



Faculty of Resource Science and Technology

Comparative Analysis of Gene Expressed in Trunking and Non-Trunking *Metroxylon sagu* Leaves Utilizing Transcriptome Sequencing

Yan Wei Jie

**Doctor of Philosophy
2023**

Comparative Analysis of Gene Expressed in Trunking and Non-Trunking
Metroxylon sagu Leaves Utilizing Transcriptome Sequencing

Yan Wei Jie

A thesis submitted

In fulfillment of the requirements for the degree of Doctor of Philosophy

(Biochemistry)

Faculty of Resource Science and Technology

UNIVERSITI MALAYSIA SARAWAK

2023

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.


.....

Signature

Name: Yan Wei Jie

Matric No.: 17010084

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Date : 24/10/2023

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to those who have contributed directly or indirectly to this study.

I'm utmost grateful to have Prof. Dr. Mohd Hasnain Md Hussain as the project leader and my supervisor who secured funding for this project and guided me throughout the study.

I would also would like to extend my thanks for the Faculty of Resource Science and Technology which provide the facility and technical supports for the completion of this study.

My sincere gratitude to the Centre for Graduate Studies, for the advice and support given during my period of study in Universiti Malaysia Sarawak.

I would like to thank the management of the Universiti Malaysia Sarawak for making it possible for me to complete my study here in Sarawak.

Special thanks to my dearest colleagues who assisted me in this project, provide me with insightful opinion and moral support.

Last but not least, my heartfelt thanks for my beloved family who provide me with living expenses and moral support so that I can have and ease of mind to complete this study.

This study will not be possible without the supports for all of this good people.

ABSTRACT

Sago palm (*Metroxylon sagu*) is an important agricultural starch-producing palm that contributes to Malaysia's economics, especially in the State of Sarawak. In this palm tree, storage-starch is located in the trunk of the plant. Under normal conditions, the sago palm develops its trunk after 4 - 5 years of being planted. However, sago palm planted on deep-peat soil failed to develop their trunk even after 17 years of being planted. This phenomenon is known as 'non-trunking', which eliminates the economic value of the palms. Numerous research has been done to address the phenomenon, but the molecular mechanisms of sago palm responding toward the responsible stresses are still lacking. Therefore, this study used the transcriptomic approach to determine differentially expressed enzymes and pathways causing the non-trunking phenomenon. A newly improved CTAB extraction protocol was used for RNA extraction, followed by next-generation transcriptome sequencing using the BGISEQ-500 platform. The sequence was de novo-assembled using TRINITY software package and the differentially expressed genes (DEGs) were identified using DESeq2 algorithm. A total of 40.11 Gb were sequenced from the sago palm leaf samples. The assembled sequence produced 102,447 Unigenes, with an N50 score of 1,809 bp and a GC ratio of 44.34%. The alignment of Unigenes with seven functional databases (NR, NT, GO, KOG, KEGG, SwissProt and InterPro) resulted in the annotation of 65,523 (63.96%) genes. A total of 936 DEGs were identified. Analysis of the DEG data and the pathways involved showed that the non-trunking phenomenon in sago palm is caused by the nutrient deficiency of nitrogen, phosphate and sulphate. The DEG analysis also indicated that nutrient starvation negatively affects the normal biological processes in the sago palm; aggravating the effect of boron and aluminium toxicity; and the malnourished sago palm is also more susceptible

to pathogenic infection; which ultimately leads to stunted growth and development of the plant.

Keywords: Transcriptome, non-trunking, *Metroxylon sagu*, RNA sequencing, differentially expressed genes

Analisis Perbandingan Gen yang Terekspres dalam Daun Metroxylon sagu Berbatang dan Tidak Berbatang Menggunakan Penjujukan Transkriptom

ABSTRAK

Pokok sagu (Metroxylon sagu) ialah tanaman pertanian yang menghasilkan kanji yang penting dan menyumbang kepada ekonomi Malaysia, terutamanya di Negeri Sarawak. Kanji yang dihasilkan oleh pokok sagu disimpan di bahagian tengah batang pokok tersebut. Dalam keadaan normal, batang pokok sagu tumbuh memanjang keatas selepas 4-5 tahun ditanam. Namun, terdapat pokok sagu yang ditanam di tanah gambut dalam tidak bertumbuh batangnya walaupun selepas 17 tahun ditanam. Fenomena ini dikenali sebagai 'Tidak Berbatang', dan ia menurunkan nilai ekonomi tanaman tersebut. Banyak kajian telah dilakukan berkaitan fenomena ini, tetapi kajian berkaitan mekanisme molekular pokok sagu yang bertanggungjawab untuk fenomena ini masih kabur. Oleh itu, dalam kajian ini, sampel daun diambil daripada pokok sagu berbatang (normal) dan tidak berbatang yang ditanam di atas tanah gambut untuk pengekstrakan RNA jumlah dari daun pokok sagu tersebut menggunakan protokol pengekstrakan CTAB yang diubahsuai, diikuti dengan penjujukan transkriptom (mRNA) generasi kedua menggunakan platform BGISEQ-500. Data mentah dibersihkan, dan susunan secara de novo dilakukan menggunakan pakej perisian TRINITY. Sebanyak 40.11 Gb telah dijujuk daripada sampel daun pokok sagu. Jujukan yang siap disusun menghasilkan 102,447 Unigenes, dengan skor N50 1,809 bp dan nisbah GC sebanyak 44.34%. Penjajaran Unigenes dengan tujuh pangkalan data bagi fungsi gen (NR, NT, GO, KOG, KEGG, SwissProt dan InterPro) memberikan anotasi kepada Unigenes sebanyak 65,523 (63.96%). Gen Pengekspresan berbeza (DEG) telah dikenal pasti menggunakan algoritma DESeq2. Sebanyak 936 DEGs pokok sagu tidak berbatang telah dikenalpasti. Analisa data DEGs menunjukkan fenomena tidak berbatang pokok sagu

adalah disebabkan oleh kekurangan zat nitrogen, fosfat dan sulfat. Analisis DEG juga menunjukkan bahawa kekurangan nutrien ini memberi kesan negatif kepada proses biologi dalam pokok sagu; memburukkan lagi kesan ketoksikan boron dan aluminium; dan pokok sagu yang kekurangan nutrien ini juga lebih mudah terdedah kepada jangkitan kuman; yang akhirnya menyebabkan pertumbuhan dan perkembangan tumbuhan pokok sagu terbantut.

Kata kunci: *Transkriptomik, tidak berbatang, Metroxylon sagu, jujukan RNA, gen pengekspressan berbeza*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvi
CHAPTER 1 INTRODUCTION	1
1.1 Study Background	1
1.2 Problem Statement	3
1.3 Objectives	4
CHAPTER 2 LITERATURE REVIEW	6
2.1 Introduction to Sago Palm	6
2.1.1 Sago Palm Taxonomy	6
2.1.2 Distribution, Adaptations and Conservation of Sago Palm	11
2.1.3 Traditional Use of Sago Palm	11
2.1.4 Sago Starch Extraction	12

2.1.5	Use of Sago Starch	13
2.1.6	Advantages of Sago Palm	13
2.2	Potential of <i>M. sagu</i> as Starch Producing Crops	14
2.3	Other Starch Producing Crops	15
2.4	Sago Palm in Malaysia	16
2.4.1	Sago Plantation in Sarawak	18
2.4.2	Research of Non-Trunking Sago Palm	20
2.5	Nucleotide Sequencing Technology	25
2.5.1	RNA Sequencing (RNA-seq)	33
2.6	RNA Extraction and Quality Assessment	33
2.7	Bioinformatics	35
CHAPTER 3	MATERIALS AND METHODS	37
3.1	Sampling	37
3.2	RNA Extraction Protocol Trial	38
3.2.1	CTAB Miniprep RNA Extraction	38
3.2.2	PCI RNA Extraction	40
3.2.3	MRIP RNA Extraction	42
3.2.4	Gel Electrophoresis	43
3.2.5	Spectrophotometry	43
3.2.6	Final RNA Extraction Protocol	44

3.3	RNA Attribute Measurement and Selection	46
3.4	RNA Sequencing	47
3.5	Quantitative Real-Time PCR (qPCR) Analysis	52
3.5.1	Gene Selections	52
3.5.2	Primer Design	52
3.5.3	Two-Step RT-qPCR	53
3.5.4	Relative Gene Expression	56
3.6	Statistical Analysis	57
3.6.1	RNA Extraction Protocol Trial	57
3.6.2	RNA Attribute Measurement and Selection	57
3.6.3	Quantitative Real-Time PCR (qPCR) Analysis	57
CHAPTER 4	RESULTS AND DISCUSSION	59
4.1	RNA Extraction Protocol Comparison	59
4.2	RNA Attribute Test	66
4.3	RNA Sequencing	69
4.3.1	Sequencing reads filtering	69
4.3.2	<i>De novo</i> assembly	70
4.3.3	Unigene Functional Annotation	71
4.3.4	Unigene Expression	73
4.3.5	Differentially Expressed Genes (DEGs) Detection	74

4.3.6	DEGs Functional Analysis by Gene Ontology (GO) of Non-Trunking Sago Palm	76
4.3.7	DEGs Pathway Analysis by KEGG.	112
4.3.8	DEGs Protein-Protein Interaction (PPI) Analysis by STRINGdb	126
4.3.9	RNA-seq Data Validation by qPCR Analysis	134
4.4	Nutrient Deficiency and Toxicity in Sago Palm	140
4.5	DEGs Comparison with Previous Transcriptome Studies in Non-Trunking Sago Palms	151
	CHAPTER 5 CONCLUSION AND RECOMMENDATIONS	153
4.1	Conclusion	153
4.2	Recommendations	154
	REFERENCES	156
	APPENDICES	180

LIST OF TABLES

	Page
Table 2.1: Taxon of <i>Metroxylon sagu</i> with Rank Description	6
Table 2.2: Morphological Characteristics of the <i>Metroxylon</i> spp.	8
Table 2.3: Other Starch Crops Starch Yield	15
Table 2.4: Comparison of Yield and Components of the <i>Metroxylon</i> spp.	17
Table 2.5: Growth Performance of Sago Palm in Shallow Peat and Deep Peat Soil 9 Years After Plantation	21
Table 2.6: Comparison of Soil Properties between Shallow Peat and Deep Peat Soil in Dalat Sago Plantation	21
Table 2.7: Comparison Between Sequencing Technology by Generation	31
Table 3.1: Sampling Time and Global Positioning System (GPS)	37
Table 3.2: CTAB Miniprep Chemicals List	38
Table 3.3: PCI Chemicals List	40
Table 3.4: MRIP Chemicals List	42
Table 3.5: Improved CTAB Extraction Protocol Chemicals List	44
Table 3.6: General Differences of Modified CTAB Method with Untergasser (2008)	45
Table 3.7: cDNA Synthesis Chemical List	54
Table 3.8: cDNA Synthesis PCR Setting	54
Table 3.9: qPCR NTC Chemical List	55
Table 3.10: qPCR Targeted Gene Chemical List	55
Table 3.11: qPCR Settings	56
Table 3.12: High Resolution Melt Setting	56
Table 4.1: Spectrophotometry Reading of the Extracted RNA Samples	59
Table 4.2: Normality Test for 260/280 Ratio, 260/230 Ratio and RNA Weight Data	61

Table 4.3:	Tukey Multiple Comparison Test of 260/280 Ratio and 260/230 Ratio Data between RNA Extraction Methods	62
Table 4.4:	Independent-Samples Kruskal-Wallis Multiple Comparison Test of RNA Weight Data between RNA Extraction Methods	64
Table 4.5:	Selected Total RNA Extracts for RNA Sequencing.	67
Table 4.6:	Clean Reads Quality Metrics	69
Table 4.7:	Quality Metrics of Transcripts	70
Table 4.8:	Quality Metrics of Unigenes	71
Table 4.9:	Annotation Summary	72
Table 4.10:	DEGs involved in Biological Processes and its Definations	77
Table 4.11:	Annotation of the DEGs involved in the Biological Processes	84
Table 4.12:	Molecular Functions Which are Significantly Different between Trunking and Non-Trunking Sago Palm, Unigenes involved and its Annotations based on Gene Ontology Analysis	95
Table 4.13:	Grouping, Annotation and Expression of DEGs which are Involved in Secondary Active Transmembrane Transporter GO-Term by Gene Ontology Analysis	103
Table 4.14:	Grouping, Annotation and Expression of DEGs Which are Involved in Ammonium Transmembrane Transporter Activity GO-Term by Gene Ontology Analysis	105
Table 4.15:	Grouping, Annotation and Expression of DEGs which are Involved in Peroxidase Activity GO-Term by Gene Ontology Analysis	106
Table 4.16:	Grouping, Annotation and Expression of DEGs Involved in Heme Binding GO-Term by Gene Ontology Analysis	109
Table 4.17:	DEGs Involved in Phenylpropanoid Biosynthesis Pathway	114
Table 4.18:	DEGs involved in Benzoxazinoid Biosynthesis Pathway	118
Table 4.19:	DEGs involved in Stilbenoid, Diarylheptanoid and Gingerol Biosynthesis Pathway	121
Table 4.20:	DEGs involved in Flavonoid Biosynthesis Pathway	123
Table 4.21:	DEGs involved in Circadian Rhythm Pathway	125
Table 4.22:	Total DEGs Protein-Protein Interaction (PPI)	127

Table 4.23: Pathway involved by the Total DEGs	128
Table 4.24: Upregulated DEGs Protein-Protein Interaction (PPI)	130
Table 4.25: Pathway involved by the Upregulated DEGs	131
Table 4.26: Downregulated DEGs Protein-Protein Interaction (PPI)	133
Table 4.27: Selected Genes for qPCR Analysis	135
Table 4.28: Primer Design of Selected Genes	136
Table 4.29: Relative Gene Expressions of Selected Genes	138
Table 4.30: Pearson's Correlation Analysis between RNA-seq Data and qPCR Data	139
Table 4.31: DEGs involved in Nitrogen Regulation	142
Table 4.32: DEGs involved in Phosphorus Regulation	144
Table 4.33: DEGs involved in Sulphur Regulation	146
Table 4.34: DEGs involved in Boron Transport	148
Table 4.35: Other DEGs involved in Boron Regulation	149
Table 4.36: DEGs involved in Aluminium Regulation	151

LIST OF FIGURES

	Page
Figure 1.1: Distribution of <i>Metroxylon</i> spp.	2
Figure 2.1: Chain Termination Method	26
Figure 2.2: Chemical Degradation Method	27
Figure 2.3: Next Generation Sequencing Workflow	28
Figure 2.4: Third Generation Sequencing Signal Detection	30
Figure 2.5: Bioinformatics Workflow of RNA Sequencing Service provided by Beijing Genome Institute	36
Figure 3.1: BGISEQ-500 Transcriptome Library Workflow	48
Figure 3.2: Bioinformatics Workflow of the Sequenced Data	50
Figure 4.1: Gel Electrophoresis for Total RNA Extract with CTAB, PCI, and MRIP	65
Figure 4.2: TBE-AGE of the 4 Batches of Total RNA Sample from 5 Trunking and 5 Non-Trunking Sago Palm Leaf Tissues	68
Figure 4.3: Distribution of the NR Database Annotated Unigenes by Species Similarity	73
Figure 4.4: Principal Component Analysis of Gene Expression between Trunking and Non-Trunking sago palm leaf	74
Figure 4.5: MA Plot of DEGs Between Trunking and Non-Trunking Sago Palm	75
Figure 4.6: Heatmap of DEGs between Samples	76
Figure 4.7: Biological Processes involving <i>Pentatricopeptide Repeat-Containing Protein (At3g42630)</i>	86
Figure 4.8: Biological Processes involved by <i>SPX Domain-Containing Protein 1 (SPX1)</i>	87
Figure 4.9: Biological Processes involved by <i>Translation Factor GUF1 Homolog, Chloroplastic (LEPA)</i>	89
Figure 4.10: Biological Processes involved by <i>MADS-box Transcription Factor 50-like (MADS50)</i>	91

Figure 4.11: Biological Processes involved by <i>Cytochrome P450 (CYPs)</i>	92
Figure 4.12: Biological Processes involved by <i>LSD ONE LIKE 1 (LOL1)</i>	94
Figure 4.13: GO Ancestor Chart of DEGs between Trunking and Non-Trunking Sago Palm Leaf Tissue based on Molecular Function	99
Figure 4.14: Phenylpropanoid Biosynthesis Pathway Map by KEGG	114
Figure 4.15: Benzoxazinoid Biosynthesis Pathway Map by KEGG	119
Figure 4.16: Stilbenoid, Diarylheptanoid and Gingerol Biosynthesis Pathway Map by KEGG	120
Figure 4.17: Flavonoid Biosynthesis Pathway Map by KEGG	122
Figure 4.18: Circadian Rhythm Pathway Map by KEGG	124
Figure 4.19: Protein-Protein Interaction Network of Total DEGs excluding DEGs without any Interactions	129
Figure 4.20: Protein-Protein Interaction Network of Upregulated DEGs excluding DEGs without any Interactions	132

LIST OF ABBREVIATIONS

A	Adenine
ACS	American Chemical Society
ADS	Acyl-Lipid Desaturase
AGE	Agarose Gel Electrophoresis
AMT	Ammonia Transmembrane Transporter
APK	APS Kinase
APR	APS Reductase
APS	Adenosine 5'-Phosphosulfate
APX	Ascorbate Peroxidases
ATP	Adenosine Triphosphate
ATPS	ATP Sulfurylase
B	Boron
B(OH) ₃	Boric Acid
BGI	Beijing Genome Institute
BOR	Boron Transporter
C	Cytosine
C4H	Cinnamic Acid 4-Hydroxylase
Ca	Calcium

CaCl ₂	Calcium Chloride
CAT	Catalases
CDS	Coding DNA Sequence
CG%	Percentage of Guanine And Cytosine In The Sequences
CHX	H(+) Antiporter
circRNA	Circular RNA
CO	CONSTANS Proteins
Co	Cobalt
cPAS	Combinatorial Probe-Anchor Synthesis
CPO	Chloroperoxidases
CPSase	Carbamoylphosphate Synthase
CRAUN	Crop Research and Application Unit
CTAB	Hexadecyltrimethylammonium Bromide
Ct-value	Cycle Threshold value
CYPs	cytochrome P proteins
ddNTP	Dideoxynucleotides Triphosphates
DEG	Differentially Expressed Gene
dH ₂ O	Distilled Water
dscDNA	Double-Strand cDNA

DTX/MATE	DETOXIFICATION/Multidrug and Toxic Compound Extrusion Proteins
EDTA	Ethylenediaminetetraacetic acid
FAO	Food and Agriculture Organization
FPKM	Fragments Per Kilobase of Transcript per Million
FT	Flowering locust T
G	Guanine
GC	Guanine/Cytosine
GEO	Gene Expression Omnibus of NCBI
GO	Gene Ontology Database by STRINGdb
GS	Glutamine Synthetase
H ₂ S	Hydrogen Sulphide
HAD	Histone Deacetylase
HCl	Hydrochloric Acid
IDT	Integrated DNA Technologies
KEGG	Kyoto Encyclopedia of Genes and Genomes Database
KOG	Eukaryotic Orthologous Groups of Proteins Database by HSLs
LCDA	Land Custody and Development Authority
LiCl	Lithium Chloride
lncRNA	Long Noncoding RNA

LSU	Response to Low Sulfur
<i>M. amicarum</i>	<i>Metroxylon amicarum</i>
<i>M. paulcoxii</i>	<i>Metroxylon paulcoxii</i>
<i>M. sagu</i>	<i>Metroxylon sagu</i>
<i>M. salomonense</i>	<i>Metroxylon salomonense</i>
<i>M. vitiense</i>	<i>Metroxylon vitiense</i>
<i>M. warburgii</i>	<i>Metroxylon warburgii</i>
Mg	Magnesium
MgCl ₂	Magnesium Chloride
MgSO ₄	Magnesium Sulphate
MIP	Major Intrinsic Protein
miRNA	MicroRNA
mRNA	Messenger Ribonucleic Acid
N	Nitrogen
N50	Sequence Length Median
N70	Highest Sequence Length Value of The Lowest 30% of the Sequence Population by Sequence Length
N90	Highest Sequence Length Value of The Lowest 10% of the Sequence Population by Sequence Length
NaCl	Sodium Chloride

NADPH	Reduced Nicotinamide Adenine Dinucleotide Phosphate
NaOAc	Sodium Acetate
NCBI	National Center for Biotechnology Information
NF-YC	Nuclear Transcription Factor Y Subunit C
NGS	Next Generation Sequencing
NIP	Nodulin-26 Like Intrinsic Protein
NiR	Nitrite Reductase
NR	Nitrate Reductase
NR	Non-Redundant Protein Sequence Database of NCBI
NRT/PTR	Nitrate Transporter/Peptide Transporter
NT	Nucleotide collection database of NCBI
NTC	No-Template Control
OD	Optical Density
Oligo(dT)	Deoxy-Thymidine Nucleotides
ORF	Open Reading Frame
P	Phosphate
PAGE	Polyacrylamide Gel Electrophoresis
PAPS	3'-Phosphoadenosine 5'-Phosphosulfate
PCA	Principal Component Analysis

PCI	Phenol-Chloroform-Isoamyl alcohol
PCR	Polymerase Chain Reaction
PER	Peroxidases
PHO	Phosphate Transporter
PHR	Phosphate Starvation Response Protein
PHT	Phosphate Transporter
PHT	Inorganic Phosphate Transporter
Pi	Inorganic Phosphate
PIP	Plasma Membrane Intrinsic Protein
piRNA	Piwi-Interacting RNA
PPI	Protein-Protein Interaction
PRR	Pseudo-Response Regulator
PSR	Pi starvation responses protein
P-value	Probabily Value
PVP	Polyvinylpyrrolidone
Q20	Equivalent to The Probability of an Incorrect Base Call 1 in 100 Times
Q30	Equivalent to The Probability of an Incorrect Base Call 1 in 1000 Times
qPCR	Quantitative Real-Time PCR

RBOH	Respiratory Burst Oxidase Homolog Protein
RG-II	Rhamnogalacturonan-II
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
rRNA	Ribosomal RNA
RT	Reverse Transcription
S ²⁻	Sulphide
SDS	Sodium Dodecyl Sulfate
SIP	Small Basic Intrinsic Protein
SIR	Sulfite Reductase
SMRT	Single-Molecule Real-Time Technology
snoRNAs	Small Nucleolar RNAs
snRNAs	Small Nuclear RNAs
SO ₂	Sulphuric gas
SO ₃ ²⁻	Sulphite
SO ₄ ²⁻	Sulphate

SOC	Suppressor of CO
sp.	Species
spp.	All the Member of the Species
SRA	Sequence Read Archive
SSR	Short Sequence Repeat
Std	Standard Deviation
SULTR	Plant Sulphur Transporter
T	Thymine
TANF	Total Aerobic Nitrogen Fixer
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TF	Transcription Factor
TIP	Tonoplast Intrinsic Protein
TRIS	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TRVAB	Total Recoverable Viable Aerobic Bacteria
tSMS	True Single-Molecule Sequencing
U	Uracil
UV	Ultra Violet

XIP	X Intrinsic Protein
ZMW	Zero Mode Waveguide Well
Zn	Zinc
ΔC_t	Relative Gene Expressions within Sample
$\Delta\Delta C_t$	Relative Gene Expressions between Samples

CHAPTER 1

INTRODUCTION

1.1 Study Background

Sago palm (*Metroxylon sagu*) is one of the starch-producing crops and is the best starch producing species within the *Metroxylon* spp. which could be found from southern Thailand to the Solomon islands (Ehara, 2018; Ellen, 2006) (**Figure 1.1**). Among the *Metroxylon* spp. only *M. sagu* is both hapaxanthic (once-flowering) and soboliferous (produce suckers) (Flach, 1997). *M. sagu* lifecycle is divided into four stages which is the rosette, bole forming, inflorescence and fruit ripening stages (Flach, 1997). The starch accumulation in the trunk starts at the bole formation stage and reaches its peak at the beginning of the inflorescence stage which is approximately nine years after plantation (Flach, 1997). The advantages of sago palm are its ability to adapt to acidic, saline soil with seasonal flooding such as peat swamp which other crops are unable to grow (Bintoro *et al.*, 2018).

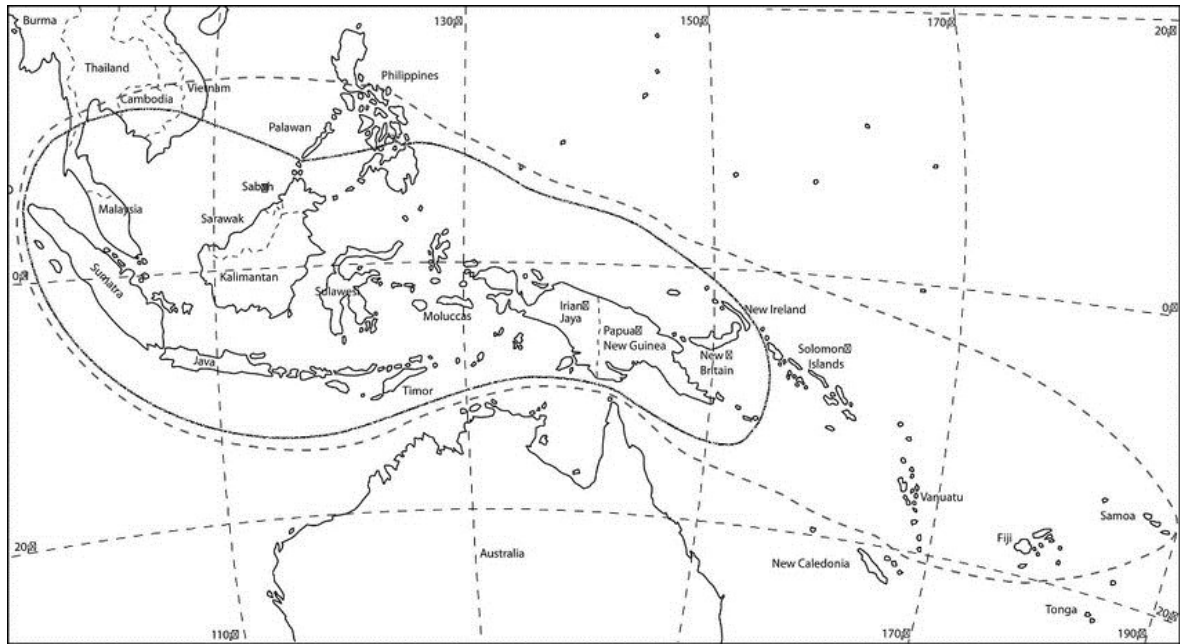


Figure 1.1: Distribution of *Metroxylon* spp.

Description: *M. sagu* are distributed from southern Thailand to the Solomon islands along the equator.

Keys: The dotted line encapsulation represents the distribution of *Metroxylon* spp. while the continuous line encapsulation represents the *M. sagu* distribution.

Source: (Ellen, 2006)

Peat soil in Malaysia consists of 2,457,730 ha (7.45%); of which Sarawak state has 1,697,847 ha (13.64%); comprising 100,418 ha (5.91%) of shallow peat soil and 1,597,429 ha (94.09%) of deep peat soil (Wetlands International, 2010). The ability of sago palm to grow in peat swamp catches Sarawak state interest to utilise its non-arable land for agriculture purposes. The Sarawak government's effort in promoting the sago industry started as early as the year 1982 by intensifying research and development by the Sarawak Department of Agriculture and the world's first commercial deep peat sago plantation was developed in Mukah district by the Land Custody and Development Authority (LCDA) of Sarawak. This effort increases the number of sago palm plantation in Sarawak, thus increases

the sago starch and flour yield per annum which ultimately lead to the increase in the export revenue of the state. In the year 2021, the export of sago starch and flour in Sarawak worth RM 80.4 million (Sarawak State Secretary Office, 2022b).

1.2 Problem Statement

In the deep peat sago plantation located in the Mukah district, there are occurrences of *M. sagu* clusters that have not developed their trunks even after more than seventeen years. This is in contrast to the normal growth pattern where sago palms typically develop their trunks within five years of being planted. This condition is referred to as "non-trunking" sago palm. (Liew, 2004; Lulie, 2000). This phenomenon occurs to sago plantation in deep peat soil which reduces the overall yield of sago log per hectare which negatively effect the stable supply of sago log in the market and the income of the farmers. The fluctuating sago log supply reduces sago starch production, raising production costs and increasing the selling price which is 35% and 20% higher than cassava and corn starches respectively (Jong, 2018). As a result, sago starch becomes less competitive compared to other starches in the market.

Numerous studies were performed to address the non-trunking sago palm problem such as soil physicochemical properties (Liew, 2004; Lulie, 2000), soil microbiome (Apun *et al.*, 2009) and molecular studies (Edward-Atit, 2013; Hussain *et al.*, 2020; Hussain *et al.*, 2012; Kamal, 2014; Kamarol, 2015; Lim *et al.*, 2021). The general outcome of the studies revealed that the mineral deficiency causes the non-trunking in sago palm, but how this deficiency affects sago development remains unanswered. Previous transcriptomic research in non-trunking sago palm utilised the differential display (DD) and representational difference analysis (RDA) to elucidate the differentially expressed genes (DEGs) (Edward-

Atit, 2013; Kamarol, 2015). The limitations of their studies are constricted selection of DEGs based on either the core sequence tested as in DD or mRNA abundance as in RDA. Other than that, the DEGs determination is based on a single reference and a target sample without replications. Their studies found only less than 20 functional annotated DEGs, which may not be sufficient to thoroughly overview the gene expression responsible for the non-trunking phenomenon in sago palm. High-throughput RNA sequencing is able to generate a complete transcriptome profile of any condition/treatment which gives a better transcript coverage compared to the previous studies (Bansal *et al.*, 2018). In conjunction with those studies, this study utilises RNA sequencing for transcriptome analysis to compare the gene expression between the trunking and non-trunking sago palm leaf tissue and highlight the DEGs that correlate with the non-trunking phenomenon in sago palm. The hypothesis if this study is the DEGs obtain by transcriptomic analysis able to explain the non-trunking phenomenon in sago palm.

1.3 Objectives

This study aims to uncover the molecular mechanisms and pathways which correspond to the non-trunking phenomenon in sago palm. This can be achieved through the following set of objectives:

1. To extract adequate quality and quantity of total RNA from sago palm leaf tissue for sequencing purpose;
2. To identify the differentially expressed genes (DEGs) in leaf tissue between the trunking (normal) and non-trunking (target of interest) sago palm;

3. To correlate the DEGs functions and pathways with the cause and effect of non-trunking phenomenon in sago palm.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to Sago Palm

2.1.1 Sago Palm Taxonomy

The word “*Metroxylon*” is a combination of two Greek words namely: “μήτρα” pronouncing “mi.tra”, meaning pith; and the word “ξύλο” pronouncing “ksilo”, meaning wood which implies a woody plant with spongy parenchyma cells storing nutrients, in this case, starch (Dransfield *et al.*, 2008). The taxa of plants were divided into fourteen ranks from the highest which is kingdom, subkingdom, superdivision, division, class, subclass, order, family, subfamily, tribes, subtribes, genus, section and species (**Table 2.1**).

Table 2.1: Taxon of *Metroxylon sagu* with Rank Description

Description: *M. sagu* can be differentiate from the rest of the *Metroxylon spp.* by 18 rows of scale on the fruits and the plant which is both hapaxanthic and soboliferous.

Rank	Scientific Name	Description
Kingdom	<i>Plantae</i>	Plants ¹
Subkingdom	<i>Tracheobionta</i>	Vascular plants ¹
Superdivision	<i>Spermatophyta</i>	Seed plants ¹
Division	<i>Magnoliophyta</i>	Flowering plants ¹
Class	<i>Liliopsida</i>	Monocotyledons ¹
Subclass	<i>Arecidae</i>	Broad net-veined petiolate leaves with numerous small flowers which are usually arranged on a spathe to form inflorescences ²
Order	<i>Arecales</i>	Woody plant ³

Table 2.1 continued

Family	<i>Arecaceae</i>	Palm family ¹
Subfamily	<i>Calamoideae</i>	Ovary and fruit covered in imbricate scales ⁴
Tribe	<i>Calamaeae</i>	Flowers not in bracts, petal not sharply pointed, less than 12 stamens, fruit lack endocarp, leaf pinnate, spiny regular or sub-regular organised in full or partial whorls ⁴
Subtribe	<i>Metroxylinae</i>	Massive tree, hermaphroditic or polygamous, pyramidal stigmas, and leaflet linear to lanceolate, acute, entire ⁴
Genus	<i>Metroxylon</i> Rottb.	Woody palm with starchy pith ⁴
Section	<i>Eumetroxylon</i> Bec.	18 rows of scales on the fruit ⁵
Species	<i>Metroxylon sagu</i> Rottb	Hapaxanthic and soboliferous ⁵

Sources:¹(USDA & NRCS, 2022), ²(Dictionary of Botany, 2003), ³(Hodgkiss, 2016),
⁴(Dransfield *et al.*, 2008), ⁵(Flach, 1997)

Genus *Metroxylon* are divided into two sections namely; *Eumetroxylon* and *Coelococcus* distinguished by the number of scales on their fruits, 18 rows and 24 - 29 rows respectively (Flach, 1997). *Eumetroxylon* consist of only single species, *M. sagu* while *Coelococcus* spp. consist of five species, *Metroxylon amicarum*, *Metroxylon vitiense*, *Metroxylon salomonense*, *Metroxylon warburgii*, and *Metroxylon paulcoxii* (Ehara, 2018) (Table 2.2)

Table 2.2: Morphological Characteristics of the *Metroxylon* spp.

Description: Characteristics of *Metroxylon* spp. which differentiate them.

Species	Distribution	Characteristics
<i>M. sagu</i> Rottb. (Sago palm) 	Thailand, Malaysia, Brunei, Philippines, Indonesia, Papua New Guinea, and Solomon Islands	Clump forming. Spineless or spiny, with varying spine length and density. Fruit is covered with 18 rows of vertically arranged scales, 5 cm in diameter
<i>M. amicarum</i> (H. Wendl.) Becc. (Caroline ivory nut palm) 	Caroline Islands and Micronesia	Lateral inflorescence. First-order branches emerge at leaf axils. Large globose fruit, 26–29 rows of scales, 7–10.5 cm in diameter, slightly flat at top, no dent at base

Table 2.2 continued

<i>M. vitiense</i> (H. Wendl.) H. Wendl. ex Benth. & Hook. f. (Fijian sago palm) 	Fiji Islands	Second-order branches on inflorescence droop and are very short (18 cm). Spines on petiole and rachis are black. Globose fruit, 27–31 rows of scales, conical with round base and wide top, 5.5–6.6 cm in length, 4.5–7 cm in diameter
<i>M. salomonense</i> (Warb.) Becc. (Solomons' sago palm) 	Solomon Islands, Northeastern part of New Guinea Island, New Britain Island, and Northern islands in Vanuatu	Inflorescence shows two or three orders of branching. Rachillae arising as both second-order branching on first-order branch of inflorescence and third-order branching after second-order branching Third-order branches (rachillae) on inflorescence droop and are very long (20–26 cm), pubescence on the adaxial side of all pedicel bracts. Spines on petiole and rachis are soft. Globose fruit, 27–28 rows of scales, 7 cm in diameter. Seed diameter: 4 cm; pericarp thickness, 5–6 mm

Table 2.2 continued

<i>M. warburgii</i> (Heim) Becc. (Vanuatu sago palm) 	Vanuatu, Fiji (Rotuma Is.), Samoa	All branches on inflorescence arising upright. No pubescence on the adaxial side of all pedicel bracts. Marked cuticle formation on the abaxial side of leaflet. Pyriform fruit, 24–28 rows of scales, 7–10 cm in length, 4–9 cm in diameter
<i>M. paulcoxii</i> McClatchey (<i>M. upoluense</i> Becc.) 	Samoa	Inflorescence shows two to three orders of branching, rachillae arising non-uniformly (drooping or upright). Second-order branches arising between the center and tip of a first-order branch form rachillae, which droop, and third-order branches arising between the center and base of a second-order branch form rachillae, which grow upright (two types of branching pattern coexist at the central segment). Small fruit narrowing toward base, length 3.3 cm, diameter 2.5 cm, 24 rows of scales

Source: (Ehara, 2018)

2.1.2 Distribution, Adaptations and Conservation of Sago Palm

Distribution of sago palm (*Metroxylon* spp.) are along 90°E - 180°E, 10°N - 10°S along the equator (Bintoro *et al.*, 2018). Sago palm capable to grow on various type of soil such as sulfaquents and sulfaquepts (sulfidic soil); sulfihemists (sulfuric soil); hidraquents (waterlogged); tropaquents (tropical climate); fluvaquents (alluvial); psammaquents (sandy soils); tropofibrists and tropohemists and troposaprists (peats); volcanic; podzolic; and grumusol (Bintoro *et al.*, 2018; Okazaki & Sasaki, 2018). Sago palm can be found up to 1200 m elevation sea level but its optimum growth is at area below 400 m elevation sea level (Bintoro *et al.*, 2018). The humidity and sunlight required for optimal growth are 90% and 900 J/cm²/day respectively (Bintoro *et al.*, 2018). The rainfall required for optimal growth of sago palm is 2000 mm/year and wet month need to be more than nine months while dry month have to be less than two months (Bintoro *et al.*, 2018). Sago palm grow in flood soil utilising its pneumatophores roots for gases exchange and able to tolerate mild salinity (less than 200 mM NaCl) enabling it to grow near ocean swamp (Bintoro *et al.*, 2018; Okazaki & Sasaki, 2018). Sago palm can also tolerate low pH soil enabling it to grow in peat soil. The roles of sago palm in environment conservation are: 1) to prevent soil subsidence where it can grows on peat soil or swampy soil without the need to drain out the water from the soil and thus preventing the release of carbon dioxide and methane by the aerobic decomposition of the organic compound in the soil 2) to retain the water in the peat soil while retaining the natural habitat of peat soil for other organisms and 3) to absorb carbon dioxide and hence reducing greenhouse gases emission (Bintoro *et al.*, 2018).

2.1.3 Traditional Use of Sago Palm

Commonly, the sago palm is used by the local communities as a food source. The sago palm contains starch in its trunk which is edible but more commonly processed into

sago starch flour before it is used to make other delicacies. Sago starch is used as the main carbohydrate source in sago palm growing communities before the introduction of other major food crops such as paddy (Nishimura, 2018; Wardis, 2018). The second common use of sago palm is as thatch material. The sago leaf could be woven to make roofs and walls for houses, leaf sheets can be used as fences, and the bark can be used to make floors or fences (Pue *et al.*, 2018). The sago palm frond can be used to make utensils where its leaflet can be woven to make baskets, midrib can be tie up as broom, and vascular bundle at the base of leaf sheet can be make into blowpipe darts (Pue *et al.*, 2018). The sago starch filtrate can be used to grow mushrooms, as animal feed, or to breed grubs larvae, which is considered local delicacy (Pue *et al.*, 2018). Sago starch was also used in bartering and ceremonial occasions in some Papua New Guinea communities. The sago starch is used as one of the wedding gifts from the groom's family to the bride's family as tradition in Melanesia communities, in ritual for health, luck and good harvest in the coming years in Kiwai community, as offering to keep the bad spirits away by Toradja community, and guest offering in funeral in Arapesh community. Sago starch is used as barter commodity by Gulf community to exchange for earthenware with Motu community (Toyoda, 2018).

2.1.4 Sago Starch Extraction

Traditional sago starch processing involves sago palm felling, transportation, debarking and crushing usually conducted using simple tools such as a knife and axe. Then the starch paste will be washed, filtered, kneaded, and trampled before storing (Darma, 2018; Flach, 1997; Jong, 2018). More advanced methods of extraction using mechanical instruments in sago starch factories are used to improve the amount and quality of the sago starch product. The modern factory is able to produce 500 – 1000 t/month of dry starch (Jong, 2018). The processes involved debarking of sago log using bark scrapper or ring

debarker, pith milling with steel nail drum rasper, starch extraction using stainless steel rotary sieves, slurry concentration with sedimentation tank, starch refining with troughs with slurry flow control and crusher, and starch drying using centrifugation and under sunlight (Jong, 2018).

2.1.5 Use of Sago Starch

Sago starch can be used in the production of food-based products such as bread, fortified sago, crackers, cake, jelly, puddings, thickeners, noodles, biscuits, cookies, sago pearl, and table sugar. Non-food-based uses of sago starch such as binders for charcoal and tablet, biodegradable polymer in the production of biodegradable plastics, textile and paper sizing, hydrogels, kojic acid, lactic acid, fermentable sugars and ethanol (Bujang, 2018; Hirao *et al.*, 2018; Singhal *et al.*, 2008).

2.1.6 Advantages of Sago Palm

There are four reasons why the sago palm was used as a local food crop before the introduction of other major commercial food crops and is still used as a staple food in certain communities in southeast Asia and Oceania; they are: 1) The availability of sago palm for harvesting throughout the year, unlike other seasonal food crops give the advantage of food security for the community at any time; 2) The ability of sago palm to grow on peat and swampy land and is more resistant to pests and infections compared to other food crops which gives the advantages of non-arable land utilisation and less manpower needed for crop management; 3) High yield of starch per plant and less manpower required for harvesting and preparation which give a good calories per manpower harvesting value of the plant; 4) The storing period of traditionally refined sago starch is up to several months which is a huge advantage for food crops (Wardis, 2018).

2.2 Potential of *M. sagu* as Starch Producing Crops

Food and Agriculture Organization (FAO) projected that the demand for food will increase towards the year 2050 due to the increase in human population. To meet the increasing food demand, food production needs to be increased by 90% reaching the year 2050. The required increase is based on the increasing human capital and the fluctuation of current agriculture yield due to the climate change. Excessive research on the major crops were done to increase its productivity and not much could be improved to further increase their yields. In order to further increase the productivity of the major crops, new agriculture land need to be exploited. Increase demand for food exert pressure on arable land-use for agriculture. Arable land constrains happened in most of the county in the world for most of the spare arable land concentrated only in few countries. Since the potential of the current major crops and arable land cannot be further exploited economically, another alternative where exploitation of underutilised potential crops which can be grown on non-arable land are on demand.

M. sagu is one of the potential underutilised crops which can be exploited to increase food production. The advantages of sago palm farming are: 1) Sago palm can be grown on soil (peat, swampy) that no other crops can grow economically (Bintoro *et al.*, 2018; Flach, 1997; Okazaki & Sasaki, 2018); 2) The sustainable yield of sago starch is around 10 matrix ton per hectare which is the highest among current starch producing crops (Jong, 2018); 3) no re-plantation and little maintenance are needed in sago palm plantation which reduces manpower and expensive machinery needed (Bujang, 2018; Flach, 1997; Wardis, 2018); 4) multiple uses of the sago palm, from leaf to the bark of the plant is usable (Toyoda, 2018); 5) Environmental friendly because sago palm plantation can be done with little disturbance

on the peat soil environment (Wardis, 2018); 6) Food security as the sago starch can be consumed during food shortage crisis (Flach, 1997; Pue *et al.*, 2018; Wardis, 2018).

2.3 Other Starch Producing Crops

Other common starch crops such as barley (*Hordeum* spp.), cassava (*Manihot esculenta*), maize (*Zea mays*), potatoes (*Solanum tuberosum*), paddy (*Oryza* spp.), rye (*Secale cereale*), sorghum (*Sorghum* spp.), sweet potatoes (*Ipomoea batatas*), wheat (*Triticum* spp.), and yams (*Dioscorea* spp.). Starch yield of the starch crops are based on the average yield of the crops, starch content and its recovery efficiency (**Table 2.3**).

Table 2.3: Other Starch Crops Starch Yield

Description: Average starch yield per hectare of land from other starch crops are from 0.06 t/ha to 3.98 t/ha.

Starch Crops	Average Yield (t/ha) ¹	Starch Content (%)	Starch recovery (%)	Starch Yield (t/ha)
barley (<i>Hordeum</i> spp.)	3.0432	58.1 - 72.2 ²	70 - 84 ¹²	1.84564
cassava (<i>Manihot esculenta</i>)	10.7163	25 ³	67 - 72 ¹³	1.928934
maize (<i>Zea mays</i>)	5.7547	70 ⁴	68.92 ¹⁴	2.776297
potatoes (<i>Solanum tuberosum</i>)	21.7688	12 - 16 ⁵	70 ¹⁵	2.438106
paddy (<i>Oryza</i> spp.)	4.6089	81 - 93 ⁶	85 ¹⁶	3.643335
rye (<i>Secale cereale</i>)	3.3781	57 - 66 ⁷	50 ¹⁷	1.114773
sorghum (<i>Sorghum</i> spp.)	1.4585	68 - 75 ⁸	40 - 55 ⁸	0.601631
sweet potato (<i>Ipomoea batatas</i>)	12.0922	14.2 - 34.1 ⁹	47.6 - 96.5 ⁹	3.97912
wheat (<i>Triticum</i> spp.)	3.4744	60 - 75 ¹⁰	60 - 75 ¹⁸	1.95435

Table 2.3 continued

yams (<i>Dioscorea</i> spp.)	8.4732	10 - 25 ¹¹	91.2 ¹¹	1.93189
-------------------------------	--------	-----------------------	--------------------	---------

Source: ¹(Food and Agriculture Organization of the United Nation, 2022), ²(Asare *et al.*, 2011), ³(Kaur *et al.*, 2016), ⁴(Cui *et al.*, 2014), ⁵(Robertson *et al.*, 2018), ⁶(Omar *et al.*, 2016), ⁷(Deleu *et al.*, 2020), ⁸(Udachan *et al.*, 2012), ⁹(Mahfuzur Rahman *et al.*, 2003), ¹⁰(Šramková *et al.*, 2009), ¹¹(Zhu, 2015), ¹²(Sharma & Tejinder, 2014), ¹³(Marder *et al.*, 1996), ¹⁴(Zhang *et al.*, 2005), ¹⁵(Hameed *et al.*, 2015), ¹⁶(Lumdubwong & Seib, 2000), ¹⁷(Autio & Eliasson, 2009), ¹⁸(Tian *et al.*, 2008)

There are other extraction methods which reported higher starch recovery such as utilising enzyme treatment to improve starch separation (Lumdubwong & Seib, 2000), increasing in blending time (Mahfuzur Rahman *et al.*, 2003), and increasing the steeping time (Pérez *et al.*, 2003) but is not economical comparing the additional starch recovery with the cost increment.

2.4 Sago Palm in Malaysia

M. sagu is the only *Metroxylon* spp. in Malaysia which are locally known as *pokok sagu* and (Flach, 1997). *M. sagu* is the best starch producing species in *Metroxylon* genus (**Table 2.4**). This palm tree is semi-cultivated in Sarawak state (51,518 ha at the year 2015) and the production of sago starch in the year 2021 worth RM 80.1 million (Sarawak State Secretary Office, 2022a, 2022b). Sago palm in Malaysia consists of only one species which is *M. sagu* which is the best starch producer among the other *Metroxylon* spp. (Ehara, 2018) (**Table 2.3**). Yield of sago palm dried starch is around 10 t/ha per annum after eight years being planted (Jong, 2018).

Table 2.4: Comparison of Yield and Components of the *Metroxylon* spp.

Description: *M. sagu* produces the highest starch yield among the *Metroxylon* spp.

Species (Survey location)	Trunk Length (m)	Trunk Diameter (cm)	Pith Density (g/m ³)	Pith Dry Matter Content (%)	Pith Dry Matter Weight (kg)	Pith Starch Content (%)	Pith Total Sugar Content (%)	Starch Yield (kg/plant)
<i>M. sagu</i> (Indonesia)	8.6	45.2	0.770	41.1	413.9	77.1	4.9 ^a	309.8
<i>M. salomonense</i> (Vanuatu)	8.5	58.0	0.850	18.5	326.0	48.9	15.3	159.4
<i>M. warbugii</i> (Vanuatu)	5.5	35.0	0.902	35.7	156.5	34.8	13.5	48.6
<i>M. amicarum</i> (Micronesia)	10.7	44.9	0.794	16.0	179.4	38.8	10.0	71.8
<i>M. vitiense</i> (Fiji)	7.6	36.7	0.894	20.8	156.1	26.7	8.4	49.4

Keys: ^a Only data of pith total sugar content was from Papua New Guinea (mean value of two folk varieties used for food).

Source: (Ehara, 2018)

2.4.1 Sago Plantation in Sarawak

In Sarawak, the sago palm plantation is mainly semi-cultivated small plantations, mostly managed by smallholder (Jong, 2018). The sago plantation mostly in peat soil, located at Dalat, Mukah (Jong, 2018). Difference in crop management applied in the plantations results in varying crops density causing suboptimal starch yields in same plantation (Jong, 2018). In Sarawak state, there are 1,697,847 ha (13.64%) peat land from the 12,398,500 ha total land area, comprising of 100,418 ha (5.91%) shallow peat soil and 1,597,429 ha (94.09%) deep peat soil (Wetlands International, 2010). Sago palm's ability to thrive in peat soil is advantageous for agriculture and non-arable land-use, particularly in deep peat soil, which makes up the majority of the peat land in Sarawak, where other crops are not suitable to grow in deep peat soil (Flach, 1997).

Despite the advantages of sago palm, there are several challenges in the sago industry. One of them is the long plant maturation period where mineral soil requires around eight years, shallow peat soil requires around nine to eleven years and deep peat soil requires more than eleven years to mature. The productivity of sago palm in peat soil is low, particularly in deep peat soil, which produces less than 2 t/ha whereas shallow peat produces 9 t/ha, comparing to 15 t/ha of sago starch produced in mineral soil (Yong *et al.*, 2018). Sago starch price in the starch market is not competitive where the price is about 35% higher than the cassava starch and about 20% higher than corn starch (Jong, 2018). Higher starch price discouraged the starch buyer in choosing sago starch as the ingredient for their production. The higher price of sago starch are resulted by the increased of production cost due to the inconsistent supply of sago log and the high sago log price which cause by the low supply over demand of sago log in the market (Jong, 2018). Land-use for sago plantation area is also a challenge where mostly peat soil and swampy area soil are allocated for sago palm

plantation by the Sarawak government and these increase logistic and infrastructure costs (Jong, 2018; Yong *et al.*, 2018). Land ownership and compensation is hard to deal with since the land use in Sarawak mostly are owned by different tenants in small sizes obscuring medium-scale or large-scale sago plantation (Jong, 2018). Because of the aforementioned challenges, investors are hesitant to invest in sago industries.

In the 1980s, the Sarawak government initiated and promoted sago palm plantation to support the growing demand of the sago industries while utilising the peat soil land located in the Mukah division which are not suitable to produce other crops (Jong, 2018; Yong *et al.*, 2018). The observation of the performance was done on the sago palm plantation over a period of 15 years and the outcome was not as expected and hence the Crop Research and Application Unit (CRAUN), which the Sarawak government funded, conducted detailed studies on the deep peat sago plantation in 2003. Two study locations were selected: 1) Sebakong Sago Plantation (SPP) in Mukah; and 2) Sungai Talau Research Station (STRS) in Dalat.

The findings from CRAUN were that the growth of the active roots in sago palm at the second year of plantation are located 1 m width and 0 – 30 cm depth where at the fourth year of plantation located 2 m width and 0 – 30 cm depth from the sago palm (Yong *et al.*, 2018). Leaf numbers in the rosette stage in deep peat soil are comparable with mineral soil but the number of leaves of deep peat sago palm were lesser than the leaves of mineral soil sago palm at bole forming stage (Yong *et al.*, 2018). Soil content of deep peat from 0 – 20 cm depth consists of sapric peat while more than 20 cm depth consist of hemic peat (Yong *et al.*, 2018). The density of deep peat soil are very loose (0.15 g/m^3) and pH of the deep peat soil range from 3.0 to 4.0 (Yong *et al.*, 2018). Sago palms mature in deep peat soil, shallow

peat soil, and mineral soil after more than 11 years, 10 years, and 9 years, respectively, with annual sago starch yields of less than 2 t/h, 9 t/h, and more than 15 t/h (Yong *et al.*, 2018). From the economic point of view, the cost requires to develop the land for sago plantation in deep peat soil, shallow peat soil and mineral soil are RM 30,800 per hectare, RM 26,800 per hectare and RM 23,600 per hectare respectively with a payback period of undefined, 14 years and 11 years respectively (Yong *et al.*, 2018). Cost of starch production from deep peat soil, shallow peat soil and mineral soil are RM3520, RM1020 and RM700, respectively, resulting in the profitability of sago plantation in deep peat soil being unmeasurable due to uncertain sago palm maturation time, 13.1% for shallow peat and 23.4% in mineral soil (Yong *et al.*, 2018).

2.4.2 Research of Non-Trunking Sago Palm

The average trunk formation growing stage are at the fifth year after plantation (Flach, 1997). Sago palm which are planted on deep peat soil and do not produce trunks even after five years are known as non-trunking sago palm. Research has been done to address the non-trunking sago palm problem in Dalat Sago Plantation such as on the physical and chemical properties of the soil, microorganisms in soil and mineral uptake of the sago palm (Apun *et al.*, 2009; Sim *et al.*, 2015; Wong, 2004). The growth of sago palm at the nine years after plantation in deep peat soil shows a non-trunking phenomenon. (**Table 2.5**) (Sim *et al.*, 2015). Soil property parameters of the deep peat soil such as: degree of humification, and water table which is lower: pH, bulk density, total ash content, conductivity, total nitrogen, available phosphorus, and total potassium, are significantly lower with 95% confidence than those in shallow peat soil (**Table 2.6**) (Apun *et al.*, 2009; Liew, 2004; Sim *et al.*, 2015; Wong, 2004). All of this suggested that the nutrients content in the shallow peat soil is significantly better than that of deep peat soil. The higher water table level aggravates

mineral leaching in deep peat soil while the lower humification of shallow peat soil are due to the new input of plant materials from the sago harvesting in shallow peat soil (Sim *et al.*, 2015).

Table 2.5: Growth Performance of Sago Palm in Shallow Peat and Deep Peat Soil 9 Years After Plantation

Description: Non-trunking sago palm growth are stunted compared to normal sago palm.

Growth Performance	Trunking (Shallow Peat)	Non-Trunking (Deep Peat)
Diameter at breast height (cm)	53.4 ± 2.7	No Trunk Formed
Height (m)	15.4 ± 1.5	4.3 ± 1.1
Number of suckers	8.0 ± 2.4	4.0 ± 1.5
Length of longest frond (m)	6.8 ± 0.5	2.4 ± 0.7

Keys: Sample size, n = 15

Source: (Sim *et al.*, 2015)

Table 2.6: Comparison of Soil Properties between Shallow Peat and Deep Peat Soil in Dalat Sago Plantation

Description: There are significant differences in soil properties between shallow peat and deep peat soil from Dala Sago Plantation.

