, *437*(7061), 1022-1026.

Guerin, C., Joët, T., Serret, J., Lashermes, P., Vaissayre, V., Agbessi, M. D. T., … Dussert, S. (2016). Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *The Plant Journal*, *87*(5), 423-441.

Guo, H., Kan, Y., & Liu, W. (2011). Differential expression of miRNAs in response to topping in flue-cured tobacco (*Nicotiana tabacum*) roots. *PLoS ONE*, *6*(12), e28565.

Guo, H. S., Xie, Q., Fei, J. F., & Chua, N. H. (2005). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *Plant Cell*, *17*(5), 1376-1386.

Guo, Q., Xiang, A. L., Yang, Q., & Yang, Z. M. (2007). Bioinformatic identification of microRNAs and their target genes from *Solanum tuberosum* expressed sequence tags. *Chinese Science Bulletin*, *52*(17), 2380-2389.

Guo, R., Chen, X., Lin, Y., Xu, X., Thu, M. K., & Lai, Z. (2016). Identification of novel and conserved miRNAs in leaves of *in vitro* grown *Citrus reticulata* "Lugan" plantlets by Solexa sequencing. *Frontiers in Plant Science*, *6*, 1212.

Guo, S., Zheng, Y., Joung, J. G., Liu, S., Zhang, Z., Crasta, O. R., … Fei, Z. (2010). Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics*, *11*, 384.

Guo, W., Wu, G., Yan, F., Lu, Y., Zheng, H., Lin, L., … Chen, J. (2012). Identification of novel *Oryza sativa* miRNAs in deep sequencing-based small RNA libraries of rice infected with *Rice stripe virus*. *PLoS ONE*, *7*(10), e46443.

Gutiérrez, L. F., Sánchez, O. J., & Cardona, C. A. (2009). Process integration possibilities for biodiesel production from palm oil using ethanol obtained from lignocellulosic residues of oil palm industry. *Bioresource Technology*, *100*(3), 1227-1237.

Habib, S. H., Saud, H. M., & Kausar, H. (2014). Efficient oil palm total RNA extraction with a total RNA extraction kit. *Genetics and Molecular Research*, *13*(2), 2359-2367.

Hartley, C. W. S. (1988). *The oil palm* (3rd ed.). London, UK: Longman Group Ltd.

Hausser, J., Syed, A. P., Bilen, B., & Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Research*, *23*, 604-615.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1-8.

Heuer, S., Lörz, H., & Dresselhaus, T. (2000). The MADS box gene ZmMADS2 is specifically expressed in maize pollen and during maize pollen tube growth. *Sexual Plant Reproduction*, *13*, 21-27.

Hibio, N., Hino, K., Shimizu, E., Nagata, Y., & Ui-Tei, K. (2012). Stability of miRNA 5′ terminal and seed regions is correlated with experimentally observed miRNA-mediated silencing efficacy. *Scientific Reports*, *2*, 996.

Hintzsche, J. D., Robinson, W. A., & Tan, A. C. (2016). A survey of computational tools to analyze and interpret whole exome sequencing data. *International Journal of Genomics*, *2016*, 7983236.

Ho, C., Kwan, Y., Choi, M., Tee, S., Ng, W., Lim, K., … Ong Abdullah, M. (2007). Analysis and functional annotation of expressed sequence tags (ESTs) from multiple tissues of oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics*, *8*, 381.

Ho, C. L., Tan, Y. C., Yeoh, K. A., Ghazali, A. K., Yee, W. Y., & Hoh, C. C. (2016). *De novo* transcriptome analyses of host-fungal interactions in oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics*, *17*, 66.

Ho, H., Gudimella, R., Ong-Abdullah, M., & Harikrishna, J. A. (2017). Expression of microRNAs during female inflorescence development in African oil palm (*Elaeis guineensis* Jacq.). *Tree Genetics & Genomes*, *13*, 35.

Ho, H., Low, J. Z., Gudimella, R., Tammi, M. T., & Harikrishna, J. A. (2015). Expression patterns of inflorescence- and sex-specific transcripts in male and female inflorescences of African oil palm (*Elaeis guineensis*). *Annals of Applied Biology*, *168*(2), 274-289.

Ho, T., Pallett, D., Rusholme, R., Dalmay, T., & Wang, H. (2006). A simplified method for cloning short interfering RNAs from *Brassica juncea* infected with Turnip mosaic potyvirus and Turnip crinkle carmovirus. *Journal of Virological Methods*, *136*(1-2), 217-223.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, *31*, 3429-3431.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie*, *125*, 167-188.

Hong, Y., & Jackson, S. (2015). Floral induction and flower formation- the role and potential applications of miRNAs. *Plant Botechnology Journal*, *13*, 282-292.

Houseley, J., & Tollervey, D. (2009). The many pathways of RNA degradation. *Cell*, *136*(4), 763-776.

Hu, J., Jin, J., Qian, Q., Huang, K., & Ding, Y. (2016). Small RNA and degradome profiling reveals miRNA regulation in the seed germination of ancient eudicot *Nelumbo nucifera*. *BMC Genomics*, *17*(1), 684.

Hu, Z., Jiang, Q., Ni, Z., Chen, R., Xu, S., & Zhang, H. (2013). Analyses of a *Glycine max* degradome library identify microRNA targets and microRNAs that trigger secondary siRNA biogenesis. *Journal of Integrative Plant Biology*, *55*(2), 160-176.

Huang, H., Xu, L., Tong, Z., Lin, E., Liu, Q., Cheng, L., & Zhu, M. (2012). *De novo* characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. *BMC Genomics*, *13*, 648.

Huang, J., Lin, C., Cheng, T., Chang, B. C., Cheng, S., Chen, Y., … Chen, F. (2015). A *de novo* floral transcriptome reveals clues into *Phalaenopsis* orchid flower development. *PLoS ONE*, *10*(5), e0123474.

Huijser, P., & Schmid, M. (2011). The control of developmental phase transitions in plants. *Development*, *138*, 4117-4129.

Hundertmark, M., & Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*, *9*, 118.

Ibrahim, H., Pertl, H., Pittertschatscher, K., Fadl-Allah, E., El-Shahed, A., Bentrup, F. W., & Obermeyer, G. (2002). Release of an acid phosphatase activity during lily pollen tube growth involves components of the secretory pathway. *Protoplasma, 219,* 176-183.

Ishida, M., & Abu Hassan, O. (1997). Utilization of oil palm fronds as cattle feed. *Japanese Agricultural Research Quarterly, 31,* 41-47.

Itaya, A., Bundschuh, R., Archual, A. J., Joung, J. G., Fei, Z., Dai, X., … Ding, B. (2008). Small RNAs in tomato fruit and leaf development. *Biochimica et Biophysica Acta, 1779*(2), 99-107.

Iwakawa, H. O., & Tomari, Y. (2013). Molecular insights into microRNA-mediated translational repression in plants. *Molecular Cell, 52,* 591-601.

Jacquemard, J. C. (1995). *Le palmier a` huile.* Paris, France: Maisonneuve & Larose.

Jaeger, K. E., Graf, A., & Wigge, P. A. (2006). The control of flowering in time and space. *Journal of Experimental Botany, 57,* 3415-3418.

Jain, S., Chittem, K., Brueggeman, R., Osorno, J. M., Richards, J., & Nelson Jr., B. D. (2016). Comparative transcriptome analysis of resistant and susceptible common bean genotypes in response to soybean cyst nematode infection. *PLoS ONE, 11*(7), e0159338.

Jalani, B. S., Kushairi, A., & Cheah, S. C. (2003). Oil palm and coconut. In D. J. Murphy & B. G. Murray (Eds.), *Encyclopedia of applied plant sciences* (pp. 960-969). London, UK: Elsevier Academic Press.

Jaligot, E., Adler, S., Debladis, É., Beulé, T., Richaud, F., Ilbert, P., … Rival, A. (2011). Epigenetic imbalance and the floral developmental abnormality of the *in vitro*-regenerated oil palm *Elaeis guineensis*. *Annals of Botany, 108*(8), 1453-1462.

Jaligot, E., Rival, A., Beulé, T., Dussert, S., & Verdeil, J. –L. (2000). Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): The DNA methylation hypothesis. *Plant Cell Reports, 19*(7), 684-690.

Jensen, S. G., Lamy, P., Rasmussen, M. H., Ostenfeld, M. S., Dyrskjøt, L., Ørntoft, T. F., & Andersen, C. L. (2011). Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. *BMC Genomics, 12,* 435.

Jha, A., & Shankar, R. (2011). Employing machine learning for reliable miRNA target identification in plants. *BMC Genomics*, *12*, 636.

Jia, L., Zhang, D., Qi, X., Ma, B., Xiang, Z., & He, N. (2014). Identification of the conserved and novel miRNAs in mulberry by high-throughput sequencing. *PloS ONE*, *9*(8), e104409.

Jiao, Y., Song, W., Zhang, M., & Lai, J. (2011). Identification of novel maize miRNAs by measuring the precision of precursor processing. *BMC Plant Biology*, *11*, 141.

Jin, J., Lee, M., Bai, B., Sun, Y., Qu, J., Rahmadsyah, … Yue, G. H. (2016). Draft genome sequence of an elite *Dura* palm and whole-genome patterns of DNA variation in oil palm. *DNA Research*, *23*(6), 527-533.

Jin, W., Li, N., Zhang, B., Wu, F., Li, W., Guo, A., & Deng, Z. (2008). Identification and verification of microRNA in wheat (*Triticum aestivum*). *Journal of Plant Research*, *121*, 351-355.

Jin, W., & Wu, F. (2015). Characterization of miRNAs associated with *Botrytis cinerea* infection of tomato leaves. *BMC Plant Biology*, *15*, 1.

