Journal of Oil Palm Research Vol. 32 (3) September 2020 p. 406-419
DOI: https://doi.org/10.21894/jopr.2020.0034

TRANSCRIPTOME ANALYSIS ACROSS VARIOUS MESOCARP DEVELOPMENTAL STAGES OF MPOB-ANGOLA *Dura*

PRISCILLA ELIZABETH MORRIS*; PEK-LAN CHAN*; LESLIE OOI CHENG-LI*; PEI-WEN ONG*; KATIALISA KAMARUDDIN*; MARHALIL MARJUNI*; MOHD DIN AMIRUDDIN*; KUANG-LIM CHAN*; ROZANA ROSLI*; LESLIE LOW ENG-TI*; RAJINDER SINGH*; NOOR AZMI SHAHARUDDIN; DENIS J MURPHY‡; MOHAMAD ARIF ABD MANAF* and SITI NOR AKMAR ABDULLAH‡‡**

ABSTRACT

Angolan germplasm materials are being evaluated for their potential use in oil palm breeding programmes. Large fruits characteristic of some Angola dura *palms have provided breeders with a source of favourable traits for introgression into the current planting materials. With the aim of understanding molecular regulation during mesocarp development, maturation and ripening of Angola* dura *palms, efforts were initiated to profile the expression of transcripts across nine different developmental stages of mesocarp tissues. A 36 675 total gene set was identified from ribonucleic acid (RNA) sequencing with 24 226 transcripts successfully annotated with the Plant Reference Sequence (RefSeq) Database using BLASTX, while 12 449 transcripts had no hit to any known genes. Pairwise T-test was performed using TIGR Multiexperiment Viewer and a total of 21 261 transcripts were identified as significant across all the pairwise comparisons. BLAST2GO analysis resulted in the annotation of 13 996 unigenes with various gene ontology (GO) terms. Transcripts associated with lipid metabolic process were highly expressed during lag phase preceding the lipid biosynthesis [10 to 12 weeks after anthesis (WAA)] and fruit maturation (18 to 20 WAA) stages. Further annotation of the unigenes with Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway resulted in the identification of 279, 757 and 142 transcripts related to lipid metabolism, carbohydrate metabolism and hormone metabolism, respectively. Detailed analysis of the expression data revealed that certain transcripts such as* KAS I, FatA, DGAT, *WRI1 and bZIP showed unique expression profiles in the MPOB-Angola* dura *as compared to the published data. The availability of these transcriptome datasets gives an insight into the transcriptional mechanisms controlling the Angolan* dura *fruit development, maturation and ripening.*

Keywords: oil palm, MPOB-Angola *dura*, mesocarp, fruit development, gene ontology.

Date received: 14 June 2019**; Sent for revision:** 18 June 2019**; Accepted:** 22 September 2019**; Available online:** 30 June 2020.

* Malaysian Palm Oil Board,
6 Persiaran Institusi, Bandar Baru Bangi,
43000 Kajang, Selangor, Malaysia.
E-mail: chanpl@mpob.gov.my

** Department of Biochemistry,
Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia.

‡ Genomics and Computational Biology Research Group,
University of South Wales, Pontypridd,
United Kingdom.

‡‡ Institute of Plantation Studies,
Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia.

INTRODUCTION

Oil palm Angolan germplasm materials were collected from the coastal areas of Angola, Africa in 1991 (Rajanaidu *et al*., 1991) and were planted at the Malaysian Palm Oil Board (MPOB) Research Station Kluang, Johor, Malaysia in 1994 (Kushairi *et al*., 2003). Angolan germplasm materials have been evaluated for their potential use in the oil palm breeding programmes to widen the genetic base of existing materials (Noh *et al*., 2002). It was reported that the fatty acid composition, iodine value and carotene content of some palms are comparable to the commercial *dura* x *pisifera* (DxP) materials,

although the variation observed for the traits is large. The availability of unsaturated oil with high oleic and carotene content will benefit the food and pharmaceutical industries (Kushairi *et al*., 2017; 2018). Angolan germplasm can be used for the genetic improvement of existing planting materials because of their favourable variation and heritability (moderate to high) for several important oil-related traits (Noh *et al*., 2002).

Advances in next-generation sequencing technologies have resulted in ribonucleic acid (RNA) sequencing (RNA-Seq) being widely used for transcriptome profiling. This approach provides more precise measurement of transcript levels and enables the discovery of gene isoforms as compared to other approaches such as microarrays and expressed sequence tags (EST) (Wang *et al*., 2009). The RNA-Seq was chosen to generate transcriptome data as it can overcome the limitations of hybridisation-based approaches, such as requirement for sequence information and the detection of higher level of background signals due to cross-hybridisation (Wang *et al.,* 2009; Okoniewski and Miller, 2006; Royce *et al.,* 2007). The EST method, on the other hand, is the least preferred due to high cost. Furthermore, a significant portion of the short tags cannot be uniquely mapped to the reference genome and isoforms cannot be distinguished from each other (Wang *et al*., 2009). In contrast to microarray and EST methods, the RNA-Seq analysis provides a wide dynamic range enabling the detection of more differentially expressed genes with higher fold change. RNA-Seq can also be used to investigate splice variants, gene isoforms, single nucleotide polymorphisms and post transcriptional modifications (Lalonde *et al*., 2011). Through 454 or Illumina sequencing, comprehensive transcriptome sequences were generated for several oil crops such as sesame (Wei *et al*., 2011), coconut (Fan *et al*., 2013), avocado (Kilaru *et al*., 2015) and oil palm (Tranbarger *et al*., 2011). These data proved useful for gene discovery and development of molecular markers.

Transcriptome analysis across oil palm mesocarp collected from *dura* palms of different origin was performed previously by Tranbarger *et al*. (2011). In this study, a 454 pyrosequencing-derived transcriptome was used to dissect the information on lipid and carotenoid biosynthesis pathways. Guerin *et al.* (2016) also detected tight transcriptional coordination of fatty acid biosynthesis (FAS) in the plastid with sugar sensing, plastidial glycolysis, transient starch storage and carbon recapture pathways via coexpression network. Recently, oil palm gene model was updated by detailed classification of the oil palm stearoyl-ACP desaturases (*SAD*) and acyl-acyl carrier protein (ACP) thioesterases (*FAT*) genes (Rosli *et al.,* 2018a). Transcriptome analysis across oil palm has enabled the discovery of various isoforms within the four families of diacylglycerol acyltransferase (*DGAT*) genes (Rosli *et al.,* 2018b). Transcript profiling of the *Elaeis guineensis DGAT* across various tissues and developmental stages showed that *DGAT* is important for flowering and fruit development in oil palm (Rosli *et al.,* 2018b).

Accumulation of up to 90% of oil in the mesocarp involves coordinated regulation of key FAS genes by WRINKLED1 (WRI1) transcription factor. Tranbarger *et al.* (2011) reported the identification of WRI1 with expression profile that correlated well with several FAS genes. The increase in the accumulation of lipid signifies common regulatory features between seeds and fruits. Although it is now confirmed that WRI1 regulates oil synthesis in both seed and non-seed tissues (Ma *et al.,* 2013; Singh *et al.,* 2013), the master regulators that control the expression of WRI1 in the oil palm mesocarp remain unclear. Common regulators in seed tissues such as LEAFY COTYLEDON (LEC1 and LEC2), ABSCISIC ACID INSENSITIVE3 (ABI3) and FUSCA3 (FUS3) genes were not detected. Bourgis *et al*. (2011) showed that the expression of a WRI1 ortholog was 57-fold up-regulated in oil palm mesocarp compared to date palm and the profile was similar to its target genes. In the work by Jin *et al.* (2017), they found that the ectopic expression of *EgWRI1-1* in *Arabidopsis* plants dramatically increased the seed mass and oil content. Their findings showed that *EgWRI1* is the key gene that contributes to the hybrid vigour of lipid biosynthesis trait in hybrid *tenera* fruit form. Apart from WRI1, two novel transcription factors (TF), termed NF-YB-1 and ZFP-1, were also found at the core of the FAS regulatory module in oil palm (Guerin *et al*., 2016).

In this study, transcript profiling was performed across the mesocarp tissues from MPOB-Angola *dura* from 5 to 24 weeks after anthesis (WAA), which includes nine different developmental stages. The availability of higher coverage of transcriptome datasets from nine different developmental stages of mesocarp provides improved insights into the transcriptional mechanisms controlling the *dura* fruit development, maturation and ripening. Identification of differentially expressed transcripts associated with lipid metabolism, transcription factors and hormone metabolism in the mesocarp of MPOB-Angola *dura* are also reported.

MATERIALS AND METHODS

Plant Materials

The second generation MPOB-Angola *dura* palms from Trial 0.481 and 0.482 were selected for this study. These palms were planted at the MPOB Research Station in Kluang, Johor. Fruits bunches were collected from these palms at nine different developmental stages, which consisted of 5, 8, 10, 12, 15, 18, 20, 22 and 24 WAA. For each developmental

stage, fruit bunches were harvested from two independent palms. However, for 24 WAA, sample was available for one palm. A total of 90 fruits from each bunch were randomly chosen to form three technical replicates, with 30 fruits representing one technical replicate. Mesocarp tissue from each fruit was immediately frozen in liquid nitrogen and stored at -80°C.