Parameters	Shallow Peat	Deep Peat	p value
pH*	3.3 - 3.7	3.2 - 3.7	0.02
Bulk density* (g cm ⁻³)	0.12 ± 0.07	0.11 ± 0.03	0.01
Total ash content* (g kg ⁻¹)	245.0 ± 158.0	38.0 ± 17.0	2.7x10 ⁻⁵
Rubbed fibre content* (g kg ⁻¹)	218.0 ± 67.0	173.0 ± 56.0	0.07
Cation exchange capacity* (meq 100g ⁻¹)	79.5 ± 13.6	79.3 ± 13.4	0.96
Total nitrogen* (g kg ⁻¹)	28.0 ± 7.6	3.8 ± 2.2	3.1x10 ⁻¹²

Table 2.6 continued

Available phosphorus* (mg kg ⁻¹)	45.3 ± 12.4	24.6 ± 10.9	4.0x10 ⁻⁵
Total potassium* (mg kg ⁻¹)	100.2 ± 34.5	48.5 ± 12.0	7.3x10 ⁻⁶
Conductivity* (dS/m)	0.0915	0.0544	1.0x10 ⁻⁴
Degree of humification*	H4 - H5	H5 - H6	No Data
Water table* (cm)	20 - 30	40 - 50	No Data
Total recoverable viable aerobic bacteria (TRVA)** (cfu/g)	9.8x10 ⁵	7.2x10 ⁵	0.20
Total aerobic nitrogen fixers (TANF)** (cfu/g)	9.1x10 ⁵	8.2x10 ⁵	0.631

Keys: *Sample size, n = 45, **Sample size, n = 6,

Source: (Apun *et al.*, 2009; Liew, 2004; Sim *et al.*, 2015; Wong, 2004)

CRAUN had also tested different formulation of fertilisers feeding into the selected deep peat soil sago plot (Yong *et al.*, 2018) (**Table 2.4**). The growth of the sago palm during rosette stage were similar between sago palm planted on mineral soil, shallow peat soil and deep peat soil but the health of the sago leaves planted on deep peat soil deteriorate on the 5th year (Yong *et al.*, 2018). Majority of the sago palm trees that were treated with the fertiliser in deep peat soil showed trunk development (80%), but starting from year 6, the number of leaves on their crowns significantly decreased compared to the sago palm trees in mineral soil. (Yong *et al.*, 2018). The test also found that the growth retardation of peat sago palm is due to the lack of essential nutrients such as; phosphate (P, - 0.01%), potassium (K, - 0.08%), magnesium (Mg, - 0.01%), manganese (Mn, - 38.68 ppm), iron (Fe, - 12.79 ppm), Zinc (Zn, - 0.92 ppm), boron (B, - 1.12 ppm) and copper (Cu, - 0.29 ppm) in the sago palm leaf at year 6 (trunk formation stage) (Yong *et al.*, 2018). Even applying fertiliser in the deep peat soil, only 36% of the sago palm reached a harvestable stage at year 11 and the starch

produced was, significantly lower, which was less than 20 kg per trunk compared to more than 150 kg per trunk from sago palm in mineral soil (Yong *et al.*, 2018).

Although the findings in soil properties and nutrient uptakes in sago palm leaves show differences, they are not directly relating to the cause of non-trunking because not all parameters were tested and can only lay out a general conclusion that non-trunking of sago palm planted in deep peat soil was caused by a lack of essential nutrients for growth. There is still a lack of understanding of the causes and effects of the non-trunking phenomenon in sago palm.

Further studies in the field of molecular biology were done to explain the cause and effects of the non-trunking sago palm such as in transcriptomics, proteomics and metabolomics in sago palm which shows differences in expression and highlighted several possibilities that may cause the non-trunking in sago palm (Edward-Atit, 2013; Kamal, 2014; Kamarol, 2015; Yan, 2016). Previous transcriptomics analysis between trunking and non-trunking sago palm revealed the genes that are downregulated related to: the utilisation of nitrogen (N); regulation of phosphate (P) and cobalt (Co); defence; plant stress mediation signalling; flowering; and growth (Edward-Atit, 2013; Kamarol, 2015). The upregulated genes in non-trunking sago palm are related to: active uptake of nitrogen (N); regulation of calcium (Ca); magnesium (Mg) and zinc (Zn); protein trafficking; plant growth; cell metabolism; salt and related osmotic stress response; sugar accumulation; and signalling (Edward-Atit, 2013; Kamarol, 2015). Their findings concluded that the plant were under nutrient stress which causes the growth retardation of sago palm.

Proteomics comparison between trunking and non-trunking leaf tissue shows that downregulated proteins in non-trunking sago palm are those that corresponded to glycolysis

(Phosphoglycerate kinase-like) and cell rescue/defence proteins (Kinesin-1-like protein, Esterase/lipase domain-containing protein, Hairpin binding protein, and Putative ascorbate peroxidase). Upregulated proteins identified from non-trunking sago palm vary from photosynthesis (Ribulose 1,5-biphosphate carboxylase-oxygenase), protein biosynthesis (Protein disulfide isomerase) and cell rescue/defence proteins (Inositol methyl transferase) (Kamal, 2014). This study shows that the non-trunking sago palm was under stress and conserving reserved energy but actively involved in photosynthesis and protein biosynthesis to produce the essential energy and components for survival.

Comparison between shallow peat soil sago palm with deep peat soil sago palm metabolomics expression shows that there are differences in the presence of oils, waxes, squalene, haloalkanes, sulphite ester, phosphonates, 2-thiophene, phosphonic acid, squalene, aldehyde and vitamin E (Yan, 2016). The metabolomics finding describes more on the effect of the deep peat soil toward the growth of sago palm rather than the cause of the non-trunking in sago palm. A comprehensive study is needed to link all the omics research in trunking and non-trunking sago palm and one of the ways is by differential expressed genes (DEGs) analysis utilising next generation sequencing (NGS) technology. It is arguable to that proteomics study are better in explaining the non-trunking phenomenon in sago palm but due to: the limitation of data validation; limited supporting bioinformatics tools; and masking effect of the high abundance proteins; making it less popular choice for whole proteins profiling comparative studies (Zhang *et al.*, 2014). RNA sequencing advancement on the other hand: enables the discovery of novel transcripts; identification of alternatively spliced genes; detection of allele-specific expression; and elucidate the functional complexity of the transcription; making it a popular choice for comparative studies (Kukurba & Montgomery, 2015).

2.5 Nucleotide Sequencing Technology

Sequencing technology can be divided to three generations. First-generation sequencing consist of the two methods; Sanger's chain termination method, which sequence the DNA strain using the A, C, G and T random ddNTPs chain termination polymerase chain reaction (PCR); and the less popular Maxam and Gilbert's chemical degradation method which sequence DNA strain using the phosphate-radioisotopes-labelled A, G+A, C+T and C chemical fragmented (Masoudi-Nejad *et al.*, 2013; Maxam & Gilbert, 1977; Sanger *et al.*, 1977) (**Figure 2.1 & Figure 2.2**).

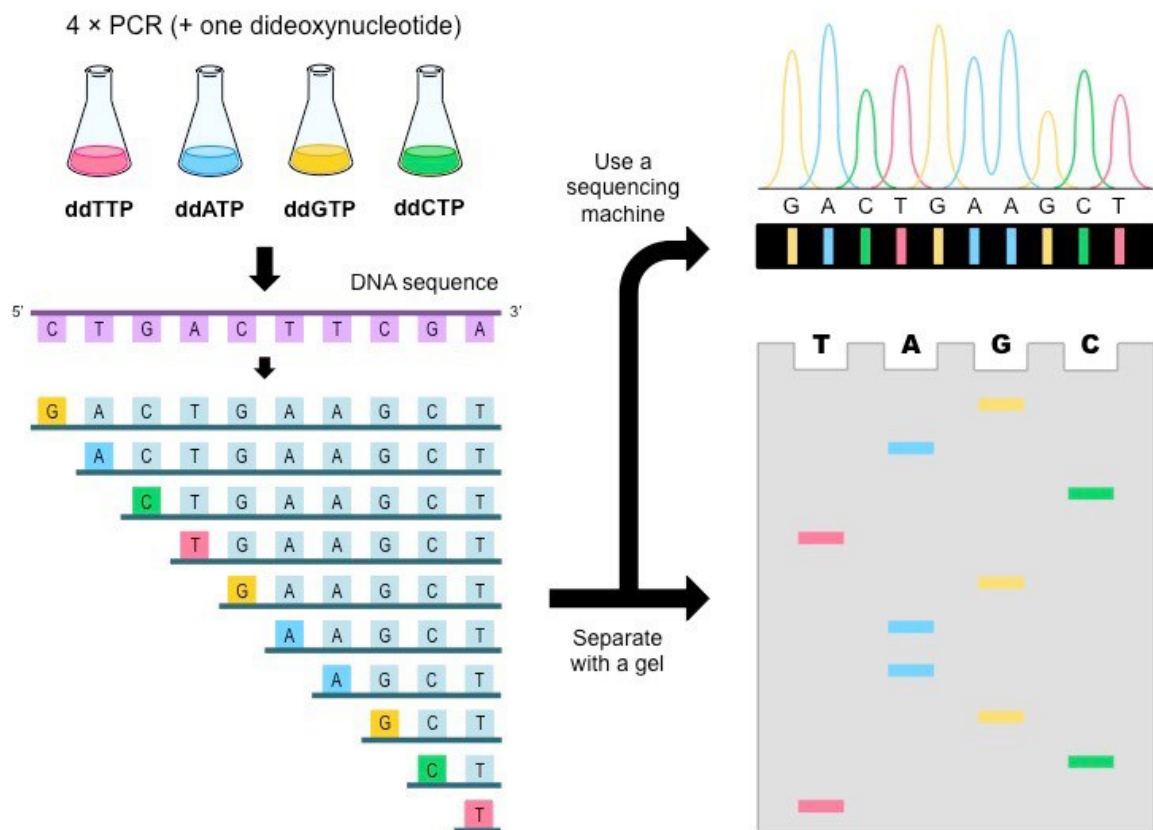


Figure 2.1: Chain Termination Method

Description: Chain termination method sequences the target sequence by comparing the migration of the amplified known 5'-end DNA fragments in gel electrophoresis.

Source: (BioNinja, n.d.)

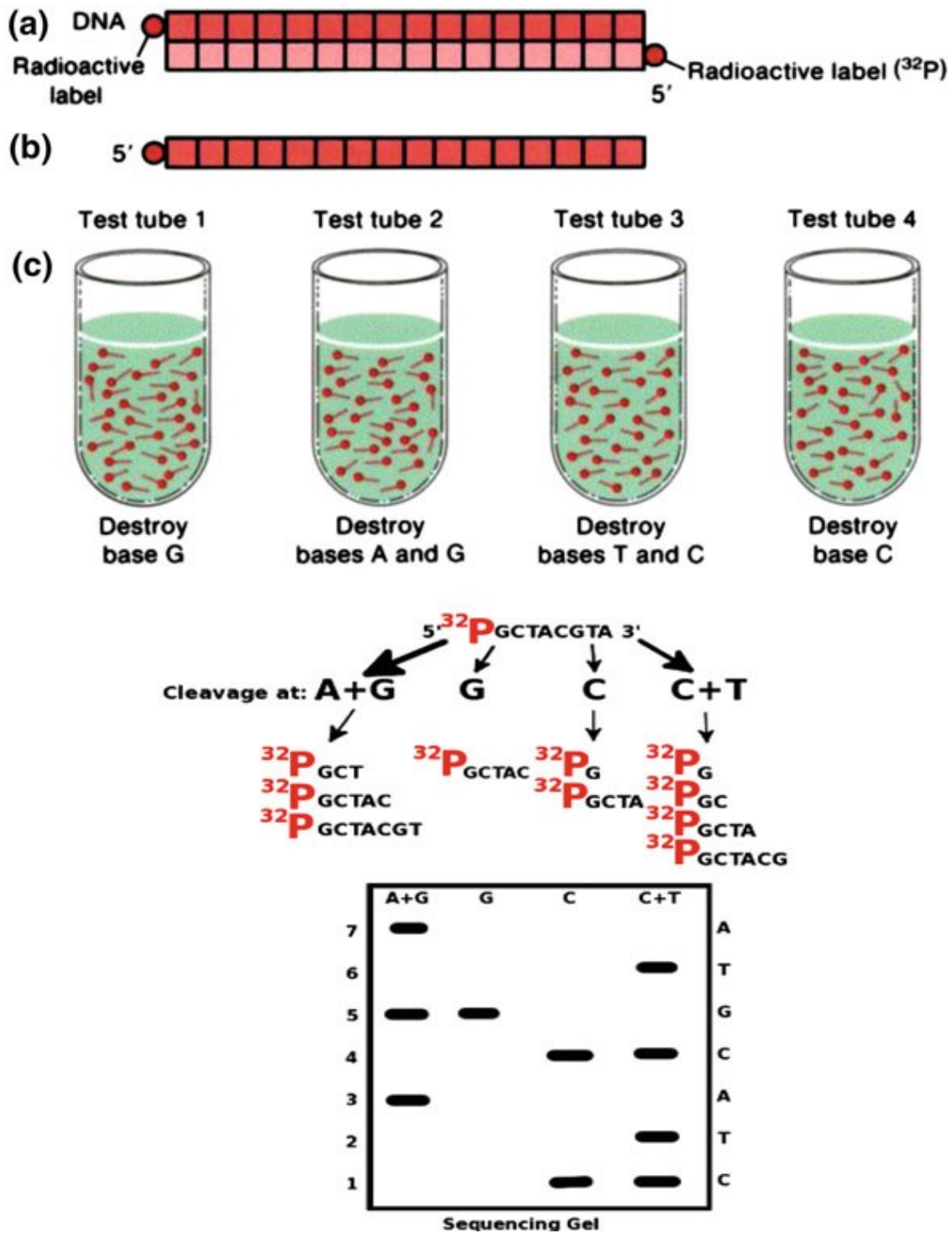


Figure 2.2: Chemical Degradation Method

Description: Chemical degradation method sequences the target sequence by comparing the migration of the radioisotope labelled targeted nucleotide degraded DNA fragments in gel electrophoresis.

Source: Reprint with permission (Masoudi-Nejad *et al.*, 2013) (License Number: 5484901135893)

Second generation sequencing or next generation sequencing (NGS) denoted as high-throughput sequencing was first developed in 1996 at the Stockholm Royal Institute of Technology and later commercially introduced in the form of Genome Sequencer GS20 in the year 2005 due to the high demand for an automated, faster and cheaper sequencer after the experience of the first human genome project that took thirteen years which was completed in 2003. The general work flow of NGS are: sample collection; library preparation, where sample sequence is digested to a shorter fragment and inserted with adapters at both ends; fixing the fragmented sample sequences on a flow cell plate and colony amplified; parallel sequencing by colony PCR on the flow cell plate; and lastly data analysis and assembly (**Figure 2.3**). Compared to first generation Sanger sequencing (700 - 1000 bps), the read length of NGS is much shorter (100 - 700 bps) but it can parallel sequence billions of colonies at a time giving a massive sequence data (Bansal *et al.*, 2018; Masoudi-Nejad *et al.*, 2013). In second generation sequencing, three different techniques are commonly used to amplify DNA fragments before sequencing which can be categorised as: emulsion PCR (Thermo Fisher, Qiagen, Roche), bridge amplification (Illumina), template walking (Thermo Fisher) and DNA nanoball generation (Beijing Genome Institute) (Goodwin *et al.*, 2016).

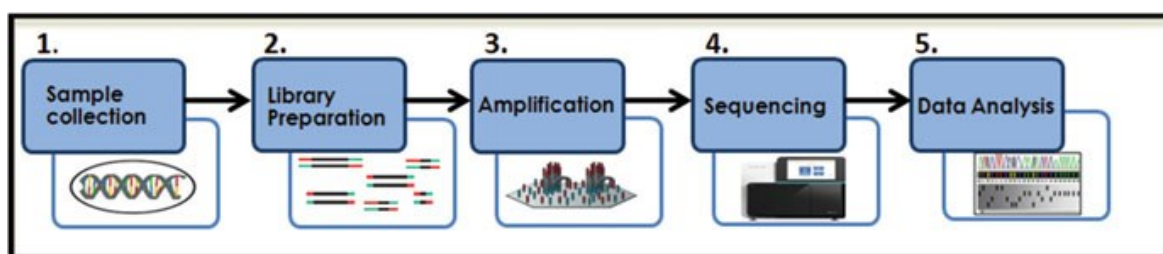


Figure 2.3: Next Generation Sequencing Workflow

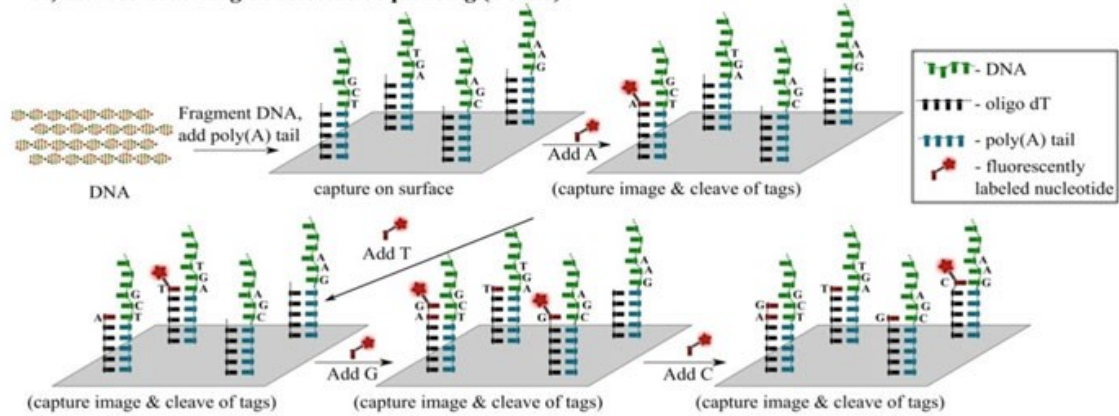
Description: General workflow of NGS in all platforms.

Source: Reprint with permission (Bansal *et al.*, 2018) (License Number: 5484910045046)

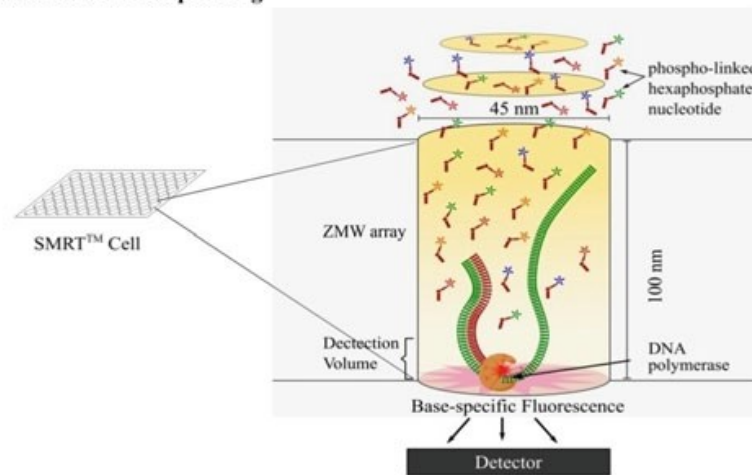
Third generation sequencing or single molecule sequencing denoted as long-read sequencing are represented by three technologies namely the Helicos BioSciences' True Single-Molecule Sequencing (tSMS), Pacific Biosciences' Single-Molecule Real-Time Technology (SMRT) and Oxford Nanopore Technologies' Nanopore Sequencing. Third generation sequencing are similar to second generation sequencing but no fragments amplification is required. General workflow of third generation sequencing consist of: a simple library preparation sequence digestion and adapter ligation; fixation on reaction cell; sequencing capturing; and finally data analysis (Bansal *et al.*, 2018; Masoudi-Nejad *et al.*, 2013) (**Figure 2.4**). Selection of which generation of sequencing technique is based on the need of the studies as each generations of sequencing technique have its own pros and cons (**Table 2.7**). For examples: If the study focus on few targeted genes, first generation sequencing will be the method of choice; second generating sequencing are commonly used for gene discovery studies because of its high throughput; and long sequence read is the highlight of the third generation sequencing and are the preferential method in eukaryotic genome construction studies (Smith, 2022).

Massively Parallel Sequencing: Third Generation

A) Helicos True Single Molecule Sequencing (TSMS)



B) Pacific Biosciences SMRT sequencing



C) Nanopore Sequencing

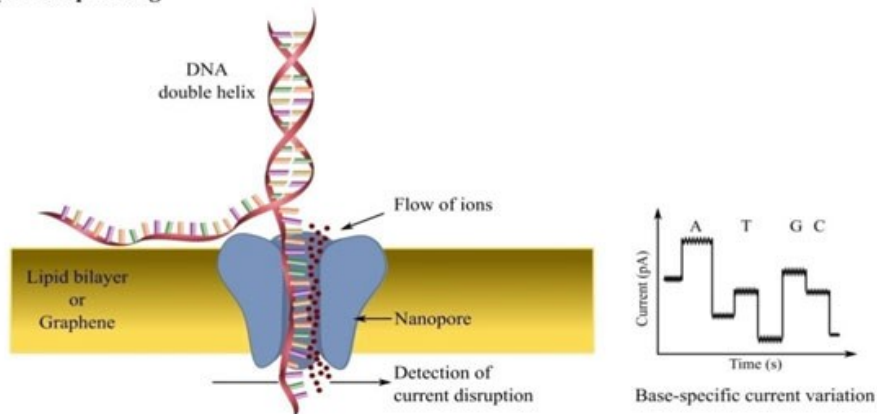


Figure 2.4: Third Generation Sequencing Signal Detection

Description: The three wellknown third generation sequencing detection technology.

Source: Reprint with permission (Bansal *et al.*, 2018) (License Number: 5484910045046)

Table 2.7: Comparison Between Sequencing Technology by Generation

Description: Second generation sequencing has the highest throughput, raw read accuracy but short sequence read length which is most suitable for total transcriptome analysis

Sequencing	First Generation	Second Generation	Third Generation
Fundamental technology	Size-separation of specifically end labelled DNA fragments, produced by sequencing by synthesis or degradation	Wash-and-scan sequencing by synthesis	sequencing by synthesis, by degradation, or direct physical inspection of the DNA molecule
Resolution	Averaged across many copies of the DNA molecule being sequenced	Averaged across many copies of the DNA molecule being sequenced	Single-molecule resolution
Current raw read accuracy	High	High	Moderate
Current read length	Moderate (800 – 1000 bps)	Short, generally much shorter than Sanger sequencing (100 – 700 bps)	Short to Long (100 - >1000 bps)
Current throughput	Low	High	Moderate
Current cost	High cost per base Low cost per run	Low cost per base High cost per run	Low-to-moderate cost per base Low cost per run

Table 2.7 continued

RNA-sequencing method	cDNA sequencing	cDNA sequencing	Direct RNA sequencing and cDNA sequencing
Time start to result	Hours	Days	Hours
Sample preparation	Moderately complex, PCR amplification not required	Complex, PCR amplification required	Ranges from complex to very simple depending on technology
Data analysis	Routine	Complex because of large data volumes and because short reads complicate assembly and alignment algorithms	Complex because of large data volumes and because technologies yield new types of information and new signal processing challenges
Primary results	Base calls with quality values	Base calls with quality values	Base calls with quality values, potentially other base information such as kinetics

Source: (Schadt *et al.*, 2010)

2.5.1 RNA Sequencing (RNA-seq)

RNA-seq is one of the applications of sequencing technology where RNAs were converted to cDNA prior to sequencing in second generation sequencing while direct RNAs sequencing in third generation sequencing to obtain the genetic information and functions of the RNAs. RNAs can be divided into coding (messenger RNA (mRNA)) and non-coding RNA. Non-coding RNA can be further divided based on its function such as ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNAs), small nucleolar RNAs (snoRNAs), microRNA (miRNA), piwi-interacting RNA (piRNA), and long noncoding RNAs (lncRNAs) (Kukurba & Montgomery, 2015). Transcriptomics are studies of the RNA expression and mRNA is one of the fields of study which relate genetic and protein expression. RNA-seq are commonly practiced using NGS platform for the purpose of transcriptomics studies as it generates a huge amount of reads to cover the wide variety of gene expressions in the samples (Schadt *et al.*, 2010). Information obtained by RNA-seq are raw truncated short sequences which need to be further processed to produce the full-length sequences and its information by bioinformatics tools.

2.6 RNA Extraction and Quality Assessment

Selection of the right RNA extraction protocol is crucial as it will affect the quality of the RNA sample use for sequencing later (Yockteng *et al.*, 2013). The importance parameters to indicate a good quality RNA extractions are purity and integrity (Heptinstall & Rapley, 2000; Imbeaud *et al.*, 2005). RNA purity is measured by the RNA/Protein contaminant (OD260/280) ratio and RNA/(polysaccharides and other organic contaminants) (OD260/230) ratio (Heptinstall & Rapley, 2000; Imbeaud *et al.*, 2005). The gold standard for a pure RNA sample indicated by the OD260/280 and OD260/230 ratio more than score 1.8 (Heptinstall & Rapley, 2000). Conventional RNA integrity test use the 28S/18S rRNA

ratio of 2.0 as the gold standard for high-quality RNA integrity indicator. The conventional test is lack of reliability because of the inconsistency of the peak areas measurement and a poor indicator of the other RNA integrity (Imbeaud *et al.*, 2005). A more recent approach of RNA integrity utilising RNA integrity number (RIN) measure the RNA integrity by multidimensional scores which include; Total RNA ratio, 28S peak height and area ratio, 18S and 28S area to the area of the fast region, linear regression at the end point of the fast region, amount of detected fragments in the fast region, presence or absence of the 18S peak, and relation of the overall mean value to the median value (Schroeder *et al.*, 2006). The measurement of RNA integrity using RIN more accurately representing the RNA integrity of the sample compared to the conventional test (Imbeaud *et al.*, 2005).

Cells in plant contain the same genetic information but the gene expression and regulation are different in respond to particular cell type, developmental stage and environmental stimuli or stresses (Suza *et al.*, 2021).

Transcriptomics studies revealed that gene expressions are dynamic but interconnected between organs where the gene expression of one organ influences the others (Guo *et al.*, 2022; Liao *et al.*, 2020). Plant leaf tissue's gene expression are widely used and able to show the cause and effect of plant stresses as it is the centre of metabolisms where in involved in; all sort of metabolites, hormones, signaling molecules, and etc. synthesis; photosynthesis; nutrient assimilation; waste deposition; gases exchange; transpiration; and etc. (Digel *et al.*, 2015; Guo *et al.*, 2022; Liao *et al.*, 2020).

2.7 Bioinformatics

Bioinformatics is an interdisciplinary field which encompass molecular biology, computer science, and statistics. It is use to: process intensive data; large-scale biological processes and interactions; and produce inferences of collected data (Can, 2014).

In RNA-seq data analysis, internal BGISEQ-500 bioinformatics tools were used to removes low quality raw short sequence reads and use sequence assemble tool such as ABySS, SOAPdenovo, Oases or Trinity, which the latter proven superior, to assemble the clean raw short sequence reads to full sequence reads for further analysis (Grabherr *et al.*, 2011). Full sequence reads can be functionally annotated by performing sequence alignment with bioinformatics databases such as National Center for Biotechnology Information (NCBI), Kyoto Encyclopedia of Genes and Genomes (KEGG), UniProt database and etc. (Ejigu & Jung, 2020). RNA expression can also be calculated using the data from RNA-seq by using bioinformatics tool such as RSEM or transcriptome comparison between samples for differentially expressed gene (DEGs) identification by using bioinformatics tool such as Bioconductor DESeq2 (Li & Dewey, 2011; Love *et al.*, 2014) (**Figure 2.5**).

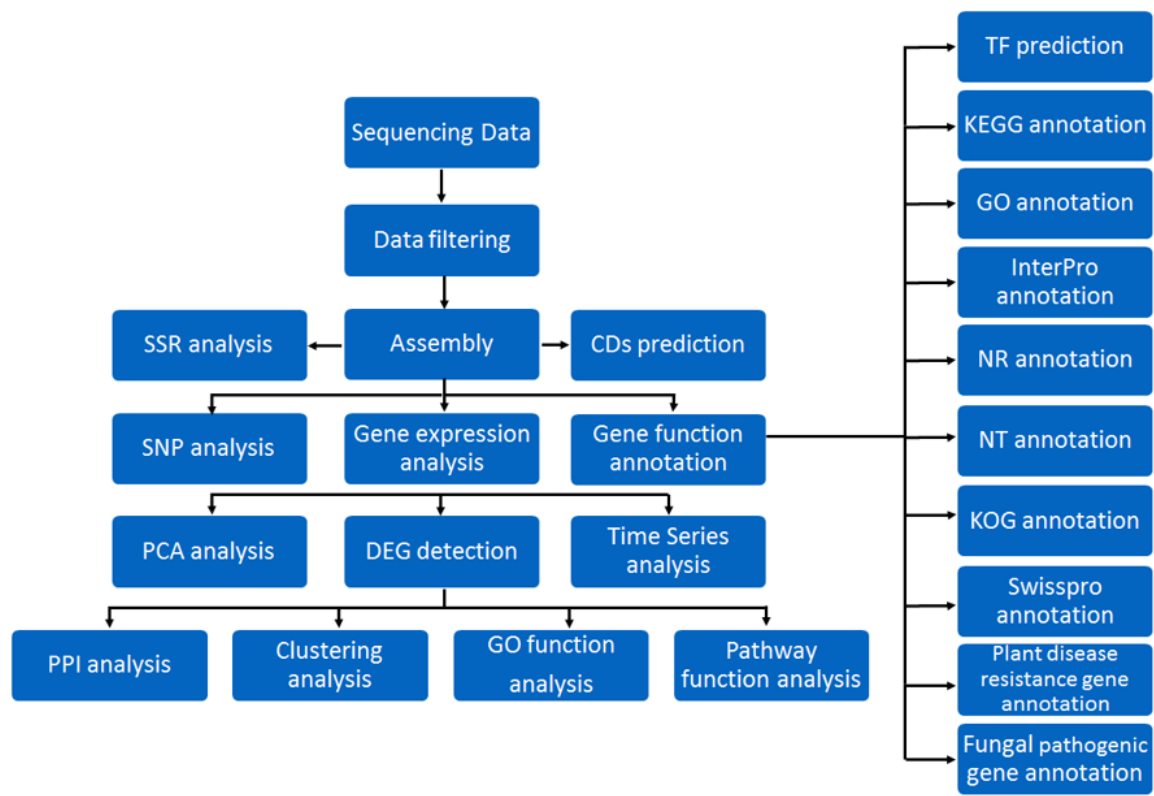


Figure 2.5: Bioinformatics Workflow of RNA Sequencing Service provided by Beijing Genome Institute

Description: Gene expression between trunking and non-trunking sago palm is compare by DEG detection followed by the functional analysis, pathway analysis, and protein-protein interaction analysis to explain the non-trunking phenomenon in sago palm.

Keys: SSR = simple sequence repeat; CDs = Coding sequences; SNP = single nucleotide polymorphism; PCA = Principal component analysis, DEG = Differentially expressed gene; PPI = Protein-protein interaction; GO = Gene ontology; TF = Transcript factor; KEGG = Kyoto Encyclopedia of Genes and Genomes, NR = NCBI protein database, NT = NCBI necleotide database; KOG = Eukaryotic orthologous groups

CHAPTER 3

MATERIALS AND METHODS

3.1 Sampling

The sampling design was based on the objective of the project which is to compare normal (trunking) and abnormal (non-trunking) *M. sago*. Five biological samples of *M. sago* leaves from trunking and non-trunking were collected. Sampling site were at Dalat Sago Plantation, Mukah, Sarawak. The sampling was conducted on October 23, 2018 (**Table 3.1**).

Table 3.1: Sampling Time and Global Positioning System (GPS)

Description: Sampling were performed in pentaplicates for both trunking and non-trunking sago palm on the same day at Dalat Sago Plantation.

Morphology	Sample ID	Collection Time (UTC +08:00)	GPS (WGS84 datum)
Non-Trunking (> 17 years old)	NT01	10am	+2° 49' 40.3'', +111° 50' 26.0''
	NT02		+2° 49' 40.5'', +111° 50' 25.8''
	NT03		+2° 49' 40.7'', +111° 50' 25.6''
	NT04		+2° 49' 40.4'', +111° 50' 26.3''
	NT05		+2° 49' 41.2'', +111° 50' 25.7''
Trunking (~ 5 years old)	T01	11am	+2° 51' 07.7'', +111° 49' 35.9''
	T02		+2° 51' 07.8'', +111° 49' 36.0''
	T03		+2° 51' 08.1'', +111° 49' 35.8''
	T04		+2° 51' 08.0'', +111° 49' 35.5''
	T05		+2° 51' 07.8'', +111° 49' 36.4''

The second or third frond from the centre of the sago palm crown was selected as the target sample. The leaflets from the selected frond were wiped with a kitchen towel applied with 70% ethanol as a disinfectant and to remove debris. The midrib of the leaflets was removed and the leaflets were rolled then inserted into the 50 mL falcon tube, which was labelled accordingly. The 50 mL falcon tube containing the samples was immediately snap-freeze in liquid nitrogen. The sample tubes were kept in Dewar tank containing liquid nitrogen until delivered to the university laboratory and stored in a -80°C freezer.

3.2 RNA Extraction Protocol Trial

RNA extraction was performed using three protocols to determine the best RNA extraction protocol of sago palm leaf tissue for the purpose of RNA sequencing. The tested protocols were CTAB Miniprep (Untergasser, 2008), PCI (Ong, 2019), and MRIP (Xiao *et al.*, 2012). 7 g sago palm leaflet was powdered with a pre-cooled mortar and pestle with liquid nitrogen and were used in the subsequent extraction trials.

3.2.1 CTAB Miniprep RNA Extraction

Chemical was prepared according to the protocol (**Table 3.2**). Extraction were performed with some adaptation to standardise initial extraction sample weight.

Table 3.2: CTAB Miniprep Chemicals List

Description: Materials required for CTAB Miniprep protocol's buffer preparation.

Chemicals	Recipe
CTAB Buffer (pH 8.0)	2% Cetyl Trimethyl Ammonium Bromide (CTAB)
	1.4 M Sodium Chloride (NaCl)
	20 mM Ethylenediaminetetraacetic acid (EDTA)

Table 3.2 continued

	100 mM tris(hydroxymethyl)aminomethane (TRIS)
	2% Polyvinylpyrrolidone (PVP40)
Tris-HCl Buffer	40 mM Tris-HCl
	10 mM MgSO ₄
	1 mM CaCl ₂
RQ1 RNase-Free DNase	Promega
β-Mercaptoethanol	Molecular Grade
Chloroform	ACS grade
Isopropanol	ACS grade
70% Ethanol	70% ACS grade absolute ethanol
	30% dH ₂ O

One-gram of powdered sago leaf was transferred into two labelled tubes each. Fifteen ml of ice-cooled CTAB with 1% β-mercaptoethanol was transferred into each tube. An equal volume of ice-cooled chloroform was added to each tube and mixed. The mixtures were partitioned by centrifuging at 13,200 rpm, 4°C for 5 minutes. The top aqueous layer from tube were transferred into new labelled centrifuge tubes. Another 15 ml ice-cooled chloroform was added into each tube and mixed. The mixtures were repartitioned by centrifuging at 13,200 rpm in 4°C for 5 minutes. The top aqueous layer from the tubes were transferred into new labelled centrifuge tubes. An equal volume of pre-cooled 2-isopropanol were added into each tubes and incubated in room temperature for 10 minutes. The total nucleotides were precipitated by centrifuging at 13,200 rpm in 4°C for 15 minutes. Supernatant from the tubes were discarded and the pellets were transferred into a new 1.5

ml microcentrifuge tubes with 1 ml of ice-cooled 70% EtOH. The total nucleotide pellets were washed twice with 70% EtOH and dried. The total nucleotides were then treated with RQ1 RNase-Free DNase according to its recommended protocol. After that, the tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes to precipitate the contaminants. Finally, the supernatant containing the total RNA were then transferred into a new tube before further analysis. The step-by-step CTAB RNA extraction protocol is available in **Appendix A**.

3.2.2 PCI RNA Extraction

Chemical was prepared according to the protocol (**Table 3.3**). Extraction were performed with some adaptation to standardise initial extraction sample weight.

Table 3.3: PCI Chemicals List

Description: Materials required for PCI protocol's buffer preparation.

Chemicals	Recipe
Extraction Buffer	0.15 M LiCl
	0.05 M Tris (pH 9)
	0.005 M EDTA
	5% SDS
PCI Buffer	50% Liquid Phenol (pH 4 - 4.5)
	48% ACS grade Chloroform
	2% ACS grade Isoamyl Alcohol
Tris-HCl Buffer	40 mM Tris-HCl
	10 mM MgSO ₄
	1 mM CaCl ₂
LiCl	8 M LiCl

Table 3.3 continued

EtOH	ACS grade absolute ethanol
NaOAc	3 M NaOAc
70% Ethanol	70% ACS grade absolute ethanol
	30% dH ₂ O
dH ₂ O	Distilled water

One-gram of powdered sago leaf was transferred into two labelled tubes each. Fifteen ml of ice-cooled extraction buffer was transferred into each tube and mixed. The mixtures were partitioned by centrifuging at 13,200 rpm, 4°C for 5 minutes and the top aqueous layer from tubes were transferred into new labelled tubes. The aqueous layer in the new tubes were then washed with 15 ml of ice-cooled PCI buffer trice. After that, the aqueous layer in the new tubes was washed with 15 ml chloroform twice. 8 M LiCl were added to the aqueous layer in the new tubes with the final concentration of 2 M LiCl and incubated at 4°C overnight (~ 24 hours). The next day, the total nucleotides were precipitated by centrifuging at 13,200 rpm, 4°C for 30 minutes. The supernatant was discarded and the pellets were transferred into a new 1.5 ml microcentrifuge tubes with 1 ml of ice-cooled 70% ethanol. The total nucleotides were reprecipitated by centrifuging at top speed at 4°C for 5 minutes. Supernatant from tubes were discarded and the pellets were dissolved in 400 µl dH₂O followed by addition of 1 ml 100% EtOH and 40 µl 3 M NaOAc. Tubes were incubated at -80°C for 20 minutes and then reprecipitated the total nucleotide by centrifuging at top speed at 4°C for 30 minutes. Supernatant were discarded and the pellets were washed with 1 ml of 70% EtOH twice and dried. The total nucleotides were then treated with RQ1 RNase-Free DNase according to its recommended protocol. After that, the tubes were centrifuged at

13,200 rpm, 4°C for 5 minutes to precipitate the contaminants. Finally, the supernatant containing the total RNA were then transferred into a new tube before further analysis. The step-by-step PCI RNA extraction protocol is available in **Appendix B**.

3.2.3 MRIP RNA Extraction

Chemical was prepared according to the protocol (**Table 3.4**). Extraction were performed with some adaptation for more initial extraction sample weight.

Table 3.4: MRIP Chemicals List

Description: Materials required for MRIP protocol's buffer preparation.

Chemicals	Recipe
MRIP buffer (pH 5.0)	0.4 M Ammonium Thiocyanate
	0.8 M Guanidine Thiocyanate
	0.1 M Sodium Acetate
	4 M Phenol
Tris-HCl buffer	40 mM Tris-HCl
	10 mM MgSO ₄
	1 mM CaCl ₂
Chloroform	ACS grade
Isopropanol	ACS grade
70% ethanol	70% ACS grade absolute ethanol
	30% dH ₂ O

One-gram of powdered sago leaf was transferred into two labelled tubes each. Fifteen ml of MRIP buffer was transferred into the tubes, mixed and incubated at room temperature

for 5 minutes. 5 ml ice-cooled chloroform were then added into tubes, mixed and incubated for another 5 minutes at room temperature. The mixtures were partitioned by centrifuged at 13,200 rpm, 4°C for 10 minutes and the top aqueous layer from tubes were transferred into new labelled tubes. Equal volume of pre-cooled isopropanol was then added into tubes, mixed and incubated on ice for 10 minutes. The total nucleotides were then precipitated by centrifuging at 13,200 rpm in 4°C for 15 minutes. Supernatant from tubes were discarded and the pellets were transferred into a new 1.5 ml microcentrifuge tubes with 1 ml of ice-cooled 70% EtOH. The nucleotides were repelleted by centrifuging at 13,200 rpm in 4°C for 5 minutes. The pellets were washed again with 1 ml of ice-cooled 70% EtOH and dried. The total nucleotides were then treated with RQ1 RNase-Free DNase according to its recommended protocol. After that, the tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes to precipitate the contaminants. Finally, the supernatant containing the total RNA were then transferred into a new tube before further analysis. The step-by-step MRIP RNA extraction protocol is available in **Appendix C**.

3.2.4 Gel Electrophoresis

One percent TBE-AGE gel was prepared for gel electrophoresis (1% Agarose in TBE (0.09 M Tris base, 0.09 M Boric acid, 0.002 M EDTA, pH 8.0)). Five microliters Thermo Scientific™ GeneRuler 1kb DNA ladder and 5 µl from all the RNA extracts which were mixed with 1.5 µl of loading dye were loaded into the gel wells. Gel electrophoresis was conducted for 35 min at 100V, 2A and viewed on an UV transilluminator.

3.2.5 Spectrophotometry

Ten times and 100x dilution were performed on each extracts from CTAB, PCI and MRIP protocols to measure the total RNA concentration and extraction purity using 260/280

and 260/230 ratios. The diluted extracts were analysed using Eppendorf Biophotometer plus. Extracts were transferred into Eppendorf UVette® for measurement and Tris-HCl buffer was used as the blank solution.

3.2.6 Final RNA Extraction Protocol

CTAB miniprep protocol was optimised to obtain the desired RNA quality and quantity for RNA sequencing. Chemical was prepared according to the optimised protocol (**Table 3.5**). The general difference between the modified CTAB method with the Untergasser (2008) CTAB miniprep are mentioned in **Table 3.6**. The optimised RNA extraction protocol was performed on four batches of all the 10 sago palm leaf tissue samples (5 trunking and 5 non-trunking) for pre-sequencing assessment.

Table 3.5: Improved CTAB Extraction Protocol Chemicals List

Description: Materials required for the final modified CTAB protocol's preparation.

Chemicals	Recipe
CTAB Buffer (pH 8.0)	2% Cetyl Trimethyl Ammonium Bromide (CTAB)
	1.4 M Sodium Chloride (NaCl)
	20 mM Ethylenediaminetetraacetic acid (EDTA)
	100 mM tris(hydroxymethyl)aminomethane (TRIS)
	2% Polyvinylpyrrolidone (PVP40)
Tris-HCl Buffer	40 mM Tris-HCl
	10 mM MgSO ₄
	1 mM CaCl ₂
20mM EDTA solution	20 mM EDTA
RQ1 RNase-Free DNase	Promega

Table 3.5 continued

β-Mercaptoethanol	Molecular Grade
Chloroform	ACS grade
Isopropanol	ACS grade
70% Ethanol	70% ACS grade absolute ethanol
	30% dH ₂ O

Table 3.6: General Differences of Modified CTAB Method with Untergasser (2008)

Description: List of modifications of the improved CTAB method compared to reference.

Features	Improved CTAB Method	Untergasser (2008) Method
RNA pellet re-solubilisation	Gentle mixing	65 °C in water for 15 minutes
DNA removal step	DNase Treatment	Did not mention which specifically
RNA precipitation	ice-cooled isopropanol	Lithium Chloride
Estimated time (time mention from the protocol only)	1 hour 53 minutes	1 hour 56 minutes

One-gram sago palm leaf tissue was powdered with a pre-cooled mortar and pestle with liquid nitrogen and transferred into a labelled tube. Fifteen ml of CTAB with 1% beta-mercaptoethanol was transferred into the tubes, mixed and incubated 5 minutes. Fifteen ml of chloroform was added into the tubes, mixed and incubated 5 minutes. The mixtures were partitioned by centrifuging at 13,200 rpm in 4°C for 5 minutes. The top aqueous layer from tubes were transferred into new labelled centrifuge tube. Fifteen ml of chloroform was added

into tubes, mixed and incubated 5 minutes. The mixtures were partitioned by centrifuging at 13,200 rpm in 4°C for 5 minutes. The top aqueous layer from tube was transferred into new labelled centrifuge tubes. Equal volume of ice-cooled isopropanol was added into the pooled tube, mixed and incubated in room temperature for 10 minutes. The total nucleotides were then precipitated by centrifuging at 13,200 rpm in 4°C for 15 minutes. Supernatant was discarded and 1 ml of 70% EtOH was added and the pellets were transferred into labelled new 1.5 ml microcentrifuge tubes. The total nucleotides were then reprecipitated by centrifuging at 13,200 rpm in 4°C for 2 minutes. Pellet was washed again with 70% EtOH and dried. The total nucleotides were then treated with RQ1 RNase-Free DNase according to its recommended protocol. After that, the tubes were centrifuged at 13,200 rpm, 4°C for 15 minutes to precipitate the contaminants. The top aqueous layer from the tubes were transferred into four new labelled tube each. Equal volume of ice-cooled isopropanol was added into the tubes, mixed and incubated in room temperature for 10 minutes. The total RNA was then reprecipitated by centrifuging at 13,200 rpm in 4°C for 15 minute. Supernatant from tube was discarded and washed with 1 ml of 70% ethanol twice. Supernatant from tube was then discarded and the total RNA pellet was dried. Finally, the total RNA pellet was suspended in 1 ml of ice-cooled absolute ethanol and stored at -80°C before use. The step-by-step CTAB RNA extraction protocol is available in **Appendix D**.

3.3 RNA Attribute Measurement and Selection

All the extracted RNA samples quality was initially verified using TBE-AGE analysis and spectroscopy analysis before shipment to Beijing Genome Institute (BGI). In BGI the samples were remeasured using Nanodrop and Agilent 2100 Bioanalyzer to identify the effect of the sample over shipment time. After that, samples selection to proceed with

RNA-seq are based on RIN, 25S/18S ratio, purity, and total mass of the total RNA (Weighted in descending order).

3.4 RNA Sequencing

For sequencing, six samples were selected with the highest scores, three biological replicates each from trunking and non-trunking sago palm leaf tissue were proceeded for RNA-seq analysis by BGI.

Library preparation was conducted on all the six selected RNA samples by BGI (**Figure 3.1**). First, the total RNA were mRNA enriched by Oligo(dT) selection to remove all the other RNAs except mRNAs. The mRNAs were then fragmentized by enzyme digestion. After that, double-strand cDNA (dscDNA) synthesis by reverse transcription using N6 random primer was performed. The cDNAs were then subjected to end-repair and 3' adenylation before adding A tailing and adaptor ligation. The quantity of the dscDNA embedded with adapter and poly-A tail was increased using PCR amplification. The PCR product were denatured by heat and the single-strand DNA is cyclised by splint oligo and DNA ligase. Finally, combinatorial probe-anchor synthesis (cPAS), utilising Phi 29 DNA polymerase for rolling circle replication, was performed to generate the DNA nanoball, which later was fixed onto the patterned array flow-cell for sequencing.

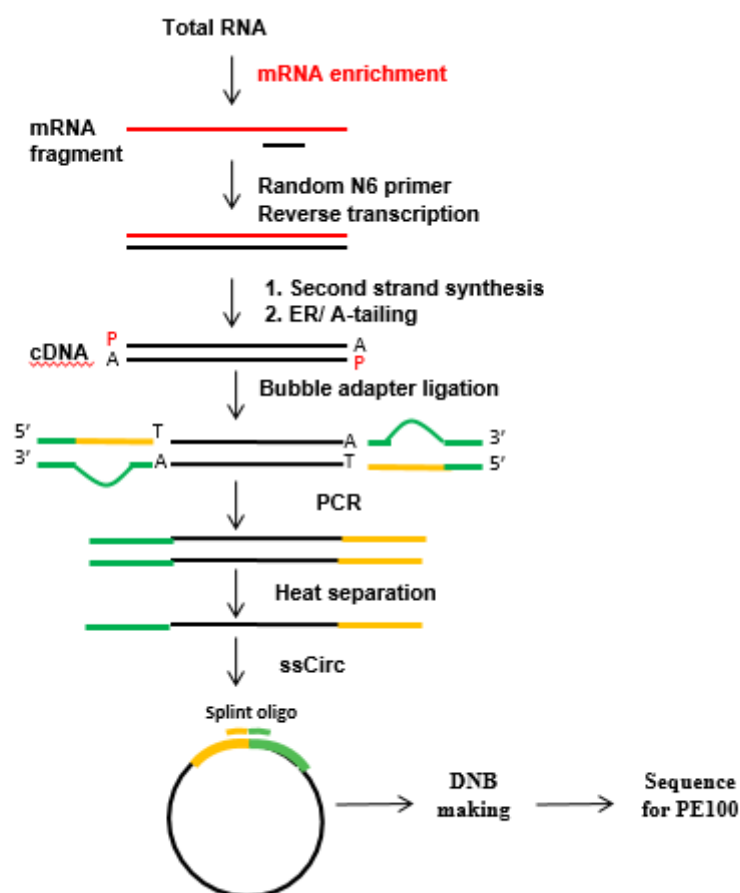


Figure 3.1: BGISEQ-500 Transcriptome Library Workflow

Description: Diagram on the sample preparation before sequencing

Source: (BGI, 2019b)

Paired-end RNA sequencing was conducted on all the prepared samples using BGISEQ-500 platform generating 6 GBs of raw data each. Bioinformatics analysis was performed on the obtained data (**Figure 3.2**). The initial 6 GBs raw sequence were filtered by an in-house developed quality control (QC) pipeline to remove low quality reads; sequences with more than 20% of the bases having phred quality score lower than 15, reads with adaptors sequence, and reads with unknown bases (N) which consist of more than 5%

of the total bases; to produce the clean reads (Fang *et al.*, 2017). *De novo* assembly were conducted using the clean reads with Trinity (Version: v2.0.6; Parameters: --min_contig_length 150 --CPU 8 --min_kmer_cov 3 --min_glue 3 --bfly_opts '-V 5 --edge-thr=0.1 --stderr) subsequently clustered the transcript using TGICL (Version: v2.0.6; Parameters: -l 40 -c 10 -v 25 -O '-repeat_stringency 0.95 -minmatch 35 -minscore 35') to determine the Unigenes.

The list of Unigenes obtained were further undergo functional annotation with seven functional databases namely; Nucleic acid database of NCBI (NR), protein database of NCBI (NT), Gene Ontology database (GO), Eukaryotic Orthologous Groups database (KOG), Kyoto Encyclopedia of Genes and Genomes database (KEGG), SwissProt database and InterPro database. The candidate Unigenes Coding DNA Sequence (CDS) were selected using TransDecoder software (Version: v3.0.1; Parameters: default) and predicted using Pfam protein homology sequence on the longest open reading frame (ORF). Unigenes which encode transcription factor (TF) and short sequence repeat (SSR) were also identified. Gene expression analysis were conducted using clean reads and mapped with Unigenes using Bowtie2 (Version: v2.2.5; Parameters: -q --phred64 --sensitive --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 -I 1 -X 1000 --no-mixed --no-discordant -p 1 -k 200) and the gene expression level were calculated using RSEM (Version: v1.2.12; Parameters: default). Differentially expressed genes (DEGs) were identified using DEseq2 (Parameters: Fold Change ≥ 2.00 and Adjusted Pvalue ≤ 0.05) and Poisson Distribution Algorithms (Parameters: Fold Change ≥ 2.00 and FDR ≤ 0.001).

The identified DEGs were further analysed using R software for hierarchical clustering using pheatmap function to identify the gene expression correlation between

trunking and non-trunking sago palm. Gene ontology (GO) classification and functional enrichment using phyper function, KEGG pathway classification and functional enrichment by phyper function, and protein-protein interaction (PPI) networks with STRING software were performed to identify the significant gene functions and pathways which related to the non-trunking phenomenon in sago palm.

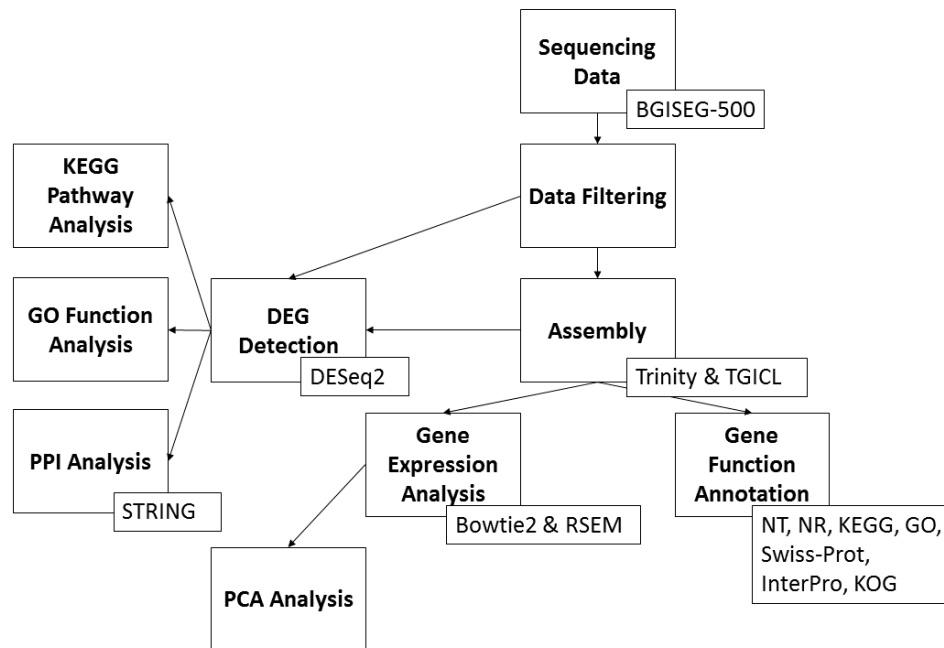


Figure 3.2: Bioinformatics Workflow of the Sequenced Data

Description: Diagram on the transcriptome analysis service provided by BGI focusing on comparative analysis

Multiple comparison tests were performed on the GO-Terms and KEGG pathways of the non-trunking sago palm DEGs. The P-Values of the GO-Terms were calculated based on hypergeometric distribution and corrected by bonferroni correction to obtained the adjusted p-value (**Equation 3.1** & **Equation 3.2**).

$$P - Value = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{N-i}}{\binom{N}{n}} \quad \text{Equation 3.1}$$

Keys; m = Number of DEGs in the GO-Term
M = Number of DEGs in the population
i = Number of DEGs in the GO-Term which is selected
N = Total Genes in the population
n = Total Genes in the GO-Term

$$adjusted\ P - Value = (P - value) \times (Total\ number\ of\ GO\ term) \quad \text{Equation 3.2}$$

The P-value of the KEGG pathways were calculated based on hypergeometric distribution and corrected by Holm–Bonferroni method to obtained the Q-Value (**Equation 3.3 & Equation 3.4**).

$$P - Value = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{N-i}}{\binom{N}{n}} \quad \text{Equation 3.3}$$

Keys; m = Number of DEGs in the pathway
M = Number of DEGs in the population
i = Number of DEGs in the pathway which is selected
N = Total Genes in the population
n = Total Genes in the pathway

$$Q - Value = \frac{(P - value) \times (Total\ number\ of\ pathways)}{Rank - 1} \quad \text{Equation 3.4}$$

Keys; Rank = Place of the pathway in P-value ascending order

Selection of the significant GO-Terms and KEGG Pathways were based on the adjusted P-Value ≤ 0.01 and Q-Value ≤ 0.01 respectively as recommended in the BGI report to keep the false discovery rate at 1% or lower. PPI analysis was performed to complement/validate the GO functional and KEGG pathway analysis.