Johnson, M. T. J., Carpenter, E. J., Tian, Z., Bruskiewich, R., Burris, J. N., Carrigan, C. T., … Wong, G. K. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS ONE*, *7*(11), e50226.

Jones-Rhoades, M. W., & Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell*, *14*(6), 787-799.

Jones-Rhoades, M. W., Bartel, D. P., & Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. *Annual Review of Plant Biology*, *57*, 19-53.

Jørgensen, H., Sanadi, A. R., Felby, C., Lange, N. E., Fischer, M., & Ernst, S. (2010). Production of ethanol and feed by high dry matter hydrolysis and fermentation of palm kernel press cake. *Applied Biochemistry and Biotechnology*, *161*(1-8), 318-332.

Joshi, T., Yan, Z., Libault, M., Jeong, D. H., Park, S., Green, P.J., … Stacey, G. (2010). Prediction of novel miRNAs and associated target genes in *Glycine max*. *BMC Bioinformatics*, *11*, S14.

Jouannic, S., Argout, X., Lechauve, F., Fizames, C., Borgel, A., Morcillo, F., … Tregear, J. (2005). Analysis of expressed sequence tags from oil palm (*Elaeis guineensis*). *FEBS Letters*, *579*(12), 2709-2714.

Jourdan, C., Michaux-Ferrière, N., & Perbal, G. (2000). Root system architecture and gravitropism in the oil palm. *Annals of Botany*, *85*, 861-868.

Juarez, M. T., Kui, J. S., Thomas, J., Heller, B. A., & Timmermans, M. C. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. *Nature*, *428*(6978), 84-88.

Kakrana, A., Hammond, R., Patel, P., Nakano, M., & Meyers, B. C. (2014). sPARTA: A parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. *Nucleic Acids Research*, *42*(18), e139.

Kalam, M. A., Masjuki, H. H., Saifullah, M. G., & Seng, T. B. (2008). Envo diesel test on automotive engine– an analysis of its performance and emissions results. *International Journal of Mechanical and Materials Engineering*, *3*(1), 55-60.

Kang, M., Zhao, Q., Zhu, D., & Yu, J. (2012). Characterization of microRNAs expression during maize seed development. *BMC Genomics*, *13*, 360.

Kantar, M., Unver, T., & Budak, H. (2010). Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. *Functional & Integrative Genomics*, *10*, 493-507.

Kater, M. M., Franken, J., Carney, K. J., Colombo, L., & Angenent, G. C. (2001). Sex determination in the monoecious species cucumber is confined to specific floral whorls. *Plant Cell*, *13*, 481-493.

Katiyar A., Smita S., Chinnusamy V., Pandey D. M., & Bansal K. (2012). Identification of miRNAs in sorghum by using bioinformatics approach. *Plant Signaling & Behavior*, *7*, 246-259.

Katiyar, A., Smita, S., Muthusamy, S. K., Chinnusamy, V., Pandey, D. M., & Bansal, K. C. (2015). Identification of novel drought-responsive microRNAs and trans-acting siRNAs from *Sorghum bicolor* (L.) Moench by high-throughput sequencing analysis. *Frontiers in Plant Science*, *6*, 506.

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., … Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, *6*, 4.

Kazama, Y., Fujiwara, M. T., Koizumi, A., Nishihara, K., Nishiyama, R., Kifune, E., … Kawano, S. (2009). A SUPERMAN-like gene is exclusively expressed in female flowers of the dioecious plant *Silene latifolia*. *Plant and Cell Physiology*, *50*, 1127-1141.

Khan Barozai, M. Y., Irfan, M., Yousaf, R., Ali, I., Qaisar, U., Maqbool, A., … Riazuddin, S. (2008). Identification of micro-RNAs in cotton. *Plant Physiology and Biochemistry*, *46*, 739-751.

Khraiwesh, B., Arif, M. A., Seumel, G. I., Ossowski, S., Weigel, D., Reski, R., & Frank, W. (2010). Transcriptional control of gene expression by microRNAs. *Cell*, *140*, 111-122.

Kidner, C. A., & Martienssen, R. A. (2005). The role of ARGONAUTE (AGO1) in meristem formation and identity. *Developmental Biology*, *280*, 504-517.

Kiefer, E., Heller, W., & Ernst, D. (2000). A simple and efficient protocol for isolation of functional RNA from plant tissues rich in secondary metabolites. *Plant Molecular Biology Reporter*, *18*, 33-39.

Kim, J. C., Laparra, H., Calderon-Urrea, A., Mottinger, J. P., Moreno, M. A., & Dellaporta, S. L. (2007). Cell cycle arrest of stamen initials in maize sex determination. *Genetics*, *177*, 2547-2551.

Kim, S., Yang, J. -Y., Xu, J., Jang, I. -C., Prigge, M. J., & Chua, N. -H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary microRNAs. *Plant and Cell Physiology*, *49*, 1634-1644.

Koushki, M., Nahidi, M., & Cheraghali, F. (2015). Physico-chemical properties, fatty acid profile and nutrition in palm oil. *Journal of Paramedical Sciences*, *6*(3), 117-134.

Kramer, M. F. (2011). Stem-loop RT-qPCR for miRNAs. *Current Protocols in Molecular Biology*, *Chapter 15*, Unit 15.10.

Kresnowati, M. T. A. P., Setiadi, T., Tantra, T. M., & Rusdi, D. (2016). Microbial production of xylitol from oil palm empty fruit bunch hydrolysate: Effects of inoculum and pH. *Journal of Engineering and Technological Sciences, 48*(5), 523-533.

Kulcheski, F. R., Oliveira, L. F., Molina, L. G., Almerão, M. P., Rodrigues, F. A., Marcolino, J., … Margis, R. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics*, *12*, 307.

Kume, T., Matsuhashi, S., Ito, H., Hashimoto, S., Ishigaki, I., Awang, M. R., … Hamdani, H. (1998). Study on upgrading of oil palm wastes to animal feeds by radiation and fermentation processing. *JAERI – Research*, *98*, 013.

Lacombe, S., Nagasaki, H., Santi, C., Duval, D., Piégu, B., Bangratz, M., … Echeverria, M. (2008). Identification of precursor transcripts for 6 novel miRNAs expands the diversity on the genomic organisation and expression of miRNA genes in rice. *BMC Plant Biology*, *8*, 123.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., & Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science*, *294*(5543), 853-858.

Lai, E. C. (2003). MicroRNAs: Runts of the genome assert themselves. *Current Biology, 13*(23), R925-R936.

Laitinen, R. A. E., Broholm, S., Albert, V. A., Teeri, T. H., & Elomaa, P. (2006). Patterns of MADS-box gene expression mark flower-type development in *Gerbera hybrida* (Asteraceae). *BMC Plant Biology*, *6*, 11.

Lal, M., & Jaiswal, V. S. (1988). Modification of flower sex and acid phosphatase activity by phthalimides in female plants of *Morus nigra* L. *Plant Growth Regulation*, *7*, 29-37.

Lanet, E., Delannoy, E., Sormani, R., Floris, M., Brodersen, P., Crété, P., … Robaglia, C. (2009). Biochemical evidence for translational repression by *Arabidopsis* microRNAs. *Plant Cell*, *21*, 1762-1768.

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

Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, *294*, 858-862.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J. U., Rätsch, G., & Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 8795-8800.

Lawson, M. J., & Zhang, L. (2006). Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biology*, *7*, R14.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, *75*(5), 843-854.

Lee, W. S., Gudimella, R., Wong, G. R., Tammi, M. T., Khalid, N., & Harikrishna, J. A. (2015). Transcripts and microRNAs responding to salt stress in *Musa acuminata* Colla (AAA Group) cv. Berangan roots. *PLoS ONE*, *10*, e0127526.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, *23*(20), 4051-4060.

Legros, S., Mialet-Serra, I., Caliman, J. P., Siregar, F. A., Clement-Vidal, A., Fabre, D., & Dingkuhn, M. (2009). Phenology, growth and physiological adjustments of oil palm (*Elaeis guineensis*) to sink limitation induced by fruit pruning. *Annals of Botany*, *104*, 1183-1194.

Lei, X., Xiao, Y., Xia, W., Mason, A. S., Yang, Y., Ma, Z., & Peng, M. (2014). RNA-seq analysis of oil palm under cold stress reveals a different C-repeat binding factor (CBF) mediated gene expression pattern in *Elaeis guineensis* compared to other species. *PLoS ONE*, *9*(12), e114482.

Lelandais-Briere, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., … Crespi, M. (2009). Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell*, *21*(9), 2780-2796.

Levy, S. E., & Myers, R. M. (2016). Advancements in next-generation sequencing. *Annual Review of Genomics and Human Genetics*, *17*, 95-115.

Li, B., Duan, H., Li, J., Deng, X. W., Yin, W., & Xia, X. (2013). Global identification of miRNAs and targets in *Populus euphratica* under salt stress. *Plant Molecular Biology*, *81*(6), 525-539.

Li, C., & Zhang, B. (2016). MicroRNAs in control of plant development. *Journal of Cellular Physiology*, *231*(2), 303-313.

Li, F., Orban, R., & Baker, B. (2012). SoMART: A web server for plant miRNA, tasiRNA and target gene analysis. *The Plant Journal*, *70*(5), 891-901.

Li, H., Dong, Y., Yin, H., Wang, N., Yang, J., Liu, X., … Li, X. (2011). Characterization of the stress associated microRNAs in *Glycine max* by deep sequencing. *BMC Plant Biology*, *11*, 170.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078-2079.

Li, J., Reichel, M., & Millar, A. A. (2014). Determinants beyond both complementarity and cleavage govern microR159 efficacy in *Arabidopsis*. *PLoS Genetics*, *10*(3), e1004232.

Li, L., Li, C., & Howe, G. A. (2001). Genetic analysis of wound signaling in tomato. Evidence for a dual role of jasmonic acid in defense and female fertility. *Plant Physiology*, *127*, 1414-1417.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., … Wang, J. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research*, *20*, 265-272.

Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., … Chen, X. (2013). MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in *Arabidopsis*. *Cell*, *153*(3), 562-574.

Li, T., Ma, L., Geng, Y., Hao, C., Chen, X., & Zhang, X. (2015). Small RNA and degradome sequencing reveal complex roles of miRNAs and their targets in developing wheat grains. *PLoS ONE*, *10*(10), e0139658.

Li, X., Lawas, L. M., Malo, R., Glaubitz, U., Erban, A., Mauleon, R., … Jagadish, K. S. (2015). Metabolic and transcriptomic signatures of rice floral organs reveal sugar starvation as a factor in reproductive failure under heat and drought stress. *Plant, Cell & Environment*, *38*(10), 2171-2192.

Li, X., Xie, X., Li, J., Cui, Y., Hou, Y., Zhai, L., … Bian, S. (2017). Conservation and diversification of the miR166 family in soybean and potential roles of newly identified miR166s. *BMC Plant Biology*, *17*(1), 32.

Li, X., & Zhang, Y. Z. (2005). Computational detection of microRNAs targeting transcription factor genes in *Arabidopsis thaliana*. *Computational Biology and Chemistry*, *29*, 360-367.

Li, Y., Li, W., & Jin, Y. X. (2005). Computational identification of novel family members of microRNA genes in *Arabidopsis thaliana* and *Oryza sativa*. *Acta Biochimica et Biophysica Sinica*, *37*(2), 75-87.

Lian, H., Li, X., Liu, Z., & He, Y. (2013). HYL1 is required for establishment of stamen architecture with four microsporangia in *Arabidopsis*. *Journal of Experimental Botany*, *64*(11), 3397-3410.

Lian, S., Cho, W. K., Kim, S., Choi, H., & Kim, K. (2016). Time-course small RNA profiling reveals rice miRNAs and their target genes in response to *Rice stripe virus* infection. *PLoS ONE*, *11*(9), e0162319.

Liang, G., Ai, Q., & Yu, D. (2015). Uncovering miRNAs involved in crosstalk between nutrient deficiencies in *Arabidopsis*. *Scientific Reports*, *5*, 11813.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., … Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, *433*, 769-773.

Lin, H. C., Morcillo, F., Dussert, S., Tranchant-Dubreuil, C., Tregear, J. W., & Tranbarger, T. J. (2009). Transcriptome analysis during somatic embryogenesis of tropical monocot *Elaeis guineensis*: Evidence for conserved gene functions on early development. *Plant Molecular Biology*, *70*, 173-192.

Lindow, M., & Krogh, A. (2005). Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics*, *6*, 119.

Liu, H., Qin, C., Chen, Z., Zuo, T., Yang, X., Zhou, H., … Pan, G. (2014b). Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. *BMC Genomics*, *15*, 25.

Liu, J., Yin, T., Ye, N., Chen, Y., Yin, T., Liu, M., & Hassani, D. (2013). Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). *PLoS ONE*, *8*(4), e60181.

Liu, J. J., Goh, C. J., Loh, C. S., Liu, P., & Pua, E. C. (1998). A method for isolation of total RNA from fruit tissues of banana. *Plant Molecular Biology Reporter*, *16*(1), 1-6.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., … Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, *2012*, 251364.

Liu, N., Tu, L., Wang, L., Hu, H., Xu, J., & Zhang, X. (2017). MicroRNA 157-targeted *SPL* genes regulate floral organ size and ovule production in cotton. *BMC Plant Biology*, *17*, 7.

Liu, Q., Wang, F., & Axtell, M. J. (2014a). Analysis of complementarity requirements for plant microRNA targeting using a *Nicotiana benthamiana* quantitative transient assay. *Plant Cell*, *26*(2), 741-753.

Liu, S., Li, J. H., Wu, J., Zhou, K. R., Zhou, H., Yang, J. H., & Qu, L. H. (2015). StarScan: A web server for scanning small RNA targets from degradome sequencing data. *Nucleic Acids Research*, *43*(W1), W480-6.

Liu, X., Huang, J., Wang, Y., Khanna, K., Xie, Z., Owen, H. A., & Zhao, D. (2010). The role of floral organs in carpels, an *Arabdopsis* loss-of-function mutation in microRNA160a, in organogenesis and the mechanism regulating its expression. *The Plant Journal*, *62*, 416-428.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, *25*, 402-408.

Llave, C., Kasschau, K. D., rector, M. A., & Carrington, J. C. (2002). Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, *14*(7), 1605-1619.

Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C., & Clarke, J. (2006). SERRATE: A new player on the plant microRNA scene. *EMBO Reports*, *7*, 1052-1058.

Low, E. L., Alias, H., Boon, S., Shariff, E. M., Tan, C. A., Ooi, L. C. L., … Singh, R. (2008). Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biology*, *8*, 62.

Low, E. L., Rosli, R., Jayanthi, N., Mohd-Amin, A. H., Azizi, N., Chang, K., … Singh, R. (2014). Analyses of hypomethylated oil palm gene space. *PLoS ONE*, *9*(1), e86728.

Luo, Y., Guo, Z., & Li, L. (2013). Evolutionary conservation of microRNA regulatory programs in plant flower development. *Developmental Biology*, *380*, 133-144.

Ma, G., Zhang, L., Yungyuen, W., Tsukamoto, I., Iijima, N., Oikawa, M., … Kato, M. (2016). Expression and functional analysis of citrus carotene hydroxylases: Unravelling the xanthophyll biosynthesis in citrus fruits. *BMC Plant Biology*, *16*, 148.

Madon, M., Clyde, M. M., & Cheah, S. C. (1998). Cytological analysis of *Elaeis guineensis* and *E. oleifera* chromosomes. *Journal of Oil Palm Research*, *10*(1), 68-91.

Makita, Y., Shimada, S., Kawashima, M., Kondou-Kuriyama, T., Toyoda, T., & Matsui, M. (2015). MOROKOSHI: Transcriptome database in *Sorghum bicolor*. *Plant and Cell Physiology*, *56*(1), e6.

Malaysian Palm Oil Board. (2016b). Export of palm oil by destination: 2016 (tonnes). Selangor, Malaysia: Author.

Malaysian Palm Oil Board. (2016a). Overview of the Malaysian oil palm industry 2015. Selangor, Malaysia: Author.

Malaysian Palm Oil Council. (2012). Palm oil and palm kernel oil applications. Selangor, Malaysia: Author.

Malaysia's pioneer bioplastics pilot plant is operational (2011, July 13). *The Star*. Retrieved from http://www.thestar.com.my/business/business-news/2011/07/13/malaysias-pioneer-bioplastics-pilot-plant-is-operational/

Mantello, C. C., Cardoso-Silva, C. B., Da Silva, C. C., De Souza, L. M., Scaloppi Junior, E. J., Goncalves, P. S., … De Souza, L. M. (2014). *De novo* assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. *PLoS ONE*, *9*, e102665.

Mao, W., Li, Z., Xia, X., Li, Y., & Yu, J. (2012). A combined approach of high-throughput sequencing and degradome analysis reveals tissue specific expression of microRNAs and their targets in cucumber. *PLoS ONE*, *7*(3), e33040.

Markham, N. R., & Zuker, M. (2008). UNAFold: Software for nucleic acid folding and hybridization. *Methods in Molecular Biology*, *453*, 3-31.

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., … Bendahmane, A. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature*, *461*, 1135-1138.

Mba, O. I., Dumont, M. J., & Ngadi, M. (2015). Palm Oil: Processing, characterization and utilization in the food industry. *Food Bioscience*, *10*, 26-41.

McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structures. *Biopolymers*, *29*, 1105-1119.

Md Nasaruddin, N., Harikrishna, K., Othman, R. Y., Lim, S. H., & Harikrishna, J. A. (2007). Computational prediction of microRNAs from oil palm (*Elaeis guineensis* Jacq.) expressed sequence tags. *Asia Pacific Journal of Molecular Biology and Biotechnology*, *15*, 107-113.

Mehrpooyan, F., Othman, R. Y., & Harikrishna, J. A. (2012). Tissue and temporal expression of miR172 paralogs and the AP2-like target in oil palm (*Elaeis guineensis* Jacq.). *Tree Genetics & Genomes*, *8*, 1331-1343.

Meister, G., & Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature*, *431*, 343-349.

Mekhilef, S., Siga, S., & Saidur, R. (2010). A review on palm oil biodiesel as a source of renewable fuel. *Renewable and Sustainable Energy Reviews*, *15*, 1937-1949.

Meléndez, M. R., & Ponce, W. P. (2016). Pollination in the oil palms *Elaeis guineensis*, *E. oleifera* and their hybrids (OxG), in tropical America. *Pesquisa Agropecuária Tropical*, *46*(1), 102-110.

Mello, C. C., & Conte, D. Jr. (2004). Revealing the world of RNA interference. *Nature*, *431*, 338-342.

Meng, F., Liu, H., Wang, K., Liu, L., Wang, S., Zhao, Y., … Li, Y. (2013). Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). *BMC Plant Biology*, *13*, 140.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., … Qi, Y. (2008). Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5 ′ terminal nucleotide. *Cell*, *133*(1), 116-127.

Mica, E., Piccolo, V., Delledonne, M., Ferrarini, A., Pezzotti, M., Casati, C., … Horner, D. S. (2010). Correction: High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in *Vitis vinifera*. *BMC Genomics*, *11*, 109.

Milev, I., Yahubyan, G., Minkov, I., & Baev, V. (2011). miRTour: Plant miRNA and target prediction tool. *Bioinformation*, *6(*6), 248-249.

Min, X. J., Butler, G., Storms, R., & Tsang, A. (2005). OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, *33*(Web Server issue), W677-80.