Total RNA Extraction, Purification and Quality Assessment

Total RNA extraction was performed using modified CTAB method with addition of phenol (Ong *et al*., 2019) on a total of 51 mesocarp tissues. The total RNA was purified using RNeasy Mini Kit and was subjected to the on-column DNase I digestion following the manufacturer's instruction (Qiagen, USA). The integrity of the purified total RNA was investigated with Agilent 2100 Bioanalyzer using a RNA Nano Labchip® (Agilent Technologies, CA, USA).

Library Construction, Sequence Generation and Quality Assessment

Transcriptome sequencing was performed for 51 total RNA samples from mesocarp. Truseq cDNA libraries were constructed and sequenced on Illumina HiSeq2000 sequencing platform following the manufacturer's instructions. A total of four paired-end cDNA libraries were sequenced on one lane of Illumina flow cell. The quality of the raw reads from Illumina sequencing was assessed using FastQC (Andrews, 2010). TruSeq Universal Adapters were removed using TagCleaner (Schmieder *et al*., 2010). Raw reads were trimmed based on base quality with a minimum Q_{phred} = 20 using Dynamic Trim (Cox *et al.,* 2010) (default algorithm implemented by SolexaQA) and reads which were shorter than 30 bp were discarded using LengthSort which is also a program under SolexaQA. Assignment of trimmed reads to paired-end and singletons were done using Select Paired.

Transcriptome Data Analysis

The processed clean reads were further analysed using Tuxedo suite pipeline which consists of Bowtie, Tophat2, Cufflinks, Cuffmerge, Cuffdiff and CummeRbund (Trapnell *et al*., 2012). Tophat was used to align RNA-Seq reads to the reference genome *Pisifera* 6.31 using Bowtie. The mapped reads were then assigned to Cufflinks to assemble individual transcripts from the RNA-Seq reads that had been aligned to the genome and inferred the splicing structure of each gene (Trapnell *et al*., 2010). Cuffmerge was then performed to merge all of the 51 mesocarp transcripts assemblies and differentially expressed genes were identified using Cuffdiff. In this study, Cuffdiff read counts varied for each gene across the two biological replicates and three technical replicates for each developmental stage and these variance estimates were used to calculate the significance of observed changes in expression. These results were reported as fragments per kilobase of exon per million reads mapped (FPKM) values for each sample for a given gene, gene- and transcript-related attributes such as common name and location in the genome. The total gene set was annotated using GenBank Plant Reference Sequence (RefSeq) Database (O'Leary *et al*., 2016) via BLASTX (Altschul *et al*., 1990).

Identification of Differentially Expressed Transcripts and Functional Classification

Eight different pairwise comparison (5-8, 8-10, 10-12, 12-15, 15-18, 18-20, 20-22, 22-24) were performed across nine different developmental stages of mesocarp using TIGR Multiexperiment Viewer (MeV) (Saeed *et al*., 2003). A transcript was considered as differentially expressed if it was identified as significantly expressed at p-value threshold of 0.01 in one of the pairwise comparisons. The list of differentially expressed transcripts in mesocarp was subjected to Blast2GO for functional classification based on gene ontology (GO) terms (Conesa *et al*., 2005). Assignment of transcripts to various pathways was performed using Kyoto Encyclopedia of Genes and Genomics (KEGG) database (Kanehisa and Goto, 2000). The fruit developmental stages discussed in this article were done according to Tranbarger *et al.* (2011). The different WAA are related to the following phases; 5-8 WAA (cell division and expansion), 10-12 WAA (lag phase), 15-18 WAA (maturation) and 20-24 WAA (ripening).

RESULTS AND DISCUSSION

Transcriptome Sequencing and Data Analysis

Transcriptomes from 51 total RNA samples from mesocarp were sequenced using the Illumina HiSeq2000 platform. A total of 4 548 075 379 raw reads were generated from the 51 samples (*Table 1*). We obtained a high percentage of clean reads (81.63%) after removal of 835 594 608 low quality reads (*Table 1*). A total of 3 408 872 726 (74.95%) reads were assigned to paired end reads leaving behind 303 608 045 (6.68%) of singletons. The clean paired-end reads were referred to as the 375 Gigabase (Gb) read set (*Table 1*). These reads were subjected to Tuxedo suite pipeline analysis. Tophat2 analysed the mapping results to identify splice junctions between exons. A total of 89.25% of the processed reads were mapped to the reference genome *Pisifera* 6.31 (*Figure 1*). The remaining 10.75 % were not mapped to the current *Pisifera* 6.31 genome.

TABLE 1. QUALITY PROCESSING OF READS FROM 51 MESOCARP TISSUES

Mesocarp library	Read number (forward + reverse)	Read size (forward + reverse)
Raw reads	4 548 075 379 (100%)	570 027 000 000 (100%)
Clean reads	3 712 480 771 (81.63%)	400 263 357 000 (70.22%)
Paired-end reads	3 408 872 726(74.95%)	374 884 197 000 (65.77%)
Orphan reads (single end)	303 608 045 (6.68%)	25 379 160 000 (4.45%)

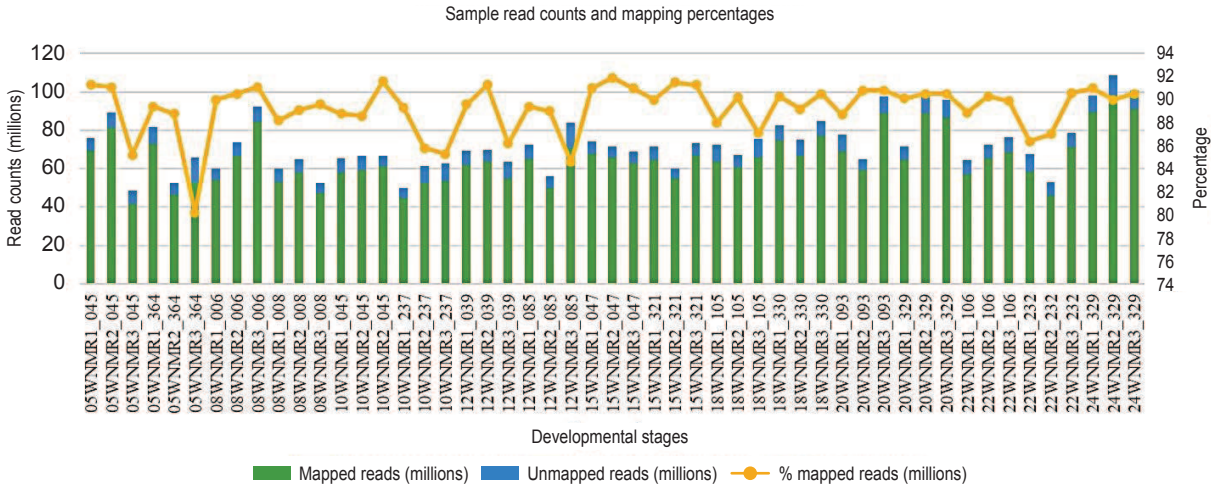


Figure 1. TopHat mapping of the reads from 51 mesocarp tissues on the Pisifera *6.31 reference genome.*

After the transcriptome assembly, 125 102 isoforms and 36 675 unigenes, with a total of 393 858 514 bp and guanine-cytosine (GC) content of 42.37% were identified. The N50 and N90 were 3989 and 1755 bps, respectively (*Table 2*). A total of 24 226 transcripts (66%) were successfully annotated to Plant RefSeq Database while 12 449 (34%) transcripts had no hit to any known genes (*Figure 2*). Notably, 34% of transcripts might be novel sequences specific to the mesocarp tissues of MPOB-Angola *dura*. However, in the work carried out by Jin *et al*. (2017), a lower percentage (41.7%) of similarity was reported when they annotated the *de novo* assembled oil palm transcriptomes from mesocarp and endosperm tissues to the *Arabidopsis thaliana* protein annotation database.

TABLE 2. ASSEMBLY STATISTICS USING TUXEDO SUITE PIPELINE

Attributes	Value
Number of transcripts	125 102
Total bases	393 858 514
Guanine-cytosine (GC) content	42.37%
Gap content	0%
Average length	3 148.29
Maximum length	39 704
Minimum length	35
N50 length	3 989
N90 length	1 755

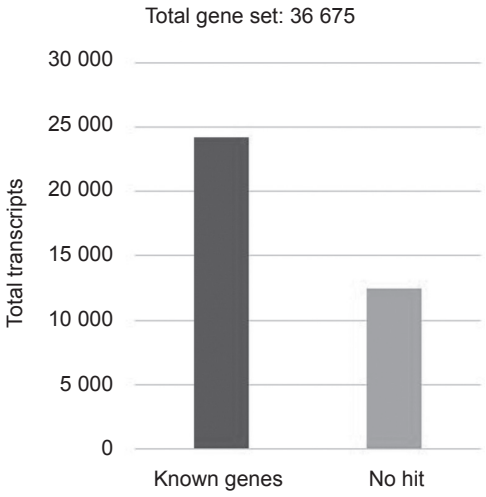


Figure 2. Summary of transcripts annotated to the Plant Reference Sequence (RefSeq) Database.