3.5 Quantitative Real-Time PCR (qPCR) Analysis

To support the finding of RNA-seq analysis, cross-referencing was conducted using qPCR analysis of selected DEGs based on its significance, identity and expression level. The test was performed with all the ten sago palm samples (5 trunking and 5 non-trunking).

3.5.1 Gene Selections

Gene selection was based on the following criteria: 1) Only unique gene without isoforms; 2) Gene which was identified in KEGG/GO; 3) Gene originated from plant source; 4) DEGs log₂foldChange between trunking and non-trunking sago palm from 1 – 5.

Only one gene was selected as a reference gene for expression normalisation as this is only used to validate the samples gene expression between trunking and non-trunking sago palm from the RNA-seq data. Proteasome subunit beta type-4 (Unigene4400_ALL) was selected as reference gene because it's expressions between all the samples were stable (Adjusted P-value: 0.99995), high expression (DESeq2 expression > 900), high sequence length (1194 bp) which reduce the chances of overlapping with other sequence, and do not have other isoform.

3.5.2 Primer Design

Primer design were performed using Primer-Blast by National Center for Biotechnology Information (NCBI). Subsequently the designed primers were ordered from Integrated DNA Technologies, Inc. (IDT). NCBI Primer-Blast were performed in default settings except: 1) “Primers for target on one template” tab was selected; 2) “PCR template” window where selected gene sequence in FASTA format was inputted; 3) “Primer Pair Specificity Checking Parameters” Database tab were set to Custom and RNA-seq gene

library in FASTA format was uploaded as template; 4) “Primer Pair Specificity Checking Parameters: Organisms” window where “*Metroxylon sagu* (taxid:93297)” was inputted

Factor in consideration for design primer selection are: 1) Low self-complementary and self-3'-complementary value < 7.00; 2) Difference of T_m values between forward and reversed primer < 3°C; 3) Product length preferred < 200 bps

3.5.3 Two-Step RT-qPCR

Stocked RNA extracts were used in the cDNA synthesis using Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit (**Table 3.7**) and qPCR analysis conducted using Applied Biosystems™ PowerUp™ SYBR™ Green Master Mix with recommended protocols (**Table 3.9 & Table 3.10**). cDNA synthesis and qPCR condition were established and performed using Bio-Rad T100 thermal cycler (**Table 3.8**) and Applied Biosystems QuantStudio 5 Real-Time PCR System respectively (**Table 3.11**). High-Resolution Melting (HRM) analysis was performed to validate the amplification in the qPCR only produced targeted product (**Table 3.12**). qPCR was performed for all ten samples (five trunking and five non-trunking) with three technical replicates and one no-template control (NTC) for five selected genes and one reference gene (proteasome subunit beta type-4).

Table 3.7: cDNA Synthesis Chemical List

Description: Materials required for the cDNA synthesis sample preparation.

Component	Volume
Total RNA extract	20.0 μ l
10x RT Buffer	4.0 μ l
25x dNTP Mix (100 mM)	1.6 μ l
10x Oligo(dT)15-VN Primers	4.0 μ l
MultiScribe™ Reverse Transcriptase	2.0 μ l
Nuclease-free water	8.4 μ l
Final Volume	40.0 μ l

Table 3.8: cDNA Synthesis PCR Setting

Description: Parameter settings for cDNA synthesis.

Step	Temperature (°C)	Time (Minute)	Cycles
Anneal primer	25	5	1
Extension	42	60	1
Inactivation	70	15	1
Hold	4	∞	1

Table 3.9: qPCR NTC Chemical List

Description: Materials required for the qPCR NTC preparation.

Component	Volume	Final Concentration
PowerUp™ SYBR™ Green Master Mix (2X)	10.0 µl	1X
Forward Primer (10 µM)	1.00 µl	0.5 µM
Reverse Primer (10 µM)	1.00 µl	0.5 µM
Nuclease-Free Water	8.00 µl	
Total	20.00 µl	

Table 3.10: qPCR Targeted Gene Chemical List

Description: Materials required for the qPCR targeted genes preparation.

Component	Volume	Final Concentration
PowerUp™ SYBR™ Green Master Mix (2X)	10.00 µl	1X
Forward Primer (10 µM)	1.00 µl	0.5 µM
Reverse Primer (10 µM)	1.00 µl	0.5 µM
DNA Template	2.00 µl	≤ 250 ng
Nuclease-Free Water	6.00 µl	
Total	20.00 µl	

Table 3.11: qPCR Settings

Description: Parameter settings for qPCR amplification.

Step	Temperature (°C)	Time (Second)	Cycles
Dual-Lock™DNA polymerase activation	95	300	1
Denaturation	95	15	40
Annealing	55	30	
Elongation	72	60	

Table 3.12: High Resolution Melt Setting

Description: Parameter settings for HRM analysis.

Function	Temperature (°C)	Time (Second)
Pre-Hold	72	15
Initial temperature	72	
Gradient temperature increases	+0.15	
Hold		1
Final Temperature	95	

3.5.4 Relative Gene Expression

Relative gene expressions between trunking and non-trunking sago palm ($\Delta\Delta C_T$) were calculated using the cycle threshold value (C_T -value) of selected genes and compared with the reference gene. The C_T -values were obtained from the qPCR result. The mean C_T of the technical replicates were calculated for all the samples. Relative gene expression within the sample (ΔC_T) was calculated for all the samples. $\Delta\Delta C_T$ was calculated for the gene

of interest and $\log_2$ (Fold-Change) of each gene of interest between trunking and non-trunking were determined (**Equation 3.5 – 3.7**) (Livak & Schmittgen, 2001). Statistical Analysis

3.6 Statistical Analysis

3.6.1 RNA Extraction Protocol Trial

IBM® SPSS® Statistics 28.0.1.0 (142) was used to analysis the 260/280 ratio, 260/230 ratio and RNA weight data obtained for RNA extraction method comparison. All the data were assessed using normality test to determine its' parametricity. Parametric data was further subjected to Multivariate analysis of variance (MANOVA) coupled with Tukey's test while not parametric data was subjected to the Kruskal-Wallis Test.

3.6.2 RNA Attribute Measurement and Selection

Comparison of the before and after shipment RNA attributes test were performed using IBM® SPSS® Statistics version 28.0.1.0 (142). All the data were assessed using normality test to determine its' parametricity. Outlier analysis were performed using boxplot. Pair t-test were performed on the RNA mass, 260/230 and 260/280 ratio to assess the changes of RNA quality and purity after shipment.

3.6.3 Quantitative Real-Time PCR (qPCR) Analysis

In order to determine the accuracy of the RNA-seq gene expression data representing the population samples, comparison of the RNA-seq gene expression with a higher sample number of qPCR gene expression data were performed. The qPCR gene expression data were grouped into two groups; group 1-10: Representing data calculated from all ten sago palm used for population estimation; and group 1,4,5,7,8,9: Representing data calculated

from the same six sago palm samples used in RNA-seq which used for sample estimation. Pearson's correlation study was performed using IBM® SPSS® Statistics 28.0.1.0 (142) to determine the relationship of the RNA-seq gene expression data with the qPCR gene expression data.

$$\Delta C_T = (\text{Mean } C_T \text{ of gene of interest}) \\ - (\text{Mean } C_T \text{ of reference gene}) \quad \textbf{Equation 3.5}$$

$$\Delta\Delta C_T = \Delta C_T (\text{Non} - \text{Trunking}) - \Delta C_T (\text{Trunking sago palm}) \quad \textbf{Equation 3.6}$$

$$\text{Log}_2 (\text{FoldChange}) = -\Delta\Delta C_T \quad \textbf{Equation 3.7}$$

CHAPTER 4

RESULTS AND DISCUSSION

4.1 RNA Extraction Protocol Comparison

Three RNA extraction protocols were tested on sago palm leaf tissues: CTAB miniprep, PCI and MRIP. One gram of powdered sago palm leaf tissue each was extracted using the three protocols, quantified and their purity were compared using Eppendorf BioPhotometer Plus for the selection of the best RNA extraction protocol of sago palm leaf tissue (**Table 4.1**).

Table 4.1: Spectrophotometry Reading of the Extracted RNA Samples

Description: CTAB protocol shows the highest RNA purity (260/230 ratio) and (260/280 ratio); and quantity (average concentration)

Protocol - Replicate	A230	A260	A280	A340	260/280	260/230	Concentration ($\mu\text{g/mL}$)	
							Total	Average
CTAB 1	0.316	0.637	0.330	0.001	1.93	2.01	255.0	177.5
CTAB 2	0.131	0.250	0.130	0.000	1.93	1.90	100.0	
PCI 1	0.299	0.270	0.161	0.008	1.67	0.90	107.8	111.7
PCI 2	0.354	0.289	0.181	0.017	1.60	0.82	115.6	
MRIP 1	0.941	0.189	0.160	0.044	1.18	0.20	75.6	79.7
MRIP 2	1.073	0.209	0.180	0.054	1.16	0.20	83.8	

The 260/280 ratio, 260/230 ratio and the RNA weight data were subjected to a normality test using IBM[®] SPSS[®] Statistics 28.0.1.0 (142). The normality test's null

hypothesis (H_0) assumes the population is normally distributed and the standard significance level (α) is 0.05. The P-value of 260/280 ratio and 260/230 ratio data are higher than α so we accept H_0 ; hence they are normally distributed (**Table 4.2**). On the other hand, P-value of RNA weight is lower than α so we reject H_0 , hence the data are not normally distributed. Since both 260/280 ratio and 260/230 ratio data are normally distributed, they were subjected to MANOVA coupled with Tukey's test while because of the RNA weight data is not normally distributed, it was subjected to independent-samples Kruskal-Wallis test using IBM® SPSS® Statistics 28.0.1.0 (142). The H_0 for the Tukey's test assumes all means being compared are from the same population and the standard significance level (α) is 0.05. The P-value between all the RNA extraction methods are lower than the α (0.05) value so we reject H_0 , hence the 260/280 ratio and 260/230 ratio data between all the RNA extraction methods are not the same (**Table 4.3**). The H_0 for the Kruskal-Wallis test assumes the distribution of the weight is the same across the RNA extraction methods and the standard significance level (α) is 0.05. The P-value of all the RNA extraction methods are higher than the α (0.05) value so we accept H_0 , hence distribution of the weight is the same across the RNA extraction methods (**Table 4.4**).

Table 4.2: Normality Test for 260/280 Ratio, 260/230 Ratio and RNA Weight Data

Description: The p-value of 260/280 ratio and 260/230 ratio data are higher than α (0.05) value hence they are normally distributed while RNA weight data p-value are lower than α (0.05) value hence it is not normally distributed with 95% confidence.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
260/280	.210	6	.200*	.859	6	.185
260/230	.219	6	.200*	.862	6	.195
Weight	.373	6	.009	.689	6	.005
*. This is a lower bound of the true significance.						
a. Lilliefors Significance Correction						

Table 4.3: Tukey Multiple Comparison Test of 260/280 Ratio and 260/230 Ratio Data between RNA Extraction Methods

Description: The P-value between all the RNA extraction methods are lower than the α (0.05) value hence the 260/280 ratio and 260/230 ratio data between all the RNA extraction methods are significantly with 95% confidence.

Multiple Comparisons							
Tukey HSD							
Dependent Variable	(I) Method	(J) Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
260/280	CTAB	MRIP	.7600*	.02972	<.001	.6358	.8842
		PCI	.2950*	.02972	.004	.1708	.4192
	MRIP	CTAB	-.7600*	.02972	<.001	-.8842	-.6358
		PCI	-.4650*	.02972	.001	-.5892	-.3408
	PCI	CTAB	-.2950*	.02972	.004	-.4192	-.1708
		MRIP	.4650*	.02972	.001	.3408	.5892
260/230	CTAB	MRIP	1.7550*	.05553	<.001	1.5230	1.9870
		PCI	1.0950*	.05553	<.001	.8630	1.3270
	MRIP	CTAB	-1.7550*	.05553	<.001	-1.9870	-1.5230

Table 4.3 continued

		PCI	-.6600*	.05553	.003	-.8920	-.4280
	PCI	CTAB	-1.0950*	.05553	<.001	-1.3270	-.8630
		MRIP	.6600*	.05553	.003	.4280	.8920
Based on observed means.							
The error term is Mean Square(Error) = .003.							
*. The mean difference is significant at the .05 level.							

Table 4.4: Independent-Samples Kruskal-Wallis Multiple Comparison Test of RNA Weight Data between RNA Extraction Methods

Description: There is no significant difference of the extracted RNA weight between the RNA extraction methods as shown by all the adjusted P-value higher than the α (0.05) value with 95% confidence.

Pairwise Comparisons of Method					
Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig. ^a
MRIP-PCI	-2.000	1.871	-1.069	.285	.855
MRIP-CTAB	4.000	1.871	2.138	.033	.098
PCI-CTAB	2.000	1.871	1.069	.285	.855
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .050.					
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.					

Based on the RNA extraction protocol comparison results, the CTAB protocol gave the highest score of purity and quantity, and most intact visible 25S and 18S rRNA bands compared to the PCI protocol and MRIP protocol (**Figure 4.1 & Table 4.1 – 4.4**). Hence, CTAB protocol was chosen for RNA extract from sago palm leaf tissue and was further optimised to obtain reproducible high-quality RNA extracts.

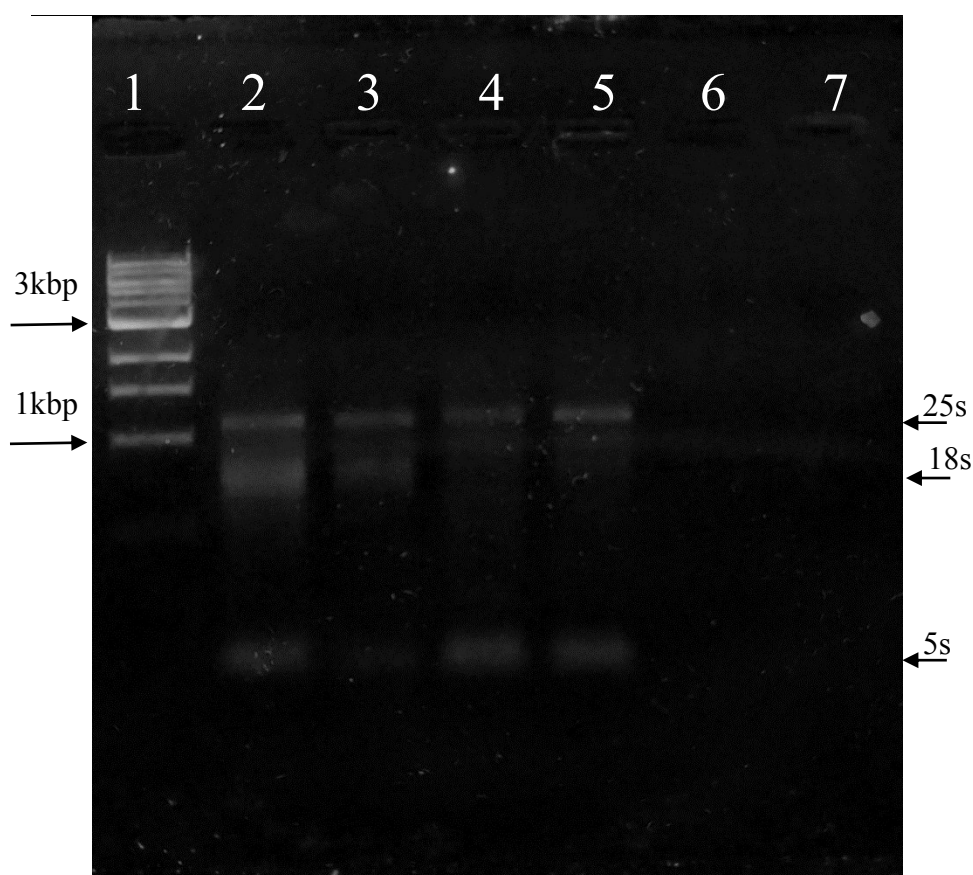


Figure 4.1: Gel Electrophoresis for Total RNA Extract with CTAB, PCI, and MRIP

Description: Lane 1; GeneRuler 1 kb DNA Ladder, lane 2; CTAB replicate 1, lane 3; CTAB replicate 2, lane 4; PCI replicate 1, lane 5; PCI replicate 2, lane 6; MRIP replicate 1, lane 7; and MRIP replicate 2. CTAB protocol gives the best RNA integrity (visible 25S and 18S rRNA bands)

Several adopted methods such as Rochester *et al.* (1986), Cbomczynski & Sacchi (1987), Chang *et al.* (1993), Schultz *et al.* (1994), Salzman *et al.* (1999), Kiefer *et al.* (2000), Hu *et al.* (2002), Hussain (2002), modified Zeng & Yang (2002), Wu *et al.* (2002), modified Gasic *et al.* (2004), Xiao *et al.* (2012) and several kits such as RNAqueous™ Total RNA Isolation Kit, FavorPrep™ Plant Total RNA Mini extraction kit, and Qiagen RNeasy Plant Mini Kit were used in sago palm leaf RNA extraction but the actual quality and quantity were not well defined. Notably, modified Zeng & Yang (2002) and modified Gasic *et al.*

(2004) methods which utilized CTAB as its extraction buffer reported to produce better results compared to others in young sago palm leaf RNA extraction (Yan *et al.*, 2022).

The major differences of the improved CTAB method based on Untergasser (2008) with modified Zeng & Yang (2002) and modified Gasic *et al.* (2004) are the extraction buffer are not pre-heated to 55°C - 65°C and the precipitation are using ice-cooled isopropanol for 10 minutes instead of overnight precipitation with lithium chloride. This change minimized the damage to the RNA due to high temperature and the extraction complete more rapidly (Fabre *et al.*, 2014).

4.2 RNA Attribute Test

Four batches of RNA extraction were performed from the ten sago palm leaf tissue samples (five trunking and five non-trunking) resulting in the total of forty RNA extracts were prepared. Purity, quantity and integrity of the samples were tested before sending for RNA attribute test to ensure the quality and quantity of the samples (**Figure 4.2 & Appendix E**). Samples were first sent to Interscience Sdn. Bhd. by dry ice shipping before being shipped to BGI Hong Kong Company Limited for further testing. RNA attribute test prior to RNA-seq sample selection was performed for all the forty samples by BGI Hong Kong Company Limited. Comparison the RNA attributes before and after shipment were performed using paired t-test. The purity of the RNA extract before and after shipment are identical while the quantity of the RNA samples is significantly lower after shipment (**Appendixes F – J**). This may be due to pipetting errors during sample replication and/or incomplete dissolution of the RNA pellet in the solubilising buffer. This error can be minimised by homogenising the stock extraction before aliquot during replicate preparation and/or sent the total RNA extract in TE buffer instead of pellet in absolute ethanol. The

selection of the samples for RNA sequencing was based on the highest scored RNA samples recommendation by BGI where Total RNA mass $\geq 1 \mu\text{g}$, OD260/280 and OD260/230 ratio > 1.8 , RIN ≥ 6.5 and 25S/18S ≥ 1.0 (BGI, 2019a). Three total RNA extracts from different sago palm were selected for both trunking and non-trunking sago palm. All selected samples exceeded the recommended parameters where the total RNA mass $> 1.0 \mu\text{g}$, OD260/280 and OD260/230 ratio > 1.8 , RIN > 6.5 and 25S/18S > 1.0 (**Table 4.5**). Samples S01T01A, S04T04B and T5 were chosen to represent trunking sago palm cluster, while samples S07NT02A, S08NT03B, and S09NT04B were chosen to represent non-trunking sago palm cluster.

Table 4.5: Selected Total RNA Extracts for RNA Sequencing.

Description: Best three sample replicates were selected from trunking and non-trunking sago palm leaf samples.

Sago Palm	Sample Name	RNA Attribute Test via Agilent 2100 Bioanalyzer & NanoDrop				
		Total Mass (μg)	OD260 /280	OD260 /230	RIN	25S/18S
Trunking	S01T01A	13.50	1.91	2.61	6.6	1.4
	S04T04B	18.7	1.88	2.23	7.4	2
	T5	37.53	1.96	2.22	7.5	1.7
Non-trunking	S07NT02A	38.34	1.99	2.62	7.7	2.0
	S08NT03B	31.45	1.99	2.38	7.4	1.8
	S09NT04B	28.73	1.99	2.41	8.3	1.8

Note: Minimum score for selection; Total Mass $\geq 1 \mu\text{g}$, OD260/280 and OD260/230 ratio > 1.8 , RIN ≥ 6.5 and 25S/18S ≥ 1.0

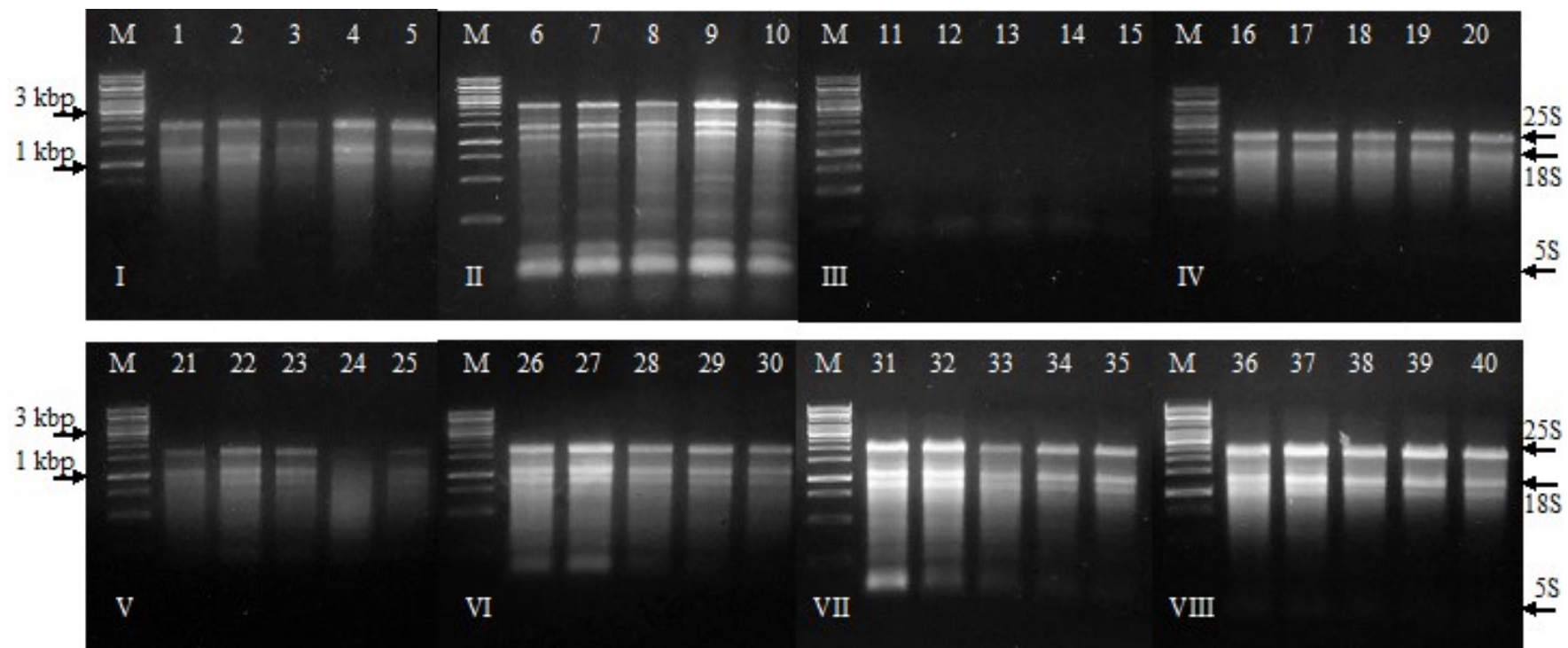


Figure 4.2: TBE-AGE of the 4 Batches of Total RNA Sample from 5 Trunking and 5 Non-Trunking Sago Palm Leaf Tissues

Keys: M: GeneRuler 1 kb DNA Ladder; 1: T1; 2: T2; 3: T3; 4: T4; 5: T5; 6: N1; 7: N2; 8: N3; 9: N4; 10: N5; 11: S01T01; 12: S02T02; 13: S03T03; 14: S04T04; 15: S05T05; 16: S06NT01; 17: S07NT02; 18: S08NT03; 19: S09NT04; 20: S10NT05; 21: S01T01A; 22: S02T02A; 23: S03T03A; 24: S04T04A; 25: S05T05A; 26: S06NT01A; 27: S07NT02A; 28: S08NT03A; 29: S09NT04A; 30: S10NT05A; 31: S01T01B; 32: S02T02B; 33: S03T03B; 34: S04T04B; 35: S05T05B; 36: S06NT01B; 37: S07NT02B; 38: S08NT03B; 39: S09NT04B; 40: S20NT05B. Gel I & II: Batch 1; Gel III & IV: Batch 2; Gel V & VI: Batch 3; Gel VII & VIII: Batch 4.

Description: Samples RNA pellet were dissolved in 100 µl. Then 5 µl of the dissolved RNA pellet were loaded into the gel.

4.3 RNA Sequencing

4.3.1 Sequencing reads filtering

Since the indication of the extraction batch is irrelevant in RNA-seq samples, labelling was simplified to ST1 representing S01T01A, ST4 representing S04T04B, ST5 representing T5, SN7 representing S07NT02A, SN8 representing S08NT03B, and SN9 representing S09NT04B. Raw sequence reads of the six RNA samples obtained from RNA-seq showed good quality for all the samples as the clean read ratio exceeded 95% and after removal of reads with > 5% unknown (N base) ratio, reads which contain adaptors, and low quality reads (> 20% of bases with Q score lower than 15), the clean reads show high accuracy reflected by Q score, Q30 (equivalent to the probability of an incorrect base call 1 in 1000 times) above 90% and Q20 (equivalent to the probability of an incorrect base call 1 in 100 times) above 95% (**Table 4.6**). The sequencing result is consider good as the Q20 and Q30 scores exceeded the good quality sequencing baseline by the BGI standard (Q20 $\geq$ 90%; Q30 $\geq$ 80%) and is comparable with some of the reported RNA sequencing in date palm (Q30 $\sim$ 93% - 95%; clean read ratio $\sim$ 96% - 99%) and oil palm (Q20 > 97%; Q30 > 93%) (Maher *et al.*, 2021; Suraninpong & Nuanlaong, 2022).

Table 4.6: Clean Reads Quality Metrics

Description: All the total RNA samples give higher than 95% clean reads ratio.

Sample	Total Raw Reads(M)	Total Clean Reads(M)	Total Clean Bases(GB)	Clean Reads Q20(%)	Clean Reads Q30(%)	Clean Reads Ratio(%)
ST1	69.96	66.82	6.68	97.88	91.45	95.51
ST4	69.96	66.97	6.7	98.09	92.07	95.72
ST5	69.96	66.68	6.67	97.76	91.14	95.31

Table 4.6 continued

SN7	69.96	66.66	6.67	98	91.78	95.27
SN8	69.41	66.62	6.66	97.94	91.59	95.99
SN9	69.96	67.33	6.73	97.98	91.63	96.24

4.3.2 *De novo* assembly

The clean reads of the fragmented mRNA were then assembled *de novo* using Trinity software and generated a reference sequence (**Table 4.7**). Cluster transcripts from the reference sequence then undergo abundance screening using TGICL software to obtain unique gene (Unigene) sequences (**Table 4.8**). The pooled Unigenes from all the 6 samples (All-Unigene) are comparable to those reported in oil palm comparative transcriptomic study (Suraninpong & Nuanlaong, 2022).

Table 4.7: Quality Metrics of Transcripts

Descriptions: Mean sequence length of all the samples assembly are higher than 700 bp.

Sample	Total Number	Total Length	Mean Length	N50	N70	N90	GC(%)
ST1	75730	56765699	749	1334	740	283	45.54
ST4	94989	73822623	777	1447	778	285	44.11
ST5	84578	65560398	775	1430	776	285	44.67
SN7	68068	55890222	821	1472	860	309	45.44
SN8	91100	65484623	718	1303	684	270	44.75
SN9	97205	78047560	802	1503	821	293	43.65

Keys: N50 is a weighted median statistic that 50% of the total length is contained in Unigenes that are equal to or larger than this value. N70 and N90 represented 70% and 90% of the total length respectively.

Table 4.8: Quality Metrics of Unigenes

Description: There are unique sequences among samples as shows by the lower total number of the sample non-redundant sequences compared to total samples non-redundant sequences.

Sample	Total Number	Total Length	Mean Length	N50	N70	N90	GC(%)
ST1	53270	45270441	849	1409	847	332	45.64
ST4	68390	59838729	874	1522	889	333	44.17
ST5	60537	52810113	872	1506	881	333	44.72
SN7	48542	44955041	926	1542	973	367	45.51
SN8	64137	52366130	816	1390	800	316	44.82
SN9	70377	63341038	900	1570	924	342	43.69
All-Unigene	102447	103410779	1009	1809	1122	382	44.34

Keys: N50 is a weighted median statistic that 50% of the total length is contained in Unigenes that are equal to or larger than this value. N70 and N90 represented 70% and 90% of the total length respectively. “All-Unigene” is the total unique genes from all the samples.

4.3.3 Unigene Functional Annotation

After assembly, the Unigenes were compared with seven functional databases namely Nucleotide Database of National Center for Biotechnology Information (NR), Protein Database of National Center for Biotechnology Information (NT), Gene Ontology (GO), EuKaryotic Orthologous Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), high quality manually annotated and non-redundant protein sequence database of UniProt Knowledgebase (SwissProt) and protein database of European Bioinformatics Institute (InterPro) (**Table 4.9**) for identification. Only 63.96% of the Unigenes were

annotated and the remaining 36.04% are novel gene. Annotated genes show high genes similarity of sago palm based on NR database with oil palm 46% and date palm 34% mainly because they are from the same *Arecaceae* family (**Figure 4.3**). Unique and overlapping Unigene annotations across protein databases were illustrated between NR, KEGG, KOG, SwissProt and InterPro (**Appendix K**). Unigene functional annotation was then performed on KOG, GO, and KEGG databases (**Appendix L - Appendix N**). This work is registered under NCBI BioProject PRJNA781491. The raw RNA sequencing data are available in Sequence Read Archive (SRA) database with accession numbers SRX13165895, SRX13165896, SRX13165897, SRX13165898, SRX13165899, and SRX13165900. Gene expression and annotation information are accessible in public functional genomics data repository Gene Expression Omnibus (GEO) with accession number GSE189085

Table 4.9: Annotation Summary

Description: Only 63.96% of the non-redundant sequences of the samples were annotated.

Values	Number	Percentage
Total	102,447	100%
Nr	56,600	55.25%
Nt	54,530	53.23%
SwissProt	42,327	41.32%
KEGG	44,057	43.00%
KOG	45,181	44.10%
InterPro	47,388	46.26%
GO	15,312	14.95%
Intersection	7,421	7.24%
Overall	65,523	63.96%

Species Distribution

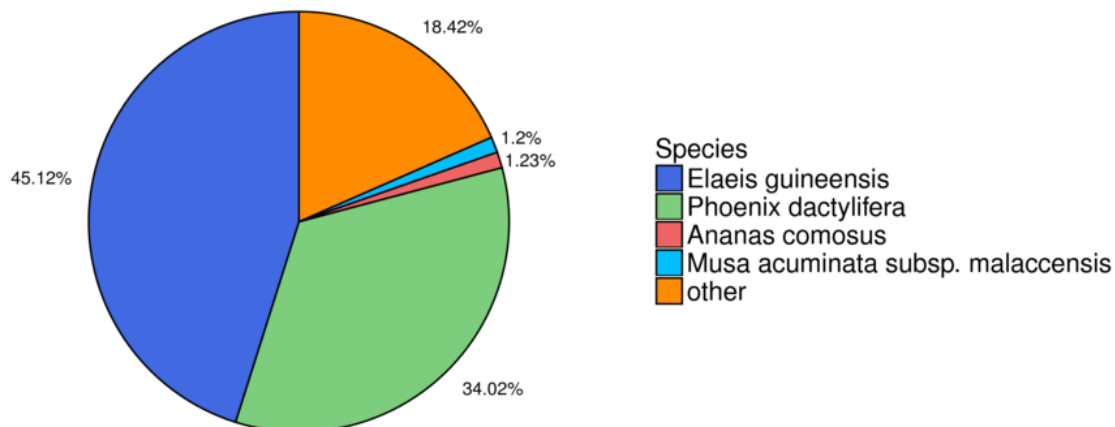


Figure 4.3: Distribution of the NR Database Annotated Unigenes by Species Similarity

Description: Sequences similarity of the six sago palm samples are highest with oil palm sequences in NCBI database.

4.3.4 Unigene Expression

After assembly, all the clean reads of each sample were mapped to the Unigenes with Bowtie2 software and gene expression level were calculated with RSEM. The correlation in gene expression between trunking sago palm (upper-end cluster) and non-trunking sago palm (bottom-end cluster) were illustrated by Principal Component Analysis (PCA) (**Figure 4.4**) which is able to cluster the two phenotypes into trunking and non-trunking groups with 94.29% data coverage even before differentially expressed gene (DEGs) analysis. The PCA analysis shows that the gene expression between trunking and non-trunking sago palm leaf are different which ensure that there are DEGs between these two groups in this dataset. Differentially expressed genes analysis was conducted to identify the genes which are differentially expressed between the two groups.

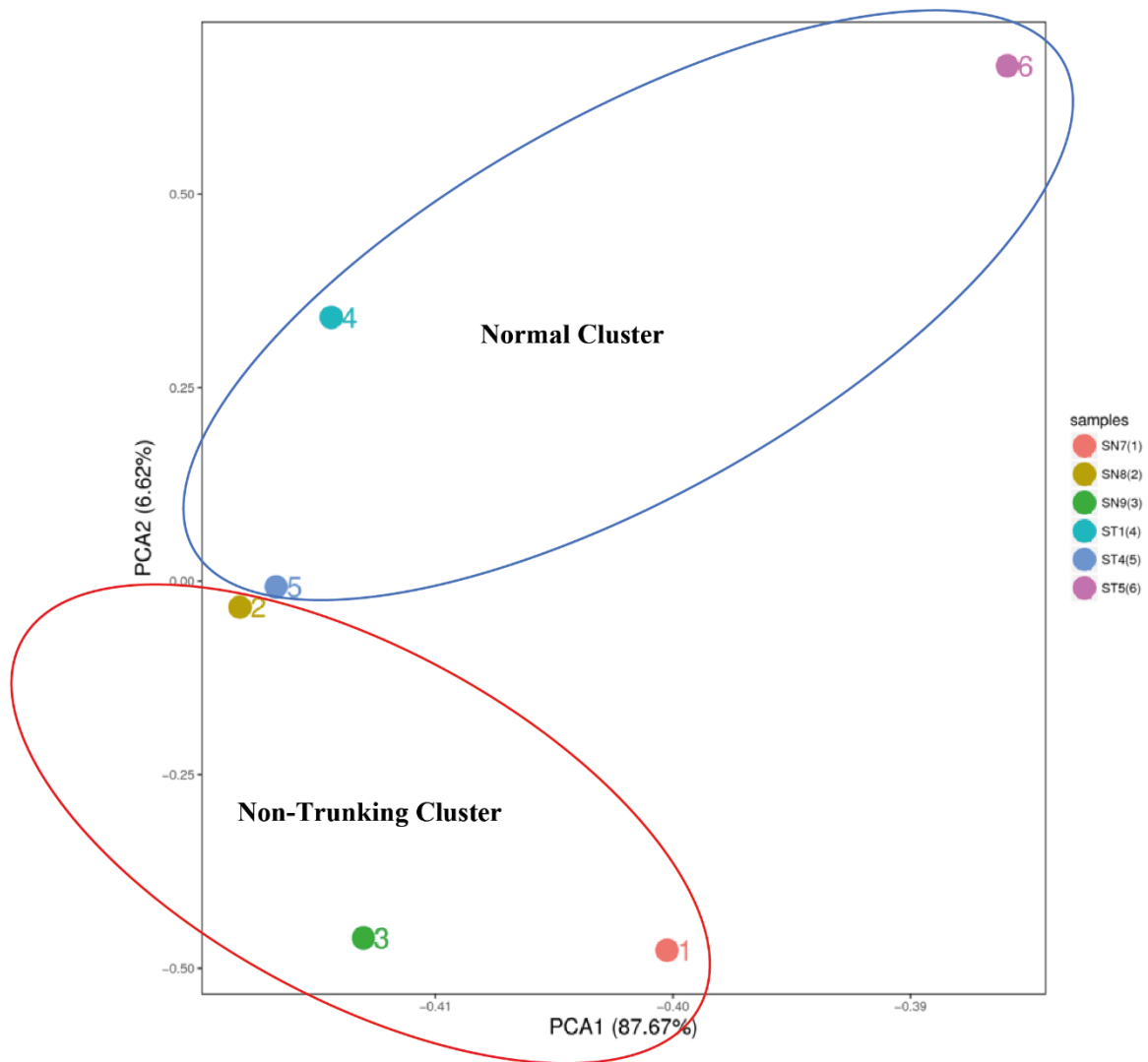


Figure 4.4: Principal Component Analysis of Gene Expression between Trunking and Non-Trunking sago palm leaf

Description: Non-trunking sago palm gene expression was clustered at the bottom cluster while the trunking sago palm gene expression was clustered at the top cluster.

4.3.5 Differentially Expressed Genes (DEGs) Detection

Gene expression between trunking and non-trunking sago palm groups were compared by DESeq2 Bioconductor package with a less than two-fold differences cut-off. A total of 936 DEGs comprising of 492 upregulated and 444 downregulated genes were

identified (**Figure 4.5**). The DEGs between the trunking and non-trunking sago palm group are well distinguished into two groups as shown in the heatmap (**Figure 4.6**). The heatmap shows that the relation in term of gene expression of sample 7 and sample 9 are more closely related than sample 8 in the non-trunking sago palm group. The heatmap also shows that the gene expression of sample 1 and sample 5 are more closely related compared to sample 4 in the trunking sago palm group. In order to identify the roles of the DEGs, functional analysis by Gene Ontology, pathway analysis by KEGG and protein-protein interaction analysis by STRINGdb were performed.

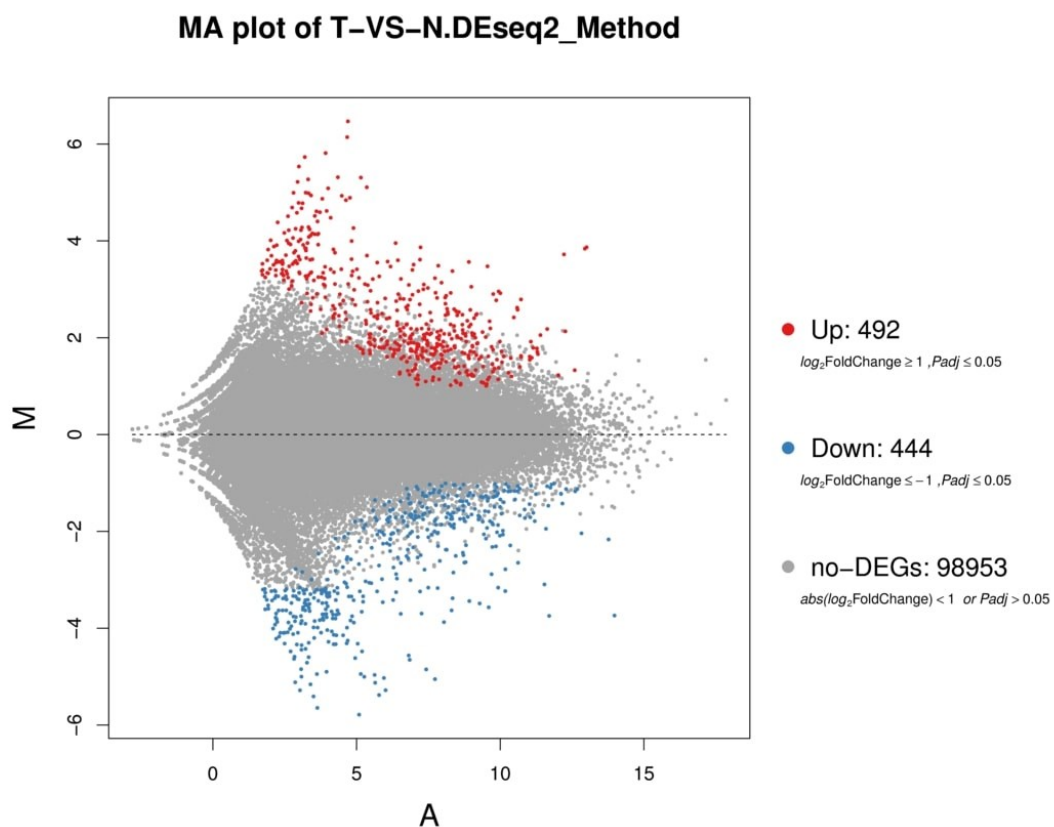


Figure 4.5: MA Plot of DEGs Between Trunking and Non-Trunking Sago Palm

Description: There are a total of 936 DEGs in non-trunking sago palm (492 upregulated + 444 downregulated genes) generated by DESeq2 using six mRNA datasets (3 trunking and 3 non-trunking).

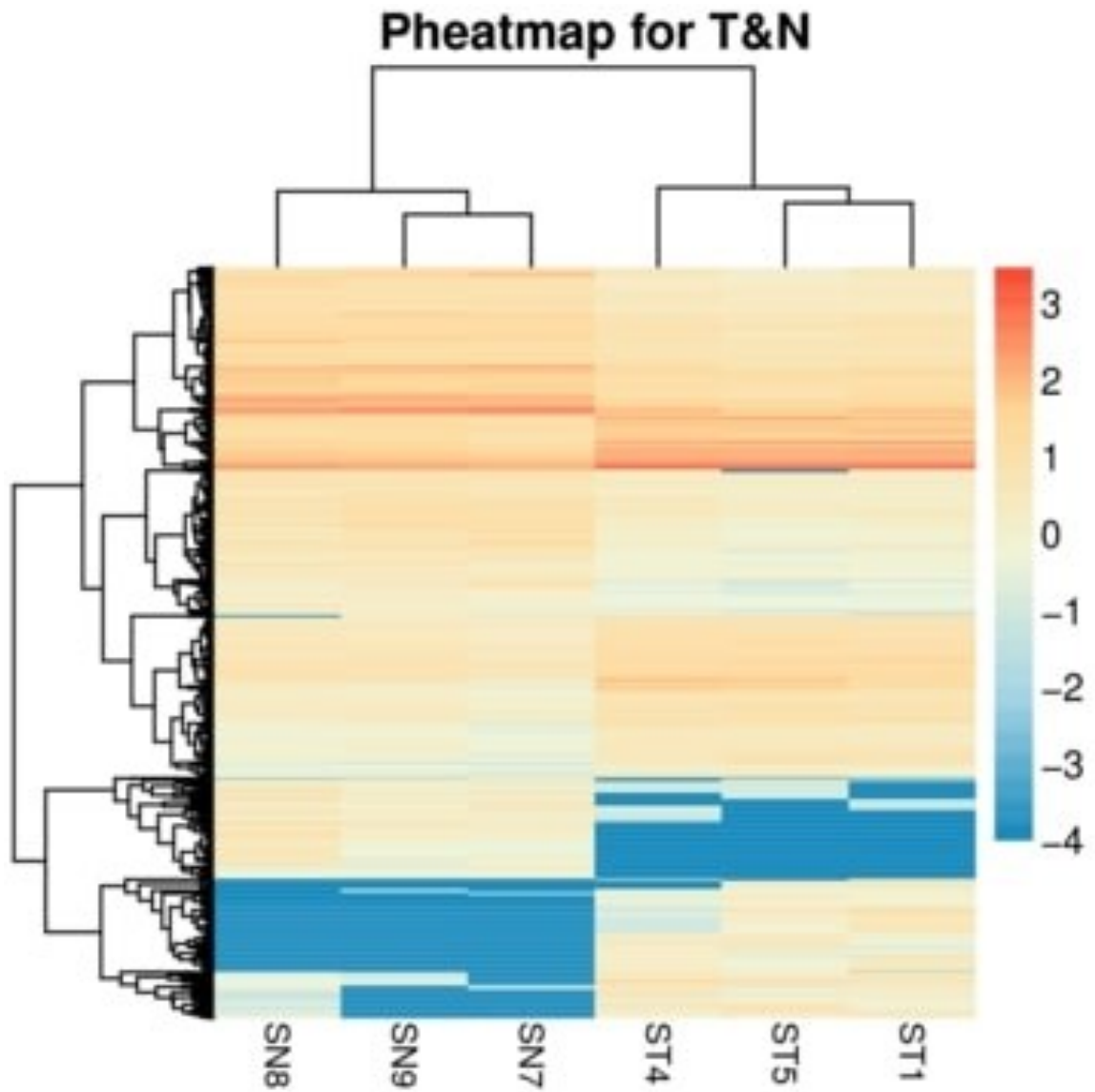


Figure 4.6: Heatmap of DEGs between Samples

Description: Sample 7 with sample 9 and sample 1 with sample 5 are more closely related compared to sample 8 and sample 4 respectively.

4.3.6 DEGs Functional Analysis by Gene Ontology (GO) of Non-Trunking Sago Palm

Functional analysis of the DEGs with GO database annotated 233 genes. The GO annotated DEGs consist of 297 genes involved in biological processes, 355 genes involved in cellular components and 284 genes involved in molecular function (**Appendix O**). The

DEGs regulations based on the GO-Term in non-trunking sago palm leaf tissue were a mixture of upregulated and downregulated genes (**Appendix P**).

Selection of the significant GO-Term are based on the term with less or equal to one percent false discovery rate (adjusted P-value ≤ 0.01) as recommended by BGI guideline. GO enrichment analysis of the DEGs portrayed several biological processes which are significantly different between trunking and non-trunking transcriptomes (**Table 4.10**, **Appendix Q**).

Table 4.10: DEGs involved in Biological Processes and its Definitions

Description: Selection of the significant GO-Terms from the DEGs obtained by DESeq2 were based on hypergeometric distribution which adjusted by the Bonferroni correction (p-adjusted ≤ 0.01)

Biological Processes	Definition	DEGs
positive regulation of photomorphogenesis GO:2000306	Any process that activates or increases the frequency, rate or extent of photomorphogenesis.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All
positive regulation of response to stimulus GO:0048584	Any process that activates, maintains or increases the rate of a response to a stimulus. Response to stimulus is a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus.	CL3840.Contig1_All CL3840.Contig2_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All

Table 4.10 continued

positive regulation of post-embryonic development GO:0048582	Any process that activates or increases the frequency, rate or extent of post-embryonic development. Post-embryonic development is defined as the process whose specific outcome is the progression of the organism over time, from the completion of embryonic development to the mature structure.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All
positive regulation of developmental process GO:0051094	Any process that activates or increases the rate or extent of development, the biological process whose specific outcome is the progression of an organism over time from an initial condition (e.g. a zygote, or a young adult) to a later condition (e.g. a multicellular animal or an aged adult).	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All
positive regulation of multicellular organismal process GO:0051240	Any process that activates or increases the frequency, rate or extent of an organismal process, any of the processes pertinent to the function of an organism above the cellular level; includes the integrated processes of tissues and organs.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All
regulation of photomorphogenesis GO:0010099	Any process that modulates the rate or extent of photomorphogenesis.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All

Table 4.10 continued

regulation of response to red or far red light GO:2000030	Any process that modulates the frequency, rate or extent of response to red or far red light.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All
positive regulation of biosynthetic process GO:0009891	Any process that activates or increases the frequency, rate or extent of the chemical reactions and pathways resulting in the formation of substances.	CL3184.Contig9_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All Unigene10336_All
positive regulation of biological process GO:0048518	Any process that activates or increases the frequency, rate or extent of a biological process. Biological processes are regulated by many means; examples include the control of gene expression, protein modification or interaction with a protein or substrate molecule.	CL3184.Contig9_All CL3840.Contig1_All CL3840.Contig2_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All

Table 4.10 continued

		Unigene10336_All
regulation of response to stimulus GO:0048583	Any process that modulates the frequency, rate or extent of a response to a stimulus. Response to stimulus is a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus.	CL3840.Contig1_All CL3840.Contig2_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All Unigene11064_All
positive regulation of cellular process GO:0048522	Any process that activates or increases the frequency, rate or extent of a cellular process, any of those that are carried out at the cellular level, but are not necessarily restricted to a single cell. For example, cell communication occurs among more than one cell, but occurs at the cellular level.	CL3184.Contig9_All CL3840.Contig1_All CL3840.Contig2_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of transcription DNA-templated GO:0045893	Any process that activates or increases the frequency, rate or extent of cellular DNA-templated transcription.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_Al

Table 4.10 continued

		CL6728.Contig7_All CL7558.Contig4_All
positive regulation of RNA biosynthetic process GO:1902680	Any process that activates or increases the frequency, rate or extent of RNA biosynthetic process.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of nucleic acid-templated transcription GO:1903508	Any process that activates or increases the frequency, rate or extent of nucleic acid-templated transcription.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of RNA metabolic process GO:0051254	Any process that activates or increases the frequency, rate or extent of the chemical reactions and pathways involving RNA.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of metabolic process GO:0009893	Any process that activates or increases the frequency, rate or extent of the chemical reactions and pathways within a cell or an organism.	CL3184.Contig9_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All

Table 4.10 continued

		CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All Unigene10336_All
positive regulation of macromolecule biosynthetic process GO:0010557	Any process that increases the rate, frequency or extent of the chemical reactions and pathways resulting in the formation of a macromolecule, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass.	CL3184.Contig9_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of nucleobase-containing compound metabolic process GO:0045935	Any cellular process that activates or increases the frequency, rate or extent of the chemical reactions and pathways involving nucleobases, nucleosides, nucleotides and nucleic acids.	CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of cellular biosynthetic process GO:0031328	Any process that activates or increases the frequency, rate or extent of the chemical reactions and pathways resulting in the formation of substances, carried out by individual cells	CL3184.Contig9_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All

Table 4.10 continued

		CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All
positive regulation of gene expression GO:0010628	Any process that increases the frequency, rate or extent of gene expression. Gene expression is the process in which a gene's coding sequence is converted into a mature gene product or products (proteins or RNA). This includes the production of	CL3184.Contig9_All CL5846.Contig4_All CL6728.Contig1_All CL6728.Contig2_All CL6728.Contig3_All
	an RNA transcript as well as any processing to produce a mature RNA product or an mRNA or circRNA (for protein-coding genes) and the translation of that mRNA or circRNA into protein. Protein maturation is included when required to form an active form of a product from an inactive precursor form	CL6728.Contig5_All CL6728.Contig7_All CL7558.Contig4_All

Source: (Binns *et al.*, 2009)

The DEGs involved in the twenty significant biological processes between trunking and non-trunking sago palm are govern by pentatricopeptide repeat-containing protein (At3g42630), SPX domain-containing protein 1 (SPX1), translation factor GUF1 homolog chloroplastic (LEPA), nuclear transcription factor Y subunit C-1 (NFYC1), MADS-box transcription factor 50-like (MADS50), cytochrome P450 (CYPs), and LSD ONE LIKE 1 (LOL1) (**Table 4.11**).

Table 4.11: Annotation of the DEGs involved in the Biological Processes

Description: There are several biological processes which are differential expressed in non-trunking sago palm.

DEGs	Unigenes	Log ₂ FoldChange
<i>pentatricopeptide repeat-containing protein (At3g42630)</i>	CL3184.Contig9_All	-1.72
<i>SPX domain-containing protein 1 (SPX1)</i>	CL3840.Contig1_All	1.64
	CL3840.Contig2_All	1.59
<i>translation factor GUF1 homolog, chloroplastic (LEPA)</i>	CL5846.Contig4_All	4.22
<i>nuclear transcription factor Y subunit C-1 (NFYC1)</i>	CL6728.Contig1_All	-3.38
	CL6728.Contig2_All	-3.70
	CL6728.Contig3_All	-3.00
	CL6728.Contig5_All	-3.49
	CL6728.Contig7_All	-3.44
<i>MADS-box transcription factor 50-like (MADS50)</i>	CL7558.Contig4_All	-1.25
<i>cytochrome P450 (CYPs)</i>	Unigene10336_All	4.84
<i>LSD ONE LIKE 1 (LOL1)</i>	Unigene11064_All	3.70

Note: DEGs was calculated by DESeq2 and annotated by GO database based on the unigenes sequences.

One of the DEGs involved in seven GO enrichment biological processes is *pentatricopeptide repeat-containing protein (At3g42630)* which is downregulated in non-trunking sago palms leaf tissue and are involved in seven biological processes in GO enrichment (**Figure 4.7**). The flows of the biological processes show that *At3g42630* involved in the positive regulations of cellular biosynthesis, macromolecule biosynthesis,

and gene expression. At3g42630 are target-specific to DNA or mRNA to modulate organelle (chloroplast and/or mitochondria) gene expression by mRNA processing and stabilising and/or regulating translation by translational protein activation or recruitment of translation related proteins (Saha *et al.*, 2007). At3g42630 involves in plant male fertility restoration, plant embryogenesis, circadian rhythmic expression and plant development processes (Saha *et al.*, 2007). However, specific function of At3g42630 have yet to be reported. The downregulation of *At3g42630* suggested that non-trunking sago palms were under growth suppression.