Mitra, R., & Bandyopadhyay, S. (2011). MultiMiTar: A novel multi objective optimization based miRNA-target prediction method. *PLoS ONE*, *6*(9), e24583.

Mohamed, W. Z., & Alimon, A. R. (2014, Oct). *Recent advances in the utilization of oil palm by-products as animal feed*. Paper presented at the International Conference on Livestock Production and Veterinary Technology (ICARD), Bogor, Indonesia.

Mohanpuria, P., & Yadav, S. K. (2012). Characterization of novel small RNAs from tea (*Camellia sinensis* L.). *Molecular Biology Reports*, *39*(4), 3977-3986.

Mohorianu, I., Schwach, F., Jing, R. C., Lopez-Gomollon, S., Moxon, S., Szittya, G., … Dalmay, T. (2011). Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *The Plant Journal*, *67*, 232-246.

Mokhtar, A., Hassan, K., Abdul Aziz, A., & Wahid, M. B. (2011). Plywood from oil palm trunks. *Journal of Oil Palm Research*, *23*(3), 1159-1165.

Morcillo, F., Gagneur, C., Adam, H., Richaud, F., Singh, R., Cheah, S. C., … Tregear, J. W. (2006). Somaclonal variation in micropropagated oil palm. Characterization of two novel genes with enhanced expression in epigenetically abnormal cell lines and in response to auxin. *Tree Physiology*, *26*(5), 585-594.

Morin, R. D., Aksay, G., Dolgosheina, E., Ebhardt, H. A., Magrini, V., Mardis, E. R., … Unrau, P. J. (2008). Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Research*, *18*, 571-584.

Motta, F. L., & Santana, M. H. (2014). Solid-state fermentation for humic acids production by a *Trichoderma reesei* strain using an oil palm empty fruit bunch as the substrate. *Applied Biochemistry and Biotechnology, 172*(4), 2205-2217.

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme-Pilcher, R. L., Moulton, V., & Dalmay, T. (2008a). Deep sequencing of tomato short RNAs identified microRNAs targeting genes involved in fruit ripening. *Genome Research*, *18*(10), 1602-1609.

Moxon, S., Schwach, F., MacLean, D., Dalmay, T., Studholme, D. J., & Moulton, V. (2008b). A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics*, *24*(19), 2252-2253.

Moyroud, E., Minguet, E. G., Ott, F., Yant, L., Posé, D., Monniaux, M., … Parcy, F. (2011). Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. *Plant Cell*, *23*(4), 1293-1306.

Muhlemann, J. K., Maeda, H., Chang, C. Y., Miguel, P. S., Baxter, I., Cooper, B., … Dudareva, N. (2012). Developmental changes in the metabolic network of snapdragon flowers. *PLoS ONE*, *7*, e40381.

Mumtaz, T., Yahaya, N. A., Abd-Aziz, S., Rahman, N. A., Yee, P. L., Shirai, Y., & Hassan, M. A. (2010). Turning waste to wealth-biodegradable plastics polyhydroxyalkanoates from palm oil mill effluent – a Malaysian perspective. *Journal of Cleaner Production*, *18*(14), 1393-1402.

Murai, K. (2013). Homeotic genes and the ABCDE model for floral organ formation in wheat. *Plants*, *2*(3), 379-395.

Murphy, D. J. (2009). Oil palm: Future prospects for yield and quality improvements. *Lipid Technology*, *21*(11-12), 257-260.

Nag, A., Yang, Y., & Jack, T. (2007). DORNROSCHEN-LIKE, an AP2 gene, is necessary for stamen emergence in *Arabidopsis*. *Plant Molecular Biology*, *65*(3), 219-232.

Nakano, M., Nobuta, K., Vemaraju, K., Tej, S. S., Skogen, J. W., & Meyers, B. C. (2006). Plant MPSS databases: Signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Research*, *34*(Database issue), D731-5.

Narsai, R., Devenish, J., Castleden, I., Narsai, K., Xu, L., Shou, H., & Whelan, J. (2013). Rice DB: An *Oryza* information portal linking annotation, sub-cellular location, function, expression, regulation and evolutionary information for rice and Arabidopsis. *The Plant Journal*, *76*(6), 1057-1073.

Nie, S., Xu, L., Wang, Y., Huang, D., Muleke, E. M., Sun, X., … Liu, L. (2015). Identification of bolting-related microRNAs and their targets reveals complex miRNA-mediated flowering-time regulatory networks in radish (*Raphanus sativus* L.). *Scientific Reports*, *5*, 14034.

Nozawa, M., Miura, S., & Nei, M. (2012). Origins and evolution of microRNA genes in plant species. *Genome Biology and Evolution*, *4*(3), 230-239.

Numnark, S., Mhuantong, W., Ingsriswang, S., & Wichadakul, D. (2012). C-mii: A tool for plant miRNA and target identification. *BMC Genomics*, *13*(Suppl 7), S16.

Ó'maoiléidigh, D. S., Graciet, E., & Wellmer, F. (2014). Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytologist*, *201*, 16-30.

Ong-Abdullah, M., Ordway, J. M., Jiang, N., Ooi, S., Kok, S., Sarpan, N., … Martienssen, R. A. (2015). Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*, *525*, 533-537.

Ossowski, S., Schwab, R., & Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *The Plant Journal*, *53*, 674-690.

Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., & Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*, *425*(6955), 257-263.

Palatnik, J. F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., … Weigel, D. (2007). Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. *Developmental Cell*, *13*(1), 115-125.

Pang, M. X., Woodward, A. W., Agarwal, V., Guan, X. Y., Ha, M. S., Ramachandran, V., … Chen, Z. J. (2009). Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biology*, *10*, R122.

Pant, B. D., Buhtz, A., Kehr, J., & Scheible, W. R. (2008). MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *The Plant Journal*, *53*, 731-738.

Papp, I., Mette, M. F., Aufsatz, W., Daxinger, L., Schauer, S. E., Ray, A., … Matzke, A. J. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiology*, *132*(3), 1382-1390.

Park, J. H., Ishikawa, Y., Yoshida, R., Kanno, A., & Kameya, T. (2003). Expression of *AODEF*, a B-functional MADS-box gene, in stamens and inner tepals of the dioecious species *Asparagus officinalis* L. *Plant Molecular Biology*, *51*, 867-875.

Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., & Poethig, R. S. (2005). Nuclear processing and export of microRNAs in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 3691-3696.

Park, W., Li, J., Song, R., Messing, J., & Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology*, *12*, 1484-1495.

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*, 1061-1067.

Pasquinelli, A. E. (2012). MicroRNAs and their targets: Recognition, regulation and an emerging reciprocal relationship. *Nature Reviews Genetics*, *13*(4), 271-282.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., … Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, *457*, 29.

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., & Yanofsky, M. F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, *405*, 200-203.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., … Quackenbush, J. (2003). TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics*, *19*, 651-652.

Pillai, R. S. (2005). MicroRNA function: Multiple mechanisms for a tiny RNA? *RNA*, *11*(12), 1753-1761.

Pootakham, W., Sonthirod, C., Naktang, C., Jomchai, N., Sangsrakru, D., & Tangphatsornruang, S. (2016). Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: A case study in oil palm (*Elaeis guineensis*). *Molecular Breeding, 36*, 154.

Prabu, G. R., & Mandal, A. K. A. (2010). Computational identification of miRNAs and their target genes from expressed sequence tags of tea (*Camellia sinensis*). *Genomics Proteomics Bioinformatics, 8*(2), 113-121.

Puspadevi, S. (2017, March 9). Global palm oil output forecast to increase by 6 million tonnes. *The Star*. Retrieved from http://www.thestar.com.my

Qi, Y., Guo, H., Li, K., & Liu, W. (2012). Comprehensive analysis of differential genes and miRNA profiles for discovery of topping-responsive genes in flue-cured tobacco roots. *FEBS Journal, 279*(6), 1054-1070.

Qiu, C. X., Xie, F. L., Zhu, Y. Y., Guo, K., Huang, S. Q., Nie, L., & Yang, Z. M. (2007). Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene, 395*, 49-61.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., … Gu, Y. (2012). A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics, 13*, 341.

Rajagopalan, R., Vaucheret, H., Trejo, J., & Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Development, 20*, 3407-3425.

Rajanaidu, N., Rao, V., Abdul Halim, H., & Ong, A. S. H. (1989). Genetic resources: New developments in oil palm breeding. *Elaeis, 1*, 1-10.

Ramachandran, V., & Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science, 321*, 1490-1492.

Rawsthorne, S. (2002). Carbon flux and fatty acid synthesis in plants. *Progress in Lipid Research, 41*(2), 182-196.

Reddy, A. S. N., Marquez, Y., Kalyna, M., & Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell, 25*, 3657-3683.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., & Bartel, D. P. (2002). MicroRNAs in plants. *Genes & Development*, *16*(13), 1616-1626.

Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278-289.

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., & Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell*, *110*(4), 513-520.

Rival, A., Aberlenc-Bertossi, F., Morcillo, F., Tregear, J., Verdeil, J. -L., & Duval, Y. (1999). Scaling-up in micropropagation of palms: The example of oil palm. In C. Oropeza, J. -L. Verdeil, G. R. Ashburner, R. Cardena & J. M. Santamaria (Eds.), *Current advances in coconut biotechnology* (pp. 407-418). Dordrecht, Netherlands: Kluwer Academic Publishers.

Rival, A., Jaligot, E., Beulé, T., & Finnegan, E. J. (2008). Isolation and expression analysis of genes encoding MET, CMT, and DRM methyltransferases in oil palm (*Elaeis guineensis* Jacq.) in relation to the 'mantled' somaclonal variation. *Journal of Experimental Botany*, *59*(12), 3271-3281.

Rival, A., Tregear, J., Verdeil, J. –L., Richaud, F., Beulé, T., Hartman, C., … Duval, Y. (1998). Molecular search for mRNA and genomic markers of the oil palm "mantled" somaclonal variation. *Acta Horticulturae*, *461*, 165-171.