Pearson correlation coefficients were calculated for the three technical replicates and the two biological replicates using the $\log_2$ (FPKM) values obtained across all developmental stages from 5, 8, 10, 12, 15, 18, 20, 22 and 24 WAA of mesocarp tissues. All of the technical replicates showed higher correlation of more than 0.84 meanwhile the Pearson correlation coefficients for the two biological replicates across the developmental stages ranged from 0.84 to 0.93. This showed that the data from technical and biological replicates are reproducible. Furthermore, the data was also plotted using boxplot function of CummeRbund package (*Figure 3*). The distribution

for $\log_{10}$ (FPKM) expression ratios across technical and biological replicates in the boxplot were similar. This further supported the reproducibility of data from technical and biological replicates.

A pairwise T-test was performed on the 36 675 genes across nine different developmental stages of mesocarp tissues. From the result of eight pairwise T-tests carried out, the total significant genes (without duplication) across all the pairwise comparisons was 21 261 (*Table 3*). The longest sequences of these genes were retrieved for functional annotation using BLAST2GO. Similarity search using BLASTX against Plant RefSeq Database at the cut off value of 1^{E-5} was carried out on the 21 261 genes before BLAST2GO was performed on these genes.

Functional Annotation of Differentially Expressed Transcripts

Functional classification by GO. GO can be defined as controlled vocabularies of defined terms corresponding to gene product properties (Conesa and Götz, 2008). It has three sub parts that describe gene products in terms of their association to Biological Processes, Cellular Components and Molecular Functions (Conesa and Götz, 2008). In this study, BLAST2GO analysis was performed separately according to the pairwise comparison and the summary of the sequences that were annotated to the GO terms are shown in *Table 3*.

Figure 4 illustrates the three GO terms associated with Biological Processes (the whole figure with all the GO terms are included in *Appendix 1*). It was observed that transcripts associated with lipid metabolic process were only expressed during transition of fruit developmental stages from 10 to 12 WAA (154 transcripts), 18 to 20 WAA (174 transcripts) and 22 to 24 WAA (328 transcripts). As 18 WAA belongs to the end of maturation phase whereas 20 WAA occurs at the onset of ripening, this might explain why lipid metabolism-related genes are highly expressed during these stages. Peak lipid biosynthesis starts during maturation and reaches a maximum at ripening stage (Teh *et al.*, 2014). On the other hand, it was observed that carbohydrate metabolic process related genes were only expressed during the transition of fruit developmental stages

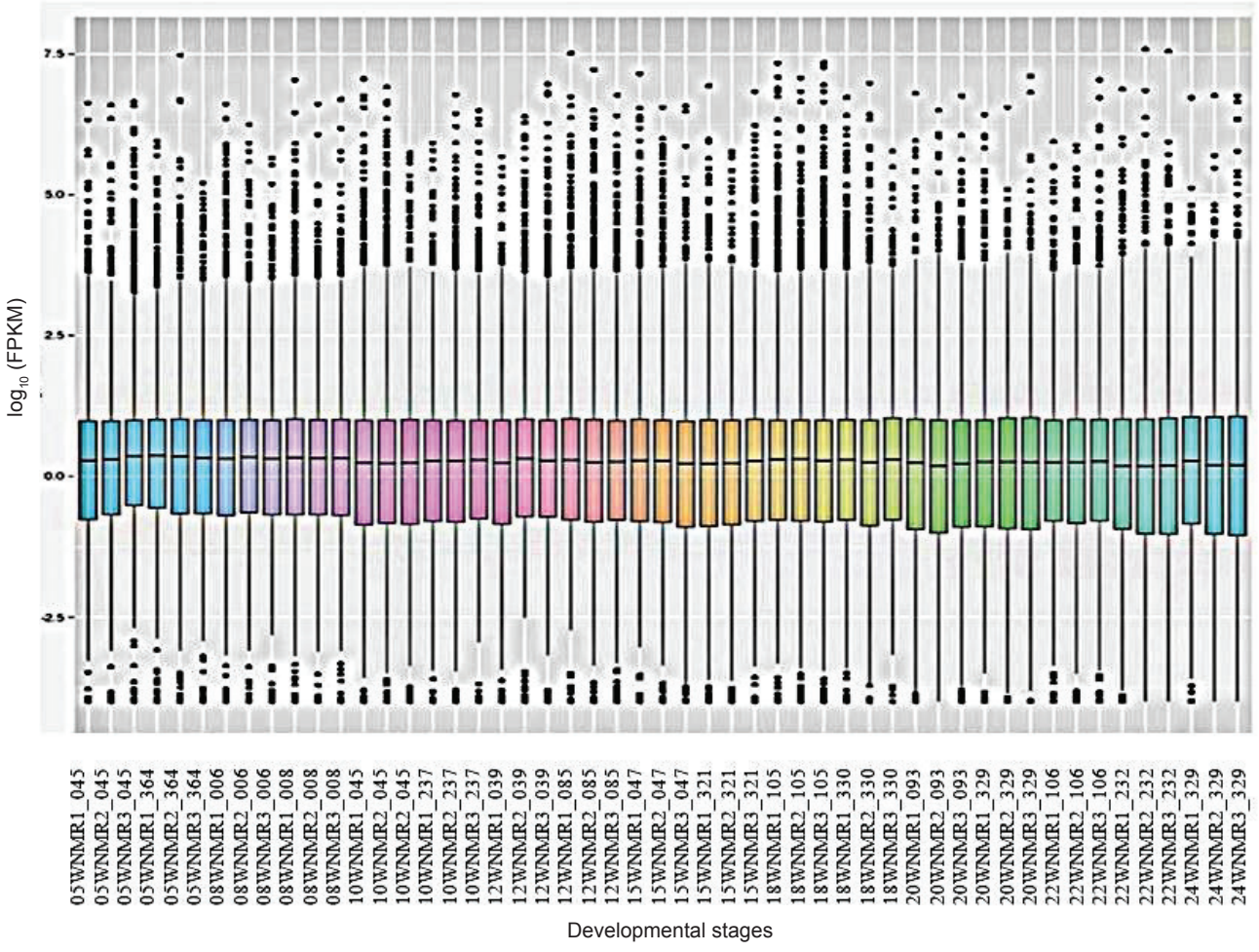


Figure 3. Boxplot representing the reproducibility of the biological and technical replicates.

from 10 to 12 WAA (150 transcripts) (*Figure 4*). This could be due to the starting point of uptake in sugar during the lag phase that precedes the onset of oil synthesis during the maturation phase. Besides that, transcripts related to response to hormone were weakly expressed during the transition of fruit developmental stages from 10 to 12 WAA (182 transcripts), 18 to 20 WAA (192 transcripts) and 20 to 22 WAA (181 transcripts), whereas they were more highly expressed during transition of fruit developmental stages from 5 to 8 WAA (381 transcripts), 12 to 15 WAA (370 transcripts) and 22 to 24 WAA (440 transcripts) (*Figure 4*). According to Tranbarger *et al.* (2011), during end of the lag phase (12 WAA), there is a decrease in auxin, gibberellic acid (GA) and cytokinin metabolites and this may explain the faintly expressed transcripts related to responses to hormones during transition period of fruit developmental stages from 10 to 12 WAA. Apart from that, the large increase in the hormones abscisic acid (ABA) and ethylene related to ripening process explains the highly expressed transcripts involved in responses to hormones during the transition period of fruit developmental process between 22 to 24 WAA.

Meanwhile, for Molecular Functions, it was observed that the number of transcripts associated with all the GO terms were highly expressed during transition of fruit developmental stages from 22 to 24 WAA (*Figure 5*). Tranbarger *et al.* (2011) had earlier revealed that transcripts involved in carbohydrate and cellular aromatic compound metabolic processes were up-regulated during the oil palm ripening stage. As such, the high expression of transcripts for GO terms under Molecular Functions (*e.g.* nucleotide binding, nucleoside phosphate binding, nucleic acid binding and ribonucleotide binding) reflect their potential role in assisting the processes reported by Tranbarger *et al.* (2011).

For Cellular Components, surprisingly, the transcripts related to the intracellular organelle lumen were only expressed in the transition period from 15 to 18 WAA (*Figure 6*). This could be due to the central role of endoplasmic reticulum (ER) in lipid and protein biosynthesis (Alberts *et al.*, 2002). It was previously described by Tranbarger *et al.* (2011) that triacylglycerol (TAG) metabolism occurred in ER and there was a paradigm shift observed with a steep rise of all the TAG related lipids from 14 to 17 WAA (maturation phase).