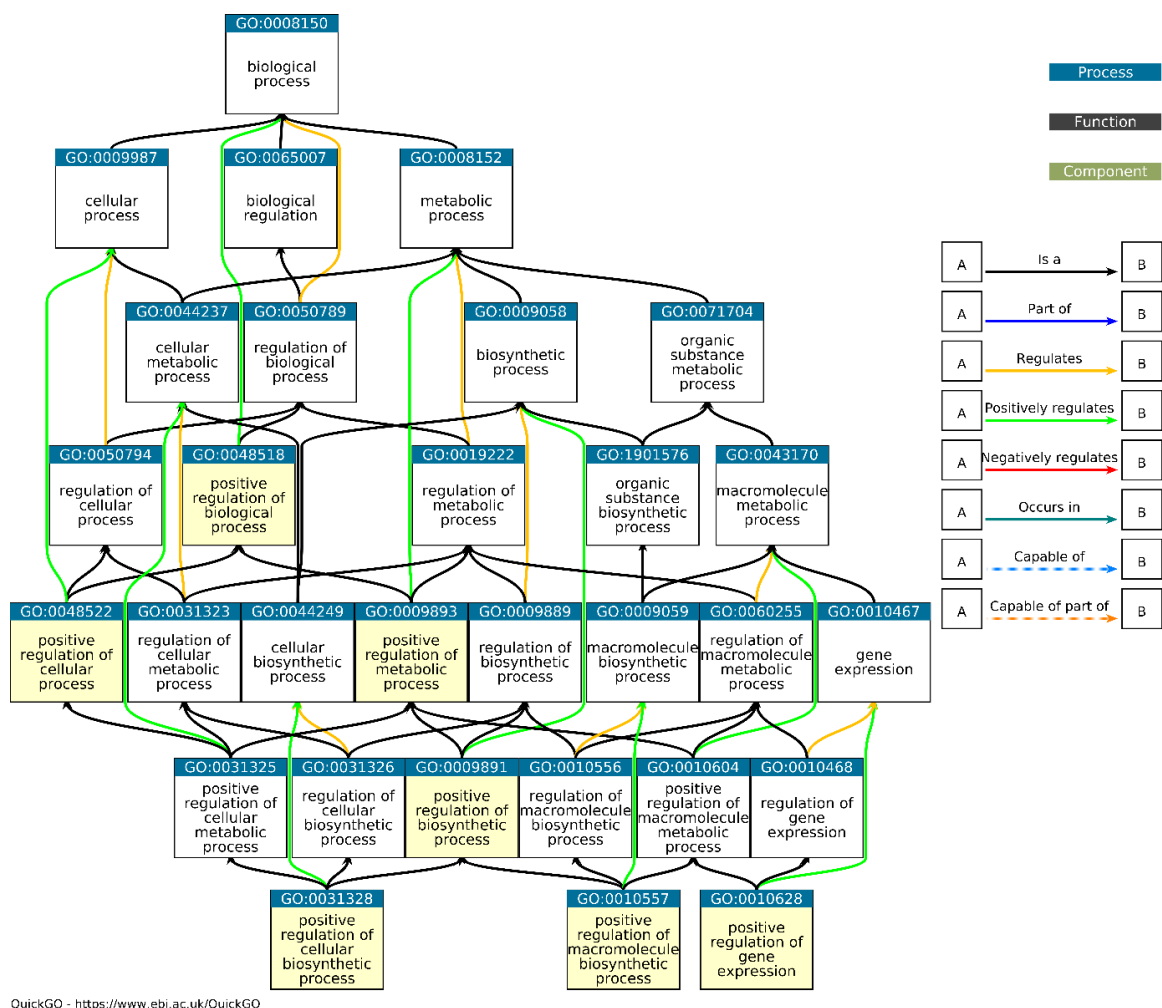


Figure 4.7: Biological Processes involving *Pentatricopeptide Repeat-Containing Protein (At3g42630)*

Description: *At3g42630* involves in positive regulation of biological processes.

Keys: The yellow colour filled box are the terms associated with the DEGs with 95% confidence.

Source: (Binns *et al.*, 2009)

SPX domain-containing protein 1 (SPX1) is upregulated in non-trunking sago palms leaf tissue and are involves in GO-Terms related to the biological processes (**Figure 4.8**). The flows of the biological processes show that *SPX1* involved in the positive regulations of cellular processes and response to stimulus. *SPX1* is located in the nucleus and regulates gene transcription by competitive binding to deactivate transcription factors (Puga *et al.*,

2014). *SPX1* are regulated by the phosphate starvation response 1 (PHR1) after activation in Pi deficiency environment (Puga *et al.*, 2014). Upregulation of *SPX1* also increase the transcription of inorganic phosphate (Pi) starvation-inducible genes, *S-like ribonuclease* (*RNS1*), *acid phosphatase 5* (*ACP5*), and *transcription factor* (*MYB90*) which facilitate Pi recycling and reactive oxygen species scavenging in Arabidopsis (Duan *et al.*, 2008). The upregulation of *SPX1* suggested that non-trunking sago palms were under phosphate starvation.

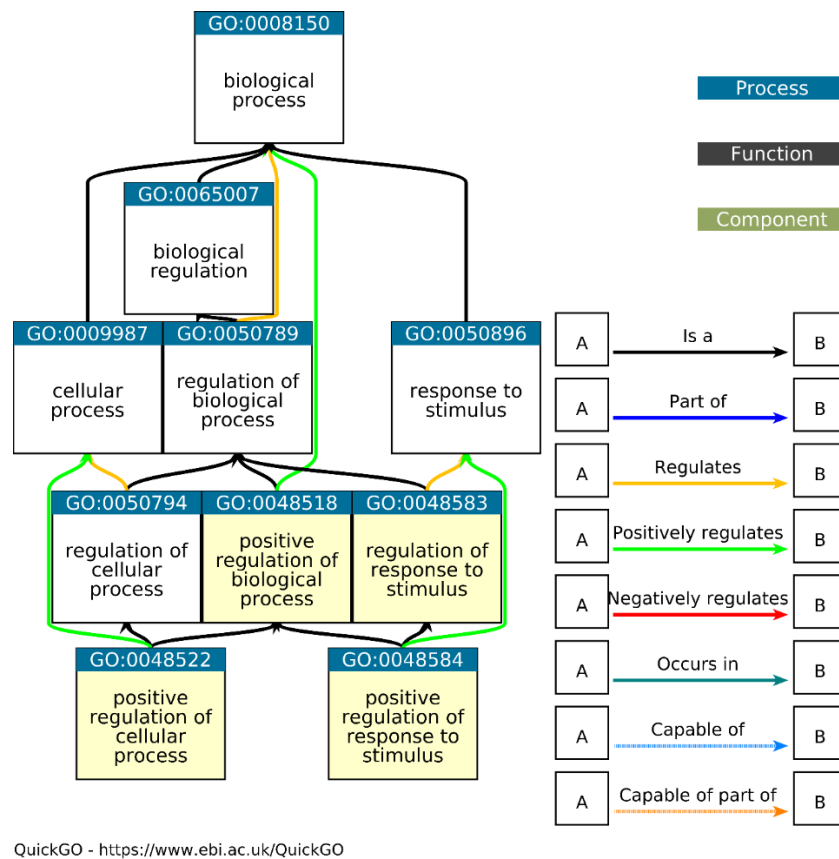


Figure 4.8: Biological Processes involved by *SPX Domain-Containing Protein 1 (SPX1)*

Description: *SPX1* involves in positive regulation of biological processes and regulation of respond to stimulus.

Keys: The yellow colour filled box are terms associated with the DEGs with 95% confidence.

Source: (Binns *et al.*, 2009)

Translation factor GUF1 homolog, chloroplastic (LEPA) is upregulated in non-trunking sago palms leaf tissue and are involved in seven biological processes in GO enrichment (**Figure 4.9**). The flows of the biological processes show that *LEPA* involved in the positive regulations of cellular biosynthesis, macromolecules biosynthesis and gene expression. *LEPA* is an essential translation factor that enhance the efficiency of protein synthesis in chloroplast (Ji *et al.*, 2012). *LEPA* were downregulated in *Arabidopsis thaliana* when infected by cabbage leaf curl virus (CaLCuV) (Ascencio-Ibáñez *et al.*, 2008). The upregulation of *LEPA* suggested that non-trunking sago palms were increased in metabolisms.

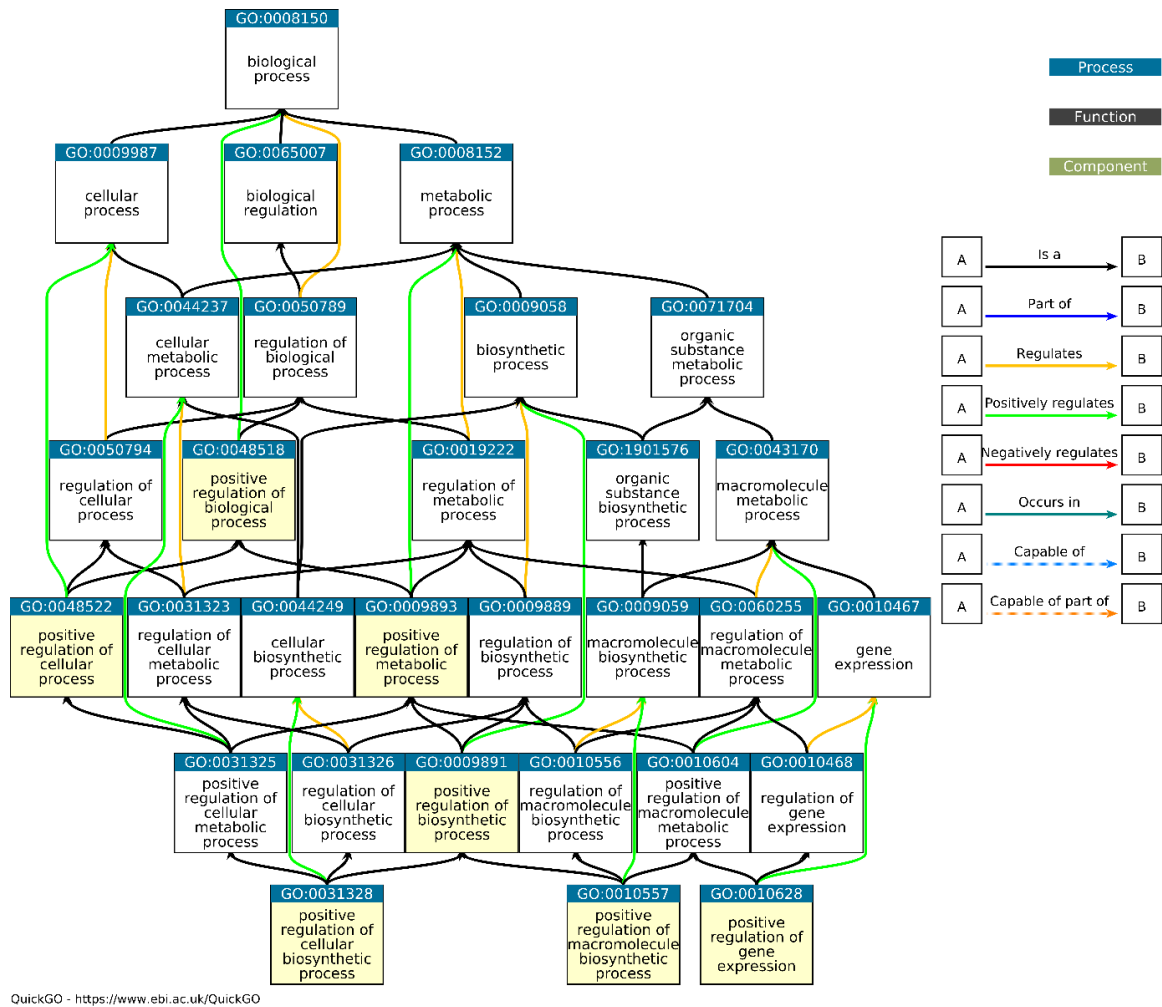


Figure 4.9: Biological Processes involved by *Translation Factor GUF1 Homolog, Chloroplastic (LEPA)*

Description: *LEPA* involves in positive regulation of biological processes.

Keys: The yellow colour filled box are the terms associated with the DEGs with 95% confidence.

Source: (Binns *et al.*, 2009)

Nuclear transcription factor Y subunit C-1 (NFYC1) are downregulated in non-trunking sago palms leaf tissue and are involved in all twenty biological processes in GO enrichment (**Table 4.11**). NFYC1 is a subunit of NFY/CCAAT complexes which bind to CCAAT box in the promoter regions of target genes via nuclear transcription factor Y subunit A (NFYA) subunit in regulating gene expression (Zhao *et al.*, 2017). Other than

NFY/CCAAT complexes, NGYC1 is able to interact with other transcription factors to regulate different gene expressions (Zhao *et al.*, 2017). There are three reported functions of nuclear transcription factor Y subunit C-1 which are positively regulate hypocotyl elongation by interacting with histone deacetylase 15 (HDA15), positively regulate flowering time by interacting with zinc finger protein CONSTANS (CO) which then upregulate the transcript levels of flowering locus T (FT) and suppressor of overexpression of CO1 (SOC1), and positively regulate freezing responses (Zhao *et al.*, 2017). The downregulation of NFYC1 suggested that the non-trunking sago palm were under growth suppression.

MADS-box transcription factor 50-like (MADS50) are downregulated in non-trunking sago palms leaf tissue and are involved in twelve biological processes in GO enrichment (**Figure 4.10**). The flows of the biological processes show that MADS50 involved in the positive regulations of DNA transcription and gene expression. MADS-domain proteins bind DNA as homo or heterodimers together with another proteins as function tetramers complexes in order to regulate transcription of target genes (Castelán-Muñoz *et al.*, 2019). MADS-box transcription factor 50-like, also known as suppressor of CONSTANS overexpression 1-like (SOC1) involved in positive regulation of bud development, negative regulation of low temperature germination, positive regulation of flowering time, and negative regulation of cold response (Castelán-Muñoz *et al.*, 2019). The downregulation of MADS-box transcription factor 50-like (MADS50) suggested that the non-trunking sago palm were under growth suppression.



Description: *MADS50* involves in positive regulation of biological processes, metabolic processes and gene expression.

Source: (Binns *et al.*, 2009)

91

biological processes show that *CYPs* involved in the positive regulations of biosynthesis processes. *CYPs* are oxidoreductase enzyme containing heme-thiolate cofactor as catalytic site (Pandian *et al.*, 2020). Plants *CYPs* involve in diverse functions such as xenobiotic metabolism, hormone regulation, secondary metabolite biosynthesis, antioxidant biosynthesis, fatty acid metabolism, and biotic and abiotic stress regulation (Pandian *et al.*, 2020). The upregulation of *CYPs* suggested that the non-trunking sago palm were under stressed and increased in metabolisms.

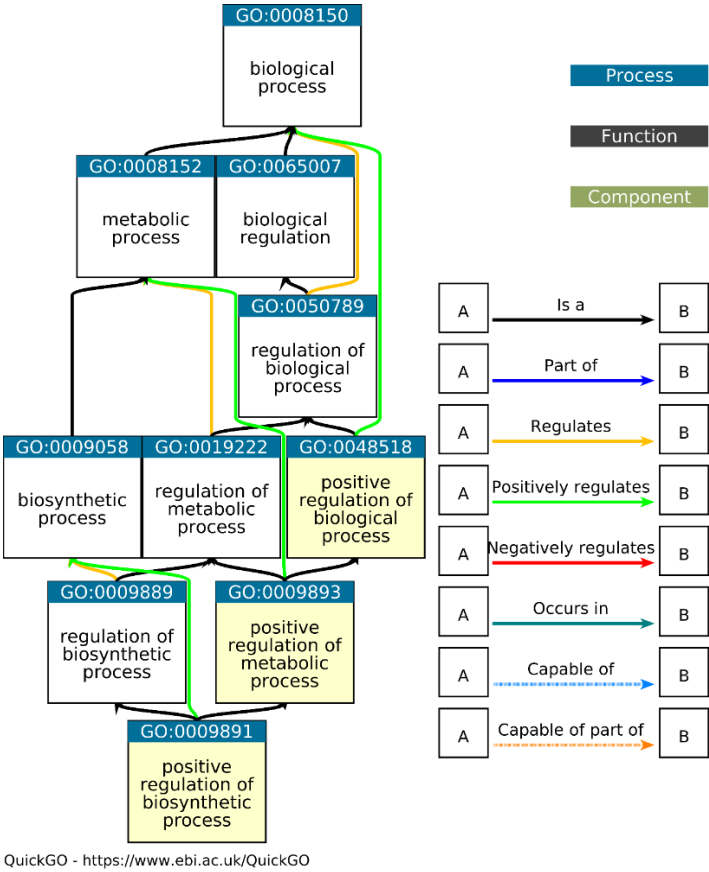


Figure 4.11: Biological Processes involved by *Cytochrome P450 (CYPs)*

Description: *CYPs* involves in positive regulation of biological processes.

Keys: The yellow colour filled box are the terms associated with the DEGs with 95% confidence.

Source: (Binns *et al.*, 2009)

LSD ONE LIKE 1 (LOL1) is upregulated in non-trunking sago palms leaf tissue and are involved in a single biological process in GO enrichment (**Figure 4.12**). The flows of the biological processes show that *LOL1* involved in the regulations of response to stimulus. *LOL1* is a transcript factor which interact with targeted gene to regulate expression (Yeh *et al.*, 2011). *LOL1* are downregulated during senescence (Ascencio-Ibáñez *et al.*, 2008), downregulated during heat stress acclimation (Larkindale & Vierling, 2008; Yeh *et al.*, 2011), downregulated during drought stress (Yong *et al.*, 2013), positive regulation of plant growth (Yeh *et al.*, 2011), and positive regulation of disease resistance (Yeh *et al.*, 2011). The upregulation of *LOL1* suggested that the non-trunking sago palm were under disease infections and plant growth.

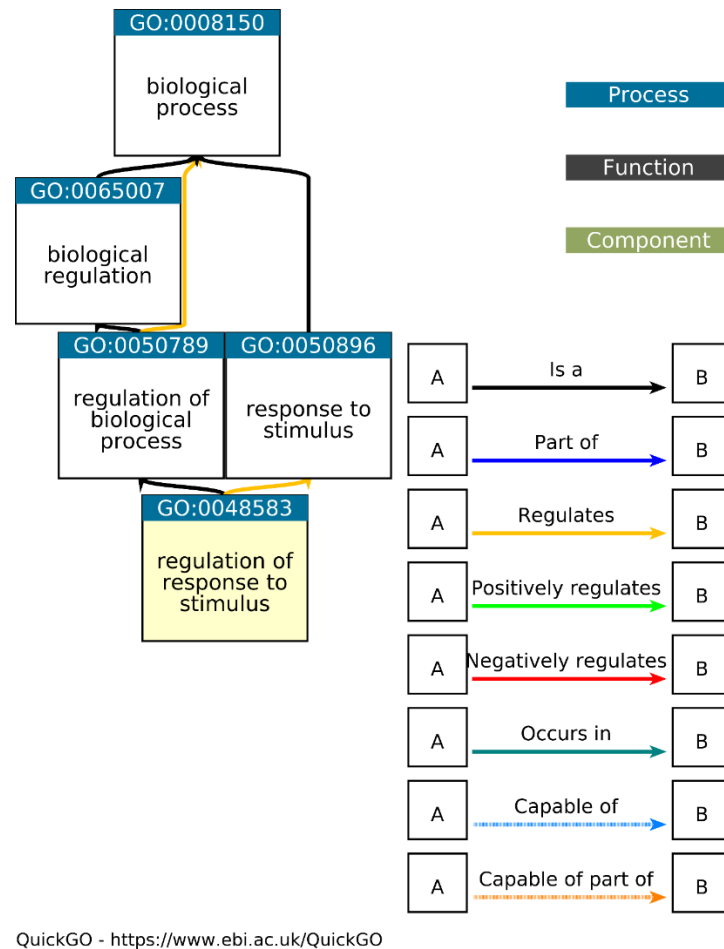


Figure 4.12: Biological Processes involved by *LSD ONE LIKE 1 (LOL1)*

Description: *LOL1* involves in regulation of response to stimulus.

Keys: The yellow colour filled box are the terms associated with the DEGs with 95% confidence.

Source: (Binns *et al.*, 2009)

The downregulation of pentatricopeptide repeat-containing protein (At3g42630), nuclear transcription factor Y subunit C-1 (NFYC1), and MADS-box transcription factor 50-like (MADS50) indicated that the non-trunking sago palms growth were suppressed. On the other hand, the upregulations of SPX domain-containing protein 1 (SPX1), translation factor GUF1 homolog chloroplastic (LEPA), cytochrome P450 (CYPs), and LSD ONE

LIKE 1 (LOL1) suggested that the non-trunking sago palms were under disease infections, increased in metabolisms and were under phosphate starvation.

The GO enrichment analysis shows genes involved in cellular component term are not significantly different between the trunking and non-trunking groups (**Appendix R**). Hence none of the GO-Term from cellular component were further discussed. DEGs of GO enrichment analysis result in molecular functions on the other hand shows that the GO-Term “secondary active transmembrane transporter activity”, “inorganic anion transmembrane transporter activity”, “inorganic anion exchanger activity”, “ammonium transmembrane transporter activity”, “heme binding”, and “peroxidase activity”, are significantly different between the trunking and non-trunking groups (**Table 4.12, Figure 4.13, & Appendix S**).

Table 4.12: Molecular Functions of the DEGs between Trunking and Non-Trunking Sago Palm, Unigenes involved and its Annotations based on Gene Ontology Analysis

Description: Selection of the significant GO-Terms from the DEGs obtained by DESeq2 were based on hypergeometric distribution which adjusted by the Bonferroni correction ($p\text{-adjusted} \leq 0.01$)

Molecular Functions	DEGs	NCBI Annotation
secondary active transmembrane transporter activity	CL2107.Contig2_All	LOW QUALITY PROTEIN: low affinity sulfate transporter 3-like
	CL2723.Contig1_All	protein DETOXIFICATION 34
	CL3065.Contig3_All	sulfate transporter 4.1, chloroplastic-like isoform X2
	CL3065.Contig4_All	sulfate transporter 4.1, chloroplastic-like isoform X2

Table 4.12 continued

	CL5711.Contig1_All	probable boron transporter 2
	CL5711.Contig2_All	probable boron transporter 2
	CL5711.Contig3_All	probable boron transporter 2
	Unigene19212_All	protein DETOXIFICATION 34
	Unigene25704_All	H(+) antiporter 18-like
	Unigene397_All	protein DETOXIFICATION 27-like
	Unigene7135_All	protein STAR1
	Unigene7260_All	inorganic phosphate transporter 2-1, chloroplastic
Inorganic anion transmembrane transporter activity	CL2107.Contig2_All	LOW QUALITY PROTEIN: low affinity sulfate transporter 3-like
	CL3065.Contig3_All	sulfate transporter 4.1, chloroplastic-like isoform X2
	CL3065.Contig4_All	sulfate transporter 4.1, chloroplastic-like isoform X2
	CL5119.Contig1_All	sulfate transporter 3.1-like
	CL5711.Contig1_All	probable boron transporter 2
	CL5711.Contig2_All	probable boron transporter 2
	CL5711.Contig3_All	probable boron transporter 2
Inorganic anion exchanger activity	CL5711.Contig1_All	probable boron transporter 2
	CL5711.Contig2_All	probable boron transporter 2
	CL5711.Contig3_All	probable boron transporter 2
Ammonium transmembrane transporter activity	CL181.Contig1_All	ammonium transporter 3 member 1-like

Table 4.12 continued

	CL9493.Contig1_All	Ammonium_transp domain-containing protein
	CL9493.Contig2_All	ammonium transporter 3 member 1-like
	CL9493.Contig3_All	ammonium transporter 3 member 1-like
Heme binding	CL10615.Contig2_All	predicted protein
	CL1983.Contig3_All	trans-cinnamate 4-monooxygenase
	CL2730.Contig1_All	peroxidase 15
	CL2730.Contig2_All	peroxidase A2-like
	CL2730.Contig3_All	peroxidase 15
	CL3975.Contig1_All	predicted protein
	CL5284.Contig1_All	cytochrome p450
	CL599.Contig1_All	catalase
	CL9602.Contig1_All	peroxidase 4
	Unigene10336_All	hypothetical protein GUITHDRAFT_79406
	Unigene17377_All	hypothetical protein H257_07130
	Unigene18429_All	catalase
	Unigene20852_All	catalase
	Unigene24588_All	cytochrome P450 711A1-like
	Unigene26343_All	L-ascorbate peroxidase 1, cytosolic isoform X1
	Unigene39385_All	peroxidase A2-like
	Unigene5008_All	Cytochrome c

Table 4.12 continued

Peroxidase activity	CL2730.Contig1_All	peroxidase 15
	CL2730.Contig2_All	peroxidase A2-like
	CL2730.Contig3_All	peroxidase 15
	CL599.Contig1_All	catalase
	CL9602.Contig1_All	peroxidase 4
	Unigene18429_All	catalase
	Unigene20852_All	catalase
	Unigene26343_All	L-ascorbate peroxidase 1, cytosolic isoform X1
	Unigene39385_All	peroxidase A2-like
	Unigene5924_All	chloroperoxidase
	Unigene8449_All	respiratory burst oxidase homolog protein A

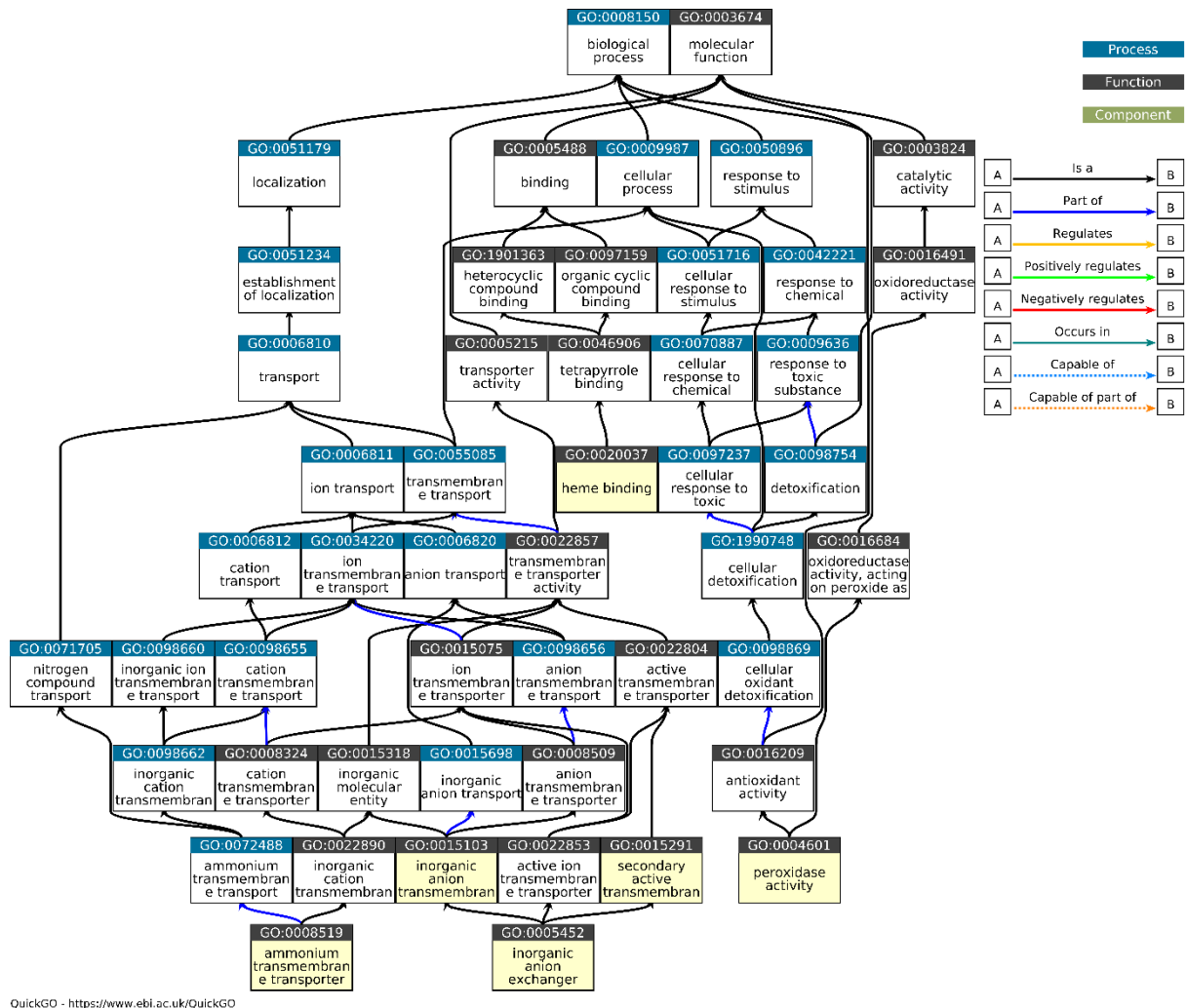


Figure 4.13: GO Ancestor Chart of DEGs between Trunking and Non-Trunking Sago Palm Leaf Tissue based on Molecular Function

Description: The molecular functions in heme binding, peroxidase activities, secondary active transmembrane transporter, inorganic anion transmembrane transporter and ammonium transmembrane transporter are differentially expressed in non-trunking sago palm.

Keys: The yellow colour filled box are the terms associated with the DEGs with 99% confidence.

Source: (Binns *et al.*, 2009)

DEGs from GO-Term “inorganic anion transmembrane transporter activity” and “inorganic anion exchanger activity” are a subset of “secondary active transmembrane transporter activity” where all the DEGs listed in the GO-Term “inorganic anion

transmembrane transporter activity” and “inorganic anion exchanger activity” presents in the “secondary active transmembrane transporter activity” DEGs list. The DEGs in “secondary active transmembrane transporter” consist of *plant sulphur transporters (SULTRs)*, *boron transporter (BOR)*, *protein DETOXIFICATION (DTX)*, *H(+) antiporter (CHX)*, *sensitive to Al rhizotoxicity1 (STAR1)* and *inorganic phosphate transporter (PHT)*.

In sulfate transporter (SLC26A/SulP transporter) group, *SULTR4;1* and *SULTR3;1* were upregulated while *SULTR2* were downregulated in non-trunking sago palm leaf tissue (**Table 4.13**). *SULTR2* facilitate the source to sink sulphate transport, *SULTR3;1* facilitate the uptake of sulphate in cytoplasm into chloroplast and *SULTR4;1* regulate the release of sulphate in vacuole to cytoplasm (Takahashi, 2019). Upregulation of *SULTR3;1* and *SULTR4;1* while downregulation of *SULTR2* in shoot suggest sulphur starvation of the plant.

DEGs in *boron transporters (BORs)* group were upregulated in non-trunking sago palm leaf tissue (**Table 4.13**). *BOR2* facilitate the efflux of borate from root to shoot under boron-deficiency stress but a lower efficiency than *BOR1* (Miwa *et al.*, 2013). *BOR2* also positively regulates cross-linking of Rhamnogalacturonan II for cell wall integrity and root elongation (Miwa *et al.*, 2013). Other than *BOR2*, expression of *BOR4* in non-trunking sago palm leaf tissue was also upregulated but not listed in the significant GO enrichment list. Function of *BOR4* antagonised *BOR2* where it involved in boron tolerance by regulating boron exclusion from tissues and cells under boron-excessive stress (Yoshinari & Takano, 2017). Upregulation of both *BOR2* and *BOR4* in non-trunking sago palms leaf tissue suggested that the plants were under boron-excessive stress and improving cell wall integrity.

DEGs in *protein DETOXIFICATION (DTX)/multidrug and toxic compound extrusion (MATE)* group were upregulated in non-trunking sago palm leaf tissue (**Table 4.13**). DTX/MATE are unidirectional membrane Na^+/H^+ antiporter which transport diverse substrates in plant (Takanashi *et al.*, 2014). DTX/MATE is involved in: xenobiotic efflux; alkaloid accumulation; flavonoid accumulation; Fe translocation; Al^{3+} detoxification; plant hormone signalling; substrates transportation; plant growth and development; negatively regulates disease resistance; positively regulates drought stress; positively regulates salt stress; and positively regulates cold stress (Lu *et al.*, 2019; Takanashi *et al.*, 2014). DTX34 involves in pollen maturation (Bock *et al.*, 2006) and germination (Wang *et al.*, 2008) in floral tissue of *Arabidopsis thaliana*. DTX27 marginally upregulated in response to cabbage leaf curl virus (CaLCuV) infection (Ascencio-Ibáñez *et al.*, 2008). Other DTXs from the DEGs which are not included in the GO-Term list are DTX21, DTX18 and DTX42. DTX21 are effector-induced herbicide-mediated protein (de Torres-Zabala *et al.*, 2007; Ramel *et al.*, 2007). DTX18 is a hydroxycinnamic acid amide transporter which co-expressed with p-coumaroyl-CoA:agmatine N4-p-coumaroyl transferase (ACT) to inhibit the germination of pathogen spore (Dobritsch *et al.*, 2016). DTX18 also upregulated in respond to osmotic stress (de Torres-Zabala *et al.*, 2007), upregulation corresponding to pathogenic defence hormones (jasmonic acid and ethylene) (Pré *et al.*, 2008) and down regulated in respond to cytokinin (Lee *et al.*, 2007). DTX42 is citrate efflux transporter which regulate aluminium toxicity and is regulated by zinc finger protein STOP1 (Liu *et al.*, 2009). Upregulation of *DTX/MATE* in non-trunking sago palms suggested that the plants were active in substrate/signal transportation to combat biotic or/and abiotic stresses.

DEGs in *cation/H(+) antiporter (CHX)* were downregulated in non-trunking sago palm leaf tissue (**Table 4.13**). CHX18 located in pre-vacuolar compartment (Salil *et al.*,

2011). CHX18 facilitates K^+ homeostasis under alkaline condition, positively regulates alkaline tolerance, positively regulates cationic drug hygromycin B tolerance (Salil *et al.*, 2011), Zn homeostasis in roots (Talke *et al.*, 2006). CHX18 express marginally during initial pollen development (microspores, bicellular pollen), root and suspension cell culture (Sze *et al.*, 2004). The cause and effect of downregulated *CHX* in non-trunking sago palms are unclear. This respond may better explain when more studies were done on the functions of CHX in plant leaf.

DEGs of *STAR1* was upregulated in non-trunking sago palm leaf tissue (**Table 4.13**). STAR1 contain a nucleotide binding domain (NBD) without transmembrane domains (TMDs). STAR1 is involved in aluminium toxicity stress by forming a ABC transporter complex with protein STAR2 to efflux uridine diphosphate glucose (UDP-glucose) in root (Huang *et al.*, 2009). Al resistance transcription factor 1 (ART1) positively regulates STAR1 expression (Tsutsui *et al.*, 2011) but are not due to aluminium induction. On the other hand, abscisic stress-ripening protein 5 (ASR5) transcription factor positively regulates STAR1 expression in respond to aluminium induction (Arenhart *et al.*, 2014). STAR1 is also involved in inorganic phosphate influx (Belal *et al.*, 2015). Upregulation of *STAR1* in non-trunking sago palms suggested that the plants were under aluminium toxicity and phosphate starvation.

DEGs of *inorganic phosphate transporter (PHT)* was upregulated in non-trunking sago palm leaf tissue (**Table 4.13**). PHT is a Pi/H^+ symporter which consists of twelve transmembrane domains, a long extracellular hydrophilic N terminus, and a hydrophilic loop between transmembrane-8 and transmembrane-9. PHT2;1 is a low affinity Pi/H^{1+} symporter which express predominantly in plant chloroplast inner membrane (Versaw & Harrison,

2002). *PHT2;1* did not directly respond to Pi starvation but modulates Pi-starvation responses and the translocation of Pi in leaves.(Daram *et al.*, 1999; Versaw & Harrison, 2002). *PHT1;5* was downregulated in non-trunking sago palm leaf tissue. *PHT1;5* involved in inorganic phosphate efflux and translocation (Mudge *et al.*, 2002). Upregulation of *PHT2;1* while downregulation of *PHT1;5* in non-trunking sago palms suggest that the plants were under phosphate starvation.

Table 4.13: Grouping, Annotation and Expression of DEGs which are Involved in Secondary Active Transmembrane Transporter GO-Term by Gene Ontology Analysis

Description: *Sulfate transporters, boron transporter, DTX/MATE, CHX, STAR1* and *PHT* are differentially expressed in non-trunking sago palm.

Gene Group	DEGs	Log ₂ (Foldchange)	Annotation
<i>SLC26A/SulP transporter (TC 2.A.53)</i>	CL2107.Contig2_All	-1.14	<i>low affinity sulfate transporter 3-like (SULTR2) (TC 2.A.53.1.4)</i>
	CL3065.Contig3_All	1.79	<i>sulfate transporter 4.1, chloroplastic-like isoform X2 (SULTR4;1) (TC 2.A.53.1.14)</i>
	CL3065.Contig4_All	1.55	
	CL5119.Contig1_All	2.18	<i>PREDICTED: sulfate transporter 3;1-like (SULTR3.1)</i>
<i>Boron Transporter (TC 2.A.31.3)</i>	CL5711.Contig1_All	2.27	<i>PREDICTED: probable boron transporter 2 [Elaeis guineensis]</i>
	CL5711.Contig2_All	3.38	
	CL5711.Contig3_All	3.49	

Table 4.13 continued

<i>protein DETOXIFICATION (DTX)/multidrug and toxic compound extrusion (MATE) protein</i>	Unigene19212_All	2.90	PREDICTED: <i>protein DETOXIFICATION 34 [<u>Elaeis guineensis</u>]</i>
	CL2723.Contig1_All	1.92	PREDICTED: <i>protein DETOXIFICATION 34 [<u>Elaeis guineensis</u>]</i>
	Unigene397_All	1.39	PREDICTED: <i>protein DETOXIFICATION 27-like [<u>Brachypodium distachyon</u>]</i>
<i>cation/H(+) antiporter (CHX)</i>	Unigene25704_All	-1.80	<i>cation/H(+) antiporter 18-like [<u>Juglans regia</u>] (TC 2.A.37)</i>
<i>sensitive to Al rhizotoxicity1 (STAR1)</i>	Unigene7135_All	1.98	PREDICTED: <i>protein STAR1 [<u>Elaeis guineensis</u>]</i>
<i>inorganic phosphate transporter (PHT)</i>	Unigene7260_All	1.45	PREDICTED: <i>inorganic phosphate transporter 2-1,</i>

Table 4.13 continued

			<i>chloroplastic [Elaeis guineensis]</i>
--	--	--	--

DEGs from ammonium transmembrane transporter activity group consist only the *ammonia transmembrane transporter (AMT3;1)*. The upregulation of *AMT3;1* indicates N-starvation in paddy (**Table 4.14**) (Li *et al.*, 2012). Upregulation of *AMT3;1* in non-trunking sago palms suggest that the plants were under nitrate starvation.

Table 4.14: Grouping, Annotation and Expression of DEGs Which are Involved in Ammonium Transmembrane Transporter Activity GO-Term by Gene Ontology Analysis

Description: *AMT3;1* is differentially expressed in non-trunking sago palm.

Gene Group	DEGs	Log ₂ (Foldchange)	Annotation
<i>Ammonia Transmembrane Transporter (AMT3;1)</i>	CL181.Contig1_All	1.63	<i>ammonium transporter 3 member 1-like</i>
	CL9493.Contig1_All	2.80	
	CL9493.Contig2_All	3.40	
	CL9493.Contig3_All	3.87	

DEGs from peroxidase activity group are peroxidases (EC 1.11.1.-), which utilises peroxide as acceptor to perform oxidoreduction process without incorporating oxygen atom into the product, are grouped into five groups namely *peroxidase (PER) family* which targeting phenolic donor, *catalase (CAT) family* which reduce peroxides, *L-ascorbate peroxidase (APX) family* which targeting L-ascorbate, *chloroperoxidase (CPO) family*

which targeted chloride donor and *respiratory burst oxidase homolog protein (RBOH)* family transporter protein which perform peroxidase and NADPH oxidase activity (**Table 4.15**). *PER* protein involves in wide range of processes such as cell wall metabolism, lignification, suberisation, auxins metabolism, wound healing, reactive oxygen species (ROS) and reactive nitrogen species (RNS) metabolism, fruit growth and ripening, defence against pathogens, plant growth, biotic and abiotic stress regulation, signalling regulation and seed germination (Pandey *et al.*, 2017). *CAT* protein involves in oxidative stress regulation (Pandey *et al.*, 2017). *APX* involved in oxidative stress regulation and photo-protection (Pandey *et al.*, 2017). *CPO* protein involves in positive regulation of fungal infection (Jacks *et al.*, 2000). *RBOH* protein involves in plant development, immunity, and abiotic stress regulation (Hu *et al.*, 2020; Kobayashi *et al.*, 2006). Upregulation of *PER*, *CAT*, *APX* and *RBOH* while downregulation of *CPO* in non-trunking sago palms suggest that the plants were under oxidative stresses and may also involve in cell wall reinforcement and pathogen defence.

Table 4.15: Grouping, Annotation and Expression of DEGs which are Involved in Peroxidase Activity GO-Term by Gene Ontology Analysis

Description: *PER*, *CAT*, *APX*, *CPO* and *RBOH* are differentially expressed in non-trunking sago palm.

Gene Group	DEGs	Log ₂ (Foldchange)	Annotation
<i>peroxidase</i> (<i>PER</i>) family. Classical plant peroxidase (class III) subfamily (EC:1.11.1.7)	CL2730.Contig1_All	2.08	PREDICTED: <i>peroxidase 15 [Phoenix dactylifera]</i>
	CL2730.Contig2_All	1.83	PREDICTED: <i>peroxidase A2-like [Musa acuminata subsp. malaccensis]</i>

Table 4.15 continued

	CL2730.Contig3_All	2.09	PREDICTED: <i>peroxidase 15 [Phoenix dactylifera]</i>
	CL9602.Contig1_All	2.91	PREDICTED: <i>peroxidase 4 [Elaeis guineensis]</i>
	Unigene39385_All	2.04	PREDICTED: <i>peroxidase A2-like [Musa acuminata subsp. malaccensis]</i>
<i>Catalase (CAT)</i> family	CL599.Contig1_All	3.28	<i>catalase [Fistulifera solaris]</i>
plant peroxidase (class I) subfamily (EC:1.11.1.6)	Unigene18429_All	-3.85	<i>catalase [Fistulifera solaris]</i>
	Unigene20852_All	-3.45	<i>catalase [Fistulifera solaris]</i>
<i>Ascorbate peroxidase (APX)</i> family plant peroxidase (class I) subfamily (EC:1.11.1.11)	Unigene26343_All	1.20	PREDICTED: <i>L-ascorbate peroxidase 1, cytosolic isoform XI [Phoenix dactylifera]</i>
<i>chloroperoxidase (CPO)</i> family (EC:1.11.1.10)	Unigene5924_All	-4.31	<i>chloroperoxidase [Phytophthora sojae]</i>

Table 4.15 continued

<i>respiratory burst oxidase homolog protein (RBOH) family (TC 5.B.1.3)</i> (EC:1.11.1.-, EC:1.6.3.-)	Unigene8449_All	3.40	PREDICTED: <i>respiratory burst oxidase homolog protein A [Elaeis guineensis]</i>
--	-----------------	------	--

DEGs from heme binding group consist of some DEGs from peroxidases group, delta-9 acyl-lipid desaturase (ADS), trans-cinnamate 4-monooxygenase/cinnamic acid 4-hydroxylase (C4H), and cytochrome P450 (CYPs) (**Table 4.16**). ADS protein involves in regulating fatty acid biosynthesis which specificity varies by subcellular compartment and the expression dependent to plant organs (Fukuchi-Mizutani *et al.*, 1998; Heilmann *et al.*, 2004; Yao *et al.*, 2003). C4H protein involve in the phenylpropanoid biosynthesis and upregulated by wound, light and pathogenic infection (Kawai *et al.*, 1996). The role of CYP and peroxidases proteins were mentioned in biological processes of GO enrichment and peroxidase activity GO-Term earlier respectively.

Table 4.16: Grouping, Annotation and Expression of DEGs Involved in Heme Binding GO-Term by Gene Ontology Analysis

Description: *ADS*, *C4H*, *PER*, *CAT*, *APX* and *CYPs* are differentially expressed in non-trunking sago palm.

Gene Group	DEGs	Log ₂ (Foldchange)	Annotation
<i>Delta-9 acyl-lipid desaturase</i> (<i>ADS</i>)	CL10615.Contig2_All	4.93	<i>Delta-9 acyl-lipid desaturase 1</i> OS= <i>Arabidopsis thaliana</i> GN=ADS1 PE=2 SV=1
	CL3975.Contig1_All	-4.48	<i>Delta-9 acyl-lipid desaturase 1</i> OS= <i>Arabidopsis thaliana</i> GN=ADS1 PE=2 SV=1
<i>Trans-cinnamate 4-monooxygenase/ cinnamic acid 4-hydroxylase</i> (<i>C4H</i>)	CL1983.Contig3_All	1.72	<i>Trans-cinnamate 4-monooxygenase</i> OS= <i>Populus kitakamiensis</i> GN=CYP73A16 PE=2 SV=1
<i>peroxidase</i> family. Classical plant (class III) peroxidase subfamily (<i>PER</i>) (EC:1.11.1.7)	CL2730.Contig1_All	2.08	PREDICTED: <i>peroxidase 15</i> [<i>Phoenix dactylifera</i>]

Table 4.16 continued

	CL2730.Contig2_All	1.83	PREDICTED: <i>peroxidase A2-like</i> [<i>Musa acuminata</i> <i>subsp. malaccensis</i>]
	CL2730.Contig3_All	2.09	PREDICTED: <i>peroxidase 15</i> [<i>Phoenix dactylifera</i>]
	CL9602.Contig1_All	2.91	PREDICTED: <i>peroxidase 4</i> [<i>Elaeis</i> <i>guineensis</i>]
	Unigene39385_All	2.04	PREDICTED: <i>peroxidase A2-like</i> [<i>Musa acuminata</i> <i>subsp. malaccensis</i>]
catalase family (EC:1.11.1.6)	CL599.Contig1_All	3.28	<i>catalase</i> [<i>Fistulifera</i> <i>solaris</i>]
	Unigene18429_All	-3.85	<i>catalase</i> [<i>Fistulifera</i> <i>solaris</i>]
	Unigene20852_All	-3.45	<i>catalase</i> [<i>Fistulifera</i> <i>solaris</i>]
Ascorbate peroxidase family (EC:1.11.1.11)	Unigene26343_All	1.20	PREDICTED: <i>L-</i> <i>ascorbate peroxidase</i> <i>1, cytosolic isoform XI</i> [<i>Phoenix dactylifera</i>]
Cytochrome P450 (CYPs)	CL5284.Contig1_All	-3.83	<i>cytochrome p450</i> [<i>Nannochloropsis</i> <i>gaditana CCMP526</i>] >EKU22991.1

Table 4.16 continued

	Unigene10336_All	4.84	<i>Cytochrome P450 71A9 OS=Glycine max GN=CYP71A9 PE=2 SV=1</i>
	Unigene17377_All	-3.77	<i>Cytochrome P450 714B2 OS=Oryza sativa subsp. japonica GN=CYP714B2 PE=1 SV=2</i>
	Unigene24588_All	4.61	PREDICTED: <i>cytochrome P450 711A1-like [Brachypodium distachyon]</i>
	Unigene5008_All	3.99	<i>Cytochrome c [Arabidopsis thaliana]</i>

From the molecular function in GO enrichment point of view, the DEGs expression in non-trunking sago palms suggested that the plants were under sulphur and nitrate starvation shown by the upregulation of *AMT3;1*, *SULTR3;1*, and *SULTR4;1* and the downregulation of *SULTR2*. The analysis also suggested that the non-trunking sago palms were under phosphate starvation and metabolism shown by the upregulation of *STAR1* and expression of *PHTs*. At the same time, the non-trunking sago palms were under boron and aluminium toxicity stress shown by the upregulation of *BOR4*, *DTXs* and *STAR1*. The analysis also suggested that the non-trunking sago palms performed cell wall reinforcement shown by the upregulation of *BOR2*, *PER* and *C4H*. Pathogenic stress was shown in the

non-trunking sago palms by the upregulation of *DTX*, *RBOH* and *PER*. Oxidative stress was shown in the non-trunking sago palms by the CAT, APX, and PER expression. The involvement of fatty acid and phenylpropanoid biosynthesis pathways in non-trunking sago palms was shown by the expression of *ADS* and *C4H*, respectively.

Based on the observation of gene expression in biological processes and molecular functions in GO enrichment, the non-trunking sago palm was under mineral starvation (phosphate, sulphate and nitrate) and toxicity (aluminium and boron), resulting in plant growth suppression and weakened its pathogenic resistance. The pathogenic susceptible non-trunking sago palm responded to the pathogenic infection by increasing the building block's metabolism for reinforcing the cell wall, increasing pathogenic suppression and wound healing.

4.3.7 DEGs Pathway Analysis by KEGG.

Pathway analysis of the DEGs with KEGG database annotated 335(35.79%) genes distributed to the 22 DEGs in cellular processes pathways, 34 DEGs in environmental information processing pathways, 75 DEGs in genetic information processing pathways, 468 DEGs in metabolism pathways and 38 DEGs in organismal system pathways (**Appendix T**). The gene expression of the DEGs involved in the top 30 KEGG pathway are inconsistent where there were upregulated and downregulated genes from the same pathways (**Appendix U**). The involved pathways were then narrow down to the top 20 based on the number of DEGs involved, Q-values and the rich factor (**Appendix V**). In order to keep the false discovery rate to 1% or lower, Q-Value > 0.01 cut-off were performed and only phenylpropanoid biosynthesis, biosynthesis of secondary metabolites, benzoxazinoid biosynthesis, stilbenoid, diarylheptanoid and gingerol biosynthesis, flavonoid biosynthesis, and circadian rhythm were selected for further discussion (**Appendix W**). The biosynthesis

of secondary metabolite pathway is not further discussed accounting for its universal coverage which make the prediction of the products involved by the DEGs impossible and hence the contribution of the DEGs from the pathway to the growth of sago palm cannot be assumed.

The KEGG pathway analysis of the DEGs of the non-trunking sago palm involves in phenylpropanoid biosynthesis shows inconsistent expression (**Table 4.17 & Figure 4.14**). The flow of the enzymatic reaction in phenylpropanoid biosynthesis suggesting the upregulated production of monolignol which is the monomer for lignin biosynthesis. Lignin is the main component of cell wall which contributes to plant growth, lodging resistance and the responses to a variety of biotic and abiotic stresses (Liu *et al.*, 2018).

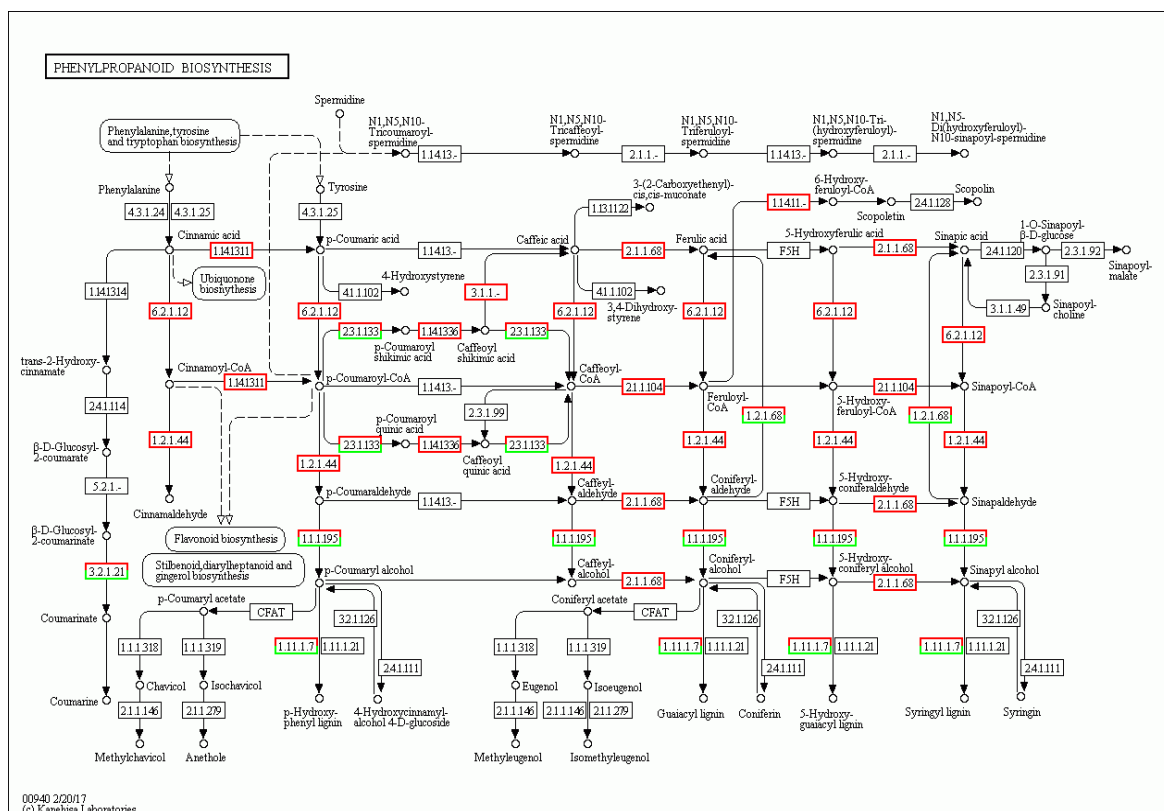


Figure 4.14: Phenylpropanoid Biosynthesis Pathway Map by KEGG

Description: The DEGs of the non-trunking sago palm shows a flow of reactions from cinnamic acid to the end product lignins.

Keys: Red outline indicates upregulated DEGs were involved while green plus red outline indicates a mixture of upregulated and downregulated DEGs were involved.

Source: (Kanehisa & Goto, 2000)

Table 4.17: DEGs Involved in Phenylpropanoid Biosynthesis Pathway

Description: The genes involved in phenylpropanoid biosynthesis shows a mixed (up or down) regulation.