Roberdi, Sobir, Yahya, S., & Toruan-Mathius, N. (2015). Differential analysis of gene related to hard bunch phenomena in oil palm (*Elaeis guineensis* Jacq.) fruits. *Emirates Journal of Food and Agriculture*, *27*(8), 596-601.

Rocheta, M., Sobral, R., Magalhães, J., Amorim, M. I., Ribeiro, T., Pinheiro, M., … Costa, M. M. R. (2014). Comparative transcriptomic analysis of male and female flowers of monoecious *Quercus suber*. *Frontiers in Plant Science*, *5*, 599.

Rogers, K., & Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell*, *25*(7), 2383-2399.

Rolland, F., Baena-Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: Conserved and novel mechanisms. *Annual Review of Plant Biology*, *57*, 675-709.

Ruan, M. B., Zhao, Y. T., Meng, Z. H., Wang, X. J., & Yang, W. C. (2009). Conserved miRNA analysis in *Gossypium hirsutum* through small RNA sequencing. *Genomics*, *94*(4), 263-268.

Rubiopińa, J. A., & Vázquezflota, F. A. (2008). Isolation of functional total RNA from *Argemone mexicana* tissues. *Electronic Journal of Biotechnology*, *11*(4).

Ruiz-Ferrer, V., & Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology*, *60*, 485-510.

Rumble, S. M., Lacroute, P., Dalca, A. V., Fiume, M., Sidow, A., & Brudno, M. (2009). SHRiMP: Accurate mapping of short color-space reads. *PLoS Computational Biology*, *5*, e1000386.

Rusinov, V., Baev, V., Minkov, I. N., & Tabler, M. (2005). MicroInspector: A web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Research*, *33*(suppl2), W696-W700.

Sanchez Mde, L., Costas, C., Sequeira-Mendes, J., & Gutierrez, C. (2012). Regulating DNA replication in plants. *Cold Spring Harbor Perspectives in Biology*, *4*(12).

Schmitz, R. J., Schultz, M. D., Urich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., … Ecker, J. R. (2013). Patterns of population epigenomic diversity. *Nature*, *495*(7440), 193-198.

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., … Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, *22*(6), 839-851.

Schwarz, S., Grande, A. V., Bujdoso, N., Saedler, H., & Huijser, P. (2008). The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in *Arabidopsis*. *Plant Molecular Biology*, *67*, 183-195.

Shi, H., & Bressan, R. (2006). RNA extraction. In J. Salinas & J. J. Sánchez-Serrano (Eds.), *Arabidopsis protocols* (pp. 345). New York, NY: Humana Press.

Shamimuzzaman, M., & Vodkin, L. (2012). Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. *BMC Genomics*, *13*, 310.

Shankar, R., Bhattacharjee, A., & Jain, M. (2016). Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Scientific Reports*, *6*, 23719.

Shearman, J. R., Jantasuriyarat, C., Sangsrakru, D., Yoocha, T., Vannavichit, A., Tragoonrung, S., & Tangphatsornruang, S. (2013). Transcriptome analysis of normal and mantled developing oil palm flower and fruit. *Genomics*, *101*, 306-312.

Sheng, L., Chai, W., Gong, X., Zhou, L., Cai, R., Li, X., … Cheng, B. (2015). Identification and characterization of novel maize miRNAs involved in different genetic background. *International Journal of Biological Sciences*, *11*(7), 781-793.

Sheppard, L. A., Brunner, A. M., Krutovskii, K. V., Rottmann, W. H., Skinner, J. S., Vollmer, S. S., & Strauss, S. H. (2000). A *DEFICIENS* homolog from the dioecious tree black cottonwood is expressed in female and male floral meristems of the two-whorled, unisexual flowers. *Plant Physiology*, *124*, 627-640.

Shu, J., Xia, Z., Li, L., Liang, E. T., Slipek, N., Shen, D., … Steer, C. J. (2012). Dose-dependent differential mRNA target selection and regulation by let-7a-7f and miR-17-92 cluster microRNAs. *RNA Biology*, *9*, 1275-1287.

Singh, R., Ong-Abdullah, M., Low, E. T., Manaf, M. A., Rosli, R., Nookiah, R., … Sambanthamurthi, R. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, *500*, 335-9.

Singh, R., Tan, S. G., Panandam, J. M., Rahman, R. A., Ooi, L. C., Low, E. T., … Cheah, S. C. (2009). Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. *BMC Plant Biology*, *9*, 114.

SIRIM Berhad. (2015). Malaysia's bioplastics transformation. Selangor, Malaysia: Author.

Skrypina, N. A., Timofeeva, A. V., Khaspekov, G. L., Savochkina, L. P., & Beabealashvilli, R. Sh. (2003). Total RNA suitable for molecular biology analysis. *Journal of Biotechnology*, *105*(1-2), 1-9.

Somyong, S., Poopear, S., Sunner, S. K., Wanlayaporn, K., Jomchai, N., Yoocha, T., … Tragoonrung, S. (2016). Acc oxidase and miRNA 159a, and their involvement in fresh fruit bunch yield (FFB) via sex ratio determination in oil palm. *Molecular Genetics and Genomics*, 2*91*(3), 1243-1257.

Song, C., Zhang, D., Zheng, L., Zhang, J., Zhang, B., Luo, W., … Han, M. (2017). miRNA and degradome sequencing reveal miRNA and target genes that may mediate shoot growth in 'Fuji Nagafu 2' apple and its bud mutant 'Yanfu 6'. *Frontiers in Plant Science*, *8*, 441.

Song, G., Zhang, R., Zhang, S., Li, Y., Gao, J., Han, X., … Li, G. (2016). Response of microRNAs to cold treatment in the young spikes of common wheat. *BMC Genomics*, *18*, 212.

Song, L., Han, M. -H., Lesicka, J., & Fedoroff, N. (2007). *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 5437-5442.

Song, Q., Liu, Y., Hu, X., Zhang, W., Ma, B., Chen, S., & Zhang, J. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology*, *11*, 5.

Song, Y., Ma, K., Ci, D., Zhang, Z., & Zhang, D. (2013). Sexual dimorphism floral microRNA profiling and target gene expression in andromonoecious Poplar (*Populus tomentosa*). *PLoS ONE*, *8*(5), e62681.

Song, Y., Tian, M., Ci, D., & Zhang, D. (2015). Methylation of microRNA genes regulates gene expression in bisexual flower development in andromonoecious poplar. *Journal of Experimental Botany*, *66*, 1891-1905.

Song, Y., Zhang, Y., Liu, J., Wang, C., Liu, M., Feng, S., & Chen, X. (2012). Comparison of gibberellin acid content and the genes relatived to GA biosynthesis between 'Changfu 2' apple (*Malus domestica* Borkh.) and its spur sport. *Scientia Agricultura Sinica*, *45*, 2668-2675.

Spanudakis, E., & Jackson, S. (2014). The role of microRNAs in the control of flowering time. *Journal of Experimental Botany*, *65*(2), 365-380.

Sparnaaij, L. D., Rees, A. R., & Chapas, L. C. (1963). Annual yield variation in the oil palm. *Journal of West African Institute for Oil Palm Research*, *4*, 111-125.

Srivastava, P. K., Moturu, T. R., Pandey, P., Baldwin, I. T., & Pandey, S. P. (2014). A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, *15*, 348.

Srivastava, R. B., & Vishnoi, T. (2012). Biodegradable polymers and polymeric composites. In A. Tiwari & R. B. Srivastava (Eds.), *Biotechnology in biopolymers: Developments, applications & challenging areas* (pp. 179). Shropshire, UK: Smithers Rapra Technology Ltd.

Starega-Roslan, J., Galka-Marciniak, P., & Krzyzosiak, W. J. (2015). Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer. *Nucleic Acids Research*, *43*(22), 10939-10951.

Stocks, M. B., Moxon, S., Mapleson, D., Woolfenden, H. C., Mohorianu, I., Folkes, L., … Moulton, V. (2012). The UEA sRNA workbench: A suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*, *28*(15), 2059-2061.

Subramaniam, V., Choo, Y. M., Muhammad, H., Zulkifli, H., Tan, Y. A., & Puah, C. W. (2010). Life cycle assessment of the production of crude palm oil (part 3). *Journal of Oil Palm Research*, *22*, 895-903.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W. B., Zhu, J. K., & Yu, O. (2008). Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics*, *9*, 160.

Sulaiman, O., Hashim, R., Wahab, R., Samsi, H. W., & Mohamed, A. H. (2008). Evaluation on some finishing properties of oil palm plywood. *Holz als Roh- und Werkstoff*, *66*, 5-10.

Sun, F., Guo, G., Du, J., Guo, W., Peng, H., Ni, Z., … Yao, Y. (2014). Whole-genome discovery of miRNAs and their targets in wheat (*Triticum aestivum* L.). *BMC Plant Biology*, *14*, 142.

Sun, Y., Mui, Z., Liu, X., Yim, A. K., Qin, H., Wong, F., … Lim, B. L. (2016). Comparison of small RNA profiles of *Glycine max* and *Glycine soja* at early developmental stages. *International Journal of Molecular Sciences*, *17*(12), 2043.

Sun, Y. H., Lu, S., Shi, R., & Chiang, V. (2011). Computational prediction of plant miRNA targets. *Methods in Molecular Biology*, *744*, 175-186.

Sunkar, R., Girke, T., Jain, P. K., & Zhu, J. K. (2005). Cloning and characterization of microRNAs from rice. *Plant Cell*, *17*, 1397-1411.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., & Zhu, J. K. (2008). Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biology*, *8*, 25.

Suzuki, S., Ono, N., Furusawa, C., Ying, B., & Yomo, T. (2011). Comparison of sequence reads obtained from three next-generation sequencing platforms. *PLoS ONE*, *6*(5), e19534.