Functional classification by KEGG. KEGG is a database resource to determine the functions and utilities of the biological system using molecular-level information (Conesa and Götz, 2008). The pathway-based analysis is useful for further understanding of the biological functions and gene interactions. To further characterise the transcriptome of the MPOB-Angola *dura* mesocarp tissues, a total of 13 996 transcripts with significant matches in the GO

TABLE 3. SUMMARY OF SEQUENCES ANNOTATED TO GENE ONTOLOGY

Pairwise comparison	Total sequences	Sequences annotated to gene ontology	Percentage of annotated sequences %	Total sequences annotated without duplication
5*vs*.8	5 691	3 943	69	-
8*vs*.10	4 366	3 114	71	-
10*vs*.12	2 819	1 918	68	-
12*vs*.15	5 686	3 919	69	-
15*vs*.18	4 610	3 225	70	-
18*vs*.20	3 219	2 168	67	-
20*vs*.22	2 868	1 955	68	-
22*vs*.24	7 127	4 769	67	-
				21 261

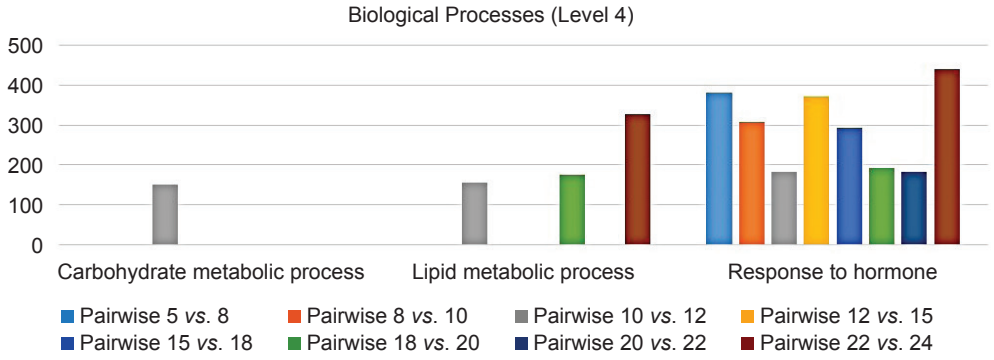


Figure 4. Three gene ontology (GO) terms associated with Biological Processes across eight pairwise comparisons.

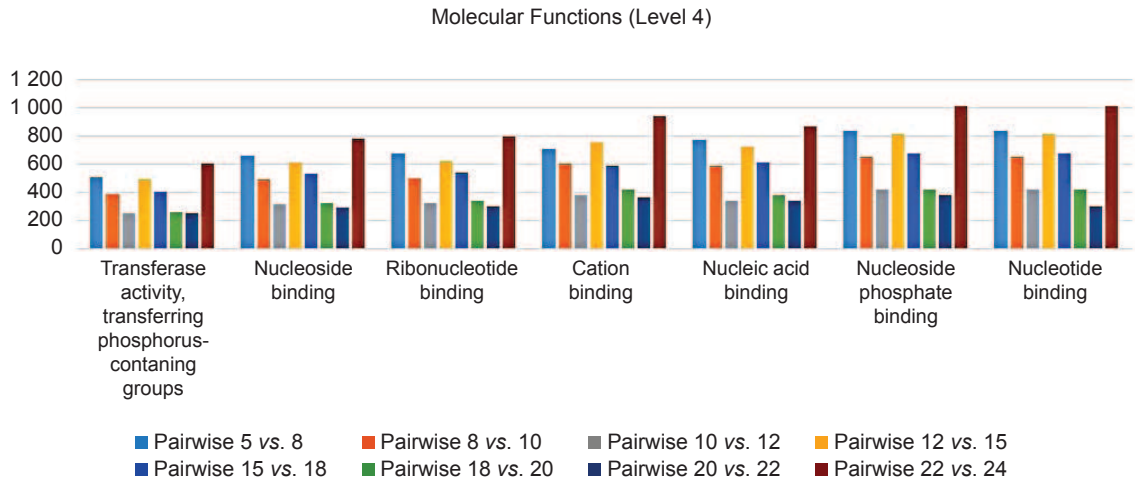


Figure 5. Gene ontology (GO) classification of assembled unigenes under Molecular Functions across eight pairwise comparisons.

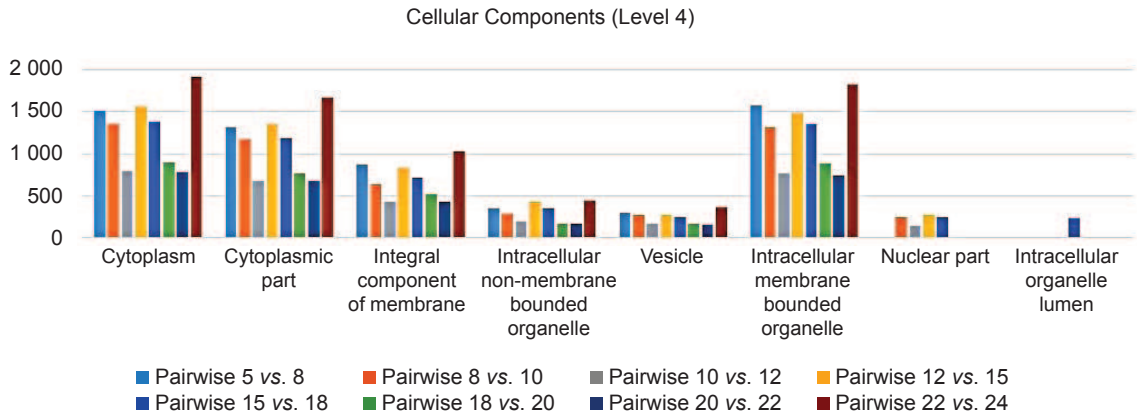


Figure 6. Gene ontology (GO) classification of assembled unigenes under Cellular Components across eight pairwise comparisons.

database were analysed using KEGG pathway database. Three metabolic processes, namely lipid, carbohydrate and hormone metabolism, were studied in detail.

A total of 57 transcripts were found to be involved in FAS, followed by 119 transcripts in fatty acid (FA) degradation, 44 transcripts in FA elongation and 59 transcripts in unsaturation of fatty acid for lipid metabolism. It was observed that 3-oxoacyl-synthase I, 3-oxoacyl-synthase III, palmitoyl-acyl carrier protein thioesterase, and 3-ketoacyl-acyl carrier protein synthase I were involved in FAS process. Meanwhile, acyl-protein thioesterase, palmitoyl-protein thioesterase 1, enolase and glyoxysomal FA beta-oxidation were associated with FA elongation. Moreover, acyl-CoA oxidase, peroxisomal 3-ketoacyl-CoA thiolase and acetyl-CoA acetyltransferase were involved in FA degradation. It was also observed that 3-oxoacyl-[acyl-carrier-protein] reductase, SAD and peroxisomal enoyl-CoA hydratase were responsible for FA desaturation.

Meanwhile, a total of 238 transcripts were linked to glycolysis, followed by 155 transcripts in pyruvate metabolism, 74 transcripts in tricarboxylic acid cycle (TCA) cycle and 290 in starch and sucrose metabolism for carbohydrate metabolism. It was observed that pyruvate kinase, pyruvate dehydrogenase, sugar isomerase and *GDSL* esterases/lipase proteins (*GELP*) were also involved in the glycolysis process. Moreover, biotin carboxylase, glycerate dehydrogenase and acetyl-CoA acetyltransferase were associated with pyruvate metabolism. Succinyl-CoA ligase, ATP citrate synthase and isocitrate dehydrogenase were involved in TCA cycle. It was also observed that beta amylase, glucan endo-1, 3-beta-glucosidase 3 precursor and cell wall invertase were responsible for starch and sucrose metabolism.

Apart from that, a total of 96 transcripts were involved in process involving alpha-linolenic acid (ALA) metabolism, 35 transcripts in steroid biosynthesis, 11 in carotenoid biosynthesis and seven transcripts in zeatin biosynthesis for hormone

metabolism. It was observed that allene oxide cyclase 4, putative 12-oxophytodienoate reductase, and acetyl-CoA acetyltransferase were involved in the production of jasmonic acid involving linolenic acid. Besides that, cycloartenol synthase was associated with steroid biosynthesis where it plays a role in producing brassinosteroid hormones (Clouse, 2002). Moreover, ABA 8′-hydroxylase and carotenoid-cleavage dioxygenase were involved in carotenoid biosynthesis which is responsible for the production of ABA. Lastly, it was observed that those cytokinin dehydrogenase and cytokinin oxidases were responsible for zeatin biosynthesis where these transcripts play an important part in the production of cytokinin hormones (Conesa and Götz, 2008). The functional classification of KEGG contributed resourceful information for evaluating certain processes, functions and pathways involved in mesocarp tissues of MPOB-Angola *dura*.

Functional Annotation of Transcripts Associated with Lipid Metabolism, Hormone Metabolism and Transcription Factors

From the list of annotated genes, FA, TAG, plant hormone metabolism genes and TF were selected to evaluate their expression profile during mesocarp development.