KEGG Annotation	DEGs	Log ₂ (Foldchange)
<i>trans-cinnamate 4-monooxygenase</i> [EC:1.14.14.91]	CL1983.Contig3_All	1.72
	CL1983.Contig4_All	1.61
<i>feruloyl-CoA 6-hydroxylase</i> [EC:1.14.11.61]	CL1696.Contig10_All	2.47
	CL1696.Contig12_All	1.92

Table 4.17 continued

	CL1696.Contig13_All	2.92
	CL1696.Contig16_All	1.88
	CL1696.Contig6_All	2.27
	CL1696.Contig7_All	1.67
	CL1696.Contig8_All	2.62
	Unigene29382_All	1.40
<i>caffeic acid 3-O-methyltransferase / acetylserotonin O-methyltransferase</i> [EC:2.1.1.68/2.1.1.4]	CL5807.Contig1_All	2.47
<i>caffeoylshikimate esterase</i> [EC:3.1.1.-]	CL9550.Contig2_All	2.24
<i>4-coumarate--CoA ligase</i> [EC:6.2.1.12]	Unigene16086_All	3.23
<i>shikimate O-hydroxycinnamoyl- transferase</i> [EC:2.3.1.133]	CL260.Contig4_All	1.87
	CL260.Contig5_All	2.19
	Unigene12362_All	3.28
	Unigene12362_All	-1.27
	CL3632.Contig3_All	-1.16
<i>5-O-(4-coumaroyl)-D-quinic 3'- monooxygenase</i> [EC:1.14.14.96]	CL4589.Contig2_All	2.59
	Unigene20574_All	3.33
<i>caffeoyl-CoA O-methyltransferase</i> [EC:2.1.1.104]	CL10313.Contig1_All	2.01
<i>coniferyl-aldehyde dehydrogenase</i> [EC:1.2.1.68]	CL404.Contig3_All	-1.72
	Unigene9586_All	1.41
<i>cinnamoyl-CoA reductase</i> [EC:1.2.1.44]	CL8093.Contig1_All	3.21
	CL8093.Contig2_All	3.60

Table 4.17 continued

	Unigene19802_All	3.46
	Unigene29636_All	3.55
<i>cinnamyl-alcohol dehydrogenase</i> [EC:1.1.1.195]	CL697.Contig2_All	1.85
	CL697.Contig3_All	1.72
	CL832.Contig1_All	2.23
	CL832.Contig2_All	2.90
	CL832.Contig3_All	2.96
	CL832.Contig4_All	2.77
	CL832.Contig6_All	2.49
	CL9634.Contig1_All	2.55
	Unigene5628_All	3.75
	Unigene17952_All	-3.61
	Unigene984_All	-1.17
<i>beta-glucosidase</i> [EC:3.2.1.21]	CL3116.Contig3_All	2.07
	CL3126.Contig2_All	2.75
	CL3768.Contig11_All	-2.25
<i>peroxidase</i> [EC:1.11.1.7]	CL1024.Contig4_All	3.85
	CL2133.Contig2_All	3.37
	CL2730.Contig1_All	2.08
	CL2730.Contig2_All	1.83
	CL2730.Contig3_All	2.09
	CL5822.Contig2_All	2.46
	CL6054.Contig7_All	1.32

Table 4.17 continued

	CL8443.Contig1_All	1.15
	CL8443.Contig2_All	1.32
	CL929.Contig3_All	1.87
	CL929.Contig4_All	3.10
	CL9602.Contig1_All	2.91
	Unigene13084_All	4.25
	Unigene23470_All	2.60
	Unigene39385_All	2.04
	Unigene46223_All	3.79
	Unigene46437_All	3.68
	CL1367.Contig2_All	-1.19
	CL3856.Contig1_All	-3.70
	CL5626.Contig2_All	-1.93
	CL8814.Contig2_All	-1.63
	Unigene24562_All	-4.44
	Unigene25869_All	-2.04
	Unigene34087_All	-3.83

DEGs from the benzoxazinoid biosynthesis pathway were also involved in the phenylpropanoid biosynthesis pathway (**Table 4.18 & Figure 4.15**). Upregulation of genes involved in the benzoxazinoid biosynthesis pathway in non-trunking sago palms indicated that the non-trunking sago palm has a higher demand or generated higher concentration of benzoxazinoid metabolites. The roles of benzoxazinoid in plants are diverse as an example

it involved in pathogenic defence, herbivory, allelopathy, plant-microbe interaction, detoxification and as a signalling molecules (Niculaes *et al.*, 2018).

Table 4.18: DEGs involved in Benzoxazinoid Biosynthesis Pathway

Description: All the genes involved in benzoxazinoid biosynthesis are upregulated.

KEGG Annotation	DEGs	Log ₂ (Foldchange)
<i>2,4-dihydroxy-1,4-benzoxazin-3-one-glucoside dioxygenase</i> [EC:1.14.11.59]	CL1696.Contig10_All	2.47
	CL1696.Contig12_All	1.92
	CL1696.Contig13_All	2.92
	CL1696.Contig16_All	1.88
	CL1696.Contig6_All	2.27
	CL1696.Contig7_All	1.67
	CL1696.Contig8_All	2.62

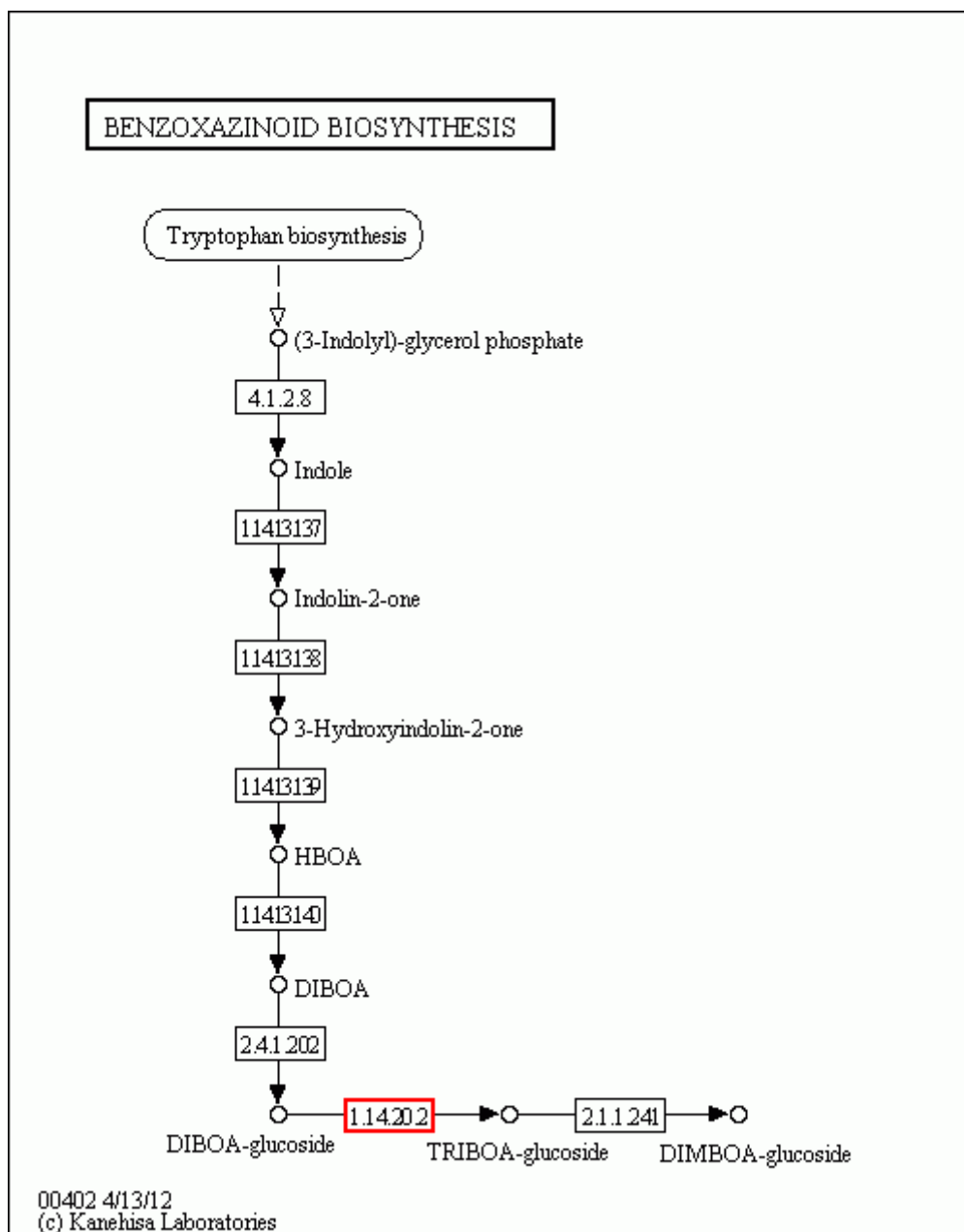


Figure 4.15: Benzoxazinoid Biosynthesis Pathway Map by KEGG

Description: Only one reaction is upregulated in benzoxazinoid biosynthesis pathway where the DIBOA-glucoside form the TRIBOA-glucoside.

Keys: Red outline indicates upregulated DEGs were involved.

Source: (Kanehisa & Goto, 2000)

The flows of enzymatic reaction involved by the DEGs from the stilbenoid, diarylheptanoid and gingerol biosynthesis pathway from cinnamoyl-coA to ferulyl-coA and

pterostilbene suggest that the part of the DEGs from this pathway are also involved in phenylpropanoid biosynthesis pathway (Table 4.19 & Figure 4.16). Pterostilbene is a phytoalexin compound which is both an antioxidant and toxic to bacteria and fungi (Duke, 2022). Pathogenic infection in plants stimulates pterostilbene production (Duke, 2022). Upregulation of genes related to pterostilbene in non-trunking sago palms suggested that the plants were under pathogenic infection.

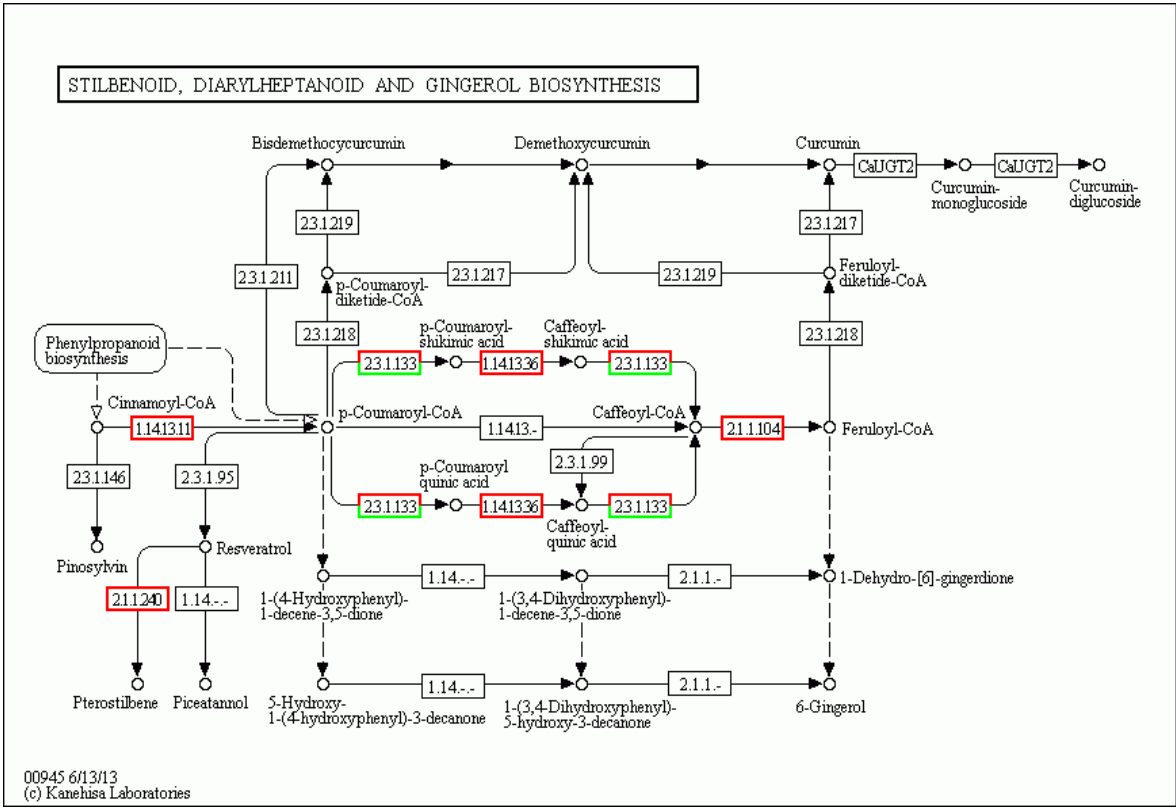


Figure 4.16: Stilbenoid, Diarylheptanoid and Gingerol Biosynthesis Pathway Map by KEGG

Description: The DEGs of the non-trunking sago palm shows a flow of reactions from cinnamoyl-CoA to the end product Feruloyl-CoA which is a sub-section of the phenylpropanoid biosynthesis pathway.

Keys: Red outline indicates upregulated DEGs were involved while green plus red outline indicates a mixture of upregulated and downregulated DEGs were involved.

Source: (Kanehisa & Goto, 2000)

Table 4.19: DEGs involved in Stilbenoid, Diarylheptanoid and Gingerol Biosynthesis Pathway

Description: Only *shikimate O-hydroxycinnamoyl-transferase* genes were mixed regulated while all the other genes related to the stilbenoid, diarylheptanoid and gingerol biosynthesis pathway are upregulated.

KEGG Annotation	DEGs	Log ₂ (Foldchange)
<i>trans-cinnamate 4-monooxygenase</i> [EC:1.14.14.91]	CL1983.Contig3_All	1.72
	CL1983.Contig4_All	1.61
<i>shikimate O-hydroxycinnamoyl-transferase</i> [EC:2.3.1.133]	CL260.Contig4_All	1.87
	CL260.Contig5_All	2.19
	Unigene12362_All	3.28
	Unigene12362_All	-1.27
	CL3632.Contig3_All	-1.16
<i>5-O-(4-coumaroyl)-D-quinic 3'-monooxygenase</i> [EC:1.14.14.96]	CL4589.Contig2_All	2.59
	Unigene20574_All	3.33
<i>caffeoyl-CoA O-methyl-transferase</i> [EC:2.1.1.104]	CL10313.Contig1_All	2.01
<i>trans-resveratrol di-O-methyl-transferase</i> [EC:2.1.1.240]	Unigene3722_All	1.86

The flows of the DEGs' enzymatic reaction involved in flavonoid production suggested that flavonoid biosynthesis in non-trunking sago palm leaf tissue is upregulated (**Figure 4.17 & Table 4.20**). Flavonoid roles in a plant are diverse: from UV-B protection; colouring compound to attracting pollinators; pathogens and herbivory defence; root nodulation; gene expression; cell wall synthesis; allelopathic action against other plants; polar transport of auxins inhibition for plant growth; antioxidant; and mineral deficiency and

toxicity regulation (Buer *et al.*, 2008; Cesco *et al.*, 2010; Falcone Ferreyra *et al.*, 2012; Liu & Murray, 2016; Liu *et al.*, 2018; Nix *et al.*, 2017; Petrusa *et al.*, 2013). Upregulation of flavonoid biosynthesis pathway in non-trunking sago palms hinted that the plants were under biotic or/and abiotic stresses.

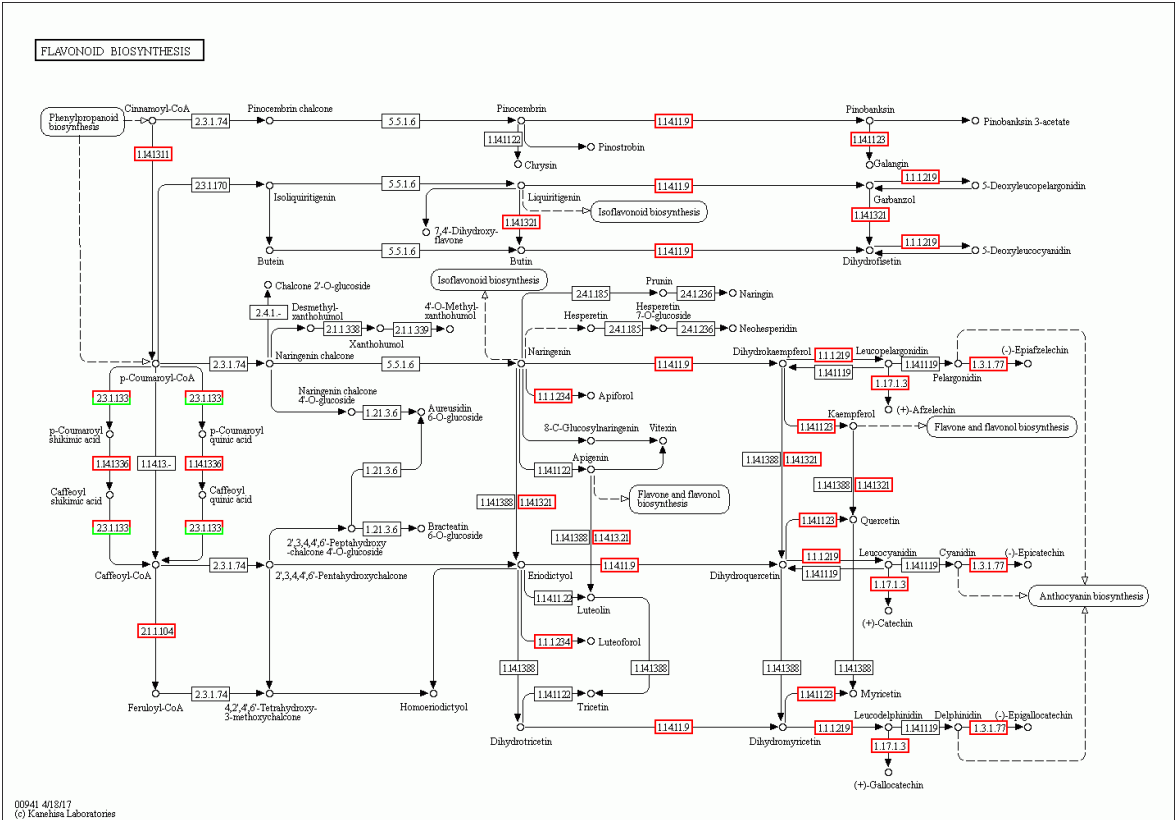


Figure 4.17: Flavonoid Biosynthesis Pathway Map by KEGG

Description: The DEGs of the non-trunking sago palm shows a flow of reactions to form flavonoid compounds.

Keys: Red outline indicates upregulated DEGs were involved while green plus red outline indicates a mixture of upregulated and downregulated DEGs were involved

Source: (Kanehisa & Goto, 2000)

Table 4.20: DEGs involved in Flavonoid Biosynthesis Pathway

Description: Only *shikimate O-hydroxycinnamoyl-transferase* genes were mixed regulated while all the other genes related to the flavonoid biosynthesis pathway are upregulated.

KEGG Annotation	DEGs	Log ₂ (Foldchange)
<i>trans-cinnamate 4-monooxygenase</i> [EC:1.14.14.91]	CL1983.Contig3_All	1.72
	CL1983.Contig4_All	1.61
<i>flavonoid 3'-monooxygenase</i> [EC:1.14.14.82]	CL8139.Contig2_All	2.32
<i>bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase</i> [EC:1.1.1.219/1.1.1.234]	CL3983.Contig2_All	2.35
<i>naringenin 3-dioxygenase</i> [EC:1.14.11.9]	Unigene29382_All	1.40
<i>flavonol synthase</i> [EC:1.14.20.6]	Unigene29382_All	1.40
<i>leucoanthocyanidin reductase</i> [EC:1.17.1.3]	CL4324.Contig2_All	2.00
<i>anthocyanidin reductase</i> [EC:1.3.1.77]	CL984.Contig4_All	2.27
<i>shikimate O-hydroxycinnamoyl-transferase</i> [EC:2.3.1.133]	CL260.Contig4_All	1.87
	CL260.Contig5_All	2.19
	Unigene12362_All	3.28
	Unigene12362_All	-1.27
	CL3632.Contig3_All	-1.16
<i>5-O-(4-coumaroyl)-D-quinic 3'-monooxygenase</i> [EC:1.14.14.96]	CL4589.Contig2_All	2.59
	Unigene20574_All	3.33
<i>caffeoyl-CoA O-methyltransferase</i> [EC:2.1.1.104]	CL10313.Contig1_All	2.01

ELF3, *PRR5*, *PRR7*, *PRR9*, *CHE*, *GI*, *FKF1*, and *CO* were downregulated while *LHY* and *HY5* were upregulated. The flow of protein reaction involved suggesting that DEGs involved in photomorphogenesis. Photomorphogenesis in plant governs the seed germination and de-etiolation such as; inhibition of stem and petiole elongation, chloroplast development, chlorophyll biosynthesis, leaf expansion, branching, flowering and senescence (Tripathi *et al.*, 2019). Upregulation of *HY5* in non-trunking sago palm suggested that the plants are inhibiting the elongation of the stem and petiole as observed in the non-trunking sago palms phenotype (Table 2.5).

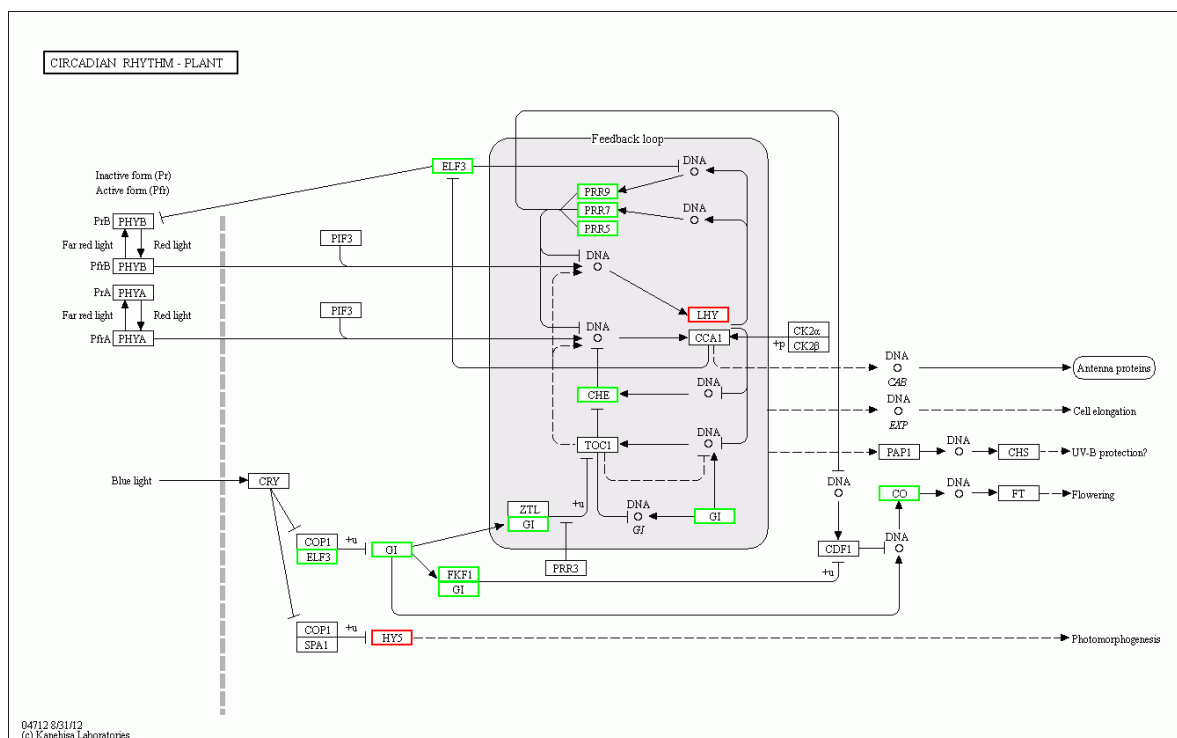


Figure 4.18: Circadian Rhythm Pathway Map by KEGG

Description: The DEGs of the non-trunking sago palm shows a down-regulation of the regulating genes except for *HY5* and *LHY* which are upregulated.

Keys: Red outline indicates upregulated DEGs were involved while green outline indicates downregulated DEGs were involved

Source: (Kanehisa & Goto, 2000)

Table 4.21: DEGs involved in Circadian Rhythm Pathway

Description: The DEGs of the non-trunking sago palm shows a down-regulation of circadian rhythm pathway genes except for *HY5* and *LHY* which are upregulated.

KEGG Annotation	DEGs	Log ₂ (Foldchange)
<i>protein EARLY FLOWERING 3 [ELF3]</i>	CL1630.Contig4_All	-1.43
<i>GIGANTEA [GI]</i>	Unigene20400_All	-2.07
	Unigene28475_All	-2.15
	Unigene28476_All	-1.85
	Unigene28480_All	-1.99
	Unigene55076_All	-2.04
<i>flavin-binding kelch repeat F-box protein 1 [FKF1]</i>	Unigene879_All	-3.75
<i>transcription factor HY5 [HY5]</i>	CL4395.Contig3_All	1.47
<i>transcription factor TCP21 (protein CCA1 HIKING EXPEDITION) [CHE]</i>	CL4267.Contig1_All	-1.66
<i>pseudo-response regulator 9 [PRR9]</i>	CL3683.Contig1_All	-4.43
	CL3683.Contig2_All	-3.15
<i>pseudo-response regulator 7 [PRR7]</i>	CL3683.Contig1_All	-4.43
	CL3683.Contig2_All	-3.15
<i>pseudo-response regulator 5 [PRR5]</i>	CL3683.Contig2_All	-3.15
	CL6620.Contig1_All	-2.99
<i>MYB-related transcription factor LHY [LHY]</i>	CL1173.Contig1_All	1.99
	CL1173.Contig10_A1 1	1.71

Table 4.21 continued

<i>zinc finger protein CONSTANS [CO]</i>	CL3683.Contig1_All	-4.43
	CL3683.Contig2_All	-3.15
	CL6840.Contig1_All	-2.74
	CL7230.Contig1_All	-3.42
	CL7230.Contig2_All	-3.26
	Unigene16572_All	-3.16

The observation of DEGs based on KEGG analysis pathway suggesting that the non-trunking sago palms were: suppressing its growth and development, shown by the upregulation of photomorphogenesis; and increased its metabolisms to regulates stresses as shown by the upregulation of lignin monomers, benzoxazinoid, pterostilbene and flavonoid biosynthesis. This suggestion in-line with the GO enrichment analysis and further clarify the cause and effect of the non-trunking phenomenon.

4.3.8 DEGs Protein-Protein Interaction (PPI) Analysis by STRINGdb

DEGs were analysed using NCBI non-redundant proteins database annotation by the stringdb.org web interface. The setting for the analysis are STRINGdb ver. 11.5, three groups of input of DEGs with NCBI protein annotation (Total DEGs, Upregulated DEGs, Downregulated DEGs) and compared with *Phoenix dactylifera* reference genome (highest annotation similarity for *Elaeis guineensis* is absent from the reference genome list). The analysis does not include predicted functional partners and with interaction score equal to or more than 400 (medium confidence) cut-off. The three groups of inputs were used to

identifies the pathways involved by the interactions of the overall DEGs, interactions of the upregulated DEGs and the interactions of the downregulated DEGs.

A total of 801 total DEGs annotated by NCBI non-redundant proteins database were input into STRINGdb with *Phoenix dactylifera* reference genome. The PPI analysis of total DEGs interaction illustrated that the DEGs involved in stilbenoid, diarylheptanoid and gingerol biosynthesis, phenylpropanoid biosynthesis, biosynthesis of secondary metabolites and metabolic pathways based on functional enrichments by KEGG pathway (**Table 4.22**, **Table 4.23**, & **Figure 4.19**). The pathways involved by the total DEGs based on PPI analysis are similar to those reported by the KEGG analysis previously.

Table 4.22: Total DEGs Protein-Protein Interaction (PPI)

Description: The STRINGdb output shows that the nodes are not random and that the observed number of edges is significant as the p-value of the enrichment are lower than 0.05.

STRINGdb output	Values
Number of nodes	162
Number of edges	74
Average node degree	0.914
Avg. local clustering coefficient	0.261
Expected number of edges	43
PPI enrichment p-value	$1.21e^{-05}$

Keys: Number of nodes = Number of Protein, Number of edges = Number of interactions, Average node degree = average interaction per node, Avg. local clustering coefficient = connection strength of all the DEGs involved, expected number of edges = expected number of edges if were selected in random, PPI enrichment p-value = Probability of the PPI enrichment to occur at random (the smaller the value the more significant the PPI enrichment)

Table 4.23: Pathway involved by the Total DEGs

Description: List of significant pathways involved by the total DEGs in non-trunking sago palm.

Pathway	Description	Count in Network	Strength	False discovery rate
pda00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3 of 32	1.18	0.0492
pda00940	Phenylpropanoid biosynthesis	8 of 146	0.95	0.00067
pda01110	Biosynthesis of secondary metabolites	18 of 1243	0.37	0.0492
pda01100	Metabolic pathways	27 of 2271	0.29	0.0492

Keys: Pathway = Pathway entry code in KEGG, Description = Name of the pathway involved, Count in Network = number of DEGs involved in the pathway over Total annotated DEGs, Strength = significance of the pathway, False discovery rate = chances of the pathway selected at random.

The PPI analysis of upregulated DEGs interaction illustrated that the DEGs involved in: stilbenoid, diarylheptanoid and gingerol biosynthesis; flavonoid biosynthesis; phenylpropanoid biosynthesis; biosynthesis of secondary metabolites; and metabolic pathways based on functional enrichments by KEGG pathway (**Table 4.24**, **Table 4.25**, & **Figure 4.20**). The pathways involved by the upregulated DEGs based on PPI analysis are similar to those reported by the KEGG analysis previously.

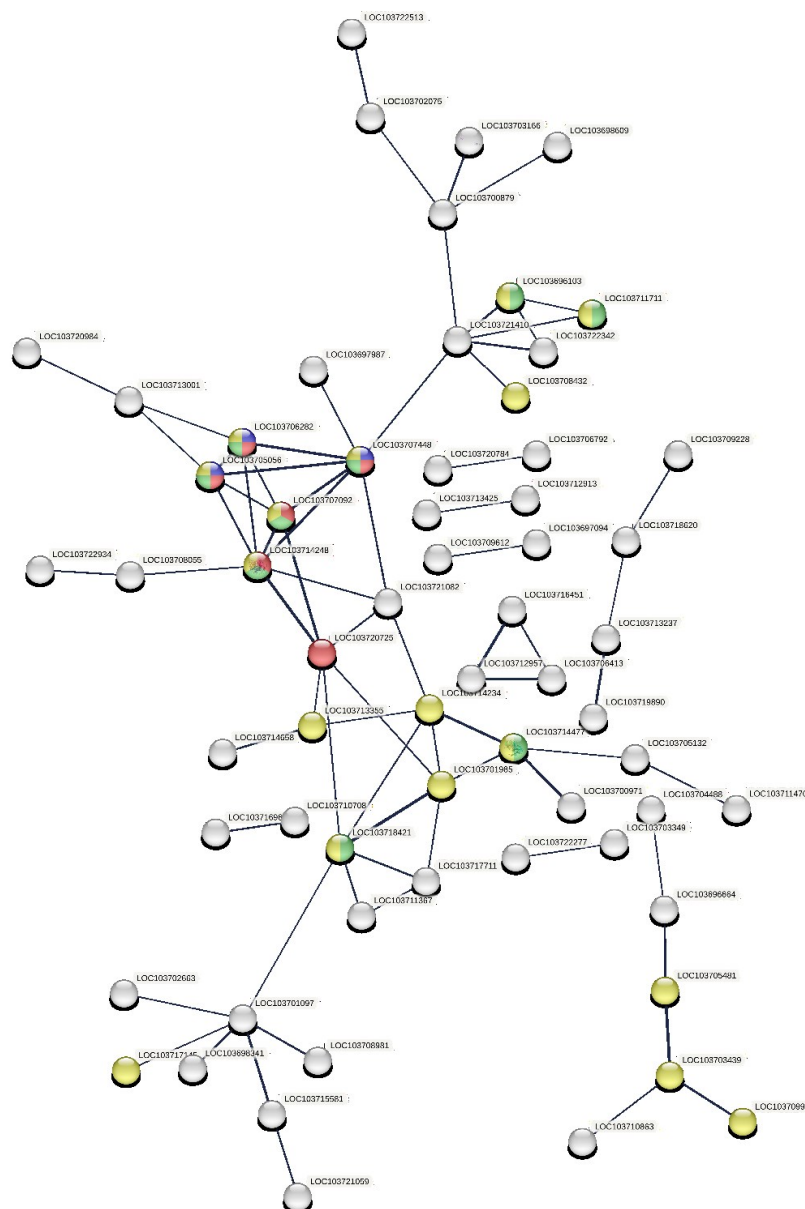


Figure 4.19: Protein-Protein Interaction Network of Total DEGs excluding DEGs without any Interactions

Description: Interactions of the DEGs proteins and the significant pathway involved.

Keys: Red colour node indicate DEGs involved in phenylpropanoid biosynthesis; Blue colour node indicate DEGs involved in stilbenoid, diarylheptanoid and gingerol biosynthesis; Green colour node indicate DEGs involved in biosynthesis of secondary metabolites; and Yellow colour node indicate DEGs involved in metabolic pathways. Line attached between the nodes indicate interaction between the DEGs. Thickness of the linking line indicate significance of the interaction, the thicker the line the more significant the interaction. Node with strings inside indicate available information on it three dimensional (3D) structure.

Source: (Szklarczyk *et al.*, 2019)

Table 4.24: Upregulated DEGs Protein-Protein Interaction (PPI)

Description: The STRINGdb output shows that the nodes are not random and that the observed number of edges is significant as the p-value of the enrichment are lower than 0.05.

STRINGdb output	Values
Number of nodes:	94
Number of edges:	26
Average node degree:	0.553
Avg. local clustering coefficient:	0.117
Expected number of edges:	14
PPI enrichment p-value:	0.00219

Keys: Number of nodes = Number of Protein, Number of edges = Number of interactions, Average node degree = average interaction per node, Avg. local clustering coefficient = connection strength of all the DEGs involved, expected number of edges = expected number of edges if were selected in random, PPI enrichment p-value = Probability of the PPI enrichment to occur at random (the smaller the value the more significant the PPI enrichment)

Table 4.25: Pathway involved by the Upregulated DEGs

Description: List of significant pathways involved by the total DEGs in non-trunking sago palm.

Pathway	Description	Count in Network	Strength	False discovery rate
pda00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3 of 32	1.42	0.0091
pda00941	Flavonoid biosynthesis	3 of 54	1.19	0.0296
pda00940	Phenylpropanoid biosynthesis	7 of 146	1.13	0.00017
pda01110	Biosynthesis of secondary metabolites	14 of 1243	0.5	0.0091
pda01100	Metabolic pathways	20 of 2271	0.39	0.0091

Keys: Pathway = Pathway entry code in KEGG, Description = Name of the pathway involved, Count in Network = number of DEGs involved in the pathway over Total annotated DEGs, Strength = significance of the pathway, False discovery rate = chances of the pathway selected at random.

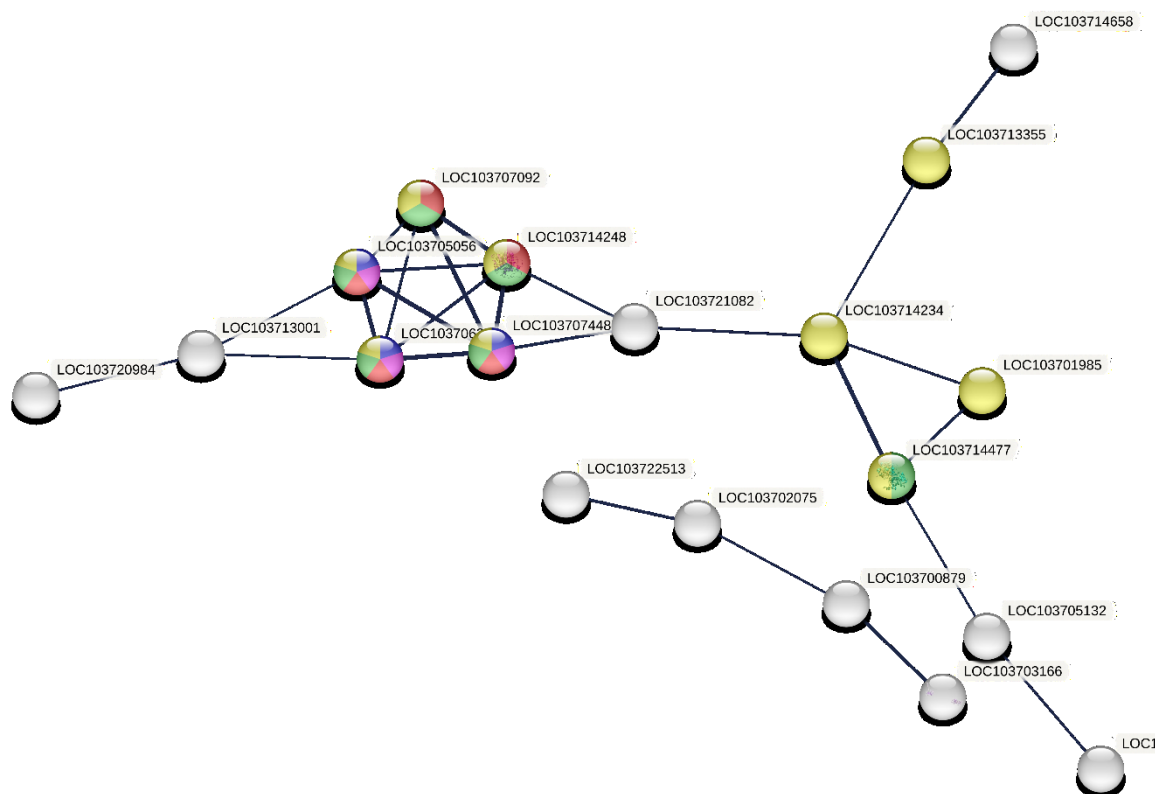


Figure 4.20: Protein-Protein Interaction Network of Upregulated DEGs excluding DEGs without any Interactions

Description: Interactions of the DEGs proteins and the significant pathway involved.

Keys: Red colour node indicate DEGs involved in phenylpropanoid biosynthesis; Blue colour node indicate DEGs involved in stilbenoid, diarylheptanoid and gingerol biosynthesis; Green colour node indicate DEGs involved in biosynthesis of secondary metabolites; and Yellow colour node indicate DEGs involved in metabolic pathways. Line attached between the nodes indicate interaction between the DEGs. Thickness of the linking line indicate significance of the interaction, the thicker the line the more significant the interaction.

Source: (Szklarczyk *et al.*, 2019)

The PPI analysis of downregulated DEGs interaction do not show any significant functional enrichment (**Table 4.23**).

Table 4.26: Downregulated DEGs Protein-Protein Interaction (PPI)

Descripton: The STRINGdb output shows that the nodes are not random and that the observed number of edges is significant as the p-value of the enrichment are lower than 0.05.

STRINGdb output	Values
number of nodes:	71
number of edges:	18
average node degree:	0.507
avg. local clustering coefficient:	0.286
expected number of edges:	8
PPI enrichment p-value:	0.00257

Keys: Number of nodes = Number of Protein, Number of edges = Number of interactions, Average node degree = average interaction per node, Avg. local clustering coefficient = connection strength of all the DEGs involved, expected number of edges = expected number of edges if were selected in random, PPI enrichment p-value = Probability of the PPI enrichment to occur at random (the smaller the value the more significant the PPI enrichment).

Analysis of the PPI enrichment data obtained from total DEGs and upregulated DEGs illustrated the similar pathways involved in non-trunking sago palm with KEGG functional enrichment analysis previously. The upregulation of the stilbenoid, diarylheptanoid and gingerol biosynthesis, flavonoid biosynthesis, phenylpropanoid biosynthesis, biosynthesis of secondary metabolites and metabolic pathways agreed with the KEGG functional enrichment analysis previously which concluded that the non-trunking sago palms were under stresses and as shown by the upregulation of: stilbenoid, diarylheptanoid and gingerol; flavonoid; phenylpropanoid; and other metabolites.

Wrapping up the analysis of DEGs by functional analysis by GO, pathway analysis by KEGG and the PPI by STRINGdb, the cause of non-trunking in sago palm are due to

phosphate, sulphate and nitrate starvation in addition to aluminium and boron toxicity resulting in plant growth suppression and weaken its pathogenic resistance. The pathogen infected non-trunking sago palm increases its metabolism of the building block for reinforcing cell wall and increase flavonoid biosynthesis for pathogenic suppression and wound healing. Mineral starvation causes: the mineral scavenging or/and translocation to young leaf; lateral roots propagation and nodulation for nitrogen-fixation rhizobia; arbuscular mycorrhizae and other symbiotic microorganisms' symbiosis; in order to increase the availability of the required mineral. Metal toxicity were addressed by the flavonoids (eg. polyalcohols, phenolics, citrate) extrusion from root to act as metal (boron and aluminium) chelators to lessen/nullify the metal uptake or/and translocation of the metal to old leaves (Cesco *et al.*, 2010; Falcone Ferreyra *et al.*, 2012; Kayihan, 2021; Yoshinari & Takano, 2017).

4.3.9 RNA-seq Data Validation by qPCR Analysis

Five genes of interest and one housekeeping gene were selected based on the mentioned criteria: DEGs without isoforms, DEGs annotated by KEGG/GO, annotated plant gene, DEGs $\log_2$ foldChange between trunking and non-trunking sago palm from 1 – 5 (**Chapter 3.5.1, & Table 4.27**). Primers of the selected genes were designed and sequenced (**Table 4.28**). Unigene4400_ALL (proteasome subunit beta type-4) was selected as the reference gene. Relative gene expressions were calculated based on the Ct-value obtained in qPCR results (**Table 4.29**). Pearson's correlation analysis was performed with the data obtained in **Table 4.29 (Table 4.30)**.

Table 4.27: Selected Genes for qPCR Analysis

Description: Information of the targeted genes and reference gene used to validate the RNA-seq result.

Gene	Length	T-Expression	N-Expression	log2FoldChange (N/T)	KEGG Annotation
Unigene7182_All NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	1995	3891.49	9777.51	1.33	gra:105767541 <i>Gossypium raimondii</i>
Unigene10336_All cytochrome P450 711A1	1810	4.63	132.58	4.84	sot:102593258 <i>Solanum tuberosum</i>
Unigene27309_All peroxidase 46	748	5.64	223.22	5.31	bdi:100822183 <i>Brachypodium distachyon</i>
Unigene7794_All RING-H2 finger protein ATL80	793	577.58	22.97	-4.65	zma:100284205 <i>Zea mays</i>
Unigene937_All receptor homology region, transmembrane domain- and RING domain-containing protein 1	1044	1865.84	826.14	-1.18	jcu:105629434 <i>Jatropha curcas</i>
Unigene4400_All* proteasome subunit beta type-4	1194	920.08	917.93	0.00	egu:105035389 <i>Elaeis guineensis</i>

Keys: *Asterisk indicates reference gene.

Table 4.28: Primer Design of Selected Genes

Desription: List of primers sequences and informations for the targeted genes and reference gene designed.

Gene ID	Primer	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self Com.	Self3' Com.
Unigene7182 _All	Forward	CTCCGGTACTCCTGGCAAAA	Plus	20	721	740	59.68	55.00	4.00	1.00
	Reverse	GCGGTCTCCAAAATTGCTCC	Minus	20	1301	1282	59.83	55.00	4.00	0.00
	Product length	581								
Unigene10336 _All	Forward	ATTGGCACAGAACCCAGAGG	Plus	20	1073	1092	59.96	55.00	3.00	1.00
	Reverse	TCTTACCGTCGTGGAAGTGC	Minus	20	1599	1580	60.04	55.00	3.00	2.00
	Product length	527								
Unigene27309 _All	Forward	TGCTGTCTGCGAATCCCTTA	Plus	20	239	258	59.10	50.00	3.00	2.00
	Reverse	CACGCCAATCGCTGTCTCT	Minus	19	439	421	60.45	57.89	3.00	0.00

Table 4.28 continued

	Product length	201								
Unigene7794 _All	Forward	TTTTAAGGCAAGAACCCGCC	Plus	20	163	182	59.04	50.00	4.00	2.00
	Reverse	ACTCCGACGTGGTGGTAATC	Minus	20	646	627	59.47	55.00	4.00	1.00
	Product length	484								
Unigene937 _All	Forward	CCCCCTGATGAGCAAGCTAC	Plus	20	419	438	60.18	60.00	4.00	2.00
	Reverse	AGGCTTTGCGCCTTCAAAAG	Minus	20	686	667	59.97	50.00	4.00	3.00
	Product length	268								
Unigene4400 _All	Forward	GCCCATTTTTTCACTCCACCG	Plus	20	647	666	59.76	55.00	2.00	2.00
	Reverse	GTACAAGGGCATGGAGCGTA	Minus	20	897	878	59.82	55.00	4.00	2.00
	Product length	251								

Table 4.29: Relative Gene Expressions of Selected Genes

Description: The RNA-seq gene expression data were more correlated with the actual 6 samples qPCR gene expression data compared with the 10 samples data as seen in the lower standard deviation

Genes	RNA-seq	$-\Delta\Delta Ct$ 1-10	$-\Delta\Delta Ct$ 1,4,5,7,8,9	Std - $\Delta\Delta Ct$ 1-10	Std $-\Delta\Delta Ct$ 1,4,5,7,8,9
Unigene7182 _All	1.3291	2.4836	2.1599	0.8163	0.5874
Unigene10336 _All	4.8401	1.3781	4.1254	2.4480	0.5054
Unigene27309 _All	5.3063	3.2266	5.6821	1.4706	0.2657
Unigene7794 _All	-4.6520	-5.5754	-6.7475	0.6529	1.4817
Unigene937 _All	-1.1754	-1.4732	-0.9441	0.2106	0.1635
Average standard deviation				1.1197	0.6007

Keys: $-\Delta\Delta Ct$ 1-10: Negative Delta Delta Cyclic Threshold of ten samples; $-\Delta\Delta Ct$ 1,4,5,7,8,9: Negative Delta Delta Cyclic Threshold of sample 1,4,5,7,8,9. Std $-\Delta\Delta Ct$ 1-10: Standard deviation of $-\Delta\Delta Ct$ 1-10 with RNA-seq log2foldchange value; Std $-\Delta\Delta Ct$ 1,4,5,7,8,9: Standard deviation of $-\Delta\Delta Ct$ 1,4,5,7,8,9 with RNA-seq log2foldchange value.

The relative gene expression value of qPCR analysis shows similar pattern of gene expression with RNA-seq analysis of trunking and non-trunking sago palm (**Table 4.29**). The relative gene expression standard deviation of the identical samples set to RNA-seq (Std $-\Delta\Delta Ct$ 1,4,5,7,8,9) is lower compared to the larger samples set (Std $-\Delta\Delta Ct$ 1-10) as expected aligned with the increment of the number of samples from six to ten. Pearson's Correlation Analysis shows that the RNA-seq gene expression data are strongly positively correlated

with the actual six samples qPCR gene expression data and the ten samples qPCR gene expression data (**Table 4.30**). The Pearson's correlation (r) between the RNA-seq data and the $-\Delta\Delta Ct$ 1,4,5,7,8,9 qPCR data is $r = 0.980$, $p\text{-value} > 0.99$ while the RNA-seq data and the $-\Delta\Delta Ct$ 1-10 qPCR data is $r = 0.909$, $p\text{-value} > 0.95$. These results of the qPCR analysis agreed with the gene expression data from the RNA-seq and also shows that the sample gene expression estimation by RNA-seq able to represent the population gene expression.

The variation of gene expression value between RNA-seq and qPCR analysis are due to the difference in count normalisation, difference in PCR efficiency between selected genes, technical error etc. Even though the qPCR analysis of relative gene expressions is not equal to that of $\log_2$ (Fold-Change) in RNA-seq analysis, it does shows similar pattern in which identified the upregulated and downregulated genes in non-trunking sago palm either in the six samples set or in the ten samples set qPCR analysis and thus supporting the finding of RNA-seq analysis.

Table 4.30: Pearson's Correlation Analysis between RNA-seq Data and qPCR Data

Description: The RNA-seq gene expression data are strongly positively correlated with the actual six samples qPCR gene expression data (0.980 , $p\text{-value} = 0.99$) compared with the ten samples data (0.909 , $p\text{-value} = 0.95$).

Correlations				
		RNAseq	$-\Delta\Delta Ct$ 1,4,5,7,8,9	$-\Delta\Delta Ct$ 1-10
RNAseq	Pearson Correlation	1	.980**	.909*
	Sig. (2-tailed)		.003	.032
	N	5	5	5

Table 4.30 continued

- $\Delta\Delta\text{Ct}$ 1,4,5,7,8,9	Pearson Correlation	.980**	1	.966**
	Sig. (2-tailed)	.003		.007
	N	5	5	5
- $\Delta\Delta\text{Ct}$ 1-10	Pearson Correlation	.909*	.966**	1
	Sig. (2-tailed)	.032	.007	
	N	5	5	5
**. Correlation is significant at the 0.01 level (2-tailed).				
*. Correlation is significant at the 0.05 level (2-tailed).				

Keys: RNAseq: RNA-seq gene expression data by DESeq2, - $\Delta\Delta\text{Ct}$ 1-10: qPCR gene expression data of the ten sago palm samples (including the six sago palm samples used in generating the RNA-seq data); - $\Delta\Delta\text{Ct}$ 1,4,5,7,8,9: qPCR gene expression data of the six sago palm samples used in generating the RNA-seq data

4.4 Nutrient Deficiency and Toxicity in Sago Palm

Since the outcome of the DEGs analysis pointed out that the cause of the non-trunking phenomenon in sago palm is due to nutrient deficiency and toxicity, a more thorough discussion concerning the DEGs was included.

Nitrogen is one of the macronutrients in plants and its uptake majorly are in the form of inorganic nitrate in alkaline soil while ammonia in acidic soil but in some cases the plant would uptake organic nitrate in the form of amino acids (Masclaux-Daubresse *et al.*, 2010). Nitrogen deficient coconut palms causes uniformly V-shaped dull green to yellowish leaves, starting from oldest to youngest (Kendra, 2013). Nitrogen-deficient solitaire palms cause a great reduction in growth, canopy and leaf size with tapered trunks (Broschat, 2020).

The uptake of nitrogen is facilitated by nitrate transporter (NRT), ammonium transporter (AMT), amino acid transporter and amino acid permease (Masclaux-Daubresse *et al.*, 2010). After the intake of nitrogen into the plant, nitrate and ammonia need to be converted to ammonium before assimilation into amino acids. Ammonia were transported into the vacuole and exported out as ammonium (Masclaux-Daubresse *et al.*, 2010). Nitrate were initially reduced to nitrite in cytoplasm by nitrate reductase (NR) before further reduce to ammonium by nitrite reductase (NiR) in plastids (Masclaux-Daubresse *et al.*, 2010). Direct incorporation of ammonium into glutamate form glutamine facilitated by glutamine synthetase (GS) while incorporation of ammonium into carbamoylphosphate facilitated by carbamoylphosphate synthase (CPSase) act as precursor for the synthesis of citrulline and arginine in plastids (Masclaux-Daubresse *et al.*, 2010). Other than nitrogen assimilation utilising external nitrogen source, nitrogen recycling and translocation from senescence leaves to growing parts or storage organ such as seeds and young leaves are among the plant strategy in optimal utilisation of nitrogen source (Masclaux-Daubresse *et al.*, 2010).

Interestingly, non-trunking sago palms leaf tissue DEGs only shows the upregulation of nitrogen transporters and none on the proteins associated with nitrogen assimilation (**Table 4.31**). This suggested that the gene regulation for nitrogen assimilation are separated from gene regulation of nitrogen starvation respond. Upregulation of nitrogen assimilation proteins maybe triggered by the high concentration of inorganic nitrogen source while upregulation of nitrogen starvation related protein triggered by low concentration of inorganic nitrogen source in sago palms.

Table 4.31: DEGs involved in Nitrogen Regulation

Description: DEGs involves in nirogen regulation in non-trunking sago palm are upregulated.

Nitrogen Assimilation	DEGs	log ₂ Fold-Change (N/T)	Description
<i>protein NRT1/ PTR FAMILY 6.3</i>	CL1039.Contig6_All	1.85	Dual affinity bidirectional nitrate transporter.
	CL1039.Contig8_All	2.07	
	CL1039.Contig2_All	2.28	
	CL1039.Contig19_All	2.52	
	CL1039.Contig12_All	2.66	
	CL1039.Contig14_All	2.82	
<i>Protein NRT1/ PTR FAMILY 3.1</i>	Unigene26483_All	1.78	Chloroplast nitrate influx transporter
	Unigene347_All	2.66	
	Unigene348_All	2.99	
<i>ammonium transporter 3 member 1-like (AMT3;1)</i>	CL181.Contig1_All	1.63	Involved in ammonium transportation. Upregulate in N-deficiency
	CL9493.Contig1_All	2.80	
	CL9493.Contig2_All	3.40	
	CL9493.Contig3_All	3.87	

Phosphorus is one of the macronutrients in plant and its uptake is in the form of inorganic phosphate (Pi). Phosphorus is involved in: protein regulation and activation; photosynthesis; and respiration in plant. Insufficient phosphorus in plant causes: purplish of leaves due to anthocyanin accumulation; stunted growth of oil palms with short fronds; small, narrow, purple or bronze discolouration and necrotic spots on older leaves in coconut

palms; affecting flowering in coconut palms; and in extreme cases, a complete cessation of growth were observed in areca palms (Broschat, 2015; Kendra, 2013; Rankine & Fairhurst, 1999).

The availability of inorganic phosphate in soil are scarce (1 - 10 mM) due to: its very low solubility and poor mobility in soil; and are pH sensitive: where it readily precipitates with aluminium and iron in acidic soil; while precipitate with calcium and magnesium in alkaline soil (Młodzińska & Zboińska, 2016). The uptake and assimilation of Pi in plants are facilitated by inorganic phosphate transporter (PHT) family, phosphate transporter (PHO) family, SPX-MFS family and phosphate translocators family (Srivastava *et al.*, 2018). Low cellular phosphate concentration reduced the SPX domain-containing protein (SPX1) bind-inactivation affinity toward phosphate starvation response 1 (PHR1) transcript factor and thus activating the transcript factor and trigger the Pi starvation responses (PSRs) (Młodzińska & Zboińska, 2016). Other than phosphorus assimilation utilising external Pi source, phosphorus recycling and translocation from senescence leaves to growing parts or storage organ such as seeds are among the plant strategy in optimal utilisation of phosphorus source.

The non-trunking sago palms leaf tissue DEGs consist of: upregulated *PHT2;1* and *SPX1*; and downregulated *PHT1;5* (**Table 4.32**). The phosphorus related DEGs reveals that significant demand of Pi in chloroplast metabolisms under Pi deficient condition as show by the upregulation of *PHT2;1* and *SPX1*. Downregulation of *PHT1;5* prevent the efflux of cellular Pi to phloem thus preventing the relocation of Pi under Pi deficient respond in young leaves.

Table 4.32: DEGs involved in Phosphorus Regulation

Description: DEGs involves in phosphorus regulation in non-trunking sago palm are mixed regulated.

Phosphorus Assimilation	DEGs	log ₂ Fold-Change (N/T)	Description
<i>inorganic phosphate transporter 2;1 (PHT2;1)</i>	Unigene7260_All	1.45	Transport Pi into chloroplast
<i>inorganic phosphate transporter 1;5 (PHT1;5)</i>	CL8793.Contig1_All	-3.43	Transport Pi into phloem
<i>SPX domain-Containing protein 1 (SPX1)</i>	CL3840.Contig1_All	1.64	Regulate the activation of PHR1
	CL3840.Contig2_All	1.59	

Sulphur is one of the secondary macronutrients in plant and its uptake are in the form of inorganic sulphate (SO_4^{2-}) facilitated by plant sulphur transporter (SULTR) in roots or in the form of sulphuric gas (SO_2) or hydrogen sulphide (H_2S) in shoot (Zhao *et al.*, 2008). Sulphur is involved in proteins biosynthesis, defence of plants against stresses and pests. Sulphur deficient young coconut palms exhibits: shorter leaves length with abnormally early splitting of leaflets; chlorosis of leaves; leaflets become grey; necrosed and curled; and stunted coconut palms growth (Shanmugam, 2016).