Syed, N. H., Kalyna, M., Marguez, Y., Barta, A., & Brown, J. W. (2012). Alternative splicing in plants--coming of age. *Trends in Plant Sciences*, *17*(10), 616-623.

Syed Alwee, S., Van Der Linden, C. G., Van Der Schoot, J., De Folter, S., Angenent, G. C., Cheah, S. C., & Smulders, M. J. M. (2006). Characterization of oil palm MADS box genes in relation to the mantled flower abnormality. *Plant Cell Tissue and Organ Culture*, *85*, 331-344.

Szittya, G., Moxon, S., Santos, D. M., Jing, R., Fevereiro, M. P., Moulton, V., & Dalmay, T. (2008). High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. *BMC Genomics*, *9*, 593.

Tang, S., Wang, Y., Li, Z., Gui, Y., Xiao, B., Xie, J., … Fan, L. (2012b). Identification of wounding and topping responsive small RNAs in tobacco (*Nicotiana tabacum*). *BMC Plant Biology*, *12*, 28.

Tang, Z., Zhang, L., Xu, C., Yuan, S., Zhang, F., Zheng, Y., & Zhao, C. (2012a). Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. *Plant Physiology*, *159*(2), 721-738.

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Münster, T., … Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Molecular Biology*, *42*, 115-149.

Thieme, C. J., Gramzow, L., Lobbes, D., & Theißen, G. (2011). SplamiR-prediction of spliced miRNAs in plants. *Bioinformatics*, *27*(9), 1215-1223.

Tian, T., You, Q., Zhang, L., Yi, X., Yan, H., Xu, W., & Su, Z. (2016). SorghumFDB: Sorghum functional genomics database with multidimensional network analysis. *Database* (Oxford), *2016*.

Ting, N. C., Mohd Zaki, N., Rosli, R., Low, E. L., Ithnin, M., Cheah, S. C., … Singh, R. (2010). SSR mining in oil palm EST database: Application in oil palm germplasm diversity studies. *Journal of Genetics*, *89*(2), 135-145.

Toyooka, K., Goto, Y., Asatsuma, S., Koizumi, M., Mitsui, T., & Matsuoka, K. (2009). A mobile secretory vesicle cluster involved in mass transport from the Golgi to the plant cell exterior. *Plant Cell*, *21*(4), 1212-1229.

Tranbarger, T. J., Dussert, S., Joët, T., Argout, X., Summo, M., Champion, A., … Morcillo, F. (2011). Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiology*, *156*(2), 564-584.

Tranbarger, T. J., Kluabmongkol, W., Sangsrakru, D., Morcillo, F., Tregear, W. J., Tragoonrung, S., & Billotte, N. (2012). SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*. *BMC Plant Biology*, *12*, 1.

Treangen, T. J., & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36-46.

Ummat, A., & Bashir, A. (2014). Resolving complex tandem repeats with long reads. *Bioinformatics*, *30*(24), 3491-3498.

United States, Department of Agriculture. (2015). Oilseeds: World markets and trade. Washington, USA: Author.

United States, Department of Agriculture. (2016). Oilseeds: World markets and trade. Washington, USA: Author.

Urasaki, N., Tarora, K., Shudo, A., Ueno, H., Tamaki, M., Miyagi, N., & Adaniya, S. (2012). Digital transcriptome analysis of putative sex-determination genes in papaya (*Carica papaya*). *PLoS ONE*, *7*(7), e40904.

Uthaipaisanwong, P., Chanprasert, J., Shearman, J. R., Sangsrakru, D., Yoocha, T., Jomchai, N., … Tangphatsornruang, S. (2012). Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). *Gene*, *500*(2), 172-180.

Van Verk, M. C., Hickman, R., Pieterse, C. M. J., & Van Wees, S. C. M. (2013). RNA-Seq: Revelation of the messengers. *Trends in Plant Sciences*, *18*(4), 175-179.

Vaucheret, H., Vazquez, F., Crété, P., & Bartel, D. P. (2004). The action of *ARGONAUTE1* in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes & Development*, *18*, 1187-1197.

Venkov, P. V., & Hadjiolov, A. A. (1969). Differential stability of 28s and 18s rat liver ribosomal ribonucleic acids. *Biochemical Journal*, *115*, 91.

Verheye, W. H. (2002). Growth and production of oil palm. In W. H. Verheye (Ed.), *Soils, plants and crop production* (Volume 2). Eolss Publishers. Retrieved from http://www.eolss.net/sample-chapters/c10/e1-05a-27-00.pdf

Vishwakarma, N. P., & Jadeja, V. J. (2013). Identification of miRNA encoded by *Jatropha curcas* from EST and GSS. *Plant Signaling & Behavior*, *8*(2), e23152.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell*, *136*, 669-687.

Wahid, M. B., Abdullah, S. N. A., & Henson, I. E. (2004). Oil palm− achievements and potential. *Proceedings of the 4th International Crop Science Congress*, Brisbane, Australia, 26 Sep-1 Oct 2004. Brisbane, Australia: The Regional Institute Ltd.

Wan, L. C., Wang, F., Guo, X. Q., Lu, S. F., Qiu, Z. B., Zhao, Y. Y., … Lin, J. X. (2012). Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. *BMC Plant Biology*, *12*, 146.

Wan, P., Wu, J., Zhou, Y., Xiao, J., Feng, J., Zhao, W., … Chen, J. Y. (2011). Computational analysis of drought stress-associated miRNAs and miRNA co-regulation network in *Physcomitrella patens*. *Genomics Proteomics Bioinformatics*, *9*(1-2), 37-44.

Wang, C. M., Liu, P., Sun, F., Li, L., Liu, P., Ye, J., & Yue, G. H. (2012a). Isolation and identification of miRNAs in *Jatropha curcas*. *International Journal of Biological Sciences*, *8*(3), 418-429.

Wang, J. -F., Zhou, H., Chen, Y. -Q., Luo, Q. -J., & Qu, L. -H. (2004b). Identification of 20 new miRNAs from *Oryza sativa*. *Nucleic Acids Research*, *32*, 1688-1695.

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

Wang, S. Z., Pan, L., Hu, K., Chen, C. Y., &Ding, Y. (2010). Development and characterization of polymorphic microsatellite markers in *Momordica charantia* (Cucurbitaceae). *American Journal of Botany*, *97*, e75-e78.

Wang, T. Z., Chen, L., Zhao, M. G., Tian, Q. Y., & Zhang, W. H. (2011). Identification of drought-responsive microRNAs in *Medicago truncatula* by genome-wide highthroughput sequencing. *BMC Genomics*, *12*, 367.

Wang, X. J., Reyes, J. L., Chua, N. H., & Gaasterland, T. (2004a). Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biology*, *5*(9), R65.

Wang, Y., Li, L., Tang, S., Liu, J., Zhang, H., Zhi, H., … Diao, X. (2016b). Combined small RNA and degradome sequencing to identify miRNAs and their targets in response to drought in foxtail millet. *BMC Genetics*, *17*, 57.

Wang, Y., Zhou, L., Yu, X., Stover, E., Luo, F., & Duan, Y. (2016a). Transcriptome profiling of Huanglongbing (HLB) tolerant and susceptible *Citrus* plants reveals the role of basal resistance in HLB tolerance. *Frontiers in Plant Science*, *7*, 933.

Wang, Y. W., Li, P. C., Cao, X. F., Wang, X. J., Zhang, A. M., & Li, X. (2009a). Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. *Biochemical and Biophysical Research Communications*, *378*, 799-803.

Wang, Z., Gerstein, M., & Snyder, M. (2009b). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57-63.

Wang, Z., Huang, J., Sun, Z., & Zheng, B. (2015). Identification of microRNAs differentially expressed involved in male flower development. *Functional & Integrative Genomics*, *15*, 225-232.

Wang, Z. J., Huang, J. Q., Huang, Y. J., Li, Z., & Zheng, B. S. (2012b). Discovery and profiling of novel and conserved microRNAs during flower development in *Carya cathayensis* via deep sequencing. *Planta*, *236*, 613-621.

Waugh, A., Gendron, P., Altman, R., Brown, J. W., Case, D., Gautheret, D., … Major, F. (2002). RNAML: A standard syntax for exchanging RNA information. *RNA*, *8*, 707-717.

Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F., & Meyerowitz, E. M. (1992). LEAFY controls floral meristem identity in *Arabidopsis*. *Cell*, *69*, 843-859.

Wen, M., Lin, X., Xie, M., Wang, Y., Shen, X., Liufu, Z., … Tang, T. (2016). Small RNA transcriptomes of mangroves evolve adaptively in extreme environments. *Scientific Reports*, *6*, 27551.

Whiteford, N., Skelly, T., Curtis, C., Ritchie, M. E., Löhr, A., Zaranek, A. W., … Brown, C. (2009). Swift: Primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics*, *25*, 2194-2199.

Wightman, B., Ha, I., & Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell, 75*(5), 855-862.

Wilhelm, B. T., & Landry, J. R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods, 48*(3), 249-257.

Williams, L., Grigg, S. P., Xie, M., Christensen, S., & Fletcher, J. C. (2005). Regulation of *Arabidopsis* shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. *Development, 132*, 3657-3668.

Willson, M. F. (1979). Sexual selection in plants. *The American Naturalist, 113*, 777-790.

Wind, J., Smeekens, S., & Hanson, J. (2010). Sucrose: Metabolite and signaling molecule. *Phytochemistry, 71*, 1610-1614.

Wink, M. (1988). Plant breeding: Importance of plant secondary metabolites for protection against pathogens and herbivores. *Theoretical and Applied Genetics, 75*(2), 225-233.

Wonkyi-Appiah, J. B. (1987). Genetic control of fertility in the oil palm (*Elaeis guineensis* Jacq.). *Euphytica, 36*, 505-511.