Lipid metabolism. For genes related to FA, we managed to identify 3-ketoacyl-acyl carrier protein synthase I (*KAS I*), 3-Oxoacyl-ACP-synthase III (*KAS III*), oleoyl-acyl carrier protein thioesterase (*FatA*) and palmitoyl-acyl carrier protein thioesterase (*FatB*). The expression profiles of these genes were plotted and compared with the one reported by Bourgis *et al*. (2011). It was observed that the expression profiles of the FA genes in this study are almost similar with those reported by Bourgis *et al.* (2011) (*Figure 7*), where an increase in expression profile throughout mesocarp development was observed. Moreover, there was a drastic increase in expression profile of *KAS I* from 15 to 18 WAA and interestingly 15 WAA is the point where mesocarp oil accumulation in this tissue enters the exponential phase (Tranbarger *et al*., 2011). The expression profiles of the FA related genes in this study generally showed increased expression until the ripening stage. This correlates well with the observations of Bourgis *et al*. (2011) that the expression levels for almost all transcripts related to FAS continue to increase until the end of oil accumulation.

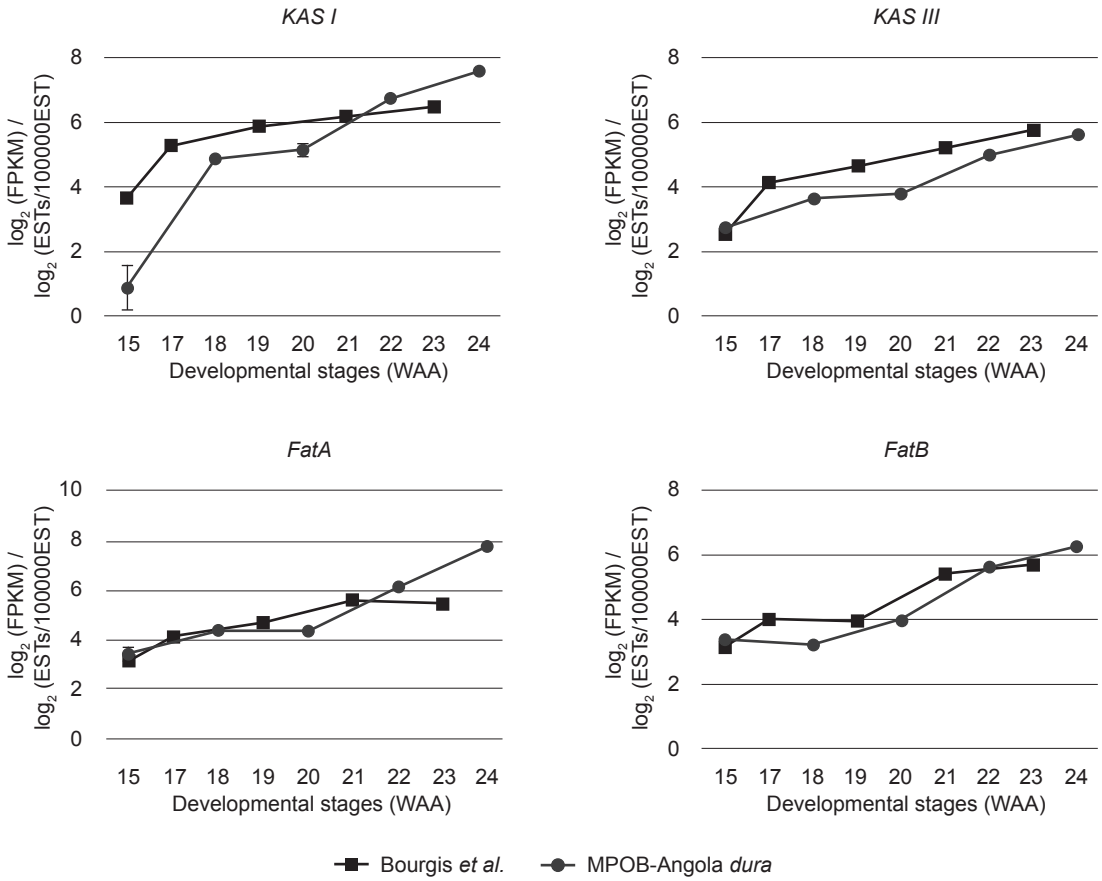


Figure 7. Comparison of expression profiles between MPOB-Angola dura *[log_2 (FPKM)] and Bourgis* et al. *(2011) [log_2 (EST/100000 EST) for selected fatty acid (FA) genes during mesocarp development. The error bars represent ± standard error (SE).*

Similarly, the expression of selected TAG biosynthesis genes throughout mesocarp development were compared with the published datasets from transcriptome sequencing by Jin *et al*. (2017). The expression profiles of the selected genes of MPOB-Angola *dura* (*Figure 8*), *DGAT*, glycerol-3-phosphate acyltransferase (*GPAT*) and phosphatidate phosphatase (*PAP*), exhibited an increase in expression signal during the ripening stage (20 to 22 WAA) where maximum lipid concentration is observed. However, the exception was *DGAT*, where Jin *et al*. (2017) reported decreasing expression towards the ripening stage. The role of *DGAT* in TAG biosynthesis is to catalyse the last (acyl-CoA dependent) acylation step to TAG. Interestingly, the expression of omega-6 fatty acid desaturase (*FAD*) for both MPOB-Angola *dura* and Jin *et al.* (2017) decreases throughout mesocarp development. Jin *et al.* (2017) opined that the role of the *FAD*-encoding enzyme is to synthesise ALA and for a production of longer FA.

Hormone metabolism. Ethylene-insensitive 3-like 1 protein (*EIN3*), ABA responsive isoform 1, GA receptor *GID1* and auxin-independent growth promoter protein were identified as major plant hormonal genes from the list of the annotated genes. The *EIN3* which is a negative regulator of ethylene action (Teh *et al*., 2014), exhibited a down regulated expression pattern (*Figure 9*) for both MPOB-Angola *dura* and Jin *et al.* (2017). GA receptor *GID1* plays an important role in the GA signal transduction (Wang *et al*., 2017). *Figure 9* shows the expression of GA receptor *GID1* in the MPOB-Angola *dura* and as reported by Jin *et al.* (2017). GA receptor *GID1* in MPOB-Angola *dura* appeared to show higher expression at 5 WAA which is during the development phase and the expression decreased at ripening phase. On the other hand, Jin *et al.* (2017) reported an opposite trend to that observed in this study, as GA receptor *GID1* genes peaked at the ripening phase (22 WAA). The expression profiles of ABA responsive isoform 1 and auxin-independent growth promoter protein were also compared to Jin *et al*. (2017) and there was similarity in the trend of expression when these two genes were compared to the published transcriptome datasets. *Figure 9* shows that ABA responsive isoform 1 exhibited an increase expression in the mesocarp as lipid accumulated during ripening. Furthermore, auxin-independent growth promoter protein also showed highest expression at 22 WAA.