Sulphate intake into the plant are facilitated by the plant sulphur transporter (SULTR family) which will be fixed into adenosine triphosphate (ATP) to form adenosine 5'-

phosphosulfate (APS) by ATP sulfurylase (ATPS) (Bohrer & Takahashi, 2016; Chan *et al.*, 2013; Takahashi, 2019). The APS will be synthesised to either 3'-phosphoadenosine 5'-phosphosulfate (PAPS) by APS kinase (APK) which act as sulphur donor during the sulphur-containing metabolite biosynthesis or will be further reduced into sulphite (SO_3^{2-}) by APS reductase (APR) and then further reduced to sulphide (S^{2-}) by sulfite reductase (SIR). Sulphide will then be incorporated into O-acetylserine by O-acetylserine(thiol)lyase to form cysteine which later can be synthesised into methionine or glutathione (Bohrer & Takahashi, 2016; Chan *et al.*, 2013; Takahashi, 2019; Zhao *et al.*, 2008). Sulphate and sulphur-containing molecules can be transported from leaves to sink organs through the phloem by SULTR2 under sulphur sufficient condition (Bohrer & Takahashi, 2016; Takahashi, 2019). Unlike plant under nitrogen or phosphate starvation, plant under sulphur starvation condition prevent the sulphur from leaving the leaves by downregulation of SULTR2 which is why the observed chlorosis start from the young leaf during sulphur starvation in plant (Shanmugam, 2016; Takahashi, 2019).

The downregulation of *SULTR2;1* in-conjunction with the upregulation of *SULTR4;1*, *SULTR3;1*, *APR1* and *Response to Low Sulfur 3-like (LSU3)* in non-trunking sago palms suggested that the plant is under sulphur starvation (**Table 4.29**). LSU3 were reported to be upregulated in *Arabidopsis thaliana* under sulphur starvation but the function is still unknown (Sirko *et al.*, 2015). Upregulation of *SULTR4;1* increases the efflux of the sulphate which stored in vacuole while upregulation of *SULTR3;1* increases the influx of the sulphate from cytoplasm into chloroplast for sulphur assimilation (Bohrer & Takahashi, 2016; Takahashi, 2019). Upregulation of APR favour the synthesis of cysteine which is important in plant primary metabolisms for protein synthesis (Bohrer & Takahashi, 2016; Chan *et al.*, 2013; Shanmugam, 2016; Zhao *et al.*, 2008).

Table 4.33: DEGs involved in Sulphur Regulation

Description: DEGs involves in sulphur regulation in non-trunking sago palm are mixed regulated.

Sulphur Assimilation	DEGs	log₂ Fold-Change (N/T)	Description
<i>SULTR2;1</i>	CL2107.Contig1_All	-1.29	Involved in sulphate efflux into xylem or phloem
	CL2107.Contig2_All	-1.14	
<i>SULTR4;1</i>	CL3065.Contig4_All	1.55	Involved in sulphate efflux from vacuole
	CL3065.Contig3_All	1.796	
<i>SULTR3;1</i>	CL5119.Contig1_All	2.18	Involved in sulphate influx into chloroplast
<i>5'-adenylylsulfate reductase 1 (APR1), chloroplastic</i>	CL1088.Contig5_All	1.92	Involved in sulphur assimilation into cysteine
<i>Response to Low Sulfur 3-like (LSU3)</i>	Unigene23088_All	1.93	Upregulated in response to sulphur deficiency with unknown function

Boron is one of the micronutrients essential in plant growth. Boron intake in the form of boric acid [B(OH)₃] via passive diffusion, facilitated diffusion by protein channels, and can be effluxes out from cytoplasm via transporters (Brdar-Jokanović, 2020; Yoshinari & Takano, 2017). Boron primary function in plant metabolism involves the stabilisation of molecules with cis-diol groups which later involved in: membrane and cell wall biosynthesis; transport of sugars across the membrane; lignin and flavonoid synthesis; and

metabolism of auxins, nitrogen compounds, and phenols (Brdar-Jokanović, 2020; Yoshinari & Takano, 2017). Boron toxicity affects cellular metabolism, causes DNA damage, which results in tissue necrosis often observed in plant starting from the tip of the older leaves that would spread inwards in prolonged toxicity and in severe toxicity causes further necrosis in leaf sheaths, stems, spikes, and awns (Brdar-Jokanović, 2020; Kumar *et al.*, 2014; Yoshinari & Takano, 2017). Boron toxicity also causes: delayed emergence and foliation; reduction in stem height and yield; decrease lateral roots growth; young shoot tip cassation; and bud abscission (Brdar-Jokanović, 2020).

Boric acid intake into the plant by facilitated diffusion are aided by major intrinsic protein (MIPs) family which further divide into nodulin-26 like intrinsic protein (NIPs), plasma membrane intrinsic proteins (PIPs), small basic intrinsic proteins (SIPs), tonoplast intrinsic proteins (TIPs), and X intrinsic proteins (XIPs). NIPs are further divided into three subclasses and only subclass II are significant in boron transport. NIP5;1 involved in boron uptake in plant roots (Yoshinari & Takano, 2017). NIP6;1 on the other hand involved in transporting boron from xylem to phloem for further distribution to sink organs (Tanaka *et al.*, 2008). NIP7;1 involved in the boron transport for microsporogenesis (Brdar-Jokanović, 2020; Yoshinari & Takano, 2017). PIP2;4 and PIP2;7 involved in boron efflux from plant which upregulate during boron toxicity (Kumar *et al.*, 2014). Boron transporter family (BOR) consist of seven paralogs involved in borate (B(OH)_4^-) efflux from plant cell (Brdar-Jokanović, 2020). BOR1 involved in the translocation of boron from root to xylem for further distribution to sink organ (Miwa *et al.*, 2013). BOR2 and BOR4 are involved in the translocation of boron from symplast to apoplast which reduced the cellular boron content and assist in the rhamnogalacturonan-II (RG-II)-borate crosslinking in cell wall biosynthesis (Miwa *et al.*, 2013).

One of the strategies of plant in overcoming boron toxicity is bypassing the boron movement in plant out of the cytoplasm. Upregulation of boron transportation gene *PIP2;4*, *BOR2* and *BOR4* while downregulation of *NIP6;1* in non-trunking sago palms leaf tissue suggested that the plant is under boron toxicity and diverting the flow of boron out of the cytoplasm and further prevent the distribution of boron to other organs by limiting the movement of boron into phloem (**Table 4.34**). Upregulation of *NIP2;1* in non-trunking sago palms leaf tissue that is responsible for silicon uptake in plant which promote the expression of *BOR2* are induced by boron toxicity (Celikkol Akcay & Erkan, 2016; Ma *et al.*, 2006) (**Table 4.35**). Another strategy of plant to overcome boron toxicity is by reducing transpiration, resulting in lesser boron uptake and reduction of aquaporin (PIPs) were one of the ways to reduce transpiration (Macho-Rivero *et al.*, 2018). The contradiction finding of upregulation of *PIP1;2* under boron toxicity may be due to the plant respond toward nitrogen, phosphate and sulphur starvation at the same time.

Table 4.34: DEGs involved in Boron Transport

Description: DEGs involves in boron transport in non-trunking sago palm are mixed regulated.

Boron Transport	DEGs	log ₂ Fold-Change (N/T)	Description
<i>probable boron transporter 2 (BOR2)</i>	CL5711.Contig1_All	2.27	Efflux borate from symplast to apoplast
	CL5711.Contig2_All	3.38	
	CL5711.Contig3_All	3.49	
<i>boron transporter 4-</i>	CL3775.Contig1_All	2.08	Efflux borate from symplast to apoplast
	CL3775.Contig3_All	2.05	

Table 4.34 continued

<i>like isoform XI (BOR4)</i>			
<i>aquaporin PIP2;4 isoform XI</i>	Unigene3535_All	3.47	Involve in boron efflux out of the plant.
<i>aquaporin NIP6;1</i>	Unigene189_All	-3.24	Involved in boron transport from xylem to phloem.
	Unigene190_All	-4.90	
	CL4818.Contig2_All	-3.23	
	CL4818.Contig3_All	-5.28	
	CL4818.Contig4_All	-5.05	

Table 4.35: Other DEGs involved in Boron Regulation

Description: DEGs involves in boron regulation in non-trunking sago palm are upregulated.

Other Boron Tolerance Related Genes	DEGs	log ₂ Fold-Change (N/T)	Description
<i>probable aquaporin PIP1;2</i>	CL5378.Contig4_All	2.92	Control water uptake in plant. Downregulated in short boron toxicity respond.
	CL5378.Contig5_All	2.19	
	CL5378.Contig6_All	2.27	
<i>aquaporin NIP2;1</i>	CL5995.Contig1_All	1.36	Silicon competitive inhibitor of boron transporter which also induce the expression of BOR2

Aluminium is not an essential nutrient in plant but it does bring benefits by promoting plant growth, increase stress tolerance and promote essential nutrient uptake but too much of aluminium causes toxicity to plant which affect its growth, development and nutrient uptake, (Bojórquez-Quintal *et al.*, 2017). Aluminium regulation in plant are not well studied hence how the aluminium enters the plant and transported is still unknown (Liu *et al.*, 2022).

One of the responses of plant toward aluminium toxicity are with rhizosphere pH modification by influx proton (H^+) and ammonium (NH_4^+) from the soil and efflux organic acid or mucilage to chelate soil aluminium (Al^{3+}) (Bojórquez-Quintal *et al.*, 2017; Liu *et al.*, 2009). Cell wall and plasma membrane component modification is among the strategy adapted by plant to reduce the binding of aluminium on it which could altered its original strength and permeability (Bojórquez-Quintal *et al.*, 2017; Huang *et al.*, 2009). Aluminium toxicity in the plant can be alleviated by aluminium chelator such as organic acids and phenolic compound to form non-toxic complexes which can later be translocated out of the plant or to senescence organ. Upregulation of metabolic pathways synthesising aluminium chelators and oxidative stress regulators are among the strategy to overcome aluminium toxicity in plant (Bojórquez-Quintal *et al.*, 2017).

Upregulation of both *STAR1* and *DTX42* suggest the respond of non-trunking sago palms leaf tissue toward aluminium toxicity where modification of cell wall preventing aluminium binding and transport of aluminium chelating agent, citrate, from source to sink organ. Citrate can be effluxed out to rhizosphere to chelates the soil aluminium which reduce aluminium intake afterward (**Table 4.36**).

Table 4.36: DEGs involved in Aluminium Regulation

Description: DEGs involves in aluminium regulation in non-trunking sago palm are upregulated.

Gene involved in Aluminium Regulation	DEGs	log ₂ Fold-Change (N/T)	Description
<i>sensitive to Al rhizotoxicity1 (STAR1)</i>	Unigene7135_All	1.98	Involved in the transport of UDP-glucose used in cell wall modification which prevent Al ³⁺ -binding
<i>DETOXIFICATION 42 (DTX42)</i>	CL2527.Contig3_All	1.60	Involved in citrate transport (aluminium chelator)
	CL2527.Contig1_All	1.82	
	CL2527.Contig5_All	2.06	
	CL591.Contig1_All	2.53	

4.5 DEGs Comparison with Previous Transcriptome Studies in Non-Trunking Sago Palms

Previous studies in transcriptome comparison between trunking and non-trunking sago palms by Anastasia (2013) and Siti (2015) revealed that the genes that were differentially expressed related to the: active uptake and utilisation of nitrogen (N); regulation of phosphate (P), cobalt (Co), calcium (Ca), magnesium (Mg) and zinc (Zn), pathogenic defence; plant stress mediation signalling; flowering and growth; protein trafficking; cell metabolism; and sugar accumulation. In order to correlate the previous finding with this study, a comparison of the DEGs obtained by previous studies was

performed. The lists of DEGs from the previous findings obtained by differential display analysis and representational difference analysis were aligned with the RNA-seq data obtained in this study and the NCBI-NR database (**Appendix X & Appendix Y**). Sequence alignment with the RNA-seq data obtained in this study shows that the list of DEGs obtained from their studies were not listed in the DEGs obtained in this study although most of it aligned with high similarity score to the list of genes obtained from this study. Sequence alignment of their findings with RNA-seq data obtained in this study shows that there are two unique sequences and one unique sequence reported by Anastasia (2013) and Siti (2015) respectively. The unique sequences from the previous study were unannotated genes and S-adenosylmethionine synthetase: which involved in various processes such as methylation reactions, production of polyamines and phytohormone ethylene, etc. (Heidari *et al.*, 2020; Sekula *et al.*, 2020). The unique sequences from their finding maybe related to young leaf growth and development since their RNA sample originated from young sago palms leaf while in this study, matured sago palms leaves were used. Sequence alignment of their finding with the newer NCBI-NR database shows a better annotation compared to their previous annotation by the older NCBI-NR database because of the more sequences uploaded and updated in the newer NCBI-NR database. The sequences alignment of the findings by Anastasia (2013) and Siti (2015) with the RNA-seq data provide a better hit compared to direct sequences alignment with NCBI-NR database because of the higher similarity of their transcripts with the RNA-seq data obtained in this study which have not been uploaded into the NCBI database (**Appendix X & Appendix Y**). Longer sequence of the RNA-seq data provides a better precision of samples sequence alignment with the NCBI-NR database. Hence the unannotated transcripts from the previous studies were able to be annotated using the sequence data obtained from this study.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

The first objective of this study, to extract adequate quality and quantity of total RNA from sago palms leaf tissue, was achieved where high-quality total RNA was extracted from sago palms leaf tissue using the improved CTAB method. The second objective, which was to identify the differentially expressed genes (DEGs) in leaf tissue between the trunking (normal) and non-trunking (target of interest) sago palm, was achieved as a total of 936 DEGs of non-trunking sago palms were identified. The third objective, which was to correlate the DEGs with the cause and effect of the non-trunking phenomenon in sago palms, was achieved. The findings of this study by functional, pathway and PPI analysis showed that transcriptomic levels of various enzymes involved in many biosynthesis and enzymatic pathways, such as phenylpropanoid biosynthesis, biosynthesis of secondary metabolite, flavonoid biosynthesis and circadian rhythm with a false discovery rate of less than 1%, are differentially expressed between the trunking and non-trunking sago palms leaf tissue. Therefore, based on the differential transcriptomic levels affecting various pathways and functions, this study concluded that nitrogen, phosphate and sulphate nutrient deficiency caused the non-trunking phenomenon in sago palm. Nutrient starvation negatively affects the normal biological processes in the sago palm; which aggravates the effect of boron and aluminium toxicity; and malnutrition in sago palms is also more susceptible to pathogenic infection; which ultimately leads to stunted growth and development of the plant. The plant responded to these stresses by negatively regulating flowering development; and aluminium and boron transport; while positively regulating phosphate, sulphate and nitrate transport;

and the photomorphogenesis for lignin, antioxidant and flavonoid biosynthesis. These responses increase the intake and assimilation of nitrate, phosphate and sulphate; while diverting the excess boron and aluminium to the apoplastic pathway; so that it will not disturb the normal cellular function and limit its distribution to other organs. These responses also increase the synthesis of metabolites involved in cell wall biosynthesis, oxidative stress alleviation, pathogenic defence and wound healing.

This work contributes towards filling the knowledge gap in the trunking and non-trunking phenomenon of sago palm by identifying 936 DEGs between trunking and non-trunking sago palm, highlighting the significant functions and pathways of the DEGs, relates the finding with the non-trunking phenomenon in sago palm and finally explained the causes and effects of the non-trunking phenomenon in sago palm.

4.2 Recommendations

Natural deep peatland is not suitable for sago palm cultivation. Proper management of the deep peatland is required to improve the growth and development of sago palm. In this context, proper fertilisation of the sago palm in deep peat soil with nitrogen, phosphate and sulphate fertiliser is required for optimal growth. The nutrient leaching properties of peatland (waterlogging and high peat depth) is a challenge which can be overcome with the application of slow-release/controlled-release fertiliser or by a more frequent but lower amount of normal fertiliser application. Studies on the optimal amount of nutrient intake by sago palm and planting strategies are needed to design an optimum fertilisation strategy. Better irrigation management for the sago palm plantation by lowering the groundwater level will also reduce nutrient leaching and promote organic matter decomposition, ultimately improving mineral deposition.

Selection of a better tolerance sub-cultivar of sago palm toward the nitrogen, phosphate and sulphate availability in soil for plantation in deep peatland is also a recommended strategy. More studies on the sub-cultivar of sago palm are needed to select the proper sub-cultivar for deep peatland sago palm cultivation.

To better understand the non-trunking phenomenon, further research on the non-trunking phenomenon can be done, such as correlating the DEGs of bole meristems and root with the finding of this study will collectively formulate a more complete causative of the non-trunking phenomenon. Other than that, the improvement on the high-throughput protein sequencing techniques; protein databases; and the bioinformatics tools; whole protein profile comparison between trunking and non-trunking sago palm could be performed. Complementing the finding in this study with the differentially expressed protein will further improve the understanding of the cause and effect of non-trunking in sago palm.

The improved understanding of the reason for non-trunking and the improvement in the planting, fertilisation and irrigation strategies makes it possible for the sago palm to thrive even in deep peat soil in the future.

REFERENCES

- Apun, K., Lihan, S., Wong, M. K., & Bilung, L. M. (2009). Microbiological Characteristics of Trunking and Non-Trunking Sago Palm Peat Soil. In K. Bujang, Z. Z. Zakaria, U. H. E. Sim, K. Apun, H. Hussain, Z. Ngaini, & I. Ipor (Eds.), *1st ASEAN Sago Symposium: Current Trend and Development in Sago Research*, Sarawak, Malaysia, 29-30 October 2009. FRST Universiti Malaysia Sarawak.
- Arenhart, R. A., Bai, Y., de-Oliveira, L. F. V., Neto, L. B., Schunemann, M., Maraschin, F. d. S., Mariath, J., Silverio, A., Sachetto-Martins, G., Margis, R., Wang, Z. Y., & Margis-Pinheiro, M. (2014). New Insights into Aluminum Tolerance in Rice: The ASR5 Protein Binds The STAR1 Promoter and Other Aluminum-Responsive Genes. *Molecular Plant*, 7(4), 709-721.
- Asare, E. K., Jaiswal, S., Maley, J., Båga, M., Sammynaiken, R., Rossnagel, B. G., & Chibbar, R. N. (2011). Barley Grain Constituents, Starch Composition, and Structure Affect Starch in Vitro Enzymatic Hydrolysis. *Journal of Agricultural and Food Chemistry*, 59(9), 4743-4754.
- Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R., & Hanley-Bowdoin, L. (2008). Global Analysis of *Arabidopsis* Gene Expression Uncovers a Complex Array of Changes Impacting Pathogen Response and Cell Cycle during Geminivirus Infection. *Plant Physiology*, 148(1), 436–454.
- Autio, K., & Eliasson, A. C. (2009). Rye Starch. In J. N. BeMiller & R. L. Whistler (Eds.), *Starch: Chemistry and Technology*, (3rd ed, pp. 579-588). Elsevier Science.

- Bansal, G., Narta, K., & Teltumbade, M. R. (2018). Next-Generation Sequencing: Technology, Advancements, and Applications. In A. Shanker (Ed.), *Bioinformatics: Sequences, Structures, Phylogeny*, (pp. 15-46). Springer Nature Singapore Pte Ltd.
- Belal, R., Tang, R., Li, Y., Mabrouk, Y., Badr, E., & Luan, S. (2015). An ABC Transporter Complex Encoded by Aluminum Sensitive 3 and NAP3 is Required for Phosphate Deficiency Responses in *Arabidopsis*. *Biochemical and Biophysical Research Communications*, 463(1-2), 18-23.
- BGI. (2019a). *DNBseq™ Service Overview: RNA-Seq (Transcriptome) Sequencing*. PN: BGIASO4_061319. [Online] Available at: <https://www.bgi.com/api/assets/01b07d24-e382-4c07-b657-3ee75c9f7c0d> [Assessed on 29 December 2022].
- BGI. (2019b). *PLAphqT [Transcriptome Denovo Project] 2019/10/21* (F19FTSAPHT0409). BGI.
- Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). QuickGO: a Web-Based Tool for Gene Ontology Searching. *Bioinformatics*, 25(22), 3045-3046.
- Bintoro, M. H., Iqbal Nurulhaq, M., Pratama, A. J., Ahmad, F., & Ayulia, L. (2018). Growing Area of Sago Palm and Its Environment. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 17-29). Springer Nature.
- BioNinja. (n.d.). *DNA Sequencing*. [Online] Available at: <https://ib.bioninja.com.au/higher-level/topic-7-nucleic-acids/71-dna-structure-and-replic/dna-sequencing.html> [Assessed on 11 March 2023].

- Bock, K. W., Honys, D., Ward, J. M., Padmanaban, S., Nawrocki, E. P., Hirschi, K. D., Twell, D., & Sze, H. (2006). Integrating Membrane Transport with Male Gametophyte Development and Function through Transcriptomics. *Plant Physiology*, 140(4), 1151–1168.
- Bohrer, A.-S., & Takahashi, H. (2016). Compartmentalization and Regulation of Sulfate Assimilation Pathways in Plants. In W. J. Kwang & L. Galluzzi (Eds.), *International Review of Cell and Molecular Biology*, (Vol. 326, pp. 1-31). Elsevier.
- Bojórquez-Quintal, E., Escalante-Magaña, C., Echevarría-Machado, I., & Martínez-Estévez, M. (2017). Aluminum, a Friend or Foe of Higher Plants in Acid Soils. *Frontiers in Plant Science*, 8, 1767.
- Brdar-Jokanović, M. (2020). Boron Toxicity and Deficiency in Agricultural Plants. *International Journal of Molecular Sciences*, 21(4), 1424.
- Broschat, T. K. (2015). Fertilization of Landscape Palms to Reduce Nitrogen and Phosphate Impacts on the Environment. *American Society for Horticultural Science*, 50(3), 469-473.
- Broschat, T. K. (2020). *Nutrient Deficiencies of Landscape and Field-Grown Palms in Florida*. Gainesville, Florida: UF/IFAS Extension.
- Buer, C. S., Muday, G. K., & Djordjevic, M. A. (2008). Implications of Long-Distance Flavonoid Movement in *Arabidopsis thaliana*. *Plant Signaling & Behavior*, 3(6), 415-417.
- Bujang, K. (2018). Production, Purification, and Health Benefits of Sago Sugar. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 299-307). Springer Nature.

- Can, T. (2014). Introduction to Bioinformatics. In Y. Malik & J. Allmer (Eds.), *miRNomics: MicroRNA Biology and Computational Analysis. Methods in Molecular Biology*, (Vol. 1107, pp. 51-71). Humana Press.
- Castelán-Muñoz, N., Herrera, J., Cajero-Sánchez, W., Arrizubieta, M., Trejo, C., García-Ponce, B., Sánchez, M. d. I. P., Álvarez-Buylla, E. R., & Garay-Arroyo, A. (2019). MADS-Box Genes are Key Components of Genetic Regulatory Networks Involved in Abiotic Stress and Plastic Developmental Responses in Plants. *Frontiers in Plant Science*, *10*, 853.
- Celikkol Akcay, U., & Erkan, I. E. (2016). Silicon Induced Antioxidative Responses and Expression of BOR2 and Two PIP Family Aquaporin Genes in Barley Grown Under Boron Toxicity. *Plant Molecular Biology Reporter*, *34*(1), 318-326.
- Cesco, S., Neumann, G., Tomasi, N., Pinton, R., & Weiskopf, L. (2010). Release of Plant-Borne Flavonoids into the Rhizosphere and Their Role in Plant Nutrition. *Plant and Soil*, *329*(1), 1-25.
- Chan, K. X., Wirtz, M., Phua, S. Y., Estavillo, G. M., & Pogson, B. J. (2013). Balancing Metabolites in Drought: The Sulfur Assimilation Conundrum. *Trends in Plant Science*, *18*(1), 18-29.
- Cui, L., Dong, S., Zhang, J., & Liu, P. (2014). Starch Granule Size Distribution and Morphogenesis in Maize (*Zea mays* L.) Grains with Different Endosperm Types. *Australian Journal of Crop Science*, *8*, 1560-1565.

- Daram, P., Brunner, S., Rausch, C., Steiner, C., Amrhein, N., & Buchera, M. (1999). Pht2;1 Encodes a Low-Affinity Phosphate Transporter from *Arabidopsis*. *The Plant Cell*, 11, 2153–2166.
- Darma. (2018). Improvement of Sago Processing Machinery. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 231-245). Springer Nature.
- de Torres-Zabala, M., Truman, W., Bennett, M. H., Lafforgue, G., Mansfield, J. W., Rodriguez Egea, P., Bögre, L., & Grant, M. (2007). *Pseudomonas syringae* pv. Tomato Hijacks the *Arabidopsis* Absciscic Acid Signalling Pathway to Cause Disease. *The European Molecular Biology Organization Journal*, 26(5), 1434-1443.
- Deleu, L. J., Lemmens, E., Redant, L., & Delcour, J. A. (2020). The Major Constituents of Rye (*Secale cereale* L.) Flour and Their Role in the Production of Rye Bread, a Food Product to Which a Multitude of Health Aspects are Ascribed. *Cereal Chemistry*, 97(4), 739-754.
- Dictionary of Botany. (2003). *Arecidae*. [Online] Available at: <http://www.botanydictionary.org/arecidae.html> [Assessed on 13 February 2022].
- Digel, B., Pankin, A., & von Korff, M. (2015). Global Transcriptome Profiling of Developing Leaf and Shoot Apices Reveals Distinct Genetic and Environmental Control of Floral Transition and Inflorescence Development in Barley. *The Plant Cell*, 27(9), 2318-2334.
- Dobritsch, M., Lübken, T., Eschen-Lippold, L., Gorzolka, K., Blum, E., Matern, A., Marillonnet, S., Böttcher, C., Dräger, B., & Rosahl, S. (2016). MATE Transporter-Dependent Export of Hydroxycinnamic Acid Amides. *The Plant Cell*, 28(2), 583-596.

- Dransfield, J., Uhl, N. W., Asmussen, C. B., Baker, W. J., Harley, M. M., & Lewis, C. E. (2008). *Genera Palmarum - The Evolution and Classification of the Palms*. Surrey, UK: Royal Botanic Gardens, Kew.
- Duan, K., Yi, K., Dang, L., Huang, H., Wu, W., & Wu, P. (2008). Characterization of a Sub-Family of *Arabidopsis* Genes with the SPX Domain Reveals Their Diverse Functions in Plant Tolerance to Phosphorus Starvation. *The Plant Journal*, 54, 965–975.
- Duke, S. O. (2022). Benefits of Resveratrol and Pterostilbene to Crops and Their Potential Nutraceutical Value to Mammals. *Agriculture*, 12(3), 368.
- Edward-Atit, A. S. (2013). *Differential Expression Gene Profiling Studies on Trunking and Non-Trunking Sago Palm (Metroxylon sagu Rottb.)* [Master's Dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak, Malaysia. Available at: <https://ir.unimas.my/id/eprint/9378/>
- Ehara, H. (2018). Genetic Variation and Agronomic Features of *Metroxylon* Palms in Asia and Pacific. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 45-59). Springer Nature.
- Ejigu, G. F., & Jung, J. (2020). Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology*, 9(9), 295.
- Ellen, R. (2006). Local Knowledge and Management of Sago Palm (*Metroxylon sagu* Rottboell) Diversity in South Central Seram, Maluku, Eastern Indonesia. *Journal of Ethnobiology* 26(2), 258–298.

- Fabre, A.-L., Colotte, M., Luis, A., Tuffet, S., & Bonnet, J. (2014). An Efficient Method for Long-Term Room Temperature Storage of RNA. *European Journal of Human Genetics*, 22(3), 379-385.
- Falcone Ferreyra, M. L., Rius, S. P., & Casati, P. (2012). Flavonoids: Biosynthesis, Biological Functions, and Biotechnological Applications. *Frontiers in Plant Science*, 3, 222-222.
- Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H., Lu, H., Luber, J. M., Xia, M., Li, W., Stein, S., Xu, X., Zhang, W., Drmanac, R., Wang, J., Yang, H., Hammarström, L., Kostic, A. D., Kristiansen, K., & Li, J. (2017). Assessment of the cPAS-bBased BGISEQ-500 Platform for Metagenomic Sequencing. *GigaScience*, 7(3), gix133.
- Flach, M. (1997). *Sago Palm: Metroxylon sagu Rottb.* Rome, Italy: Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute.
- Food and Agriculture Organization of the United Nation. (2022). *Crops and Livestock Products* [Database]. FAOSTAT. Available at: <https://www.fao.org/faostat/en/#data/QCL> [Assessed on 13 February 2022].
- Fukuchi-Mizutani, M., Tasaka, Y., Tanaka, Y., Ashikari, T., Kusumi, T., & Murata, N. (1998). Characterization of $\delta A9$ Acyl-Lipid Desaturase Homologues from *Arabidopsis thaliana*. *Plant and Cell Physiology*, 39(2), 247-253.
- 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.

- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011). Full-Length Transcriptome Assembly from RNA-Seq Data Without a Reference Genome. *Nature Biotechnology*, 29(7), 644-652.
- Guo, H., Pu, X., Jia, H., Zhou, Y., Ye, G., Yang, Y., Na, T., & Wang, J. (2022). Transcriptome Analysis Reveals Multiple Effects of Nitrogen Accumulation and Metabolism in the Roots, Shoots, and Leaves of Potato (*Solanum tuberosum* L.). *BioMed Central Plant Biology*, 22, 282.
- Hameed, M., Malik, S. R., Iqbal, M. F., & Mehmood, M. (2015). Extraction of Starch from Potato by Enzymatic Process. *Science International (Lahore)*, 27(6), 6049-6052.
- Heidari, P., Mazloomi, F., Nussbaumer, T., & Barcaccia, G. (2020). Insights into the SAM Synthetase Gene Family and Its Roles in Tomato Seedlings under Abiotic Stresses and Hormone Treatments. *Plants*, 9(5), 586.
- Heilmann, I., Pidkowich, M. S., Girke, T., & Shanklin, J. (2004). Switching Desaturase Enzyme Specificity by Alternate Subcellular Targeting. *Proceedings of the National Academy of Sciences of the United States of America*, 101(28), 10266-10271.
- Heptinstall, J., & Rapley, R. (2000). Spectrophotometric Analysis of Nucleic Acids. In R. Rapley (Ed.), *The Nucleic Acid Protocols Handbook*, (pp. 57-60). Humana Press.

- Hirao, K., Kondo, T., Kainuma, K., & Takahashi, S. (2018). Starch Properties and Uses as Food for Human Health and Welfare. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 285-298). Springer Nature.
- Hodgkiss, R. J. (2016). *Order Arecales - Palm Trees*. [Online] Available at: <https://succulent-plant.com/families/arecales.html> [Assessed on 13 February 2022].
- Hu, C. H., Wang, P. Q., Zhang, P. P., Nie, X. M., Li, B. B., Tai, L., Liu, W. T., Li, W. Q., & Chen, K. M. (2020). NADPH Oxidases: The Vital Performers and Center Hubs During Plant Growth and Signaling. *Cells*, 9(2), 437.
- Huang, C. F., Yamaji, N., Mitani, N., Yano, M., Nagamura, Y., & Ma, J. F. (2009). A Bacterial-Type ABC Transporter is Involved in Aluminum Tolerance in Rice. *The Plant Cell*, 21(2), 655-667.
- Hussain, H., Wei-Jie, Y., Zainab, N., Julaihi, N., Rina, T., & Showkat, A. B. (2020). Differential Metabolites Markers from Trunking and Stressed Non-Trunking Sago Palm (*Metroxylon sagu* Rottb.). *Current Chemical Biology*, 14(4), 262-278.
- Hussain, M. H., Kamal, M. M., Edward-Atit, A. S., Kamarol, S. I. L., Yan, W. J., Ngaini, Z., Al-Obaidi, J., & Mohd-Yusuf, Y. (2012). Proteomics of Trunking and Nontrunking Sago Palm (*Metroxylon Sagu* Rottb.). In N. Busri, Z. C. Abdullah, C. N. Hipolito, M. H. Hussain, H. A. Roslan, U. H. E. Sim, S. Adeni, A. Husaini, & N. Humrawali (Eds.), *2nd ASEAN Sago Symposium 2012: Advances in Sago Research and Development*, Sarawak, Malaysia, 29 - 31 October 2012. FRST Universiti Malaysia Sarawak.
- Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Eveno, E., Mueller, O., Schroeder, A., & Auffray, C. (2005). Towards Standardization of RNA Quality

- Assessment using User-Independent Classifiers of Microcapillary Electrophoresis Traces. *Nucleic Acids Research*, 33(6), e56.
- Jacks, T. J., De-Lucca, A. J., Rajasekaran, K., Stromberg, K., & van Pée, K. (2000). Antifungal and Peroxidative Activities of NonHeme Chloroperoxidase in Relation to Transgenic Plant Protection. *Journal of Agricultural and Food Chemistry*, 48(10), 4561–4564.
- Ji, D. L., Lin, H., Chi, W., & Zhang, L. X. (2012). CpLEPA Is Critical for Chloroplast Protein Synthesis Under Suboptimal Conditions in *Arabidopsis thaliana*. *Public Library of Science One*, 7(11), e49746.
- Jong, F. S. (2018). An Overview of Sago Industry Development, 1980s–2015. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 75-89). Springer Nature.
- Kamal, M. M. (2014). *Identification of Differentially Expressed Protein in Trunking and Non-Trunking Sago Palm Leaves* [Master's dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak, Malaysia. Available at: <https://ir.unimas.my/id/eprint/8350/>
- Kamarol, S. I. L. (2015). *Representational Difference Analysis (RDA) for Identification of Molecular Factors Contributing to Trunking and Non Trunking Sago Palm (Metroxylon sagu)* [Master's Dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak, Malaysia. Available at: <https://ir.unimas.my/id/eprint/9389/>
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30.

- Kaur, K., Ahluwalia, P., & Singh, H. (2016). Cassava: Extraction of Starch and Utilization of Flour in Bakery Products. *International Journal of Food and Fermentation Technology* 6(2), 351-355.
- Kawai, S., Mori, A., Shiokawa, T., Kajita, S., Katayama, Y., & Morohoshi, N. (1996). Isolation and Analysis of Cinnamic Acid 4-Hydroxylase Homologous Genes from a Hybrid Aspen, *Populus kitakamiensis*. *Bioscience, Biotechnology, and Biochemistry*, 60(10), 1586-1597.
- Kayihan, C. (2021). The Involvement of the Induction of Anthocyanin Biosynthesis and Transport in Toxic Boron Responsive Regulation in *Arabidopsis thaliana*. *Turkish Journal of Botany*, 45(3), 181-191.
- Kendra, K. V. (2013). *A Guide to Identification and Management of Nutrient Deficiencies & Cultural Practices in Plantation Crops*. Goa, India: ICAR Research Complex for Goa.
- Kobayashi, M., Kawakita, K., Maeshima, M., Doke, N., & Yoshioka, H. (2006). Subcellular Localization of Strboh Proteins and NADPH-dependent O₂-Generating Activity in Potato Tuber Tissues. *Journal of Experimental Botany*, 57(6), 1373-1379.
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 2015(11), 951-969.
- Kumar, K., Mosa, K. A., Chhikara, S., Musante, C., White, J. C., & Dhankher, O. P. (2014). Two Rice Plasma Membrane Intrinsic Proteins, OsPIP2;4 and OsPIP2;7, are Involved in Transport and Providing Tolerance to Boron Toxicity. *Planta*, 239(1), 187-198.

- Larkindale, J., & Vierling, E. (2008). Core Genome Responses involved in Acclimation to High Temperature. *Plant Physiology*, 146(2), 748-761.
- Lee, D. J., Park, J.-Y., Ku, S.-J., Ha, Y.-M., Kim, S., Kim, M. D., Oh, M.-H., & Kim, J. (2007). Genome-Wide Expression Profiling of ARABIDOPSIS RESPONSE REGULATOR 7 (ARR7) Overexpression in Cytokinin response. *Molecular Genetics and Genomics*, 277(2), 115-137.
- Li, B., & Dewey, C. N. (2011). RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BioMed Central Bioinformatics*, 12(1), 323.
- Li, S. M., Li, B. Z., & Shi, W. M. (2012). Expression Patterns of Nine Ammonium Transporters in Rice in Response to N Status. *Pedosphere*, 22(6), 860–869.
- Liao, W., Mei, Z., Miao, L., Liu, P., & Gao, R. (2020). Comparative Transcriptome Analysis of Root, Stem, and Leaf Tissues of *Entada phaseoloides* Reveals Potential Genes Involved in Triterpenoid Saponin Biosynthesis. *BioMed Central Genomics*, 21, 639.
- Liew, S. H. (2004). *Groundwater Characteristics of Peat Soils Under Trunking and Non-Trunking Sago Palms in the Dalat Sago Plantation* [Master's Dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak, Malaysia. Available at: <https://ir.unimas.my/id/eprint/14781/>
- Lim, L. W. K., Chung, H. H., Hussain, M. H., & Gan, H. M. (2021). Genome Survey of Sago Palm (*Metroxylon sagu* Rottboll). *Plant Gene*, 28, 100341.
- Liu, C.-W., & Murray, J. D. (2016). The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. *Plants*, 5(3), 33.

- Liu, H., Zhu, R., Shu, K., Lv, W., Wang, S., & Wang, C. (2022). Aluminum Stress Signaling, Response, and Adaptive Mechanisms in Plants. *Plant Signaling & Behavior*, 17(1), e2057060.
- Liu, J., Magalhaes, J. V., Shaff, J., & Kochian, L. V. (2009). Aluminum-Activated Citrate and Malate Transporters from the MATE and ALMT Families Function Independently to Confer Arabidopsis Aluminum Tolerance. *The Plant Journal*, 57, 389–399.
- Liu, Q., Luo, L., & Zheng, L. (2018). Lignins: Biosynthesis and Biological Functions in Plants. *International Journal of Molecular Sciences*, 19(2), 335.
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2. *Genome Biology*, 15(12), 550.
- Lu, P., Magwanga, R. O., Kirungu¹, J. N., Hu, Y., Dong, Q., Cai, X., Zhou, Z., Wang, X., Zhang, Z., Hou, Y., Wang, K., & Liu, F. (2019). Overexpression of Cotton a DTX/MATE Gene Enhances Drought, Salt, and Cold Stress Tolerance in Transgenic *Arabidopsis*. *Frontiers in Plant Science*, 10, 299.
- Lulie, M. (2000). *Dalat & Mukah Sago Plantation Peat Soil Study: Final report*. Sarawak, Malaysia: Sarawak Department of Agriculture, Soils Branch.
- Lumdubwong, N., & Seib, P. A. (2000). Rice Starch Isolation by Alkaline Protease Digestion of Wet-Milled Rice Flour. *Journal of Cereal Science*, 31(1), 63-74.

- Ma, J. F., Tamai, K., Yamaji, N., Mitani, N., Konishi, S., Katsuhara, M., Ishiguro, M., Murata, Y., & Yano, M. (2006). A Silicon Transporter in Rice. *Nature*, *440*, 688-691.
- Macho-Rivero, M. A., Herrera-Rodríguez, M. B., Brejcha, R., Schäffner, A. R., Tanaka, N., Fujiwara, T., González-Fontes, A., & Camacho-Cristóbal, J. J. (2018). Boron Toxicity Reduces Water Transport from Root to Shoot in *Arabidopsis* Plants. Evidence for a Reduced Transpiration Rate and Expression of Major PIP Aquaporin Genes. *Plant and Cell Physiology*, *59*(4), 841-849.
- Maher, M., Ahmad, H., Nishawy, E., Li, Y., & Luo, J. (2021). Novel Transcriptome Study and Detection of Metabolic Variations in UV-B-Treated Date Palm (*Phoenix dactylifera* cv. Khalas). *International Journal of Molecular Sciences*, *22*(5), 2564.
- Mahfuzur Rahman, S. M., Wheatley, C., & Rakshit, S. K. (2003). Selection of Sweet Potato Variety for High Starch Extraction. *International Journal of Food Properties*, *6*(3), 419-430.
- Marder, R. C., Cruz, R. d. A., Moreno, M. A., Curran, A., & Trim, D. S. (1996). Investigating Sour Starch Production in Brazil. In D. Dufour, G. M. O'Brien, & R. Best (Eds.), *Cassava Flour and Starch: Progress in Research and Development*, (pp. 247-258). International Center for Tropical Agriculture.
- Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., & Suzuki, A. (2010). Nitrogen Uptake, Assimilation and Remobilization in Plants: Challenges for Sustainable and Productive Agriculture. *Annals of Botany*, *105*(7), 1141-1157.

- Masoudi-Nejad, A., Narimani, Z., & Hosseinkhan, N. (2013). Next-Generation Sequencing Methodologies. In A. Masoudi-Nejad, Z. Narimani, & N. Hosseinkhan (Eds.), *Next Generation Sequencing and Sequence Assembly: Methodologies and Algorithms*, (pp. 1-9). Springer.
- Maxam, A. M., & Gilbert, W. (1977). A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), 560-564.
- Miwa, K., Wakuta, S., Takada, S., Ide, K., Takano, J., Naito, S., Omori, H., Matsunaga, T., & Fujiwara, T. (2013). Roles of BOR2, a Boron Exporter, in Cross Linking of Rhamnogalacturonan II and Root Elongation under Boron Limitation in *Arabidopsis*. *Plant Physiology*, 163(4), 1699-1709.
- Młodzińska, E., & Zboińska, M. (2016). Phosphate Uptake and Allocation - A Closer Look at *Arabidopsis thaliana* L. and *Oryza sativa* L. *Frontiers in Plant Science*, 7, 1198-1198.
- Mudge, S. R., Rae, A. L., Diatloff, E., & Smith, F. W. (2002). Expression Analysis Suggests Novel Roles for Members of the PHT1 Family of Phosphate Transporters in *Arabidopsis*. *The Plant Journal*, 31(3), 341-353.
- Niculaes, C., Abramov, A., Hannemann, L., & Frey, M. (2018). Plant Protection by Benzoxazinoids—Recent Insights into Biosynthesis and Function. *Agronomy*, 8(8), 143.
- Nishimura, Y. (2018). Sago Starch: Transformation of Extraction and Consumption Processes in Traditional Indonesian Societies. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 221-229). Springer Nature.

- Nix, A., Paull, C., & Colgrave, M. (2017). Flavonoid Profile of the Cotton Plant, *Gossypium hirsutum*: A Review. *Plants (Basel, Switzerland)*, 6(4), 43.
- Okazaki, M., & Sasaki, Y. (2018). Soil Environment in Sago Palm Forest. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 193-206). Springer Nature.
- Omar, K. A., Salih, B. M., Abdulla, N. Y., Hussin, B. H., & Rassul, S. M. (2016). Evaluation of Starch and Sugar Content of Different Rice Samples and Study their Physical Properties. *Indian Journal of Natural Sciences*, 6(36), 11084-11093.
- Ong, Z. P. (2019). *Rapid Amplification of cDNA Ends of Differentially Expressed Genes from Trunking and Non-Trunking Sago Palm (Metroxylon sagu Rottb.)* [Master's Dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak Malaysia. Available at: <https://ir.unimas.my/id/eprint/25237/>
- Pandey, V. P., Awasthi, M., Singh, S., Tiwari, S., & Dwivedi, U. N. (2017). A Comprehensive Review on Function and Application of Plant Peroxidases. *Biochemistry & Analytical Biochemistry*, 6(1), 308.
- Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., & Jugulam, M. (2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants*, 9, 454.
- Pérez, O. E., Haros, M., Suárez, C., & Rosell, C. M. (2003). Effect of Steeping Time on the Starch Properties from Ground Whole Corn. *Journal of Food Engineering*, 60, 281-287.
- Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., & Vianello, A. (2013). Plant Flavonoids--Biosynthesis, Transport and Involvement in Stress Responses. *International Journal of Molecular Sciences*, 14(7), 14950-14973.

- Pré, M., Atallah, M., Champion, A., De Vos, M., Pieterse, C. M. J., & Memelink, J. (2008). The AP2/ERF Domain Transcription Factor ORA59 Integrates Jasmonic Acid and Ethylene Signals in Plant Defense. *Plant Physiology*, 147(3), 1347-1357.
- Pue, A. G., Fletcher, M. T., Blaney, B., Greenhill, A. R., Warner, J. M., Atagazli, L., & Ng, J. C. (2018). Addressing Food Insecurity in Papua New Guinea Through Food Safety and Sago Cropping. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 123-137). Springer Nature.
- Puga, M. I., Mateos, I., Charukesi, R., Wang, Z., Franco-Zorrilla, J. M., de Lorenzo, L., Irigoyen, M. L., Masiero, S., Bustos, R., Rodríguez, J., Leyva, A., Rubio, V., Sommer, H., & Paz-Ares, J. (2014). SPX1 is a Phosphate-Dependent Inhibitor of Phosphate Starvation Response 1 in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(41), 14947-14952.
- Ramel, F., Sulmon, C., Cabello-Hurtado, F., Taconnat, L., Martin-Magniette, M.-L., Renou, J.-P., El Amrani, A., Couée, I., & Gouesbet, G. (2007). Genome-Wide Interacting Effects of Sucrose and Herbicide-Mediated Stress in *Arabidopsis Thaliana*: Novel Insights into Atrazine Toxicity and Sucrose-Induced Tolerance. *BioMed Central Genomics*, 8, 450-450.
- Rankine, I., & Fairhurst, T. H. (1999). Management of Phosphorus, Potassium and Magnesium in Mature Oil Palm. *Better Crops International*, 13, 10-15.
- Robertson, T. M., Alzaabi, A. Z., Robertson, M. D., & Fielding, B. A. (2018). Starchy Carbohydrates in a Healthy Diet: The Role of the Humble Potato. *Nutrients*, 10(11), 1764.

- Saha, D., Prasad, A. M., & Srinivasan, R. (2007). Pentatricopeptide Repeat Proteins and their Emerging Roles in Plants. *Plant Physiology and Biochemistry* 45, 521-534.
- Salil, C., Yong, X. L., Padmanaban, S., Nanatani, K., Uozumi, N., Rao, R., & Sze, H. (2011). Plant-Specific Cation/H⁺ Exchanger 17 and Its Homologs Are Endomembrane K⁺ Transporters with Roles in Protein Sorting. *Journal of Biological Chemistry*, 286(39), 33931–33941.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467.
- Sarawak State Secretary Office. (2022a). *Sago Areas - Sago Starch 10 Years Performance* [Dataset]. Available at: <https://data.sarawak.gov.my/home/data/resource/0363ebd1-c8c5-4b40-9333-f8523ca4b0ac> [Assessed on 12 February 2022].
- Sarawak State Secretary Office. (2022b). *Sarawak Principal Exports 2017 - Q2 2022 (April)* [Dataset]. Available at: [Assessed on 12/07/2022].
- Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A Window into Third-Generation Sequencing. *Human Molecular Genetics*, 19(R2), R227-R240.
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., & Ragg, T. (2006). The RIN: An RNA Integrity Number for Assigning Integrity Values to RNA Measurements. *BioMed Central Molecular Biology*, 7(1), 3.
- Sekula, B., Ruszkowski, M., & Dauter, Z. (2020). S-Adenosylmethionine Synthases in Plants: Structural Characterization of Type I and II Isoenzymes from *Arabidopsis*

- thaliana* and *Medicago truncatula*. *International Journal of Biological Macromolecules*, 151, 554-565.
- Shanmugam, K. S. (2016). Sulphur Nutrition of Coconut. *Indian Coconut Journal* (October 2016), 24-28.
- Sharma, P., & Tejinder, S. (2014). Extraction of Starch from Hulled and Hull-Less Barley with Papain and Aqueous Sodium Hydroxide. *Journal of Food Science and Technology*, 51(12), 3870-3877.
- Sim, S. F., Khan, A. J., Mohamed, M., & Mohamed, A. M. D. (2015). The Relationship between Peat Soil Characteristics and the Growth of Sago Palm (*Metroxylon sagu*). *Sago Palm* 13(2), 9-16.
- Singhal, R. S., Kennedy, J. F., Gopalakrishnan, S. M., Kaczmarek, A., Knill, C. J., & Akmar, P. F. (2008). Industrial Production, Processing, and Utilization of Sago Palm-Derived Products. *Carbohydrate Polymers*, 72(1), 1-20.
- Sirko, A., Anna Wawrzyńska, Rodríguez, M. C., & Sęktas, P. (2015). The Family of LSU-Like Proteins. *Frontiers in Plant Science*, 5, 774.
- Smith, C. (2022). *Guide to Choosing a Sequencing Technique*. [Online] (March 10, 2022) Available at: <https://www.biocompare.com/Editorial-Articles/584105-Guide-to-Choosing-a-Sequencing-Technique/> [Assessed on 19 December 2022].
- Šramková, Z., Gregová, E., & Šturdík, E. (2009). Chemical Composition and Nutritional Quality of Wheat Grain. *Acta Chimica Slovaca*, 2(1), 115-138.

- Srivastava, S., Upadhyay, M. K., Srivastava, A. K., Abdelrahman, M., Suprasanna, P., & Tran, L.-S. P. (2018). Cellular and Subcellular Phosphate Transport Machinery in Plants. *International Journal of Molecular Sciences*, 19(7), 1914.
- Suraninpong, P., & Nuanlaong, S. (2022). Comparative Transcriptome Profiling and Molecular Marker Development for Oil Palm Fruit Color. *Scientific Reports*, 12(1), 15507.
- Suza, W., Lee, D., Becraft, P., & Hanneman, M. (2021). Regulation of Gene Expression. In W. Suza & D. Lee (Eds.), *Genetics, Agriculture, and Biotechnology*, (pp. 86-97). Iowa State University Digital Press.
- Sze, H., Padmanaban, S., Cellier, F., Honys, D., Cheng, N. H., Bock, K. W., Conéjéro, G., Li, X. Y., Twell, D., Ward, J. M., & Hirschi, K. D. (2004). Expression Patterns of a Novel AtCHX Gene Family Highlight Potential Roles in Osmotic Adjustment and K⁺ Homeostasis in Pollen Development. *Plant Physiology*, 136(1), 2532–2547.
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Research*, 47(D1), D607-D613.
- Takahashi, H. (2019). Sulfate Transport Systems in Plants: Functional Diversity and Molecular Mechanisms Underlying Regulatory Coordination. *Journal of Experimental Botany*, 70(16), 4075–4087.

- Takanashi, K., Shitan, N., & Yazaki, K. (2014). The Multidrug and Toxic Compound Extrusion (MATE) Family in Plants. *Plant Biotechnology* 31, 417–430.
- Talke, I. N., Hanikenne, M., & Krämer, U. (2006). Zinc-Dependent Global Transcriptional Control, Transcriptional Deregulation, and Higher Gene Copy Number for Genes in Metal Homeostasis of the Hyperaccumulator *Arabidopsis halleri*. *Plant Physiology*, 142(1), 148–167.
- Tanaka, M., Wallace, I. S., Takano, J., Roberts, D. M., & Fujiwara, T. (2008). NIP6;1 is a Boric Acid Channel for Preferential Transport of Boron to Growing Shoot Tissues in *Arabidopsis*. *The Plant Cell*, 20(10), 2860-2875.
- Tian, Y. H., Zhang, C. H., Jiang, D., Dai, T. B., Jing, Q., & Cao, W. X. (2008). Separation and Purification Process of Wheat Flour Starch Granule. China. Patent No. CN101418048B.
- Toyoda, Y. (2018). Life and Livelihood in Sago-Growing Areas. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 31-42). Springer Nature.
- Tripathi, S., Hoang, Q. T. N., Han, Y. J., & Kim, J. I. (2019). Regulation of Photomorphogenic Development by Plant Phytochromes. *International Journal of Molecular Sciences*, 20(24), 6165.
- Tsutsui, T., Yamaji, N., & Ma, J. F. (2011). Identification of a Cis-Acting Element of ART1, a C2H2-type Zinc-Finger Transcription Factor for Aluminum Tolerance in Rice. *Plant Physiology*, 156(2), 925-931.
- Udachan, I., Sahoo, A. K., & Hend, G. M. (2012). Extraction and Characterization of Sorghum (*Sorghum bicolor* L. Moench) Starch. *International Food Research Journal*, 19, 315-319.

- Untergasser, A. (2008). *RNA Miniprep using CTAB*. [Online] Available at: http://www.molbi.de/protocols/miniprep_rna_ctab_v1_0.htm [Assessed on 13 January 2016].
- USDA, & NRCS. (2022). *Metroxylon sagu Rottb. - Sago Palm in The PLANTS Database* [Database]. Available at: <https://plants.sc.egov.usda.gov/home/plantProfile?symbol=MESA7> [Assessed on 13 February 2022].
- Versaw, W. K., & Harrison, M. J. (2002). A Chloroplast Phosphate Transporter, PHT2;1, Influences Allocation of Phosphate within the Plant and Phosphate-Starvation Responses. *The Plant Cell*, 14, 1751–1766.
- Wang, Y., Zhang, W. Z., Song, L. F., Zou, J. J., Su, Z., & Wu, W. H. (2008). Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in *Arabidopsis*. *Plant Physiology*, 148(3), 1201–1211.
- Wardis, G. (2018). Feasibility of Small-Scale Sago Industries in the Maluku Islands, Indonesia. In H. Ehara, Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm*, (pp. 109-121). Springer Nature.
- Wetlands International. (2010). *A Quick Scan of Peatlands in Malaysia*. Petaling Jaya, Malaysia: Wetlands International-Malaysia.
- Wong, M. K. (2004). *Comparative Study of Microorganisms in Soil of Trunking and Non-Trunking Sago Palms Plantation* [Master's Dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak, Malaysia. Available at: <https://ir.unimas.my/id/eprint/14829/>

- Xiao, Y., Yang, Y., Cao, H., Fan, H., Ma, Z., Lei, X., Mason, A. S., Xia, Z., & Huang, X. (2012). Efficient Isolation of High Quality RNA from Tropical Palms for RNA-seq Analysis. *Plant Omics*, 5(6), 584-589.
- Yan, W.-J., Pendi, F. H., & Hussain, M. H. (2022). Improved CTAB Method for RNA Extraction of Thick Waxy Leaf Tissues from Sago Palm (*Metroxylon sagu* Rottb.). *Chemical and Biological Technologies in Agriculture*, 9(1), 63.
- Yan, W. J. (2016). *Metabolites Comparison between Trunking and Non-Trunking Sago Palm (Metroxylon Sagu)* [Master's Dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository. Sarawak, Malaysia. Available at: <https://ir.unimas.my/id/eprint/30670/>
- Yao, K., Bacchetto, R. G., Lockhart, K. M., Friesen, L. J., Potts, D. A., Covello, P. S., & Taylor, D. C. (2003). Expression of the Arabidopsis ADS1 Gene in *Brassica juncea* Results in a Decreased Level of Total Saturated Fatty Acids. *Plant Biotechnology Journal*, 1(3), 221-229.
- Yeh, S. H., Lin, C. S., Wu, F. H., & Wang, A. Y. (2011). Analysis of the Expression of BohLOL1, which Encodes an LSD1-like Zinc Finger Protein in *Bambusa oldhamii*. *Planta*, 234, 1179–1189.
- Yockteng, R., Almeida, A. M., Yee, S., Andre, T., Hill, C., & Specht, C. D. (2013). A Method for Extracting High-Quality RNA from Diverse Plants for Next-Generation Sequencing and Gene Expression Analyses. *Applications in Plant Sciences*, 1(12), apps.1300070.
- Yong, C. M. R., Sobeng, Y., Zaini, F., & Busri, N. (2018). Suitability of Peat Swamp Areas for Commercial Production of Sago Palms: The Sarawak Experience. In H. Ehara,

- Y. Toyoda, & D. V. Johnson (Eds.), *Sago Palm : Multiple Contributions to Food Security and Sustainable Livelihoods*, (pp. 91-108). Springer Nature.
- Yong, D., Ning, L., Virilouvet, L., Riethoven, J.-J., Fromm, M., & Avramova, Z. (2013). Four Distinct Types of Dehydration Stress Memory Genes in *Arabidopsis thaliana*. *BioMed Central Plant Biology*, 13, 229.
- Yoshinari, A., & Takano, J. (2017). Insights into the Mechanisms Underlying Boron Homeostasis in Plants. *Frontiers in Plant Science*, 8, 1951.
- Zhang, Z., Wu, S., Stenoien, D. L., & Paša-Tolić, L. (2014). High-Throughput Proteomics. *Annual Review of Analytical Chemistry*, 7(1), 427-454.
- Zhang, Z., Xian, N., Eckhoff, S. R., & Feng, H. (2005). Sonication Enhanced Cornstarch Separation. *Starch*, 57(6), 240–245.
- Zhao, F. J., Tausz, M., & Kok, L. J. D. (2008). Role of Sulfur for Plant Production in Agricultural and Natural Ecosystems. In R. Hell, C. Dahl, D. Knaff, & T. Leustek (Eds.), *Sulfur Metabolism in Phototrophic Organisms*, (pp. 417–435). Springer, Dordrecht.
- Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H., & Li, G. (2017). The *Arabidopsis thaliana* Nuclear Factor Y Transcription Factors. *Frontiers in Plant Science*, 7, 2045.
- Zhu, F. (2015). Isolation, Composition, Structure, Properties, Modifications, and Uses of Yam Starch. *Comprehensive Reviews in Food Science and Food Safety*, 14(4), 357-386.