Wright, P. R., Georg, J., Mann, M., Sorescu, D. A., Richter, A. S., Lott, S., … Backofen, R. (2014). CopraRNA and IntaRNA: Predicting small RNA targets, networks and interaction domains. *Nucleic Acids Research, 42*(Web Server issue), W119-W123.

Wu, G., Park, M. Y., Conway, S. R., Wang, J. W., Weigel, D., & Poethig, R. S. (2009b). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell, 138*, 750-759.

Wu, H., Ma, Y., Chen, T., Wang, M., & Wang, X. (2012). PsRobot: A web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Research, 40*(Web Server issue), W22-W28.

Wu, H. X., Jia, H. M., Ma, X. W., Wang, S. B., Yao, Q. S., Xu, W. T., … Zhan, R. L. (2014). Transcriptome and proteomic analysis of mango (*Mangifera indica* Linn) fruits. *Journal of Proteomics, 105*, 19-30.

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., & Qi, Y. (2009a). Rice microRNA effector complexes and targets. *Plant Cell*, *21*(11), 3421-3435.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., & Qi, Y. (2010). DNA methylation mediated by a microRNA pathway. *Molecular Cell*, *38*(3), 465-475.

Wu, X., Ding, D., Shi, C., Xue, Y., Zhang, Z., Tang, G., & Tang, J. (2016). microRNA-dependent gene regulatory networks in maize leaf senescence. *BMC Plant Biology*, *16*, 73.

Xia, R., Zhu, H., An, Y. Q., Beers, E. P., & Liu, Z. (2012). Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biology*, *13*, 1-18.

Xian, Z., Huang, W., Yang, Y., Tang, N., Zhang, C., Ren, M., & Li, Z. (2014). miR168 influences phase transition, leaf epinasty, and fruit development via SlAGO1 in tomato. *Journal of Experimental Botany*, *65*, 6655-6666.

Xiao, Y., Xia, W., Yang, Y., Mason, A. S., Lei, X., & Ma, Z. (2013). Characterization and evolution of conserved microRNA through duplication events in date palm (*Phoenix dactylifera*). *PLoS ONE*, *8*(8), e71435.

Xiao, Y., Yang, Y., Cao, H., Fan, H., Ma, Z., Lei, X., … Huang, X. (2012). Efficient isolation of high quality RNA from tropical palms for RNA-seq analysis. *Plant Omics*, *5*, 584-558.

Xiao, Y., Zhou, L., Xia, W., Mason, A. S., Yang, Y., Ma, Z., & Peng, M. (2014). Exploiting transcriptome data for the development and characterization of gene-based SSR markers related to cold tolerance in oil palm (*Elaeis guineensis*). *BMC Plant Biology*, *14*, 384.

Xie, F., & Zhang, B. (2010). Target-align: A tool for plant microRNA target identification. *Bioinformatics*, *26*(23), 3002-3003.

Xie, F. L., Frazier, T. P., & Zhang, B. H. (2011). Identification, characterization and expression analysis of microRNAs and their targets in the potato (*Solanum tuberosum*). *Gene*, *473*, 8-22.

Xie, F. L., Huang, S. Q., Guo, K., Xiang, A. L., Zhu, Y. Y., Nie, L., & Yang, Z. M. (2007). Computational identification of novel microRNAs and targets in *Brassica napus*. *FEBS Letters*, *581*, 1464-1474.

Xie, K., Shen, J., Hou, X., Yao, J., Li, X., Xiao, J., & Xiong, L. (2012). Gradual increase of miR156 regulates temporal expression changes of numerous genes during leaf development in rice. *Plant Physiology*, *158*(3), 1382-1394.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., & Carrington, J. C. (2005). Expression of Arabidopsis MIRNA genes. *Plant Physiology*, *138*(4), 2145-2154.

Xin, C., Liu, W., Lin, Q., Zhang, X., Cui, P., Li, F., … Yu, J. (2015a). Profiling microRNA expression during multi-staged date palm (*Phoenix dactylifera* L.) fruit development. *Genomics*, *105*(4), 242-251.

Xin, M., Yang, G., Yao, Y., Peng, H., Hu, Z., Sun, Q., … Ni, Z. (2015b). Temporal small RNA transcriptome profiling unraveled partitioned miRNA expression in developing maize endosperms between reciprocal crosses. *Frontiers in Plant Science*, *6*, 744.

Xing, L., Zhang, D., Li, Y., Zhao, C., Zhang, S., Shen, Y., … Han, M. (2014). Genome-wide identification of vegetative phase transition-associated microRNAs and target predictions using degradome sequencing in *Malus hupehensis*. *BMC Genomics*, *15*, 1125.

Xing, S., Salinas, M., Garcia-Molina, A., Höhmann, S., Berndtgen, R., & Huijser, P. (2013). SPL8 and miR156-targeted SPL genes redundantly regulate Arabidopsis gynoecium differential patterning. *The Plant Journal*, *75*(4), 566-577.

Xu, D., Long, H., Liang, J., Zhang, J., Chen, X., Li, J., … Yu, M. (2012b). *De novo* assembly and characterization of the root transcriptome of *Aegilops variabilis* during an interaction with the cereal cyst nematode. *BMC Genomics*, *13*, 133.

Xu, J., Li, Y., Wang, Y., Liu, X., & Zhu, X. (2017). Altered expression profiles of microRNA families during de-etiolation of maize and rice leaves. *BMC Research Notes*, *10*, 108.

Xu, M. Y., Dong, Y., Zhang, Q. X., Zhang, L., Luo, Y. Z., Sun, J., … Wang, L. (2012a). Identification of miRNAs and their targets from *Brassica napus* by high-throughput sequencing and degradome analysis. *BMC Genomics*, *13*, 421.

Xu, S., Liu, N., Mao, W., Hu, Q., Wang, G., & Gong, Y. (2016). Identification of chilling-responsive microRNAs and their targets in vegetable soybean (*Glycine max* L.). *Scientific Reports*, *6*, 26619.

Yaish, M. W., Sunkar, R., Zheng, Y., Ji, B., Al-Yahyai, R., & Farooq, S. A. (2015). A genome-wide identification of the miRNAome in response to salinity stress in date palm (*Phoenix dactylifera* L.). *Frontiers in Plant Science, 6*, 946.

Yang, J., Li, J., Shao, P., Zhou, H., Chen, Y., & Qu, L. (2010). starBase: A database for exploring microRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research, 39*(Database issue), D202-9.

Yang, L., Liu, Z., Lu, F., Dong, A., & Huang, H. (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *The Plant Journal, 47*, 841-850.

Yang, X., & Li, L. (2012). Analyzing the microRNA transcriptome in plants using deep sequencing data. *Biology, 1*, 297-310.

Yang, X., Wang, L., Yuan, D., Lindsey, K., & Zhang, X. (2013). Small RNA and degradome sequencing reveal complex miRNA regulation during cotton somatic embryogenesis. *Journal of Experimental Botany, 64*(6), 1521-1536.

Yang, X., Zhang, H., & Li, L. (2011). Global analysis of gene-level microRNA expression in *Arabidopsis* using deep sequencing data. *Genomics, 98*(1), 40-46.

Yao, Y., Guo, G., Ni, Z., Sunkar, R., Du, J., Zhu, J. K., & Sun, Q. (2007). Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biology, 8*, R96.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., … Wang, J. (2006). WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research, 34*(Web Server issue), W293-7.

Yeap, W. C., Loo, J. M., Wong, Y. C., & Kulaveerasingam, H. (2013). Evaluation of suitable reference genes for qRT-PCR gene expression normalization in reproductive, vegetative tissues and during fruit development in oil palm. *Plant Cell, Tissue and Organ Culture, 116*(1), 55-66.

Yew, C. W., & Kumar, S. V. (2012). Isolation and cloning of microRNAs from recalcitrant plant tissues with small amounts of total RNA: A step-by step approach. *Molecular Biology Reports, 39*(2), 1783-1790.

Yin, Z., & Shen, F. (2010). Identification and characterization of conserved microRNAs and their target genes in wheat (*Triticum aestivum*). *Genetics and Molecular Research, 9*, 1186-1196.

Yin, Z. J., Li, C. H., Han, X. L., & Shen, F. F. (2008). Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). *Gene, 414*, 60-66.

Yoo, S. K., Chung, K. S., Kim, J., Lee, J. H., Hong, S. M., Yoo, S. J., … Ahn, J. H. (2005). CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in *Arabidopsis*. *Plant Physiology, 139*, 770-778.

Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., … Chen, X. (2008). The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 10073-10078.

Yu, B., & Wang, H. (2010). Translational inhibition by microRNAs in plants. *Progress in Molecular and Subcellular Biology, 50*, 41-57.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, … Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science, 307*, 932-935.

Yue, J., Li, J., Wang, D., Araki, H., Tian, D., & Yang, S. (2010). Genome-wide investigation reveals high evolutionary rates in annual model plants. *BMC Plant Biology, 10*, 242.

Zahn, L. M., Leenbens-Mack, J., dePamphilis, C. W., Ma, H., & Theissen, G. (2005). To B or not to B a flower: The role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. *Journal of Heredity, 96*, 225-240.

Zakaria, M. R., Ariffin, H., Johar, N. A. M., Abd-Aziz, S., Nishida, H., Shirai, Y., & Hassan, M. A. (2010a). Biosynthesis and characterization of poly (3-hydroxybutyrate-co-3-hydroxyvalerate) copolymer from wild-type *Comamonas* sp. EB172. *Polymer Degradation and Stability, 95*, 1382-1386.

Zakaria, M. R., Tabatabaei, M., Mohamad Ghazali, F., Abd-Aziz, S., Shirai, Y., & Hassan, M. A. (2010b). Polyhydroxyalkanoate production from anaerobically treated palm oil mill effluent by new bacterial strain *Comamonas* sp. EB172. *World Journal of Microbiology and Biotechnology*, *26*(5), 767-774.