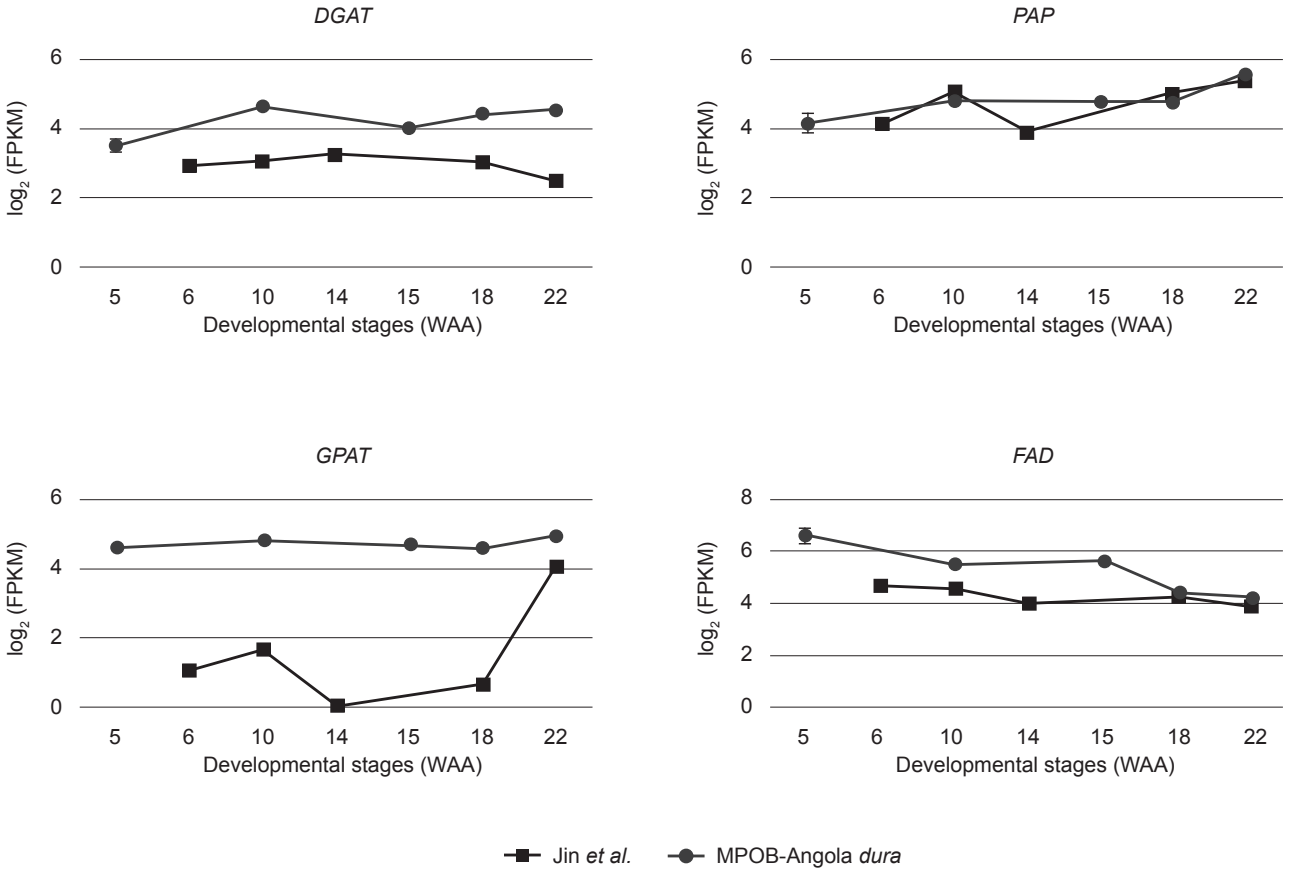


Figure 8. Comparison of expression profiles between MPOB-Angola dura *[log_2 (FPKM)] and Jin* et al. *(2017) [log_2 (FPKM)] for selected triacylglycerol (TAG) genes during mesocarp development. The error bars represent ± standard error (SE).*

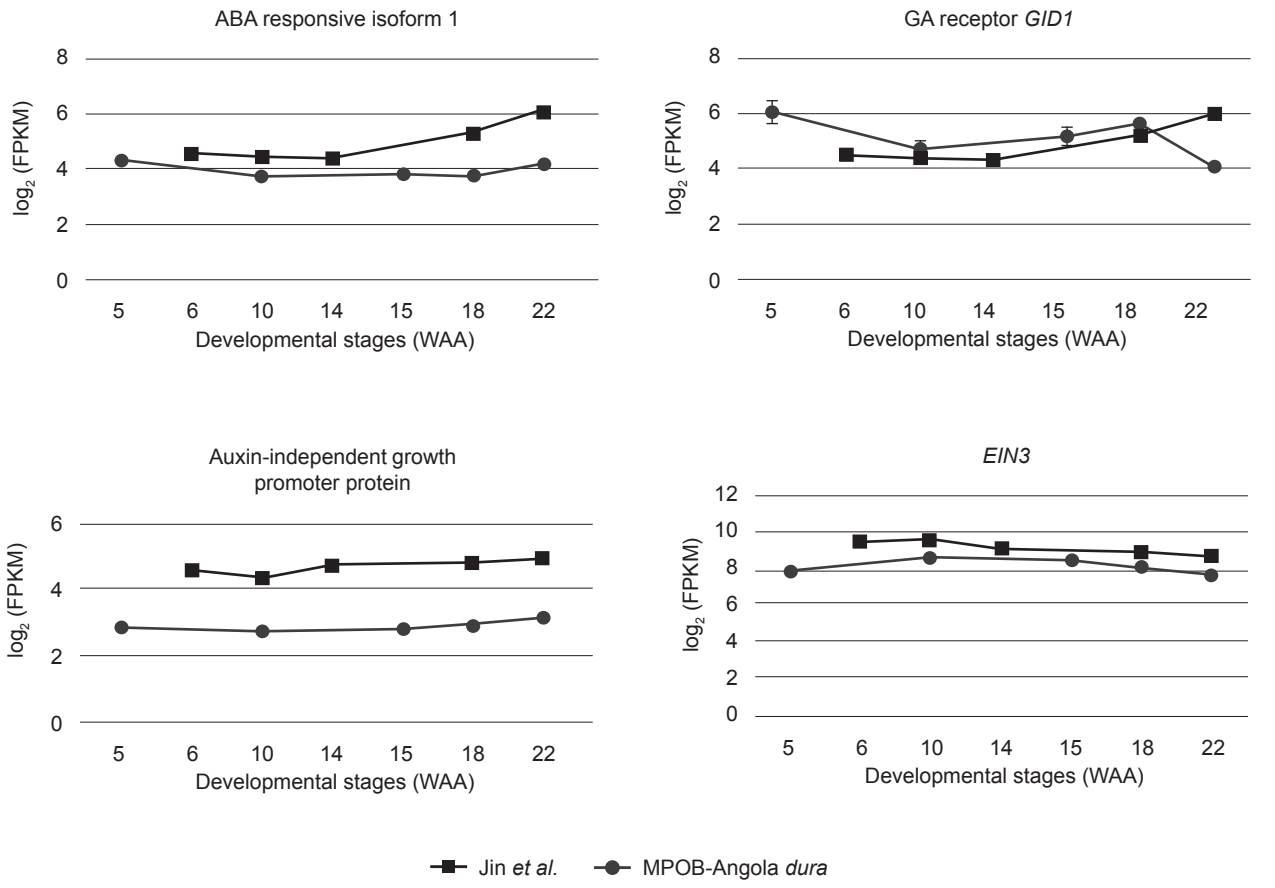


Figure 9. Comparison of expression profiles between MPOB-Angola dura *[log_2 (FPKM)] and Jin* et al. *(2017) [log_2 (FPKM)] for genes associated with hormone metabolism during mesocarp development. The error bars represent ± standard error (SE).*

Transcription factors. TF are proteins which are involved in regulating gene expression (Takatsuji, 1998). Here, we identified WRI1, basic leucine zipper (bZIP), nuclear transcription factor Y subunit alpha (NF-YA) and ethylene responsive transcription factors (ERF). These selected TF were compared to the published transcriptome datasets (*Figure 10*). The expression profile of WRI1 in the MPOB-Angola *dura* is similar to that published by Tranbarger *et al.* (2011) during lag and ripening phases. The levels of WRI1 were low but gradually increased throughout the whole lag phase and then increased rapidly at the time of ripening. In contrast, Tranbarger *et al.* (2011) reported that the expression of WRI1 drastically decreased at maturation phase while this study observed that the WRI1 in MPOB-Angola *dura* was elevated throughout the maturation phase. The up-regulated expression of WRI1 during the ripening phase suggests that WRI1 is a transcriptional enhancer of FAS genes such as ATP-citrate synthase (*ACS*), pyruvate dehydrogenase e1 component subunit β (*PDHE1b*), biotin carboxyl carrier protien of acetyl coenzyme a carboxylase 1 (*BCCP*), acetyl coenzyme a carboxylase (*ACCase*), acyl-carrier protein 1 (*ACP1*), *KAS III*, enoyl- acyl carrier protein reductase (*EAR*) and long chain fatty acid coenzyme a ligase 4 (*LACS*) (Yeap *et al*., 2017).

The expression profiles of NF-YA and ERF observed in this study are similar to Tranbarger *et al.* (2011) and Jin *et al.* (2017), respectively. Contrary to WRI1, the levels of NF-YA were high in the early maturation phase, declined during the maturation phase, peaking at early ripening stage, followed by a decline until 24 WAA. Yeap *et al.* (2017) reported that NF-YA is a transcriptional activator of FAS, FA modification and TAG assembly genes. Similarly, the level of expression of ERF was high during early development stage and the expression remained high until the end of ripening stage. Tranbarger *et al.* (2011) reported that ethylene production in oil palm mesocarp is marked by a coordinated increase in a large number of ethylene related transcripts especially those encoding for ERF. Nonetheless, bZIP exhibited differences in expression profiles when MPOB-Angola *dura* was compared to the results of Tranbarger *et al.* (2011) (*Figure 10*). We observed that the expression pattern of bZIP in MPOB-Angola *dura* showed a minimal increase from 15 until 24 WAA whereas bZIP reported by Tranbarger *et al.* (2011) showed a drastic increase from 14 to 20 WAA and falling off towards 23 WAA. Despite the differences of expression pattern, bZIP was identified as one of the differentially expressed transcripts in the transcriptome analysis.