APPENDICES

Appendix A: CTAB Miniprep RNA Extraction Steps

Steps	Procedures
1	Oak Ridge High-Speed centrifuge tubes were labelled.
2	7 g sample was powdered with a pre-cooled mortar and pestle with liquid nitrogen.
3	1 g of powdered sample was transferred into two labelled tubes each.
4	15 ml of ice-cooled basic CTAB with 1% β -mercaptoethanol was transferred into each tube.
5	15 ml of ice-cooled chloroform was added into each tube and mixed.
6	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
7	13 ml of the top aqueous layer from tube were transferred into new labelled Oak Ridge High-Speed centrifuge tubes.
8	15 ml of ice-cooled chloroform was added into each tube and mixed.
9	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
10	11 ml of the top aqueous layer from tubes were transferred into new labelled Oak Ridge High-Speed centrifuge tubes.
11	11 ml of pre-cooled 2-isopropanol were added into tubes and incubated in room temperature for 10 minutes.
12	Tubes were centrifuged at 13,200 rpm, 4°C for 15 minutes.
13	Supernatant from tubes were discarded and the pellets were transferred into 1.5 ml microcentrifuge tubes respectively by 1 ml of ice-cooled 70% ethanol.
14	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
15	Supernatant from tubes were discarded and 1 ml of ice-cooled 70% ethanol was added.

Appendix A continued

16	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
17	Supernatant from tubes were discarded and the dried pellets were dissolved in 99 µl of Tris-HCl buffer. 1 µl of RQ1 RNase-Free DNase were added and mixed.
18	Tubes were incubated for 30 minutes at room temperature.
19	10 µl of RQ1 RNase-Free DNase Stop solution were added and mixed.
20	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
21	Supernatant were transferred into a new tube before further analysis.

Appendix B: PCI RNA Extraction Steps

Steps	Procedures
1	Oak Ridge High-Speed centrifuge tubes were labelled.
2	7 g sample was powdered with a pre-cooled mortar and pestle with liquid nitrogen.
3	1 g of powdered sample was transferred into two labelled tubes each.
4	15 ml of ice-cooled extraction buffer was transferred into each tube and mixed.
5	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
6	14 ml of top aqueous layer from tubes were transferred into new labelled Oak Ridge High-Speed centrifuge tubes.
7	15 ml of ice-cooled PCI buffer were added into tubes and mixed.
8	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
9	13 ml of the top aqueous layer from tubes were transferred into a new labelled Oak Ridge High-Speed centrifuge tube.
10	15 ml of ice-cooled PCI buffer were added into tubes and mixed.
11	Tubes were centrifuged at 13,200 rpm, 4°C for 10 minutes.

Appendix B continued

12	12 ml of the top aqueous layer from tubes were transferred into a new labelled Oak Ridge High-Speed centrifuge tube.
13	15 ml of ice-cooled PCI were added into tubes and mixed.
14	Tubes were centrifuged at 13,200 rpm, 4°C for 15 minutes.
15	11 ml of the top aqueous layer from tubes were transferred into a new labelled Oak Ridge High-Speed centrifuge tubes and 15 ml ice-cooled chloroform were added and mixed.
16	Tubes were centrifuged at 13,200 rpm, 4°C for 15 minutes.
17	9 ml of the top aqueous layer from tubes were transferred into a new labelled Oak Ridge High-Speed centrifuge tubes and 15 ml of chloroform were added and mixed.
18	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
19	7 ml of the top aqueous layer from tubes were transferred into a new labelled Oak Ridge High-Speed centrifuge tubes and 2.33 ml of 8M LiCl were added.
20	Tubes were incubated overnight at 4°C.
21	Tubes were centrifuged at 13,200 rpm, 4°C for 30 minutes.
22	Supernatant from tubes were discarded and the pellets were transferred into 1.5 ml microcentrifuge tubes respectively by 1 ml of ice-cooled 70% ethanol.
23	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
24	Supernatant from tubes were discarded and the pellets were dissolved with 400 µl dH ₂ O followed by addition of 1 ml 100% EtOH and 40 µl 3 M NaOAc.
25	Tubes were incubated at -80°C for 20 minutes.
26	Tubes were centrifuged at 13,200 rpm, 4°C for 30 minutes.
27	Supernatant from tubes were discarded and 1 ml of 70% ethanol were added.
28	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.

Appendix B continued

29	Supernatant from tubes were discarded and 1 ml of 70% ethanol were added.
30	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
31	Supernatant from tubes were discarded and the dried pellets were dissolved in 99 µl of Tris-HCl buffer. 1 µl of RQ1 RNase-Free DNase were added and mixed.
32	Tubes were incubated for 30 minutes at room temperature.
33	10 µl of RQ1 RNase-Free DNase Stop solution were added and mixed.
34	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
35	Supernatant were transferred into a new tube before further analysis.

Appendix C: MRIP RNA Extraction Steps

Steps	Procedures
1	Oak Ridge High-Speed centrifuge tubes were labelled.
2	7 g sample was powdered with a pre-cooled mortar and pestle with liquid nitrogen.
3	1 g of powdered sample was transferred into two labelled tubes each.
4	15 ml of MRIP buffer was transferred into the tubes and mixed.
5	Tubes were incubated at room temperature for 5 minutes.
6	5 ml ice-cooled chloroform were added into tubes and mixed.
7	Tubes were incubated for 5 minutes at room temperature.
8	Tubes were centrifuged at 13,200 rpm, 4°C for 10 minutes.
9	10 ml of the top aqueous layer from tubes were transferred into a new labelled Oak Ridge High-Speed centrifuge tube.
10	10 ml of pre-cooled isopropanol were added into tubes and mixed.
11	Tubes were incubated on ice for 10 minutes.

Appendix C continued

12	Tubes were centrifuged at 13,200 rpm, 4°C for 15 minutes.
13	Supernatant from tubes were discarded and the pellets were transferred into 1.5 ml microcentrifuge tubes with 1 ml of ice-cooled 70% ethanol.
14	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
15	Supernatant from tubes were discarded and 1 ml of ice-cooled 70% ethanol was added.
16	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
17	Supernatant from tubes were discarded and the dried pellets were dissolved in 99 µl of Tris-HCl buffer. 1 µl of RQ1 RNase-Free DNase were added and mixed.
18	Tubes were incubated for 30 minutes at room temperature.
19	10 µl of RQ1 RNase-Free DNase Stop solution were added and mixed.
20	Tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
21	Supernatant were transferred into a new tube before further analysis.

Appendix D: RNA Extraction Protocol RNA Extraction Steps

Steps	Procedures
1	2.2 g sample was powdered with a pre-cooled mortar and pestle with liquid nitrogen.
2	1 g powdered sample was transferred into two labelled tubes each.
3	15 ml of CTAB with 1% beta-mercaptoethanol was transferred into the tubes, mixed and incubated 5 minutes.
4	15 ml of chloroform was added into the tubes, mixed and incubated 5 minutes.
5	The tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.

Appendix D continued

6	13 ml of the top aqueous layer from tubes were transferred into new labelled Oak Ridge High-Speed centrifuge tubes.
7	15 ml of chloroform was added into tubes, mixed and incubated for 5 minutes.
8	The tubes were centrifuged at 13,200 rpm, 4°C for 5 minutes.
9	10 ml of the top aqueous layer from tubes were pooled into new labelled Oak Ridge High-Speed centrifuge tubes.
10	20 ml of ice-cooled isopropanol were added into the pooled tube, mixed and incubated in room temperature for 10 minutes.
11	The tube was centrifuged at 13,200 rpm, 4°C for 15 minutes.
12	Supernatant from tube was discarded and 1 ml of 70% ethanol was added and the pellets were transferred into labelled new 1.5 ml microcentrifuge tubes.
13	The tube was centrifuged at 13,200 rpm, 4°C for 2 minutes
14	Supernatant from tube was discarded and 1 ml of 70% ethanol was added.
15	The tube was centrifuged at 13,200 rpm, 4°C for 2 minutes
16	Supernatant from tube was discarded and the dried pellet was dissolved in 1 ml of Tris-HCl buffer.
17	2 µl of RQ1 RNase-Free DNase were added and mixed.
18	The tubes were incubated for 30 minutes at room temperature.
19	100 µl of 20 mM EDTA solution was added and mixed.
20	The tube was centrifuged at 13,200 rpm, 4°C for 15 minutes.
21	540 µl top aqueous layer from the tubes were transferred into four new labelled tube each. 540 µl of ice-cooled isopropanol was added into the tubes, mixed and incubated in room temperature for 10 minutes.
22	The tubes were centrifuged at 13,200 rpm, 4°C for 15 minutes.
23	Supernatant from tubes were discarded and 1 ml of 70% ethanol was added.

Appendix D continued

24	The tubes were centrifuged at 13,200 rpm, 4°C for 2 minutes
25	Supernatant from tubes were discarded and 1 ml of 70% ethanol was added.
26	The tubes were centrifuged at 13,200 rpm, 4°C for 2 minutes
27	Supernatant from tubes were discarded and 1 ml of ice-cooled absolute ethanol was added into the dried pellets.
28	Samples were stored at -80°C before use.

Appendix E: RNA Attribute Test Prior Shipment and Qualification by Agilent 2100 Bioanalyzer

Sample Batch	Sample Name	In-house RNA Attribute Test			RNA Attribute Test via Agilent 2100 Bioanalyzer & NanoDrop				
		Total Mass (µg)	OD260 /280	OD260 /230	Total Mass (µg)	OD260 /280	OD260 /230	RIN	25S/18S
Batch 1 (Extracted on January 25-28, 2019)	T1	63.89	1.92	2.22	32.40	1.92	2.22	4.6	1.5
	T2	73.13	1.92	2.17	18.09	1.92	2.17	3.4	1.3
	T3	42.19	1.94	2.20	44.28	1.94	2.20	2.6	2.4
	T4	76.02	1.96	2.20	45.36	1.96	2.20	6.5	1.6
	T5*	75.78	1.96	2.22	37.53	1.96	2.22	7.5	1.7
	N1	42.01	2.14	2.36	30.51	2.14	2.36	5.9	1.7
	N2	75.99	2.00	2.29	44.55	2.00	2.29	7.7	1.9
	N3	55.71	2.07	2.30	52.65	2.07	2.30	5.3	1.6
	N4	108.64	1.88	2.08	42.39	1.88	2.08	7.6	1.8
	N5	77.46	2.03	2.33	26.46	2.03	2.33	7.3	1.9

Appendix E continued

Batch 2 (Extracted on May 11-15, 2019)	S01T01	36.065	1.88	2.18	24.03	1.88	2.18	5.1	1.4
	S02T02	32.558	1.75	1.98	29.70	1.75	1.98	3.4	0.7
	S03T03	32.580	1.84	2.01	24.84	1.84	2.01	3.0	1.4
	S04T04	26.404	1.88	2.12	23.49	1.88	2.12	7.3	1.9
	S05T05	10.992	1.91	2.13	16.20	1.91	2.13	4.7	1.7
	S06NT01	44.183	1.80	2.07	63.99	1.80	2.07	7.1	1.7
	S07NT02	59.466	1.71	1.90	75.33	1.71	1.90	7.4	1.8
	S08NT03	41.502	1.47	1.44	45.90	1.47	1.44	7.4	1.7
	S09NT04	26.760	1.84	2.02	39.15	1.84	2.02	7.8	2.0
	S10NT05	19.448	1.74	1.90	23.49	1.74	1.90	7.6	1.8
Batch 3 (Extracted on May 18-21, 2019)	S01T01A*	28.399	1.91	2.61	13.50	1.91	2.61	6.6	1.4
	S02T02A	44.293	1.93	2.66	20.79	1.93	2.66	3.5	1.2
	S03T03A	22.622	1.93	2.68	9.72	1.93	2.68	2.4	0
	S04T04A	26.808	1.90	2.51	18.90	1.90	2.51	7.1	1.8
	S05T05A	15.159	1.91	2.57	12.15	1.91	2.57	4.1	1.9

Appendix E continued

	S06NT01A	67.105	1.92	2.40	56.97	1.92	2.40	6.8	2.3
	S07NT02A*	52.728	1.99	2.62	38.34	1.99	2.62	7.7	2.0
	S08NT03A	41.896	1.98	2.45	22.68	1.98	2.45	5.5	1.9
	S09NT04A	48.511	1.99	2.50	34.02	1.99	2.50	8.0	2.0
	S10NT05A	51.902	1.99	2.47	34.29	1.99	2.47	7.2	2.0
Batch 4 (Extracted on August 20-24, 2019)	S01T01B	62.84	1.94	2.53	30.6	1.94	2.53	2.8	2.6
	S02T02B	71.53	1.93	2.43	35.02	1.93	2.43	4.8	2.4
	S03T03B	36.61	1.91	2.28	12.75	1.91	2.28	2.7	2.1
	S04T04B*	32.56	1.88	2.23	18.7	1.88	2.23	7.4	2
	S05T05B	34.73	1.93	2.48	15.13	1.93	2.48	6.6	1.9
	S06NT01B	59.43	2.03	2.55	31.45	2.03	2.55	3.4	2.7
	S07NT02B	61.12	2.01	2.45	35.87	2.01	2.45	7.5	1.9
	S08NT03B*	43.61	1.99	2.38	31.45	1.99	2.38	7.4	1.8
	S09NT04B*	50.17	1.99	2.41	28.73	1.99	2.41	8.3	1.8
	S10NT05B	56.25	1.97	2.47	40.46	1.97	2.47	6.8	2.1

Note: Asterisk (*) sample are chosen for subsequent RNA-seq analysis

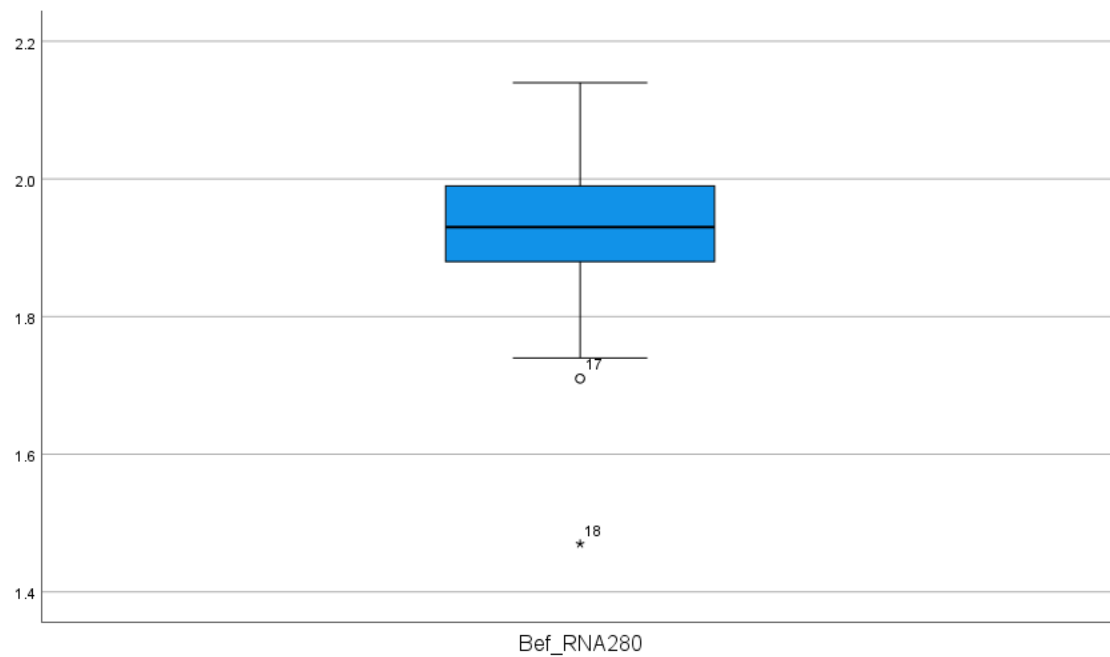
Appendix F: Normality Test of the RNA parameters Before and After Shipment to BGI

Tests of Normality						
	Kolmogorov-Smirnov^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Bef_Mass	.101	40	.200*	.971	40	.399
Bef_RNA280	.196	40	<.001	.864	40	<.001
Bef_RNA230	.086	40	.200*	.942	40	.039
Aft_Weight	.071	40	.200*	.953	40	.095
Aft_RNA280	.196	40	<.001	.864	40	<.001
Aft_RNA230	.086	40	.200*	.942	40	.039
*. This is a lower bound of the true significance.						
a. Lilliefors Significance Correction						

Keys: Bef_Mass: Total RNA weight before shipment; Aft_Mass: Total RNA weight after shipment; Bef_RNA280: 260/280 ratio before shipment; Aft_RNA280: 260/280 ratio after shipment; Bef_RNA230: 260/230 ratio before shipment; Aft_RNA230: 260/230 ratio after shipment.

Note: Since degree of freedom (df) of the tested data is < 50, Shapiro-Wilk result will be used for normality test. Only total RNA weight dataset are normally distributed while the 260/280 ratio and 260/230 ratio are not normally distributed.

Appendix G: Boxplot of the RNA280 Dataset Before Shipment to BGI



Key: Bef_RNA280: 260/280 ratio before shipment; ○: Outlier; ★: Extreme Outlier

Note: Data number 17 is outlier of the 260/280 ratio and data number 18 is extreme outlier.

The extreme outlier data were removed for paired sample t-test.

Appendix H: Normality Test of the RNA parameters Before and After Shipment to BGI
After Removal of Extreme Outliers

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Bef_Mass	.091	39	.200*	.972	39	.429
Bef_RNA280	.134	39	.074	.958	39	.153
Bef_RNA230	.092	39	.200*	.970	39	.382
Aft_Weight	.079	39	.200*	.946	39	.062
Aft_RNA280	.134	39	.074	.958	39	.153
Aft_RNA230	.092	39	.200*	.970	39	.382
*. This is a lower bound of the true significance.						
a. Lilliefors Significance Correction						

Keys: Bef_Mass: Total RNA weight before shipment; Aft_Mass: Total RNA weight after shipment; Bef_RNA280: 260/280 ratio before shipment; Aft_RNA280: 260/280 ratio after shipment; Bef_RNA230: 260/230 ratio before shipment; Aft_RNA230: 260/230 ratio after shipment.

Note: Since degree of freedom (df) of the tested data is < 50, Shapiro-Wilk result will be used for normality test. After removal of the extreme outliers, the Total RNA weight, 260/280 ratio, and 260/230 ratio datasets are normally distributed.

Appendix I: Paired Sample T-Test of the RNA parameters Before and After Shipment to BGI's Statistics

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Bef_Mass	48.3987	39	20.82633	3.33488
	Aft_Mass	31.6913	39	14.48404	2.31930
Pair 2	Bef_RNA280	1.9282 ^a	39	.08611	.01379
	Aft_RNA280	1.9282 ^a	39	.08611	.01379
Pair 3	Bef_RNA230	2.3169 ^a	39	.21373	.03422
	Aft_RNA230	2.3169 ^a	39	.21373	.03422
a. The correlation and t cannot be computed because the standard error of the difference is 0.					

Keys: Bef_Mass: Total RNA weight before shipment; Aft_Mass: Total RNA weight after shipment; Bef_RNA280: 260/280 ratio before shipment; Aft_RNA280: 260/280 ratio after shipment; Bef_RNA230: 260/230 ratio before shipment; Aft_RNA230: 260/230 ratio after shipment.

Note: Pair 2: Before and After RNA280 ratio; and Pair 3: Before and After RNA230 standard error of the difference is 0, shows that the samples purity are the same before and after shipment. Only Pair 1: Comparison between the Mass before and after shipment, using t-test was performed.

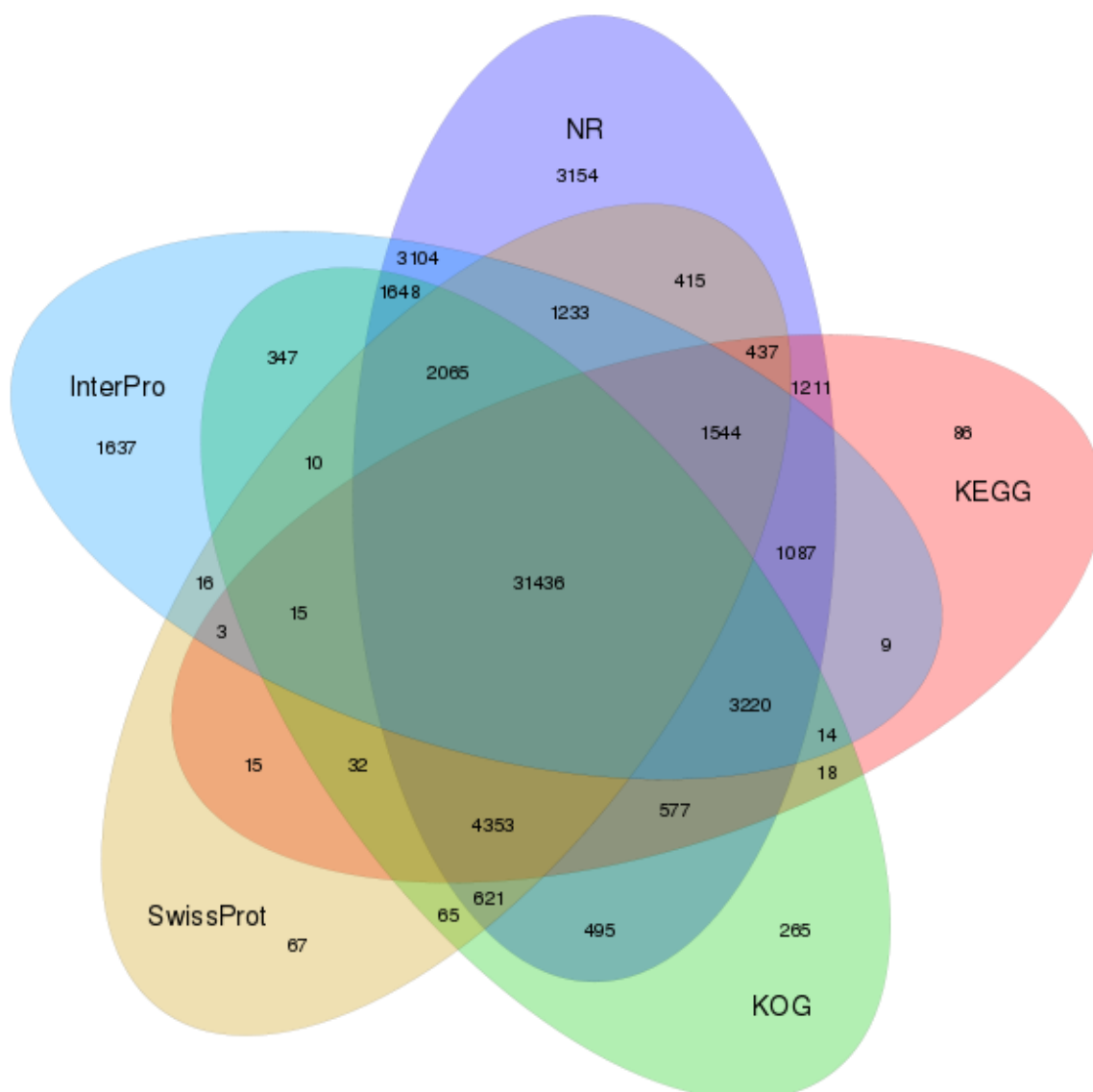
Appendix J: Paired Sample T-Test of the RNA parameters Before and After Shipment to BGI Results

Paired Samples Test										
		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				One-Sided p	Two-Sided p
					Lower	Upper				
Pair 1	Bef_Mass - Aft_Mass	16.71	18.30	2.93	10.77	22.64	5.70	38	<.001	<.001

Keys: Bef_Mass: Total RNA weight before shipment; Aft_Mass: Total RNA weight after shipment

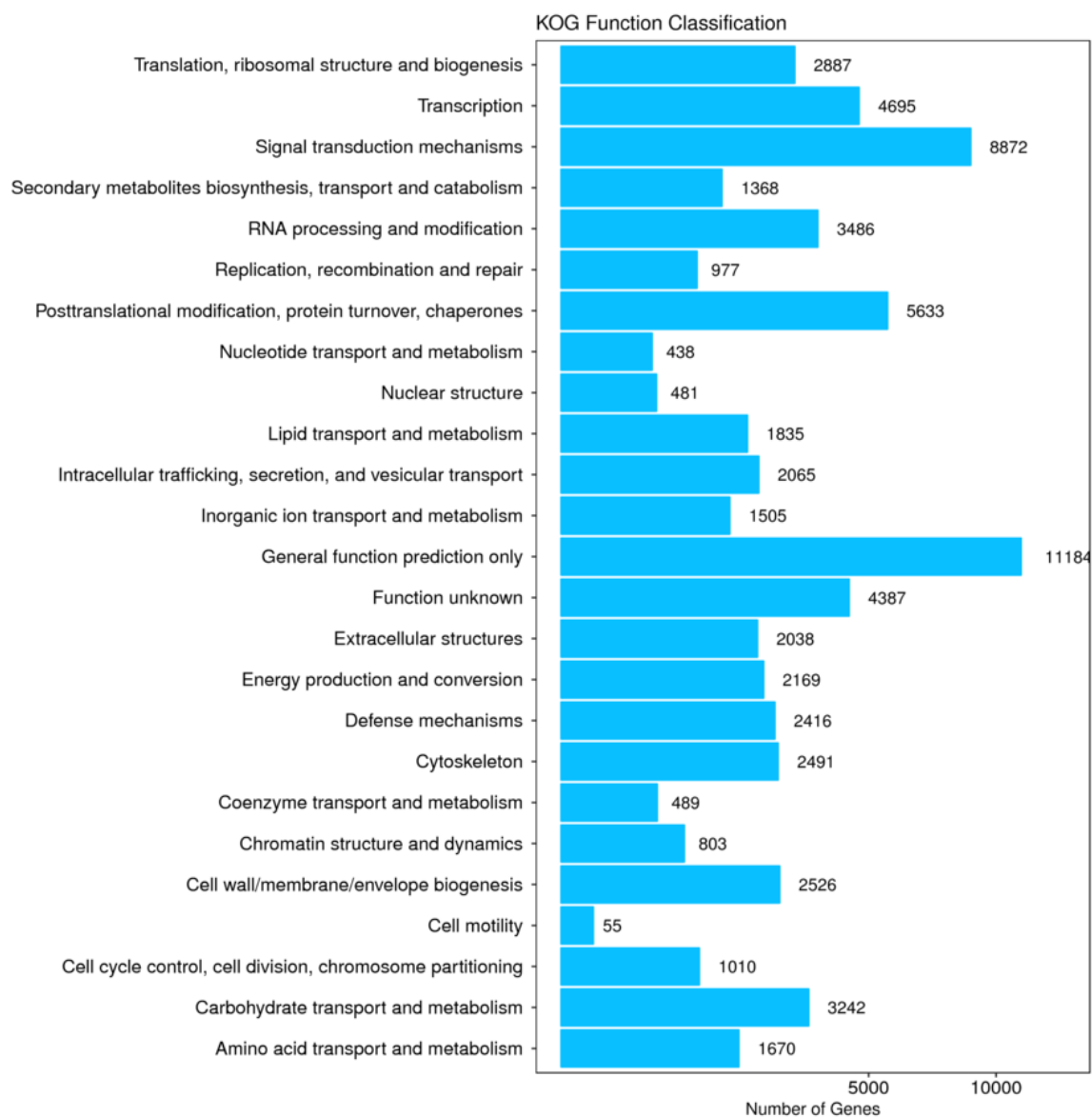
Note: Pair 1: One sided significance value (p-value) is less than 0.001 (p-value < 0.05), hence the Total RNA weight before and after shipment are significantly different.

Appendix K: Venn Diagram of Annotation Result between NR, KOG, KEGG, SwissProt and InterPro Databases of the Unigenes

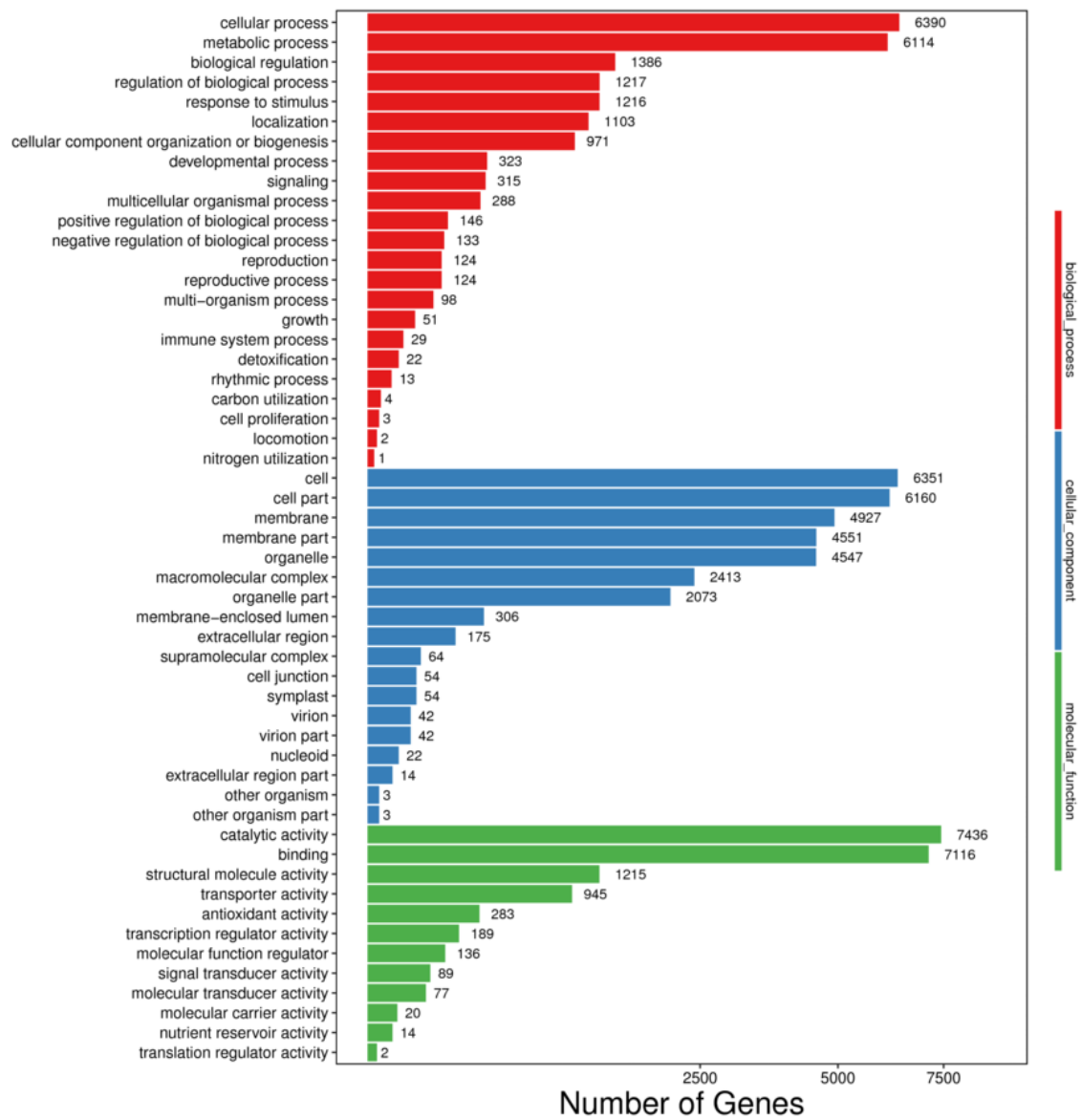


Note: The Venn diagram shows that there is no one database which able to annotated all the Unigenes and the 5 databases complimenting each other to maximise the annotation of the Unigenes.

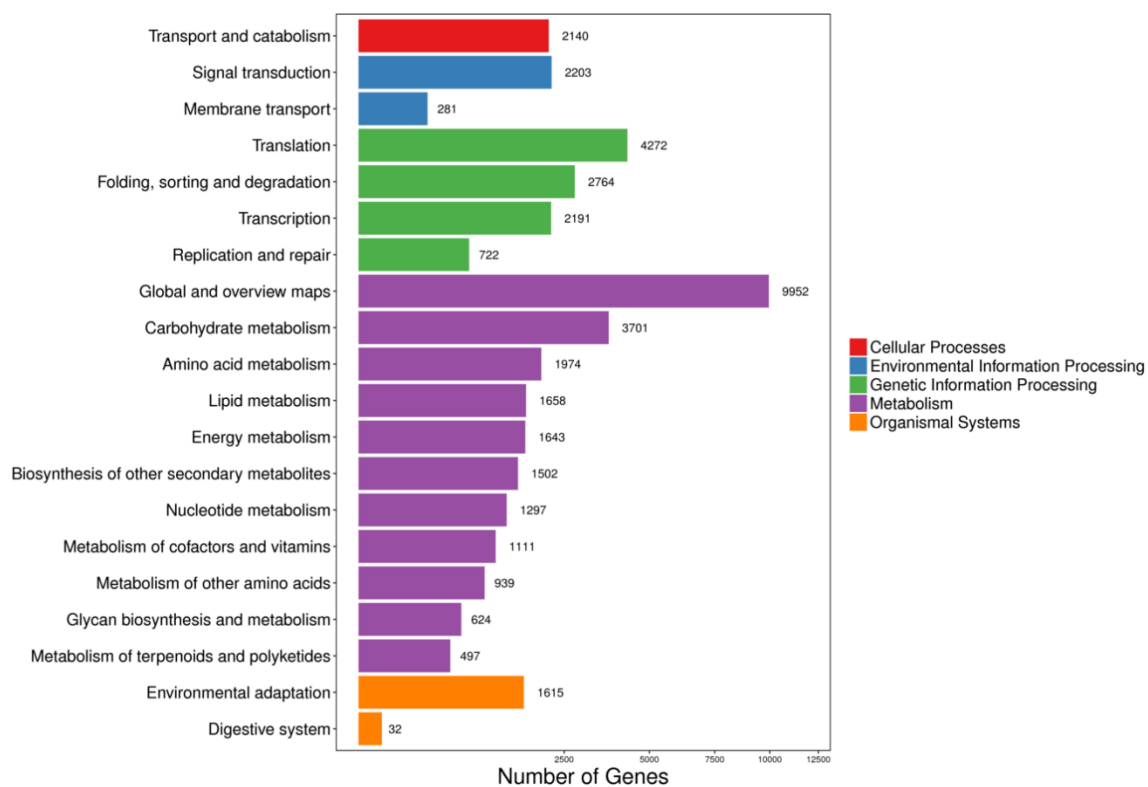
Appendix L: Functional Distribution of KOG Annotated Genes from Sago Palm Leaf



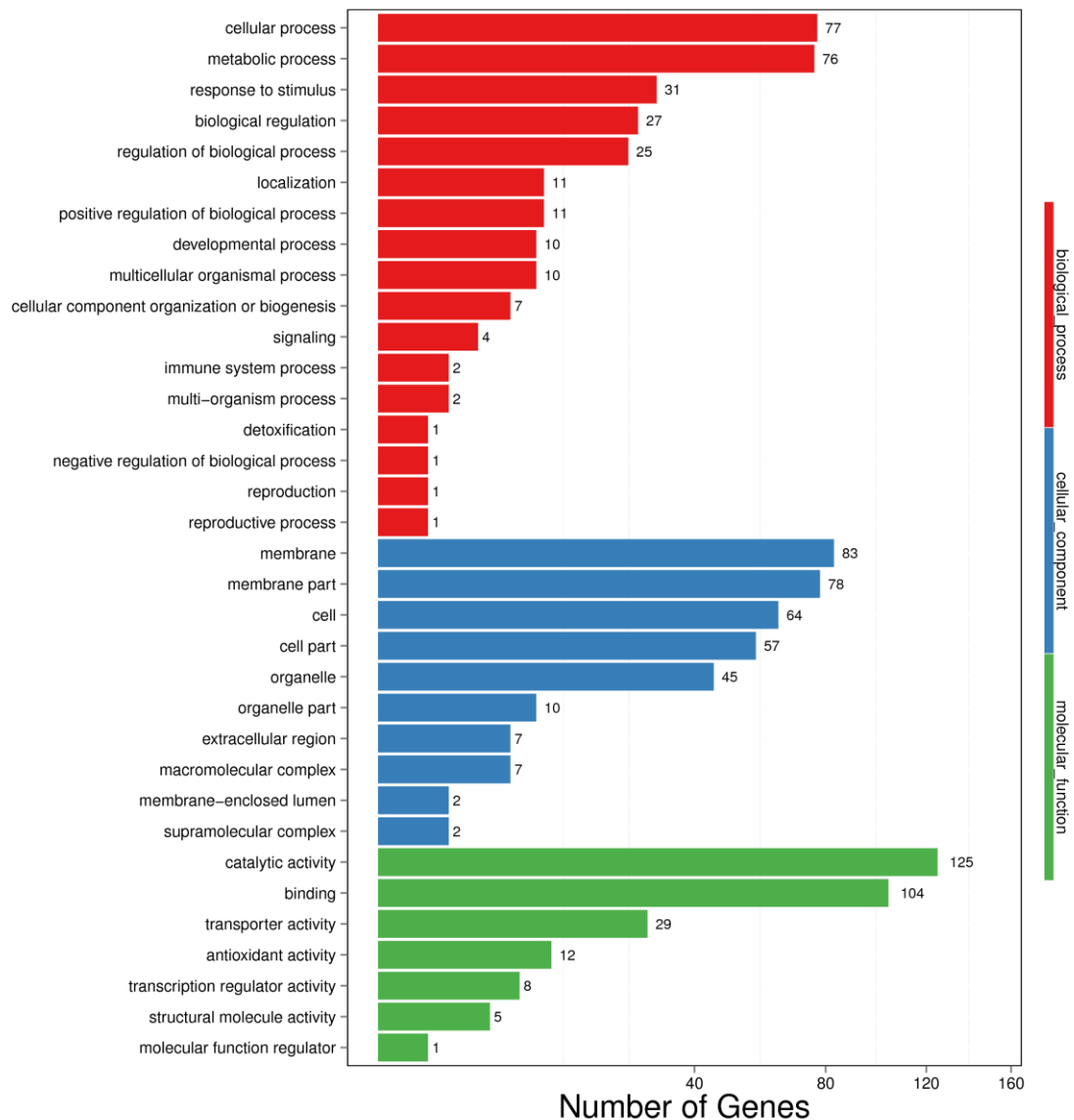
Appendix M: Functional Distribution of GO Annotated Genes in Sago Palm Leaf



Appendix N: Functional Distribution of KEGG Annotated Genes in Sago Palm Leaf

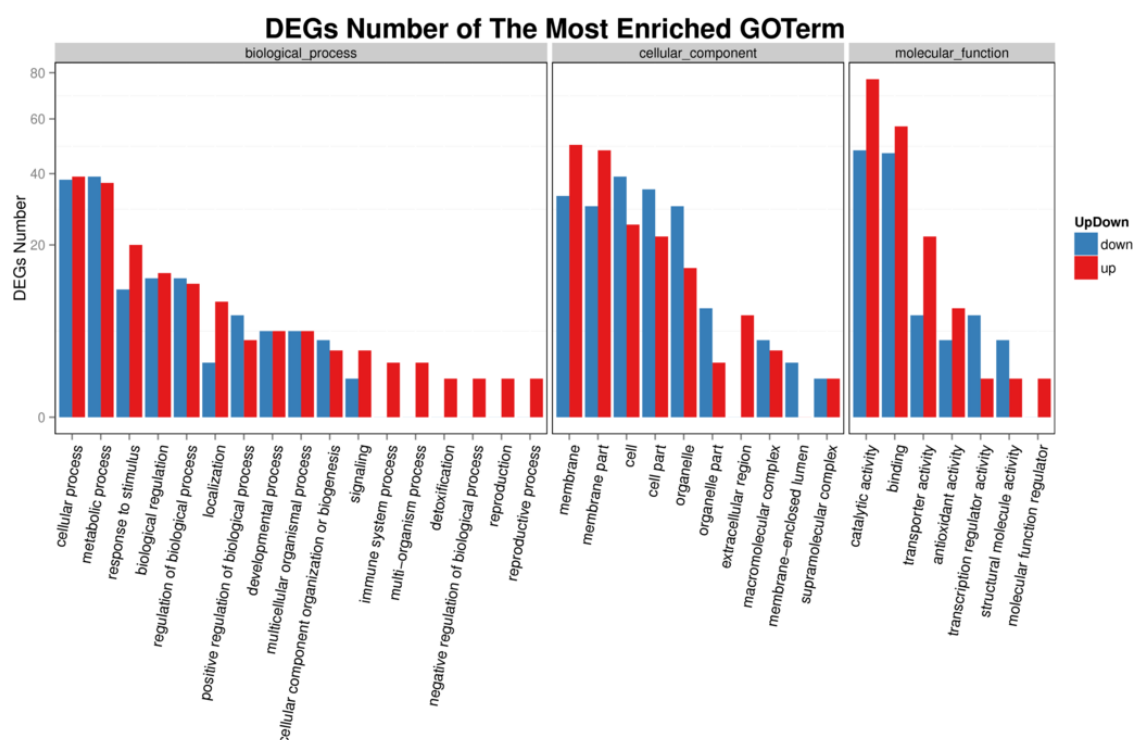


Appendix O: GO Classification of DEGs



Note: There are a total of 297 DEGs involves in biological processes, DEGs involves in cellular component and 284 DEGs involves in molecular function which involved in growth and development in plant.

Appendix P: Upregulated and Downregulated DEGs in the most Enriched GO-Term of the Non-Trunking Sago Palm Leaf



Note: Unigenes involved in immune system process, multi- organism process, detoxification, negative regulation of biological process, reproduction, reproductive process, extracellular region and molecular function regulator were upregulated while unigenes involved in membrane-enclosed lumen were downregulated.

Appendix Q: GO Enrichment Analysis of the DEGs for Biological Processes

Gene Ontology Term	Cluster frequency	Genome frequency of use	Corrected P-value
positive regulation of photomorphogenesis	5 of 107	7 of 8273	2.12E-06
positive regulation of response to stimulus	7 of 107	23 of 8273	3.22E-06
positive regulation of post-embryonic development	5 of 107	10 of 8273	2.47E-05
positive regulation of developmental process	5 of 107	11 of 8273	4.49E-05
positive regulation of multicellular organismal process	5 of 107	11 of 8273	4.49E-05
regulation of photomorphogenesis	5 of 107	12 of 8273	7.62E-05
regulation of response to red or far red light	5 of 107	12 of 8273	7.62E-05
positive regulation of biosynthetic process	9 of 107	91 of 8273	0.00073
positive regulation of biological process	11 of 107	146 of 8273	0.0008
regulation of response to stimulus	8 of 107	73 of 8273	0.00126
positive regulation of cellular process	10 of 107	128 of 8273	0.00171
positive regulation of transcription, DNA-templated	6 of 107	38 of 8273	0.00251
positive regulation of RNA biosynthetic process	6 of 107	38 of 8273	0.00251
positive regulation of nucleic acid-templated transcription	6 of 107	38 of 8273	0.00251
positive regulation of RNA metabolic process	6 of 107	40 of 8273	0.00342
positive regulation of metabolic process	9 of 107	111 of 8273	0.00382

Appendix Q continued

positive regulation of macromolecule biosynthetic process	8 of 107	86 of 8273	0.00434
positive regulation of nucleobase-containing compound metabolic process	6 of 107	42 of 8273	0.00458
positive regulation of cellular biosynthetic process	8 of 107	88 of 8273	0.00515
positive regulation of gene expression	8 of 107	90 of 8273	0.00608
hydrogen peroxide metabolic process	8 of 107	99 of 8273	0.01221
hydrogen peroxide catabolic process	8 of 107	99 of 8273	0.01221
positive regulation of nitrogen compound metabolic process	8 of 107	99 of 8273	0.01221
drug catabolic process	10 of 107	161 of 8273	0.01284
positive regulation of macromolecule metabolic process	8 of 107	101 of 8273	0.01411
positive regulation of cellular metabolic process	8 of 107	104 of 8273	0.01742
regulation of developmental process	7 of 107	83 of 8273	0.0289
reactive oxygen species metabolic process	8 of 107	115 of 8273	0.03558
regulation of response to extracellular stimulus	2 of 107	2 of 8273	0.05204
positive regulation of response to extracellular stimulus	2 of 107	2 of 8273	0.05204
regulation of response to nutrient levels	2 of 107	2 of 8273	0.05204
positive regulation of response to nutrient levels	2 of 107	2 of 8273	0.05204

Appendix Q continued

cellular response to cold	2 of 107	2 of 8273	0.05204
positive regulation of cellular response to phosphate starvation	2 of 107	2 of 8273	0.05204
cofactor catabolic process	8 of 107	131 of 8273	0.08776
regulation of post-embryonic development	5 of 107	50 of 8273	0.13801
phosphate ion transmembrane transport	2 of 107	3 of 8273	0.1548
regulation of cellular response to stress	3 of 107	13 of 8273	0.1719
positive regulation of cell communication	2 of 107	4 of 8273	0.30698
tryptophan catabolic process to kynurenine	2 of 107	4 of 8273	0.30698
kynurenine metabolic process	2 of 107	4 of 8273	0.30698
regulation of multicellular organismal development	5 of 107	61 of 8273	0.3463
regulation of multicellular organismal process	5 of 107	64 of 8273	0.43044
response to oxidative stress	9 of 107	210 of 8273	0.48269
positive regulation of response to external stimulus	2 of 107	5 of 8273	0.50732
inorganic anion transmembrane transport	2 of 107	6 of 8273	0.75456
tryptophan catabolic process	2 of 107	7 of 8273	1
amine catabolic process	2 of 107	7 of 8273	1
cellular biogenic amine catabolic process	2 of 107	7 of 8273	1
indole-containing compound catabolic process	2 of 107	7 of 8273	1
indolalkylamine catabolic process	2 of 107	7 of 8273	1
cellular catabolic process	16 of 107	590 of 8273	1

Appendix Q continued

cofactor metabolic process	16 of 107	602 of 8273	1
tryptophan metabolic process	2 of 107	9 of 8273	1
indolalkylamine metabolic process	2 of 107	9 of 8273	1
hormone transport	2 of 107	9 of 8273	1
auxin polar transport	2 of 107	9 of 8273	1
auxin transport	2 of 107	9 of 8273	1
regulation of cellular process	24 of 107	1095 of 8273	1
regulation of response to stress	3 of 107	29 of 8273	1
cellular response to starvation	3 of 107	30 of 8273	1
cellular ketone metabolic process	4 of 107	59 of 8273	1
response to starvation	3 of 107	31 of 8273	1
response to stimulus	25 of 107	1178 of 8273	1
regulation of response to external stimulus	2 of 107	11 of 8273	1
indole-containing compound metabolic process	2 of 107	11 of 8273	1
cellular response to phosphate starvation	2 of 107	12 of 8273	1
regulation of biological process	25 of 107	1217 of 8273	1
cell redox homeostasis	8 of 107	237 of 8273	1
phosphate ion transport	2 of 107	13 of 8273	1
cellular response to nutrient levels	3 of 107	37 of 8273	1
benzene-containing compound metabolic process	2 of 107	14 of 8273	1
response to nutrient levels	3 of 107	39 of 8273	1
intein-mediated protein splicing	4 of 107	73 of 8273	1

Appendix Q continued

protein splicing	4 of 107	73 of 8273	1
glycerol ether metabolic process	3 of 107	40 of 8273	1
ether metabolic process	3 of 107	40 of 8273	1
biological regulation	27 of 107	1386 of 8273	1
supramolecular fiber organization	3 of 107	42 of 8273	1
drug metabolic process	16 of 107	699 of 8273	1
cellular response to extracellular stimulus	3 of 107	44 of 8273	1
cellular response to external stimulus	3 of 107	44 of 8273	1
response to stress	16 of 107	708 of 8273	1
cellular nitrogen compound catabolic process	4 of 107	81 of 8273	1
response to extracellular stimulus	3 of 107	46 of 8273	1
catabolic process	17 of 107	780 of 8273	1
aromatic amino acid family catabolic process	2 of 107	18 of 8273	1
protein processing	5 of 107	125 of 8273	1
vitamin biosynthetic process	4 of 107	85 of 8273	1
vitamin metabolic process	4 of 107	89 of 8273	1
aromatic amino acid family metabolic process	3 of 107	52 of 8273	1
protein maturation	5 of 107	134 of 8273	1
cellular homeostasis	8 of 107	285 of 8273	1
small molecule biosynthetic process	12 of 107	517 of 8273	1
coenzyme biosynthetic process	5 of 107	140 of 8273	1
cellular biogenic amine metabolic process	2 of 107	23 of 8273	1
cellular amine metabolic process	2 of 107	23 of 8273	1

Appendix Q continued

alpha-amino acid catabolic process	3 of 107	56 of 8273	1
homeostatic process	8 of 107	303 of 8273	1
fatty acid biosynthetic process	4 of 107	104 of 8273	1
inorganic anion transport	2 of 107	27 of 8273	1
regulation of biological quality	11 of 107	483 of 8273	1
regulation of biosynthetic process	10 of 107	430 of 8273	1
phyloquinone biosynthetic process	2 of 107	29 of 8273	1
phyloquinone metabolic process	2 of 107	29 of 8273	1
regulation of cell communication	2 of 107	30 of 8273	1
regulation of hormone levels	2 of 107	32 of 8273	1
fatty acid metabolic process	4 of 107	118 of 8273	1
monocarboxylic acid biosynthetic process	4 of 107	119 of 8273	1
organic acid catabolic process	4 of 107	120 of 8273	1
carboxylic acid catabolic process	4 of 107	120 of 8273	1
amine metabolic process	2 of 107	34 of 8273	1
cellular amino acid catabolic process	3 of 107	75 of 8273	1
serine family amino acid metabolic process	2 of 107	35 of 8273	1
response to external stimulus	4 of 107	124 of 8273	1
actin filament organization	2 of 107	36 of 8273	1
cell communication	8 of 107	347 of 8273	1
heterocycle catabolic process	3 of 107	79 of 8273	1
pyridine-containing compound biosynthetic process	2 of 107	37 of 8273	1

Appendix Q continued

actin filament-based process	2 of 107	38 of 8273	1
actin cytoskeleton organization	2 of 107	38 of 8273	1
anion transmembrane transport	2 of 107	39 of 8273	1
regulation of cellular macromolecule biosynthetic process	9 of 107	418 of 8273	1
regulation of macromolecule biosynthetic process	9 of 107	419 of 8273	1
cytoskeleton organization	3 of 107	83 of 8273	1
ketone biosynthetic process	2 of 107	40 of 8273	1
quinone metabolic process	2 of 107	40 of 8273	1
quinone biosynthetic process	2 of 107	40 of 8273	1
regulation of cellular biosynthetic process	9 of 107	427 of 8273	1
aromatic compound catabolic process	3 of 107	87 of 8273	1
regulation of primary metabolic process	10 of 107	493 of 8273	1
small molecule catabolic process	5 of 107	193 of 8273	1
cofactor biosynthetic process	5 of 107	193 of 8273	1
one-carbon metabolic process	2 of 107	43 of 8273	1
organic cyclic compound catabolic process	3 of 107	90 of 8273	1
organic hydroxy compound biosynthetic process	2 of 107	45 of 8273	1
small molecule metabolic process	28 of 107	1741 of 8273	1
coenzyme metabolic process	8 of 107	385 of 8273	1
positive regulation of cellular amide metabolic process	2 of 107	48 of 8273	1

Appendix Q continued

positive regulation of translation	2 of 107	48 of 8273	1
regulation of transcription, DNA-templated	7 of 107	326 of 8273	1
regulation of nucleic acid-templated transcription	7 of 107	326 of 8273	1
regulation of RNA biosynthetic process	7 of 107	326 of 8273	1
lipid biosynthetic process	6 of 107	272 of 8273	1
response to cold	2 of 107	51 of 8273	1
tissue development	2 of 107	54 of 8273	1
regulation of RNA metabolic process	7 of 107	343 of 8273	1
water-soluble vitamin biosynthetic process	2 of 107	55 of 8273	1
regulation of gene expression	9 of 107	474 of 8273	1
regulation of nitrogen compound metabolic process	9 of 107	474 of 8273	1
carbohydrate metabolic process	13 of 107	748 of 8273	1
phospholipid biosynthetic process	2 of 107	57 of 8273	1
cellular aldehyde metabolic process	4 of 107	167 of 8273	1
unsaturated fatty acid biosynthetic process	2 of 107	58 of 8273	1
unsaturated fatty acid metabolic process	2 of 107	58 of 8273	1
cellular lipid metabolic process	6 of 107	290 of 8273	1
water-soluble vitamin metabolic process	2 of 107	59 of 8273	1
shoot system development	2 of 107	59 of 8273	1
positive regulation of cellular protein metabolic process	2 of 107	60 of 8273	1