Zhang, B., Pan, X., Cobb, G. P., & Anderson, T. A. (2006b). Plant microRNA: A small regulatory molecule with big impact. *Developmental Biology, 289*, 3-16.

Zhang, B., Pan, X., Wang, Q., Cobb, G. P., & Anderson, T. A. (2006c). Computational identification of microRNAs and their targets. *Computational Biology and Chemistry*, *30*(6), 395-407.

Zhang, B. H., Pan, X. P., & Anderson, T. A. (2006e). Identification of 188 conserved maize microRNAs and their targets. *FEBS Letters*, *580*, 3753-3762.

Zhang, B. H., Pan, X. P., Cannon, C. H., Cobb, G. P., & Anderson, T. A. (2006a). Conservation and divergence of plant microRNA genes. *The Plant Journal*, *46*, 243-259.

Zhang, B. H., Pan, X. P., Cox, S. B., Cobb, G. P., & Anderson, T. A. (2006d). Evidence that miRNAs are different from other RNAs. *Cellular and Molecular Life Sciences*, *63*(2), 246-254.

Zhang, B. H., Pan, X. P., Wang, Q. L., Cobb, G. P., & Anderson, T. A. (2005). Identification and characterization of new plant microRNAs using EST analysis. *Cell Research*, *15*, 336-360.

Zhang, B. H., Wang, Q. L., Wang, K. B., Pan, X. P., Liu, F., Guo, T. L., ... Anderson, T. A. (2007). Identification of cotton microRNAs and their targets. *Gene*, *397*, 26-37.

Zhang, J., Liu, J., & Ming, R. (2014). Genomic analyses of the CAM plant pineapple. *Journal of Experimental Botany*, *65*(13), 3395-3404.

Zhang, J. G., Zeng, R., Chen, J. S., Liu, X., & Liao, Q. S. (2008). Identification of conserved microRNAs and their targets from *Solanum lycopersicum* Mill. *Gene*, *423*, 1-7.

Zhang, L., Chia, J. M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., ... Ware, D. (2009b). A genome-wide characterization of microRNA genes in maize. *PLoS Genetics*, *5*(11), e1000716.

Zhang, L., Zheng, Y., Jagadeeswaran, G., Li, Y. F., Gowdu, K., & Sunkar, R. (2011). Identification and temporal expression analysis of conserved and novel microRNAs in *Sorghum*. *Genomics*, *98*, 460-468.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., … Chen, Z. J. (2015a). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*, *33*, 531-537.

Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., … Jin, H. (2010a). Multiple distinct small RNAs originate from the same microRNA precursors. *Genome Biology*, *11*, R81.

Zhang, W., Luo, Y., Gong, X., Zeng, W., & Li, S. (2009a). Computational identification of 48 potato microRNAs and their targets. *Computational Biology and Chemistry*, *33*, 84-93.

Zhang, Y. (2005). miRU: An automated plant miRNA target prediction server. *Nucleic Acids Research*, *33*, W701-W704.

Zhang, Y., Zhang, S., & Ling, L. (2015c). *De novo* transcriptome analysis to identify flavonoid biosynthesis genes in *Stellera chamaejasme*. *Plant Gene*, *4*, 64-68.

Zhang, Z., Jiang, L., Wang, J., Gu, P., & Chen, M. (2015b). MTide: An integrated tool for the identification of miRNA-target interaction in plants. *Bioinformatics*, *31*, 290–291.

Zhang, Z., Yu, J., Li, D., Liu, F., Zhou, X., Wang, T., … Su, Z. (2010b). PMRD: Plant microRNA database. *Nucleic Acids Research*, *38*, D806-13.

Zhao, C. Z., Xia, H., Frazier, T. P., Yao, Y. Y., Bi, Y. P., Li, A. Q., ... Wang, X. J. (2010). Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). *BMC Plant Biology*, *10*, 3.

Zheng, Y., Li, Y., Sunkar, R., & Zhang, W. (2011). SeqTar: An effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Research*, *40*, e28.

Zhou, J., Cheng, Y., Yin, M., Yang, E., Gong, W., Liu, C., … Ren, Z. (2015). Identification of novel miRNAs and miRNA expression profiling in wheat hybrid necrosis. *PLoS ONE*, *10*(2), e0117507.

Zhou, M., Gu, L., Li, P., Song, X., Wei, L., Chen, Z., & Cao, X. (2010). Degradome sequencing reveals endogenous small RNA targets in rice (*Oryza sativa* L. ssp. indica). *Frontiers in Biology*, *5*(1), 67-90.

Zhou, R., Wang, Q., Jiang, F., Cao, X., Sun, M., Liu, M., & Wu, Z. (2016a). Identification of miRNAs and their targets in wild tomato at moderately and acutely elevated temperatures by high-throughput sequencing and degradome analysis. *Scientific Reports*, *6*, 33777.

Zhou, Y., Xu, Z., Duan, C., Chen, Y., Meng, Q., Wu, J., … Li, X. (2016b). Dual transcriptome analysis reveals insights into the response to *Rice black-streaked dwarf virus* in maize. *Journal of Experimental Botany*, *67*(15), 4593-4609.

Zhu, G., Wu, A., Xu, X. J., Xiao, P. P., Lu, L., Liu, J., … Zhao, X. M. (2016). PPIM: A protein-protein interaction database for maize. *Plant Physiology*, *170*(2), 618-626.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S. H., Liou, L. W., … Zhang, X. (2011). *Arabidopsis* Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell*, *145*(2), 242-256.

Zhu, Q., Upadhyaya, N. M., Gubler, F., & Helliwell, C. A. (2009). Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (*Oryza sativa*). *BMC Plant Biology*, *9*, 149.

Zhu, Q. H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., & Helliwell, C. (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Research*, *18*, 1456-1465.

Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y., & Robb, G. B. (2012). Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. *Nucleic Acids Research*, *40*(7), e54.

Zhuang, W., Gao, Z., Wen, L., Huo, X., Cai, B., & Zhang, Z. (2015). Metabolic changes upon flower bud break in Japanese apricot are enhanced by exogenous GA_4. *Horticulture Research*, *2*, 15046.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406-3415.

Zuker, M., & Jacobson, A. B. (1998). Using reliability information to annotate RNA secondary structures. *RNA*, *4*(6), 669-679.

Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, *9*(1), 133-148.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Journal Papers (ISI-cited)

Ho, H., Low, J. Z., Gudimella, R., Tammi, M. T., & Harikrishna, J. A. (2016). Expression patterns of inflorescence- and sex-specific transcripts in male and female inflorescences of African oil palm (*Elaeis guineensis*). *Annals of Applied Biology*, *168*(2), 274-289.

Ho, H., Gudimella, R., Ong-Abdullah, M., & Harikrishna, J. A. (2017). Expression of microRNAs during female inflorescence development in African oil palm (*Elaeis guineensis* Jacq.). *Tree Genetics & Genomes*, *13*, 35.

Conference Papers

Ho, H. L., Loong, H. Y., Md Nasaruddin, N., & Harikrishna, J. A. (2008, June). *Study of microRNAs in floral development of oil palm (Elaeis guineensis Jacq.)*. Poster presented at the 17th Malaysian Society of Molecular Biology & Biotechnology (MSMBB) Scientific Meeting, Kuala Lumpur, Malaysia.

Ho, H., Othman, R. Y., Ong-Abdullah, M., & Harikrishna, J. A. (2009, August). *Study of miRNA 156 in oil palm (Elaeis guineensis Jacq.)*. Poster presented at the 18th Malaysian Society of Molecular Biology & Biotechnology (MSMBB) Scientific Meeting, Kuala Lumpur, Malaysia.

Ho, H. L., Othman, R. Y., & Harikrishna, J. A. (2010, November). *High-throughput sequencing of microRNAs associated with maturation of oil palm female flowers*. Poster presented at the Asian Regional Conference on Systems Biology, Kuala Lumpur, Malaysia.

Ho, H., & Harikrishna, J. A. (2011, December). *Thirty-three orthologous microRNA families are differentially expressed in flowers of oil palm (Elaeis guineensis Jacq.)*. Poster presented at the 16thBiological Sciences Graduate Congress (BSGC), National University of Singapore, Singapore.

Ho, H., Othman, R. Y., Ong-Abdullah, M., & Harikrishna, J. A. (2012, April). *Thirty-three differentially expressed orthologous microRNA families in male and female flowers of oil palm (Elaeis guineensis Jacq.)*. Poster presented at the University of Malaya Researchers' Conference, Kuala Lumpur, Malaysia.

Ho, H. L., Othman, R. Y., & Harikrishna, J. A. (2012, November). *Identification of novel transcripts of Elaeis guineensis Jacq. (African oil palm) from transcriptomes of immature and mature female flowers*. Poster presented at the 19thMalaysian Society of Molecular Biology & Biotechnology (MSMBB) Scientific Meeting, Kuala Lumpur, Malaysia.

Ho, H. L., Gudimella, R., Low, J. Z. B., Tammi, M., & Harikrishna, J. A. (2013, June). *Identification of novel transcripts of Elaeis guineensis Jacq. (African oil palm) from the transcriptomes of female inflorescences, male inflorescences and mesocarps*. Paper presented at the 20thMalaysian Society of Molecular Biology & Biotechnology (MSMBB) Scientific Meeting, Kuala Lumpur, Malaysia.

Ho, H. L., Gudimella, R., Low, J. Z. B., Tammi, M., & Harikrishna, J. A. (2013, November). *Comparative transcriptional profiles of flower to fruit development in African oil palm (Elaeis guineensis Jacq.)*. Poster presented at the University of Malaya Researchers' Conference, Kuala Lumpur, Malaysia.