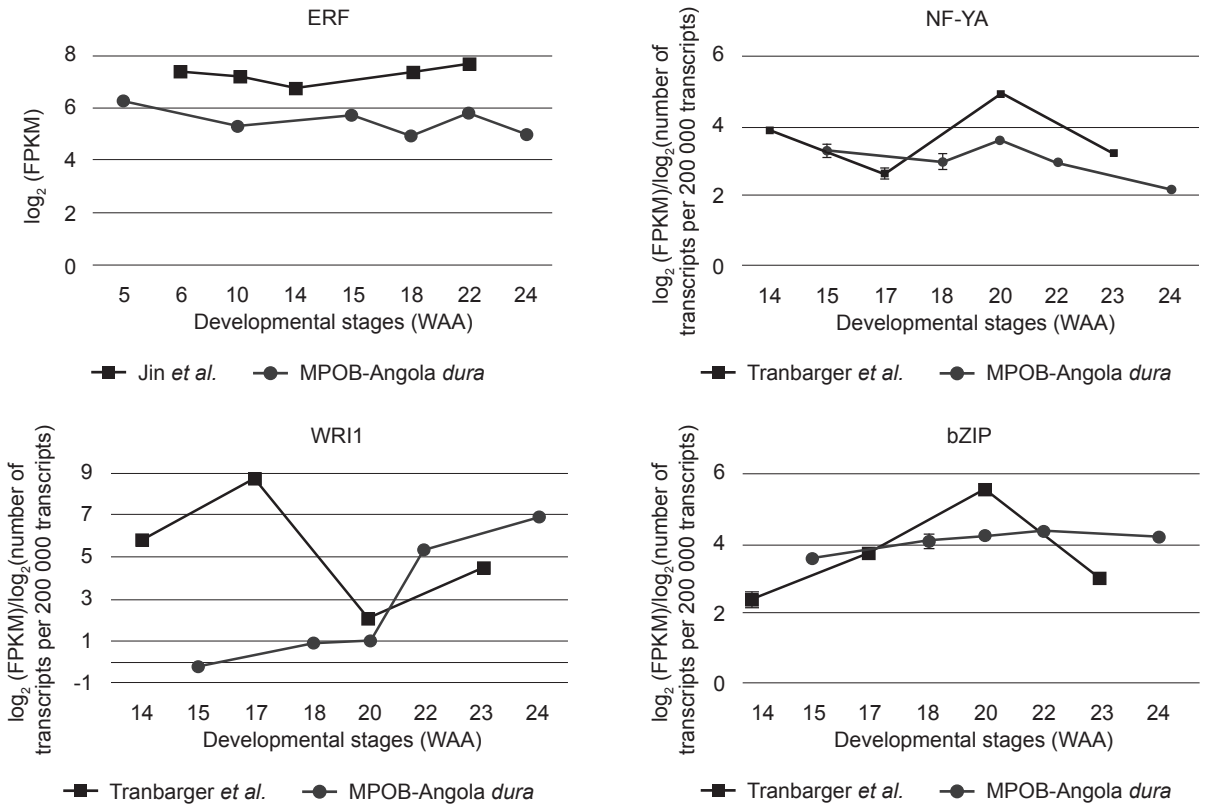


Figure 10. Comparison of expression profiles between MPOB-Angola dura *[log_2 (FPKM)] with Jin* et al. *(2017) [log_2 (FPKM)] and Tranbarger* et al. *(2011) [log_2 (number of transcripts per 200 000 transcripts)] for selected transcription factors during mesocarp development. The error bars represent ± standard error (SE).*

CONCLUSION

In this study, we analysed the transcriptome of MPOB-Angola *dura* mesocarp tissues from nine different developmental stages. A total of 36 675 gene sets were identified from RNA-Seq with 24 226 of the transcripts successfully annotated with Plant RefSeq Database. From this number, a total of 21 261 transcripts were found to be significantly differentially expressed, with 66% of the transcripts annotated with significant matches in the GO database. To our knowledge, this is the first attempt to profile the expression of transcripts across MPOB-Angola *dura* fruits, an important germplasm being used for introgression into advanced breeding lines. The transcriptomes in this study provide valuable information of the transcriptional mechanisms controlling the Angolan *dura* fruit development, maturation and ripening.

ACKNOWLEDGEMENT

The authors wish to thank the Director-General of MPOB, Deputy Director-General (R&D), colleagues of Bioinformatics and Genomics Units, Advanced Biotechnology and Breeding Centre, for their support and technical assistance. The first author would also like to acknowledge MPOB for the Graduate Student Assistantship Scheme (GSAS) sponsorship. The work was carried out using internal MPOB funding.

REFERENCES

Alberts, B; Johnson, A; Lewis, J; Raff, M; Roberts, K and Walter, R (2002). The endoplasmic reticulum. *Molecular Biology of the Cell*. 4th edition. https://www.ncbi.nlm.nih.gov/books/NBK26841/, accessed on 29 January 2019. 712 pp.

Altschul, S F; Gish, W; Miller, W; Myers, E W and Lipman, D J (1990). Basic local alignment search tool. *J. Molecular Biology*, *215(3)*: 403-410.

Andrews, S (2010). FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc, accessed on 16 February 2017.

Bourgis, F; Kilaru, A; Xia, C; Ebounge, G F N; Drira, N; Ohlrogge, J E and Arondel, V (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. of the National Academy of Sciences of the United States of America*, *108(30)*: 12527-12532.

Clouse, S D (2002). Brassinosteroids. *The Arabidopsis Book*. p. e0009. DOI:10.1199/tab.0009.

Conesa, A and Götz, S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics*, *2008*: 619832. DOI: 10.1155/2008/619832.

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

Cox, M P; Peterson, D A and Biggs, P J (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, *11*: 485.

Fan, H; Xiao, Y; Yang, Y; Xia, W; Mason, A S; Xia, Z; Qiao, F; Zhao, S and Tang, H (2013). RNA-Seq analysis of *Cocos nucifera*: Transcriptome sequencing and *de novo* assembly for subsequent functional genomics approaches. *PLoS ONE*, *8(3)*: e59997.

Guerin, C; Joët, T; Serret, J; Lashermes, P; Vaissayre, V; Agbessi, M D; Beulé, T; Severac, D; Amblard, P; Tregear, J; Durand-Gasselin, T; Morcillo, F and Dussert, S (2016). Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *The Plant J., 87(5)*: 423-441.

Jin, J; Sun, Y; Qu, J; Syah, R; Lim, C H; Alfiko, Y; Rahman, N E B; Suwanto, A;Yue, G; Wong, L; Chua, N H and Ye, J (2017). Transcriptome and functional analysis reveals hybrid vigor for oil biosynthesis in oil palm. *Scientific Reports*, *7*: 439.

Kanehisa, M and Goto, S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research, 28(1)*: 27-30.

Kilaru, A; Cao, X; Dabbs, P B; Sung, H-J; Rahman, M M; Thrower, N; Zynda, G; Podicheti, R; Ibarra-Laclette, E; Herrera-Estrella, L; Mockaitis, K and Ohlrogge, J B (2015). Oil biosynthesis in a basal angiosperm: Transcriptome analysis of *Persea Americana* mesocarp. *BMC Plant Biology*, *15:* 203.

Kushairi, A; Loh, S K; Azman, I; Hishamuddin, E; Ong-Abdullah, M; Mohd Noor Izuddin, Z B; Razmah, G; Sundram, S and Parveez, G K A (2018). Oil palm economic performance in Malaysia and R&D progress in 2017. *J. Oil Palm Res. Vol. 30(2)*: 163-195.

Kushairi, A; Singh, R and Ong-Abdullah, M (2017). The oil palm industry in Malaysia: Thriving with transformative technologies. *J. Oil Palm Res. Vol. 29(4)*: 431-439.

Kushairi, A; Rajanaidu, N; Mohd Din, A; Isa, Z A and Noh, A (2003). Selection of some economically important traits from MPOB's Tanzanian and Angolan oil palm germplasm collections. *Proc. of the 2003 PIPOC International Palm Oil Congress: The Power-House for the Global Oils & Fats Economy – Agriculture Conference*. MPOB, Bangi. p. 665-689.

Lalonde, E; Ha, K C; Wang, Z; Bemmo, A; Kleinman, C L; Kwan, T; Pastinen, T and Majewski, J (2011). RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Research*, *21(4)*: 545-554.

Ma, W; Kong, Q; Arondel, V; Kilaru, A; Bates, P D; Thrower, N A; Benning, C and Ohlrogge, J B (2013). *WRINKLED1*, a ubiquitous regulator in oil accumulating tissues from *Arabidopsis* embryos to oil palm mesocarp. *PLoS One*, *8(7)*: e68887.

Noh, A; Rajanaidu, N; Kushairi, A; Mohd Rafil, Y; Mohd Din, A; Mohd Isa, Z A and Saleh, G (2002). Variability in fatty acid composition, iodine value and carotene content in the MPOB oil palm germplasm collection from Angola. *J. Oil Palm Res. Vol. 14(2)*: 18-23.

O'Leary, N A; Wright, M W; Brister, J R; Ciufo, S; Haddad, D; McVeigh, R; Rajput, B; Robbertse, B; Smith-White, B; Ako-Adjei, D; Astashyn, A; Badretdin, A; Bao, Y; Blinkova, O; Brover, V; Chetvernin, V; Choi, J; Cox, E; Ermolaeva, O; Farrell, C M; Goldfarb, T; Gupta, T; Haft, D; Hatcher, E; Hlavina, W; Joardar, V S; Kodali, V K; Li, W; Maglott, D; Masterson, P; McGarvey, K M; Murphy, M R; O'Neill, K; Pujar, S; Rangwala, S H; Rausch, D; Riddick, L D; Schoch, C; Shkeda, A; Storz, S S; Sun, H; Thibaud-Nissen, F; Tolstoy, I; Tully, R E; Vatsan, A R; Wallin, C; Webb, D; Wu, W; Landrum, M J; Kimchi, A; Tatusova, T; DiCuccio, M; Kitts, P; Murphy, T D and Pruitt, K D (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research, 44(Database issue)*: D733-D745.