Appendix Q continued

regulation of nucleobase-containing compound metabolic process	7 of 107	360 of 8273	1
positive regulation of protein metabolic process	2 of 107	61 of 8273	1
anion transport	2 of 107	63 of 8273	1
regulation of cellular metabolic process	9 of 107	501 of 8273	1
regulation of metabolic process	10 of 107	575 of 8273	1
anatomical structure morphogenesis	2 of 107	66 of 8273	1
multicellular organismal process	5 of 107	247 of 8273	1
organic hydroxy compound metabolic process	3 of 107	124 of 8273	1
plant organ development	2 of 107	70 of 8273	1
cellular response to stimulus	10 of 107	589 of 8273	1
alcohol metabolic process	2 of 107	73 of 8273	1
regulation of macromolecule metabolic process	9 of 107	529 of 8273	1
phospholipid metabolic process	2 of 107	76 of 8273	1
pyridine-containing compound metabolic process	5 of 107	266 of 8273	1
system development	3 of 107	140 of 8273	1
organic acid biosynthetic process	6 of 107	345 of 8273	1
carboxylic acid biosynthetic process	6 of 107	345 of 8273	1
response to biotic stimulus	2 of 107	83 of 8273	1
monocarboxylic acid metabolic process	7 of 107	416 of 8273	1

Appendix Q continued

regulation of translation	2 of 107	84 of 8273	1
regulation of cellular amide metabolic process	2 of 107	85 of 8273	1
sulfur compound biosynthetic process	3 of 107	151 of 8273	1
multi-organism process	2 of 107	89 of 8273	1
posttranscriptional regulation of gene expression	2 of 107	91 of 8273	1
defense response	2 of 107	95 of 8273	1
response to temperature stimulus	2 of 107	98 of 8273	1
organophosphate biosynthetic process	5 of 107	307 of 8273	1
pentose-phosphate shunt	2 of 107	102 of 8273	1
glucose 6-phosphate metabolic process	2 of 107	102 of 8273	1
glyceraldehyde-3-phosphate metabolic process	2 of 107	103 of 8273	1
pyridine nucleotide metabolic process	4 of 107	250 of 8273	1
nicotinamide nucleotide metabolic process	4 of 107	250 of 8273	1
protein folding	4 of 107	252 of 8273	1
anatomical structure development	4 of 107	256 of 8273	1
protein complex biogenesis	2 of 107	112 of 8273	1
oxidoreduction coenzyme metabolic process	4 of 107	258 of 8273	1
NADP metabolic process	2 of 107	114 of 8273	1
regulation of cellular protein metabolic process	2 of 107	116 of 8273	1
sulfur compound metabolic process	4 of 107	264 of 8273	1

Appendix Q continued

developmental process	4 of 107	267 of 8273	1
protein complex subunit organization	2 of 107	123 of 8273	1
alpha-amino acid metabolic process	4 of 107	276 of 8273	1
regulation of protein metabolic process	2 of 107	125 of 8273	1
lipid metabolic process	6 of 107	430 of 8273	1
proteolysis	5 of 107	357 of 8273	1
carboxylic acid metabolic process	12 of 107	902 of 8273	1
pyruvate metabolic process	2 of 107	131 of 8273	1
signal transduction	4 of 107	293 of 8273	1
signaling	4 of 107	293 of 8273	1
oxoacid metabolic process	12 of 107	919 of 8273	1
multicellular organism development	3 of 107	222 of 8273	1
glucose metabolic process	2 of 107	142 of 8273	1
organic acid metabolic process	12 of 107	940 of 8273	1
organic substance catabolic process	8 of 107	629 of 8273	1
cellular amino acid metabolic process	5 of 107	397 of 8273	1
oxidation-reduction process	5 of 107	399 of 8273	1
cellular response to stress	3 of 107	247 of 8273	1
hexose metabolic process	2 of 107	167 of 8273	1
carbohydrate biosynthetic process	2 of 107	177 of 8273	1
monosaccharide metabolic process	2 of 107	179 of 8273	1
inorganic ion transmembrane transport	2 of 107	180 of 8273	1
response to abiotic stimulus	3 of 107	274 of 8273	1

Appendix Q continued

establishment of localization in cell	3 of 107	275 of 8273	1
protein localization	4 of 107	364 of 8273	1
aromatic compound biosynthetic process	9 of 107	789 of 8273	1
organophosphate metabolic process	8 of 107	706 of 8273	1
organelle organization	5 of 107	467 of 8273	1
cellular carbohydrate metabolic process	2 of 107	202 of 8273	1
macromolecule localization	4 of 107	391 of 8273	1
transcription, DNA-templated	4 of 107	392 of 8273	1
RNA biosynthetic process	4 of 107	392 of 8273	1
nucleic acid-templated transcription	4 of 107	392 of 8273	1
organic cyclic compound biosynthetic process	9 of 107	836 of 8273	1
ion transmembrane transport	2 of 107	219 of 8273	1
cellular amino acid biosynthetic process	2 of 107	221 of 8273	1
heterocycle biosynthetic process	8 of 107	771 of 8273	1
nucleotide biosynthetic process	2 of 107	225 of 8273	1
nucleoside phosphate biosynthetic process	2 of 107	225 of 8273	1
organonitrogen compound catabolic process	3 of 107	333 of 8273	1
phosphate-containing compound metabolic process	8 of 107	789 of 8273	1
cellular localization	3 of 107	338 of 8273	1
protein transport	3 of 107	340 of 8273	1
peptide transport	3 of 107	349 of 8273	1
amide transport	3 of 107	350 of 8273	1

Appendix Q continued

establishment of localization	11 of 107	1079 of 8273	1
establishment of protein localization	3 of 107	354 of 8273	1
nucleobase-containing compound biosynthetic process	6 of 107	648 of 8273	1
localization	11 of 107	1099 of 8273	1
transmembrane transport	3 of 107	372 of 8273	1
nitrogen compound transport	3 of 107	373 of 8273	1
intracellular transport	2 of 107	270 of 8273	1
phosphorus metabolic process	8 of 107	851 of 8273	1
nucleotide metabolic process	5 of 107	585 of 8273	1
nucleoside phosphate metabolic process	5 of 107	589 of 8273	1
transport	10 of 107	1063 of 8273	1
ion transport	2 of 107	298 of 8273	1
macromolecular complex subunit organization	2 of 107	299 of 8273	1
ribose phosphate metabolic process	4 of 107	520 of 8273	1
cellular component biogenesis	4 of 107	522 of 8273	1
carbohydrate derivative biosynthetic process	2 of 107	321 of 8273	1
metabolic process	70 of 107	5895 of 8273	1
organic substance transport	3 of 107	436 of 8273	1
cellular component organization	6 of 107	760 of 8273	1
cellular component assembly	2 of 107	342 of 8273	1
response to chemical	2 of 107	354 of 8273	1

Appendix Q continued

nucleobase-containing small molecule metabolic process	5 of 107	687 of 8273	1
ATP metabolic process	2 of 107	358 of 8273	1
cellular protein modification process	2 of 107	367 of 8273	1
protein modification process	2 of 107	367 of 8273	1
cellular component organization or biogenesis	7 of 107	904 of 8273	1
purine nucleoside monophosphate metabolic process	2 of 107	375 of 8273	1
purine ribonucleoside monophosphate metabolic process	2 of 107	375 of 8273	1
purine nucleoside triphosphate metabolic process	2 of 107	383 of 8273	1
purine ribonucleoside triphosphate metabolic process	2 of 107	383 of 8273	1
nucleoside monophosphate metabolic process	2 of 107	385 of 8273	1
ribonucleoside monophosphate metabolic process	2 of 107	385 of 8273	1
cellular aromatic compound metabolic process	17 of 107	1855 of 8273	1
ribonucleoside triphosphate metabolic process	2 of 107	392 of 8273	1
nucleoside triphosphate metabolic process	2 of 107	397 of 8273	1
macromolecule modification	3 of 107	523 of 8273	1
purine ribonucleotide metabolic process	2 of 107	409 of 8273	1

Appendix Q continued

purine nucleotide metabolic process	2 of 107	414 of 8273	1
ribonucleotide metabolic process	2 of 107	419 of 8273	1
carbohydrate derivative metabolic process	5 of 107	766 of 8273	1
purine-containing compound metabolic process	2 of 107	457 of 8273	1
generation of precursor metabolites and energy	2 of 107	461 of 8273	1
organic cyclic compound metabolic process	16 of 107	1896 of 8273	1
cellular process	71 of 107	6221 of 8273	1
heterocycle metabolic process	15 of 107	1821 of 8273	1
RNA metabolic process	5 of 107	838 of 8273	1
nucleic acid metabolic process	6 of 107	977 of 8273	1
cellular biosynthetic process	24 of 107	2757 of 8273	1
organic substance biosynthetic process	25 of 107	2861 of 8273	1
biosynthetic process	26 of 107	2994 of 8273	1
cellular metabolic process	50 of 107	4944 of 8273	1
nucleobase-containing compound metabolic process	11 of 107	1664 of 8273	1
organic substance metabolic process	52 of 107	5216 of 8273	1
gene expression	13 of 107	2041 of 8273	1
cellular nitrogen compound biosynthetic process	12 of 107	1991 of 8273	1
organonitrogen compound biosynthetic process	11 of 107	1894 of 8273	1

Appendix Q continued

protein metabolic process	11 of 107	1934 of 8273	1
organonitrogen compound metabolic process	23 of 107	3118 of 8273	1
cellular macromolecule biosynthetic process	9 of 107	1798 of 8273	1
macromolecule biosynthetic process	9 of 107	1813 of 8273	1
translation	4 of 107	1226 of 8273	1
peptide biosynthetic process	4 of 107	1226 of 8273	1
primary metabolic process	42 of 107	4776 of 8273	1
amide biosynthetic process	4 of 107	1240 of 8273	1
peptide metabolic process	4 of 107	1272 of 8273	1
cellular amide metabolic process	4 of 107	1321 of 8273	1
cellular nitrogen compound metabolic process	20 of 107	3063 of 8273	1
cellular protein metabolic process	7 of 107	1769 of 8273	1
macromolecule metabolic process	18 of 107	2950 of 8273	1
nitrogen compound metabolic process	29 of 107	4050 of 8273	1
cellular macromolecule metabolic process	14 of 107	2742 of 8273	1

Appendix R: GO Enrichment Analysis of the DEGs for Cellular Component

Gene Ontology Term	Cluster frequency	Genome frequency of use	Corrected P-value
intrinsic component of membrane	77 of 141	4257 of 9699	0.42384
integral component of membrane	76 of 141	4248 of 9699	0.63498
extracellular region	7 of 141	175 of 9699	0.91236
membrane part	78 of 141	4551 of 9699	1
membrane	83 of 141	4927 of 9699	1
intrinsic component of plasma membrane	2 of 141	36 of 9699	1
apoplast	2 of 141	49 of 9699	1
actin cytoskeleton	2 of 141	50 of 9699	1
plasma membrane part	2 of 141	51 of 9699	1
supramolecular complex	2 of 141	64 of 9699	1
supramolecular polymer	2 of 141	64 of 9699	1
supramolecular fiber	2 of 141	64 of 9699	1
polymeric cytoskeletal fiber	2 of 141	64 of 9699	1
nucleus	21 of 141	1266 of 9699	1
cytoskeleton	3 of 141	136 of 9699	1
mitochondrial matrix	2 of 141	80 of 9699	1
chloroplast	12 of 141	760 of 9699	1
cell wall	2 of 141	103 of 9699	1
external encapsulating structure	2 of 141	103 of 9699	1
plastid	12 of 141	796 of 9699	1
cytoskeletal part	2 of 141	115 of 9699	1

Appendix R continued

cell periphery	7 of 141	503 of 9699	1
chloroplast thylakoid membrane	2 of 141	153 of 9699	1
plastid thylakoid membrane	2 of 141	155 of 9699	1
chloroplast stroma	3 of 141	233 of 9699	1
plastid stroma	3 of 141	240 of 9699	1
plasma membrane	5 of 141	414 of 9699	1
chloroplast envelope	2 of 141	186 of 9699	1
thylakoid membrane	2 of 141	187 of 9699	1
chloroplast thylakoid	2 of 141	189 of 9699	1
plastid thylakoid	2 of 141	191 of 9699	1
plastid envelope	2 of 141	194 of 9699	1
photosynthetic membrane	2 of 141	200 of 9699	1
chloroplast part	5 of 141	455 of 9699	1
plastid part	5 of 141	461 of 9699	1
mitochondrion	8 of 141	704 of 9699	1
thylakoid part	2 of 141	213 of 9699	1
thylakoid	2 of 141	259 of 9699	1
membrane-enclosed lumen	2 of 141	306 of 9699	1
organelle lumen	2 of 141	306 of 9699	1
intracellular organelle lumen	2 of 141	306 of 9699	1
intracellular membrane-bounded organelle	39 of 141	3244 of 9699	1
membrane-bounded organelle	39 of 141	3287 of 9699	1
mitochondrial part	2 of 141	360 of 9699	1

Appendix R continued

organelle envelope	3 of 141	503 of 9699	1
envelope	3 of 141	503 of 9699	1
organelle subcompartment	2 of 141	396 of 9699	1
catalytic complex	2 of 141	406 of 9699	1
cytosol	2 of 141	611 of 9699	1
intracellular organelle	45 of 141	4545 of 9699	1
organelle	45 of 141	4547 of 9699	1
protein complex	2 of 141	805 of 9699	1
non-membrane-bounded organelle	7 of 141	1552 of 9699	1
intracellular non-membrane-bounded organelle	7 of 141	1552 of 9699	1
ribosome	3 of 141	1203 of 9699	1
organelle part	10 of 141	2073 of 9699	1
intracellular organelle part	10 of 141	2073 of 9699	1
cell	64 of 141	6351 of 9699	1
intracellular ribonucleoprotein complex	3 of 141	1378 of 9699	1
ribonucleoprotein complex	3 of 141	1378 of 9699	1
cytoplasmic part	22 of 141	3509 of 9699	1
cell part	57 of 141	6160 of 9699	1
intracellular part	48 of 141	5656 of 9699	1
intracellular	51 of 141	5890 of 9699	1
macromolecular complex	7 of 141	2413 of 9699	1
cytoplasm	25 of 141	4296 of 9699	1

Appendix S: GO Enrichment Analysis of the DEGs for Molecular Function

Gene Ontology Term	Cluster frequency	Genome frequency of use	Corrected P-value
secondary active transmembrane transporter activity	13 of 198	135 of 12242	3.91E-05
inorganic anion transmembrane transporter activity	7 of 198	38 of 12242	3.00E-04
inorganic anion exchanger activity	3 of 198	3 of 12242	6.00E-04
ammonium transmembrane transporter activity	4 of 198	12 of 12242	4.27E-03
heme binding	17 of 198	348 of 12242	7.19E-03
peroxidase activity	11 of 198	159 of 12242	7.76E-03
secondary active sulfate transmembrane transporter activity	4 of 198	15 of 12242	1.13E-02
sulfate transmembrane transporter activity	4 of 198	15 of 12242	0.01134
sulfur compound transmembrane transporter activity	4 of 198	15 of 12242	0.01134
tetrapyrrole binding	17 of 198	378 of 12242	0.01979
oxidoreductase activity, acting on peroxide as acceptor	11 of 198	185 of 12242	0.03002
oxidoreductase activity	53 of 198	2079 of 12242	0.04701
active transmembrane transporter activity	14 of 198	305 of 12242	0.06332
transporter activity	29 of 198	945 of 12242	0.08717
6-phosphogluconolactonase activity	2 of 198	3 of 12242	0.11124
arylformamidase activity	2 of 198	4 of 12242	0.22012

Appendix S continued

FAD binding	3 of 198	15 of 12242	0.23654
cofactor binding	31 of 198	1107 of 12242	0.25079
antioxidant activity	12 of 198	283 of 12242	0.31584
anion transmembrane transporter activity	7 of 198	118 of 12242	0.43597
phosphorylase activity	2 of 198	6 of 12242	0.53865
glycogen phosphorylase activity	2 of 198	6 of 12242	0.53865
linear malto-oligosaccharide phosphorylase activity	2 of 198	6 of 12242	0.53865
SHG alpha-glucan phosphorylase activity	2 of 198	6 of 12242	0.53865
DNA binding transcription factor activity	8 of 198	156 of 12242	0.55299
sequence-specific DNA binding	7 of 198	125 of 12242	0.6001
monooxygenase activity	7 of 198	138 of 12242	1
transcription regulator activity	8 of 198	189 of 12242	1
protein disulfide oxidoreductase activity	5 of 198	88 of 12242	1
transmembrane transporter activity	20 of 198	742 of 12242	1
protein heterodimerization activity	5 of 198	104 of 12242	1
inorganic phosphate transmembrane transporter activity	2 of 198	16 of 12242	1
transferase activity, transferring acyl groups other than amino-acyl groups	6 of 198	145 of 12242	1
antiporter activity	4 of 198	74 of 12242	1
drug transmembrane transporter activity	3 of 198	47 of 12242	1
disulfide oxidoreductase activity	5 of 198	121 of 12242	1

Appendix S continued

oxidoreductase activity, acting on a sulfur group of donors	6 of 198	175 of 12242	1
isomerase activity	10 of 198	361 of 12242	1
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	7 of 198	222 of 12242	1
ion transmembrane transporter activity	14 of 198	580 of 12242	1
transferase activity, transferring acyl groups	6 of 198	202 of 12242	1
serine-type endopeptidase activity	5 of 198	158 of 12242	1
metal ion binding	44 of 198	2289 of 12242	1
carboxylic ester hydrolase activity	2 of 198	37 of 12242	1
protein dimerization activity	6 of 198	211 of 12242	1
catalase activity	3 of 198	78 of 12242	1
cation binding	44 of 198	2316 of 12242	1
coenzyme binding	13 of 198	572 of 12242	1
peptidyl-prolyl cis-trans isomerase activity	4 of 198	132 of 12242	1
iron ion binding	6 of 198	232 of 12242	1
vitamin binding	4 of 198	135 of 12242	1
cis-trans isomerase activity	4 of 198	136 of 12242	1
carbon-oxygen lyase activity	4 of 198	137 of 12242	1
inorganic molecular entity transmembrane transporter activity	11 of 198	501 of 12242	1
DNA binding	15 of 198	721 of 12242	1
pyridoxal phosphate binding	3 of 198	101 of 12242	1

Appendix S continued

vitamin B6 binding	3 of 198	101 of 12242	1
calmodulin binding	2 of 198	57 of 12242	1
O-methyltransferase activity	2 of 198	57 of 12242	1
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	2 of 198	57 of 12242	1
stearoyl-CoA 9-desaturase activity	2 of 198	58 of 12242	1
acyl-CoA desaturase activity	2 of 198	58 of 12242	1
serine-type peptidase activity	5 of 198	207 of 12242	1
serine hydrolase activity	5 of 198	207 of 12242	1
endopeptidase activity	7 of 198	322 of 12242	1
catalytic activity	125 of 198	7436 of 12242	1
transferase activity, transferring hexosyl groups	4 of 198	177 of 12242	1
oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water	2 of 198	75 of 12242	1
peptidase activity	12 of 198	660 of 12242	1
transition metal ion binding	17 of 198	959 of 12242	1
flavin adenine dinucleotide binding	3 of 198	144 of 12242	1
zinc ion binding	11 of 198	619 of 12242	1
electron transfer activity	5 of 198	275 of 12242	1
carbohydrate binding	2 of 198	103 of 12242	1
lyase activity	6 of 198	351 of 12242	1

Appendix S continued

NAD binding	3 of 198	167 of 12242	1
hydrolase activity, hydrolyzing O-glycosyl compounds	3 of 198	170 of 12242	1
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	2 of 198	108 of 12242	1
transferase activity, transferring glycosyl groups	5 of 198	306 of 12242	1
translation initiation factor activity	2 of 198	115 of 12242	1
hydro-lyase activity	2 of 198	115 of 12242	1
transferase activity, transferring alkyl or aryl (other than methyl) groups	2 of 198	119 of 12242	1
cytoskeletal protein binding	2 of 198	121 of 12242	1
hydrolase activity, acting on glycosyl bonds	3 of 198	190 of 12242	1
magnesium ion binding	2 of 198	138 of 12242	1
cation transmembrane transporter activity	6 of 198	415 of 12242	1
protein serine/threonine kinase activity	5 of 198	349 of 12242	1
ATPase activity	4 of 198	284 of 12242	1
oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	2 of 198	148 of 12242	1
peptidase activity, acting on L-amino acid peptides	8 of 198	567 of 12242	1
oxidoreductase activity, acting on the aldehyde or oxo group of donors	2 of 198	168 of 12242	1
protein binding	11 of 198	811 of 12242	1

Appendix S continued

hydrolase activity	31 of 198	2221 of 12242	1
ATPase activity, coupled	2 of 198	212 of 12242	1
metallopeptidase activity	2 of 198	213 of 12242	1
translation factor activity, RNA binding	2 of 198	217 of 12242	1
purine nucleoside binding	3 of 198	320 of 12242	1
GTP binding	3 of 198	320 of 12242	1
guanyl nucleotide binding	3 of 198	320 of 12242	1
purine ribonucleoside binding	3 of 198	320 of 12242	1
guanyl ribonucleotide binding	3 of 198	320 of 12242	1
nucleoside binding	3 of 198	324 of 12242	1
ribonucleoside binding	3 of 198	324 of 12242	1
methyltransferase activity	2 of 198	241 of 12242	1
GTPase activity	2 of 198	245 of 12242	1
ion binding	64 of 198	4507 of 12242	1
transferase activity, transferring one-carbon groups	2 of 198	255 of 12242	1
catalytic activity, acting on a protein	22 of 198	1765 of 12242	1
calcium ion binding	2 of 198	266 of 12242	1
hydrolase activity, acting on ester bonds	4 of 198	451 of 12242	1
nucleoside-triphosphatase activity	6 of 198	619 of 12242	1
hydrogen ion transmembrane transporter activity	2 of 198	277 of 12242	1
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	2 of 198	279 of 12242	1

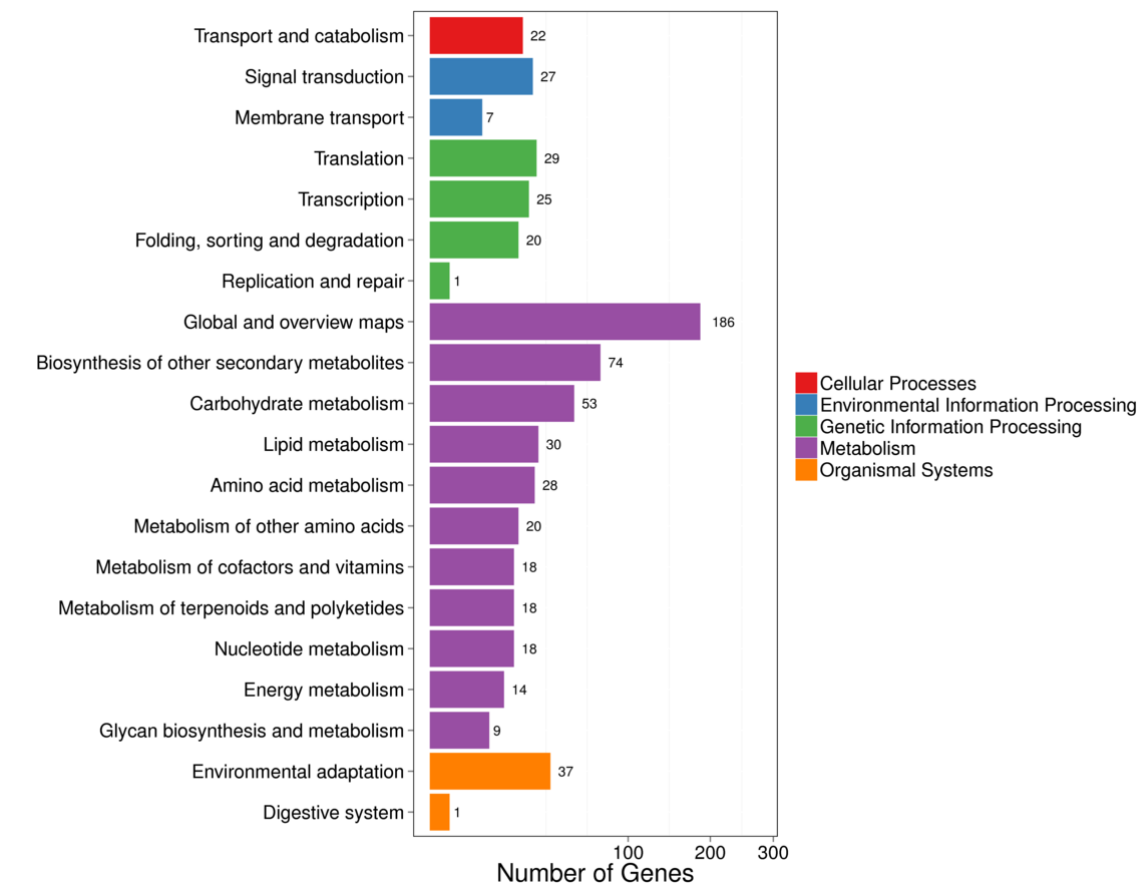
Appendix S continued

protein kinase activity	6 of 198	634 of 12242	1
binding	104 of 198	7116 of 12242	1
pyrophosphatase activity	6 of 198	651 of 12242	1
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	6 of 198	655 of 12242	1
hydrolase activity, acting on acid anhydrides	6 of 198	658 of 12242	1
monovalent inorganic cation transmembrane transporter activity	2 of 198	314 of 12242	1
oxidoreductase activity, acting on CH-OH group of donors	2 of 198	317 of 12242	1
organic cyclic compound binding	66 of 198	4851 of 12242	1
heterocyclic compound binding	66 of 198	4851 of 12242	1
phosphotransferase activity, alcohol group as acceptor	6 of 198	756 of 12242	1
inorganic cation transmembrane transporter activity	2 of 198	378 of 12242	1
nucleic acid binding	23 of 198	2097 of 12242	1
drug binding	18 of 198	1805 of 12242	1
kinase activity	6 of 198	899 of 12242	1
ATP binding	15 of 198	1680 of 12242	1
small molecule binding	27 of 198	2592 of 12242	1
transferase activity	22 of 198	2241 of 12242	1
anion binding	24 of 198	2426 of 12242	1
adenyl ribonucleotide binding	15 of 198	1747 of 12242	1

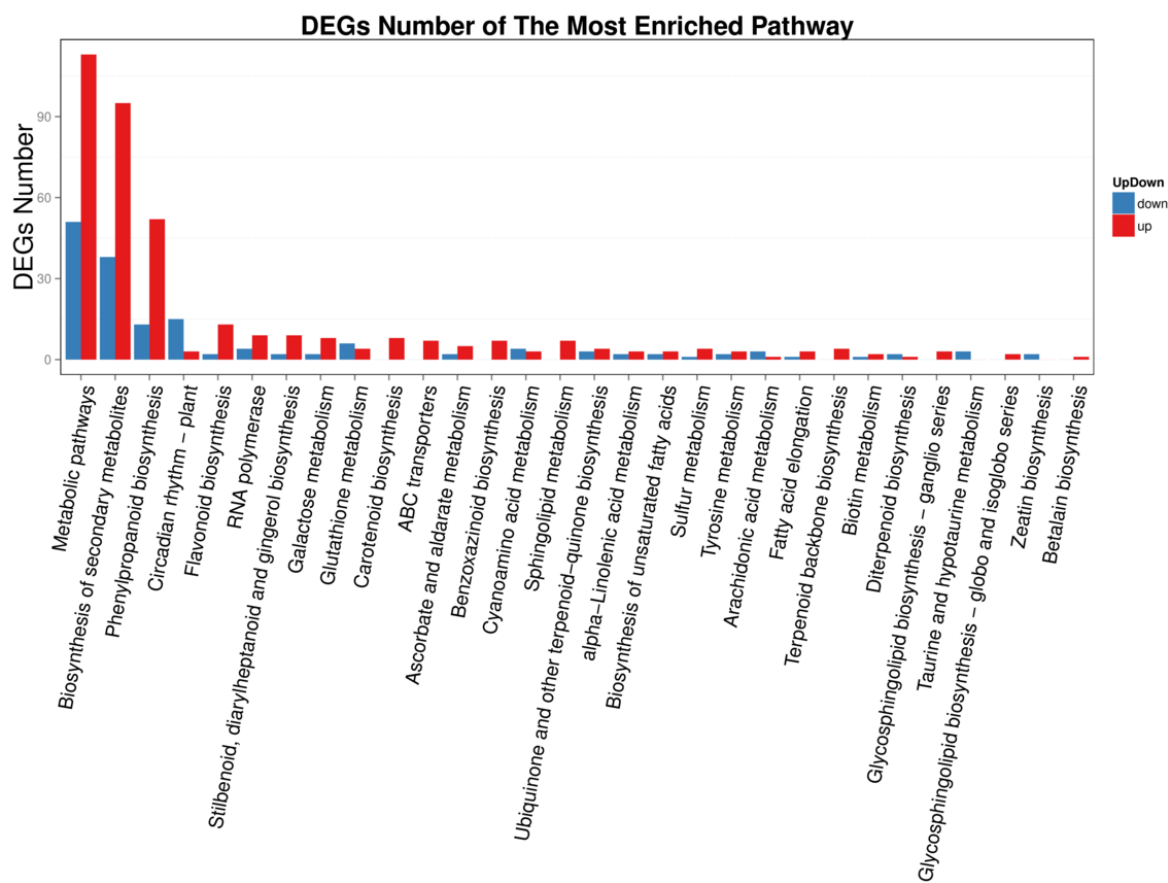
Appendix S continued

adenyl nucleotide binding	15 of 198	1751 of 12242	1
purine ribonucleoside triphosphate binding	18 of 198	1994 of 12242	1
RNA binding	8 of 198	1164 of 12242	1
nucleotide binding	24 of 198	2464 of 12242	1
nucleoside phosphate binding	24 of 198	2464 of 12242	1
purine ribonucleotide binding	18 of 198	2061 of 12242	1
purine nucleotide binding	18 of 198	2065 of 12242	1
transferase activity, transferring phosphorus-containing groups	6 of 198	1061 of 12242	1
ribonucleotide binding	18 of 198	2113 of 12242	1
carbohydrate derivative binding	18 of 198	2130 of 12242	1
structural molecule activity	5 of 198	1215 of 12242	1
structural constituent of ribosome	3 of 198	1125 of 12242	1

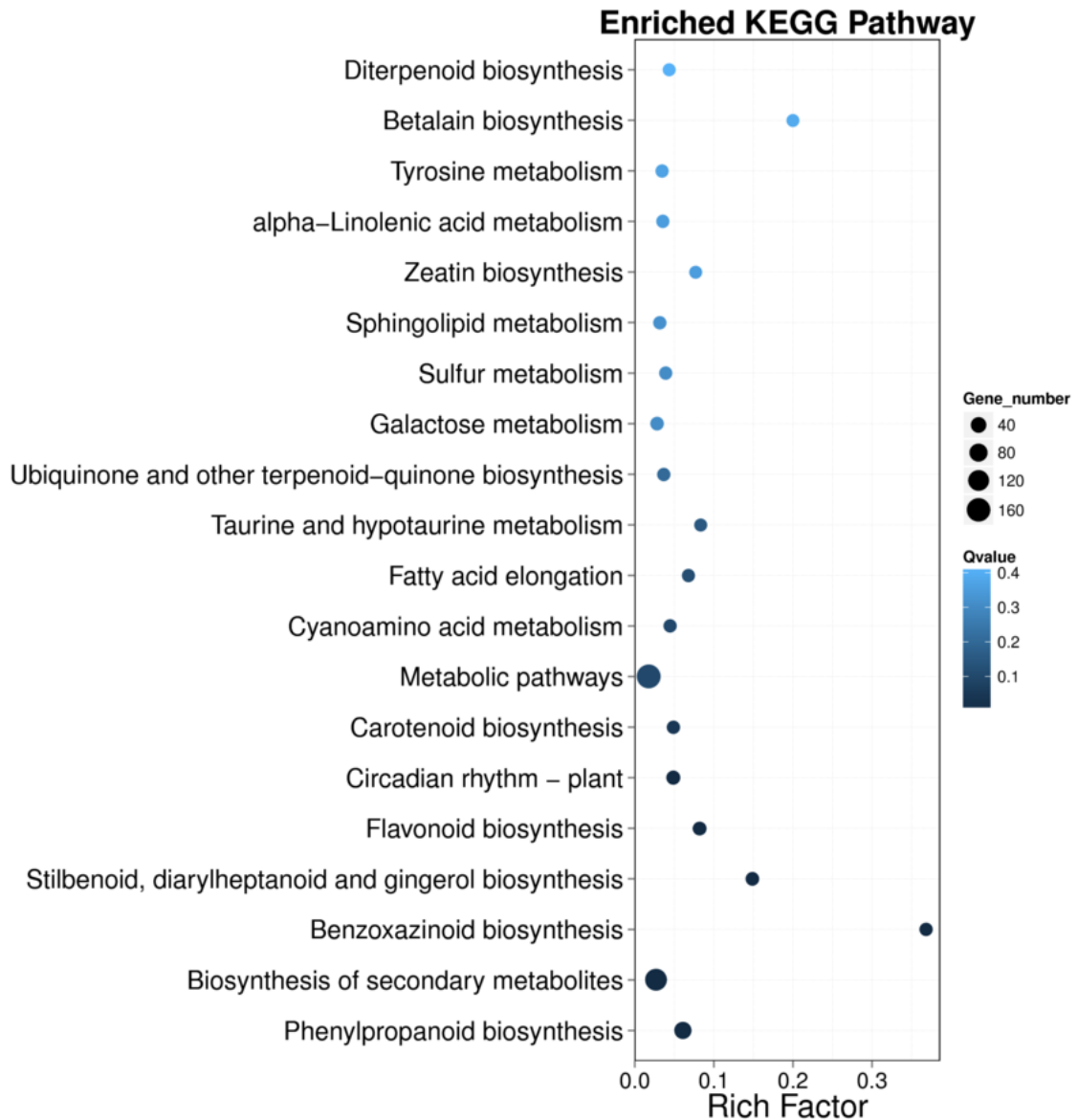
Appendix T: KEGG Pathway Classification of DEGs



Appendix U: Top 30 of the most Enriched KEGG Pathway of the DEGs



Appendix V: Top 20 KEGG pathway Functional Enrichment of DEGs



Keys: Q-value = $(1 - \text{adjusted P-value})$, the higher Q-value the more significant the enriched pathway. Rich factor = ratio of (DEGs involved in the pathway) over (total gene involved in the pathway), the higher the Rich Factor the more significant the enriched pathway.

Appendix W: KEGG Pathway Analysis of the DEGs

Gene Ontology Term	DEGs genes (648)	All genes (44057)	Q-value
Phenylpropanoid biosynthesis	65 (10.03%)	1071 (2.43%)	7.08E-20
Biosynthesis of secondary metabolites	133 (20.52%)	4965 (11.27%)	2.63E-10
Benzoxazinoid biosynthesis	7 (1.08%)	19 (0.04%)	2.16E-07
Stilbenoid, diarylheptanoid and gingerol biosynthesis	11 (1.7%)	74 (0.17%)	3.01E-07
Flavonoid biosynthesis	15 (2.31%)	183 (0.42%)	2.18E-06
Circadian rhythm - plant	18 (2.78%)	370 (0.84%)	2.06E-04
Carotenoid biosynthesis	8 (1.23%)	164 (0.37%)	4.60E-02
Metabolic pathways	164 (25.31%)	9397 (21.33%)	1.02E-01
Cyanoamino acid metabolism	7 (1.08%)	157 (0.36%)	1.02E-01
Fatty acid elongation	4 (0.62%)	59 (0.13%)	1.16E-01
Taurine and hypotaurine metabolism	3 (0.46%)	36 (0.08%)	1.49E-01
Ubiquinone and other terpenoid-quinone biosynthesis	7 (1.08%)	191 (0.43%)	2.04E-01
Galactose metabolism	10 (1.54%)	355 (0.81%)	3.06E-01
Sulfur metabolism	5 (0.77%)	128 (0.29%)	3.06E-01
Sphingolipid metabolism	7 (1.08%)	221 (0.5%)	3.19E-01
Zeatin biosynthesis	2 (0.31%)	26 (0.06%)	3.54E-01
alpha-Linolenic acid metabolism	5 (0.77%)	141 (0.32%)	3.54E-01
Tyrosine metabolism	5 (0.77%)	145 (0.33%)	3.68E-01
Betalain biosynthesis	1 (0.15%)	5 (0.01%)	3.91E-01
Diterpenoid biosynthesis	3 (0.46%)	69 (0.16%)	4.09E-01

Appendix W continued

Glutathione metabolism	10 (1.54%)	409 (0.93%)	4.09E-01
Ascorbate and aldarate metabolism	7 (1.08%)	264 (0.6%)	4.57E-01
Glycosphingolipid biosynthesis - globo and isoglobo series	2 (0.31%)	39 (0.09%)	4.91E-01
Arachidonic acid metabolism	4 (0.62%)	127 (0.29%)	4.91E-01
Biotin metabolism	3 (0.46%)	82 (0.19%)	4.91E-01
ABC transporters	7 (1.08%)	281 (0.64%)	4.91E-01
RNA polymerase	13 (2.01%)	623 (1.41%)	5.08E-01
Glycosphingolipid biosynthesis - ganglio series	3 (0.46%)	87 (0.2%)	5.08E-01
Terpenoid backbone biosynthesis	4 (0.62%)	142 (0.32%)	5.65E-01
Biosynthesis of unsaturated fatty acids	5 (0.77%)	198 (0.45%)	5.84E-01
Glycosaminoglycan degradation	3 (0.46%)	105 (0.24%)	6.76E-01
Phenylalanine metabolism	4 (0.62%)	163 (0.37%)	7.14E-01
Nicotinate and nicotinamide metabolism	3 (0.46%)	116 (0.26%)	7.68E-01
Vitamin B6 metabolism	2 (0.31%)	69 (0.16%)	8.06E-01
Selenocompound metabolism	3 (0.46%)	124 (0.28%)	8.06E-01
Starch and sucrose metabolism	13 (2.01%)	729 (1.65%)	8.06E-01
Flavone and flavonol biosynthesis	1 (0.15%)	25 (0.06%)	8.57E-01
Linoleic acid metabolism	2 (0.31%)	77 (0.17%)	8.57E-01
Purine metabolism	18 (2.78%)	1086 (2.46%)	8.83E-01
Other glycan degradation	4 (0.62%)	200 (0.45%)	8.83E-01
Isoquinoline alkaloid biosynthesis	2 (0.31%)	87 (0.2%)	9.30E-01
Cholesterol metabolism	1 (0.15%)	32 (0.07%)	9.35E-01

Appendix W continued

Arginine and proline metabolism	5 (0.77%)	288 (0.65%)	9.69E-01
Fatty acid degradation	4 (0.62%)	227 (0.52%)	9.69E-01
Phosphonate and phosphinate metabolism	1 (0.15%)	39 (0.09%)	9.69E-01
beta-Alanine metabolism	4 (0.62%)	231 (0.52%)	9.69E-01
Fatty acid biosynthesis	3 (0.46%)	166 (0.38%)	9.69E-01
Monobactam biosynthesis	1 (0.15%)	40 (0.09%)	9.69E-01
Riboflavin metabolism	2 (0.31%)	105 (0.24%)	9.73E-01
Plant-pathogen interaction	19 (2.93%)	1246 (2.83%)	9.75E-01
Glycolysis / Gluconeogenesis	11 (1.7%)	721 (1.64%)	1.00E+00
Fatty acid metabolism	5 (0.77%)	332 (0.75%)	1.00E+00
Brassinosteroid biosynthesis	1 (0.15%)	53 (0.12%)	1.00E+00
MAPK signaling pathway - plant	16 (2.47%)	1112 (2.52%)	1.00E+00
Phenylalanine, tyrosine and tryptophan biosynthesis	3 (0.46%)	202 (0.46%)	1.00E+00
Ether lipid metabolism	2 (0.31%)	133 (0.3%)	1.00E+00
N-Glycan biosynthesis	3 (0.46%)	209 (0.47%)	1.00E+00
Cysteine and methionine metabolism	6 (0.93%)	435 (0.99%)	1.00E+00
Peroxisome	7 (1.08%)	512 (1.16%)	1.00E+00
Glycerophospholipid metabolism	5 (0.77%)	367 (0.83%)	1.00E+00
Pyrimidine metabolism	14 (2.16%)	1018 (2.31%)	1.00E+00
Cutin, suberine and wax biosynthesis	1 (0.15%)	73 (0.17%)	1.00E+00
Carbon metabolism	19 (2.93%)	1402 (3.18%)	1.00E+00
RNA transport	20 (3.09%)	1493 (3.39%)	1.00E+00

Appendix W continued

Glyoxylate and dicarboxylate metabolism	6 (0.93%)	473 (1.07%)	1.00E+00
Biosynthesis of amino acids	16 (2.47%)	1207 (2.74%)	1.00E+00
Thiamine metabolism	1 (0.15%)	83 (0.19%)	1.00E+00
Glycerolipid metabolism	5 (0.77%)	411 (0.93%)	1.00E+00
Pentose phosphate pathway	5 (0.77%)	412 (0.94%)	1.00E+00
Tryptophan metabolism	4 (0.62%)	336 (0.76%)	1.00E+00
Carbon fixation in photosynthetic organisms	5 (0.77%)	424 (0.96%)	1.00E+00
Plant hormone signal transduction	13 (2.01%)	1086 (2.46%)	1.00E+00
Pantothenate and CoA biosynthesis	1 (0.15%)	117 (0.27%)	1.00E+00
Inositol phosphate metabolism	3 (0.46%)	305 (0.69%)	1.00E+00
RNA degradation	6 (0.93%)	557 (1.26%)	1.00E+00
Alanine, aspartate and glutamate metabolism	3 (0.46%)	306 (0.69%)	1.00E+00
Nitrogen metabolism	2 (0.31%)	218 (0.49%)	1.00E+00
Arginine biosynthesis	2 (0.31%)	228 (0.52%)	1.00E+00
Autophagy - other	2 (0.31%)	232 (0.53%)	1.00E+00
Folate biosynthesis	1 (0.15%)	137 (0.31%)	1.00E+00
Endocytosis	12 (1.85%)	1085 (2.46%)	1.00E+00
Histidine metabolism	1 (0.15%)	145 (0.33%)	1.00E+00
Phosphatidylinositol signaling system	3 (0.46%)	346 (0.79%)	1.00E+00
Basal transcription factors	2 (0.31%)	257 (0.58%)	1.00E+00
Pentose and glucuronate interconversions	3 (0.46%)	372 (0.84%)	1.00E+00
mRNA surveillance pathway	10 (1.54%)	1007 (2.29%)	1.00E+00

Appendix W continued

Citrate cycle (TCA cycle)	2 (0.31%)	303 (0.69%)	1.00E+00
Glycine, serine and threonine metabolism	1 (0.15%)	214 (0.49%)	1.00E+00
2-Oxocarboxylic acid metabolism	1 (0.15%)	231 (0.52%)	1.00E+00
Lysine degradation	1 (0.15%)	244 (0.55%)	1.00E+00
Proteasome	1 (0.15%)	251 (0.57%)	1.00E+00
Pyruvate metabolism	3 (0.46%)	492 (1.12%)	1.00E+00
Aminoacyl-tRNA biosynthesis	1 (0.15%)	254 (0.58%)	1.00E+00
Fructose and mannose metabolism	1 (0.15%)	262 (0.59%)	1.00E+00
Valine, leucine and isoleucine degradation	1 (0.15%)	283 (0.64%)	1.00E+00
Protein processing in endoplasmic reticulum	9 (1.39%)	1145 (2.6%)	1.00E+00
Homologous recombination	1 (0.15%)	320 (0.73%)	1.00E+00
Spliceosome	10 (1.54%)	1322 (3%)	1.00E+00
Phagosome	1 (0.15%)	359 (0.81%)	1.00E+00
Ribosome biogenesis in eukaryotes	1 (0.15%)	368 (0.84%)	1.00E+00
Ubiquitin mediated proteolysis	4 (0.62%)	777 (1.76%)	1.00E+00
Amino sugar and nucleotide sugar metabolism	2 (0.31%)	634 (1.44%)	1.00E+00
Oxidative phosphorylation	2 (0.31%)	697 (1.58%)	1.00E+00
Ribosome	4 (0.62%)	1742 (3.95%)	1.00E+00

Appendix X: Anastasia (2013) Non-Trunking Sago Palm Leaf Tissue Transcriptome Comparison

Anastasia (2013) Gene ID	Anastasia (2013) Gene Annotation	E- value	RNA-seq Alignment Hits	Score	E- value	RNA-seq Gene Annotation	NR Blast Score	NR Blast E- value	NR Blast Annotation
1nt1			-						
1t1a			CL6052.Contig2 _ALL	393	1.00E -109	PREDICTED: probable N-acetyl- gamma-glutamyl- phosphate reductase, chloroplastic			
1t1b			Unigene16129_ ALL	370	1.36E -102	PREDICTED: peptide-N(4)-(N- acetyl-beta- glucosaminy)aspar agine amidase			

Appendix X continued

1t1c			CL6052.Contig2 _ALL	333	6.60E -92	PREDICTED: probable N-acetyl- gamma-glutamyl- phosphate reductase, chloroplastic			
1t1d			CL6052.Contig2 _ALL	402	2.04E -112	PREDICTED: probable N-acetyl- gamma-glutamyl- phosphate reductase, chloroplastic			
1t1e			Unigene4992_A LL	38.3	0.012	NA			
1t1f			Unigene29431_ ALL	207	6.25E -54	NA			

Appendix X continued

1t2			Unigene16129_ ALL	370	1.31E -102	PREDICTED: peptide-N(4)-(N- acetyl-beta- glucosaminyl)aspar agine amidase			
2t1			CL8828.Contig3 _ALL	228	1.19E -60	PREDICTED: proteasome subunit beta type-7-B			
2nt1	Unannotated protein (Populus trichocarpa)	5.00E -09	CL3012.Contig1 _ALL	650	0	hypothetical protein CICLE_v10023137 mg	72.4	8.00E -14	hypothetical protein COCNU_02G00 3200
2nt2b			CL8028.Contig3 _ALL	82.4	3.07E -17	PREDICTED: casein kinase 1-like isoform X2			
2nt3a			Unigene19393_ ALL	77	1.30E -15	PREDICTED: bZIP transcription factor 11-like			

Appendix X continued

2nt3b			CL7690.Contig1 _ALL	35.6	0.002	PREDICTED: 40S ribosomal protein S9-2			
2nt3c			Unigene7926_A LL	91.5	9.92E -20	NA			
3nt1			CL8205.Contig2 _ALL	221	1.73E -58	NA			
4nt2	Calmodulin-like protein (Arabidopsis thaliana)	3.00E -09	Unigene10281_ ALL	578	2.23E -165	PREDICTED: probable calcium- binding protein CML36	84.3	2.00E -17	probable calcium-binding protein CML36
4nt3			CL5035.Contig2 _ALL	452	1.49E -127	PREDICTED: sorbitol dehydrogenase			

Appendix X continued

5t1a			Unigene54499_ ALL	993	0	PREDICTED: uncharacterized protein LOC103707595			
5t1b			Unigene46053_ ALL	1083	0	PREDICTED: uncharacterized protein LOC103707595			
8t1	Cytosolic ascorbate peroxidase (Elaeis guaneensis)	3.00E -94	CL3892.Contig1 _ALL	1406	0	L-ascorbate peroxidase, cytosolic	270	1.00E -87	ascorbate peroxidase
9nt2a			CL797.Contig12 _ALL	535	2.29E -152	PREDICTED: uncharacterized protein LOC103718967			

Appendix X continued

9nt2b			CL797.Contig6_ ALL	545	4.49E -155	PREDICTED: uncharacterized protein LOC103718967			
9nt3			CL7680.Contig2 _ALL	321	4.28E -88	PREDICTED: protein SRC2-like			
9nt4a			CL4672.Contig2 _ALL	221	2.22E -58	PREDICTED: calcium-dependent protein kinase 16- like			
9nt4b			Unigene25153_ ALL	379	2.55E -105	PREDICTED: argininosuccinate synthase, chloroplastic			
9nt6			Unigene19262_ ALL	35.6	0.007	NA			
10nt2			-						

Appendix X continued

11nt1			CL8974.Contig3 _ALL	299	1.46E -81	PREDICTED: uncharacterized protein LOC105034073 isoform X1	54.3	4.00E -06	formamidase-like isoform X1
11t1	Inorganic phosphate transporter 3 (Z. mays)	8.00E -21	CL5041.Contig3 _ALL	640	0	PREDICTED: probable inorganic phosphate transporter 1-8	96.7	3.00E -20	inorganic phosphate transporter 1-4
13t2			-						
14t1			Unigene34680_ ALL	40.1	0.003	NA			
14t3			Unigene42994_ ALL	36.5	0.026	NA			
14nt1a			CL6142.Contig3 _ALL	341	4.61E -94	PTR FAMILY 6.4	58.9	5.00E -08	protein NRT1/ PTR FAMILY 6.4

Appendix X continued

14nt1b	Non-symbiotic hemoglobin 1 (<i>Pyrus communis</i>)	1.00E -06	Unigene46887_ ALL	150	1.08E -36	PREDICTED: non- symbiotic hemoglobin 1	57.8	2.00E -07	hypothetical protein GH714_008539
15t1			Unigene16010_ ALL	350	1.14E -96	hypothetical protein OsI_13666			
15t1b			Unigene16010_ ALL	370	1.33E -102	hypothetical protein OsI_13666			
5nt1	Thioredoxin M3 (<i>O. sativa</i>)	3.00E -25	Unigene25418_ ALL	776	0	PREDICTED: thioredoxin M3, chloroplastic-like isoform X1	134	1.00E -36	thioredoxin M3, chloroplastic
17t1			CL8807.Contig2 _ALL	444	1.68E -125	PREDICTED: signal recognition particle 54 kDa protein, chloroplastic			

Appendix X continued

19t2	CBS domain- containing protein (<i>A. thaliana</i>)	5.00E -21	CL3943.Contig2 _ALL	767	0	PREDICTED: CBS domain-containing protein CBSX3, mitochondrial	107	7.00E -26	CBS domain- containing protein CBSX3, mitochondrial
------	---	--------------	------------------------	-----	---	--	-----	--------------	--

Appendix Y: Siti (2015) Non-Trunking Sago Palm Leaf Tissue Transcriptome Comparison

Siti (2016) Gene ID	Siti (2016) Gene Annotation	E- value	RNA-seq Alignment Hits	Score	E- value	RNA-seq Gene Annotation	NR Blast E- value	NR Blast Score	NR Blast Annotation
NT01 (271bp)	Amino acid binding protein	1.00E -29	CL4727.Contig1 _ALL	427	6.24E -120	ACT domain- containing protein DS12, chloroplastic	1.00E -31	122	ACT domain-containing protein DS12, chloroplastic-like
NT02 (282bp)	Unannotated protein	8.00E -26	Unigene15791_ ALL	425	2.69E -119	BnaCnng12640D	1.27E -30	115	hypothetical protein C4D60_Mb00t14590
NT03 (283bp)	Ribulose biphosphate carboxylase large chain	2.00E -07	Unigene15791_ ALL	434	5.21E -122	BnaCnng12640D	8.16E -44	152	hypothetical protein DEO72_LG11g1671
NT04 (368bp)	Cell wall- associated hydrolase, partial	3.00E -29	Unigene15786_ ALL	581	7.37E -166	uncharacterized protein LOC110281477	1.49E -26	114	hypothetical protein LSAT_6X94260

Appendix Y continued

NT05 (285bp)	DNA J	7.00E -32	CL9442.Contig2 _ALL	421	2.25E -118	PREDICTED: dnaJ protein homolog 2 isoform X2	1.05E -39	146	chaperone protein dnaJ A6 isoform X2
NT06 (351bp)			Unigene53123_ ALL	254	1.09E -67	NA			
NT07 (431bp)	Unannotated protein	7.00E -37	CL10611.Contig 3_ALL	672	0	BnaCnng12640D	9.26E -45	142	hypothetical protein CR513_02839
NT08 (351bp)			Unigene53123_ ALL	277	9.54E -75	NA			
NT09 (350bp)			Unigene53123_ ALL	272	4.04E -73	NA			
NT10 (737bp)	Beta- glucosidase	2.00E -21	Unigene48174_ ALL	517	1.30E -146	NA	7.45E -46	171	beta-glucosidase 22 isoform X4
NT11 (350bp)			Unigene53123_ ALL	272	4.04E -73	NA			
NT12 (737bp)	Beta- glucosidase	2.00E -28	Unigene48174_ ALL	517	1.30E -146	NA	7.45E -46	171	beta-glucosidase 22 isoform X4

Appendix Y continued

NT13 (483bp)	Unannotated protein	3.00E -29	CL859.Contig12 _ALL	794	0	hypothetical protein GLYMA_13G02010 0	2.57E -57	203	hypothetical protein TEA_028899
NT14 (322bp)	Cytochrome P450	2.00E -17	Unigene28180_ ALL	490	8.83E -139	hypothetical protein GLYMA_13G01250 0	1.82E -26	106	hypothetical protein Ae201684_019105
NT15 (425bp)	Cell wall- associated hydrolase	7.00E -23	CL10611.Contig 3_ALL	682	0	BnaCnng12640D	3.37E -46	148	hypothetical protein CR513_02839
NT16 (239bp)	PsbA	6.00E -37	Unigene16208_ ALL	354	3.33E -98	PsbA (chloroplast)	2.47E -36	132	unnamed protein product
T01 (218bp)	Unannotated protein	5.00E -28	Unigene26199_ ALL	300	5.76E -82	PREDICTED: chaperone protein ClpD2, chloroplastic	1.41E -27	115	hypothetical protein SASPL_120434

Appendix Y continued

T02 (264bp)	S-adenosylmethionine synthase	1.00E-41	Unigene38929_ ALL	36.5	0.039	aspartic proteinase A1	8.74E-42	150	S-adenosylmethionine synthase
T03 (280bp)	Unannotated protein	2.00E-15	CL6489.Contig1 _ALL	410	4.70E-115	PREDICTED: uncharacterized protein LOC105044176	3.75E-31	122	uncharacterized protein LOC103715452
T04 (199bp)	Peptidyl-prolyl cistrans isomerase CYP20-2,	5.00E-15	Unigene1213_A LL	274	7.32E-74	PREDICTED: photosynthetic NDH subunit of lumenal location 5, chloroplastic	1.31E-14	77.8	hypothetical protein HYC85_010733
T05 (250bp)	Annexin	2.00E-16	Unigene26027_ ALL	37.4	0.012	PREDICTED: annexin D1-like	4.29E-17	85.5	hypothetical protein CERZMDRAFT_10556 3

Appendix Y continued

T06 (280bp)	Unannotated protein	5.00E -16	CL6489.Contig1 _ALL	410	4.70E -115	PREDICTED: uncharacterized protein LOC105044176	3.75E -31	122	uncharacterized protein LOC103715452
T07 (190bp)	Ribosomal protein	8.00E -24	Unigene6299_A LL	252	2.28E -67	RNA polymerase alpha subunit (plastid)	2.85E -24	98.2	ribosomal protein L14
T08 (176bp)	S- adenosylmet hionine synthase-like protein	3.00E -21	-				5.59E -20	92	S-adenosylmethionine synthetase