Okoniewski, M J and Miller, C J (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics*, *7*: 276.

Ong, P W; Chan, P-L; Ooi, L C L and Singh, R (2019). Isolation of high quality total RNA from various tissues of oil palm (*Elaeis guineensis*) for reverse transcription quantitative real-time PCR (RT-qPCR). *J. Oil Palm Res. Vol. 31(2)*: 195-203.

Rajanaidu, N; Jalani, B K and Domingos, M (1991). Collection of oil palm (*Elaeis guineensis*) germplasm in Angola. *ISOPB Newsletter*, *8(2)*: 2-3.

Rosli, R; Amiruddin, N; Ab Halim, M A; Chan, P-L; Chan, K-L; Azizi, N; Morris, P E; Leslie, Low E-T; Ong-Abdullah, M; Sambanthamurthi, R; Singh, R and Murphy, D J (2018a). Comparative genomic and transcriptomic analysis of selected fatty acid biosynthesis genes and CNL disease resistance genes in oil palm. *PLoS ONE*, *13(4)*: e0194792.

Rosli, R; Chan, P-L; Chan, K-L; Amiruddin, N; Low, E-T L; Singh, R; Harwood, J L and Murphy, D J (2018b). *In silico* characterization and expression profiling of the diacylglycerol acyltransferase gene family (DGAT1, DGAT2, DGAT3 and WS/DGAT) from oil palm, *Elaeis guineensis*. *Plant Science*, *275*: 84-96.

Royce, T E; Carriero, N J and Gerstein, M B (2007). An efficient pseudomedian filter for tiling microarrays. *BMC Bioinformatics*, *8*: 186.

Saeed, A I; Sharov, V; White, J; Li J; Liang, W; Bhagabati, N; Braisted, J; Klapa, M; Currier, T; Thiagarajan, M; Sturn, A; Snuffin, M; Rezantsev, A; Popov, D; Ryltsov, A; Kostukovich, E; Borisovsky, I; Liu, Z; Vinsavich, A; Trush, V and Quackenbush, J (2003). TM4: A free, open-source system for microarray data management and analysis. *Biotechniques*, *34(2)*: 374-378.

Schmieder, R; Lim, Y W; Rohwer, F and Edwards, R (2010). TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics*, *11*: 341.

Singh, R; Ong-Abdullah, M; Low, E-T L; Abdul Manaf, M A; Rosli, R; Rajanaidu, N; Ooi, L C-L; Ooi, S-E; Chan, K-L; Halim, M A; Azizi, N; Nagappan, J; Bacher, B; Lakey, N; Smith, S W; He, D; Hogan, M; Budiman, M A; Lee, E K; Desalle, R; Kudrna, D; Goicoechea, J L; Wing, R A; Wilson, R A; Fulton, R S; Ordway, J M; Martienssen, R A and Sambanthamurthi, R (2013). Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature*, *500*: 335-339.

Takatsuji, H (1998). Zinc-finger transcription factors in plants. *Cellular and Molecular Life Sciences*, *54*: 582-596.

Teh, H F; Neoh, B K; Wong, Y C; Kwong, Q B; Ooi, T E K; Lee, T M N; Soon, H T; Yoke, J S L; Danial, A D; Ersad, M A; Kulaveerasingam, H and Appleton, D R (2014). Hormones, polyamines, and cell wall metabolism during oil palm fruit mesocarp development and ripening. *J. Agricultural and Food Chemistry*, *62*: 8143-8152.

Tranbarger, T J; Dussert, S; Joet, T; Argout, X; Summo, M; Champion, A; Cros, D; Omore, A; Nouy, B and Morcillo, F (2011). Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiology*, *156*: 564-584.

Trapnell, C; Williams, B A; Pertea, G; Mortazavi, A; Kwan, G; Baren, M J V; Salzberg, S L; Wold, B J and Pachter, L (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*: 511-515.

Trapnell, C; Roberts, A; Goff, L; Pertea, G; Kim, D; Kelley, D R; Pimentel, H; Salzberg, S L; Rinn, J L and Pachter, L (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols, 7(3)*: 562-578.

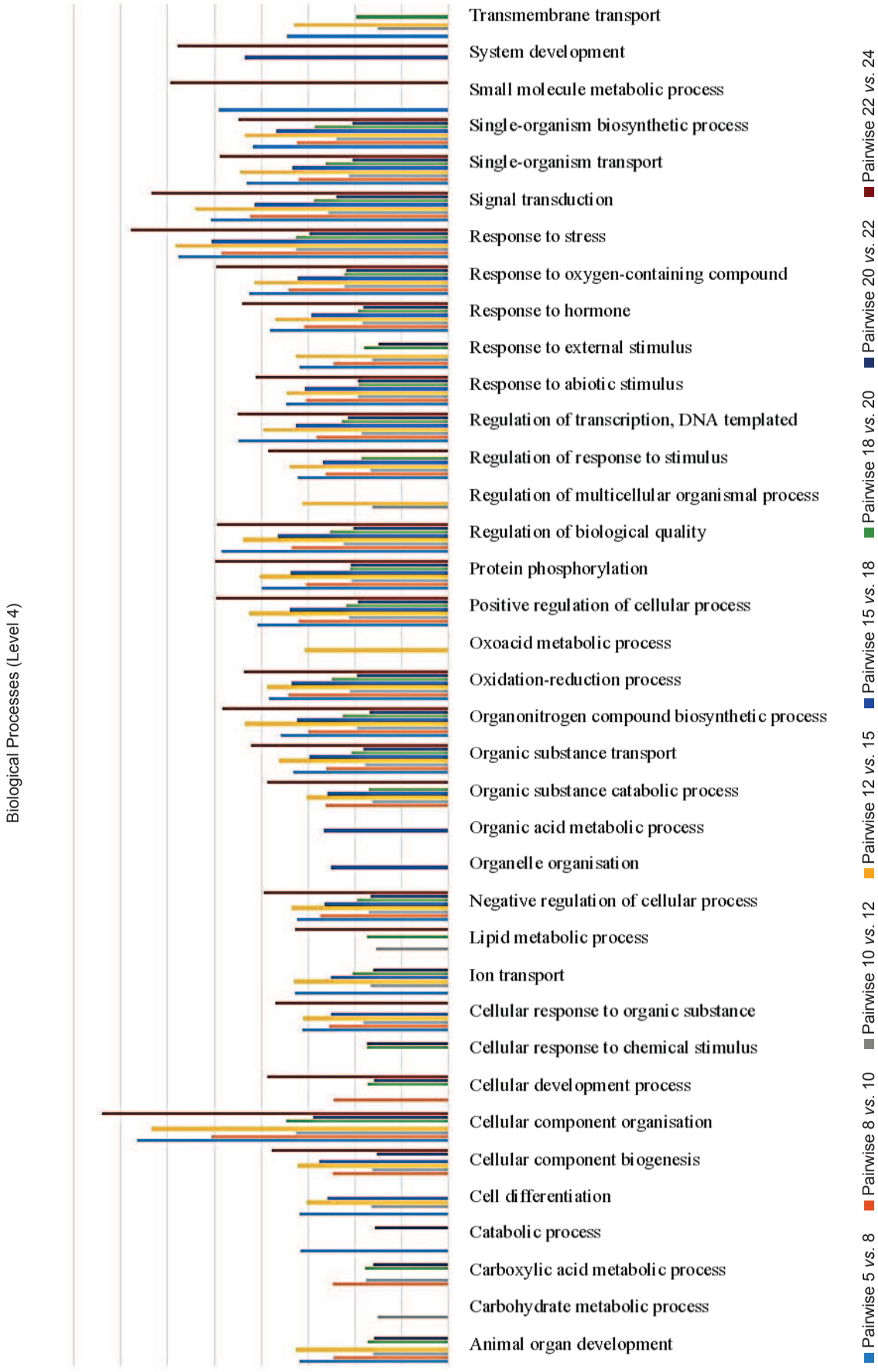
Wang, X; Li, J; Ban, L; Wu, Y; Wu, X; Wang, Y; Wen, H; Chapurin, V; Dzyubenko, N; Li, Z; Wang, Z and Gao, H (2017). Functional characterization of a gibberellin receptor and its application in alfalfa biomass improvement. *Scientific Reports*, *7*: 41296.

Wang, Z; Gerstein, M and Snyder, M (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10(1)*: 57-63.

Wei, W; Qi, X; Wang, L; Zhang, Y; Hua,W; Li, D; Lv, H and Zhang, X (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics*, *12*: 451.

Yeap, W-C; Lee, F-C; Shan, D K S; Musa, H; Appleton, D R and Kulaveerasingam, H (2017). *WRI1-1*, *ABI5*, *NF-YA3* and *NF-YC2* increase oil biosynthesis in coordination with hormonal signalling during fruit development in oil palm. *The Plant J.*, *91*: 97-113.

Appendix 1



Gene ontology (GO) classification of assembled unigenes under Biological Processes across eight pairwise comparisons.